Lectures on effective field theory

David B. Kaplan

February 26, 2016

Abstract

Lectures delivered at the ICTP-SAFIR, Sao Paulo, Brasil, February 22-26, 2016.

Contents

1 Effective Quantum Mechanics

1.1 What is an effective field theory?

The uncertainty principle tells us that to probe the physics of short distances we need high momentum. On the one hand this is annoying, since creating high relative momentum in a lab costs a lot of money! On the other hand, it means that we can have predictive theories of particle physics at low energy without having to know everything about physics at short distances. For example, we can discuss precision radiative corrections in the weak interactions without having a grand unified theory or a quantum theory of gravity. The price we pay is that we have a number of parameters in the theory (such as the Higgs and fermion masses and the gauge couplings) which we cannot predict but must simply measure. But this is a lot simpler to deal with than a mess like turbulent fluid flow where the physics at many different distance scales are all entrained together.

The basic idea behind effective field theory (EFT) is the observation that the non-analytic parts of scattering amplitudes are due to intermediate process where physical particles can exist on shell (that is, kinematics are such that internal propagators $1/(p^2 - m^2 + i\epsilon)$ in Feynman diagrams can diverge with $p^2 = m^2$ so that one is sensitive to the $i\epsilon$ and sees cuts in the amplitude due to logarithms, square roots, etc). Therefore if one can construct a quantum field theory that correctly accounts for these light particles, then all the contributions to the amplitude from virtual heavy particles that cannot be physically created at these energies can be Taylor expanded p^2/M^2, where M is the energy of the heavy particle. (By "heavy" I really mean a particle whose energy is too high to create; this might be a heavy particle at rest, but it equally well applies to a pair of light particles with high relative momentum.) However, the power of of this observation is not that one can Taylor expand parts of the scattering amplitude, but that the Taylor expanded amplitude can be computed directly from a quantum field theory (the EFT) which contains *only* light particles, with local interactions between them that encode the small effects arising from virtual heavy particle exchange. Thus the standard model does not contain X gauge bosons from the GUT scale, for example, but can be easily modified to account for the very small effects such particles could have, such as causing the proton to decay, for example.

So in fact, all of our quantum field theories are EFTs; only if there is some day a Theory Of Everything (don't hold your breath) will we be able to get beyond them. So how is a set of lectures on EFT different than a quick course on quantum field theory? Traditionally a quantum field theory course is taught from the point of view that held sway from when it was originated in the late 1920s through the development of nonabelian gauge theories in the early 1970s: one starts with a ϕ^4 theory at tree level, and then computes loops and encounters renormalization; one then introduces Dirac fermions and Yukawa interactions, moves on to QED, and then nonabelian gauge theories. All of these theories have operators of dimension four or less, and it is taught that this is necessary for renormalizability. Discussion of the Fermi theory of weak interactions, with its dimension six four-fermion operator, is relegated to particle physics class. In contrast, EFT incorporates the ideas of Wilson and others that were developed in the early 1970s and completely turned on its head how we think about UV (high energy) physics and renormalization, and how we interpret the results of calculations. While the ϕ^4 interaction used to be considered one only few well-defined (renormalizable) field theories and the Fermi theory of weak interactions was viewed as useful but nonrenormalizable and sick, now the scalar theory is considered sick, while the Fermi theory is a simple prototype of all successful quantum field theories. The new view brings with it its own set of problems, such as an obsession with the fact the our universe appears to be fine-tuned. Is the modern view the last word? Probably not, and I will mention unresolved mysteries at the end of my lectures.

There are three basic uses for effective field theory I will touch on in these lectures:

- Top-down: you know the theory to high energies, but either you do not need all of its complications to arrive at the desired description of low energy physics, or else the full theory is nonperturbative and you cannot compute in it, so you construct an EFT for the light degrees of freedom, constraining their interactions from your knowledge of the symmetries of the more complete theory;
- Bottom-up: you explore small effects from high dimension operators in your low energy EFT to gain cause about what might be going on at shorter distances than you can directly probe;
- Philosophizing: you marvel at how "fine-tuned" our world appears to be, and pondering whether the way our world appears is due to some missing physics, or because we live in a special corner of the universe (the anthropic principle), or whether we live at a dynamical fixed point resulting from cosmic evolution. Such investigations are at the same time both fascinating — and possibly an incredible waste of time!

To begin with I will not discuss effective field theories, however, but effective quantum mechanics. The essential issues of approximating short range interactions with point-like interactions have nothing to do with relativity or many-body physics, and can be seen in entirety in non-relativistic quantum mechanics. I thought I would try this introduction because I feel that the way quantum mechanics and quantum field theory are traditionally taught it looks like they share nothing in common except for mysterious ladder operators, which is of course not true. What this will consist of is a discussion of scattering from delta-function potentials in different dimensions.

1.2 Scattering in 1D

1.2.1 Square well scattering in 1D

We have all solved the problem of scattering in 1D quantum mechanics, from both square barrier potentials and delta-function potentials. Consider scattering of a particle of mass m from an attractive square well potential of width Δ and depth $\frac{\alpha^2}{2m\Delta^2}$,

$$V(x) = \begin{cases} -\frac{\alpha^2}{2m\Delta^2} & 0 \le x \le \Delta \\ 0 & \text{otherwise} \end{cases} . \tag{1}$$

Here α is a dimensionless number that sets the strength of the potential. It is straight forward to compute the reflection and transmission coefficients at energy E (with $\hbar = 1$)

$$R = (1 - T) = \left[\frac{4\kappa^2 k^2 \csc^2(\kappa\Delta)}{(k^2 - \kappa^2)^2} + 1 \right]^{-1} , \tag{2}$$

where

$$k = \sqrt{2mE} \ , \qquad \kappa = \sqrt{k^2 + \frac{\alpha^2}{\Delta^2}} \ . \tag{3}$$

For low k we can expand the reflection coefficient and find

$$R = 1 - \frac{4}{\alpha^2 \sin^2 \alpha} \Delta^2 k^2 + O(\Delta^4 k^4) \tag{4}$$

Note that $R \to 1$ as $k \to 0$, meaning that the potential has a huge effect at low enough energy, no matter how weak...we can say the interaction is very *relevant* at low energy.

1.2.2 Relevant δ-function scattering in 1D

Now consider scattering off a δ-function potential in 1D,

$$V(x) = -\frac{g}{2m\Delta} \delta(x) \ , \tag{5}$$

where the length scale Δ was included in order to make the coupling g dimensionless. Again one can compute the reflection coefficient and find

$$R = (1 - T) = \left[1 + \frac{4k^2\Delta^2}{g^2}\right]^{-1} = 1 - \frac{4k^2\Delta^2}{g^2} + O(k^4) \ . \tag{6}$$

By comparing the above expression to eq. (4) we see that at low momentum the δ function gives the same reflection coefficient to up to $O(k^4)$ as the square well, provided we set

$$g = \alpha \sin \alpha \ . \tag{7}$$

In the EFT business, the above equation is called a "matching condition"; this matching condition is shown in Fig. 1, and interpreting the structure in this figure – in particular the sign changes for g – is one of the problems at the end of the lecture. For small α the matching condition is simply $g \simeq \alpha^2$.

1.3 Scattering in 3D

Now let's see what happens if we try the same thing in 3D (three spatial dimensions), choosing the strength of a δ-function potential to mimic low energy scattering off a square well potential. Why this fixation with δ-function potentials? They are not particularly special in non-relativistic quantum mechanics, but in a relativistic field theory they are the *only* instantaneous potential which can be Lorentz invariant. That is why we always formulate quantum field theories as interactions between particles only when they are at the same point in spacetime. All the issues of renormalization in QFT arise from the singular nature of these δ-function interactions. So I am focussing on δ-function potentials in quantum mechanics in order to illustrate what is going on in the relativistic QFT.

First, a quick review of a few essentials of scattering theory in 3D, focussing only on s-wave scattering.

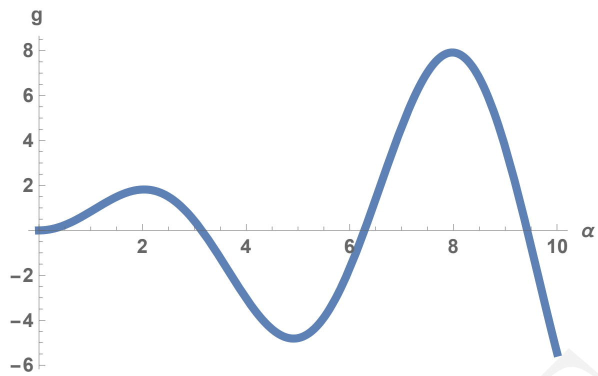

Figure 1: *The matching condition in 1D: the appropriate value of g in the effective theory for a given α in the full theory.*

A scattering solution for a particle of mass m in a finite range potential must have the asymptotic form for large $|\mathbf{r}|$

$$\psi \xrightarrow{r\to\infty} e^{ikz} + \frac{f(\theta)}{r} e^{ikr} \ . \tag{8}$$

representing an incoming plane wave in the z direction, and an outgoing scattered spherical wave. The quantity f is the scattering amplitude, depending both on scattering angle θ and incoming momentum k, and $|f|^2$ encodes the probability for scattering; in particular, the differential cross section is simply

$$\frac{d\sigma}{d\theta} = |f(\theta)|^2 \ . \tag{9}$$

For scattering off a spherically symmetric potential, both $f(\theta)$ and $e^{ikz} = e^{ikr\cos\theta}$ can be expanded in Legendre polynomials ("partial wave expansion"); I will only be interested in s-wave scattering (angle independent) and therefore will replace $f(\theta)$ simply by f — independent of angle, but still a function of k. For the plane wave we can average over θ when only considering s-wave scattering, replacing

$$e^{ikz} = e^{ikr\cos\theta} \xrightarrow[s-\text{wave}]{} \frac{1}{2}\int_0^\pi d\theta\, \sin\theta\, e^{ikr\cos\theta} = j_0(kr) \ . \tag{10}$$

Here $j_0(z)$ is a regular spherical Bessel function; we will also meet its irregular partner $n_0(z)$, where where

$$j_0(z) = \frac{\sin z}{z} \ , \qquad n_0(z) = -\frac{\cos z}{z} \ . \tag{11}$$

These functions are the s-wave solutions to the free 3D Schrödinger equation $z = kr$.

So we are interested in a solution to the Schrödinger equation with asymptotic behavior

$$\psi \xrightarrow[s-\text{wave}]{r\to\infty} j_0(kr) + \frac{f}{r} e^{ikr} = j_0(kr) + kf\left(ij_0(kr) - n_0(kr)\right) \qquad (s\text{-wave}) \tag{12}$$

Since outside the the range of the potential ψ is an exact s-wave solution to the free Schrödinger equation, and the most general solutions to the free radial Schrödinger equation are the spherical Bessel functions $j_0(kr)$, $n_0(kr)$, the asymptotic form for ψ can also be written as

$$\psi \xrightarrow{r\to\infty} A\left(\cos\delta\, j_0(kr) - \sin\delta\, n_0(kr)\right) \ . \tag{13}$$

where A and δ are real constants. The angle δ is called the phase shift, and if there was no potential, boundary conditions at $r = 0$ would require $\delta = 0$...so nonzero δ is indicative of scattering. Relating these two expressions eq. (12) and eq. (13) we find

$$f = \frac{1}{k\cot\delta - ik} \ . \tag{14}$$

So solving for the phase shift δ is equivalent to solving for the scattering amplitude f, using the formula above.

The quantity $k\cot\delta$ is interesting, since one can show that for a finite range potential it must be analytic in the energy, and so has a Taylor expansion in k involving only even powers of k , called "the effective range expansion":

$$k\cot\delta = -\frac{1}{a} + \frac{1}{2}r_0 k^2 + O(k^4) \ . \tag{15}$$

The parameters have names: a is the scattering length and r_0 is the effective range; these terms dominate low energy (low k) scattering. Proving the existence of the effective range expansion is somewhat involved and I refer you to a quantum mechanics text; there is a low-brow proof due to Bethe and a high-brow one due to Schwinger.

And the last part of this lightning review of scattering: if we have two particles of mass M scattering off each other it is often convenient to use Feynman diagrams to describe the scattering amplitude; I denote the Feynman amplitude – the sum of all diagrams – as $i\mathcal{A}$. The relation between $\mathcal{A}$ and f is

$$\mathcal{A} = \frac{4\pi}{M} f \ , \tag{16}$$

where f is the scattering amplitude for a single particle of reduced mass $m = M/2$ in the inter-particle potential. This proportionality is another result that can be priced together from quantum mechanics books, which I won't derive.

1.3.1 Square well scattering in 3D

We consider s-wave scattering off an attractive well in 3D,

$$V = \begin{cases} -\frac{\alpha^2}{m\Delta^2} & r < \Delta \\ 0 & r > \Delta \ . \end{cases} \tag{17}$$

We have for the wave functions for the two regions $r < \Delta$, $r > \Delta$ are expressed in terms of spherical Bessel functions as

$$\psi_<(r) = j_0(\kappa r) \ , \qquad \psi_>(r) = A\left[\cos\delta\, j_0(kr) - \sin\delta\, n_0(kr)\right] \tag{18}$$

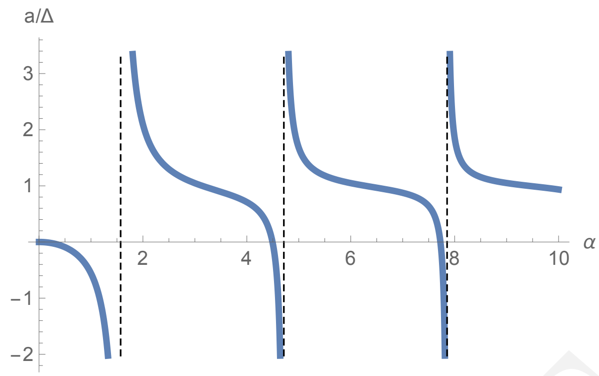

Figure 2: *a/Δ vs. the 3D potential well depth parameter α, from eq. (21).*

where $\kappa = \sqrt{k^2 + \alpha^2/\Delta^2}$ as in eq. (3) and δ is the s-wave phase shift. Equating ψ and ψ' at the edge of the potential at $r = \Delta$ allows us to solve for δ in terms of k, α, Δ, with the result

$$k \cot \delta = \frac{k(k \sin \kappa\Delta + \kappa \cot k\Delta \cos \kappa\Delta)}{k \cot k\Delta \sin \kappa\Delta - \kappa \cos \kappa\Delta} \ . \tag{19}$$

With a little help from Mathematica we can expand this in powers of k^2 and find

$$k \cot \delta = \frac{1}{\Delta}\left(\frac{\tan\alpha}{\alpha} - 1\right)^{-1} + O(k^2) \tag{20}$$

where on comparing with eq. (15) we can read off the scattering length from the k^2 expansion,

$$a = -\Delta\left(\frac{\tan\alpha}{\alpha} - 1\right) \ , \tag{21}$$

a relation shown in Fig. 2. The singularities one finds for the scattering length as the strength of the potential α increases correspond to the critical values $\alpha_c = (2n+1)\pi/2$, $n = 0, 1, 2\ldots$ where a new bound state appears.

1.3.2 Irrelevant δ-function scattering in 3D

Now we look at reproducing the above scattering length from scattering in 3D off a delta function potential. At first look this seems hopeless: note that the result for a square well of width Δ and coupling $\alpha = O(1)$ gives a scattering length that is $a = O(\Delta)$; this is to be expected since a is a length, and Δ is the natural length scale in the problem. Therefore if you extrapolate to a potential of zero width (a δ function) you would conclude that the scattering length would go to zero, and the scattering amplitude would vanish for low k. This is an example of an *irrelevant* interaction.

On second look the situation is even worse: since $-\delta^3(\mathbf{r})$ scales as $-1/r^3$ while the kinetic $-\nabla^2$ term in the Schrödinger equation only scales as $1/r^2$ you can see that the system does not have a finite energy ground state. For example if you performed a variational calculation, you could lower the energy without bound by scaling the wave function to smaller and smaller extent. Therefore the definition of a δ-function has to be modified in 3D – this is the essence of renormalization.

These two features go hand in hand: typically singular interactions are "irrelevant" and at the same time require renormalization. We can sometimes turn an irrelevant interaction into a relevant one by fixing a certain renormalization condition which forces a fine tuning of the coupling to a critical value, and that is the case here. For example, consider defining the δ-function as the $\rho \to 0$ limit of a square well of width ρ and depth $V_0 = \bar{\alpha}^2(\rho)/(m\rho^2)$, while adjusting the coupling strength $\bar{\alpha}(\rho)$ to keep the scattering length fixed to the desired value of a given in eq. (21). We find

$$a = \rho\left(1 - \frac{\tan\bar{\alpha}(\rho)}{\bar{\alpha}(\rho)}\right) \tag{22}$$

as $\rho \to 0$. There are an infinite number of solutions, corresponding to $\alpha \simeq \alpha_c = (2n+1)\pi/2$, $n = 0, 1, 2\ldots$, and the $n = 0$ possibility is

$$\bar{\alpha}(\rho) \xrightarrow{\rho\to 0} \frac{\pi}{2} + \frac{2\rho}{\pi a} + O(\rho^2)\ . \tag{23}$$

in other words, we have to tune this vanishingly thin square well to have a single bound state right near threshold ($\alpha \simeq \pi/2$). However, note that while naively you might think a potential $-g\delta^3(\mathbf{r})$ would be approximated by a square well of depth $V_0 \propto 1/\rho^3$ as $\rho \to 0$, but we see that instead we get $V_0 \propto 1/\rho^2$. This is sort of like using a potential $-r\delta^3(\mathbf{r})$ instead of $-\delta^3(\mathbf{r})$.

We have struck a delicate balance: A naive δ function potential is too strong and singular to have a ground state; a typical square well of depth $\alpha^2/m\rho^2$ becomes irrelevant for fixed α in the $\rho \to 0$ limit; but a strongly coupled potential of form $V \simeq -(\pi/2)^2/m\rho^2$ can lead to a relevant interaction so long as we tune α its critical value $\alpha_c = \pi/2$ in precisely the right way as we take $\rho \to 0$.

This may all seem more familiar to you if I to use field theory methods and renormalization. Consider two colliding particles of mass M in three spatial dimensions with a δ-function interaction; this is identical to the problem of potential scattering when we identify m with the reduced mass of the two particle system,

$$m = \frac{M}{2}\ . \tag{24}$$

We introduce the field ψ for the scattering particles (assuming they are spinless bosons) and the Lagrange density

$$\mathcal{L} = \psi^\dagger\left(i\partial_t + \frac{\nabla^2}{2M}\right)\psi - \frac{C_0}{4}\left(\psi^\dagger\psi\right)^2\ . \tag{25}$$

Here $C_0 > 0$ implies a repulsive interaction. As in a relativistic field theory, ψ annihilates particles and $\psi^\dagger$ creates them; unlike in a relativistic field theory, however, there are no anti-particles.

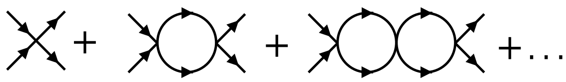

Figure 3: *The sum of Feynman diagrams giving the exact scattering amplitude for two particles interaction via a δ-function potential.*

The kinetic term gives rise to the free propagator

$$G(E, \mathbf{p}) = \frac{i}{E - \mathbf{p}^2/(2M) + i\epsilon} \ , \tag{26}$$

while the interaction term gives the vertex $-iC_0$. The total Feynman amplitude for two particles then is the sum of diagrams in Fig. 3, which is the geometric series

$$i\mathcal{A} = -iC_0 \left[1 + (C_0 B(E)) + (C_0 B(E))^2 + \dots\right] = \frac{i}{-\frac{1}{C_0} + B(E)} \ , \tag{27}$$

where B is the 1-loop diagram, which in the center of momentum frame (where the incoming particles have momenta $\pm\mathbf{p}$ and energy $E/2 = \mathbf{k}^2/(2M)$) is given by

$$B(E) = -i \int \frac{d^4q}{(2\pi)^4} \frac{i}{\left(\frac{E}{2} + q_0 - \frac{\mathbf{q}^2}{2M} + i\epsilon\right)} \frac{i}{\left(\frac{E}{2} - q_0 - \frac{\mathbf{q}^2}{2M} + i\epsilon\right)} = \int \frac{d^3q}{(2\pi)^3} \frac{1}{E - \frac{\mathbf{q}^2}{M} + i\epsilon} \ . \tag{28}$$

The B integral is linearly divergent and so I will regulate it with a momentum cutoff and renormalize the coupling C_0:

$$\begin{aligned} B(E, \Lambda) &= \int^\Lambda \frac{d^3q}{(2\pi)^3} \frac{1}{E - \frac{\mathbf{q}^2}{M} + i\epsilon} \\ &= -\frac{M\left(\Lambda - \sqrt{-EM - i\epsilon}\tan^{-1}\left(\frac{\Lambda}{\sqrt{-EM - i\epsilon}}\right)\right)}{2\pi^2} \\ &= -\frac{M\Lambda}{2\pi^2} + \frac{M}{4\pi}\sqrt{-ME - i\epsilon} + O\left(\frac{1}{\Lambda}\right) \\ &= -\frac{M\Lambda}{2\pi^2} - i\frac{Mk}{4\pi} + O\left(\frac{1}{\Lambda}\right) \ . \end{aligned} \tag{29}$$

Thus from eq. (27) we get the Feynman amplitude

$$\mathcal{A} = \frac{1}{-\frac{1}{C_0} + B(E)} = \frac{1}{-\frac{1}{C_0} - \frac{M\Lambda}{2\pi^2} - i\frac{Mk}{4\pi}} = \frac{4\pi}{M} \frac{1}{\left(-\frac{4\pi}{M\overline{C}_0} - ik\right)} \tag{30}$$

where

$$\frac{1}{\overline{C}_0} = \frac{1}{C_0} - \frac{M\Lambda}{2\pi^2} \ . \tag{31}$$

$\overline{C}_0$ is our renormalized coupling, C_0 is our bare coupling, and Λ is our UV cutoff. Since in in 3D we have (eq. (14), eq. (16))

$$\mathcal{A} = \frac{4\pi}{M}\frac{1}{k\cot\delta - ik} \ , \qquad k\cot\delta = -\frac{1}{a} + \frac{1}{2}r_0 k^2 + \dots \tag{32}$$

we see that this theory relates $\overline{C}_0$ to the scattering length as

$$\overline{C}_0 = \frac{4\pi a}{M} \ . \tag{33}$$

Therefore we can reproduce square well scattering length eq. (21) by taking

$$\overline{C}_0 = -\frac{4\pi\Delta}{M}\left(\frac{\tan\alpha}{\alpha} - 1\right) \ . \tag{34}$$

With our simple EFT we can reproduce the scattering length of the square well problem, but not the next term in the low k expansion, the effective range. There is a one-to-one correspondence between the number of terms we can fit in the effective range expansion and the number of operators we include in the EFT; to account for the effective range we would have to include a new contact interaction involving two derivatives and match its coefficient.

What have we accomplished? We have shown that one can reproduce low energy scattering from a finite range potential in 3D with a δ-function interaction, with errors of $O(k^2\Delta^2)$ with the caveat that renormalization is necessary if we want to make sense of the theory.

However there is second important and subtle lesson: We can view eq. (31) plus eq. (33) to imply a fine tuning of the inverse bare coupling $1/C_0$ coupling as $\Lambda \to \infty$: $M\Lambda C_0/(2\pi^2)$ must be tuned to $1 + O(1/a\Lambda)$ as $\Lambda \to \infty$. This is the same lesson we learned looking at square wells: if C_0 didn't vanish at least linearly with the cutoff, the interaction would be too strong to makes sense; while if ΛC_0 went to zero or a small constant, the interaction would be irrelevant. Only if ΛC_0 is fine-tuned to a critical value can we obtain nontrivial scattering at low k.

1.4 Scattering in 2D

1.4.1 Square well scattering in 2D

Finally, let's look at the intermediary case of scattering in two spatial dimensions, where we take the same potential as in eq. (17). This is not just a tour of special functions — something interesting happens! The analogue of eq. (35) for the two dimensional square well problem is

$$\psi_<(r) = J_0(\kappa r) \ , \qquad \psi_>(r) = A\left[\cos\delta\, J_0(kr) - \sin\delta\, Y_0(kr)\right] \tag{35}$$

where κ is given in eq. (3) and J, Y are the regular and irregular Bessel functions. Equating ψ and ψ' at the boundary $r = \Delta$ gives[1]

$$\begin{aligned} \cot\delta &= \frac{kJ_0(\Delta\kappa)Y_1(\Delta k) - \kappa J_1(\Delta\kappa)Y_0(\Delta k)}{kJ_0(\Delta\kappa)J_1(\Delta k) - \kappa J_1(\Delta\kappa)J_0(k\Delta)} \\ &= \frac{2\left(\frac{J_0(\alpha)}{\alpha J_1(\alpha)} + \log\left(\frac{\Delta k}{2}\right) + \gamma_E\right)}{\pi} + O\left(k^2\right) \end{aligned} \tag{36}$$

[1] In the following expressions $\gamma_E = 0.577\ldots$ is the Euler constant.

This result looks very odd because of the logarithm that depends on k! The interesting feature of this expression is not that $\cot\delta(k) \to -\infty$ for $k \to 0$: that just means that the phase shift vanishes at low k. What is curious is that for our attractive potential, the function $J_0(\alpha)/(\alpha J_1(\alpha))$ is strictly positive, and therefore $\cot\delta$ changes sign at a special value for k,

$$k = \Lambda \simeq \frac{2e^{-\frac{J_0(\alpha)}{\alpha J_1(\alpha)}-\gamma_E}}{\Delta} \ , \tag{37}$$

where the scale Λ is *exponentially* lower than our fundamental scale Δ for weak coupling, since then $J_0(\alpha)/\alpha J_1(\alpha) \sim 2/\alpha^2 \gg 1$. This is evidence in the scattering amplitude for a bound state of size $\sim 1/\Lambda$...exponentially larger than the size of the potential!

On the other hand, if the interaction is repulsive, the $J_0(\alpha)/\alpha J_1(\alpha)$ factor is replaced by $-I_0(\alpha)/\alpha I_1(\alpha) < 0$, I_n being one of the other Bessel functions, and the numerator in eq. (36) is always negative, and there is no bound state.

1.4.2 Marginal δ-function scattering in 2D & asymptotic freedom

If we now look at the Schrödinger equation with a δ-function to mock up the effects of the square well for low k we find something funny: the equation is scale invariant. What that means is that the existence of any solution $\psi(r)$ to the equation

$$\left[-\frac{1}{2m}\nabla^2 + \frac{g}{m}\delta^2(\mathbf{r})\right]\psi(r) = E\psi(r) \tag{38}$$

implies a continuous family of solutions $\psi_\lambda(r) = \psi(\lambda r)$ – the same functional form except scaled smaller by a factor of λ – with energy $E_\lambda = \lambda^2 E$. Thus it seems that all possible energy eigenvalues with the same sign as E exist and there are no discrete eigenstates...which is OK if only positive energy scattering solutions exist, the case for a repulsive interaction — but not OK if there are bound states: it appears that if there is any one negative energy state, then there is an unbounded continuum of negative energy states and no ground state. The problem is that ∇^2 and $\delta^2(\mathbf{r})$ have the same dimension, 1/length^2, and so there is no inherent scale to the left side of the equation. Since the scaling property of $\delta^D(\mathbf{r})$ changes with dimension D, while the scaling property of ∇^2 does not, $D = 2$ is special.

Since the δ-function interaction seems to be scale invariant, we say that it is neither relevant (dominating IR physics, as in 1D) nor irrelevant (unimportant to IR physics, as in 3D) but apparently of equal important at all scales, which we call *marginal*. However, we know that (i) the δ function description appears to be sick, and (ii) from our exact analysis of the square well that the IR description of the full theory is not really scale invariant, due to the logarithm. Therefore it is a reasonable guess that our analysis of the δ-function is incorrect due to its singularity, and that we are going to have to be more careful, and renormalize.

We can repeat the Feynman diagram approach we used in 3D, only now in 2D. Now the loop integral in eq. (29) is required in $d = 3$ spacetime dimensions instead of $d = 4$. It is still divergent, but now only log divergent, not linearly divergent. It still needs regularization, but this time instead of using a momentum cutoff I will use dimensional regularization, to make it look even more like conventional QFT calculations. Therefore we keep the number of spacetime dimensions d arbitrary in computing the integral, and subsequently expand about $d = 3$ (for scattering in $D = 2$ spatial dimensions)[2]. We take for our action

$$S = \int dt \int d^{d-1}x \left[\psi^\dagger \left(i\partial_t + \frac{\nabla^2}{2M} \right) \psi - \mu^{d-3} \frac{C_0}{4} \left(\psi^\dagger \psi \right)^2 \right] . \tag{39}$$

where the renormalization scale μ was introduced to keep C_0 dimensionless (see problem). Then the Feynman rules are the same as in the previous case, except for the factor of μ^{d-3} at the vertices, and we find

$$\begin{aligned} B(E) &= \mu^{3-d} \int \frac{d^{d-1}q}{(2\pi)^{d-1}} \frac{1}{E - \frac{\mathbf{q}^2}{M} + i\epsilon} \\ &= -M \left(-ME - i\epsilon \right)^{\frac{d-3}{2}} \Gamma\left(\frac{3-d}{2} \right) \frac{\mu^{3-d}}{(4\pi)^{(d-1)/2}} \\ &\xrightarrow{d\to 3} \frac{M}{2\pi} \frac{1}{(d-3)} + \frac{M}{4\pi} \left(\gamma_E - \ln 4\pi + \ln \frac{k^2}{\mu^2} - i\pi \right) + O(d-3) \end{aligned} \tag{40}$$

where $k = \sqrt{ME}$ is the magnitude of the momentum of each incoming particle in the center of momentum frame, and the scattering amplitude is therefore

$$\mathcal{A} = \frac{1}{-\frac{1}{C_0} + B(E)} = \left[-\frac{1}{C_0} + \frac{M}{2\pi} \frac{1}{(d-3)} + \frac{M}{4\pi} \left(\gamma_E - \ln 4\pi + \ln \frac{k^2}{\mu^2} - i\pi \right) \right]^{-1} \tag{41}$$

At this point it is convenient to define the dimensionless coupling constant g:

$$C_0 \equiv g \frac{4\pi}{M} . \tag{42}$$

Given the definition of our Lagrangian, $g > 0$ corresponds to a repulsive potential, and $g < 0$ is attractive. so that the amplitude is

$$\mathcal{A} = \frac{4\pi}{M} \left[-\frac{1}{g} - \frac{2}{(d-3)} + \gamma_E - \ln 4\pi + \ln \frac{k^2}{\mu^2} - i\pi \right]^{-1} \tag{43}$$

To make sense of this at $d = 3$ we have to renormalize g with the definition:

$$\frac{1}{g} = \frac{1}{\bar{g}(\mu)} + \frac{2}{(d-3)} + \gamma_E - \ln 4\pi , \tag{44}$$

where $\bar{g}(\mu)$ is the renormalized running coupling constant, and so the amplitude is given by

$$\mathcal{A} = \frac{4\pi}{M} \left[-\frac{1}{\bar{g}(\mu)} + \ln \frac{k^2}{\mu^2} - i\pi \right]^{-1} \tag{45}$$

[2]If you are curious why I did not use dimensional regularization for the $D = 3$ case: dim reg ignores power divergences, and so when computing graphs with power law divergences using dim reg you do not explicitly notice that you are fine-tuning the theory. This happens in the standard model with the quadratic divergence of the Higgs mass[2]...every few years someone publishes a preprint saying there is no fine-tuning problem since one can compute diagrams using dim reg, where there is no quadratic divergence, which is silly. I used a momentum cutoff in the previous section so we could see the fine-tuning of C_0.

Since this must be independent of μ it follows that

$$\mu \frac{d}{d\mu}\left(-\frac{1}{\bar{g}(\mu)} + \ln\frac{k^2}{\mu^2}\right) = 0 \tag{46}$$

or equivalently,

$$\mu \frac{d\bar{g}(\mu)}{d\mu} = \beta(\bar{g}) \ , \qquad \beta(\bar{g}) = 2\bar{g}(\mu)^2 \ . \tag{47}$$

If we specify the renormalization condition $\bar{g}(\mu_0) \equiv \bar{g}_0$, then the solution to this renormalization group equation is

$$\bar{g}(\mu) = \frac{1}{\frac{1}{\bar{g}_0} + 2\ln\frac{\mu_0}{\mu}} \ . \tag{48}$$

Note that this solution $\bar{g}(\mu)$ blows up at

$$\mu = \mu_0 e^{1/(2\bar{g}_0)} \equiv \Lambda \ . \tag{49}$$

For $0 < g_0 \ll 1$ (weak repulsive interaction) we have $\Lambda \gg \mu_0$, while for $-1 \ll g_0 < 0$ (weak attractive interaction) Λ is an infrared scale, $\Lambda \ll \mu_0$. If we set $\mu_0 = \Lambda$ in eq. (48) and $g_0 = \infty$, we find

$$\bar{g}(\mu) = \frac{1}{\ln\frac{\Lambda^2}{\mu^2}} \ , \tag{50}$$

and the amplitude as

$$\mathcal{A} = \frac{4\pi}{M}\frac{1}{\ln\frac{k^2}{\Lambda^2} + i\pi} \ , \tag{51}$$

or equivalently,

$$\cot\delta = -\frac{1}{\pi}\ln\frac{k^2}{\Lambda^2} \ . \tag{52}$$

Now just have to specify Λ instead of $\bar{g}_0$ to define the theory ("dimensional transmutation").

Finally, we can match this δ-function scattering amplitude to the square well scattering amplitude at low k by equating eq. (52) with our expression eq. (36), yielding the matching condition

$$\ln\frac{k^2}{\Lambda^2} = 2\left(\frac{J_0(\alpha)}{\alpha J_1(\alpha)} + \log\left(\frac{\Delta k}{2}\right) + \gamma_E\right) \tag{53}$$

from which the k dependence drops out and we arrive at and expression for Λ in terms of the coupling constant α of the square well:

$$\Lambda = \frac{2e^{-\frac{J_0(\alpha)}{\alpha J_1(\alpha)} - \gamma}}{\Delta} \tag{54}$$

If $g_0 < 0$ (attractive interaction) the scale Λ is in the IR ($\mu \ll \mu_0$ if g_0 is moderately small) and we say that the interaction is asymptotically free, with Λ playing the same role as Λ_{QCD} in the Standard Model – except that here we are not using perturbation theory, the β-function is exact, and we can take $\mu < \Lambda$ and watch $\bar{g}(\mu)$ change from $+\infty$ to $-\infty$ as we scale through a bound state. If instead $g_0 > 0$ (repulsive interaction) then Λ is in the UV, we say the theory is asymptotically unfree, and Λ is similar to the Landau pole in QED. So we see that while the Schrödinger equation *appeared* to have a scale invariance and therefore no discrete states, in reality when one makes sense of the singular interaction, a scale Λ seeks into the theory, and it is no longer scale invariant.

1.5 Lessons learned

We have learned the following by studying scattering from a finite range potential at low k in various dimensions:

- A contact interaction (δ-function) is more irrelevant in higher dimensions;
- marginal interactions are characterized naive scale invariance, and by logarithms of the energy and running couplings when renormalization is accounted for; they can either look like relevant or irrelevant interactions depending on whether the running is asymptotically free or not; and in either case they are characterized by a mass scale Λ exponentially far away from the fundamental length scale of the interaction, Δ.
- Irrelevant interactions and marginal interactions typically require renormalization; an irrelevant interaction can sometimes be made relevant if its coefficient is tuned to a critical value.

All of these lessons will be pertinent in relativistic quantum field theory as well.

1.6 Problems for lecture I

I.1) Explain Fig. 1: how do you interpret those oscillations? Similarly, what about the cycles in Fig. 2?

I.2) Consider dimensional analysis for the non-relativistic action eq. (39). Take momenta p to have dimension 1 by definition in any spacetime dimension d; with the uncertainty principle $[x,p] = i\hbar$ and $\hbar = 1$ we then must assign dimension -1 to spatial coordinate x. Write this as

$$[p] = [\partial_x] = 1 \ , \qquad [x] = -1 \ . \tag{55}$$

Unlike in the relativistic theory we can treat M as a dimensionless parameter under this scaling law. If we do that, use eq. (39) with the factor of μ omitted to figure out the scaling dimensions

$$[t] \ , \quad [\partial_t] \ , \quad [\psi] \ , \quad [C_0] \tag{56}$$

for arbitrary d, using that fact that the action S must be dimensionless (after all, in a path integral we exponentiate $S/\hbar$, which would make no sense if that was a dimensional quantity). What is special about $[C_0]$ at $d = 3$? Confirm that including the factor of μ^{d-3}, where μ has scaling dimension 1 ($[\mu] = 1$) allows C_0 to maintain its $d = 3$ scaling dimension for any d.

I.3) In eq. (52) the distinction between attractive and repulsive interactions seems to have been completely lost since that equation holds for both cases! By looking at how the 2D matching works in describing the square well by a δ-function, explain how the low energy theory described by eq. (52) behaves differently when the square well scattering is attractive versus repulsive. Is there physical significance to the scale Λ in the effective theory for an attractive interaction? What about for a repulsive interaction?

2 EFT at tree level

To construct a relativistic effective field theory valid up to some scale Λ, we will take for our action made out of all light fields (those corresponding to particles with masses or energies much less that Λ) including all possible local operators consistent with the underlying symmetries that we think govern the world. All UV physics that we are not including explicitly is encoded in the coefficients of these operators, in the same way we saw in the previous section that a contact interaction (δ-function potential) was able to reproduce the scattering length for scattering off a square well if its coefficient was chosen appropriately (we "matched" it to the UV physics). However, in the previous examples we just tried matching the scattering lengths; we could have tried to also reproduce $O(k^2\Delta^2)$ effects, and so on, but to do so would have required introducing more and more singular contributions to the potential in the effective theory, such as $\nabla^2\delta(\mathbf{r})$, $\nabla^4\delta(\mathbf{r})$, and so on. Going to all orders in k^2 would require an infinite number of such terms, and the same is true for a relativistic EFT. Such a theory is not "renormalizable" in the historical sense: there is typically no finite set of coupling constants that can be renormalized with a finite pieces of experimental data to render the theory finite. Instead there are an infinite number of counterterms need to make the theory finite, and therefore an infinite number of experimental data needed to fix the finite parts of the counterterms. Such a theory would be unless there existed some sort of expansion that let us deal with only a finite set of operators at each order in that expansion.

Wilson provided such an expansion. The first thing to accept is that the EFT has an intrinsic, finite UV cutoff Λ. This scale is typically the mass of the lightest particles *omitted* from the theory. For example, in the Fermi theory of the weak interactions, $\Lambda = M_W$. With a cutoff in place, all radiative corrections in the theory are finite, even if they are proportional to powers or logarithms of Λ. The useful expansion then is a momentum expansion, in powers of k/Λ, where k is the external momentum in some physical process of interest, such as a particle decay, two particle scattering, two particle annihilation, etc. This momentum expansion is the key tool that makes EFTs useful. To understand how this works, we need to develop the concept of operator dimension. In this lecture we will only consider the EFT at tree level.

2.1 Scaling in a relativistic EFT

As a prototypical example of an EFT, consider the Lagrangian (in four dimensional Euclidean spacetime, after a Wick rotation to imaginary time) for relativistic scalar field with a $\phi \to -\phi$ symmetry:

$$\mathcal{L}_E = \frac{1}{2}(\partial\phi)^2 + \frac{1}{2}m^2\phi^2 + \frac{\lambda}{4!}\phi^4 + \sum_{n=1}^{\infty}\left(\frac{c_n}{\Lambda^{2n}}\phi^{4+2n} + \frac{d_n}{\Lambda^{2n}}(\partial\phi)^2\phi^{2n} + \dots\right) \tag{57}$$

We are setting $\hbar = c = 1$ so that momenta have dimension of mass, and spacetime coordinates have dimension of inverse mass. I indicate this as

$$[p] = 1\ , \qquad [x] = [t] = -1\ , \qquad [\partial_x] = [\partial_t] = 1\ . \tag{58}$$

Since the action is dimensionless, then — in $d = 4$ spacetime dimensions — from the kinetic term for ϕ we see that ϕ has dimension of mass:

$$[\phi] = 1 \ . \tag{59}$$

That means that the operator ϕ^6 is dimension 6, and the contribution to the action $\int d^4x\, \phi^6$ has dimension 2, and so its coupling constant must have dimension -2, or $1/\text{mass}^2$. The operator $\phi^2(\partial^2\phi)^2$ is dimension 8 and must have a coefficient which is dimension -4, or $1/\text{mass}^4$. In eq. (57) I have introduced the cutoff scale Λ explicitly into the Lagrangian in such a way as to make the the couplings λ, c_n and d_n all dimensionless, with no loss of generality. I will assume here that $\lambda \ll 1$, $c_n \ll 1$ and $d_n \ll 1$ so that a perturbative expansions in these couplings is reasonable.

You might ask why we do things this way — why not rescale the ϕ^6 operator to have coefficient 1 instead of the kinetic term, and declare ϕ to have dimension 2/3? The reason why is because the kinetic term is more important and determines the size of quantum fluctuations for a relativistic excitation. To see this, consider the path integral

$$\int D\phi\, e^{-S_E} \ , \qquad S_E = \int d^4x\, \mathcal{L}_E \ . \tag{60}$$

Now consider a particular field configuration contributing to this path integral that looks like the "wavelet" pictured in Fig. 4, with wavenumber $|k_\mu| \sim k$, localized to a spacetime volume of size L^4, where $L \simeq 2\pi/k$, and with amplitude ϕ_k. Derivatives acting on such a configuration give powers of k, while spacetime integration gives a factor of $L^4 \simeq (4\pi/k)^4$. With this configuration, the Euclidean action is given by

$$\begin{aligned} S_E &\simeq \left(\frac{2\pi}{k}\right)^4 \left[\frac{k^2\phi_k^2}{2} + \frac{1}{2}m^2\phi_k^2 + \frac{\lambda}{4!}{\phi_k}^4 + \sum_{n=1}^{\infty}\left(\frac{c_n}{\Lambda^{2n}}{\phi_k}^{4+2n} + \frac{d_n k^2}{\Lambda^{2n}}{\phi_k}^{2+2n} + \dots\right)\right] \\ &= (2\pi)^4\left[\frac{\widehat{\phi}_k^2}{2} + \frac{1}{2}\frac{m^2}{k^2}\widehat{\phi}_k^2 + \frac{\lambda}{4!}\widehat{\phi}_k^4 + \sum_n\left(c_n\left(\frac{k^2}{\Lambda^2}\right)^n\widehat{\phi}_k^{4+2n} + d_n\left(\frac{k^2}{\Lambda^2}\right)^n\widehat{\phi}_k^{2+2n} + \dots\right)\right] \ , \end{aligned} \tag{61}$$

where in the second line I have rescaled the amplitude by k,

$$\widehat{\phi}_k \equiv \phi_k/k \ . \tag{62}$$

In the above expression, the factor of $\left(\frac{2\pi}{k}\right)^4$ in front is the spacetime volume occupied by our wavelet and comes from $\int d^4x$, while for every operator I have substituted $\partial \to k$ and $\phi \to k\widehat{\phi}_k$. Now for the path integral, consider ordinary integration over the amplitude $\widehat{\phi}_k$ for this particular mode:

$$\int d\widehat{\phi}_k\, e^{-S_E} \ . \tag{63}$$

The integral is dominated by those values of $\widehat{\phi}_k$ for which $S_E \lesssim 1$, because otherwise $\exp(-S_E)$ is very small. Which are the important terms in S_E in this region? First, assume that the particle is relativistic, $m \ll k \ll \Lambda$ so that both m^2/k^2 and k^2/Λ^2 are very small,

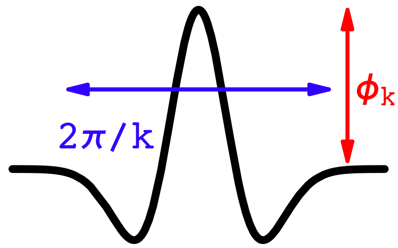

Figure 4: *sample configuration contributing to the path integral for the scalar field theory in eq. (57). Its amplitude is ϕ_k and has wave number $\sim k$ and spatial extent $\sim 2\pi/k$.*

and assume the dimensionless couplings λ, c_n, d_n are $\lesssim O(1)$. Then as one increases the amplitude $\widehat{\phi}_k$ from zero, the first term in S_E to become become $O(1)$ is the kinetic term, $(2\pi)^4\widehat{\phi}_k^2$, which occurs for $\phi_k = k\widehat{\phi}_k \sim k/(2\pi)^2$. It is because the kinetic term controls the fluctuations of the scalar field that we "canonically normalize" the field such that the kinetic term is $\frac{1}{2}(\partial\phi)^2$, and perturb in the coefficients of the other operators in the theory. Is this conclusion always true? No. For low enough k, for example, the mass term with its factor of m^2/k^2 will eventually dominate and a different scaling regime takes over. Also, if some of the dimensionless couplings λ, c_n, d_n are large, those terms may dominate and the theory will change its nature dramatically. In the first lecture we looked at scattering off a δ-function in $D = 3$ and saw that if the coupling was tuned to a particular strong value its effects could dominate low energy scattering, even though it is naively an irrelevant interaction.

What happens as we consider different momenta k? We see from eq. (61) that as k is reduced, the c_n and d_n terms, proportional to $(k^2/\Lambda^2)^n$, get smaller. Such operators are "irrelevant" operators in Wilson's language, because they become unimportant in the infrared (low k). In contrast, the mass term becomes more important; it is called a "relevant" operator. The kinetic term and the $\lambda\phi^4$ interaction do not change; such operators are called "marginal". It used to be thought that the irrelevant operators were dangerous, making the theory nonrenormalizable, while the relevant operators were safe – "superrenormalizable". As we consider radiative corrections later we will see that Wilson flipped this entirely on its head, so that irrelevant operators are now considered safe, while the existence of relevant operators is thought to be a serious problem to be solved.

In practice, when working with a relativistic theory in d spacetime dimensions with small dimensionless coupling constants, the operators with dimension d are the marginal ones, those with higher dimension are irrelevant, and those with lower dimension are relevant. The bottom line is that we can analyze the theory in a momentum expansion, working to a particular order and ignoring irrelevant operators above a certain dimension. The ability to do so will persist even when we include radiative corrections.

2.1.1 Dimensional analysis: Fermi's theory of the weak interactions

To see why dimensional analysis has practical consequences, first consider Fermi's theory of the weak interactions. Originally this was a "bottom-up" sort of EFT — Fermi did not have a complete UV description of the weak interactions, and so constructed the theory as a phenomenological modification of QED to account for neutron decay. Now we have the SM, and so we think of the Fermi theory as a "top-down" EFT: not necessary for doing calculations since we have the SM, but very practical for processes at energies far below the W mass, such as β-decay (Fermi's original application of his theory) and low energy neutral current scattering due to Z exchange (about which Fermi knew nothing).

The weak interactions refer to processes mediated by the $W^\pm$ or Z^0 bosons, whose masses are approximately 80 GeV and 91 GeV respectively. The couplings of these gauge bosons to quarks and leptons can be written in terms of the electromagnetic current

$$j^\mu_{\rm em} = \frac{2}{3}\bar{u}_i\gamma^\mu u_i - \frac{2}{3}\bar{d}_i\gamma^\mu d_i - \bar{e}_i\gamma^\mu e_i \tag{64}$$

where $i = 1, 2, 3$ runs over families, and the left-handed $SU(2)$ currents

$$j^\mu_a = \sum_\psi \overline{\psi}\gamma^\mu\left(\frac{1-\gamma_5}{2}\right)\frac{\tau_a}{2}\psi\ , \qquad a = 1, 2, 3\ , \tag{65}$$

where the $\overline{\psi}, \psi$ fields in the currents are either the lepton doublets

$$\begin{pmatrix}\nu_e\\ e\end{pmatrix}\ ,\quad \begin{pmatrix}\nu_\mu\\ \mu\end{pmatrix}\ ,\quad \begin{pmatrix}\nu_\tau\\ \tau\end{pmatrix}\ , \tag{66}$$

or the quark doublets

$$\psi = \begin{pmatrix}u\\ d'\end{pmatrix}\ ,\quad \begin{pmatrix}c\\ s'\end{pmatrix}\ ,\quad \begin{pmatrix}t\\ b'\end{pmatrix}\ , \tag{67}$$

with the "flavor eigenstates" d', s' and b' being related to the mass eigenstates d, s and b by the unitary Cabibbo-Kobayashi-Maskawa (CKM) matrix[3]:

$$q'_i = V_{ij}q_j\ . \tag{68}$$

The SM coupling of the heavy gauge bosons to these currents is

$$\mathcal{L}_{\rm J} = \frac{e}{\sin\theta_w}\left(W^+_\mu J^\mu_- + W^-_\mu J^\mu_+\right) + \frac{e}{\sin\theta_w\ \cos\theta_w}Z_\mu\left(j^\mu_3 - \sin^2\theta_w j^\mu_{\rm em}\right) \tag{69}$$

where

$$J^\mu_\pm = \frac{j^\mu_1 \mp i\, j^\mu_2}{\sqrt{2}}\ . \tag{70}$$

Tree level exchange of a W boson then gives the amplitude at low momentum exchange

$$i\mathcal{A} = \left(-i\frac{e}{\sin\theta_w}\right)^2 J^\mu_- J^\nu_+ \frac{-ig_{\mu\nu}}{q^2 - M_W^2} = -i\frac{e^2}{\sin^2\theta_w M_W^2}J^\mu_- J_{\mu+} + O\left(\frac{q^2}{M_W^2}\right)\ . \tag{71}$$

[3]The elements of the CKM matrix are named after which quarks they couple through the charged current, namely $V_{11} \equiv V_{ud}$, $V_{12} \equiv V_{us}$, $V_{21} \equiv V_{cd}$, etc.

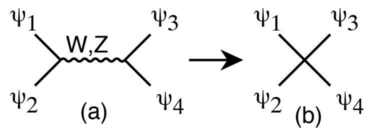

Figure 5: *(a) Tree level W and Z exchange between four fermions. (b) The effective vertex in the low energy effective theory (Fermi interaction).*

This amplitude can be reproduced to lowest order in q^2/M_W^2 by a low energy EFT with a contact interaction, Fig. 5,

$$\mathcal{L}_{\rm F} = -\frac{e^2}{\sin^2\theta_w M_W^2} J_-^\mu J_{\mu+} = \frac{8}{\sqrt{2}} G_F J_-^\mu J_{\mu+} \ , \tag{72}$$

$$G_F \equiv \frac{\sqrt{2}}{8}\frac{e^2}{\sin^2\theta_w M_W^2} = 1.166\times 10^{-5}\ \ {\rm GeV}^2\ . \tag{73}$$

This is our matching condition, analogous to the matching we did for δ-function scattering in order to reproduce the low energy behavior of square well scattering. This charged current interaction, written in terms of leptons and nucleons instead of leptons and quarks, was postulated by Fermi to explain neutron decay; the $8/\sqrt{2}$ numerical factor looks funny here because I am normalizing the currents in the way they appear in the SM, while weak currents are historically (pre-SM) normalized differently. Neutral currents were proposed in the 60's and discovered in the 70's.

Since the four-fermion Fermi interaction has dimension 6, it is an irrelevant interaction, according to our previous discussion, explaining why we say the interactions are "weak" and neutrinos are "weakly interacting". Consider, for example, some low energy neutrino scattering cross section σ. Since neutrinos only interact via W and Z exchange, the cross-section σ must be proportional to G_F^2 which has dimension -4. But a cross section has dimensions of area, or mass dimension -2. Since the only other scale around is the center of mass energy $\sqrt{s}$, on purely dimensional grounds σ must scale with energy as

$$\sigma_\nu \simeq G_F^2 s\ , \tag{74}$$

This explains why low energy neutrinos are so hard to detect, and the weak interactions are weak; at LHC energies, however, where the effective field theory has broken down, the weak interactions are marginal and characterized by the $SU(2)$ coupling constant $g \simeq 0.6$, about twice as strong as the electromagnetic coupling. It is a simple result for which one does not need the full machinery of the SM to derive.

It looks like the neutrino cross section grows with s without bound, but remember that this EFT is only valid up to $s \simeq M_W$.

2.1.2 Dimensional analysis: the blue sky

Another top-down application of EFT is to answer the question of why the sky is blue. More precisely, why low energy light scattering from neutral atoms in their ground state (Rayleigh scattering) is much stronger for blue light than red[4]. The physics of the scattering process could be analyzed using exact or approximate atomic wave functions and matrix elements, but that is overkill for low energy scattering. Let's construct an "effective Lagrangian" to describe this process. This means that we are going to write down a Lagrangian with all interactions describing elastic photon-atom scattering that are allowed by the symmetries of the world — namely Lorentz invariance and gauge invariance. Photons are described by a field A_μ which creates and destroys photons; a gauge invariant object constructed from A_μ is the field strength tensor $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$. The atomic field is defined as ϕ_v, where ϕ_v destroys an atom with four-velocity v_μ (satisfying $v_\mu v^\mu = 1$, with $v_\mu = (1,0,0,0)$ in the rest-frame of the atom), while $\phi_v^\dagger$ creates an atom with four-velocity v_μ. In this case we should use relativistic scaling, since we are interested in on-shell photons, and are uninterested in recoil effects (the kinetic energy of the atom):

$$[x] = [t] = -1, \quad [p] = [E] = [A_\mu] = 1 \;, \qquad [\phi] = \frac{3}{2} \;, \tag{75}$$

where the atomic field ϕ destroys an atom with four-velocity v_μ (satisfying $v_\mu v^\mu = 1$, with $v_\mu = (1,0,0,0)$ in the rest-frame of the atom), while $\phi^\dagger$ creates an atom with four-velocity v_μ.

So what is the most general form for $\mathcal{L}_{\rm eft}$? Since the atom is electrically neutral, gauge invariance implies that ϕ can only be coupled to $F_{\mu\nu}$ and not directly to A_μ. So $\mathcal{L}_{\rm eft}$ is comprised of all local, Hermitian monomials in $\phi^\dagger\phi$, $F_{\mu\nu}$, v_μ, and ∂_μ. Certain combinations we needn't consider for the problem at hand — for example $\partial_\mu F^{\mu\nu} = 0$ for radiation (by Maxwell's equations); also, if we define the energy of the atom at rest in it's ground state to be zero, then $v^\mu \partial_\mu \phi = 0$, since $v_\mu = (1,0,0,0)$ in the rest frame, where $\partial_t \phi = 0$. Similarly, $\partial_\mu \partial^\mu \phi = 0$. Thus we are led to consider the interaction Lagrangian

$$\begin{aligned}\mathcal{L}_{\rm eft} &= c_1 \phi^\dagger \phi F_{\mu\nu} F^{\mu\nu} + c_2 \phi^\dagger \phi v^\alpha F_{\alpha\mu} v_\beta F^{\beta\mu} \\ &\quad + c_3 \phi^\dagger \phi (v^\alpha \partial_\alpha) F_{\mu\nu} F^{\mu\nu} + \ldots\end{aligned} \tag{76}$$

The above expression involves an infinite number of operators and an infinite number of unknown coefficients! Nevertheless, dimensional analysis allows us to identify the leading contribution to low energy scattering of light by neutral atoms.

With the scaling behavior eq. (75), and the need for $\mathcal{L}$ to have dimension 4, we find the dimensions of our couplings to be

$$[c_1] = [c_2] = -3 \;, \qquad [c_3] = -4 \;. \tag{77}$$

Since the c_3 operator has higher dimension, we will ignore it. What are the sizes of the coefficients $c_{1,2}$? To do a careful analysis one needs to go back to the full Hamiltonian for the

[4]By "low energy" I mean that the photon energy E_γ is much smaller than the excitation energy ΔE of the atom, which is of course much smaller than its inverse size or mass:

$$E_\gamma \ll \Delta E \ll r_0^{-1} \ll M_{atom}$$

where r_0 is the atomic size, roughly the Bohr radius. Thus the process is necessarily elastic scattering, and to a good approximation we can ignore that the atom recoils, treating it as infinitely heavy.

atom in question interacting with light, and "match" the full theory to the effective theory. Here I will just estimate the sizes of the c_i coefficients, rather than doing some atomic physics calculations. Note that extremely low energy photons cannot probe the internal structure of the atom, and so the cross-section ought to be classical, only depending on the size of the scatterer, which I will denote as r_0, roughly the Bohr radius. Since such low energy scattering can be described entirely in terms of the coefficients c_1 and c_2, we conclude that

$$c_1 \simeq c_2 \simeq r_0^3 \ .$$

The effective Lagrangian for low energy scattering of light is therefore

$$\mathcal{L}_{\rm eft} = r_0^3 \left(a_1 \phi^\dagger \phi F_{\mu\nu} F^{\mu\nu} + a_2 \phi^\dagger \phi v^\alpha F_{\alpha\mu} v_\beta F^{\beta\mu} \right) \tag{78}$$

where a_1 and a_2 are dimensionless, and expected to be $\mathcal{O}(1)$. The cross-section (which goes as the amplitude squared) must therefore be proportional to r_0^6. But a cross section σ has dimensions of area, or $[\sigma] = -2$, while $[r_0^6] = -6$. Therefore the cross section must be proportional to

$$\sigma \propto E_\gamma^4 r_0^6 \ , \tag{79}$$

growing like the fourth power of the photon energy. Thus blue light is scattered more strongly than red, and the sky far from the sun looks blue. The two independent coefficients in this calculation must correspond to the electric and magnetic polarizabilities of the atom.

Is the expression eq. (79) valid for arbitrarily high energy? No, because we ignored higher dimension terms in the effective Lagrangian we used, terms which become more important at higher energies — and at sufficiently high energy these terms are all in principle equally important and the EFT breaks down. To understand the size of corrections to eq. (79) we need to know the size of the c_3 operator (and the rest we ignored). Since $[c_3] = -4$, we expect the effect of the c_3 operator on the scattering amplitude to be smaller than the leading effects by a factor of E_γ/Λ, where Λ is some energy scale. But does Λ equal M_{atom}, $r_0^{-1} \sim \alpha m_e$ or $\Delta E \sim \alpha^2 m_e$? The latter — the energy required to excite the atom — is the smallest energy scale and hence the most important. We expect our approximations to break down as $E_\gamma \to \Delta E$ since for such energies the photon can excite the atom. Hence we predict

$$\sigma \propto E_\gamma^4 r_0^6 \left(1 + \mathcal{O}(E_\gamma/\Delta E)\right). \tag{80}$$

The Rayleigh scattering formula ought to work pretty well for blue light, but not very far into the ultraviolet. Note that eq. eq. (80) contains a lot of physics even though we did very little work. More work is needed to compute the constant of proportionality.

2.2 Accidental symmetry and BSM physics

Now let's switch tactics and use dimensional analysis to talk about bottom-up applications of EFT. We would like to have clues of physics beyond the SM (BSM). Evidence we currently have for BSM physics are the existence of gravity, neutrino masses and dark matter. Hints for additional BSM physics include circumstantial evidence for Grand Unification and for inflation, the absence of a neutron electric dipole moment, and the baryon number asymmetry of the universe. Great puzzles include the origin of flavor and family structure, why the electroweak scale is so low compared to the Planck scale (but not so far from the QCD scale), and why we live in an epoch where matter, dark matter, and dark energy all have have rather similar densities.

In order to make progress we would like to have more data, and looking for subtle effects due to irrelevant operators can in some cases give us a much farther experimental reach than can collider physics, exactly in the same way the Fermi interaction provided a critical clue which eventually led to the SM. Those cases are necessarily ones where the irrelevant operators violate symmetries that are preserved by the marginal and irrelevant operators in the SM, and are therefore the leading contribution to certain processes. We call these symmetries "accidental symmetries": they are not symmetries of the UV theory, but they are approximate symmetries of the IR theory.

A simple and practical example of an accidental symmetry is $SO(4)$ symmetry in lattice QCD — the Euclidian version of the Lorentz group. Lattice QCD formulates QCD on a 4d hypercubic lattice, and then looks in the IR on this lattice, focusing on modes whose wavelengths are so long that they are insensitive to the discretization of spacetime. But why is it obvious that a hypercubic lattice will yield a continuum Lorentz invariant theory?

The reason lattice field theory works is because of accidental symmetry: Operators on the lattice are constrained by gauge invariance and the hypercubic symmetry of the lattice. While it is possible to write down operators which are invariant under these symmetries while violating the $SO(4)$ Lorentz symmetry, such operators all have high dimension and are not relevant. For example, if A_μ is a vector field, the $SO(4)$-violating operator $A_1A_2A_3A_4$ is hypercubic invariant and marginal (dimension 4) and so could spoil the continuum limit we desire; however, the only vector field in lattice QCD is the gauge potential, and such an operator is forbidden because it is not gauge invariant. In the quark sector the lowest dimension operator one can write which is hypercubic symmetric but Lorentz violating is

$$\sum_{\mu=1}^{4} \bar\psi\gamma_\mu D_\mu^3\psi \tag{81}$$

which is dimension six and therefore irrelevant. Thus Lorentz symmetry is automatically restored in the continuum limit.

Accidental symmetries in the SM notably include baryon number B and lepton number L: if one writes down all possible dimension ≤ 4 gauge invariant and Lorentz invariant operators in the SM, you will find they all preserve B and L. It is possible to write down dimension five $\Delta L = 2$ operators and dimension six $\Delta B = \Delta L = 1$ operators, however. That means that no matter how completely B and L are broken in the UV, at our energies these irrelevant operators become...irrelevant, and B and L appear to be conserved, at least to high precision. So perhaps B and L are not symmetries of the world at all – they just look like good symmetries because the scale of new physics is very high, so that the irrelevant B and L violating operators have very little effect at accessible energies. We will look at these different operators briefly in turn.

2.2.1 BSM physics: neutrino masses

The most important irrelevant operators that could be added to the SM are dimension 5. Any such operator should be constructed out of the existing fields of the SM and be invariant under the $SU(3) \times SU(2) \times U(1)$ gauge symmetry. Recall that the matter fields in the SM have the gauge quantum numbers

$$\begin{aligned} \text{LH fermions:} \quad & Q = (3,2)_{\frac{1}{6}} \ , \quad U = (\bar{3},1)_{-\frac{2}{3}} \ , \quad D = (\bar{3},1)_{\frac{1}{3}} \ , \quad L = (1,2)_{-\frac{1}{2}} \ , \quad E = (1,1)_1 \ , \\ \text{Higgs:} \quad & H = (1,2)_{-\frac{1}{2}} \ , \quad \widetilde{H} = (1,2)_{\frac{1}{2}} \ , \end{aligned} \tag{82}$$

where $\widetilde{H}_i = \epsilon_{ij} H_j^*$ is not an independent field. The gauge fields transform as adjoints under their respective gauge groups.

The only gauge invariant dimension 5 operator one can write down is the $\Delta L = 2$ operator (violating lepton number by two units) [5]:

$$\mathcal{L}_{\Delta L=2} = -\frac{1}{\Lambda}(L\widetilde{H})(L\widetilde{H}) \ , \qquad L = \begin{pmatrix} \nu \\ \ell^- \end{pmatrix} \ , \qquad \widetilde{H} = \begin{pmatrix} h^+ \\ h^0 \end{pmatrix} \ , \qquad \langle \widetilde{H} \rangle = \frac{v}{\sqrt{2}} \begin{pmatrix} 0 \\ 1 \end{pmatrix} \ , \tag{83}$$

where $v = 250$ GeV. There is only one independent operator (ignoring lepton flavor) since the two Higgs fields ($\widetilde{H}\widetilde{H}$) cannot be antisymmetrized and therefore must be in an $SU(2)$-triplet. An operator coupling LL in a weak triplet to HH in a weak triplet can be rewritten in the above form, where the combination $(L\widetilde{H})$ is a weak singlet,

$$(L\widetilde{H}) = \left(\nu h^0 - \ell^- h^+\right) \quad \longrightarrow \quad \frac{\nu v}{\sqrt{2}} \ . \tag{84}$$

Therefore after spontaneous symmetry breaking by the Higgs, the operator gives a contribution to the neutrino mass,

$$\mathcal{L}_{\Delta L=2} = -\frac{1}{2} m_\nu \nu\nu \ , \qquad m_\nu = \frac{v^2}{\Lambda} \ , \tag{85}$$

a $\Delta L = 2$ Majorana mass for the neutrino. A mass of $m_\nu = 10^{-2}$ eV corresponds to $\Lambda = 6 \times 10^{15}$ GeV, an interesting scale, being near the scale of GUT models, and far beyond the reach of accelerator experiments. Or: if $\Lambda = 10^{19}$ GeV, the Planck scale, then $m_\nu = 10^{-5}$ eV. This operator provides a possible and rather compelling explanation for the smallness of observed neutrino masses: they arise as Majorana masses because lepton number is not a symmetry of the universe, but are very small because lepton number becomes an accidental symmetry below a high scale. Of course, we could have the spectrum of the low energy theory wrong: perhaps there is a light right-handed neutrino and neutrinos only have L-preserving Dirac masses like the charged leptons, small simply because of a very small Yukawa coupling to the Higgs. Neutrinoless double beta decay experiments are searching for lepton number violation in hopes of establishing the Majorana mass scenario.

In any case, it is interesting to imagine what sort of UV physics could give rise to the operator in eq. (83). Three possibilities present themselves for how such an operator could

[5] One might expect to be able to write down magnetic dipole operators of the form $\bar{\psi}\sigma_{\mu\nu}F^{\mu\nu}\psi$, but such operators have the chiral structure of a mass term and require an additional Higgs field to be gauge invariant, making them dimension 6.

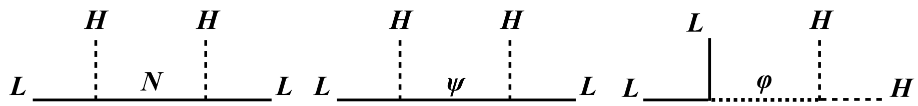

Figure 6: *Three ways the dimension 5 operator for neutrino masses in eq. (83) could arise from tree level exchange of a heavy $SU(2) \times U(1)$ singlet fermion N, a heavy $SU(2) \times U(1)$ triplet fermion ψ, or else from exchange of a massive $SU(2)$ triplet scalar ϕ.*

arise from a high energy theory at tree level, shown in Fig. 6 – either through exchange of a heavy $SU(2) \times U(1)$ singlet fermion N (a "right handed neutrino"), through exchange of a heavy $SU(2) \times U(1)$ triplet fermion ψ, or else via exchange of a heavy scalar with quantum number 3_1 under $SU(2) \times U(1)$. The fact that the resultant light neutrino mass is inversely proportional to the new scale of physics (called the "see-saw mechanism") simply results from the fact that a neutrino mass operator in the SM is an irrelevant dimension-5 operator. Note that just as G_F is proportional to g^2/M_W^2, and therefore knowing G_F was not sufficient for predicting the W mass, the scale Λ is not necessarily the mass of a new particle, as it will be inversely proportional to coupling constants about which we know nothing except in the context of some particular UV candidate theory.

A fourth possibility for neutrino masses does not arise in EFT at all: the neutrino mass might be Dirac, with right-handed neutrino we have not detected (since it is neutral under gauge symmetries) which couples to the lepton doublet L via an extremely small Yukawa coupling to the Higgs. In this case lepton symmetry is not violated. Seeing neutrinoless double β decay would be a signal of a $\Delta L = 2$ process and would be evidence in favor of a seesaw origin for neutrino masses.

2.2.2 BSM physics: proton decay

At dimension 6 one can write down very many new operators in the SM, including interesting CP violating electric dipole moment operators for fermions. One particularly interesting set of dimension 6 operators in the SM are those that violate B; they all consist of three quark fields (for color neutrality) and a lepton field, and are therefore all $\Delta B = 1$, $\Delta L = 1$ operators which conserve the combination $B - L$. These are very interesting because (i) B is a particularly good symmetry, since the proton appears to be very stable, and (ii) B violation is a prerequisite for any theory of baryogenesis. Below the QCD scale one needs to match the three quark operator onto hadron fields. An example of such an operator would be

$$\frac{1}{\Lambda^2} \epsilon_{abc} \epsilon_{\alpha\beta} \epsilon_{\gamma\delta} (d_L^{a\alpha} u_L^{b\beta})(u_L^{c\gamma} e_L^{\delta} - d_L^{c\gamma} \nu_L^{\delta}) \ , \tag{86}$$

where a, b, c are color indices and $\alpha, \beta, \gamma, \delta$ are $SU(2)$ Lorentz indices for the left-handed Weyl spinors; the terms in parentheses are weak $SU(2)$ singlets, and the whole operator is neutral under weak hypercharge. Below the QCD scale one has to match the three-quark operator onto hadrons fields. Thus roughly speaking $uud \to Z_1 p + Z_2(p\pi^0 + n\pi^+) + \ldots$. We can assume that the Z factors are made up of pure numbers times the appropriate powers of the strong interaction scale, such as $f_\pi \simeq 100$ MeV, the pion decay constant. The Z_1 term will lead to positron-proton mixing and cannot lead to proton decay, but the Z_2 term can via the processes $p \to e^+\pi^0$, or $p \to \pi^+\nu$. We can make a crude estimate of the width (inverse lifetime) to be

$$\Gamma \simeq \frac{M_p^5}{\Lambda^4}\frac{1}{8\pi} \tag{87}$$

where I used dimensional analysis to estimate the M_p^5/Λ^4 factor, assuming that the strong interaction scale in Z_2 as well as powers of momenta from phase space integrals could be approximated by the proton mass M_p , and I inserted a typical 2-body phase space factor of $1/8\pi$. For a bound on the proton lifetime of $\tau_p > 10^{34}$ years, this crude estimate gives us $\Lambda \gtrsim 10^{16}$ GeV, not so far off the bound one finds from a more sophisticated calculation. If proton decay is discovered, that will tell us something about the scale of new physics, and then the task will be to construct the full UV theory from what we learn about proton decay, much as the SM was discovered starting from the Fermi theory.

2.3 BSM physics: "partial compositeness"

This next topic does not have to do with accidental symmetry violation, but instead picks up on an interesting feature of the baryon number violating interaction we just discussed, as it suggests a mechanism for quarks and leptons to acquire masses without a Higgs. In estimating the effects of the dimension six $\Delta B = 1$ operator in the previous section I said that the 3-quark operator could be expanded as $uud \to Z_1 p + Z_2(p\pi^0 + n\pi^+) + \ldots$, and then focussed on the Z_2 term. But what about the Z_1 term? By dimensions, $Z_1 \sim \Lambda_{QCD}^3$, and so that term gives rise to a peculiar mass term of the form

$$\frac{\Lambda_{QCD}^3}{\Lambda^2} pe + h.c. \tag{88}$$

which allows a proton to mix with a positron, and the anti-proton to mix with the electron. This is not experimentally interesting, but it is an interesting phenomenon for a theorist to contemplate. Imagine eliminating the Higgs doublet from the SM. The proton would still get a mass from chiral symmetry breaking in QCD even though the quarks would remain massless, and due to the above term, and even though there would not be an electron mass directly from the weak interactions, there would be a above contribution to the proton mass due to the above positron-proton mixing. For the two component system one would have a mass matrix looking something like

$$\begin{pmatrix} M_p & \frac{\Lambda_{QCD}^3}{\Lambda^2} \\ \frac{\Lambda_{QCD}^3}{\Lambda^2} & 0 \end{pmatrix} \tag{89}$$

and to the extent that $\Lambda \gg \Lambda_{QCD}$ we find the mass eigenvalues to be

$$m_1 \simeq M_p\ , \qquad m_2 \simeq \frac{\Lambda_{QCD}^6}{M_p \Lambda^4} \tag{90}$$

so for $\Lambda = 10^{16}$ GeV the positron gets a mass of $m_e \simeq 10^{-64}$ GeV. Yes, this is a ridiculously small mass of no interest, but it is curious that the positron (electron) got a mass at all, without there being any Higgs field! It must be that QCD when QCD spontaneously breaks chiral symmetry, it has also broken $SU(2) \times U(1)$, without a Higgs, and that this proton decay operator has somehow taken the place of a Higgs Yukawa coupling. Quark fields and QCD has assumed both of the the roles that the Higgs field plays in the SM. Therefore it is worth asking whether this example be modified somehow to obtain more interesting masses for quarks and leptons?

In a later lecture we will examine how QCD breaks the weak interactions, and how a scaled up version called technicolor, with the analogue of the pion decay constant f_π being up at the 250 GeV scale instead of 93 MeV, could properly account for the spontaneous breaking of $SU(1) \times U(1)$ without a Higgs. Here I will just comment that such a theory would be expected to have TeV mass "technibaryons", which could carry color and charge. With an appropriate dimension 6 operator such as our proton decay operator, but with techniquarks in place of quarks, and all the standard model fermions in place of the positron field, in principle one could give masses to all the SM fermions through their mixing with the technibaryons. This is the idea of "partial compositeness", which in its original formulation [1] was not especially useful for model building, but which has become more interesting in the context of composite Higgs [2] – more about composite Higgs later too.

,

2.4 Problems for lecture II

II.1) What is the dimension of the operator ϕ^{10} in a $d = 2$ relativistic scalar field theory?

II.2) One defines the "critical dimension" d_c for an operator to be the spacetime dimension for which that operator is marginal. How will that operator behave in dimensions d when $d > d_c$ or $d < d_c$? In a theory of interacting relativistic scalars, Dirac fermions, and gauge bosons, determine the critical dimension for the following operators:

1. A ϕ^3 interaction;
2. A gauge coupling to either a fermion or a boson through the covariant derivative in the kinetic term;
3. A Yukawa interaction, $\phi\bar{\psi}\psi$;
4. An anomalous magnetic moment coupling $\bar{\psi}\sigma_{\mu\nu}F^{\mu\nu}\psi$ for a fermion;
5. A four fermion interaction, $(\bar{\psi}\psi)^2$.

II.3) Derive the analogue of Fermi's theory in eq. (72) for tree level Z exchange, expressing your answer in terms of G_F using the fact that $M_Z^2 = M_W^2/\cos^2\theta_w$.

II.4) How would one write down an electric dipole operator in QED, and what dimension does it have? What would you have to do to make a gauge invariant electric dipole operator in the SM that is invariant under the full $SU(3) \times SU(2) \times U(1)$ gauge symmetry?

II.5) Show that the operator

$$\epsilon_{\alpha\beta}\left(L_{\alpha i}(\sigma_2\sigma^a)_{ij}L_{\beta j}\right)\left(H_k(\sigma_2\sigma^a)_{k\ell}H_\ell\right)$$

is equivalent up to a factor of two to

$$\epsilon_{\alpha\beta}(L_{\alpha i}(\sigma_2)_{ij}H_j)(L_{\beta k}(\sigma_2)_{k\ell}H_\ell)$$

where α, β are Weyl spinor indices, and i, j, k, ℓ are the $SU(2)$ gauge group indices. Write down the two high energy theories that could give rise to the neutrino mass operator as in Fig. 6. How do I see that these theories break lepton number by two units?

3 EFT and radiative corrections

Up to now we have ignored quantum corrections in our effective theory. A Lagrangian such as eq. (57) is what used to be termed a "nonrenormalizable" theory, and to be shunned. The problem was that the theory needs an infinite number of counterterms to subtract all infinities, and was thought to be unpredictive. In contrast, a "renormalizable" theory contained only marginal and relevant operators, and needed only a finite number of counterterms, one per marginal or relevant operator allowed by the symmetries. (A "superrenormalizable" theory contained only relevant operators, and was finite beyond a certain order in perturbation theory.) However Wilson changed the view of renormalization. In a perturbative theory, irrelevant operators are renormalized, but stay irrelevant. On the other hand, the coefficients of relevant operators are renormalized to take on values proportional to powers of the cutoff, unless forbidden by symmetry. Thus in Wilson's view the relevant operators are the problem, since giving them small coefficients requires fine-tuning – unless a symmetry forbids corrections that go as powers of the cutoff. Relevant operators protected by symmetry include fermion masses and Goldstone boson masses, but for a general interacting scalar, the natural mass is $m^2 \simeq \alpha\Lambda^2$ — which means one should never see such scalars in the low energy theory.

In this lecture I discuss the techniques used to create top-down EFTs beyond tree level, as well as an example of an EFT with a marginal interaction with asymptotic freedom and an exponentially small IR scale.

3.1 Matching

I will consider a toy model for UV physics with a light scalar ϕ and a heavy scalar S :

$$\mathcal{L}_{\mathrm{UV}} = \frac{1}{2}\left((\partial\phi)^2 - m^2\phi^2 + (\partial S)^2 - M^2 S^2 - \kappa\phi^2 S\right) \tag{91}$$

The parameter κ has dimension of mass, and I will assume $\kappa \lesssim M$ and that $\langle S\rangle = 0$. This is a pretty sill model, but it is useful as an example since the Feynman diagrams are very simple; never mind that the vacuum energy is unboundedbelow, as one won't see this in perturbation theory. Suppose we are interested in $2\phi \to 2\phi$ scattering at energies much below the S mass M, and want to construct the EFT with the heavy S field "integrated out". Formally:

$$e^{-i\frac{\mathcal{S}_{\mathrm{eft}}(\phi)}{\hbar}} = \int [dS] e^{-i\frac{\mathcal{S}(\phi,S)}{\hbar}} \tag{92}$$

where I use $\mathcal{S}$ to denote the action, to distinguish it from the field S. There are good reasons for doing so: if you try to compute observables in this theory at some low momentum $k << M$ you are typically going to run into large logarithms of the form $\ln k^2/M^2$ that will spoil perturbation theory. They are easily taken care of in an EFT where you integrate out S at the scale $\mu = M$, matching the EFT to the full theory to ensure that you are reproducing the same physics. Then within the EFT you run the couplings from $\mu = M$

down to $\mu = k$ before doing your calculation. The renormalization group running sums up these large logs for you.

It is convenient to perform the matching in an $\hbar$ expansion, meaning that first you make sure that tree diagrams agree in the two theories, as in our derivation of the Fermi theory of weak interactions from the SM. Then you make sure that the two theories agree at $O(\hbar)$, etc. What sort of graphs does this matching entail, and why is this justified? Consider the original theory, in which $\hbar$ only appears in the explicit factor in $\exp -i\mathcal{S}/\hbar$, and look at a graph with P propagators, V vertices and E external legs. Euler's formula tells us that $L = P - V + 1$ [6]. Since $\hbar$ enters the path integral through $\exp(iS/\hbar)$, every propagator brings a power of $\hbar$ and every vertex brings a power of $\hbar^{-1}$; it follows that a graph is proportional to $\hbar^{P-V} = \hbar^{L-1}$. Since the graph is providing a vertex in $S_{\rm eft}/\hbar$, the L-loop matching involves contributions to $\mathcal{L}_{\rm eft}$ at $O(\hbar^L)$. An $\hbar$ expansion is always justified when a perturbative expansion is justified. To see that, consider a graph in a theory with a single type of interaction vertex involving n fields. In this case we have $(E+2P) = nV$, since one end of every external line and two ends of every internal line must end on a vertex, and there must be n lines coming in to each vertex. Putting this together with Euler's equation we have $V = (2L+E-2)/(n-2)$, which shows that for a given number of external lines, the number of vertices (and hence the power of the coupling constant) grows with the number of loops, so a loop expansion (or equivalently, an $\hbar$ expansion) is justified if a perturbative expansion is justified. This can be generalized to a theory with several types of vertices.

So we match the UV theory eq. (91) to the EFT order by order in an $\hbar$ expansion with

$$\mathcal{L}_{\rm eft} = \mathcal{L}^0_{\rm eft} + \hbar\mathcal{L}^1_{\rm eft} + \hbar^2\mathcal{L}^2_{\rm eft} + \dots \tag{93}$$

What makes this interesting is that we start introducing powers of $\hbar$ into the coupling constants of the EFT, so in the EFT the powers of $\hbar$ in a graph can be higher than the number of loops; the example below should make this clear. Since the EFT is expressed in terms of local operators, the matching also involves performing an expansion in powers of external momenta, with a contribution at p^n matching onto an n-derivative operator. We only match amplitudes involving light particles on external legs.

Tree level matching. At $\hbar^0$ we have to match the two theories at tree level. There are an infinite number of tree level graphs one can write down in the full theory, but the only ones we have to match are those that do not fall apart when I cut a light particle propagator...these I will call "1LPI" diagrams, for "1 Light Particle Irreducible". The other graphs will be automatically accounted for in the EFT by connecting the vertices with light particle propagators. That means we can fully determine the EFT by computing the three tree diagrams of the UV theory on the left side of Fig. 7. Because we are doing a momentum expansion, these will determine an infinite number of operator coefficients in the EFT. To compute the 4-point vertices in the EFT at this order, we equate the graphs shown in Fig. 7. Here I do not compute the graphs, but just indicate their general size,

[6] One way to derive this is to think of some Green function represented by a Feynman graph: it involves L integrations over loop momenta and one overall momentum conserving δ-function, $\delta^4(p_{\rm tot})$; on the other hand it also equals one momentum integration for each propagator and a momentum conserving δ-function at each vertex. From that observation one finds $(\int d^4p)^L\delta^4(p) \sim (\int d^4p)^P(\delta^4(p))^V$ or $(L-1) = (P-V)$.

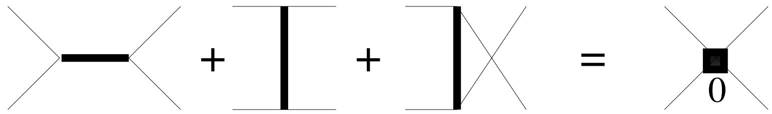

Figure 7: *Matching at $O(\hbar^0)$ between the UV theory and the EFT: on the left, integrating out the heavy scalar S (dark propagator); on the right, all contributions of four-point vertices the tree level EFT $\mathcal{L}^0_{eft}$. Equating the two sides allows on to solve for this vertices.*

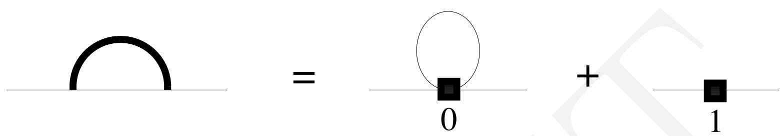

Figure 8: *Matching the 2-point function in the EFT at $O(\hbar)$. On the left, the 1-loop 1LPI graph contributing in the full theory, and on the right, graphs from the EFT include 1-loop graphs involving the 4-point vertices from $\mathcal{L}^0_{eft}$, as well as $O(\hbar)$ tree-level contributions from ϕ^2 operators in $\mathcal{L}^1_{eft}$, including the mass and kinetic term, as well as the infinite number of operators induced at this order with more derivatives. When working to a given order in a low momentum expansion, one does not need to compute all of these higher derivative operator coefficients.*

with the result

$$\mathcal{L}^0_{\rm eft} = \frac{1}{2}(\partial\phi)^2 - \frac{1}{2}m^2\phi^2 - c_0\frac{\kappa^2}{M^2}\frac{\phi^4}{4!} - d_0\frac{\kappa^2}{M^4}\frac{(\partial\phi)^2\phi^2}{4} + \dots\ , \tag{94}$$

where c_0, d_0 etc. are going to be $O(1)$ dimensionless numbers and the ellipses refers to operators with four powers of ϕ and more powers of derivatives. The subscript 0 indicates that these coupling constants are $O(\hbar^0)$. The factors of κ^2 comes from the two vertices on the LHS of Fig. 7, and expanding the heavy scalar propagator in powers of the light field's momentum gives terms of the form $(p^2/M^2)^n \times 1/M^2$.

One loop matching. At $O(\hbar^1)$ we have to compute all 1-loop 1PLI graphs in the UV theory with arbitrary numbers of external legs, in a Taylor expansion in all powers of external momenta, and equate the result to all diagrams in the EFT that are order $\hbar$; the latter include (i) all 1-loop diagrams from $\mathcal{L}^0_{\rm eft}$ (since its couplings are $O(\hbar^0)$ and a loop brings in a power of $\hbar$), plus (ii) all tree diagrams from $\mathcal{L}^1_{\rm eft}$, since the couplings of $\mathcal{L}^1_{\rm eft}$ are $O(\hbar)$. The matching conditions for the two-point functions are shown in Fig. 8, and those for the four-point functions are shown in Fig. 9. All loop diagrams are most easily renormalized using the $\overline{MS}$ with renormalization scale set to the matching scale, e.g. $\mu = M$, so that the $\ln M^2/\mu^2$ terms that will arise vanish. The result one will find is

$$\hbar\mathcal{L}^1_{\rm eft} \quad = \quad \frac{1}{2}\left(a_1\frac{\kappa^2}{16\pi^2 M^2}\right)(\partial\phi)^2 - \frac{1}{2}\left(b_1\frac{\kappa^2}{16\pi^2} + b_1'\frac{m^2}{16\pi^2}\right)\phi^2$$

$$-c_1\left(\frac{\kappa^4}{16\pi^2 M^4}\right)\frac{\phi^4}{4!} - d_1\left(\frac{\kappa^4}{16\pi^2 M^6}\right)\frac{(\partial\phi)^2\phi^2}{4} + \ldots \tag{95}$$

where the coefficients a_1, b_1, b_1', c_1, d_1 are going to be $\mathcal{O}(\hbar)$. In the above expression the b_1 term arises from the loop on the left of the equal sign in Fig. 8, while the b_1' term arises from the loop to the right of the equal sign. In addition at this order there are higher n-point vertices generated in the EFT, such as ϕ^6, $(\phi\partial^2\phi)^2$, etc. This Lagrangian can be used to compute $2\phi \to 2\phi$ scattering up to 1 loop. One can perform an a_1-dependent rescaling of the ϕ field to return to a conventionally normalized kinetic term.

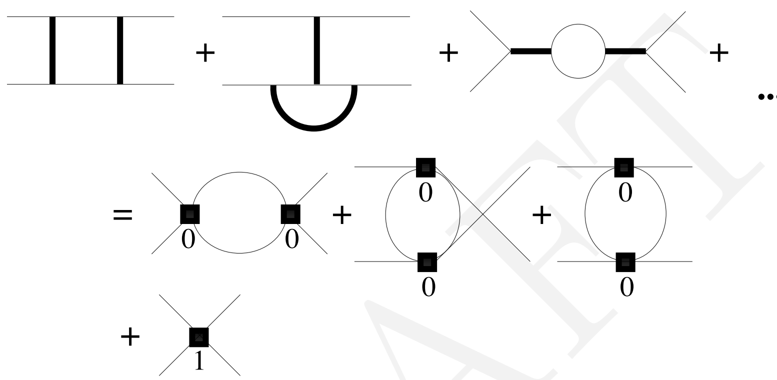

Figure 9: *Matching the 4-point function in the EFT at $O(\hbar)$. On the left, the 1LPI graphs in the full theory (with the ellipsis indicating other topologies), and on the right the $O(\hbar)$ contribution from the EFT, including 1-loop graphs involving the 4-point vertices from $\mathcal{L}^0_{EFT}$ and tree level contributions from 4-point vertices in $\mathcal{L}^1_{EFT}$, which are determined from this matching condition.*

Let me close this section with several comments about the above example:

- Notice that the loop expansion is equivalent to an expansion in $(\kappa^2/16\pi^2 M^2)$. To the extent that this is a small number, perturbation theory and the loop expansion make sense.
- We see that the matching correction to the scalar mass^2 includes a term proportional to κ^2 instead of m^2, so that even if we took $m \ll \kappa$, it is "unnatural" for the physical mass to be $\ll \frac{\kappa^2}{16\pi^2}$. In fact, in order for the meson to have a very light physical mass would require a finely tuned conspiracy between m^2 and κ^2.
- The coefficients of operators in the effective field theory are regularization scheme dependent. Their values differ for different schemes, but physical predictions do not (e.g, the relative cross sections for $2\phi \to 2\phi$ at two different energies).
- In the matching conditions the graphs in both theories have pieces depending non-analytically on light particle masses and momenta (eg, $\ln m^2$ or $\ln p^2$)...these terms cancel on both sides of the matching condition so that the interactions in $\mathcal{L}_{\rm eft}$ have a local expansion in inverse powers of $1/M$. This is an important and generic property of effective field theories.

One can now use the effective theory one has constructed to compute low energy $\phi - \phi$ scattering. If one is interested in physics at scales far below the cutoff of the EFT one might have to renormalization group improve the answer. For example, in a GUT one integrates out particles at the scale $M_{\rm GUT} \sim 10^{15}$ GeV and then computes scattering at energies around 1 GeV. Without RG improvement, the perturbative expansion will involve terms such as $\alpha(M_{\rm GUT}) \ln 10^{15} \sim 0.6$. So one eliminates the log by computing in terms of $\alpha(1\ \text{GeV})$. To do that one just needs to compute the β functions in the EFT, ignoring the heavy particles integrated out at $M_{\rm GUT}$. This procedure requires using a mass-independent renormalization scheme so one can solve the RG equations. Now since both the proton and top quark exist in the EFT, one risks getting terms such as $\alpha(1\ \text{GeV}) \ln m_t/m_p$, for example, which also isn't so good if α is the strong coupling. So one can construct a ladder of EFTs, integrating out GUT scale particles at M_{GUT}, the top quark at the scale m_t, the W and Z at their mass scale, etc. At each stage in this ladder of EFTs one uses the β function appropriate for the light particles that remain as explicit degrees of freedom in the EFT.

Matching computations like this are used for predicting the low energy gauge couplings in the SM as predicted by Grand Unified Theories (GUTS), integrating out the heavy particles at the GUT scale $M_{\rm GUT}$ and matching onto the SM as the EFT. At tree level matching, the gauge couplings in the EFT at the scale $\mu = M_{\rm GUT}$ are equal (when suitably normalizing the $U(1)$ coupling), and then one runs them down to low energy, each gauge coupling running in the SM with its own 1-loop β-function. This is the classic calculation of Georgi, Quinn and Weinberg [3] and can be used to predict $\alpha_s(M_Z)$, since the input are two unknowns (the scale $M_{\rm GUT}$ and the GUT gauge coupling $g(\mu)$ at $\mu = M_{\rm GUT}$) while the output are the three parameter of the SM α, $\alpha_s(M_Z)$, and $\sin^2\theta_w$. However, if you want greater precision you must match the GUT to the EFT at one loop, which generates small and unequal shifts in the SM gauge coupling at $\mu = M_{\rm GUT}$, and then one scales them down using the 2-loop β-functions.

3.2 Relevant operators and naturalness

We have seen that a scalar field mass typically gets large additive quantum corrections, so that in a theory with new physics at scales much larger than the weak scale (e.g. any GUT theory, and probably any theory with gravity!) it seems unnatural to have light scalars in the low energy effective theory. This would seem to be a potential problem for the SM, where we know there exists a a Higgs with a weak scale mass, far below the Planck and GUT scales. A natural size for the Higgs mass^2 would be $\frac{\alpha}{4\pi}\Lambda^2$ where Λ is the cutoff of the ET we call the SM. But what is Λ? We have seen that for proton decay and lepton violation we expect $\Lambda \gtrsim 10^{15}$ GeV; if that is the appropriate Λ to use in the Higgs mass estimate, then the SM must be fine-tuned to ~ 13 orders of magnitude! In fact, the Higgs mass appears not to be fine-tuned only if $\Lambda \sim 1$ TeV – which is one reason to think the LHC might still discover new physics.

Before thinking about what could explain the light Higgs mass, it is interesting to ask whether there are other relevant operaotrs in the SM. Fermion masses in the SM arise from dimension 4 operators (Yukawa interactions) and are hence marginal above the weak scale. You might think they are relevant below the weak scale, appearing as dimension 3 operators $m\bar\psi\psi$ – but that is misleading. Since a fermion mass term breaks the chiral symmetry $\psi \to e^{i\theta\gamma_5}\psi$, a fermion mass can only be renormalized multiplicatively: any divergent diagram must contain chiral symmetry breaking, and hence must be proportional to m. Thus renormalizations of the fermion mass are not additive: ($\delta m \sim \alpha\Lambda$) but multiplicative: ($\delta m \sim m \ln \Lambda$). Thus light fermion masses are natural since a logarithmic dependence on the cutoff is very weak and the cutoff may be exponentially larger than the fermion mass without requiring fine tuning. Similarly, since gauge boson masses break gauge symmetry, they too do not receive additive mass contributions. However there is one other relevant operator which is problematic: the operator 1, the vacuum energy or cosmological constant. This operator's coefficient receives additive contributions proportional to Λ^4, while we know that the cosmological constant needs to be $\sim (10^{-3}\text{ eV})^4$ to be consistent with cosmological observations – a scale much lower than many known physical scales. A number of new TeV physics scenarios have been proposed to solve or partially solve the Higgs fine-tuning problem: technicolor, supersymmetry, composite Higgs, extra dimensions... however no one has found a dynamical theory to explain the small cosmological constant.

It could be that new physics is around the corner, but one has to wonder whether the naturalness argument isn't missing something, especially because we see a small cosmological constant, and we see a light Higgs mass, but we don't see a host of not-very-irrelevant higher dimension operators that one might expect to be generated by new physics, and whose effects we would expect to have seen already if its scale were low. Various ideas have been suggested for alternatives to naturalness. One popular one is the anthropic principle, the idea that there are many places in the universe with different parameters, most of which are "natural" but in which life is impossible. Therefore we exist in those very peculiar fine-tuned places where life *is* possible and we shouldn't worry that it looks like a bizarre world. To make these arguments sensible you have to (i) have a UV theory for the possible values and correlations between parameters (for example: is it possible to find a place where the up quark is heavy and the down quark is light, or do they have to scale together?), as well as a sensible theory for the a priori probability distribution that they take; then (ii) one has to have a good understanding about how these parameters affect our existence. I have only seen two examples where these two criteria are met at all: anthropic arguments for the cosmological constant [4], and anthropic arguments for the axion in an inflationary universe (see [5] and references therein). And in fact, the anthropic solution is the only plausible explanation for the small cosmological constant that has been proposed. I will return to the anthropic axion in my last lecture.

Another idea is that the world is fine-tuned because of its dynamical evolution. Whereas a creationist might say that human eye is a miraculous finely-tuned apparatus, the evolutionist says it is the product of a billion years of evolution. Perhaps the light Higgs is due to some dynamical path that the universe has taken since the Big Bang? A very creative idea along these lines recently appeared in [6], again involving the axion. See also the parable I have reprinted below (but will not include in my lecture) from my 1997 TASI lectures; I wrote that after being at a conference where I thought that SUSY advocates were being way too smug about the theory!

3.3 Aside – a parable from TASI 1997

I used to live in San Diego near a beach that had high cliffs beside it. The cliffs were composed of compressed sand, and sand was always sprinkling down to the beach below.

At the base of these cliffs there was always a little ramp of sand. One day I was walking down the beach with a physicist friend of mine, and she remarked on the fact that each of these ramps of sand was at precisely the same angle.

"How peculiar!" she said, and I had to agree, but thought no further of it. However, she had a more inquiring mind than I, and called me up that evening:

"I've been conducting an experiment," she said. "I take a box of sand and tilt it until it avalanches, which occurs at an angle θ_c. You won't believe it, but θ_c is precisely the same angle as the ramps of sand we saw at the beach! Isn't that amazing?"

"Indeed," I said. "Apparently someone has performed the same experiment as you, and has sent someone out to adjust all the sand piles to that interesting angle. Perhaps it's the Master's project of some Fine Arts student."

"That's absurd!" she said. "I could believe it if the artist did this *once*, but just think — the wind is always blowing sand onto some piles and off of other piles! The artist would have to fix the piles continuously, day and night!"

As unlikely as this sounded, I had to insist that there seemed to be no alternative. However, the next morning when we met again, my friend was jubilant.

"I figured out what is going on!" she exclaimed. "I have deduced the existence of *swind*. Every time the wind blows and moves the sand, swind blows and moves it back! I call this 'Swindle Theory'. "

"That's absurd!" I exclaimed. "Wind is made of moving molecules, what in the world is swind made of, and why haven't we seen it?"

"Smallecules," she replied, "they're too small to see."

"But you still need the art student to come by in the beginning and fix the sand at precisely the angle θ_c, right?" I asked.

"True"

"So why should I believe in Swindle Theory? You've hardly explained anything!"

"Well look," she retorted, "see how beautiful the Navier-Stokes equations become when generalized to include swind?"

Indeed, the equations were beautiful, so beautiful that I felt compelled to believe in Swindle Theory, although I occasionally still have my doubts...

3.4 Landau liquid versus BCS instability

We have discussed irrelevant operators and rare symmetry violating processes; relevant operators and the naturalness problem; let us now turn to the fascinating case of asymptotically free marginal interactions.

In our discussion of 2D quantum mechanics we encountered asymptotic freedom and the dynamical generation of an exponentially small scale in the IR. This is a possibility in theories with marginal interactions that are pushed into relevancy by small radiative corrections; however it is known that for relativistic QFTs it only actually happens in nonabelian gauge theories – the most famous example being QCD. Asymptotic freedom explains why $\Lambda_{\rm QCD}$ is naturally so much smaller than the GUT or Planck scales. However, the same physics is responsible for the large Cooper pairs found in superconducting materials, which I describe here, following the work of Polchinski [7]. I like this example because it emphasizes that you should not have a fixed idea what an EFT has to look like, but should be able to adapt its use to widely different theories.

A condensed matter system can be a very complicated environment; there may be various types of ions arranged in some crystalline array, where each ion has a complicated electron shell structure and interactions with neighboring ions that allow electrons to wander around the lattice. Nevertheless, the low energy excitation spectrum for many diverse systems can be described pretty well as a "Landau liquid", whose excitations are fermions with a possibly complicated dispersion relation but no interactions. Why this is the case can be simply understood in terms of effective field theories, modifying the scaling arguments to account for the existence of the Fermi surface.

Let us assume that the low energy spectrum of the condensed matter system has fermionic excitations with arbitrary interactions above a Fermi surface characterized by the fermi energy ϵ_F ; call them "quasi-particles". Ignoring interactions at first, the action can be written as

$$S_{free} = \int dt \int d^3p \sum_{s=\pm\frac{1}{2}} \Big[\psi_s(p)^\dagger i\partial_t\psi_s(p) - (\epsilon(p) - \epsilon_F)\psi_s^\dagger(p)\psi_s(p)\Big] \tag{96}$$

where an arbitrary dispersion relation $\epsilon(p)$ has been assumed.

To understand how important interactions are, we wish to repeat some momentum space version of the scaling arguments I introduced in the first lecture. In the present case, a low energy excitation corresponds to one for which $(\epsilon(p) - \epsilon_F)$ is small, which means that $\mathbf{p}$ must lie near the Fermi surface. So in momentum space, we will want our scaling variable to vary the distance we sit from the Fermi surface, and not to rescale the overall momentum $\mathbf{p}$. After all, here a particle with $\mathbf{p} = 0$ is a high energy excitation.

This situation is a bit reminiscent of HQET where we wrote $p_\mu = mv_\mu + k_\mu$, with k_μ being variable that is scaled, measuring the "off-shellness" of the heavy quark. So in the present case we will write the momentum as

$$\mathbf{p} = \mathbf{k} + \boldsymbol{\ell} \tag{97}$$

where $\mathbf{k}$ lies on the Fermi surface and $\boldsymbol{\ell}$ is perpendicular to the Fermi surface (shown in Fig. 10 for a spherical Fermi surface). Then $\boldsymbol{\ell}$ is the quantity we vary in experiments and so we define the dimension of operators by how they must scale so that the theory is unchanged when we change $\boldsymbol{\ell} \to r\boldsymbol{\ell}$. If an object scales as r^n, then we say it has dimension n. Then $[k] = 0$, $[\ell] = 1$, and $[\int d^3p = \int d^2k d\ell] = 1$. And if we define the Fermi velocity as $\mathbf{v}_F(\mathbf{k}) = \boldsymbol{\nabla}_k\epsilon(\mathbf{k})$, then for $\ell \ll k$,

$$\epsilon(\mathbf{p}) - \epsilon_F = \boldsymbol{\ell}\cdot\mathbf{v}_F(\mathbf{k}) + \mathcal{O}(\ell^2) \ , \tag{98}$$

and so $[\epsilon - \epsilon_f] = 1$ and $[\partial_t] = 1$. Given that the action eq. (96) isn't supposed to change under this scaling,

$$[\psi] = -\frac{1}{2} \ . \tag{99}$$

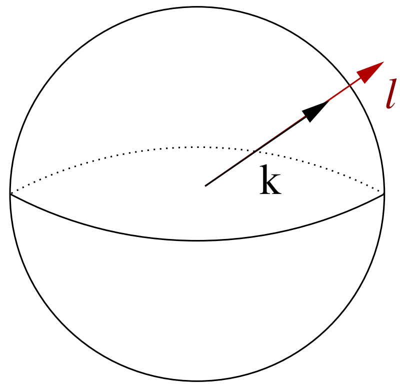

Figure 10: *The momentum* $\mathbf{p}$ *of an excitation is decomposed as* $\mathbf{p} = \mathbf{k} + \boldsymbol{\ell}$, *where* $\mathbf{k}$ *lies on the Fermi surface (which does not have to be a sphere), and* $\boldsymbol{\ell}$ *is perpendicular to the Fermi surface. Small* $|\boldsymbol{\ell}|$ *corresponds to a small excitation energy.*

Now consider an interaction of the form

$$S_{int} = \int dt \int \prod_{i=1}^{4} (d^2\mathbf{k}_i d\ell_i)\delta^3(\mathbf{P}_{tot})C(\mathbf{k}_1, \ldots, \mathbf{k}_4)\psi_s^\dagger(\mathbf{p}_1)\psi_s(\mathbf{p}_2)\psi_{s'}^\dagger(\mathbf{p}_3)\psi_{s'}(\mathbf{p}_4) \ . \qquad (100)$$

This will be relevant, marginal or irrelevant depending on the dimension of C. Apparently we have the scaling dimension $[\delta^3(\mathbf{P}_{tot})C] = -1$. So how does the δ function by itself scale? For generic $\mathbf{k}$ vectors, $\delta(\mathbf{P}_{tot})$ is a constraint on the $\mathbf{k}$ vectors that doesn't change much as one changes ℓ, so that $[\delta^3(\mathbf{P}_{tot})] = 0$. It follows that $[C] = -1$ and that the four fermion interaction is irrelevant...and that the system is adequately described in terms of free fermions (with an arbitrary dispersion relation). This is why Landau liquid theory works and is related to why in nuclear physics Pauli blocking allows a strongly interacting system of nucleons to have single particle excitations.

This is not the whole story though, or else superconductivity would never occur. Let us look more closely at the conclusion above $[\delta^3(\mathbf{P}_{tot})] = 0$. Consider the case when all the $\boldsymbol{\ell}_i = 0$, and therefore the $\mathbf{p}_i = \mathbf{k}_i$ and lie on the Fermi surface. Suppose we fix the two incoming momenta $\mathbf{k}_1$ and $\mathbf{k}_2$. The $\delta^3(\mathbf{P}_{tot})$ then constrains the sum $\mathbf{k}_3 + \mathbf{k}_4$ to equal $\mathbf{k}_1 + \mathbf{k}_2$, which generically means that the vectors $\mathbf{k}_3$ and $\mathbf{k}_4$ are constrained up to point to opposite points on a circle that lies on the Fermi surface (Fig. 11b). Thus one free parameter remains out of the four independent parameters needed to describe the vectors $\mathbf{p}_3$ and $\mathbf{p}_4$. So we see that in this generic case, $\delta^3(\mathbf{P}_{tot})$ offers three constraints, even when $\boldsymbol{\ell}_i = 0$. Therefore $\delta^3(\mathbf{P}_{tot}) = \delta^3(\mathbf{K}_{tot})$ is unaffected when $\boldsymbol{\ell}$ is scaled, and we find the above assumption $[\delta^3(\mathbf{P}_{tot})] = 0$ to be true, and Landau liquid theory is justified.

However now look at the special case when the collisions of the incoming particles are nearly head-on, $\mathbf{k}_1 + \mathbf{k}_2 = 0$. Now $\delta^3(\mathbf{P}_{tot})$ constrains the outgoing momenta to satisfy $\mathbf{k}_3 + \mathbf{k}_4 = 0$. But as seen in Fig. 11a, this only constrains $\mathbf{k}_3$ and $\mathbf{k}_4$ to lie on opposite sides of the Fermi surface. Thus $\delta^3(\mathbf{P}_{tot})$ seems to be only constraining two degrees of freedom, and could be written as $\delta^2(\mathbf{k}_3 + \mathbf{k}_4)\delta(0)$. This singularity obviously arose because the set the $\boldsymbol{\ell}_i = 0$, and so $\delta^3(\mathbf{P}_{tot})$ must be scaling as an inverse power of $\boldsymbol{\ell}$. For nonzero $\boldsymbol{\ell}$ the $\delta(0)$

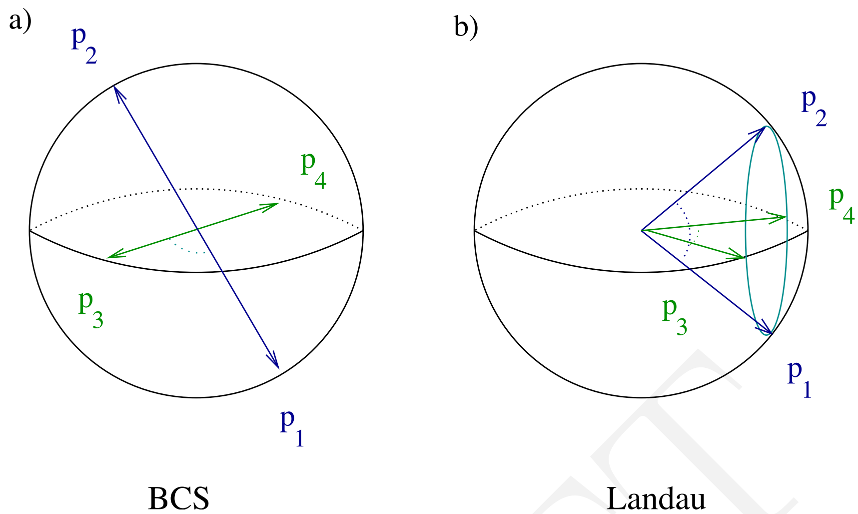

Figure 11: *Fermions scattering near the Fermi surface. (a) Head-on collisions: With* $\mathbf{k}_1 + \mathbf{k}_2 = 0$, *only two degrees of freedom in the outgoing momenta* $\mathbf{k}_3$ *and* $\mathbf{k}_4$ *are constrained, as they can point to any two opposite points on the Fermi surface. (b) The generic Landau liquid case, where the incoming particles do not collide head-on, and three degrees of freedom in the outgoing momenta* $\mathbf{k}_3$ *and* $\mathbf{k}_4$ *are constrained, as they must point to opposite sides of a particular circle on the Fermi surface. Figure from ref. [8], courtesy of Thomas Schäfer.*

becomes $\delta(\boldsymbol{\ell}_{tot})$, and as a result, the δ function scales with $\boldsymbol{\ell}^{-1}$: $[\delta^3(\mathbf{P}_{tot})] = -1$. But since $[\delta^3(\mathbf{P}_{tot})C] = -1$, it follows that for these head-on collisions we must have $[C] = 0$, and the interaction is marginal!

We have already seen that quantum corrections make a marginal interaction either irrelevant or relevant; it turns out that for an attractive interaction, the interaction becomes relevant, and for a repulsive interaction, it becomes irrelevant, just as we found for the δ-function interaction in two dimensions.

Therefore, an attractive contact interaction between quasiparticles becomes strong exponentially close to the Fermi surface (since the coupling runs logarithmically), and can lead to pairing and superconductivity just as the asymptotically free QCD coupling leads to quark condensation and chiral symmetry breaking. The BCS variational calculation shows that the pairing instability does indeed occur; the effective field theory analysis explains why Cooper pairs are exponentially large compared to the lattice spacing in superconductors. The difference between superconductors and metals that behave as Landau liquids depends on the competition between Coulomb repulsion and phonon mediated attraction in the particular material, which determines the sign of the C coupling.

3.5 Problems for lecture III

III.1) A small fermion mass can be considered natural, in contrast to a small scalar mass. This has to do with the fact that if a fermion becomes massless, usually the symmetry of the theory is enhanced by a $U(1)$ chiral symmetry $\psi \to e^{i\alpha\gamma_5}\psi$. Thus at $m = 0$, there cannot be any renormalization of the fermion mass. A corollary is that at nonzero mass m, any renormalization must be proportional to m. Can you explain why this makes the fermion mass behave like a marginal operator rather than a relevant one? Can you construct an example of a theory where it is *not* natural to have a light fermion?

4 Chiral perturbation theory

4.1 Chiral symmetry in QCD

QCD is the accepted theory of the strong interactions. At large momentum transfer, as in deep inelastic scattering processes and the decays of heavy particles such as the Z, the theory is perturbative due to asymptotic freedom. The flip side is that in the infrared, the theory becomes nonperturbative. This is good in the sense that we know that the light hadrons don't look at all like a collection of quarks weakly interacting via gluon exchange. But it does mean that QCD is not of much help in quantitatively understanding this phenomenology without resorting to lattice QCD and a computer. However, there does exist an effective field theory which is very powerful for treating analytically the interactions of the lightest hadrons, the pseudoscalar octet, consisting of the π, K, $\overline{K}$ and η.

The reason that the pseudoscalar octet mesons are lighter is because they are the pseudo-Goldstone bosons (PGBs) that arise from the spontaneous breaking of an approximate symmetry in QCD.

Consider the QCD Lagrangian, keeping only the three lightest quarks, u, d and s:

$$\mathcal{L} = \sum_{i=1}^{3} \left(\bar{q}_i i \not{D} q_i - m_i \bar{q}_i q_i\right) - \frac{1}{2}\mathrm{Tr}G_{\mu\nu}G^{\mu\nu} \ , \tag{101}$$

where $D_\mu = \partial_\mu + igA_\mu$ is the covariant derivative, $A_\mu = A_\mu^a T_a$ are the eight gluon fields with T_a being $SU(3)$ generators in the 3 representation, and $G_{\mu\nu}$ being the gluon field strength. Note that if I write the kinetic term in terms of right-handed and left-handed quarks, projected out by $(1 \pm \gamma_5)/2$ respectively, then the kinetic term may be written as

$$\sum_i \bar{q}_i i\not{D} q_i = \sum_i \left(\bar{q}_{Li} i\not{D} q_{Li} + \bar{q}_{Ri} i\not{D} q_{Ri}\right) \ . \tag{102}$$

This term by itself evidently respects a $U(3)_L \times U(3)_R$ symmetry, where I rotate the three flavors of left-handed and right-handed quarks by independent unitary matrices. One combination of these transformations, the $U(1)_A$ transformation where $q_i \to e^{i\alpha\gamma_5} q_i$ is in fact not a symmetry of the quantum theory, due to anomalies; it is a symmetry of the action but not of the measure of the path integral. This leaves us with a $U(1)_V \times SU(3)_L \times SU(3)_R$ symmetry. The $U(1)_V$ is just baryon number, under which both left- and right-handed quarks of all flavors pick up a common phase. The remaining $SU(3)_L \times SU(3)_R$ symmetry, under which $q_{Li} \to L_{ij} q_{Lj}$ and $q_{Rj} \to R_{ij} q_{Rj}$, where R and L are independent $SU(3)$ matrices, is called "chiral symmetry".

$SU(3)_L \times SU(3)_R$ is not an exact symmetry of QCD, however. The quark mass terms may be written as

$$\sum_i m_i \bar{q}_i q_i = \sum_{i,j} \bar{q}_{Ri} M_{ij} q_{Lj} + h.c. \ , \qquad M = \begin{pmatrix} m_u & & \\ & m_d & \\ & & m_s \end{pmatrix} , \tag{103}$$

where the quark masses m_i are called "current masses", not to be confused with the much bigger constituent quark masses in the quark model. Since the mass term couples left- and right-handed quarks, it is not invariant under the full chiral symmetry. Several observations:

- Note that *if* the mass matrix M were a dynamical field, transforming under $SU(3)_L \times SU(3)_R$ as

$$M \to RML^\dagger \ , \tag{104}$$

then the Lagrangian *would* be chirally invariant. Thinking of the explicit breaking of chiral symmetry as being due to spontaneous breaking due to a field M which transforms as above makes it simple to understand how M must appear in the effective theory, which will have to be chirally invariant given the above transformation. This is called treating M as a "spurion".

- The symmetry is broken to the extent that $M \neq RML^\dagger$. Since m_u and m_d are much smaller than m_s, $SU(2)_L \times SU(2)_R$ is not broken as badly as $SU(3)_L \times SU(3)_R$;
- If all three quark masses were equal but nonzero, then QCD would respect an exact $SU(3)_V \subset SU(3)_L \times SU(3)_R$ symmetry, where one sets $L = R$. This is the $SU(3)$ symmetry of Gell-Mann.
- Since $m_d - m_u$ is small, $SU(2)_V \subset SU(3)_V$, where $L = R$ and they act nontrivially only on the u and d quarks, is quite a good symmetry...also known as isospin symmetry.
- Independent vectorlike phase rotations of the three flavors of quarks are exact symmetries...these three $U(1)$ symmetries are linear combinations of baryon number, I_3 isospin symmetry, and Y (hypercharge). The latter two are violated by the weak interactions, but not by the strong or electromagnetic forces.

We know that this still is not the whole story though. An added complication is that the QCD vacuum spontaneously breaks the chiral $SU(3)_L \times SU(3)_R$ symmetry down to Gell-Mann's $SU(3)_V$ via the quark condensate:

$$\langle 0|\bar{q}_{Rj} q_{Li}|0\rangle = \Lambda^3 \delta_{ij} \ , \tag{105}$$

which transforms as a $(3, \bar{3})$ under $SU(3)_L \times SU(3)_R$. Here Λ has dimensions of mass. If one redefines the quark fields by a chiral transformation, the Kronecker δ-function above gets replaced by a general $SU(3)$ matrix,

$$\langle 0|\bar{q}_{Rj} q_{Li}|0\rangle \to L_{im} \langle 0|\bar{q}_{Rn} q_{Lm}|0\rangle R^\dagger_{nj} = \Lambda^3 \, (LR^\dagger)_{ij} \equiv \Lambda^3 \, \Sigma_{ij} \ . \tag{106}$$

Note that Σ is an $SU(3)$ matrix. If $L = R$ (an $SU(3)_V$ transformation, Gell-Mann's $SU(3)$), then $\Sigma_{ij} = \delta_{ij}$ which shows that the condensate leaves unbroken the $SU(3)_V$ symmetry. For $L \neq R$, Σ_{ij} represents a different vacuum from eq. (105), and if it wasn't for the explicit breaking of $SU(3)_L \times SU(3)_R$ by quark masses in the QCD Lagrangian, these two different vacua would be degenerate. By Goldstone's theorem therefore, there would have to be eight exact Goldstone bosons — one for each of the eight broken generators — corresponding to long wavelength, spacetime dependent rotations of the condensate. We will parametrize these excitations by replacing

$$\Sigma \to \Sigma(x) \equiv e^{2i\boldsymbol{\pi}(x)/f} \ , \qquad \boldsymbol{\pi}(x) = \pi_a(x) T_a \tag{107}$$

where the T_a are the $SU(3)$ generators ($a = 1, \ldots, 8$) in the defining representation normalized to

$$\mathrm{Tr} T_a T_b = \frac{1}{2}\delta_{ab} \ , \tag{108}$$

f is a parameter with dimension of mass which we will relate to the pion decay constant f_π, and the π_a are eight mesons transforming as an octet under $SU(3)_V$. This meson octet corresponds to long wavelength excitations of the vacuum, and the interactions of these mesons are represented by the chiral Lagrangian; it is an EFT because the mesons are the lightest excitations of QCD, and we ignore all heavier states, such as the ρ meson. In order to construct this theory we are going to be guided by symmetry: under an $SU(3)_L \times SU(3)_R$ transformation we have

$$\Sigma(x) \to L\Sigma(x)R^\dagger \ , \tag{109}$$

for arbitrary $SU(3)$ matrices L and R, and we are going to require that the chiral Lagrangian be invariant under this transformation. What makes the story interesting is that $SU(3)_L \times SU(3)_R$ is not an exact symmetry, but i broken by nonzero quark masses and by the electric charges of the quarks, and we will have to incorporate those effects.

If you are somewhat overwhelmed by this amazing mix of symmetries that are gauged, global, exact, approximate, spontaneously broken and anomalous (and each individual symmetry has usually more than one of these attributes at the same time), rest assured that it took a decade and many physicists to sort it all out (the 1960's).

4.2 Quantum numbers of the meson octet

It is useful to use the basis for $SU(3)$ generators $T_a = \frac{1}{2}\lambda_a$, where λ_a are Gell Mann's eight matrices. The meson matrix $\boldsymbol{\pi} \equiv \pi_a T_a$ appearing in the exponent of Σ is a traceless 3×3 matrix. We know that under and $SU(3)_V$ transformation $L = R = V$,

$$\Sigma \to V\Sigma V^\dagger = e^{2iV\boldsymbol{\pi}V^\dagger/f} \ , \tag{110}$$

implying that under $SU(3)_V$ the mesons transform as an octet should, namely

$$\boldsymbol{\pi} \to V\boldsymbol{\pi}V^\dagger \ . \tag{111}$$

Then by restricting V to be an I_3 (T_3) or a Y (T_8) rotation we can read off the quantum numbers of each element of the $\boldsymbol{\pi}$ matrix and identify them with real particles:

$$\boldsymbol{\pi} = \frac{1}{\sqrt{2}} \begin{pmatrix} \frac{\pi^0}{\sqrt{2}} + \frac{\eta}{\sqrt{6}} & \pi^+ & K^+ \\ \pi^- & -\frac{\pi^0}{\sqrt{2}} + \frac{\eta}{\sqrt{6}} & K^0 \\ K^- & \overline{K}^0 & -\frac{2\eta}{\sqrt{6}} \end{pmatrix} \tag{112}$$

An easy way to understand the normalization is to check that

$$\mathrm{Tr}(\boldsymbol{\pi}\boldsymbol{\pi}) = \frac{1}{2}\sum_a (\pi_a)^2 = \frac{1}{2}(\pi^0)^2 + \frac{1}{2}\eta^2 + \pi^+\pi^- + K^+K^- + K^0\overline{K}^0 \ . \tag{113}$$

which is a manifestly $SU(3)_V$ invariant operator.

4.3 The chiral Lagrangian

4.3.1 The leading term and the meson decay constant

We are now ready to write down the effective theory of excitations of the chiral condensate (the chiral Lagrangian), ignoring all the other modes of QCD. This is analogous to the quantization of rotational modes of a diatomic molecule, ignoring the vibrational modes. We are guided by two basic principles of effective field theory: (i) The chiral Lagrangian must exhibit the same approximate chiral symmetry as QCD, which means that it must be invariant under $\Sigma \to L\Sigma R^\dagger$ for arbitrary $SU(3)_L \times SU(3)_R$ matrices L, R. We will also be able to incorporate symmetry breaking effects by including the matrix M, requiring that the chiral Lagrangian be invariant under the chiral symmetry if M were to transform as in eq. (104). (ii) The other principle is that the effective theory be an expansion of local operators suppressed by powers of a cutoff Λ, which is set by the scale of physics we are ignoring, such as the ρ, $K*$, ω, and η' mesons (with masses $m_\rho = 770$ MeV, $m_{K^*} = 892$ MeV, $m_\omega = 782$ MeV and $m_{\eta'} = 958$ MeV). In practice, the cutoff seems to be at $\Lambda \simeq 1$ GeV in many processes. Our calculations will involve an expansion in powers of momenta or meson masses divided by Λ. This cutoff is to be compared with $m_{\pi^\pm} = 140$ MeV, $m_{K^+} = 494$ MeV and $m_\eta = 548$ MeV. For purely mesonic processes, meson masses always appear squared, which helps. Nevertheless, one can surmise that chiral perturbation theory will work far better for pions than kaons or the η. This is a reflection of the fact that $SU(2)_L \times SU(2)_R$ is a much better symmetry of QCD than $SU(3)_L \times SU(3)_R$.

The lowest dimension chirally symmetric operator we can write down is

$$\mathcal{L}_0 = \frac{f^2}{4}\mathrm{Tr}\partial\Sigma^\dagger\partial\Sigma = \mathrm{Tr}\partial\boldsymbol{\pi}\partial\boldsymbol{\pi} + \frac{1}{3f^2}\mathrm{Tr}[\partial\boldsymbol{\pi}, \boldsymbol{\pi}]^2 + \dots \tag{114}$$

Note that the $f^2/4$ prefactor is fixed by requiring that the mesons have canonically normalized kinetic terms. Thus we have an infinite tower of operators involving a single unknown parameter, f. From the above Lagrangian, it would seem that the only way to determine f is by looking at $\pi\pi$ scattering. However there is a better way: by looking at the charged pion decay $\pi \to \mu\nu$. This occurs through the "semi-leptonic" weak interaction eq. (72), namely the operator

$$\frac{1}{\sqrt{2}}G_F V_{ud}\left(\bar{u}\gamma^\mu(1-\gamma_5)d\right)\left(\bar{\mu}\gamma_\mu(1-\gamma_5)\nu_\mu\right) + \text{h.c.} \tag{115}$$

The matrix element of this operator sandwiched between $|\mu\nu\rangle$ and $\langle\pi|$ factorizes, and the leptonic part is perturbative. We are left with the nonperturbative part,

$$\langle 0|\bar{u}\gamma^\mu(1-\gamma_5)d|\pi^-(p)\rangle \equiv i\sqrt{2}\, f_\pi p^\mu \ . \tag{116}$$

The pion decay constant f_π is determined from the charged pion lifetime to be $f_\pi = 92.4 \pm .25$ MeV.

Even though QCD is nonperturbative, we can easily match this charged current operator onto an operator in the chiral Lagrangian. That is because we can write

$$\bar{u}\gamma^\mu(1-\gamma_5)d = 2\left(j^\mu_{L1} + ij^\mu_{L2}\right) \ , \tag{117}$$

where j^μ_{La} are the eight $SU(3)_L$ currents

$$j^\mu_{La} \equiv \bar{q}\gamma^\mu\left(\frac{1-\gamma_5}{2}\right)T_a q \ . \tag{118}$$

To compute these currents in the effective theory is easy, since we know that under $SU(3)_L$ transformations $\Sigma \to L\Sigma$, or $\delta_{La}\Sigma = iT_a\Sigma$, and can just compute the left-handed currents from the Lagrangian $\mathcal{L}_0$ in eq. (114) using Noethers theorem. The result is:

$$j^\mu_{La} = -i\frac{f^2}{2}\mathrm{Tr}T_a\Sigma^\dagger\partial^\mu\Sigma = f\,\mathrm{Tr}T_a\partial^\mu\boldsymbol{\pi} + O(\boldsymbol{\pi}^2) \ . \tag{119}$$

In particular,

$$2\left(j^\mu_{L1} + ij^\mu_{L2}\right) = 2f\ \mathrm{Tr}\begin{pmatrix}0 & 1 & 0\\ 0 & 0 & 0\\ 0 & 0 & 0\end{pmatrix}\partial^\mu\boldsymbol{\pi} + O(\boldsymbol{\pi}^2) = \sqrt{2}f\partial^\mu\pi^- + O(\pi^2) \ , \tag{120}$$

were I made use of eq. (112). Comparing this equation with eq. (116) we see that to this order,

$$f = f_\pi = 93 \text{ MeV} \ . \tag{121}$$

In general it is not possible to exactly match quark operators with operators in the chiral Lagrangian; it was possible for the semileptonic decays simply because the weak operator factorized into a leptonic matrix element and a hadronic matrix element of an $SU(3)_L$ symmetry current. For a purely hadronic weak decay, such as $K \to \pi\pi$ the four quark operator cannot be factorized, and matching to operators in the chiral Lagrangian involves coefficients which can only be computed on a lattice. Even for these processes the chiral Lagrangian can be predictive, relating weak decays with different numbers of mesons in the final state.

4.3.2 Explicit symmetry breaking

Up to now, I have only discussed operators in the chiral Lagrangian which are invariant. Note that that there are no chirally invariant operators which do *not* have derivatives (other than the operator 1). For example, one cannot write down a chirally invariant mass term for the pions. Recall that without explicit chiral symmetry breaking in the QCD Lagrangian, there would be an infinite number of inequivalent degenerate vacua corresponding to different constant values of the matrix Σ; therefore the energy (and the Lagrangian) can only have operators which vanish when Σ is constant, up to an overall vacuum energy independent of Σ. In fact, rotating $\Sigma \to \Sigma' = \Sigma + id\theta_a\,T_a\,\Sigma$ is an exact symmetry of the theory ($SU(3)_L$), and corresponds to *shifting* the pion fields $\pi_a \to \pi_a + d\theta_a\,f/2 + O(\pi^2)$. Derivative interactions are a result of this shift symmetry. (In the literature, this is called a *nonlinearly realized* symmetry, which is to say, a spontaneously broken symmetry). A theory of massless particles with nontrivial interactions at zero momentum transfer (such as QCD) would suffer severe infrared divergences, and so if the interactions had not been purely derivative, the theory would either not make sense, or would become nonperturbative like QCD.

This all changes when explicit chiral symmetry breaking is included. Now not all vacua are equivalent, the massless Goldstone bosons become massive "pseudo-Goldstone bosons" (PGBs), and acquire nonderivative interactions. In pure QCD, the only sources of explicit chiral symmetry breaking are instantons (which explicitly break the $U(1)_A$ symmetry, and the quark mass matrix. Electromagnetic interactions also introduce chiral symmetry breaking, as do weak interactions.

Quark masses. To include the effect of quark masses, we need to include the mass matrix M, recalling that if it transformed as in eq. (104), then the theory would have to be invariant. Just as with derivatives, each power of M will be accompanied by $1/\Lambda$. The leading operator we can write down is

$$\mathcal{L}_M = \Lambda^2 f^2 \left(\frac{c}{2} \frac{1}{\Lambda} \mathrm{Tr} M\Sigma + \mathrm{h.c.} \right) \equiv \frac{1}{2} f^2 \mathrm{Tr}(\widetilde{\Lambda} M)\Sigma + \mathrm{h.c.} \ , \tag{122}$$

where c is an unknown dimensionless coefficient, and I defined

$$c\Lambda \equiv \widetilde{\Lambda} = O(\Lambda) \ . \tag{123}$$

Expanding to second order in the $\boldsymbol{\pi}$, I get

$$\mathcal{L}_M = -m_\pi^2 \pi^+\pi^- - m_{K^+}^2 K^+K^- - m_{K^0}^2 K^0 \overline{K}^0 - \frac{1}{2} \begin{pmatrix} \pi^0 & \eta \end{pmatrix}_i \left[M_0^2 \right]_{ij} \begin{pmatrix} \pi^0 \\ \eta \end{pmatrix}_j \ , \tag{124}$$

with

$$m_\pi^2 = \widetilde{\Lambda}(m_u + m_d) \ , \qquad m_{K^+}^2 = \widetilde{\Lambda}(m_u + m_s) \ , \qquad m_{K^0}^2 = \widetilde{\Lambda}(m_d + m_s) \ , \tag{125}$$

and

$$M_0^2 = \widetilde{\Lambda} \begin{pmatrix} (m_u + m_d) & (m_u - m_d) \\ (m_u - m_d) & \frac{1}{3}(m_u + m_d + 4m_s) \end{pmatrix} \tag{126}$$

Note that (i) the *squares* of the meson masses are proportional to quark masses; (ii) $\pi^0 - \eta$ mixing is isospin breaking and proportional to $(m_u - m_d)$; (iii) expanding in powers of $(m_u - m_d)$, m_η^2 and $m_{\pi^0}^2$ are given by the diagonal entries of M_0^2, up to corrections of $O\left((m_u - m_d)^2\right)$; (iv) we cannot directly relate quark and meson masses because of the unknown coefficient $\widetilde{\Lambda}$.

Ignoring isospin breaking due to electromagnetism and the difference $m_u \neq m_d$, the masses we have computed obey the Gell-Mann Okuba formula

$$3m_\eta^2 + m_\pi^2 = 4m_K^2 \ . \tag{127}$$

The two sides of the above equation are satisfied experimentally to better than 1% accuracy.

Electromagnetism. To include electromagnetism into the chiral Lagrangian, we have to first go back to QCD and ask what currents out of our eight j^μ_{La} and j^μ_{Ra} couple to the photon. That is easy: the photon couples to the electromagnetic current which can be written as

$$J^\mu_{\rm em} = e\bar{q}\gamma^\mu P_L Q_L q + e\bar{q}\gamma^\mu P_R Q_R q\ , \qquad Q_L = Q_R = \begin{pmatrix} \frac{2}{3} & & \\ & -\frac{1}{3} & \\ & & -\frac{1}{3} \end{pmatrix}\ , \tag{128}$$

a simple linear combination of octet currents. So symmetry determines the covariant derivative in the chiral Lagrangian to be

$$D_\mu \Sigma = \partial_\mu \Sigma - ieA_\mu \left(Q_L\Sigma - \Sigma Q_R\right) \tag{129}$$

since $\Sigma \to L\Sigma R^\dagger$ under $SU(3)_L \times SU(3)_R$. Note that when we set the Σ field to its vacuum value, $\Sigma = 1$, the photon term drops out of the covariant derivative, which is to say that the vacuum does not break electromagnetism spontaneously. Also note that the $Q_{L,R}$ matrices are $SU(3)_L \times SU(3)_R$ spurions: in order to have unbroken chiral symmetry we would need $D_\mu\Sigma$ to have the transformation property $D_\mu\Sigma \to LD_\mu\Sigma R^\dagger$ when $\Sigma \to L\Sigma R^\dagger$, which would require the transformation properties

$$Q_L \to L^\dagger Q_L L^\dagger\ , \qquad Q_R \to RQ_R R^\dagger\ , \tag{130}$$

which is to say that Q_L transforms as part of the adjoint (octet) of $SU(3)_L$ and a singlet under $SU(3)_R$, and conversely for Q_R. We would have muddled this if we had not taken care to distinguish $Q_{L,R}$ from each other from the start, even though for the photon then end up being the same matrix.

If we now want to compute the electromagnetic contribution to the $\pi^+ - \pi^0$ mass splitting to order α we naturally look at the two one-loop diagrams we encounter in scalar QED. These are quadratically divergent, which means they need a counter term which would contribute to the pion mass^2 approximately $\sim \alpha/4\pi\Lambda^2 \sim e^2f^2$. From the transformation properties eq. (130) we see that we can add such a counter term operator to the chiral Lagrangian of the form

$$\mathcal{L}_\alpha = \xi f^4 \frac{\alpha}{4\pi} {\rm Tr} Q_L \Sigma Q_R \Sigma^\dagger \tag{131}$$

where would would expect $c = O(1)$, but which needs to be fit to data or computed using lattice QCD. If we use the $\overline{MS}$ scheme in Landau gauge, them the 1-loop diagrams vanish and we are left only with the direct contribution from the above operator in . Expanding it to second order in meson fields we get

$$\mathcal{L}_\alpha = -\xi f^4 e^2 \frac{2}{f^2} {\rm Tr} Q_L \left[\boldsymbol{\pi}, [\boldsymbol{\pi}, Q_R]\right] = -2\xi e^2 f^2 \left(\pi^+\pi^- + K^+K^-\right) \tag{132}$$

a simple result which says that the meson mass^2 gets shifted by a constant amount proportional to its charge squared.

Thus to leading order in α and the quark masses, our formula for the meson masses take the form

$$m^2_{\pi^+} = \widetilde{\Lambda}(m_u + m_d) + \frac{\alpha}{4\pi}\Delta^2\ , \qquad m^2_{K^+} = \widetilde{\Lambda}(m_u + m_s) + \frac{\alpha}{4\pi}\Delta^2\ , \tag{133}$$

with the neutral particle masses unchanged, where $\Delta^2 = 2\xi(4\pi f)^2$ is a parameter whose value we cannot predict. Following Weinberg, however, we can manipulate our formulas to make predictions for ratios of combinations of meson masses chosen so that the unknown parameters Δ and $\widetilde{\Lambda}$ drop out:

$$\frac{(m_{K^+}^2 - m_{K^0}^2) - (m_{\pi^+}^2 - m_{\pi^0}^2)}{m_{\pi^0}^2} = \frac{m_u - m_d}{m_u + m_d} , \qquad \frac{3m_\eta^2 - m_{\pi^0}^2}{m_{\pi^0}^2} = \frac{4m_s}{m_u + m_d} . \tag{134}$$

Plugging in the measured meson masses, these formulas let us determine the ratios of quark masses in QCD:

$$\frac{m_u}{m_d} \simeq \frac{1}{2} , \qquad \frac{m_d}{m_s} \simeq \frac{1}{20} . \tag{135}$$

To specify the quark masses themselves, one must perform a lattice QCD calculation and designate a renormalization scheme. Lattice simulations typically find m_s renormalized at $\mu = 2$ GeV in the $\overline{MS}$ scheme lies in the $80 - 100$ MeV range, from which one infers from the above ratios $m_d \sim 5$ MeV, $m_u \sim 2.5$ MeV in the same scheme. Evidently most of the mass of baryons and vector mesons does *not* come from the intrinsic masses of the quarks.

4.4 Loops and power counting

What makes the chiral Lagrangian an EFT and not just another model of the strong interactions is that it consists of all local operators consisted with the symmetries of QCD, and that there exists a power counting scheme that allows one to work to a given order, and to be able to make a reliable estimate of the errors arising from neglecting the subsequent order.

Beyond the leading term is an infinite number of chirally invariant operators one can write down which are higher powers in derivatives, as well as operators with more insertions of the quark mass matrix M. The derivative expansion is in powers of ∂/Λ, where Λ is the "chiral symmetry breaking scale", which should be $O(1\text{ GeV})$, since that is the mass scale of all the lowest hadronic resonances *not* included in the theory. This power counting is consistent with the leading operator eq. (114), if you consider the chiral Lagrangian to have a prefactor of $\Lambda^2 f^2$, then even in the leading operator derivatives enter as ∂/Λ:

$$\mathcal{L}_0 = \Lambda^2 f^2 \left(\frac{1}{4\Lambda^2} \mathrm{Tr}\, \partial\Sigma^\dagger\, \partial\Sigma \right) . \tag{136}$$

Since we have found that meson octet masses scale as $m_{\boldsymbol{\pi}}^2 \simeq (\widetilde{\Lambda} M)$, and since for on-shell pions $p^2 \sim m^2$, it follows that one insertion of the quark mass matrix is equivalent to two derivatives in the effective field theory expansion. This leads us to write the chiral Lagrangian as a function of (∂/Λ) and $\widetilde{\Lambda} M/\Lambda^2$. Including electromagnetism is straightforward as well: since a derivative $\partial\Sigma$ becomes a covariant derivative $D_\mu\Sigma = \partial_\mu\Sigma - ieA_\mu\,[Q,\Sigma]$, and so the covariant derivative should come with a $1/\Lambda$ factor, and hence the photon field enters as eA_μ/Λ. Operators arising from electromagnetic loops involve two insertions of the quark charge matrix Q in the proper way, along with a loop factor $\alpha/(4\pi)$. Therefore the action for the chiral theory takes the form

$$S = \int d^4x\, \Lambda^2 f^2 \widehat{\mathcal{L}}\left[\Sigma,\, D_\mu/\Lambda,\, \widetilde{\Lambda} M/\Lambda^2,\, (\alpha/4\pi)Q^2\right] \ , \tag{137}$$

where $\widehat{\mathcal{L}}$ is a dimensionless sum of all local, chirally invariant operators (treating M and Q as spurions), where the coefficient of each term (except $\mathcal{L}_0$) is preceded by a dimensionless coefficient to be fit to experiment... which we expect to be $O(1)$, but which may occasionally surprise us! That last assumption is what allows one to estimate the size of higher order corrections.

It should be clear now in what sense the u, d and s are light quarks and can be treated in chiral perturbation theory, while the c, b and t quarks are not: whether the quarks are light or heavy is relative to the scale Λ, namely the mass scale of resonances in QCD. Since the c has a mass ~ 1.5 GeV there is no sensible way to talk about an approximate $SU(4) \times SU(4)$ chiral symmetry and include D, D_s and η_c mesons in our theory of pseudo-Goldstone bosons[7] Of course, you might argue that the strange quark is sort of heavy and should be left out as well, but if we don't live dangerously sometimes, life is too boring.

4.4.1 Subleading order: the $O(p^4)$ chiral Lagrangian

It is a straightforward exercise to write down subleading operators of the chiral Lagrangian. These are operators of $O(p^4)$, $O(p^2 M)$ and $O(M^2)$, where M is the quark mass matrix. This was first done by Gasser and Leutwyler, and their choice for the set of operators has become standard:

$$\begin{aligned}
\mathcal{L}_{p^4} \quad = \quad & L_1 \left(\mathrm{Tr}\left(\partial_\mu \Sigma^\dagger \partial^\mu \Sigma\right)\right)^2 \\
& + L_2 \mathrm{Tr}\left(\partial_\mu \Sigma^\dagger \partial_\nu \Sigma\right) \mathrm{Tr}\left(\partial^\mu \Sigma^\dagger \partial^\nu \Sigma\right) \\
& + L_3 \mathrm{Tr}\left(\partial_\mu \Sigma^\dagger \partial^\mu \Sigma \partial_\nu \Sigma^\dagger \partial^\nu \Sigma\right) \\
& + L_4 \mathrm{Tr}\left(\partial_\mu \Sigma^\dagger \partial^\mu \Sigma\right) \mathrm{Tr}\left(\chi \Sigma + \mathrm{h.c.}\right) \\
& + L_5 \mathrm{Tr}\left(\left(\partial_\mu \Sigma^\dagger \partial^\mu \Sigma\right)\left(\chi \Sigma + \mathrm{h.c.}\right)\right) \\
& + L_6 \left(\mathrm{Tr}\left(\chi \Sigma + \mathrm{h.c.}\right)\right)^2 \\
& + L_7 \left(\mathrm{Tr}\left(\chi \Sigma - \mathrm{h.c.}\right)\right)^2 \\
& + L_8 \mathrm{Tr}\left(\chi \Sigma \chi \Sigma + \mathrm{h.c.}\right) \ ,
\end{aligned} \tag{138}$$

where $\chi \equiv 2\widetilde{\Lambda} M$, where $\widetilde{\Lambda}$ entered in eq. (122). Additional operators involving $F_{\mu\nu}$ need be considered when including electromagnetism.

Note that according to our power counting, we expect the L_i to be of size

$$L_i \sim \frac{\Lambda^2 f^2}{\Lambda^4} = \frac{f^2}{\Lambda^2} \sim 10^{-2} \ . \tag{139}$$

[7] This does not mean that an effective theory for $D - \pi$ interactions is impossible. However, the D mesons must be introduced as heavy matter fields, similar to the way we will introduce baryon fields later, as opposed to approximate Goldstone bosons.

4.4.2 Calculating loop effects

Now consider loop diagrams in the effective theory. These are often divergent, and so the first issue is how to regulate them. It is easy to show that a momentum cutoff on the pion momentum violates chiral symmetry and should be avoided. Consider the chiral transformation $\Sigma \to L\Sigma R^\dagger$ with $L = \exp i\boldsymbol{\ell}$, $R = \exp i\mathbf{r}$, where ℓ and $\mathbf{r}$ are traceless, hermitian 3×3 matrices, and write $\Sigma = \exp i\boldsymbol{\theta}$, where $\boldsymbol{\theta} = 2\boldsymbol{\pi}/f$. Then

$$e^{i\boldsymbol{\theta}} \to e^{i\boldsymbol{\ell}} e^{i\boldsymbol{\theta}} e^{-i\mathbf{r}} \qquad \Longrightarrow \qquad \boldsymbol{\theta} \to -i \ln \left[e^{i\boldsymbol{\ell}} e^{i\boldsymbol{\theta}} e^{-i\mathbf{r}} \right] \tag{140}$$

where by the Baker-Campbell-Haussdorf theorem the logarithm is a nonlinear function of $\boldsymbol{\theta}$ whenever $\boldsymbol{\ell} \neq \mathbf{r}$. Thus if $\boldsymbol{\theta}$ is a plane wave associated with momentum k, then under a chiral transformation it must turn into a field that depends on all harmonics of k. For this reason a momentum cutoff on the pion momentum clearly violates chiral symmetry. While it is possible to restore the symmetry with appropriately chosen counterterms, by far the simplest regularization method is dimensional regularization with a mass independent subtraction scheme, such as MS.

Recall that one of the lessons we learned from matching our toy scalar model in powers of $\hbar$ was that an EFT really has potentially two or more expansion parameters: one is in powers of momenta over the cutoff, ∂/Λ which is the main expansion parameter of the EFT; but there are also the coupling constants of the UV theory that underly the size of radiative corrections. The beauty of the construction, especially evident when using the MS scheme, is that the even when perturbation theory fails in the latter, it can hold in the former. That is because in the $\bar{M}S$ scheme operaotrs at a given power in the ∂/Λ expansion only renormalize operators at higher order in ∂/Λ. So even if a loop corrects the coefficient of a higher order operator by $O(1)$, the effects at low momentum are small.

So how big are radiative corrections in the chiral Lagrangian? This is easy to see if we rescale the terms in eq. (137) such that $\hat{x} = \Lambda x$, $\hat{M} = M\widetilde{\Lambda}/\Lambda^2$, $\hat{A}_\mu = A_\mu/\Lambda$, in which case the action takes the form

$$S = \frac{f^2}{\Lambda^2} \int d^4\hat{x}\, \widehat{\mathcal{L}} \left[\Sigma,\, \hat{D}_\mu,\, \hat{M},\, (\alpha/4\pi)Q^2 \right] \ , \tag{141}$$

and so we see that $\hbar$ is multiplied by Λ^2/f^2 in this theory. As we have seen, we get a power of $\hbar$ with each loop, and so an L-loop diagram in the chiral theory will be proportional to

$$\frac{\Lambda^2}{(4\pi f)^2} \ , \tag{142}$$

where I have included the standard $1/16\pi^2$ contribution from a loop. This factor $\frac{\Lambda^2}{(4\pi f)^2}$ controls the size of quantum corrections to operator coefficients in the theory. If we expect perturbative control of the theory to break down completely for momenta at the cutoff, $p \sim \Lambda$ — which is equivalent to saying that the theory for $p \sim \Lambda$ or $m_\pi \sim \Lambda$ the interactions of pions are completely nonperturbative — then we would expect

$$\Lambda \sim 4\pi f_\pi \ . \tag{143}$$

Estimating the size of operator coefficients in the chiral Lagrangian by this relation is called "naive dimensional analysis". It does not have to be true, but seems to work in practice for most operators (a notable exception being those implicated in the $\Delta I = 1/2$ enhancement). In contrast, for QCD at large N_c for example, $f_\pi^2 = O(N_c)$ while $\Lambda = O(1)$ and so radiative corrections in the chiral Lagrangian become small.

The $\overline{MS}$ scheme introduces a renormalization scale μ, usually chosen to be $\mu = \Lambda$. However, unlike with cutoff regularization, one never gets powers of the renormalization scale μ when computing a diagram: μ can only appear in logarithms. Consider, for example, the $O(\pi^4)$ operator from $\mathcal{L}_0$, of the form $\frac{1}{f^2}(\partial\pi)^2\pi^2$, and contract the two pions in $(\partial\pi)^2$; this one-loop graph will renormalize the pion mass. However, since the diagram is proportional to $1/f^2$, and no powers of the renormalization scale μ can appear, dimensional analysis implies that any shift in the pion mass from this graph must be proportional to $\delta m_\pi^2 \sim (m_\pi^0)^4/(4\pi f)^2$, times a possible factor of $\ln(m_\pi/\mu)$, where $(m_\pi^0)^2 \sim \tilde{\Lambda} M$ is the mass squared of the meson at leading order. Ignoring the logarithm, compare with this contribution to the pion mass contribution from the $O(p^4)$ chiral Lagrangian, which yields $\delta m_\pi^2 \sim (\tilde{\Lambda} M)^2/\Lambda^2 \sim (m_\pi^0)^4/\Lambda^2$. We see that as expected, so long as

$$4\pi f_\pi \gtrsim \Lambda \;, \tag{144}$$

then the contribution from the radiative correction from the lowest order operator is comparable to or smaller than the second order tree-level contribution, up to $\ln m_\pi^2/\mu^2$ corrections.

What about the logarithm? Note that $\ln(\Lambda^2/m_\pi^2) \simeq 4$ for $\mu = 1$ GeV. Therefore a term with a logarithm is somewhat enhanced relative to the higher order tree-level contributions. It is therefore common to see in the literature a power counting scheme of the form

$$p^2 > p^4 \ln \mu^2/p^2 > p^4 \ldots \tag{145}$$

which means that in order of importance, one computes processes in the following order:

1. Tree level contributions from the $O(p^2)$ chiral Lagrangian;
2. Radiative corrections to the $O(p^2)$ chiral Lagrangian, keeping only $O(p^4 \ln p^2)$ terms;
3. Tree level terms from the $O(p^4)$ chiral Lagrangian, as well as $O(p^4)$ radiative contributions from the $O(p^2)$ chiral Lagrangian;

and so forth. Keeping the logs and throwing out the analytic terms in step #2 is equivalent to saying that most of the $O(p^4)$ chiral Lagrangian renormalized at $\mu = m_\pi$ would come from running induced by the $O(p^2)$ Lagrangian in scaling down from $\mu = \Lambda$ to $\mu = m_\pi$, and not from the initial values of couplings in the $O(p^4)$ Lagrangian renormalized at $\mu = \Lambda$. This procedure would *not* be reasonable in the large N_c limit) but seems to work reasonably well in the real world.

4.4.3 Renormalization of $\langle 0|\bar{q}q|0\rangle$

As an example of a simple calculation, consider the computation of the ratios of the quark condensates,

$$x = \frac{\langle 0|\bar{u}u|0\rangle}{\langle 0|\bar{s}s|0\rangle} \;. \tag{146}$$

Since the operator $\bar{q}q$ gets multiplicatively renormalized, $\langle 0|\bar{q}_i q_i|0\rangle$ is scheme dependent, but the ratio x is not. The QCD Hamiltonian density is given by $\mathcal{H} = \ldots + \bar{q}Mq + \ldots$, and so it follows from the Feynman-Hellman theorem[8] that

$$\langle 0|\bar{q}_i q_i|0\rangle = \frac{\partial}{\partial m_i}\langle 0|\mathcal{H}|0\rangle = \frac{\partial \mathcal{E}_0}{\partial m_i} \ , \tag{147}$$

where $\mathcal{E}_0$ is the vacuum energy density. We do not know what is $\mathcal{E}_0$, but we do know its dependence one the quark mass matrix; from eq. (124)

$$\mathcal{E}_0 = \text{const.} - \frac{1}{2}f^2\text{Tr}(\widetilde{\Lambda}M)\Sigma + \text{h.c.} + O(M^2\ln M)\Bigg|_{\Sigma_{ij}=\delta_{ij}} = f^2\widetilde{\Lambda}\text{Tr}M + \ldots \ , \tag{148}$$

from which it follows that this scheme

$$\langle 0|\bar{q}_i q_i|0\rangle = \widetilde{\Lambda}f^2 \ , \tag{149}$$

and that in any scheme the leading contributions to x is

$$x = 1 \ . \tag{150}$$

Well good — this is what we started with for massless QCD in eq. (105)! To get the subleading logarithmic corrections, we need to compute the $O(m^2\ln m^2)$ one-loop correction to the vacuum energy. This loop with no vertices's is the Feynman diagram for which Feynman rules don't work – there is no propagator if there are no vertices! As easily seen in a Euclidean path integral, the vacuum energy density in a box of 4-volume VT for a real, noninteracting scalar is just

$$\mathcal{E}_0 = -\frac{1}{VT}\ln\left(\det(-\Box + m^2)\right)^{-1/2} = \frac{1}{VT}\frac{1}{2}\text{Tr}\ln(-\Box + m^2) \ . \tag{151}$$

In $d = (4-2\epsilon)$ Euclidean dimensions this just involves evaluating for each mass eigenstate the integral

$$\frac{\mu^{4-d}}{2}\int\frac{d^dk}{(2\pi)^d}\ln(k^2+m^2) \ . \tag{152}$$

where the prefactor of μ^{4-d} was included to keep the mass dimension to equal 4.

Let us first perform the differentiation with respect to quark mass. Then in this scheme we get the correction

$$\delta\,\langle 0|\bar{q}_i q_i|0\rangle = \frac{1}{2}\sum_{a=\pi,K,\eta}\frac{\partial m_a^2}{\partial m_i}\mu^{4-d}\int\frac{d^dk}{(2\pi)^d}\frac{1}{k^2+m_a^2} \quad \xrightarrow[\overline{MS}]{} -\sum_a\frac{\partial m_a^2}{\partial m_i}\left(\frac{m_a^2\ln m_a^2/\mu^2}{32\pi^2}\right) \tag{153}$$

where a is summed over the meson mass eigenstates, and m_i is the mass of the i^{th} flavor of quark. The final result was arrived at after performing the $\overline{MS}$ subtraction (where you only keep the $\ln m^2$ term in the $\epsilon \to 0$ limit.

[8]The substance of the Feynman-Hellman theorem is that in first order perturbation theory, the wave function doesn't change while the energy does.

To the order we are working, the quark condensate ratios are therefore given by

$$\frac{\langle 0|\bar{q}_i q_i|0\rangle}{\langle 0|\bar{q}_j q_j|0\rangle} = 1 - \frac{1}{32\pi^2\widetilde{\Lambda} f^2} \sum_{a=\pi,K,\eta} m_a^2 \ln m_a^2/\mu^2 \left(\frac{\partial m_a^2}{\partial m_i} - \frac{\partial m_a^2}{\partial m_j}\right) . \tag{154}$$

Using the masses given in eq. (125) and eq. (126), ignoring $\pi^0 - \eta$ mixing, we find

$$x \;=\; \frac{\langle 0|\bar{u}u|0\rangle}{\langle 0|\bar{s}s|0\rangle} = 1 - 3g_\pi + 2g_{K^0} + g_\eta + O(m^4) , \tag{155}$$

where

$$g_P \equiv \frac{1}{32\pi^2 f^2} m_P^2 \ln\left(\frac{m_P^2}{\mu^2}\right) \tag{156}$$

with $P = \pi$, K^+, K^0, η. The answer is μ dependent, since I have neglected to include the $O(p^4)$ Lagrangian contributions at tree-level, and in fact it is precisely those operators that serve as counterterms for the $1/\epsilon$ poles subtracted in $\overline{MS}$. However, in the usual practice of chiral perturbation theory, we take $\mu = \Lambda$, and assume the contributions from the $O(p^4)$ Lagrangian are small compared to the chiral logs I have included. Plugging in numbers with $\mu = 1$ GeV I find

$$g_\pi \simeq -0.028 , \quad g_K \simeq -0.13 , \quad g_\eta \simeq -0.13 \tag{157}$$

implying that $x \simeq 0.70$ — a 30% correction from the leading result $x = 1$. This is typical of any chiral correction that involves the strange quark, since $m_K^2/\Lambda^2 \simeq 25\%$. Corrections to $\langle\bar{u}u\rangle/\langle\bar{d}d\rangle$ will be *much* smaller, since they depend on isospin breaking, of which a typical measure is $(m_{K^0}^2 - m_{K^+}^2)/\Lambda^2 \simeq 0.004$.

4.4.4 Using the Chiral Lagrangian

Chiral Lagrangians have very many applications of interest, and thousands of papers have been written that involve chiral perturbation theory. Some early applications were studies of meson-meson scattering, as well as weak mesonic decays. The baryon octet can be included in the theory; although baryons are heavy, processes like hadronic hyperon decay (e.g. $\Sigma \to N\pi$) do not release much energy to the pion and can be treated sensibly in chiral perturbation theory. The chiral Lagrangian can be used as a mean field theory for discussing the possibility of the Bose-Einstein condensation of pions or kaons in dense matter, such as in neutron stars. A nonperturbative extension of chiral perturbation theory can be used to discuss nucleon-nucleon scattering and the properties of light nuclei. An analogue of the chiral Lagrangian can be constructed for dense superconducting quark matter. The chiral Lagrangian and analogues of it can be useful for discussing possible BSM theories, such as Composite Higgs theories and Technicolor.

4.5 Problems for lecture IV

IV.1) Verify eq. (112).

IV.2) How does Σ transform under P (parity)? What does this transformation imply for the intrinsic parity of the π_a mesons? How does Σ transform under C (charge conjugation)? Which of the mesons are eigenstates of CP, and are they CP even or odd? Recall that under P and C the quarks transform as

$$\begin{aligned} P: &\quad q \to \gamma^0 q \ , \\ C: &\quad q \to C\bar{q}^T \ , \qquad C = C^\dagger = C^{-1} = -C^T \ , \quad C\gamma_\mu C = -\gamma_\mu^T \ , \qquad C\gamma_5 C = \gamma_5 \ . \end{aligned} \tag{158}$$

IV.3) How do we know that c, and hence $\widetilde{\Lambda}$, is positive in eq. (122)? How would the world look different if it were negative? Hint: consider what Σ matrix would minimize the vacuum energy, and its implications for the spectrum of the theory.

5 Effective field theory with baryons

5.1 Transformation properties and meson-baryon couplings

It is interesting to include the baryon octet into the mix. This is reasonable so long as we consider processes with momentum transfer $\ll \Lambda$. Thus we might consider the weak decay $\Lambda \to N\pi$, but not the annihilation $N\overline{N} \to \pi\pi$. There are two separate issues here: (i) How do we figure out how baryons transform under the chiral $SU(3) \times SU(3)$ symmetry so that we can couple them to Σ, and (ii) do we need or desire the Dirac spinor formulation if we are only going to consider low momentum transfer processes? I will address the first question first, using Dirac spinors. Then I will introduce the heavy baryon formalism of Jenkins and Manohar, replacing the Dirac spinors.

First consider a world where the u, d and s are massless. We know that the baryons transform as an octet under the unbroken $SU(3)_V$ symmetry, but how do they transform under $SU(3) \times SU(3)$? The answer is: just about any way you want. To see this, consider a baryon field B written as a 3×3 traceless matrix of Dirac spinors, transforming as an octet under $SU(3)_V$:

$$B \to VBV^\dagger \ . \tag{159}$$

By considering T_3 and T_8 transformations it is possible to determine the form of the matrix B, just as we did for the meson matrix $\boldsymbol{\pi}$:

$$B = \begin{pmatrix} \frac{\Sigma^0}{\sqrt{2}} + \frac{\Lambda}{\sqrt{6}} & \Sigma^+ & p \\ \Sigma^- & -\frac{\Sigma^0}{\sqrt{2}} + \frac{\Lambda}{\sqrt{6}} & n \\ \Xi^- & \Xi^0 & -\frac{2\Lambda}{\sqrt{6}} \end{pmatrix} \ . \tag{160}$$

Now construct the left- and right-handed baryons $B_{R,L} = \frac{1}{2}(1 \pm \gamma_5)B$. Suppose that B transformed as the $(8,1) \oplus (1,8)$ representation under $SU(3) \times SU(3)$, namely

$$B_R \to RB_RR^\dagger \ , \qquad B_L \to LB_LL^\dagger \ . \tag{161}$$

But then I could define

$$B'_R \equiv \Sigma B_R \ , \qquad B'_L = B_L\Sigma \ . \tag{162}$$

The new field B' works equally well as a baryon field. However B' transforms as a $(3, \bar{3})$ under $SU(3) \times SU(3)$:

$$B' \to LB'R^\dagger \ . \tag{163}$$

Note that both B and B' both transform properly as an $SU(3)_V$ octet when $R = L \equiv V$, as in eq. (159). What we are seeing is that when you have massless pions around, you can't tell the difference between a baryon, and a superposition of that baryon with a bunch of zero momentum massless pions, and yet the two will have different $SU(3) \times SU(3)$ transformation properties. This is not a problem — rather it is liberating. It means we can choose whatever $SU(3) \times SU(3)$ transformation rule we wish for the baryons, so long as eq. (159) still holds.

While the basis eq. (161) looks appealing, it has its drawbacks. For example it allows the interaction

$$m_0 \mathrm{Tr}\overline{B}_L \Sigma B_R + \text{h.c.} = m_0 \mathrm{Tr}\overline{B}B + \frac{2m_0}{f}\mathrm{Tr}\overline{B}i\gamma_5 \boldsymbol{\pi} B + \dots \tag{164}$$

which makes it *look* like the pion can have non-derivative couplings...which ought to be impossible for a Goldstone boson. In fact, Dirac's analysis for nonrelativistic spinors shows that the γ_5 coupling is in fact a derivative coupling at low momentum transfer, made obscure.

A better basis is the following. Define

$$\xi = e^{i\boldsymbol{\pi}/f} = \sqrt{\Sigma}\ . \tag{165}$$

One can show that under an $SU(3)\times SU(3)$ transformation

$$\xi \to L\xi U^\dagger = U\xi R^\dagger\ , \tag{166}$$

where U is a uniquely defined matrix which is in general a function of the constant L and R matrices characterizing the $SU(3)\times SU(3)$ transformation, as well as the $\boldsymbol{\pi}(x)$ field. For the special case of $SU(3)_V$ transformations, $R = L = V$ and it is easy to show that $U = V$ as well, independent of $\boldsymbol{\pi}$. For more general $SU(3)\times SU(3)$ transformations $U(\boldsymbol{\pi})$ is a mess, but we will not need to know its exact form. Now if we take the basis eq. (161) and replace $B_L \to \xi^\dagger B_L \xi$, $B_R \to \xi B_R \xi^\dagger$, we get a new basis where left- and right-handed components of B transform the same way, namely

$$B \to UBU^\dagger\ . \tag{167}$$

Given the above transformation rule, B cannot couple directly to Σ, but we can define the axial and vector currents and chiral covariant derivative:

$$\begin{aligned}
A_\mu &\equiv \frac{i}{2}\left(\xi^\dagger\partial_\mu\xi - \xi\partial_\mu\xi^\dagger\right) \xrightarrow[SU(3)\times SU(3)]{} UA_\mu U^\dagger\ ,\\
V_\mu &= \frac{1}{2}\left(\xi^\dagger\partial_\mu\xi + \xi\partial_\mu\xi^\dagger\right) \xrightarrow[SU(3)\times SU(3)]{} UV_\mu U^\dagger + U\partial_\mu U^\dagger\ ,\\
D_\mu B &\equiv \left(\partial_\mu B + [V_\mu, B]\right) \xrightarrow[SU(3)\times SU(3)]{} U(D_\mu B)U^\dagger\ .
\end{aligned} \tag{168}$$

Note that D_μ, called the "chiral covariant derivative" looks like a gauge govariant derivative with the meson vector current V_μ playing the role of the gauge field.

Armed with this formalism we can write down an effective theory for meson-baryon interactions, whose first few chirally symmetric terms are

$$\mathcal{L}_1 = \mathrm{Tr}\overline{B}(i\gamma^\mu D_\mu - m_0)B - D\,\mathrm{Tr}\overline{B}\gamma^\mu\gamma_5\{A_\mu, B\} - F\,\mathrm{Tr}\overline{B}\gamma^\mu\gamma_5[A_\mu, B]\ . \tag{169}$$

Note that the common octet mass m_0 is chirally symmetric, independent of the quark masses. The parameters D and F are the traditional choice for couplings of baryons to the axial current.

Expanding A_μ and V_μ in the meson fields,

$$A_\mu = -\frac{1}{f}\partial_\mu \boldsymbol{\pi} + O(\boldsymbol{\pi}^3)\ , \qquad V_\mu = \frac{1}{2f^2}\boldsymbol{\pi}\overleftrightarrow{\partial}_\mu \boldsymbol{\pi} + O(\boldsymbol{\pi}^4)\ . \tag{170}$$

For non-relativistic baryons, the Dirac analysis implies that $\overline{B}\gamma^0 B$ and $\overline{B}\vec{\gamma}\gamma_5 B$ are big, equal to 1 and $\vec{S}$ (the baryon spin) respectively; in contrast, the bilinears $\overline{B}\vec{\gamma}B$ and $\overline{B}\gamma^0\gamma_5 B$ are small, given by $\vec{q}/m_0$ and $(\vec{S}\cdot\vec{q})/m_0$ respectively, where $\vec{q}$ is the 3-momentum transfer. Therefore the leading meson-baryon interactions for nonrelativistic baryons (written as 2-component spinors) is

$$\frac{1}{2f^2}\,\mathrm{Tr}\left(B^\dagger[\boldsymbol{\pi}\dot{\boldsymbol{\pi}} - \dot{\boldsymbol{\pi}}\boldsymbol{\pi}, B]\right) + D\,\mathrm{Tr}\left(B^\dagger\vec{\sigma}\cdot\{\vec{\nabla}\boldsymbol{\pi}, B\}\right) + F\,\mathrm{Tr}\left(B^\dagger\vec{\sigma}\cdot\left[\vec{\nabla}\boldsymbol{\pi}, B\right]\right)\ . \tag{171}$$

The vector current interaction is also called the Weinberg-Tomazawa term; it does not involve any unknown parameters and is required by chiral symmetry. The axial current interaction involves two new couplings D and F that may be fit to semi-leptonic baryon decay, using the same wonderful fact exploited in relating f to f_π: namely that weak charged currents happen to be $SU(3)$ currents, and so can be unambiguously computed in the effective theory. The combination $(D+F) = g_A = 1.25$ is derived from neutron decay. From hyperon decay, one determines $F \simeq 0.44$, $D \simeq 0.81$. Because the axial interactions involve derivatives of the mesons, they contribute to p-wave scattering, but not s-wave; in contrast, the vector interaction contributes to s-wave scattering.

The above Lagrangian is $O(p)$, involving single derivatives on the meson fields. Therefore symmetry breaking terms involving the quark mass matrix M are subleading, as we have seen that $M \sim p^2$ in our power counting. There are three such terms, are

$$\mathcal{L}_2 = a_1 \mathrm{Tr}\overline{B}(\xi^\dagger M \xi^\dagger + \text{h.c.})B + a_2 \mathrm{Tr}\overline{B}B(\xi^\dagger M\xi^\dagger + \text{h.c.}) + a_3 \mathrm{Tr}\,(M\Sigma + \text{h.c.})\,\mathrm{Tr}\overline{B}B\ . \tag{172}$$

A combination of three of the mass parameters m_0, a_1, a_2, and a_3 may be determined from the octet masses; to fit the fourth combination requires additional data, but can be done). At the same order one should consider chirally symmetric terms with two derivatives acting on the mesons, such as subleading terms in the nonrelativistic expansion of $\mathcal{L}_1$, as well as new terms such as $\mathrm{Tr}\overline{B}A_\mu A^\mu B$. However, I believe that there are too many such subleading terms to fit to low energy meson-baryon scattering data.

5.2 An EFT for nucleon-nucleon scattering

Weinberg was the one who first suggested that perhaps nuclei and nuclear matter could be described by an effective theory — the chiral Lagrangian providing the pion-nucleon couplings responsible for the long-range part of the nuclear interaction, and contact interactions playing the role of the short distance interactions [9, 10]. This contrasts with the traditional approach of constructing a nucleon-nucleon potential, tweaking it until it is capable of fitting all of the nucleon-nucleon scattering phase shift data, and then using that potential to perform an N-body calculation, adding three-body interactions as needed to explain the data. Both methods make use of phenomenological parameters to fit the data, both methods are predictive in the sense that there are far more data than free parameters. So how would an effective field theory treatment be any better than using a potential model? There are several ways:

i. The EFT provides a natural framework for including known long-distance physics, such as one- and two-pion exchange, consistent with chiral symmetry;

ii. An EFT allows one to systematically identify the interactions important at any given level of accuracy.

iii. Unlike potential models, EFT never suffers from any ambiguity about "on-shell" and "off-shell" interactions.

iv. Relativistic effects, such as time retardation are simple to include in an EFT, but not in a potential model.

v. Dynamical processes, scattering, inelastic collisions — all of these are more simply treated in an EFT.

There are reasons to be dubious, however. After all, nuclei are non-perturbative (they are bound states). Furthermore, there has to be small ratio of scales for an EFT to work, and with the Fermi momentum in nuclear matter about 280 MeV it is not clear that this criterion pertains. For nucleon-nucleon scattering with momentum transfer far below the pion mass, the situation is clearer than for nuclear matter: then one can use a pion-less EFT, where the only interactions are n-body nucleon contact interactions (as well as electromagnetism).

A lot of work by many people has been done on both the pion-less EFT for very low energy, and the EFT with pions for higher energy processes and nuclei. In this lecture I will only have time to describe the pion-less theory, which is a systematic generalization of the effective range expansion, but it is interesting in its own right because of its nonperturbative nature. There are many reviews of EFT with pions; see for example [11].

5.3 The pion-less EFT for nucleon-nucleon interactions

Given that the pion-less theory consists only of nonrelativistic nucleons, the Lagrangian is quite simple. In $(4-D)$ dimensions it takes the form

$$\begin{aligned}\mathcal{L}_{eff} &= N^\dagger\left(i\partial_t + \nabla^2/2M\right)N \\ &\quad + (\mu/2)^{4-D}\left[C_0(N^\dagger N)^2 + \frac{C_2}{8}\left[(NN)^\dagger(N\overleftrightarrow{\nabla}^2 N) + h.c\right] + ...\right] \ ,\end{aligned} \tag{173}$$

where

$$\overleftrightarrow{\nabla}^2 \equiv \overleftarrow{\nabla}^2 - 2\overleftarrow{\nabla}\cdot\overrightarrow{\nabla} + \overrightarrow{\nabla}^2 \ . \tag{174}$$

I have suppressed everywhere the spin and isospin indices and their contractions which project these interactions onto the various scattering channels ($^1S_0$, $^3S_1$, etc.). The ellipsis indicates higher derivative operators, and $(\mu/2)$ is an arbitrary mass scale introduced to allow the couplings C_{2n} multiplying operators containing ∇^{2n} to have the same dimension for any D. I focus on the s-wave channel (generalization to higher partial waves is straightforward), and assume that M is very large so that relativistic effects can be ignored. The form of the C_2 operator is fixed by Galilean invariance, which implies that when all particle momenta are boosted $\mathbf{p} \to \mathbf{p} + M\mathbf{v}$, the Lagrangian must remain invariant. There exists another two derivative operator for p-wave scattering which I will not be discussing.

The usual effective field theory expansion requires one to identify a class of diagrams to sum which gives the amplitude $i\mathcal{A}$ to the desired order in a p/Λ expansion. For nonrelativistic scattering, the scattering amplitude is related to the S-matrix by

$$S = 1 + i\frac{Mp}{2\pi}\mathcal{A} \ , \tag{175}$$

where $p = \sqrt{ME_{\rm cm}}$ is the magnitude of the momentum that each nucleon has in the center of momentum frame. For s-wave scattering, we have seen that $\mathcal{A}$ is related to the phase shift δ by

$$\mathcal{A} = \frac{4\pi}{M}\frac{1}{p\cot\delta - ip} \ . \tag{176}$$

Recall from the first lecture that for a short-range two-body potential $V(r)$ that dies off exponentially fast for $r\Lambda > 1$, it is not $\mathcal{A}$ which in general has a good Taylor expansion in p/Λ, but rather the quantity $p\cot\delta$ which is expanded in the effective range expansion:

$$p\cot\delta = -\frac{1}{a} + \frac{1}{2}\Lambda^2\sum_{n=0}^{\infty} r_n\left(\frac{p^2}{\Lambda^2}\right)^{n+1} \ , \tag{177}$$

where a is the scattering length, and r_0 is the effective range. At best, our effective theory for nucleon-nucleon scattering should reproduce the effective range expansion. Of course, if this were the whole story, it would be boring, reproducing well known results! What will make it more interesting is when we incorporate electromagnetic and weak interactions into the theory. But first we need to understand the power counting of our EFT, since Feynman diagrams give one the amplitude $\mathcal{A}$ and not the quantity $p\cot\delta$.

A Taylor expansion of $\mathcal{A}$ in powers of p yields

$$\mathcal{A} = -\frac{4\pi a}{M}\left[1 - iap + (ar_0/2 - a^2)p^2 + O(p^3/\Lambda^3)\right] \ , \tag{178}$$

For a generic short-range potential, the coefficients r_n are generally $O(1/\Lambda)$ for all n. However, a can take on any value, which is problematic since for $1/|a| > \Lambda$, the above momentum expansion of the amplitude has a radius of convergence set by $1/|a|$ and not by Λ. A general property of the scattering length is that $1/a$ is negative for a weakly attractive potential, vanishes for a more attractive potential which possesses a bound state at threshold, and becomes positive for an even more attractive potential with a deep bound state. (For example: if one considers an attractive Yukawa potential for the form

$$V(r) = -\frac{g^2}{4\pi}\frac{e^{-\Lambda r}}{r} \tag{179}$$

then a bound state at threshold appears for the critical coupling $\eta \equiv g^2M/(4\pi\Lambda) \simeq 1.7$, at which point the scattering length a diverges.) First I consider the situation where the scattering length is of natural size $|a| \sim 1/\Lambda$, and then I discuss the case $|a| \gg 1/\Lambda$, which is relevant for realistic NN scattering.

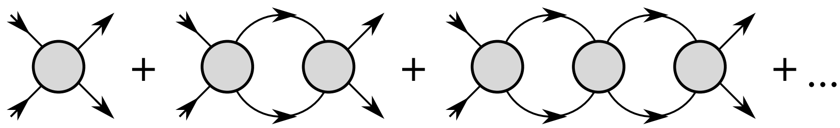

Figure 12: *The bubble chain arising from local operators. The vertex is given by the tree level amplitude, eq. (180).*

5.3.1 The case of a "natural" scattering length: $1/|a| \simeq \Lambda$

In the regime $|a| \sim 1/\Lambda$ and $|r_n| \sim 1/\Lambda$, the expansion of the amplitude $\mathcal{A}$ in eq. (178) converges up to momenta $p \sim \Lambda$, and it is this expansion that we wish to reproduce in an effective field theory.

The complete tree level s partial wave amplitude in the center of mass frame arising from $\mathcal{L}_{eff}$ is

$$i\mathcal{A}^{(cm)}_{\rm tree} = -i(\mu/2)^{4-D}\sum_{n=0}^{\infty} C_{2n}(\mu)p^{2n} \ , \tag{180}$$

where the coefficients $C_{2n}(\mu)$ are the couplings in the Lagrangian of operators with $2n$ gradients contributing to s-wave scattering. One may always trade time derivatives for spatial gradients, using the equations of motion when computing S-matrix elements, and so I will ignore such operators.

Beyond tree level one encounters the loop diagrams shown in Fig. 12. Formally, these are *all* the diagrams one encounters in a nonrelativistic theory...if you cut the diagrams in half somewhere in the middle, you can only encounter the two original particles and no additional particle-antiparticle pairs. The loop integrals one encounters are all of the form

$$\begin{aligned} I_n &\equiv i(\mu/2)^{4-D}\int \frac{\mathrm{d}^D q}{(2\pi)^D}\, \frac{\mathbf{q}^{2n}}{\left(E/2+q_0-\frac{\mathbf{q}^2}{2M}+i\epsilon\right)\left(E/2-q_0-\frac{\mathbf{q}^2}{2M}+i\epsilon\right)} \\ &= (\mu/2)^{4-D}\int \frac{\mathrm{d}^{(D-1)}\mathbf{q}}{(2\pi)^{(D-1)}}\,\mathbf{q}^{2n}\left(\frac{1}{E-\mathbf{q}^2/M+i\epsilon}\right) \\ &= -M(ME)^n(-ME-i\epsilon)^{(D-3)/2}\Gamma\left(\frac{3-D}{2}\right)\frac{(\mu/2)^{4-D}}{(4\pi)^{(D-1)/2}} \ . \end{aligned} \tag{181}$$

In order to define the theory, one must specify a subtraction scheme; different subtraction schemes amount to a reshuffling between contributions from the vertices and contributions from the the UV part of the loop integration. How does one choose a subtraction scheme that is useful? I am considering the case $|a|, |r_n| \sim 1/\Lambda$, and wish to reproduce the expansion of the amplitude eq. (178). In order to do this via Feynman diagrams, it is convenient if any Feynman graph with a particular set of operators at the vertices only contributes to the expansion of the amplitude at a particular order. Since the the expansion eq. (178) is a strict Taylor expansion in p, it is it is therefore very convenient if each Feynman graph yields a simple monomial in p. Obviously, this won't be true in a random subtraction scheme. A scheme that fulfills this criterion is the minimal subtraction scheme (MS) which amounts to subtracting any $1/(D-4)$ pole before taking the $D \to 4$ limit. As the integral eq. (181) doesn't exhibit any such poles, the result in MS is simply

$$I_n^{MS} = (ME)^n \left(\frac{M}{4\pi}\right) \sqrt{-ME - i\epsilon} = -i\left(\frac{M}{4\pi}\right) p^{2n+1} \ . \tag{182}$$

Note the nice feature of this scheme that the factors of q arising from the vertices get converted into factors of of the external momentum p when performing the integral. Similarly, a factor of the equations of motion, $i\partial_t + \nabla^2/2M$, acting on one of the internal legs at the vertex, causes the loop integral to vanish. Therefore one can use the on-shell, tree level amplitude eq. (180) as the internal vertex in loop diagrams. Summing the bubble diagrams in the center of mass frame gives

$$\mathcal{A} \quad = \quad -\frac{\sum C_{2n} p^{2n}}{1 + i(Mp/4\pi)\sum C_{2n} p^{2n}} \ . \tag{183}$$

Since for this process there are no poles at $D = 4$ in the MS scheme, the coefficients C_{2n} are independent of the subtraction point μ. The power counting in the MS scheme is particularly simple, as promised:

1. Each nucleon propagator counts as $1/p^2$;
2. Each loop integration $\int \mathrm{d}^4 q$ counts as p^5 (since $q_0 \sim \mathbf{q}^2/2M$);
3. Each vertex $C_{2n}\nabla^{2n}$ contributes p^{2n}.

In this simple theory we have managed to compute the exact amplitude, but we want to see how we derive power counting rules for the EFT, so we expand it in powers of p as

$$\mathcal{A} = \sum_{n=0}^{\infty} \mathcal{A}_n \ , \qquad \mathcal{A}_n \sim O(p^n) \tag{184}$$

where the $\mathcal{A}_n$ each arise from graphs with $L \leq n$ loops and can be equated to the low energy scattering data eq. (178) in order to fit the C_{2n} couplings. In particular, $\mathcal{A}_0$ arises from the tree graph with C_0 at the vertex; $\mathcal{A}_1$ is given by the 1-loop diagram with two C_0 vertices; $\mathcal{A}_2$ is gets contributions from both the 2-loop diagram with three C_0 vertices, as well as the tree diagram with one C_2 vertex, and so forth. Thus the first three terms are

$$\mathcal{A}_0 = -C_0 \ , \qquad \mathcal{A}_1 = iC_0^2 \frac{Mp}{4\pi} \ , \qquad \mathcal{A}_2 = C_0^3 \left(\frac{Mp}{4\pi}\right)^2 - C_2 p^2 \ . \tag{185}$$

Comparing eqs. (178, 185) I find for the first two couplings of the effective theory

$$C_0 = \frac{4\pi a}{M} \ , \qquad C_2 = C_0 \frac{a r_0}{2} \ . \tag{186}$$

In general, when the scattering length has natural size,

$$C_{2n} \sim \frac{4\pi}{M\Lambda} \frac{1}{\Lambda^{2n}} \ . \tag{187}$$

Note that the effective field theory calculation in this scheme is completely perturbative even though the underlying short-distance physics need not be. Also note that our choice of subtraction scheme (MS), while not changing the physics, made the power counting transparent. A feature of the fact that we are computing consistently to a given order in momentum is that fact that our results are independent of the renormalization scale μ.

5.3.2 The realistic case of an "unnatural" scattering length

One might guess that the results of the previous section would apply to low energy NN scattering, with role of Λ played by m_π or $m_\pi/2$. However, while it is true that the pion is the lightest hadron exchanged between nucleons, the EFT is much more interesting than the above scenario, as the NN scattering lengths are unnaturally large. For example, the ${}^1S_0$ scattering length is $a_0 = -23.714 \pm .013$ fm $\simeq 1/(8$ MeV$)$, which is *much* bigger than $1/m_\pi \simeq 1/(140$ MeV$)$.

For a nonperturbative interaction with a bound state near threshold, the expansion of $\mathcal{A}$ in powers of p is of little practical value, as it breaks down for momenta $p \gtrsim 1/|a|$, far below Λ. In the above effective theory, this occurs because the couplings C_{2n} are anomalously large, $C_{2n} \sim 4\pi a^{n+1}/M\Lambda^n$. However, the problem is not with the effective field theory method, but rather with the subtraction scheme chosen.

Instead of reproducing the expansion of the amplitude shown in eq. (178), one needs to expand in powers of p/Λ while retaining ap to all orders:

$$\mathcal{A} = -\frac{4\pi}{M}\frac{1}{(1/a+ip)}\left[1 + \frac{r_0/2}{(1/a+ip)}p^2 + \frac{(r_0/2)^2}{(1/a+ip)^2}p^4 + \frac{(r_1/2\Lambda^2)}{(1/a+ip)}p^4 + \ldots\right] \tag{188}$$

Note that for $p > 1/|a|$ the terms in this expansion scale as $\{p^{-1}, p^0, p^1, \ldots\}$. Therefore, the expansion in the effective theory should take the form

$$\mathcal{A} = \sum_{n=-1}^{\infty} \mathcal{A}_n \ , \qquad \mathcal{A}_n \sim O(p^n) \tag{189}$$

beginning at $n = -1$ instead of $n = 0$, as in the expansion eq. (184). Comparing with eq. (188), we see that

$$\begin{aligned} \mathcal{A}_{-1} &= -\frac{4\pi}{M}\frac{1}{(1/a+ip)} \ , \\ \mathcal{A}_0 &= -\frac{4\pi}{M}\frac{r_0 p^2/2}{(1/a+ip)^2} \ , \end{aligned} \tag{190}$$

and so forth. Again, the task is to compute the $\mathcal{A}_n$ in the effective theory, and equate to the appropriate expression above, thereby fixing the C_{2n} coefficients. As before, the goal is actually more ambitious: each particular graph contributing to $\mathcal{A}_n$ should be $O(p^n)$, so that the power counting is transparent.

As any single diagram in the effective theory is proportional to positive powers of p, computing the leading term $\mathcal{A}_{-1}$ must involve summing an infinite set of diagrams. It is easy to see that the leading term $\mathcal{A}_{-1}$ can be reproduced by the sum of bubble diagrams with C_0 vertices which yields in the MS scheme

$$\mathcal{A}_{-1} = \frac{-C_0}{\left[1 + \frac{C_0 M}{4\pi} ip\right]} \ . \tag{191}$$

Comparing this with eq. (190) gives $C_0 = 4\pi a/M$, as in the previous section. However, there is no expansion parameter that justifies this summation: each individual graph in the bubble sum goes as $C_0(C_0 Mp)^L \sim (4\pi a/M)(iap)^L$, where L is the number of loops. Therefore each graph in the bubble sum is bigger than the preceding one, for $|ap| > 1$, while they sum up to something small.

This is an unpleasant situation for an effective field theory; it is important to have an expansion parameter so that one can identify the order of any particular graph, and sum the graphs consistently. Without such an expansion parameter, one cannot determine the size of omitted contributions, and one can end up retaining certain graphs while dropping operators needed to renormalize those graphs. This results in a model-dependent description of the short distance physics, as opposed to a proper effective field theory calculation.

Since the sizes of the contact interactions depend on the renormalization scheme one uses, the task becomes one of identifying the appropriate subtraction scheme that makes the power counting simple and manifest. The MS scheme fails on this point; however this is not a problem with dimensional regularization, but rather a problem with the minimal subtraction scheme itself. A momentum space subtraction at threshold behaves similarly.

Consider an alternative regularization and renormalization scheme, namely to using a momentum cutoff equal to Λ. Then for large a one finds $C_0 \sim (4\pi/M\Lambda)$, and each additional loop contributes a factor of $C_0(\Lambda + ip)M/4\pi \sim (1 + ip/\Lambda)$. The problem with this scheme is that for $\Lambda \gg p$ the term ip/Λ from the loop is small relative to the 1, and ought to be ignorable; however, neglecting it would fail to reproduce the desired result eq. (190). This scheme suffers from significant cancellations between terms, and so once again the power counting is not manifest.

Evidently, since $\mathcal{A}_{-1}$ scales as $1/p$, the desired expansion would have each individual graph contributing to $\mathcal{A}_{-1}$ scale as $1/p$. As the tree level contribution is C_0, I must therefore have C_0 be of size $\propto 1/p$, and each additional loop must be $O(1)$. This can be achieved by using dimensional regularization and the *PDS* (power divergence subtraction) scheme. The *PDS* scheme involves subtracting from the dimensionally regulated loop integrals not only the $1/(D-4)$ poles corresponding to log divergences, as in MS, but also poles in lower dimension which correspond to power law divergences at $D = 4$. The integral I_n in eq. (181) has a pole in $D = 3$ dimensions which can be removed by adding to I_n the counterterm

$$\delta I_n = -\frac{M(ME)^n \mu}{4\pi(D-3)}, \tag{192}$$

so that the subtracted integral in $D = 4$ dimensions is

$$I_n^{PDS} = I_n + \delta I_n = -(ME)^n \left(\frac{M}{4\pi}\right)(\mu + ip). \tag{193}$$

In this subtraction scheme

$$\mathcal{A} \;=\; -\frac{M}{4\pi}\left[\frac{4\pi}{M\sum C_{2n}p^{2n}}+\mu+ip\right]^{-1} \ . \tag{194}$$

By performing a Taylor expansion of the above expression, and comparing with eq. (188), one finds that for $\mu \gg 1/|a|$, the couplings $C_{2n}(\mu)$ scale as

$$C_{2n}(\mu) \sim \frac{4\pi}{M\Lambda^n \mu^{n+1}} \ . \tag{195}$$

Eqs. (194,195) imply that the appropriate power counting entails $\mu \sim p$, $C_{2n}(\mu) \sim 1/p^{n+1}$. This is very different than the example of the "natural" scattering length discussed in the previous section; the strong interactions that give rise to a large scattering length have significantly altered the scaling of all the operators in the theory. A factor of ∇^{2n} at a vertex scales as p^{2n}, while each loop contributes a factor of p. The power counting rules for the case of large scattering length are therefore:

1. Each propagator counts as $1/p^2$;
2. Each loop integration $\int d^4q$ counts as p^5;
3. Each vertex $C_{2n}\nabla^{2n}$ contributes p^{n-1}.

We see that this scheme avoids the problems encountered with the choices of the MS ($\mu = 0$) or momentum cutoff ($\mu \sim \Lambda$) schemes. First of all, a tree level diagram with a C_0 vertex is $O(p^{-1})$, while each loop with a C_0 vertex contributes $C_0(\mu)M(\mu+ip)/4\pi \sim 1$. Therefore each term in the bubble sum contributing to $\mathcal{A}_{-1}$ is of order p^{-1}, unlike the case for $\mu = 0$. Secondly, since $\mu \sim p$, it makes sense keeping both the μ and the ip in eq. (193) as they are of similar size, unlike what we found in the $\mu = \Lambda$ case. The *PDS* scheme retains the nice feature of MS that powers of q inside the loop.

The above counting rules and their precursors were first discussed in [12–16] and are referred to in the literature as "KSW" counting, to be contrasted with "Weinberg counting" which is different. Using them one finds that the leading order contribution to the scattering amplitude $\mathcal{A}_{-1}$ scales as p^{-1} and consists of the sum of bubble diagrams with C_0 vertices; contributions to the amplitude scaling as higher powers of p come from perturbative insertions of derivative interactions, dressed to all orders by C_0. The first three terms in the expansion are

$$\begin{aligned}
\mathcal{A}_{-1} &= \frac{-C_0}{\left[1+\frac{C_0 M}{4\pi}(\mu+ip)\right]} \ , \\
\mathcal{A}_{0} &= \frac{-C_2p^2}{\left[1+\frac{C_0 M}{4\pi}(\mu+ip)\right]^2} \ , \\
\mathcal{A}_{1} &= \left(\frac{(C_2p^2)^2 M(\mu+ip)/4\pi}{\left[1+\frac{C_0 M}{4\pi}(\mu+ip)\right]^3} - \frac{C_4p^4}{\left[1+\frac{C_0 M}{4\pi}(\mu+ip)\right]^2}\right) \ ,
\end{aligned} \tag{196}$$

where the first two correspond to the Feynman diagrams in Fig. 13. The third term, $\mathcal{A}_1$, comes from graphs with either one insertion of $C_4\nabla^4$ or two insertions of $C_2\nabla^2$, dressed to all orders by the C_0 interaction.

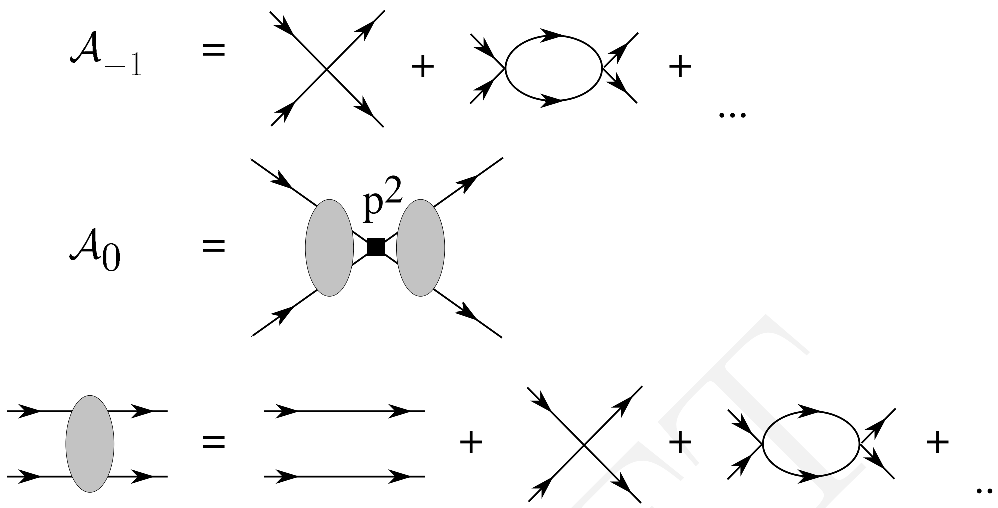

Figure 13: *Leading and subleading contributions arising from local operators. The unmarked vertex is the C_0 interaction, which is summed to all orders; the one marked "p^2" is the C_2 interaction, etc.*

Comparing eq. (196) with the expansion of the amplitude eq. (188), the couplings C_{2n} are related to the low energy scattering data a, r_n:

$$\begin{aligned} C_0(\mu) &= \frac{4\pi}{M}\left(\frac{1}{-\mu+1/a}\right) , \\ C_2(\mu) &= \frac{4\pi}{M}\left(\frac{1}{-\mu+1/a}\right)^2 \frac{r_0}{2} , \\ C_4(\mu) &= \frac{4\pi}{M}\left(\frac{1}{-\mu+1/a}\right)^3 \left[\frac{1}{4}r_0^2 + \frac{1}{2}\frac{r_1}{\Lambda^2}(-\mu+1/a)\right] . \end{aligned} \tag{197}$$

Note that assuming $r_n \sim 1/\Lambda$, these expressions are consistent with the scaling law in eq. (195).

This is pretty weird power counting! What is going on? It is easier to understand if we define the dimensionless coupling

$$\hat{C}_0 \equiv \frac{M\mu}{4\pi} C_0 . \tag{198}$$

Given our expression for C_0 above, it is easy to check that $\hat{C}_0$ obeys the RG equation

$$\mu \frac{d\hat{C}_0}{d\mu} = \hat{C}_0(1+\hat{C}_0) . \tag{199}$$

this has one fixed point at $\hat{C}_0 = 0$, corresponding to no interaction and $a = 0$, and another fixed point at $\hat{C}_0 = -1$, corresponding to $a = \pm\infty$, the case of the "unitary fermion". The

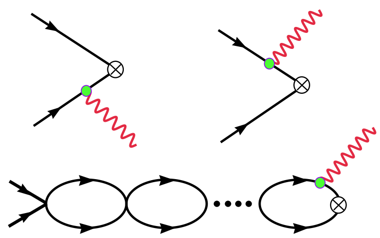

Figure 14: *The leading order contribution to $np \to d\gamma$. Solid lines denote nucleons, wavy lines denote photons. The photon coupling is through the nucleon anomalous magnetic moment operator in $\mathcal{L}_B$; the resummed unmarked vertex is the C_0 interaction. The crossed circle represents and insertion of the deuteron interpolating field $\mathcal{D}_i$. The bubble chain without photon insertions (not shown) is used to compute the wave function renormalization Z, and to fit C_0 to get the correction deuteron binding energy. See [17–19].*

usual perturbative expansion and power counting is formulated around the trivial fixed point, $\hat{C}_0 = 0$, while the KSW counting is formulated about the unitary limit, $\hat{C}_0 = -1$. Around that fixed point operators apparently do not have their naive scaling dimension and the power counting is peculiar looking.

5.3.3 Beyond the effective range expansion

So far, we have developed an elaborate machinery to just reproduce the effective range expansion! The payoff comes when one includes electromagnetic and weak interactions. The example I will briefly describe the application of the pion-less effective theory to here is the application of the pion-less effective theory to radiative capture process $np \to d\gamma$. At leading order, the ingredients to the calculation are the following:

i. One starts with the the nucleon kinetic two-nucleon C_0 interaction for the ${}^3S_1$ channel, written as

$$\mathcal{L} = \ldots - C_0 (N^T P_i N)^\dagger (N^T P_i N) \ , \tag{200}$$

where N is the nucleon doublet, and P_i is the projection operator onto the ${}^3S_1$ channel:

$$P_i = \frac{1}{\sqrt{8}} \sigma_2 \sigma_i \tau_2 \ , \quad \mathrm{Tr} P_i P_j = \frac{1}{2} \delta_{ij} \ , \tag{201}$$

where the σ_i act on spin and the τ_i act on isospin.

ii. One uses the convenient interpolating field $\mathcal{D}_i(x) \equiv N^T P_i N(x)$ to be the operator that creates a deuteron at the point x. The coupling C_0 can be fixed by ensuring that the pole in $\mathcal{A}_{-1}$ occurs at the deuteron binding energy. The leading order wave function normalization Z is extracted by looking at the residue at the pole. ($\sqrt{Z}$ is just the amplitude for our operator $\mathcal{D}_i$ to create a physical deuteron.)

iii. $np \to d\gamma$ occurs by emitting a magnetic photon, and so one needs to include in the Lagrangian the anomalous magnetic moment interaction of the nucleons:

$$\mathcal{L}_B = \frac{e}{2M_N} N^\dagger(\kappa_0 + \kappa_1\tau_3)\boldsymbol{\sigma}\cdot\mathbf{B}N \ , \tag{202}$$

where $\kappa_0 = \frac{1}{2}(\kappa_p + \kappa_n)$ and $\kappa_1 = \frac{1}{2}(\kappa_p - \kappa_n)$ are the isoscalar and isovector nucleon magnetic moments with $\kappa_p = 2.79$, $\kappa_n = -1.91$.

iv. Then at leading order one sums up the bubble chain with one insertion of the magnetic moment operator, as shown in Fig. 14.

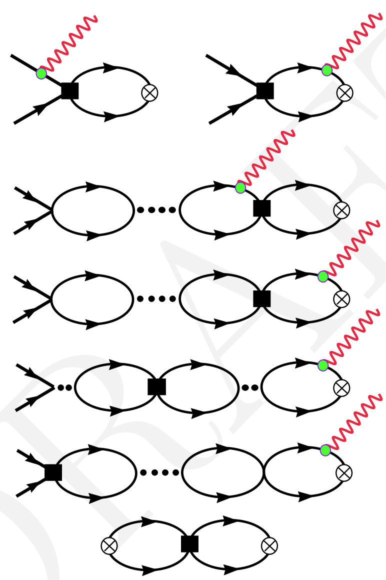

Figure 15: *The graphs contributing to $np \to d\gamma$ at NLO. The black square corresponds to an insertion of a C_2 interaction, the circle to the nucleon anomalous magnetic moment, and the resummed unmarked vertex to the C_0 interaction. The last graph is the contribution to wave function renormalization at this order. Figure from ref. [19].*

From these graphs one finds the capture cross section

$$\sigma = \frac{8\pi\alpha\gamma^5\kappa_1^2 a_0^2}{vM_N^5}\left(1 - \frac{1}{\gamma a_0}\right)^2 , \tag{203}$$

where α is the fine structure constant and v is the magnitude of the neutron velocity (in the proton rest frame), $a_0 = -23.714 \pm .013$ fm is the ${}^1S_0$ scattering length and $\gamma \equiv \sqrt{M_N B}$, where B is the deuteron binding energy. This agrees with old results of Bethe and Longmire when terms in their expression involving the effective range (which are higher order in our expansion) are neglected.

At next-to-leading order ("NLO") one needs to sum all relevant diagrams involving a single insertion of a C_2 vertex (that is, a 2-derivative contact interaction, whose value is fit

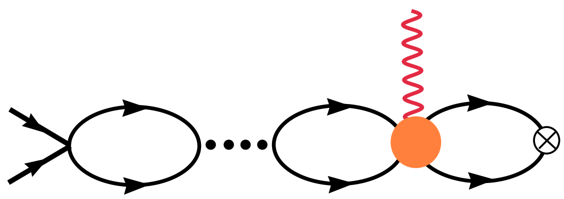

Figure 16: *An additional graph at NLO including an insertion of the L_1 operator. From ref. [19], courtesy of M. Savage.*

to the experimental effective range in NN scattering) for both the ${}^1S_0$ and ${}^3S_1$ channels, as in Fig. 15.

However this is not all. At the same order one finds a new contact interaction which cannot be fit to NN scattering data. It is a 2-body interaction with a magnetic photon attached, involving a new coupling constant L_1:

$$\mathcal{L}_{L_1} = eL_1(N^T P_i N)^\dagger (N^T P_3 N) B_i \ . \tag{204}$$

A gauge field is power counted the same as a derivative, and so the B field counts as two spatial derivatives. Thus graphs with one L_1 insertion and an infinite number of C_0 insertions comes in at the same order as the $\kappa\,\gamma NN$ vertex summed with an infinite number of C_0 vertices and one C_2 insertion. So at NLO one needs also to include the graph in Fig. 16.

The NLO result deviates from the old effective range calculations, since the L_1 operator is a completely new ingredient, and it also changes the dependency of the answer upon the effective range. This new coupling L_1 can be fit to data at one particular neutron velocity, and then one has a highly accurate prediction for neutron capture at any low velocity. The state of the art is presently an N^4LO calculation for the related breakup process $\gamma d \to np$ by Gautam Rupak [20]. His results are shown in Fig. 17.

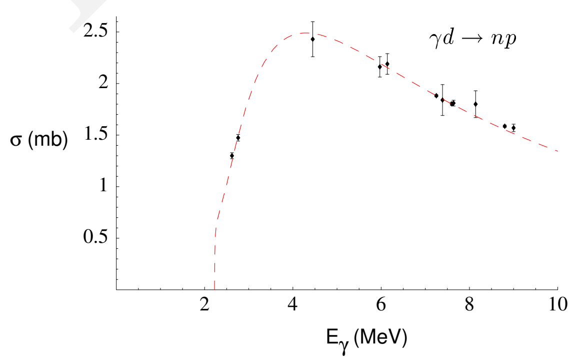

Figure 17: *Cross section for $\gamma d \to np$ breakup as a function of photon energy E_γ. The dashed line is the theoretical calculation of ref. [20], as is this figure, courtesy of G. Rupak. The data are from ref. [21]*

5.4 Including pions

The above examples are very convincing, but without pions one cannot describe nucleons with high enough relative momenta to describe nuclei, unfortunately. We saw at the beginning of this lecture that one can include baryons in the chiral Lagrangian, thereby determining how mesons couple to them in a momentum expansion. This allows one to include not only the nucleon contact interactions discussed above, but also one and multiple pion exchange between nucleons.

However, the KSW power counting i have described for the pion-less theory generalizes to treat pion exchange perturbatively, and the expansion does not converge well in the spin-triplet channel where pion exchange is quite strong. Instead, the original expansion proposed by Weinberg is commonly used, where instead of performing an EFT expansion of the scattering amplitude, one computes the nucleon-nucleon interaction in the EFT expansion, and then uses that interaction potential in the Schrödinger equation. Weinberg counting has various drawbacks, such as requiring one to tune a UV cutoff to lie in a narrow window and not remove it, but this approach appears to be the only game in town at the present. This is the method used in [11].

6 Chiral lagrangians for BSM physics

6.1 Technicolor

Chiral lagrangians are the tool for studying the physics of Goldstone bosons. The pions in particular are the ones we have been looking at, but we know that there are at least three other Goldstone bosons in reality: namely the three that are "eaten" via the Higgs mechanism and become the longitudinal degrees of freedom in the $W^\pm$ and the Z^0 gauge bosons in the SM. Let us start by asking what the world would look like if the Higgs doublet was omitted from the SM; to make this toy model simpler, let's also imagine that the world only has one family of fermions, u, d, e, ν_e, and in the following discussion I will ignore the leptons.

Without the Higgs the $SU(2) \times U(1)$ gauge bosons are all massless down at the QCD scale and have to be included in the chiral Lagrangian. Therefore we need only consider an $SU(2) \times SU(2)/SU(2)$ chiral Lagrangian, where the Σ field represents the three pions and is written as

$$\Sigma = e^{2iT_a\pi_a/f_\pi} \ , \qquad T_a = \frac{\sigma_a}{2} \ , \tag{205}$$

but how do we incorporate the gauge fields? Recall that in the SM the gauge charges of the quarks under $SU(2) \times U(1)$ are

$$\begin{pmatrix} u \\ d \end{pmatrix}_L = 2_{\frac{1}{6}} \ , \qquad u_R = 1_{\frac{2}{3}} \ , \qquad d_R = 1_{-\frac{1}{3}} \ . \tag{206}$$

Note that the $SU(2)_L$ of the chiral Lagrangian is exactly the same group as the $SU(2)_L$ that is gauged in the SM; the gauged $U(1)$ charge Y can be written as

$$Y = \frac{1}{2} B + T_{3,R} \tag{207}$$

where B is baryon number, the $U(1)_V$ symmetry that acts the same on all LH and RH quarks, and $T_{3,R} \in SU(2)_R$. The $U(1)_V$ part is uninteresting in the chiral Lagrangian since the pions do not carry that quantum number; however we have to use covariant derivatives for the $SU(2)_L \times U(1)_R$ part of the gauge group:

$$\partial_\mu \Sigma \quad \Longrightarrow \quad D_\mu \Sigma = \partial_\mu \Sigma + igW^a_\mu T_a \Sigma - ig' B_\mu \Sigma T_3 \ , \tag{208}$$

where I used the fact that under $SU(2) \times SU(2)$ the Σ field transforms as $\Sigma \to L\Sigma R^\dagger$; the dagger on R is what is responsible for the minus sign in front of the $U(1)$ gauge boson piece above.

Something peculiar is going on: now that $SU(2)_L$ is gauged, the matrix L is promoted to a spacetime dependent matrix $L(x)$, and I can choose $L(x) = \Sigma^\dagger(x)$...and a gauge transformation can turn Σ into the unit matrix! This is because the quark condensate $\langle \bar{q}_L q_R \rangle$ has broken the weak interactions spontaneously, the W and B get masses, and we just rediscovered unitary gauge.

Now let's look at the kinetic term for the Σ field in unitary gauge:

$$\frac{f^2}{4} \mathrm{Tr} D_\mu \Sigma (D^\mu \Sigma)^\dagger = \frac{f^2}{4} \left[g^2 W^a_\mu W^{\mu b} \mathrm{Tr} T_a T_b - 2gg' W^3_\mu B^\mu \mathrm{Tr} T_3 T_3 + g'^2 \mathrm{Tr} T_3 T_3 \right] \ . \tag{209}$$

This gives us a mass^2 matrix for the gauge bosons

$$M^2 = \frac{f^2}{4} \begin{pmatrix} g^2 & & & \\ & g^2 & & \\ & & g^2 & -gg' \\ & & -gg' & g'^2 \end{pmatrix} \tag{210}$$

with eigenvalues

$$m^2 = \left\{0,\ \frac{g^2 f^2}{4},\ \frac{g^2 f^2}{4},\ \frac{(g^2 + g'^2) f^2}{4}\right\} \equiv \left\{m_\gamma^2,\ m_w^2,\ m_w^2,\ mz^2\right\} \tag{211}$$

where

$$m_\gamma^2 = 0\ , \qquad m_w^2 = \frac{e^2 f_\pi^2}{4 \sin^2\theta_w}\ , \qquad m_z^2 = \frac{m_w^2}{\cos^2\theta_w} \tag{212}$$

with the conventional definitions $g = e/\sin\theta_w$, $g' = e/\cos\theta_w$.

This would look precisely like the SM if instead of $f_\pi = 93$ MeV we took $f_\pi = 250$ GeV! This was the observation of both Weinberg and Susskind [22–24]: that the strong interactions would provide the Higgs mechanism in the SM without the Higgs doublet field. While the W and Z would get massive, as in the SM, in this theory there might not be an actual Higgs boson (it is not in the chiral Lagrangian, and QCD does not have any narrow 0^+ resonance), and the order parameter for the symmetry breaking is a quark bilinear now instead of a fundamental scalar, and so that there is no naturalness problem: the weak scale is set by the QCD scale, which can be naturally much smaller than the GUT or Planck scales due to asymptotic freedom, just as Cooper pairs in superconductivity are so much larger than the crystal lattice spacing.

However, $f_\pi \neq 250$ GeV and we do not want to give away the pion. So instead we posit a new gauge interaction, just like color but at a higher scale called technicolor, and new fermions that carry this gauge quantum number, U, D, called techniquarks. Now the techniquarks condense with a technipion decay constant $F_\Pi = 250$ GeV, and we have explained the SM without the Higgs!

This is great, except for three problems: (i) the Higgs boson has been discovered, long after the invention of technicolor, while there is no analogously light 0^+ meson in the strong interactions; (ii) it is difficult to construct a technicolor theory where precision radiative corrections to the weak interactions agree with what is measured; (iii) as described here, the techniquark condensate only fulfills one of the roles played by the Higgs serves in the SM: giver of mass to the gauge bosons. To realize the other role — giver of mass to the quarks and leptons — requires new interactions added to be technicolor (e.g., Extended Technicolor), more complications, and serious problems with flavor changing neutral currents. There are more complex versions of the original theory still being explored, but the compelling beauty of the original concept is no longer there.

6.2 Composite Higgs

What is so beautiful about technicolor is that by replacing the fundamental Higgs scalar field with a quark bilinear, the fine tuning problem is avoided. Do we have to abandon that idea simply because we have seen a Higgs boson that looks fundamental? No, that has happened before: the pions looked fundamental when discovered, and only over a decade later was it decided that they were composite bound states. Furthermore, they are relativistic bound states – at least in the chiral limit, their compositeness scale is much smaller than their Compton wavelengths. Relativistic bound states are in general baffling things we have no handle on as they always require strong coupling. One of the few exceptions when we understand what is going on exception is when the bound states are Goldstone bosons, like the pions in QCD. So let us return to the QCD chiral Lagrangian for inspiration, this time not looking at the pions as candidates for the longitudinal W and Z, but at the kaons as candidates for the Higgs doublet [25, 26].

A Higgs doublet field has four real degrees of freedom, so we need to look at a chiral Lagrangian with at least four Goldstone bosons. The $SU(2)\times SU(2)/SU(2)$ chiral Lagrangian only has three pions, so that won't do; however the $SU(3) \times SU(3)/SU(3)$ example has eight mesons. Furthermore, there is a ready made $SU(2)$ doublet in that theory – the kaon. In fact, note that under the unbroken isospin and strong hypercharge $SU(2) \times U(1)$ global symmetry of QCD, the kaon doublet transforms just like the Higgs doublet does under gauged electroweak $SU(2) \times U(1)$ symmetry in the SM. So what if we scaled up a version of QCD with three massless flavors, gauge the vector-like $SU(2) \times U(1)$ symmetry, and identify the kaon doublet with the Higgs doublet?

In such a theory the Higgs would get a mass from $SU(2) \times U(1)$ gauge interactions, just as the π^+ gets a mass contribution from electromagnetic contributions (see problem 6). Unfortunately this contribution to the Higgs mass^2 is positive, and so there is no spontaneous symmetry breaking in the model so far. What is needed is some effect that contributes a negative mass^2. Various examples have been found that will do this: one is an axial gauge group under which the constituents of the composite Higgs are charged [27]; another and more compelling source of destabilization of the composite Higgs potential is the top quark [2]. The basic idea is that operators can be induced in the chiral Lagrangian which favor the vacuum $\Sigma \neq 1$ and prefer a small nonzero value corresponding to the composite Higgs doublet getting a vev v which can be computable in terms of f and coupling constants, such as the top Yukawa coupling. The example of vacuum destabilization from an axial gauge interaction is easy to understand. We have seen in eq. (4.3.2) that the operator which gives the positive electromagnetic contribution to the π^+ and K^+ mass^2 is:

$$\mathcal{L}_\alpha = \xi f^4 \frac{\alpha}{4\pi} \mathrm{Tr} Q_L \Sigma Q_R \Sigma^\dagger = \xi \frac{\alpha}{6\pi} f^4 - 2\xi \frac{\alpha}{9\pi} f^2 \left(\pi^+\pi^- + K^+K^-\right) + \dots \tag{213}$$

where $Q_R = Q_L = \mathrm{diag}\{2/3, -1/3, -1/3\}$. If we had gauged an "axial electromagnetic" $U(1)$ symmetry instead with $Q_R = -Q_L$, then this contribution to the kaon mass^2 would be $m^2_{K^+} = -2\xi \frac{\alpha}{9\pi} f^2$ — and if there were no quark masses, this would be the leading contribution and would signal that the vacuum with $\langle K \rangle = 0$ is unstable. A vev for the kaon implies a vev for Σ which is not the unit matrix: the quark condensate misaligns from the $SU(2) \times U(1)$ symmetry preserving direction, breaking the weak interactions. If there are competing positive and negative contributions to the kaon mass^2 it is possible to select whatever value one wants for v/f, and if $v/f \ll 1$, then the composite Higgs will look very much like the SM Higgs – one will need to probe energies on the order of f to see its composite nature.

This scale v is set to 250 GeV and is the usual Higgs vev; the theory then predicts at a higher scale $\Lambda \simeq 4\pi f$ the Higgs will reveal itself to be composite. The smaller the ratio v/f, the more fundamental the Higgs will appear at low energy. For a recent discussion of the viability of the composite Higgs idea and accelerator tests, see [28].

6.2.1 The axion

A problem that afflicts QCD is that one can add to the action the operator

$$-\theta \frac{g^2}{16\pi^2} \mathrm{Tr}\, G_{\mu\nu}\widetilde{G}_{\mu\nu} \tag{214}$$

and even though it is a total derivative, the topology of the vacuum allows it to have a physical effect. Because of the Levi-Civita symbol in the definition of $\widetilde{G}_{\mu\nu}$, this operator violates P and T, while conserving C. Thus it is CP violating, and it can contribute to CP-violating observables in hadrons, such as the neutron electric dipole moment. Experimental bounds on the neutron EDM imply that $\theta \lesssim 10^{-9}$. This fine tuning is called the strong CP problem. If we consider the vacuum energy $E(\theta)$ as a function of θ it is intriguing that its minimum is at $\theta = 0$ [29]. However θ is like a coupling constant and cannot change its value to minimize the energy. The brilliant observation by Peccei and Quinn [30] was that it is possible to construct a theory where θ gets replaced by a dynamical field which can relax to zero; this field gives rise to a light particle [31,32] called the axion.

It was discovered in the late 1960s and early 1970s that classical symmetries can be broken by quantum effects involving gauge fields, called anomalies. Furthermore, if that symmetry is spontaneously broken and the gauge fields are not strongly interacting, the resultant pseudo Goldstone boson will couple to gauge fields very much like the θ parameter above. For example, the linear combination of $SU(3)_L \times SU(3)_R$ currents $\bar{q}\gamma_\mu\gamma_5 T^3 q$ has an electromagnetic anomaly since $\mathrm{Tr} T^3 Q^2 \neq 0$, and it is spontaneously broken by the quark condensate as well, giving rise to the π^0 pseudo Goldstone boson, which as a consequence of the anomaly, has a coupling to photons of the form

$$\frac{\alpha}{4\pi}\frac{\pi^0}{f_\pi} F\widetilde{F} \tag{215}$$

So the Peccei Quinn mechanism involves introducing a global symmetry that has a color anomaly and which is spontaneously broken at high energy; this will result in a pseudo Goldstone boson coupling like θ. For example, consider adding a Dirac color-triplet quark $\mathcal{Q}$ to the standard model which is neutral under the weak interactions, and add a complex scalar field Φ with the Lagrangian

$$\mathcal{L} = \overline{\mathcal{Q}} i \not{D} \mathcal{Q} + |\partial\Phi|^2 - V(|\Phi|^2) - y\overline{\mathcal{Q}}_L \Phi \mathcal{Q}_R + h.c. \, , \tag{216}$$

where D_μ is the $SU(3)$ covariant derivative. This Lagrangian has a global symmetry

$$\mathcal{Q} \to e^{i\omega\gamma_5}\mathcal{Q} \, , \qquad \Phi \to e^{-2i\omega}\Phi \, , \tag{217}$$

which has a color anomaly. If the scalar potential breaks this symmetry with $\langle\Phi\rangle = f_a/\sqrt{2}$, where $f_a \gg \Lambda_{QCD}$ has dimension of mass, then the exotic quark $\mathcal{Q}$ gets very heavy with mass $yf_a/\sqrt{2}$ and the only light mode that survives is a pseudo Goldstone boson $a(x)$, the axion, which couples to glue via the anomaly:

$$\mathcal{L}_{\rm eft} = -\frac{a(x)}{f_a}\frac{g^2}{16\pi^2}\mathrm{Tr}\, G_{\mu\nu}\widetilde{G}_{\mu\nu} \tag{218}$$

So we see that θ is replaced with the dynamical quantity $a(x)/f_a$. If there was a nonzero θ in the orginal theory, we have shifted the axion field to absorb it. Now the question is, is $\langle a\rangle = 0$?

How do we figure out how the axion couples to matter? In general we do not know how to analytically compute the effect of gluon operators, but in the present case we can. Consider QCD with the u, d, s quarks included in the theory:

$$\mathcal{L} = \bar{q}i\not{D}q - \bar{q}_R M q_L + h.c. - \frac{a(x)}{f_a}\frac{g^2}{16\pi^2}\mathrm{Tr}\, G_{\mu\nu}\widetilde{G}_{\mu\nu}\ , \qquad q = \begin{pmatrix} u \\ d \\ s \end{pmatrix}\ . \tag{219}$$

If we perform a spacetime dependent axial rotation on the quarks

$$q \to e^{-i\beta(x)Q_{\rm ax}\gamma_5} q \tag{220}$$

where $Q_{\rm ax}$ is some diagonal 3×3 matrix to be determined, then in terms of the new fields the theory is

$$\begin{aligned}\mathcal{L} \;&=\; \bar{q}i\not{D}q - i\partial_\mu\beta\bar{\gamma}^\mu\gamma_f Q_{\rm ax}q - \bar{q}_R M e^{2i\beta(x)Q_{\rm ax}} q_L + \text{h.c.} \\ &\quad - \left[\frac{a(x)}{f_a} - \beta(x)\mathrm{Tr}Q_{\rm ax}\right]\frac{g^2}{16\pi^2}\mathrm{Tr}\, G_{\mu\nu}\widetilde{G}_{\mu\nu} + \frac{\alpha}{2\pi}\beta(x)(3\mathrm{Tr}Q_{\rm ax}Q^2)\, F\widetilde{F}\ .\end{aligned} \tag{221}$$

where the shift in the $G\widetilde{G}$ ($F\widetilde{F}$) is due to the color (electromagnetic) anomaly of the transformation, which for the color anomaly is proportional to $\mathrm{Tr}Q_{\rm ax}T_aT_b = \frac{1}{2}\delta_{ab}\mathrm{Tr}Q_{\rm ax}$, where the T_a are color generators, with the analogous expression with $T_{a,b}$ replaced by the electric charge matrix for the electromagnetic anomaly, and a factor of three for color. So if we choose

$$\mathrm{Tr}Q_{\rm ax} = 1\ , \qquad \beta(x) = a(x)/(f_a)\ , \tag{222}$$

we can remove the $G\widetilde{G}$ coupling entirely and obtain

$$\mathcal{L} \;=\; \bar{q}i\not{D}q - i\frac{\partial_\mu a}{2f_a}\bar{q}\gamma^\mu\gamma_5 Q_{\rm ax}q - \bar{q}_R M e^{i2a(x)Q_{\rm ax}/f_a} q_L + \text{h.c.} + \frac{\alpha}{2\pi}\frac{a}{f_a}\mathrm{Tr}Q_{\rm ax}Q^2\, F\widetilde{F} \tag{223}$$

We know how to match this Lagrangian onto the chiral Lagrangian because (i) the derivative couplings of the axion match onto symmetry currents in the chiral Lagrangian; (ii) the mass coupling just involves replacing the quark mass matrix $M \to Me^{2ia(x)Q_{\rm ax}/f_a}$ everywhere in our previous analysis; (iii) the electormagnetic coupling of the axion is the same in the chiral Lagrangian, not involving any strong interactions.

Note that we have not fixed $Q_{\rm ax}$ yet, beyond requiring $\mathrm{Tr}Q_{\rm ax} = 1$. We will do so shortly.

I will ignore the derivative and electromagnetic couplings of the axion and focus on its potential, arising from the mass term. By modifying eq. (122) we now have in the chiral Lagrangian the term

$$\begin{aligned}\mathcal{L}_M &= \frac{1}{2} f_\pi^2 \widetilde{\Lambda} \mathrm{Tr} M e^{i2a(x)Q_{\rm ax}/f_a} \Sigma + \text{h.c.} \\ &= \frac{1}{2} f_\pi^2 \widetilde{\Lambda} \mathrm{Tr} M \left[(1 + 2iaQ_{\rm ax}/f_a - 2a^2 Q_{\rm ax}^2/f_a^2 + \ldots)(1 + 2i\boldsymbol{\pi}/f_\pi - 2\boldsymbol{\pi}^2/f_\pi^2 + \ldots) \right] + \text{h.c.}\end{aligned} \tag{224}$$

Note that the terms linear in either a or $\boldsymbol{\pi}$ vanish because of the "+h.c.". However, in general there will be a mass mixing term between the meson and the axion field (more precisely, between a, π^0 and η). We can eliminate this term from the start by choosing

$$\begin{aligned}Q_{\rm ax} &= \frac{M^{-1}}{\mathrm{Tr} M^{-1}} = \frac{1}{m_u m_d + m_d m_s + m_s m_u} \begin{pmatrix} m_d m_s & & \\ & m_s m_u & \\ & & m_u m_d \end{pmatrix} \\ &= \frac{1}{m_u + m_d} \begin{pmatrix} m_d & & \\ & m_u & \\ & & 0 \end{pmatrix} + O\left(\frac{m_u}{m_s}\right) + O\left(\frac{m_d}{m_s}\right) .\end{aligned} \tag{225}$$

Note that this has unit trace as we required; it also eliminates the mixing term by cancelling off the factor of M in the trace, since $\mathrm{Tr}\,\boldsymbol{\pi} = 0$. With this choice for $Q_{\rm ax}$ we can now easily compute the axion mass and find

$$\mathcal{L}_M = -2a^2 \frac{f_\pi^2 \widetilde{\Lambda}}{f_a^2} \mathrm{Tr} M Q_{\rm ax}^2 + \ldots = -\frac{1}{2} a^2 \frac{4 f_\pi^2 \widetilde{\Lambda}}{f_a^2 \mathrm{Tr} M^{-1}} \tag{226}$$

If we use our previous result $m_\pi^2 = \widetilde{\Lambda}(m_u + m_d)$ we can read off the axion mass as

$$m_a^2 = \frac{4 f_\pi^2 \widetilde{\Lambda}}{f_a^2 \mathrm{Tr} M^{-1}} \simeq \frac{f_\pi^2 m_\pi^2}{f_a^2} \left(\frac{4 m_u m_d}{(m_u + m_d)^2} \right) \tag{227}$$

where I have dropped terms proportional to $1/m_s$ in the expression on the right. With $m_u/m_d \simeq 1/2$, the factor on the right is very close to $1/4$. Note that the axion mass scales as our undetermined parameter $1/f_a$. You can easily show that the axion is odd under CP, and since a positive mass^2 for the axion implies the minimum of the axion potential is at $\langle a \rangle = 0$, and since we have no θ parameter, the strong interactions are now CP conserving as observed in nature.

If we ignore isospin breaking and set $m_u = m_d$, then we get the full axion potential very simply,

$$V(a) = -m_a^2 f_a^2 \cos\frac{a}{f_a} = -m_\pi^2 f_\pi^2 \cos\frac{a}{f_a} \tag{228}$$

The axion model is beautiful, and really is the only known soltuion to the strong CP problem that is viable. Now the ugly facts about the model: it is critical that the entire axion potential come from QCD so that the minimum is at $\langle a \rangle = 0$. This is equivalent to say that the only operator in the theory that does not respect the global PQ symmetry is the gluon anomaly term, eq. (218). However, it is believed that quantum gravity violates global charges: if you through global charge into a black hole, it is lost. So one might expect high dimension operators that violate the PQ symmetry, suppressed by the appropriate powers of the Planck mass M_p, such as $c\Phi^6/M_p^2 +$ h.c.. However, this will typically favor a nonzero vev for the axion. It turns out one must forbid such operators up to dimension ten or so, for no obvious good reason, to ensure that the PQ mechanism solves the strong CP problem.

6.2.2 Axion cosmology and the anthropic axion

The axion potential vanishes at high temperature (compared to $\Lambda_{\rm QCD}$ because (i) chiral symmetry is restored, and (ii) instantons become rapidly unimportant at high temperature, in which case the $G\widetilde{G}$ term is unimportant. The Peccei-Quinn (PQ) symmetry breaking scale f_a is high compared to $\Lambda_{\rm QCD}$ and therefore when Φ gets a vev, it sees a completely degenerate choice of minima, meaning a vanishing potential for the axion. Thus $\langle a/f_a\rangle$ will select some random value in the early universe, but at $T\sim\Lambda_{\rm QCD}$ will find it is not at the minimum of its potential, and when its Compton wavelength comes within the horizon, it will start oscillating, and there will be a lot of energy in the field. If it is spatially uniform and oscillating in time, it looks like a Bose-Einstein condensate of axions at rest, and will dilute as $R9t0^{-3}$, where $R(t)$ is the cosmic scale factor, like any nonrelativistic matter. Because of the weak couplings to ordinary matter, suppressed by $1/f_a$, the axions will not thermalize and become relativistic. When one computes the relic energy in the axions today, on finds that it increases with f_a, and that axions could be the dark matter if $f_a\sim 10^{12}$ GeV; if f_a is larger, then too large a fraction of the mass budget is in dark matter.

One might think that if the axion just happened to be close to the right minimum from the start, then the dark matter that results could be made small for any f_a. However, if PQ symmetry breaking occurs below the reheat temperature after inflation, today's horizon results from many causally disconnected horizons at the PQ symmetry breaking scale, and it would violate causality to force a/f_a to have initial values near zero over all of these regions.

If the PQ symmetry breaks before inflation, then the initial horizon gets stretch over a volume much larger than today's horizon, and it is possible for the axion to have an accidentally small initial value, allowing large f_a without problem. It seems that we have substituted one fining problem (why θ is small) with a similar one (why the initial value for $\langle a/f_a\rangle$ is small in the patch that becomes our horizon). However here one can actually make a convincing anthropic argument:

- We know that all possible values of a/f_a will occur below the PQ scale for $0\le a/f_a\le 2\pi$ with equal probability;
- We know how to compute the dark matter abundance today in an axion model; unless the initial value of a/f_a is very close to zero, most of Ω today will come from dark matter;
- In a heavily dark matter dominated universe, black holes are created instead of galaxies, and life is impossible;

- Therefore while patches of the universe with $\Omega_{\rm dm}$ not very close to unity are very rare, life only can exist in such patches, which is why we think the initial condition for the axion was fine-tuned.

In my mind, this is the only solid anthropic argument particle physicists know how to make; see [33]. This example does not really have anything to do with the fine tuning issues associated with irrelevant operators, but since less solid anthropic arguments have been made to explain the cosmological constant and the Higgs mass, I thought I would show you this example of anthropic reasoning. It is the most solid example I know because it is the only example where we have a thorough understanding of the *ab initio* probability distribution for some fundamental parameter, since we have a complete dynamical model for how the axion comes about.

6.2.3 The Relaxion

A very clever dynamical explanation for why the Higgs is light, sort of along the lines of the parable of the previous lecture, is the relaxion model of [6]. I like this example because it shows one possible way out from the fine tuning problem associated with a relevant operator – the Achilles heel of the EFT approach to quantum field theory – in this case the Higgs mass term.

The basic idea is to take the axion and very weakly break the PQ symmetry explicitly by coupling it to the Higgs. For the Higgs-axion sector we take the Lagrangian

$$\mathcal{L} = (-M^2 + g\phi)|h|^2 - V(g\phi) - \frac{\alpha_s}{16\pi^2}\frac{\phi}{f}\mathrm{Tr}G\widetilde{G} \ , \tag{229}$$

where h is the Higgs, ϕ is the axion, and g is the small symmetry breaking parameter with dimension of mass; the mass M is taken to be the cutoff in the theory (very large compared to the weak scale). The potential V is assumed to be a function only of $g\phi$,with dimensions made up by powers of the cutoff M:

$$V \sim M^2 g\phi + g^2\phi^2 + \frac{g^3\phi^3}{M^2} + \dots \tag{230}$$

You should check that this is not an unnatural potential: none of the terms are small compared to their radiative corrections. This theory does not solve the strong CP problem any more, and the cosmology is somewhat problematic, but it has an very interesting potential for the axion. Suppose the initial value for ϕ is large and negative. Then the Higgs has a large positive mass, gets integrated out of the theory, and the quarks in the SM are massless. In this case the pion is massless and the QCD generated potential for the axion eq. (228) vanishes. So the axion sees the $-M^2 g\phi$ term in its potential and starts rolling toward positive values. As it does so, it makes the Higgs lighter, and eventually it reaches the point where the Higgs mass^2 goes negative and the Higgs gets a vev. At this point the pion becomes massive and the potential eq. (228) turns on, adding ripples to the potential. If it keeps rolling, then the Higgs vev keeps getting bigger, making the potential wiggles bigger. Eventually the axion hits a barrier it can't get over and stops, a point where the $M^2 g\phi$ term balances against the $m_\pi^2 f_\pi^2/f^2 \cos\phi/f$ potential. With appropriately chosen parameters, this can naturally be at a Higgs vev $\langle h\rangle \ll M$.

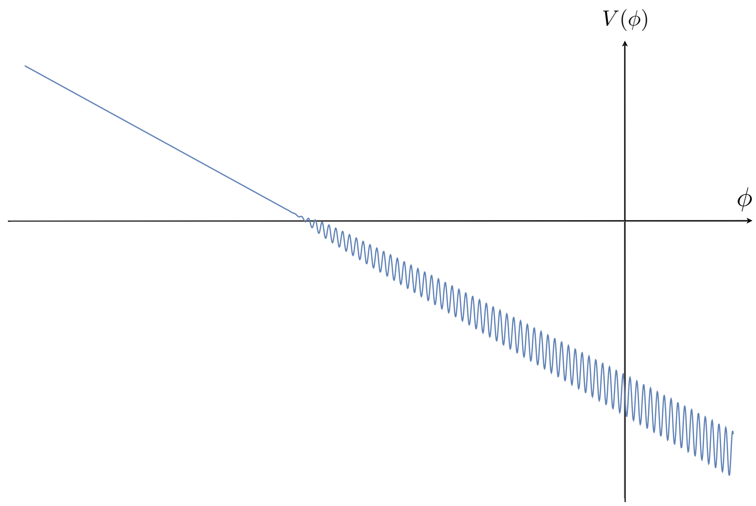

Figure 18: *The relaxion potential, from Ref. [6].*

Note that if the initial condition for ϕ was that it was large and positive initially, then the universe would be stuck in a vacuum with large Higgs vev, perhaps $O(M)$. So the smallness of the Higgs vev in our world, in this model, depends on the dynamical path taken by the ϕ field. There have been a number of papers published about the cosmology and phenomenology of this model, and I encourage you to look at them since this is a new and refreshing idea to address the naturalness problem posed by the light mass of the Higgs field.

References

[1] D. B. Kaplan, *Flavor at SSC energies: A New mechanism for dynamically generated fermion masses*, *Nucl.Phys.* **B365** (1991) 259–278.

[2] K. Agashe, R. Contino, and A. Pomarol, *The Minimal composite Higgs model*, *Nucl.Phys.* **B719** (2005) 165–187, [hep-ph/0412089].

[3] H. Georgi, H. R. Quinn, and S. Weinberg, *Hierarchy of Interactions in Unified Gauge Theories*, *Phys.Rev.Lett.* **33** (1974) 451–454.

[4] S. Weinberg, *Anthropic Bound on the Cosmological Constant*, *Phys.Rev.Lett.* **59** (1987) 2607.

[5] D. B. Kaplan and A. E. Nelson, *Inflationary Axion Cosmology Beyond Our Horizon*, arXiv:0809.1206.

[6] P. W. Graham, D. E. Kaplan, and S. Rajendran, *Cosmological Relaxation of the Electroweak Scale*, *Phys. Rev. Lett.* **115** (2015), no. 22 221801, [arXiv:1504.07551].

[7] J. Polchinski, *Effective Field Theory and the Fermi Surface*, hep-th/9210046.

[8] T. Schaefer, *Quark Matter*, hep-ph/0304281.

[9] S. Weinberg, *Nuclear Forces from Chiral Lagrangians*, *Phys.Lett.* **B251** (1990) 288–292.

[10] S. Weinberg, *Effective Chiral Lagrangians for Nucleon - Pion Interactions and Nuclear Forces*, *Nucl.Phys.* **B363** (1991) 3–18.

[11] E. Epelbaum, H.-W. Hammer, and U.-G. Meissner, *Modern Theory of Nuclear Forces*, *Rev. Mod. Phys.* **81** (2009) 1773–1825, [arXiv:0811.1338].

[12] U. van Kolck, *Nucleon-nucleon interaction and isospin violation*, hep-ph/9711222. [Lect. Notes Phys.513,62(1998)].

[13] P. F. Bedaque and U. van Kolck, *Nucleon Deuteron Scattering from an Effective Field Theory*, *Phys.Lett.* **B428** (1998) 221–226, [nucl-th/9710073].

[14] U. van Kolck, *Effective field theory of short range forces*, *Nucl. Phys.* **A645** (1999) 273–302, [nucl-th/9808007].

[15] D. B. Kaplan, M. J. Savage, and M. B. Wise, *A New Expansion for Nucleon-Nucleon Interactions*, *Phys.Lett.* **B424** (1998) 390–396, [nucl-th/9801034].

[16] D. B. Kaplan, M. J. Savage, and M. B. Wise, *Two Nucleon Systems from Effective Field Theory*, *Nucl.Phys.* **B534** (1998) 329–355, [nucl-th/9802075].

[17] D. B. Kaplan, M. J. Savage, and M. B. Wise, *A Perturbative Calculation of the Electromagnetic Form-Factors of the Deuteron*, *Phys.Rev.* **C59** (1999) 617–629, [nucl-th/9804032].

[18] D. B. Kaplan, M. J. Savage, R. P. Springer, and M. B. Wise, *An effective field theory calculation of the parity violating asymmetry in polarized $np \to d\gamma$*, *Phys.Lett.* **B449** (1999) 1–5, [nucl-th/9807081].

[19] J.-W. Chen, G. Rupak, and M. J. Savage, *Nucleon-Nucleon Effective Field Theory without Pions*, *Nucl.Phys.* **A653** (1999) 386–412, [nucl-th/9902056].

[20] G. Rupak, *Precision Calculation of N P —≫ D Gamma Cross-Section for Big Bang Nucleosynthesis*, *Nucl.Phys.* **A678** (2000) 405–423, [nucl-th/9911018].

[21] H. Arenhovel and M. Sanzone, *Photodisintegration of the deuteron: A review of theory and experiment*, . Springer-Verlag (1991).

[22] S. Weinberg, *Implications of Dynamical Symmetry Breaking*, *Phys.Rev.* **D13** (1976) 974–996.

[23] S. Weinberg, *Implications of Dynamical Symmetry Breaking: An Addendum*, *Phys.Rev.* **D19** (1979) 1277–1280.

[24] L. Susskind, *Dynamics of Spontaneous Symmetry Breaking in the Weinberg-Salam Theory*, *Phys.Rev.* **D20** (1979) 2619–2625.

[25] D. B. Kaplan and H. Georgi, *$SU(2) \times$ U(1) Breaking by Vacuum Misalignment*, *Phys.Lett.* **B136** (1984) 183.

[26] M. J. Dugan, H. Georgi, and D. B. Kaplan, *Anatomy of a Composite Higgs Model*, *Nucl.Phys.* **B254** (1985) 299.

[27] T. Banks, *CONSTRAINTS ON SU(2) x U(1) BREAKING BY VACUUM MISALIGNMENT*, *Nucl.Phys.* **B243** (1984) 125.

[28] R. Contino, C. Grojean, D. Pappadopulo, R. Rattazzi, and A. Thamm, *Strong Higgs Interactions at a Linear Collider*, *JHEP* **1402** (2014) 006, [arXiv:1309.7038].

[29] C. Vafa and E. Witten, *Parity Conservation in QCD*, *Phys. Rev. Lett.* **53** (1984) 535.

[30] R. D. Peccei and H. R. Quinn, *Constraints Imposed by CP Conservation in the Presence of Instantons*, *Phys. Rev.* **D16** (1977) 1791–1797.

[31] S. Weinberg, *A New Light Boson?*, *Phys. Rev. Lett.* **40** (1978) 223–226.

[32] F. Wilczek, *Problem of Strong p and t Invariance in the Presence of Instantons*, *Phys. Rev. Lett.* **40** (1978) 279–282.

[33] M. Tegmark, A. Aguirre, M. Rees, and F. Wilczek, *Dimensionless constants, cosmology and other dark matters*, *Phys. Rev.* **D73** (2006) 023505, [astro-ph/0511774].