HOW TO RENORMALIZE THE SCHRÖDINGER EQUATION

Lectures at the VIII Jorge André Swieca Summer School (Brazil, Feb. 1997)

G. P. LEPAGE
Newman Laboratory of Nuclear Studies, Cornell University
Ithaca, NY 14853
E-mail: gpl@mail.lns.cornell.edu

These lectures illustrate the key ideas of modern renormalization theory and effective field theories in the context of simple nonrelativistic quantum mechanics and the Schrödinger equation. They also discuss problems in QED, QCD and nuclear physics for which rigorous potential models can be derived using renormalization techniques. They end with an analysis of nucleon-nucleon scattering based effective theory.

1 Renormalization Revisited

These lectures are about effective field theories — low-energy approximations to arbitrary high-energy physics — and therefore they are about modern renormalization theory.[1]

Despite the complexity of most textbook accounts, renormalization is based upon a very familiar and simple idea: a probe of wavelength λ is insensitive to details of structure at distances much smaller than λ. This means that we can mimic the *real* short-distance structure of the target and probe by *simple* short-distance structure. For example, a complicated current source $\mathbf{J}(\mathbf{r}, t)$ of size d that generates radiation with wavelengths $\lambda \gg d$ is accurately mimicked by a sum of point-like multipole currents ($E1$, $M1$, etc). In thinking about the long-wavelength radiation it is generally much easier to treat the source as a sum of multipoles than to deal with the true current directly. This is particularly true since usually only one or two multipoles are needed for sufficient accuracy. The multipole expansion is a simple example of a renormalization analysis.

In a quantum field theory, QED for example, the quantum fluctuations probe arbitrarily short distances. This is evident when one computes radiative corrections in perturbation theory. Ultraviolet divergences, coming from loop momenta $k \to \infty$ (or wavelengths $\lambda \to 0$), result in infinite contributions — radiative corrections seem infinitely sensitive to short distance behavior. Even ignoring the infinities, this poses a serious conceptual problem since we don't really know what happens as $k \to \infty$. For example, there might be new supersymmetric interactions, or superstring properties might become important, or electrons and muons might have internal structure. The situation is saved

by renormalization theory which tells us that we don't really need to know what happens at very large momenta in order to understand low-momentum experiments. As in the multipole expansion, we can mimic the complex high-momentum, short-distance structure of the real theory, whatever it is, by a generic set of simple point-like interactions.

The transformation from the real theory to a simpler effective theory, valid for low-momentum processes, is achieved in two steps. First we introduce a momentum cutoff Λ that is of order the momentum at which new as yet unknown physics becomes important. Only momenta $k < \Lambda$ are retained when calculating radiative corrections.[a] This means that our radiative corrections include only physics that we understand, and that there are no longer infinities. Of course we don't really know the scale Λ at which new physics will be discovered, but, as we shall see, results are almost independent of Λ provided it is much larger than the momenta in the range being probed experimentally.

The second step is to add local interactions to the lagrangian (or hamiltonian). These mimic the effects of the true short-distance physics. Any radiative correction that involves momenta above the cutoff is necessarily highly virtual, and, by the uncertainty principle, must occur over distances of order $1/\Lambda$ or less. Such corrections will appear to be local to low-momentum probes whose wavelengths $\lambda \approx 1/p$ are large compared with $1/\Lambda$. Thus the correct lagrangian for cutoff QED consists of the normal lagrangian together with a series of correction terms:

$$\begin{aligned}\mathcal{L}^{(\Lambda)} &= \overline{\Psi}\left(i\partial\cdot\gamma - e(\Lambda)A\cdot\gamma - m(\Lambda)\right)\Psi - \tfrac{1}{4}F_{\mu\nu}F^{\mu\nu} \\ &+ \frac{e(\Lambda)c_1(\Lambda)}{\Lambda}\,\overline{\Psi}\sigma_{\mu\nu}F^{\mu\nu}\Psi \\ &+ \frac{e(\Lambda)c_2(\Lambda)}{2\Lambda^2}\,\overline{\Psi}i\partial_\mu F^{\mu\nu}\gamma_\nu\Psi + \frac{d_2(\Lambda)}{\Lambda^2}\,\left(\overline{\Psi}\gamma_\mu\Psi\right)^2 + \cdots,\end{aligned} \tag{1}$$

where couplings $e(\Lambda)$, $c_1(\Lambda)$, $c_2(\Lambda)$ and $d_2(\Lambda)$ are dimensionless. The correction terms are nonrenormalizable, but that does not lead to problems because we keep the cutoff finite. These new interactions are far simpler to work with than the supersymmetric/superstring/...interactions that they simulate.

The correction terms in cutoff QED modify the predictions of the theory. The modifications, however, are small if Λ is large. Contributions from the $\sigma_{\mu\nu}F^{\mu\nu}$ term, for example, are suppressed by p/Λ, where p is the typical momentum in the process under study. The next two terms are suppressed by $(p/\Lambda)^2$, and so on. In principle there are infinitely many correction terms

[a] For clarity's sake we adopt a simple cutoff as our regulator here. In actual calculations one generally tailors the regulator to optimize the calculation.

in $\mathcal{L}^{(\Lambda)}$, forming a series in $1/\Lambda$; but, when working to a given precision (ie, to a given order in p/Λ), only a finite number of these terms is important. Indeed none of these correction terms seems important in any high-precision test of QED. This indicates that the scale for new physics, Λ, is quite large — probably of order a few TeV or larger.

Expansion (1) provides a useful parameterization for the effects of new physics on low-momentum processes. The form of the cutoff lagrangian is independent of the new physics. It is only the numerical values of the couplings $c(\Lambda)$, $d(\Lambda)\ldots$ that contain information about the new physics. (The couplings are analogous to the multipole moments of a current in our example above.) Thus the implications of a high-precision test of QED can be expressed in a model-independent way as limits on or values for these couplings.

In these lectures I illustrate the powerful techniques of modern renormalization theory in a series of fully worked-out examples. These examples are all based upon the standard Schrödinger equation; they require nothing more than elementary quantum mechanics and some simple numerical analysis. And yet, as I discuss, several important problems in QED, QCD and nuclear physics can be rigorously formulated as potential models using renormalization techniques.

I begin, in Section 2, with an illustration of both nonperturbative and perturbative renormalization in the context of the Schrödinger equation. We will explore many aspects of renormalization familiar from applications in quantum field theory. In particular we will see in detail how to design an effective theory to model a particular set of low-energy data. In Section 3 I discuss the physical conditions that lead to potential models, and, briefly, how they are used in QED and QCD. Finally, in Section 4, I describe how to use our effective potential theory in a systematic analysis of low-energy nucleon-nucleon interactions.

2 Renormalizing the Schrödinger Equation

In this section I illustrate the construction and use of an effective theory by designing one that reproduces a given collection of low-energy data. I begin by discussing the data we will use. I then describe an obvious, well-known procedure for modelling such data, and its limitations. Next I construct a somewhat less obvious effective theory that overcomes these limitations. This modern approach allows us to model the data with arbitrary precision at low energies. My goal throughout is to demonstrate how to use effective theories; I am less interested in proving the formalism mathematically correct. Consequently most examples are solved numerically. Nevertheless the insight gained from analytic calculations is ultimately indispensable. Thus I end this section

with a very simple analytic calculation, using one-loop perturbation theory, that illustrates several of the key ideas.

2.1 Synthetic Data

To begin we need a collection of low-energy data that describe a particular physical system. While we could use real experimental data from a physically interesting system, it is better here to avoid the complexities associated with experimental error by generating "synthetic data." I illustrate the application to real problems later in the lectures, once we have worked through the formalism.

I generated synthetic data for use in these lectures by inventing a simple physical system, and then solving the Schrödinger equation that describes it. I obtained binding energies, low-energy phase shifts, and matrix elements. For the system, I chose the familiar one-particle Coulombic atom, but with a short-range potential $V_s(\mathbf{r})$ in addition to the Coulomb potential:

$$H = \frac{\mathbf{p}^2}{2m} + V(\mathbf{r}) \tag{2}$$

where

$$V(r) = -\frac{\alpha}{r} + V_s(\mathbf{r}). \tag{3}$$

I arbitrarily set $m=1$ and $\alpha=1$.

For our purposes, the short-range potential may be anything; I made one up. The whole point of effective theories is that we can systematically design them directly from low-energy data, with no knowledge of the short-distance dynamics. Thus the form of $V_s(\mathbf{r})$ is irrelevant to our analysis. To underscore this point I will not reveal the functional form of the $V_s(\mathbf{r})$ I used. In what follows we need only know that it has finite range.

Such short-range interactions are common in real Coulombic atoms. The effect of the proton's finite size on the hydrogen atom's spectrum is an example. Another is the weak interaction between the electron and proton, which generates very short-range potentials.

Having chosen a particular $V_s(\mathbf{r})$, I wrote a simple computer code to numerically solve for the radial wavefunctions of energy eigenstates. I used this to compute a variety of binding energies, phase shifts and matrix elements. Some of the S-state binding energies are listed in Table 1. Note that the energies would have been given by $-\alpha^2 m/2n^2 = -1/2n^2$, with $n = 1, 2\ldots$, had there been no V_s. Thus the $1S$ energy is more than doubled by the short-range potential; V_s is not a small perturbation. Also the short-range nature of V_s

Table 1: Binding energies for various S-wave eigenstates of the synthetic hamiltonian.

level	binding energy	level	binding energy
$1S$	1.28711542	$6S$	0.0155492598
$2S$	0.183325753	...	
$3S$	0.0703755485	$10S$	0.00534541931
$4S$	0.0371495726	$20S$	0.00129205010
$5S$	0.0229268241		

Table 2: S-wave phase shifts for the synthetic hamiltonian. Phase shifts are computed for $r = 50$.

energy	phase shift	energy	phase shift
10^{-10}	-0.000421343353	.03	1.232867297
10^{-5}	-0.133227246	.07	-0.579619620
.001	-1.319383451	.1	-1.156444634
.003	0.900186195	.3	-0.106466466
.007	-0.146570028	.7	-1.457426179
.01	-0.654835316	1	1.160634967

is evident in this data since the energies approach the Coulombic energies for the very low-energy, large-n states.

Sample S-wave phase shifts are given in Table 2. These phase shifts depend upon the radius at which they are measured because of the long Coulomb tail in the potential $V(\mathbf{r})$; I arbitrarily chose $r = 50$ for my phase-shift measurements, this being much larger than the Bohr radius, $1/\alpha\, m$, of my atom, and also much larger than the range of $V_s(\mathbf{r})$. (Alternatively, one can compute the phase shift with and without $V_s(\mathbf{r})$, and take the difference; the Coulomb divergence cancels in the difference.) I also computed $\langle \mathbf{p}^4 \rangle$ for several S-states, as well as the wavefunctions at the origin; these are tabulated in later sections.

The numerical analysis required to generate such data is minimal. The computation can easily be done using standard numerical analysis packages or symbolic manipulation programs on a personal computer. I strongly suggest that you invent your own $V_s(\mathbf{r})$ and subject your own synthetic data to the following analysis.

Our challenge is to design a simple theory that reproduces our low-energy data with arbitrarily high precision. We must do this using only the data and knowledge of the long-range structure of the theory (that is, we are given m

and α).

2.2 A Naive Approximation

A standard textbook approach to modelling our data is to approximate the unknown short-range potential by a delta function, whose effect is computed using first-order perturbation theory. The approximate hamiltonian is

$$H_{\rm app} = \frac{\mathbf{p}^2}{2m} - \frac{\alpha}{r} + c\,\delta^3(\mathbf{r}), \tag{4}$$

where c is a parameter. Using first-order perturbation theory, the energy levels in our approximate theory are

$$\begin{aligned} E_n^{\rm app} &= E_n^{\rm coul} + c\left|\psi_n^{\rm coul}(0)\right|^2 \\ &= -\frac{1}{2n^2} + c\,\frac{\delta_{l,0}}{\sqrt{\pi}\,n^3}. \end{aligned} \tag{5}$$

Our approximation has only a single parameter, c. This we must determine from the data. We do this by fitting formula (5) to our lowest-energy data, the $20S$ binding energy; this implies $c = -.5963$. We use the lowest-energy state because that is the state for which the replacement $V_s(\mathbf{r}) \to c\,\delta^3(\mathbf{r})$ is most accurate.

In Figure 1 I show the relative errors in several S-wave binding energies, plotted versus binding energy, obtained using just a Coulomb potential ($c=0$) and using our approximate formula, Eq. (5), with $c = -.5963$. Both approximations become more accurate as the binding energy decreases, but adding the delta function in first order gives substantially better results. Our approximation to V_s is quite successful.

The limitations of this approximation become evident if we seek greater accuracy. We have made two approximations. First we used only first-order perturbation theory. There will be large contributions from second and higher orders in perturbation theory if the short-range potential is strong. So we might wish to compute the second-order contribution to $E_n^{\rm app}$:

$$\sum_{m\neq n} \frac{\langle n|c\,\delta^3(\mathbf{r})|m\rangle\ \langle m|c\,\delta^3(\mathbf{r})|n\rangle}{E_n - E_m}. \tag{6}$$

Unfortunately this expression gives an infinite shift; the sum over scattering eigenstates diverges as scattering momentum $\mathbf{p}\to\infty$. The delta function is too singular to be meaningful beyond first-order perturbation theory.

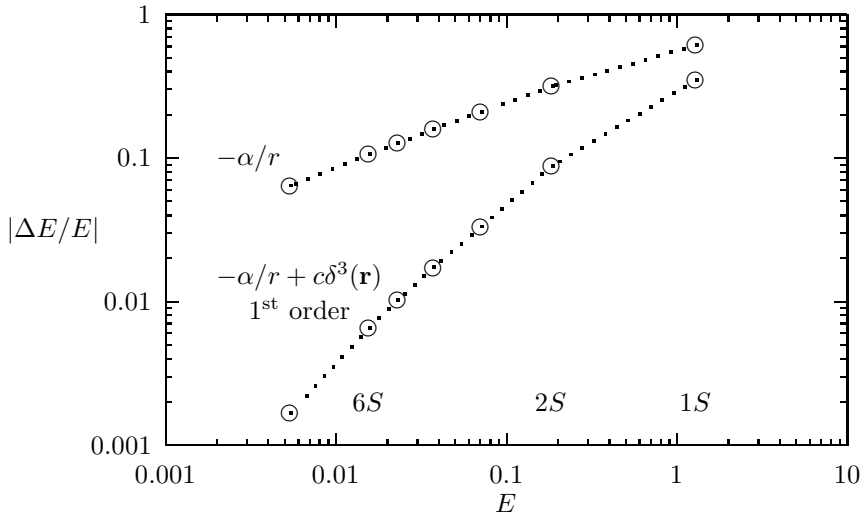

Figure 1: Relative errors in the S-wave binding energies are plotted versus binding energy for the Coulomb theory, and for the Coulomb theory augmented with a delta function in first-order perturbation theory.

The second approximation is replacing V_s by a delta function. The nature of this approximation can be appreciated by Fourier transforming V_s. Since V_s has a very short range, its transform $v_s(q^2)$ depends only weakly on the momentum transfer q. Thus we might try approximating v_s by the first few terms in its Taylor expansion when calculating low-energy matrix elements:

$$v_s(q^2) = v_s(0) + q^2\, v_s'(0) + \cdots . \tag{7}$$

This is an economical parameterization of the short-distance dynamics since it replaces a function, $v_s(q^2)$, by a small set of numbers: $v_s(0)$, $v_s'(0)\ldots$. Keeping just the first term is equivalent to approximating $V_s(r)$ by a delta function. The additional terms correspond to derivatives of a delta function, and correct for the fact that the range of V_s is not infinitely small — the expansion is in powers of q times the range. This suggests that we might improve our approximation by adding a second parameter, again to be tuned to reproduce data:

$$V_s(\mathbf{r}) \to c\,\delta^3(\mathbf{r}) + d\,\nabla^2\delta^3(\mathbf{r}). \tag{8}$$

Unfortunately the additional first-order shift in $E_n^{\rm app}$, $\langle n|d\,\nabla^2\delta^3(\mathbf{r})|n\rangle$, is infinite. Again the delta function is too singular at short distances for such matrix elements to make sense.

Our first attempt to refine the approximate model has failed. We replaced the true, but unknown short-distance behavior of the potential by behavior that is pathologically singular as $r \to 0$. Conventional wisdom, often to be seen in books and even recent conference proceedings, is that our approximations must be abandoned once infinities appear, and that only knowledge of the true potential can get us beyond such difficulties. This wisdom is incorrect, as I now show.

2.3 Effective Theory

The infinities discussed in the previous section are exactly analogous to those found in relativistic quantum field theories. Among other things, they indicate that even low-energy processes are sensitive to physics at short distances. Modern renormalization theory, however, tells us that the low-energy (infrared) behavior of a theory is independent of the *details* of the short-distance (ultraviolet) dynamics. Insensitivity to the short-distance details means that there are infinitely many theories that have the same low-energy behavior; all are identical at large distances but each is quite different from the others at short distances. Thus we can generally replace the short-distance dynamics of a theory by something different, and perhaps simpler, without changing the low-energy behavior.

The freedom to redesign at short-distances allows us to create effective theories that model arbitrary low-energy data sets with arbitrary precision. There are three steps:

1. Incorporate the correct long-range behavior: The long-range behavior of the underlying theory must be known, and it must be built into the effective theory.

2. Introduce an ultraviolet cutoff to exclude high-momentum states, or, equivalently, to soften the short-distance behavior: The cutoff has two effects. First it excludes high-momentum states, which are sensitive to the unknown short-distance dynamics; only states that we understand are retained. Second, it makes all interactions regular at $r = 0$, thereby avoiding the infinities that plague the naive approach of the previous section.

3. Add local correction terms to the effective hamiltonian: These mimic the effects of the high-momentum states excluded by the cutoff in step 2. Each correction term consists of a theory-specific coupling constant, a number, multiplied by a theory-independent local operator. The correction terms systematically remove dependence on the cutoff. Their locality

implies that only a finite number of corrections is needed to achieve any given level of precision.

We now apply this algorithm to design an effective hamiltonian that describes our data.

To begin, our effective theory is specified by a hamiltonian,

$$H_{\rm eff} = \frac{\mathbf{p}^2}{2m} + V_{\rm eff}(\mathbf{r}), \tag{9}$$

where the effective potential, $V_{\rm eff}$, must become Coulombic at large r: $V_{\rm eff}(r) \to -\alpha/r$, with $\alpha = 1$, for large r. We also need an ultraviolet cutoff. I chose to introduce a cutoff into the Coulomb potential through its Fourier transform:

$$\begin{aligned} \frac{1}{r} &\stackrel{\text{F.T.}}{\longrightarrow} \frac{4\pi}{q^2} \\ &\stackrel{\text{cutoff}}{\longrightarrow} \frac{4\pi}{q^2}\, \mathrm{e}^{-q^2a^2/2} \\ &\stackrel{\text{F.T.}}{\longrightarrow} \frac{\mathrm{erf}(r/\sqrt{2}a)}{r}, \end{aligned} \tag{10}$$

where

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \mathrm{e}^{-t^2} dt \tag{11}$$

is the standard error function. The new, regulated potential is finite at $r = 0$, but goes to $1/r$ for $r \gg a$. It inhibits momentum transfers of order $\Lambda \equiv 1/a$ or larger. The exact form of the cutoff is irrelevant; there are infinitely many choices all of which give similar results.

It is not really necessary to regulate the Coulomb potential since it is only mildly singular at the origin. Nevertheless, in many applications, it is still a good idea. For example the potential and the wavefunctions are analytic at $r = 0$ if we use the regulator described above. Numerical techniques are often much more accurate or convergent for analytic functions.

Short-distance dynamics is explicitly excluded from the effective theory by the cutoff. We mimic the effects of the true short-distance structure by adding local correction terms. As discussed in the previous section, the low-momentum behavior of any short-range potential is efficiently described in terms of the Taylor expansion in momentum space. Transforming back to coordinate space gives a series that is a polynomial in the momentum operator $\mathbf{p} \equiv \boldsymbol{\nabla}/\mathrm{i}$ multiplied by a delta function. We need an ultraviolet cutoff to avoid infinities, and therefore we smear the delta function over a volume whose

radius is approximately the cutoff distance a. I chose a smeared delta function defined by

$$\delta_a^3(\mathbf{r}) \equiv \frac{\mathrm{e}^{-r^2/2a^2}}{(2\pi)^{3/2}\, a^3}, \tag{12}$$

but, again, the detailed structure of this function is irrelevant; other choices work just as well.

Remarkably, the structure of the correction terms is now completely determined, even though we have yet to examine the data. The effective potential must have the form

$$\begin{aligned} V_{\text{eff}}(\mathbf{r}) = &-\frac{\alpha}{r}\,\mathrm{erf}(r/\sqrt{2}a) \\ &+c\,a^2\,\delta_a^3(\mathbf{r}) \\ &+d_1\,a^4\,\nabla^2\delta_a^3(\mathbf{r}) + d_2\,a^4\,\boldsymbol{\nabla}\cdot\delta_a^3(\mathbf{r})\boldsymbol{\nabla} \\ &+\cdots \\ &+g\,a^{n+2}\,\nabla^n\delta_a^3(\mathbf{r}) \\ &+\cdots, \end{aligned} \tag{13}$$

where coupling constants c, d_1, $d_2\ldots$ are dimensionless. It consists of a long-range part together with a series of local "contact" potentials. The contact terms are indistinguishable, to a low-momentum particle, from the the true short-distance potential, provided the coupling constants are properly tuned. The potential is nonrenormalizable in the traditional sense, but that is not a problem here since the cutoff prevents infinities.

Generally the effective potential reflects the symmetries of the true theory. Here our effective potential is rotationally invariant because our data are. When the data are not rotationally invariant, additional terms like $a^3\,\mathbf{g}\cdot\boldsymbol{\nabla}\,\delta_a^3(\mathbf{r})$ must be included in V_{eff}, where now the coupling constants include vectors, tensors... that characterize the rotational asymmetries of the true theory.

One might worry that our Taylor expansion of the short-distance dynamics would fail when we computed physical quantities like the scattering amplitude. This is because high-momentum states affect even low-energy processes through quantum fluctuations. For example, in the true theory the scattering amplitude is

$$\langle f|T(E)|i\rangle = \langle f|V|i\rangle + \sum_n \frac{\langle f|V|n\rangle\,\langle n|V|i\rangle}{E-E_n} + \cdots. \tag{14}$$

The sums over intermediate states in second-order and beyond include states with arbitrarily large momentum. A momentum-space Taylor expansion of the

potential would not converge in matrix elements involving such states. Furthermore when we replace the exact potential V by our effective potential V_{eff}, our ultraviolet cutoff in effect limits the sums to states with momenta less than the cutoff $1/a$; contributions from high-momentum intermediate states apparently are completely absent from our effective theory. Our effective theory is saved by the fact that high-momentum intermediate states are necessarily highly virtual if, as we assume, the initial and final states have a low energy. By the uncertainty principle, such states cannot propagate for long times or over large distances. Thus any contribution from these states is very local and can be incorporated into the effective theory using the same set of (smeared) delta-function potentials already included in V_{eff}. In this way, the high-momentum states that are explicitly excluded, or badly distorted, by the cutoff are included implicitly through the coupling constants. Note that this means the coupling constants in our effective theory depend nonlinearly on the true potential $V(r)$.

While the true theory is obviously independent of the value of the cutoff a, results computed in the effective theory are only approximately a independent. The residual a dependence in such a result is typically a power series in qa, where q is the characteristic momentum associated with the process under study (for example, the initial momentum or the momentum transfer in a scattering amplitude). The contact terms remove these a-dependent errors order-by-order in a. Thus, for example, the $\mathcal{O}(a^2)$ term in V_{eff} removes errors of order $(qa)^2$ and is the most important correction. Errors of order a^4 are removed by the $\mathcal{O}(a^4)$ terms, of which there may be only two since there are only two independent ways of combining two momentum operators $\boldsymbol{\nabla}/\mathrm{i}$ with the smeared delta function. Obviously only a finite number of contact terms is needed to remove errors through any finite order in a.

The coupling constants vary with a since they must account for quantum fluctuations excluded by the cutoff. More or less is excluded as a is increased or decreased, and therefore the coupling constants must be adjusted to compensate. They are said to be "running coupling constants." When the short-distance potential is weak, the coupling to high-momentum intermediate states is weak and the coupling constants change only slowly with a.[b] On the other hand, the coupling constants are often very cutoff dependent when the short-distance interactions are strong. The dependence of the effective theory on its coupling constants becomes highly nonlinear in this case. In fact this

[b]This behavior is modified if the cutoff distance a is very large. When a is larger than the range of the *long-range* potential, or when there is no long-range potential, the coefficients of the contact operators tend to go to a constant. Thus for example, c in V_{eff} will decrease like $1/a^2$ as a increases so that the coefficient of the delta function operator becomes constant.

behavior can be so strong that the relative importance of different contact terms can change: an a^4 operator can act more like an a^2 operator, and vice versa. The physical systems I discuss here do not suffer from this complication, although it is easy to study the phenomenon by modifying the synthetic data appropriately.

Highly nonlinear behavior also results when the cutoff distance a is made too small. From the previous section we know that taking $a \to 0$ is a bad idea. In general it makes little sense to reduce a below the range r_s of the true potential. By assumption, our data involves energies that are too low — wavelengths that are too long — to probe the true structure of the theory at distances as small as r_s. When $a < r_s$, high-momentum states are included that are sensitive to structure at distances smaller than r_s. But the structure they see there is almost certainly wrong. Thus taking a smaller than r_s cannot improve results obtained from the effective theory. In fact, as the nonlinearities develop for small a's, results often degrade, or, in more extreme cases, the theory may become unstable or untunable.

Finally I should comment on the physical significance of the cutoff. A common practice in applications of potential models, for example to nuclear or atomic physics, is to use a cutoff like our's to account for some physical effect, for example finite nuclear size. Then one must worry about whether the functional form of the cutoff is physically correct. We are *not* doing this here. Our cutoff is only a cutoff; physical effects like finite nuclear size are incorporated systematically through the contact terms. We need never worry, for example, about what the true charge distribution is inside a nucleus. Indeed this is the great strength of the effective theory. The effective theory gives us a simple, *universal* parameterization for the effects of short-distance structure. The form of the contact terms is theory independent. Only the numerical values of the coupling constants c, $d_1 \ldots$ are theory specific; for our (low-energy) purposes, everything that we need to know about the short-distance dynamics of the true theory is contained in these coupling constants. As discussed earlier, this situation is conceptually similar to a multipole analysis of a classical field. The couplings here are the analogues of the multipole moments, while the delta-function potentials are analogous to the multipoles that generate the field.

The exact form of the cutoff, and the exact definition of the smeared delta function are irrelevant. I encourage you to make up your own versions of these and repeat the analysis given here. You will get similar results for binding energies and phase shifts, although all of your coupling constants will probably be different, and possibly quite different.

2.4 Tuning the Effective Theory; Results

We now tune the parameters of our effective theory so that it reproduces our low-energy data through order a^4. For simplicity, we examine only S-wave properties. Thus we need only the c and d_1 terms in $V_{\rm eff}$, Eq. (13); we can drop the d_2 term in $V_{\rm eff}(\mathbf{r})$ since it is important only for P-wave states.[c]

To generate the results discussed here I first chose a reasonable value for the cutoff distance: to begin with, $a = 1$, the Bohr radius for the Coulombic part of the interaction. Then I varied the coupling constants c and d_1 until the S-wave phase shifts from the effective hamiltonian agreed with the data at energies $E = 10^{-5}$ and 10^{-10}. I found $c/4\pi = -3.18$ and $d/4\pi = -0.199$. Having found the coupling constants, I then generated binding energies, phase shifts, and matrix elements using the effective theory. I also generated results with only the a^2 correction ($d_1 = 0$); the effective theory requires a different value of c in this case, since the two contact terms affect each other.

I tuned the couplings using phase shifts at very low energies. This was to minimize the effect of the a^6 errors, which arise because we truncated our effective potential at order a^4. These "truncation" errors are smallest for the lowest-energy data, since they are generally proportional to qa raised to some positive power. In general one should use the most infrared data available when tuning. Alternatively one might attempt a global fit to all of the data; but then it is crucial to give greater weight in the fit to low-energy data than to high-energy data: some estimate of the $(qa)^n$ error due to truncation must be added to the experimental errors used to weight the data in the fit.[2]

So long as $a > 0$, $V_{\rm eff}(\mathbf{r})$ is a simple, nonsingular function of r. Consequently the effective theory is simple to solve numerically. To compute the results here, I reused the code that I wrote to generate the synthetic data, but with the true potential $V(\mathbf{r})$ replaced by the effective potential $V_{\rm eff}(\mathbf{r})$. The effective theory is no harder to solve than the other; in particular, because of the cutoff, there are no infinities.

In Figure 2 I compare the errors in the binding energies obtained from the tuned effective theory with those obtained in Section 2.2 from first-order perturbation theory. Even with only the a^2 correction there is a sizable reduction in errors, due to contributions from second and higher orders in the correction — our numerical solution of the effective theory is nonperturbative, and so includes all orders. The accuracy at low energies is further enhanced by the

[c] It is obvious that the d_2 term couples only to P-waves in the limit $a \to 0$. When a is nonzero, however, this term has a small residual coupling to S-waves. This results in an interaction of order a^6, which may be ignored to the order we are working here. One can easily design contact terms that contribute only to a single channel in orbital angular momentum. For example, our smeared delta function could be replaced by a local potential that is separable.

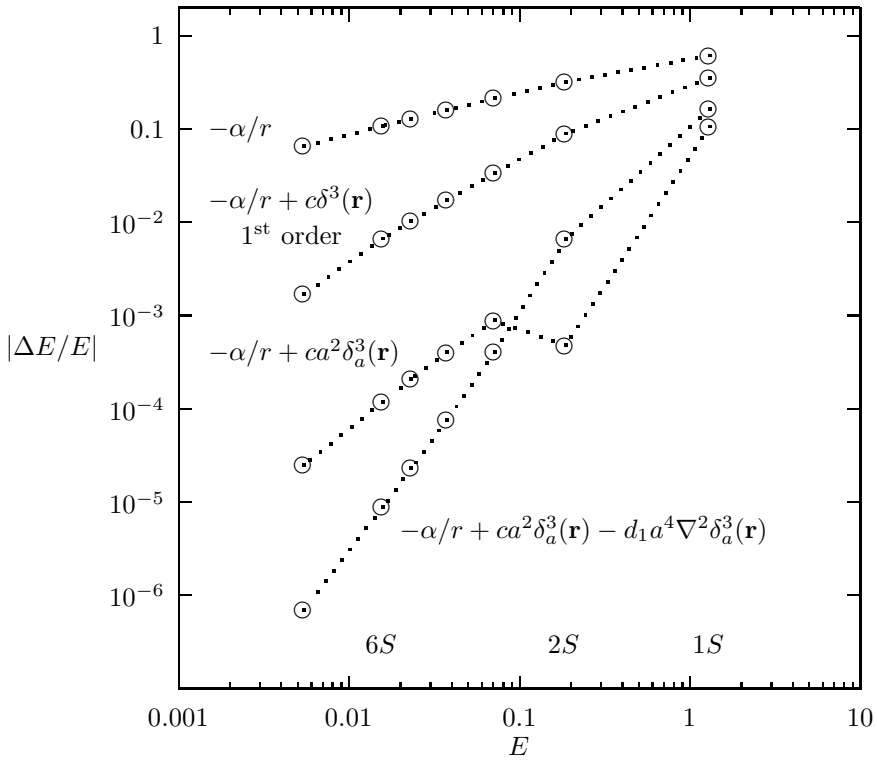

Figure 2: Relative errors in the S-wave binding energies are plotted versus binding energy for the Coulomb theory, the Coulomb theory augmented with a delta function in first-order perturbation theory, the nonperturbative effective theory through a^2, and the effective theory through a^4.

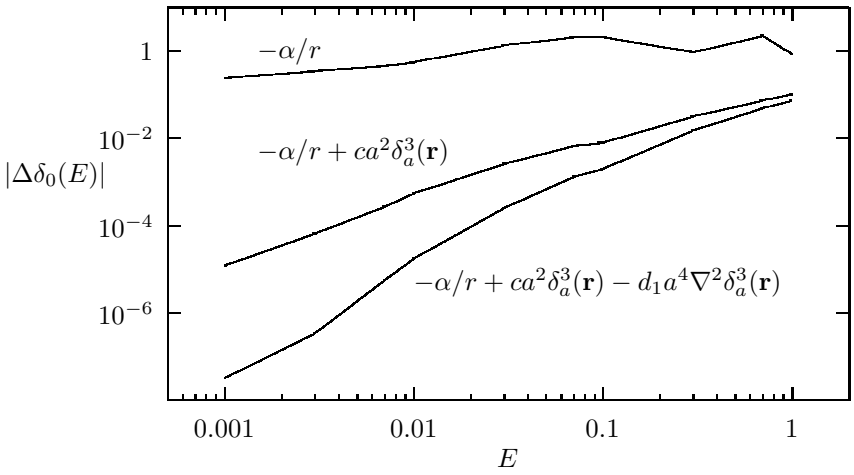

Figure 3: Errors in S-wave phase shifts (in radians) computed with the effective theory corrected through order a^2 and a^4. Values are for $a = 1$ and are plotted versus energy. Results are also shown for the theory with no contact terms.

a^4 correction, which begins to account for the finite range of the short-range interaction. The crossing of the curves for the a^2 and a^4 theories around $E=0.1$ has no significance; the error in the a^2 theory changes sign there, passing through zero.

The phase shifts tell a similar story; see Figure 3. At low energies, the errors decrease steadily as the correction terms are added, order-by-order in a^2. The slope of the error curve changes as each new correction is added, getting steeper by one power of $E \propto p^2$ each time the order of the error is increased by a^2, as expected. Similar behavior is apparent in the previous figure, which shows the errors in the binding energies.

Both of these figures show that the correction terms have little effect at energies $E>1$. At these energies, the particle's wavelength is sufficiently short that the particle can probe the detailed structure of $V_s(\mathbf{r})$. Effective theories are useless in this limit. To go much beyond $E=1$ in this example, one must somehow uncover the true short-distance structure of the theory.

I redid the analysis for a variety of different a's to explore cutoff dependence. Figure 4 shows how the error in the phase shift depends on a. As a is reduced from $10\to3\to1$, the errors decrease at all but the highest energies. For this range, a is larger than the range r_s of the true short-distance interaction, and therefore the errors are proportional to a power of qa, which decreases with

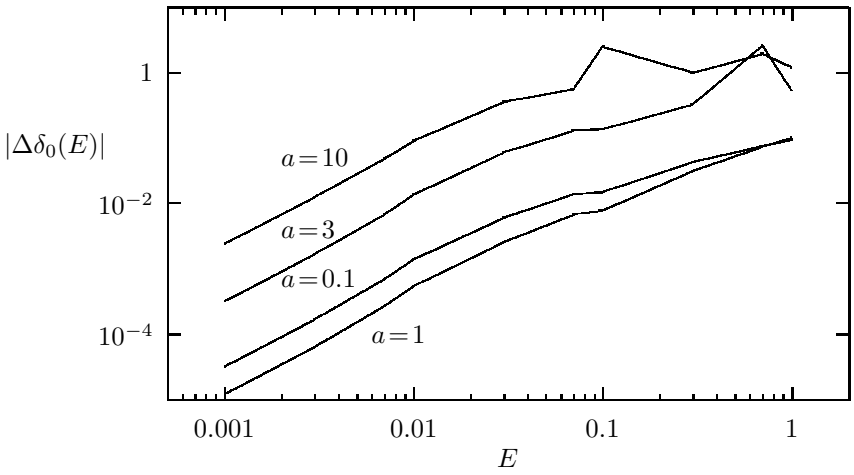

Figure 4: The dependence on cutoff of the$S$-wave phase shift errors plotted against energy is shown. Errors decrease as a is reduced until $a \approx 1$, beyond which point there is no improvement. The a^2 theory was used here ($d_1 = 0$).

decreasing a. The errors stop decreasing for $a < 1$, becoming almost insensitive to a. From our discussion in the previous section, this suggests that $r_s \approx 1$. The coupling constants also vary with a in the manner expected if $r_s = 1$. For example, with $d_1 = 0$, coupling constant $c/4\pi$ is roughly -3 and only weakly dependent on a when $a \geq 1$. It grows rapidly as a is reduced below one; for example, $c = -74.0$ when $a = 0.01$.

I also looked for multiple solutions for the coupling constants.[3] Setting $d_1 = 0$ and $a = 1$, for example, I found that the phase shift at $E = 10^{-10}$ is correct when $c/4\pi$ equals -12.4 or 1.6, as well as when it equals -3.18, the value used above. These extra solutions give reasonably good results for the phase shifts and binding energies, but not as good as our original value. The reason is that with these new c's the theory's short-distance structure is significantly modified. In particular, with $c = 24.7$, there is an extra bound state with a very high binding energy (≈ 6). We do not expect our effective theory to work well at high energies, so there is nothing wrong with having the extra state; the usual $1S$, $2S \ldots$ states are all still present, each with an extra node in its wavefunction at short distances. When $c/4\pi = 1.6$, the $1S$ state disappears, while the other states are missing a node at short distances.

I find it remarkable that a simple two-parameter effective theory, tuned using only two very low energy phase shifts, can reproduce the full range of

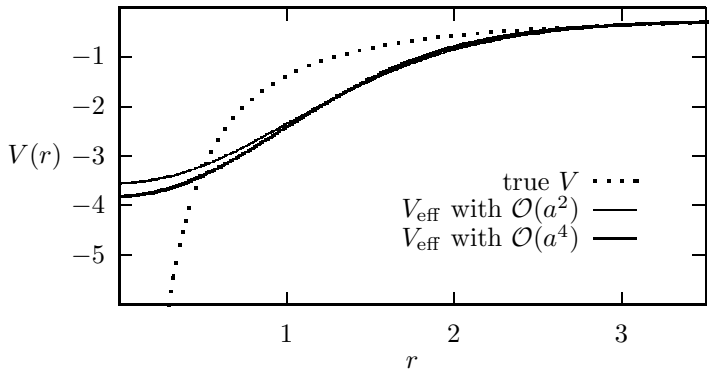

Figure 5: The true potential is plotted versus r together with the a^2 and a^4 effective potentials ($a=1$).

our data so successfully. The binding energy of the $2S$ state, for example, is accurately predicted by the effective theory to better than 1%, despite the fact that the cutoff scale $a=1$ is almost a quarter the radius of the atom. Our most infrared data is reproduced to better than six digits.

2.5 Misconceptions

It is worth pausing for a moment to address common misconceptions concerning the procedure outlined here. One common notion is that our effective potential will eventually become identical to the true potential, as more and more higher-order corrections are added. This is false. There are infinitely many theories that share the low-energy behavior of our true theory; the low-energy data that we have, no matter how precise, does not contain enough information to specify completely the high-energy structure of the underlying theory. Our procedure allows us to construct one of the infinitely many theories; it is extremely unlikely that we would happen onto the true theory.

In Figure 5 I compare the true potential, $V(\mathbf{r})$, with the $V_{\rm eff}$ having only the a^2 correction, and with that having both a^2 and a^4 corrections. The two $V_{\rm eff}$'s look almost identical, and show no sign of converging to V. Nevertheless the a^4 theory gives much more accurate results than the a^2 theory.

Another misconception, suggested by the discussion leading to Eq. (8), is that the Fourier transform of $V_{\rm eff}$ converges to that of V for momenta $q \ll 1/a$. This is not the case because, as we discussed, the effective potential also corrects for high-momentum intermediate states omitted by the cutoff, and so

Table 3: Expectation values of $\mathbf{p}^4$ for S-states in the true theory, and in the effective theory ($a = 1$). These are compared with matrix elements of the renormalized operator in the effective theory that corresponds to $\mathbf{p}^4$ in the true theory.

level	$\langle \mathbf{p}^4 \rangle$	$\langle \mathbf{p}^4 \rangle_{\text{eff}}$	$\langle Z\mathbf{p}^4 + \gamma\delta_a^3/a + \cdots \rangle_{\text{eff}}$
$1S$	69	5.9	28
$2S$	5.50	1.7	5.34
$3S$	1.309	0.4	1.306
$4S$	0.5070	0.17	0.5068
$5S$	0.24740	0.08	0.24738
$6S$	0.138784	0.05	0.138780

is related in a nonlinear way to the true potential.

Another observation that is sometimes made is that our procedure is mere curve fitting — if you add more parameters, you get better answers. Indeed it is curve fitting, but highly optimized curve fitting and systematic: errors are removed order-by-order in a. At each order, renormalization theory tells us how many parameters are needed, and what form the corrections should have. That is, it tells us which parameters are worth tuning. So while one could treat our a^2 theory as a two-parameter theory, varying c and a to get the best fit, renormalization theory says that a much better two-parameter theory is obtained by freezing a at an appropriate value, and then tuning c and d_1 in our a^4 theory. And it tells us roughly how much better as well.

2.6 Operators and the Operator Product Expansion

So far we have used our effective theory to compute binding energies and phase shifts. We now examine quantities that depend in detail on the wavefunctions. Consider, for example, the matrix element $\langle n|\mathbf{p}^4|n\rangle$, which might be important if we wished to include relativistic corrections in our potential model. In Table 3 I list values of this matrix element for several S-states both for the true theory, and for our corrected theory (with $a=1$). The values disagree by more than a factor of two, even for very low-energy states, despite the fact that the two theories agree on the corresponding binding energies to several digits.

The problem is that the operator in the effective theory that corresponds to $\mathbf{p}^4$ in the true theory is not $\mathbf{p}^4$. As is true of the hamiltonian, there are local corrections to $\mathbf{p}^4$ in the effective theory. Thus, for any S state, we expect

$$\langle \mathbf{p}^4 \rangle_{\text{true}} = Z \langle \mathbf{p}^4 \rangle_{\text{eff}} + \frac{\gamma}{a} \langle \delta_a^3(\mathbf{r}) \rangle_{\text{eff}} + \eta\, a \langle \nabla^2 \delta_a^3(\mathbf{r}) \rangle_{\text{eff}} + \mathcal{O}(a^3) \qquad (15)$$

Table 4: Wavefunctions evaluated at $r=0$ for S-states in the true theory, and in the effective theory ($a=1$). These are compared with the renormalized expression in the effective theory that corresponds to the wavefunction at the origin in the true theory.

level	$\psi(0)$	$\psi_{\rm eff}(0)$	$\overline{\gamma}\int\psi_{\rm eff}\delta_a^3+\cdots$
$1S$	1.50	0.53	-3.4
$2S$	0.383	0.19	0.369
$3S$	0.1837	0.09	0.1830
$4S$	0.11353	0.06	0.11344
$5S$	0.079005	0.04	0.078986
$6S$	0.059031	0.03	0.059025

where Z, γ and η are dimensionless constants that are independent of the state. I tuned these constants so that Eq. (15) gave correct values for the $10S$, $15S$ and $20S$ bound states, and found $Z=1$, $\gamma=-96.2$ and $\eta=-140.6$ when $a=1$. This formula gives excellent results, as shown in Table 3. Just as for the binding energies, the relative error decreases steadily with decreasing energy.

This example illustrates the subtlety of the equivalence between the effective theory and the true theory. Although the infrared spectra and phase shifts are quite similar, the short-distance structure of corresponding wavefunctions in the two theories may be completely different. This means that matrix elements of operators sensitive to short distances, like $\mathbf{p}^4$, are not equal, even for very low-energy states. Renormalization theory, however, tells us how to define correction terms for the operators in the effective theory that correct for the differences in short-distance behavior between the two theories. Just as in the hamiltonian, the corrections are local operators with state-independent coupling constants, and, again, only a finite number of terms is needed to any given order in a.

Another example of the same idea is provided by the wavefunction at the origin, which is often needed in phenomenological work. In Table 4 I list values of the wave function at the origin for several S-wave bound states. Again the effective theory gives results that are completely different from the true theory. Renormalization theory, however, implies that

$$\begin{aligned}\psi_{\rm true}(r=0) = \overline{\gamma}\int d^3r\,\psi_{\rm eff}\,\delta_a^3(\mathbf{r}) \\ +\,\overline{\eta}\,a^2\int d^3r\,\psi_{\rm eff}\,\nabla^2\delta_a^3(\mathbf{r}) \\ +\,\mathcal{O}(a^4)\end{aligned} \tag{16}$$

where $\overline{\gamma}$ and $\overline{\eta}$ are dimensionless, state-independent constants. For $a = 1$, I found $\overline{\gamma} = -28.53$ and $\overline{\eta} = -3.615$ by tuning this formula to give correct results for the $20S$ and $15S$ bound states. With these values, the results from the effective theory for $\psi_{\rm true}(0)$ are given in the last column of Table 4. Again the agreement is excellent for all but the most ultraviolet of the states, and it gets better as the states are more infrared — just as expected.

The last formula is a special case of one valid for any $r \le a$:

$$\begin{aligned}\psi_{\rm true}(r) = \overline{\gamma}(r) & \int d^3r\, \psi_{\rm eff}\, \delta_a^3(\mathbf{r}) \\ & + \overline{\eta}(r)\, a^2 \int d^3r\, \psi_{\rm eff}\, \nabla^2 \delta_a^3(\mathbf{r}) \\ & + \mathcal{O}(a^4),\end{aligned} \tag{17}$$

where coefficient functions $\overline{\gamma}(r)$ and $\overline{\eta}(r)$ are state-independent. This formula is striking. It says that given only the two coefficient functions, we can use the effective theory to compute the small r behavior of *any* low-energy S-state, bound or otherwise. All the information needed concerning the short-distance behavior of the true theory is summarized in the coefficient functions; everything we need to know about the long-distance structure of the state is given by two numbers, the values of the two integrals. This separation into short-distance and long-distance factors is often referred to as "factorization."

Factorization is most useful when either the short-distance or long-distance physics is particularly simple. For example, Eq. (17) is valid for the quark-antiquark wavefunction describing an ψ/J meson. The long-distance structure of the wavefunction cannot be easily computed using QCD, but the short-distance coefficient functions are readily calculated using perturbation theory, since QCD becomes perturbative at short distances. Thus we know the small r behavior of the wavefunction up to two parameters — the values of the integrals over the unknown wavefunction. This is enough information to allow us to compute the momentum-dependence of form factors involving the Υ at high momentum.[4]

Eq. (17) may be modified to include a term $\overline{Z}\,\psi_{\rm eff}(r)$ on the right-hand side. Then the relation works for large r's as well, since $\psi_{\rm eff}(r) \propto \psi(r)$ for $r \gg a$.

Formula (17) is an example of the "operator product expansion" or OPE, which is an important tool in modern field theory.[d] It is reminiscent of a Taylor expansion: the small x behavior of an analytic function $f(x)$ may be specified

[d] The operators forming the "product" here are the field operators associated with the two bound particles. In quantum field theory, the wavefunction is a matrix element of the product of these operators.

by a set of known functions, $1, x, x^2 \ldots$, and a set of constants, $f(0), f'(0) \ldots$. But our expansion is not a Taylor expansion. The wavefunctions in our true theory are not analytic at $r=0$: $\nabla^2\psi$ is infinite there because, as it happens, $V(0) = -\infty$. Indeed the wavefunction diverges at the origin if, for example, the potential includes an attractive $1/r^2$ potential. Eq. (17) still works in such cases; the divergent behavior is captured by the coefficient functions. In this particular case one can show that the leading coefficient function contain a powerlaw divergence of the form $(a/r)^\xi$. Parameter ξ is referred to as an "anomalous dimension."

2.7 Perturbative Derivation

Effective theories are useful in two common situations. In one, the short-distance behavior of the true theory is unknown, as above, but low-energy data is available from which we can systematically construct the low-energy effective theory. In the other, the short-distance behavior is completely understood, but a low-energy approximation is sought in order to simplify calculations of low-energy behavior. This latter situation is encountered in applications of both QCD and QED. In this section we illustrate the second situation: We construct an effective theory to model the low-energy behavior of a theory whose short-distance, high-momentum behavior is perturbative. We can calculate the coupling constants in the effective theory using perturbation theory because the short-distance dynamics is perturbative. Our simple analysis illustrates how and why effective theories work.

To keep things as simple as possible, we take the Schrödinger-Coulomb problem as our underlying theory,

$$H = \frac{\mathbf{p}^2}{2m} - \frac{\alpha}{r} \tag{18}$$

with $m = 100$ and $\alpha = .01$. We model the low-energy behavior of this theory using our standard effective theory:

$$H_{\text{eff}} = \frac{\mathbf{p}^2}{2m} - \frac{\alpha}{r}\,\text{erf}(r/\sqrt{2}\,a) - 2\pi\alpha\, c\, a^2\, \delta_a^3(\mathbf{r}). \tag{19}$$

The challenge is to determine the value of coupling constant c using perturbation theory.

The technique by which we determine c is called "perturbative matching," and is standard in QED and QCD applications. We compute the scattering amplitude $T(\mathbf{p}, \mathbf{q})$, where $\mathbf{p}$ is the initial momentum and $\mathbf{q}$ the momentum transfer, using perturbation theory in the underlying theory. We then compute the same scattering amplitude using perturbation theory in the effective

theory, to obtain an amplitude $T_{\rm eff}(\mathbf{p},\mathbf{q},c,a)$ that also depends on the unknown coupling constant c and the cutoff a. We determine c by requiring

$$T(\mathbf{p},\mathbf{q}) = T_{\rm eff}(\mathbf{p},\mathbf{q},c,a) \times \left(1+\mathcal{O}((p\,a)^4)\right). \tag{20}$$

Given a, this equation can be solved for c in the limit $p\,a\to 0$, order-by-order in α.

Note that perturbation theory doesn't converge for either of these amplitudes in the limit $p\to 0$, because of the (infrared) Coulomb threshold singularity.[e] Coupling constant c, however, has a convergent perturbation series, since it is sensitive only to the short-distance behavior in the theory. Thus, for our purposes, we can compute the scattering amplitudes using ordinary perturbation theory, ignoring the infrared problems of the expansion. When we solve for c, the infrared parts of T and $T_{\rm eff}$ cancel completely, leaving behind a convergent, infrared-free perturbation expansion. This works because, although the infrared parts of the perturbative expansions for T or $T_{\rm eff}$ may be completely incorrect, the ultraviolet parts are fine, and that is all we need to compute c. This point is essential in dealing with QCD where, because of confinement, perturbation theory is always inapplicable to low-energy on-shell quark scattering — except in calculations of the sort we are doing here.

In lowest-order perturbation theory (the Born approximation), the two amplitudes are

$$T^{(1)}(\mathbf{q}) = -\frac{4\pi\alpha}{q^2} \tag{21}$$

and

$$\begin{aligned} T^{(1)}_{\rm eff}(\mathbf{q}) &= -\frac{4\pi\alpha}{q^2}\,{\rm e}^{-q^2a^2/2}\,\left(1+c\,q^2a^2/2\right) \\ &= -\frac{4\pi\alpha}{q^2}\,\left(1+(c-1)\,q^2a^2/2+\mathcal{O}(q^4a^4)\right). \end{aligned} \tag{22}$$

Thus the two amplitudes match to the requisite order if $c=1$. The lowest-order analysis is complete. We can verify our result using the same computer code used in the previous sections, with suitably modified potentials. I calculated the error in the $1S$ binding energy when $a=.01$ using the effective theory with different values of c. The result,

$$\text{error in } 1S \text{ energy} = \begin{cases} .04\% & \text{for } c=0 \\ .0004\% & \text{for } c=1 \end{cases}, \tag{23}$$

[e] Perturbation theory in a potential $V(r)$ generally fails near threshold if either potential V or potential $-V$ has bound states.

shows substantial improvement when the lowest-order contact term is included.

In second-order perturbation theory, the scattering amplitude in the full theory is

$$T^{(2)} = \int \frac{d^3k}{(2\pi)^3}\, T^{(1)}(\mathbf{k}-\mathbf{q})\, \frac{2m}{p^2-(\mathbf{k}+\mathbf{p})^2+\mathrm{i}\epsilon}\, T^{(1)}(\mathbf{k}), \tag{24}$$

while $T^{(2)}_{\rm eff}$ is given by the same expression but with $T^{(1)} \to T^{(1)}_{\rm eff}$. We must compare these two amplitudes. When loop momentum k is small compared with $1/a$, the integrands of the two amplitudes become almost identical since, by construction,

$$T^{(1)}(\mathbf{k}) = T^{(1)}_{\rm eff}(\mathbf{k})\left(1+(k^4a^4)\right). \tag{25}$$

The two integrands, however, differ significantly when k is of order $1/a$ or larger. Thus we can neglect p and q relative to k when computing the difference between $T^{(2)}_{\rm eff}$ and $T^{(2)}$, thereby significantly simplifying the calculation:

$$\begin{aligned} T^{(2)}_{\rm eff} - T^{(2)} &\approx \int \frac{d^3k}{(2\pi)^3} \left[\left(\frac{-4\pi\alpha}{k^2}\,\frac{-2m}{k^2}\,\frac{-4\pi\alpha}{k^2}\right) \mathrm{e}^{-k^2a^2}(1+k^2a^2/2)^2 \right. \\ &\qquad \left. - \left(\frac{-4\pi\alpha}{k^2}\,\frac{-2m}{k^2}\,\frac{-4\pi\alpha}{k^2}\right)\right] \\ &= -2\pi\alpha^2\, a\, m\, 0.940 \end{aligned} \tag{26}$$

The second-order amplitudes do not agree, and therefore we need an extra correction in the effective theory. Since the difference between the amplitudes is momentum independent, we can correct for it by modifying the value of coupling constant c in the effective theory:

$$c = 1 + 0.94\,\alpha\, a\, m. \tag{27}$$

Then there is an $\mathcal{O}(\alpha^2)$ contribution from the tree-level amplitude, Eq. (22), that cancels the error in $T^{(2)}_{\rm eff}$. The coupling constant is said to have been "renormalized" by one-loop quantum fluctuations. Again we can verify our one-loop correction numerically; the result,

$$\text{error in } 1S \text{ energy} = \begin{cases} 4\times10^{-2}\,\% & \text{for } c=0 \\ 4\times10^{-4}\,\% & \text{for } c=1 \\ 8\times10^{-7}\,\% & \text{for } c=1.0094 \end{cases}, \tag{28}$$

is greatly improved.

Our analysis illustrates why the renormalization of c depends on high-energy details of the theory: in Eq. (26), the entire contribution is from loop

momenta of order $1/a$ or larger. This analysis also shows why the form of the correction terms is universal. The difference $T^{(2)}_{\rm eff} - T^{(2)}$ comes from loop momenta $k \approx 1/a \gg p, q$. Thus in general we can Taylor expand the integrand in powers of the external momenta to obtain ultimately

$$T^{(2)}_{\rm eff} - T^{(2)} = c_0 + c_2\, q^2 a^2 + \cdots . \tag{29}$$

The c_0 term is corrected by a $\delta(\mathbf{r})$ potential, the c_2 term by a $\nabla^2\delta(\mathbf{r})$ potential, and so on. The form of these potential is dictated by Taylor's theorem, and is independent of the actual short-distance physics.

Note that the integral in Eq. (26) becomes infrared divergent if one expands the integrand beyond leading order in p/k or q/k. This is because we have only matched the $T^{(1)}$'s through order q^2a^2, while the next-to-leading error in $T^{(2)}_{\rm eff}$ contributes in order q^4a^4. This is inconsistent, and consistency is essential. In general the expansion of a one-loop correction works if carried out to the same order in a as the tree-level analysis, but not beyond that order. The appearance of an infrared divergence in the calculation of a coupling constant is a sure sign that something has been omitted from the effective theory at lower order.

Our earlier analyses of operators like $\mathbf{p}^4$, and of the wavefunction at small r can be repeated for the perturbative theory, using perturbation theory to calculate the coupling constants and coefficient functions. I invite you to try it.

3 Rigorous Potential Models

Potential models are widely used as approximations to more complex quantum field theories. For example, the QCD dynamics of a heavy-quark meson like the ψ/J or the Υ is described by a phenomenological quark potential model. Similarly the dynamics of the electron and positron in positronium are well approximated by a simple Schrödinger equation. Another example, which we discuss in a later section, is the nucleon-nucleon force, which is often described in terms of a potential that somehow must come from QCD. In most such applications the potential model is regarded as a useful approximation, but not one that might serve as the basis for a rigorous, systematic analysis. In fact, potential models can often be given a rigorous formulation as effective theories.

3.1 Positronium

The defining characteristic of a potential model is that it treats a system as though it has a fixed number of constituents. Taking positronium as our example, the potential model describes the atom as a bound state of an electron and a positron. In reality, however, positronium is far more complex. In addition to the $e\overline{e}$ component, a positronium state has many other components, each with its own wavefunction:

$$|\mathrm{Ps}\rangle = \sum_{e\overline{e}} |e\overline{e}\rangle\, \psi_{e\overline{e}} + \sum_{e\overline{e}\gamma} |e\overline{e}\gamma\rangle\, \psi_{e\overline{e}\gamma} + \sum_{e\overline{e}\gamma\gamma} |e\overline{e}\gamma\gamma\rangle\, \psi_{e\overline{e}\gamma\gamma} + \cdots. \tag{30}$$

The probability for finding the atom in any one of these states can be computed from the corresponding wavefunction in the usual way, with the constraint that the sum of all probabilities is one:

$$1 = \sum_{e\overline{e}} |\psi_{e\overline{e}}|^2 + \sum_{e\overline{e}\gamma} |\psi_{e\overline{e}\gamma}|^2 + \sum_{e\overline{e}\gamma\gamma} |\psi_{e\overline{e}\gamma\gamma}|^2 + \cdots. \tag{31}$$

One wonders why, and to what extent, we are allowed to ignore all but the first term in this "Fock state" expansion of the positronium state.

The bulk of the energy in a positronium atom is in the rest masses of the electron and positron. The kinetic energies and three momenta of these particles are much smaller — typically

$$K_e \approx \alpha^2 m \quad \text{and} \quad p_e \approx \alpha m \approx 1/(\text{atom size}), \tag{32}$$

respectively. The electron and positron constantly exchange photons, which provide the binding. The exchanged photons have momenta q_γ that are typically the same size as the electron's three momentum:

$$q_\gamma \approx p_e. \tag{33}$$

The photon's energy, however, is the same size as its three momentum, since it is massless, and therefore its energy is much larger than the electron's kinetic energy ($1/\alpha$ times larger):

$$E_\gamma = q_\gamma \gg K_e. \tag{34}$$

This means that the exchanged photon is highly virtual, and therefore the exchange is effectively instantaneous on the time scale important to the electron and positron. Consequently photon exchange can be modelled with an instantaneous potential: the Coulomb potential to leading order in v_e/c, and the Breit interaction in the next order.

A potential model exists for positronium because of the disparity between the electron-positron kinetic energy and the photon's energy. We can formally derive the potential model as an effective theory by introducing an ultraviolet cutoff that restricts only particle energies. If we choose Λ_E between K_e and E_γ, then the exchanged photons are excluded from the effective theory. Their effect is reintroduced through correction terms, but, since the cutoff is on energy and not three momentum, the correction terms are local in t but nonlocal in r. That is, the correction terms are instantaneous potentials $V(\mathbf{p}, \mathbf{r})$. We have derived the potential model using renormalization theory.

The analysis of positronium is not complete because it is possible to have very low momentum photons whose energy is below the cutoff:

$$q < \Lambda_E \ll p_e. \tag{35}$$

These photons, however, have wavelengths that are much larger than the size of the atom, and so their interactions can be expanded, in the effective theory, using a multipole expansion. This means that they couple only weakly to the atom; the probability of finding a (soft) photon with the $e\overline{e}$ in a positronium atom is

$$P(e\overline{e}\gamma) \approx \alpha\, v_e^2 \approx \alpha^3. \tag{36}$$

Positronium can be described accurately by a potential model only insofar as contributions from such states are negligible. The $e\overline{e}\gamma$ states in the effective theory have energy of order $\alpha^2 m$, and so the shift positronium binding energies by $P(e\overline{e}\gamma)\,\alpha^2 m \approx \alpha^5 m$, which is relatively very small. The Lamb shift is the most famous example of such a contribution.

There are two ways of systematically incorporating soft photons. The standard approach is to use the cutoff effective theory to compute electron-positron potentials that include the contributions due to soft photons. Since the dynamical time scale for these photons is the same as for the $e\overline{e}$, the potentials are energy dependent and have divergent expansions in perturbation theory. Techniques exist for dealing with the divergent expansions for Coulombic systems like positronium. The extent of the energy dependence in the potential is directly related to the probability carried by Fock states with soft photons. This is obvious from the standard normalization condition for the wavefunction when the potential is energy dependent:

$$1 = \int d^3r\, \psi^\dagger(\mathbf{r})\,(1 - \partial V/\partial E)\,\psi(\mathbf{r}). \tag{37}$$

The second term on the right-hand side gives the probability carried by all Fock states other than the $e\overline{e}$ state; that is, it equals the sum of all terms on the right-hand side of Eq. (31) after the first.

A second approach worth investigating is to treat the effective theory as a coupled channel problem, where now the wave function has two or more components,

$$\Psi(\mathbf{r}) = \begin{pmatrix} \psi_{e\overline{e}}(\mathbf{r}) \\ \psi_{e\overline{e}\gamma}(\mathbf{r}, q) \\ \vdots \end{pmatrix}, \tag{38}$$

one for each Fock state that is important to the analysis. For most high-precision QED work only the $e\overline{e}$ and $e\overline{e}\gamma$ states are relevant; additional soft photons are too highly suppressed to contribute. The components of this wavefunction satisfy coupled Schrödinger equations whose potentials are instantaneous (energy independent) and calculable from *convergent* perturbation series in QED. The formalism is dramatically simplified by the multipole expansion, which is valid because of the cutoff. This expansion means that the $e\overline{e}\gamma$ wavefunction, for example, is a function of the $e\overline{e}$ separation, and of the magnitude q of the photon momentum; the wavefunction is independent of the direction of $\mathbf{q}$. This last fact makes a numerical treatment feasible.

Effective field theories already play a big role in high-precision studies of positronium and other simple QED atoms.[5] NRQED, for example, is an effective field theory in which the electron and positron dynamics is nonrelativistic. In the most common approach, the boundstate equation in NRQED is a Schrödinger equation with energy-dependent potentials describing soft photon contributions, as above. This equation is usually solved perturbatively. A different approach would be to compute effective potentials for the Schrödinger equation using QED or NRQED perturbation theory and the renormalization techniques outlined in these lectures. Then the boundstate equations could be solved numerically, as described in the earlier parts of these lectures. Once the potentials are determined, the properties of any state can be extracted from a simple numerical calculation. It is perhaps remarkable that there are effective potentials which include all the effects of full QED. If these potentials were computed, they could be used for multielectron atoms or in solidstate analyses — a relatively painless way of introducing QED effects.

3.2 Heavy Quark Mesons

Much of the discussion above for positronium can also be applied to mesons, like the Υ, composed of heavy quarks. The b quark and antiquark in an Υ are very heavy, and therefore they are moderately nonrelativistic: $v_b^2 \approx 0.1$. Thus the dominant gluons binding the quarks have energies that are much larger than the quarks' kinetic energy, and so can be rigorously modelled using an instantaneous potential. This is the basis for the phenomenological quark

model, which has been very successful in describing the Υ and its excitations. But, as in the case of positronium, the quark model is not the whole story. States with soft gluons in addition to the $b\overline{b}$ also contribute. Again the gluon coupling to the quarks can be expanded in a multipole expansion, and therefore the soft gluons are suppressed: the probability of the $b\overline{b}g$ state is $\alpha_s v^2 \approx 0.1$. Nevertheless there are important phenomena in which these additional Fock states play a leading role.[6] It would be worthwhile to explore a coupled-channel approach to formulating the dynamics of this Fock state, similar to that outlined above for positronium.

3.3 Potential Models in General

Potential models arise as effective theories, low-energy approximations to more fundamental quantum field theories, when the important energies in a system are much smaller than the three momenta. Then, when modelling that system, it is natural to impose a stronger cutoff on energies than on three momenta in the effective theory. This results in effective interactions that are potentials — that is, they are local in time but not in space. A system of interacting non-relativistic particles always has smaller kinetic energies than three momenta, smaller by v/c, and so is a prime candidate for a potential model.

The utility of a potential model depends upon the importance of retardation corrections. Consider, for example, the case where the interaction between the nonrelativistic particles is mediated by a massless particle, like the photon in positronium. Here exchanges in which the massless particle is very soft are allowed by the energy cutoff, and lead to retardation effects that can greatly complicate the potential. Nevertheless, as in positronium, these interactions can often be simplified using multipole expansions, and are often strongly suppressed.

Retardation effects are much smaller if the exchanged particle is massive. Indeed, there are no retardation effects at all if the mass of the exchanged particle is larger than the energy cutoff in the effective theory; the exchanged particle is always highly virtual, no matter what its three momentum, and so can always be described by an instantaneous potential. In such cases simple potential models provide a framework for calculations of arbitrary precision. We now turn to such a system.

4 Nucleon-Nucleon Interactions

Another example of an effective theory is the pion-nucleon theory of low-energy nuclear interactions. In this section I discuss the design of an effective theory

that describes low-energy nuclear interactions. I tune the couplings to reproduce phase shifts for np scattering. I also review a problem encountered in a recent analysis of this system that uses dimensional regularization. Finally I discuss briefly the general implications of effective theory for low-energy nuclear physics.

The analysis I present here is pedagogical. Others have completed analyses that are more thorough, although slightly different in detail from what follows here.[7]

4.1 What is a Hadron?

The true theory of the nuclear force is QCD: nucleons are made of quarks and they interact by exchanging quarks and gluons. Nevertheless the electric form factor of the proton, for example, is given roughly by

$$F_p(Q^2) \approx \left(\frac{1}{1 + (Q/800\,\mathrm{MeV})^2} \right)^2 \tag{39}$$

which goes to 1, the value for a pointlike particle, for momentum transfers $Q \ll 800\,\mathrm{MeV}$. This and other experimental results suggest that nucleons appear approximately point-like to probes with momenta less than 500–1000 MeV. Therefore an effective theory that describes nucleons and other hadrons in this low-momentum region should treat hadrons as elementary particles. Quark substructure cannot be resolved at such momenta. Since nucleon momenta are of order 300 MeV or less in nuclear matter, an effective theory of this sort might provide an excellent framework for the systematic study of traditional low-energy nuclear physics.

Our effective theory, therefore, should have a cutoff of order the color resolution scale, $\Lambda_c \approx 500$–$1000\,\mathrm{MeV}$. (I switch here to designating the cutoff in momentum space, Λ, rather than in coordinate space, $a = 1/\Lambda$, because the former is more conventional.) Heavy hadrons, like the proton, neutron, Δ, ρ, and so on, have masses of order the cutoff or larger and so cannot be relativistic. These are most efficiently treated as pointlike nonrelativistic particles, with kinetic energies given by

$$K = \frac{\mathbf{p}^2}{2m} - \frac{\mathbf{p}^4}{8m^3} + \cdots . \tag{40}$$

Because of the cutoff, the effective theory contains no pair creation for these hadrons; particle and antiparticle decouple.

The light hadrons — that is, the pion and possibly also the kaon — can be relativistic and so must be included as pointlike particles described by a full

relativistic quantum field theory. This field theory inherits a chiral symmetry from QCD that implies that all interactions are suppressed by powers of p/Λ_c or m_π/Λ_c. Consequently, even though pions are strongly interacting particles, their interactions at low momenta are weak, and can be analyzed using perturbation theory. The Feynman rules for this theory are given in many standard texts.[8]

The usual chiral theory is a quantum field theory. Potential models, however, provide a natural framework for describing nonrelativistic systems, as we discussed in the previous section. This is particularly true for the nucleon-nucleon problem, since the pion and kaon masses are larger than the typical nucleon energy in nuclear matter. Thus retardation due to soft pion and kaon exchange plays only a small role in our effective nucleon-nucleon theory, and can be hidden inside energy-independent potentials. Our effective theory for low-energy nucleon-nucleon physics, therefore, is specified by a hamiltonian:

$$H = \sum_{\text{nucleons}} \left(\frac{\mathbf{p}^2}{2m} - \frac{\mathbf{p}^4}{8m^3} \right) + V(\mathbf{r}, \mathbf{p}, \boldsymbol{\sigma}_i, \boldsymbol{\tau}_i, \mathbf{L}) \tag{41}$$

where the potential is a function of the positions, momenta, spins, isospins and angular momentum of the two nucleons.

4.2 The Nucleon-Nucleon Potential

The potential in our nucleon-nucleon hamiltonian can be determined by "matching" scattering amplitudes computed in the potential model with the same amplitudes computed using the chiral field theory, as in Section 2.7. There are both long-range parts, involving the exchange of one or more pions, and short-range or "contact" potentials like the delta function potentials in Section 2.

The leading-order long-range potential is due to the exchange of a single pion; it gives

$$V_{1\pi} = \alpha_\pi \, \boldsymbol{\tau}_1 \cdot \boldsymbol{\tau}_2 \, \frac{\boldsymbol{\sigma}_1 \cdot \boldsymbol{\nabla} \, \boldsymbol{\sigma}_2 \cdot \boldsymbol{\nabla}}{m_\pi^2} \, v_\Lambda(r) \tag{42}$$

where $v_\Lambda(r)$ is a Yukawa potential at large r,

$$v_\Lambda(r) \stackrel{r \text{ large}}{\longrightarrow} \frac{\mathrm{e}^{-m_\pi r}}{r}, \tag{43}$$

and the $\boldsymbol{\sigma}$'s and $\boldsymbol{\tau}$'s are Pauli matrix operators for the nucleon spin and isospin respectively. Chiral symmetry implies that the dimensionless coupling α_π is related to the pion decay constant, $f_\pi \approx 93\,\text{MeV}$, the nucleon's axial vector

coupling, $g_A \approx 1.25$, and the pion mass m_π:

$$\alpha_\pi \approx \frac{g_A^2\, m_\pi^2}{16\pi\, f_\pi^2} \approx 0.06\text{–}0.08. \tag{44}$$

The coupling is small because it is proportional to m_π^2/Λ_c^2, Λ_c being of order $4\sqrt{\pi} f_\pi$. Projecting the one-pion potential onto different spin-orbit eigenstates, we obtain

$$V_{1\pi} \to -\alpha_\pi \left[b\, v_\Lambda(r) + b_T\, v_T(r) \right] \tag{45}$$

where b and b_T are constants that depend upon the channel, and

$$v_T(r) \equiv \frac{1}{m_\pi^2}\left(v_\Lambda'' - \frac{v_\Lambda'}{r} \right) \tag{46}$$

$$\stackrel{r\,\text{large}}{\longrightarrow} \frac{\mathrm{e}^{-m_\pi r}}{r}\left[1 + \frac{3}{m_\pi r} + \frac{3}{m_\pi^2 r^2} \right]. \tag{47}$$

Here I have dropped a part of the potential that is indistinguishable from the contact terms we add below. We introduce an ultraviolet cutoff in the Fourier transform, as in Section 2.3, to obtain

$$v_\Lambda(r) \equiv \frac{1}{2r}\left[\mathrm{e}^{-m_\pi r}\mathrm{erfc}\Big(-(r\Lambda - m_\pi/\Lambda)/\sqrt{2} \Big) - (r \to -r) \right], \tag{48}$$

where $\mathrm{erfc}(x) = 1 - \mathrm{erf}(x)$.

Higher-order (in p/Λ_c) contributions to the long-range potential come from two-pion exchange. I will not include these here; they are described in the literature.[7]

The short-distance terms in the potential are smeared delta functions, with $a = 1/\Lambda$, multiplied by powers of the momentum operator $\mathbf{p} = \boldsymbol{\nabla}/\mathrm{i}$, and the spin and isopin operators. It is convenient for our analysis to classify these by the spin-orbit channels that they affect.

The ${}^1S_0$ and ${}^3S_1$ channels have Λ^{-2} terms with no $\boldsymbol{\nabla}$'s; there is one for each channel, each with its own coupling constant. In order Λ^{-4} there are two contact potentials, one for each channel, proportional to $\mathbf{p}^2\delta^3_{1/\Lambda}(\mathbf{r}) + \delta^3_{1/\Lambda}(\mathbf{r})\mathbf{p}^2$. The form of this potential can be simplified by adding $-2\,\mathbf{p}\cdot\delta^3_{1/\Lambda}(\mathbf{r})\mathbf{p}$, which doesn't couple to S-states in this order; the resulting potential is then proportional to $(\nabla^2\delta^3_{1/\Lambda})$, just as in Section 2.3.

The P states do not couple to contact terms in $\mathcal{O}(\Lambda^{-2})$. Thus our effective theory predicts that P states are more accurately described by just the long-range potential than S-states. The leading P-wave contact terms are $\mathcal{O}(\Lambda^{-4})$, and, after projecting out the orbital and spin angular momenta, are

all proportional to $\mathbf{p} \cdot \delta^3_{1/\Lambda}(\mathbf{r})\mathbf{p}$. This operator can be simplified by adding $-1/2(\mathbf{p}^2\delta^3_{1/\Lambda}(\mathbf{r}) + \delta^3_{1/\Lambda}(\mathbf{r})\mathbf{p}^2)$, which doesn't couple to P states in this order, to obtain again a potential proportional to $(\nabla^2\delta^3_{1/\Lambda})$. There is one such potential for each of the four spin-orbit P-wave channels, and each has its own coupling constant.

There are no contact terms in $\mathcal{O}(\Lambda^{-2}, \Lambda^{-4})$ that couple just to D states; thus D states should be even more accurately described by the long-range potential than P states. The only contact term in order Λ^{-4} that affects a D states is one that couples the ${}^3D_1$ and ${}^3S_1$ states.

There are nine contact terms in all through order Λ^{-4}: two Λ^{-2} terms and two Λ^{-4} terms for the S-states, four Λ^{-4} terms for the P states, and one Λ^{-4} term coupling S and D states. Each of these has its own coupling constant. This seems like a lot of coupling constants, but it is only one or two per channel.

Note that our effective theory does not include contributions from the exchange of heavy mesons like the ρ or ω. This is neither an oversight nor an error. It makes no sense, physically or mathematically, to include such exchanges in our effective theory. For example, a ρ exchanged by nonrelativistic nucleons is off energy-shell by an amount of order its mass, and therefore, by the uncertainty principle, can exist only for a short time — far too short a time for its constituent quark and antiquark to realize that they are in a ρ. Even if it made physical sense to think of ρ exchange here, its contribution would be indistinguishable from a nucleon-nucleon contact term; nonrelativistic nucleons cannot tell the difference. We must include the contact terms in any case; thus it is most efficient to use them to account for any heavy-meson exchange as well.

Our effective potential also is instantaneous; it has no energy-dependent terms. We design our potentials by matching on-shell scattering amplitudes from the potential theory with those from the chiral field theory. Since the amplitudes are on-shell, all energies can be expressed in terms of three momenta and so the potentials are naturally and rigorously energy-independent.[f] I find that this simplifies the analysis and design of effective theories.

[f] Thus, for example, the energy-dependent term in Eq. (17) of reference [7] vanishes on-shell and so would not be included in our tree-level potential. The effects of pion retardation are small, and in our formalism would be automatically included with the one-loop contributions to the potential.

4.3 Relativistic Kinetic Energies

Our discussion has focused on corrections to the potential in effective theories. In general there are corrections to the kinetic energy as well. Thus, in our nucleon-nucleon theory, there is a $-\mathbf{p}^4/4m^3$ correction due to relativity. This term is too singular to include in ordinary Schrödinger equations, but can be handled in our effective theories because of the cutoff. It is important, however, that the cutoff be low enough to prevent the $\mathbf{p}^4$ term, which is negative, from overwhelming the $\mathbf{p}^2$ term and destabilizing the effective theory.

While it is quite possible to include the $\mathbf{p}^4$ correction directly, I have found it numerically simpler to transform it into a potential, albeit an energy-dependent one. Starting with the Schrödinger equation, which has the form

$$\left[E - V + \frac{\nabla^2}{2m_r} + \frac{\nabla^4}{32m_r^3}\right]\psi = 0 \tag{49}$$

where m_r is the reduced mass, we introduce a transformed wave function ϕ where $\psi \equiv \Omega\phi$ and

$$\Omega = 1 + \frac{2m_r(E-V) - \nabla^2}{32m_r^2}. \tag{50}$$

Multiplying the Schrödinger equation on the left by Ω, we obtain

$$\begin{aligned} 0 &= \Omega\left[E - V + \frac{\nabla^2}{2m_r} + \frac{\nabla^4}{32m_r^3}\right]\Omega\phi \\ &= \left[E - V + \frac{(E-V)^2}{8m_r} + \frac{\nabla^2}{2m_r} + \mathcal{O}\left(\frac{p^6}{m_r^5}\right)\right]\phi. \end{aligned} \tag{51}$$

Solving the last equation, without the p^6 terms, is equivalent to solving the original Schrödinger equation up to corrections of order p^6, which are higher-order and therefore negligible here. We can recover the original wavefunction ψ by undoing the transformation:

$$\begin{aligned} \psi &= \Omega\phi \\ &= \left[1 + \frac{2m_r(E-V) - \nabla^2}{32m_r^2}\right]\phi \qquad (52) \\ &= \left[1 + \frac{E-V}{8m_r} + \mathcal{O}\left(\frac{p^4}{m_r^4}\right)\right]\phi. \qquad (53) \end{aligned}$$

I used the Schrödinger equation to obtain the last equation; the p^4 term is higher-order and can be neglected here. Note that the correct normalization

for ϕ, given that the ψ's are normalized to one, is

$$\int d^3r\, \phi^\dagger(\mathbf{r}) \left[1 + \frac{E - V}{4m_r}\right] \phi(\mathbf{r}) = 1. \tag{54}$$

The procedure used in the previous paragraph to transform $\mathbf{p}^4$ into a potential is often described as "using the equations of motion." In effect we replaced $\mathbf{p}^4$ by $4m_r^2(V-E)^2$. We could have guessed the result by noting that the Schrödinger equation, the "equation of motion" here, implies that $\mathbf{p}^2 \approx 2m_r(V-E)$ when acting on the wavefunction of an energy eigenstate. Thus it is plausible that $\mathbf{p}^4$ is the square of the same operator. Transformations of the wavefunction are a powerful tool for simplifying effective field theories. They also underscore the fact that there are infinitely many theories that have equivalent low-energy behavior. Useful transformations Ω typically are close to unity at low energies, so that complicated cross terms are high-order and can be neglected, as above.

I have included this discussion to show how one deals with relativistic corrections to the kinetic energy. However, I have checked and found that $\mathbf{p}^4$ corrections have little effect on the quantities I discuss below. Consequently I chose to omit them from that analysis.

4.4 Comparison with Data

Having outlined the design of a rigorous potential model for nonrelativistic nucleon-nucleon interactions, we now compare this model with experimental data. For simplicity, I restricted my study to *np* scattering, thereby avoiding the problem of Coulomb distortion in *pp* scattering. I chose to fit phase shifts for different channels, rather than the full scattering amplitude, because this greatly simplifies the analysis: there are only one or two couplings that need be tuned for each channel, and tuning can be done one channel at a time. The disadvantage is that the phase shifts are not really data; they must be extracted from the data using some sort of theoretically motivated procedure. This makes it difficult to assign experimental errors to the phase shifts. My goal, however, is more to illustrate the procedure than to obtain definitive theoretical results. The effective theory would fit exact data just as well as it fits the phase shift solutions I use here. I use phase-shifts from the multi-energy fits by the Nijmegen group.[9] Multi-energy fits are preferable here since they enforce a smooth energy dependence.

We begin with the ${}^1S_0$ phase shifts. The potential, including one-pion

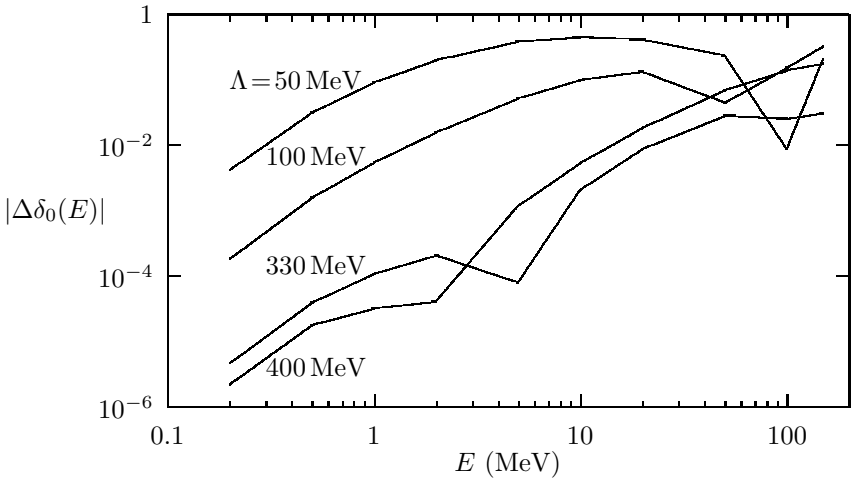

Figure 6: Errors in the ${}^1S_0$ phase shifts (in radians) versus energy for the full effective theory with different values of the cutoff Λ.

exchange and contact terms through order Λ^{-4}, is

$$V(r) = -\alpha_\pi v_\Lambda(r) + c\,\frac{\delta^3_{1/\Lambda}(\mathbf{r})}{\Lambda^2} - d\,\frac{\nabla^2\delta^3_{1/\Lambda}(\mathbf{r})}{\Lambda^4}, \tag{55}$$

where I use $\alpha_\pi = 0.075$ and $m_\pi = 140\,\text{MeV}$. Coupling constants c and d must be tuned to the data. I used the Nijmegen phase shifts at $E = 0.05\,\text{MeV}$ and $0.1\,\text{MeV}$ to tune c and d for Λ's ranging from $50\,\text{MeV}$ to $1000\,\text{MeV}$. Here and elsewhere in this section E refers to the (nonrelativistic) center-of-mass energy.

In Figure 6 I show the errors obtained in the phase shift as a function of center-of-mass energy E for a variety of different values of the cutoff Λ. As in our synthetic example, the errors start out large when Λ is small, and decrease steadily until $\Lambda \approx 300\,\text{MeV}$. No further improvement is seen for larger Λ's.

In Figure 7 I compare the errors obtained from the Λ^{-2} theory ($d = 0$) with those from the Λ^{-4} theory. At low energies the errors in the Λ^{-2} theory grow approximately like $E^{0.5}$ with energy. The errors in the Λ^{-4} theory are much smaller, but grow like $E^{1.5}$ — that is faster by two powers of p, exactly as expected. So here, as in our synthetic data, we see clear evidence that the theory is improved by adding higher-order corrections, and that it is improved in just the manner predicted. This convinces me that the effective theory is working.

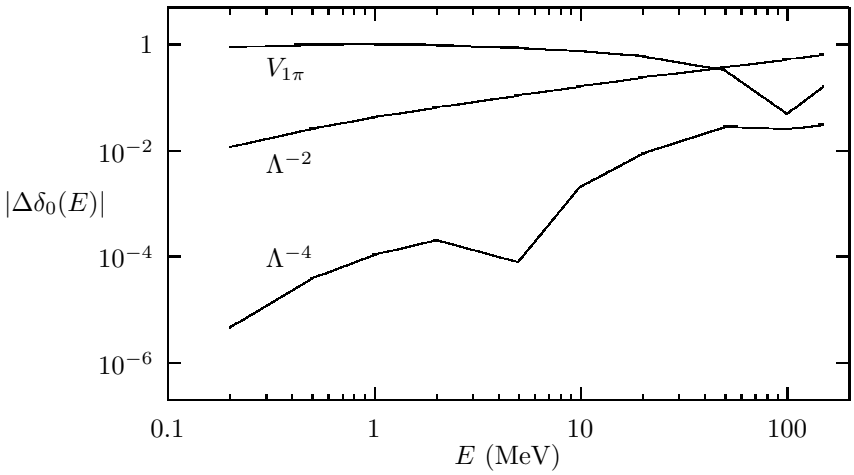

Figure 7: Errors in the ${}^1S_0$ phase shifts (in radians) versus energy for the effective theory through orders Λ^{-2} and Λ^{-4}. Results from the theory with just pion exchange ($V_{1\pi}$) are also shown. The cutoff was $\Lambda = 330\,\mathrm{MeV}$ in each case.

The curves in the logarithmic plots of $\Delta\delta_0(E)$ versus E dip at various different points. These dips result when the error changes sign. They should be ignored.

I also show errors in the last figure for the effective theory with one-pion exchange but no contact terms ($V_{1\pi}$). Evidently one-pion exchange contributes little to the phase shift at most energies. This is surprising since the leading contact term and the one-pion potential are both $\mathcal{O}(1/\Lambda^2)$. The difference is that one-pion exchange is order m_π^2/Λ^2, while the contact term is order q^2/Λ^2. Thus the data suggest that momenta $q \gg m_\pi$ play an important role during the scattering process, even for low incident energies. This is possible if the interaction is fairly strong; the phase shifts below 10 MeV depend quite nonlinearly on the coupling constants which indicates that this is case.

In Figure 8 I show the actual phase shifts, together with the data. The results from the full theory are almost independent of the cutoff. The variation at high energies gives an indication of the size of the Λ^{-6} corrections yet to be included.

It is slightly mysterious that the full effective theory stops improving when $\Lambda \approx 300$–$400\,\mathrm{MeV}$. We had expected hadron structure and other QCD effects to appear at 500–1000 MeV, not 300 MeV. One possible explanation is that we have omitted two-pion exchange from our potential. Two-pion contribu-

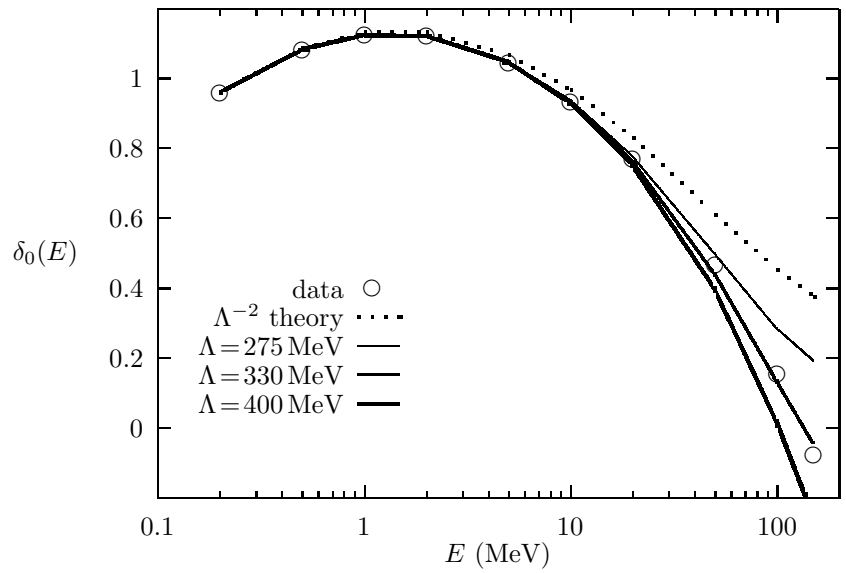

Figure 8: The $^1S_0$ phase shifts (in radians) versus energy for the full effective theory with various values of the cutoff Λ. Results for the Λ^{-2} version of the theory ($d = 0$) are also shown ($\Lambda = 200\,\mathrm{MeV}$).

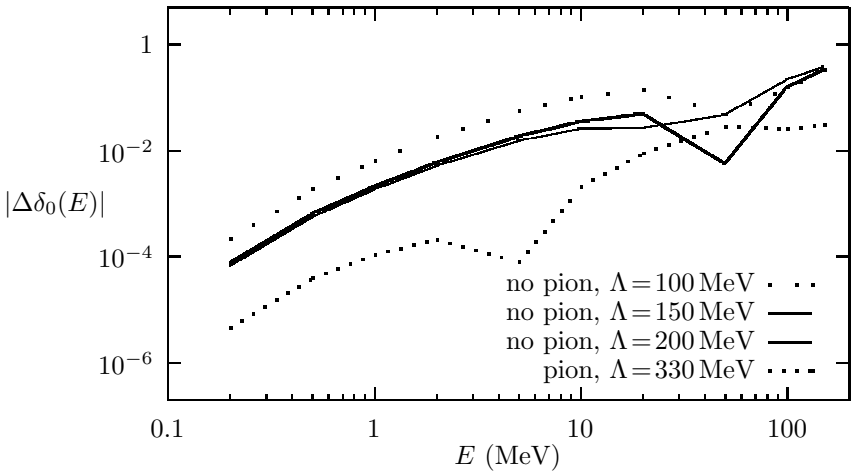

Figure 9: Errors in ${}^1S_0$ phase shifts (in radians) versus energy for the effective theory with no pion exchange, with various values of the cutoff Λ. Results for the full theory, including pion exchange, are also shown.

tions are nominally the same order as the Λ^{-4} contact term; to be consistent, we should include them in our analysis. We don't expect a particularly large contribution from two-pion exchange since one-pion exchange is so small. Nevertheless the errors are small at $\Lambda = 300\,\mathrm{MeV}$, and so we do not need much from two pions.

One way to test whether our analysis should be sensitive to a small two-pion contribution is see what happens if we drop one-pion exchange. I retuned the effective theory to the data, keeping only the contact terms. The results are summarized in Figure 9. I compare the phase shift errors from the theory without pions, for different Λ's, to the best fit that includes pion exchange. The theory without pions stops improving at $\Lambda \approx 150\,\mathrm{MeV}$, rather than at 300 MeV, and so is substantially less accurate at *all* energies below 100 MeV. This indicates that there is new physics in the data, physics that cannot be modelled by the contact potentials, at a momentum scale of order 150 MeV. The new physics, of course, is one-pion exchange. The contact potentials are unable to model one-pion exchange accurately once momenta of order m_π are admitted into the effective theory, despite the fact that the one-pion contribution is relatively quite small. We can move the cutoff to 300–400 MeV only by including one-pion exchange in the potential; perhaps we can move the cutoff further by including two-pion exchange.

There are other possible explanations for the 300 MeV scale in our analysis. For example, the inverse radius of a proton is about 240 MeV. If quarks can be resolved at this scale, then it is likely that we will be unable to raise the cutoff without switching to QCD. The radius of the proton is not necessarily the same as the color resolution scale since some of the radius could be due to a pion cloud. Other issues that I haven't examined with sufficient thoroughness include: tuning α_π; the differences in mass and α_π between charged and neutral pions; contributions from Δ's; and, of course, possible differences between the phase shifts and real data.

In Figure 10 I show results for the ${}^1P_1$ and ${}^3D_2$ phase shifts. The effective potentials for these two cases are, through order Λ^{-4},

$$V({}^1P_1) = 3\,\alpha_\pi\, v_\Lambda(r) + c({}^1P_1)\,\frac{\nabla^2 \delta^3_{1/\Lambda}(\mathbf{r})}{\Lambda^4}, \tag{56}$$

and

$$V({}^3D_2) = -\alpha_\pi\,[\,v_\Lambda(r) + 2\,v_T(r)]\,. \tag{57}$$

The data confirm that the long-range potential plays a much more important role for P-waves, which have no Λ^{-2} contact term, than for S-waves. And the D-wave analysis has no contact terms at all until order Λ^{-6}. In both these cases the phase shifts are fairly linear in the couplings, unlike the low-energy S-wave phase shifts.

The relative errors are much smaller for our S-wave phase shifts than for the others. The absolute magnitude of the errors in each channel, however, is roughly the same when the effective theories for each channel are corrected to the same order in $1/\Lambda$. Also the energy dependence of these errors is the same. This is illustrated in Figure 11. It is exactly what is expected based upon the effective theory.

4.5 Multinucleon Systems

Our nucleon-nucleon analysis is readily extended to systems of three or more nucleons. The potentials tuned using nucleon-nucleon scattering can be carried over without change. In addition new contact terms involving three or more nucleons can be added, but these tend to be very high order in Λ^{-1}, and so can often be neglected. For example, the leading three-body contact potential is order Λ^{-5}. Of course there are in addition various long-range three-body potentials predicted by the chiral theory. This subject is discussed in detail in the literature.[10]

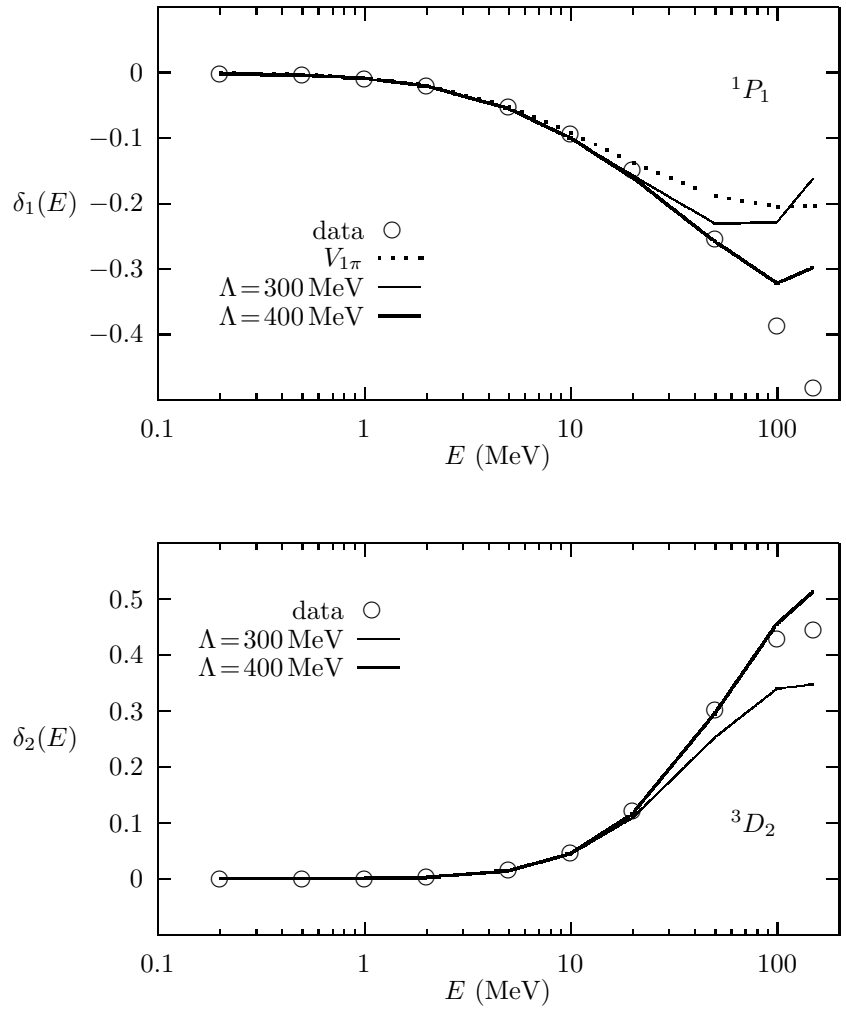

Figure 10: Phase shifts (in radians) for different channels from the effective theory. P-wave results are for the uncorrected ($V_{1\pi}$) theory, and for the full effective theory with different values of Λ. D-wave results are for the uncorrected theory with different values of Λ; there are no Λ^{-2} or Λ^{-4} corrections for this D-wave channel.

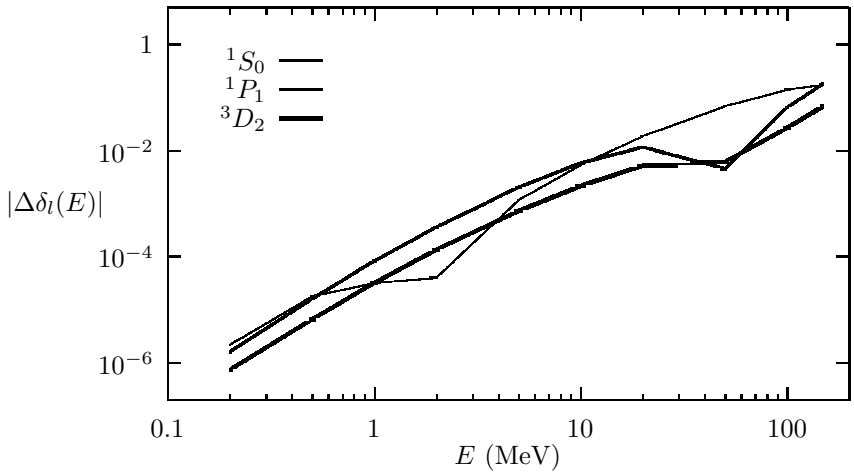

Figure 11: Errors in the phase shifts (in radians) for different channels from the effective theory corrected through order Λ^{-4}.

4.6 The Problem with Dimensional Regularization

I have emphasized in this paper that the exact details of the cutoff are unimportant; just about any choice should give results equivalent to those presented here. Dimensional regularization, however, is an exception.[g] Dimensional regularization is a popular scheme developed for perturbative analyses of gauge theories. A recent paper has uncovered a pathology when dimensional regularization is applied to the nucleon-nucleon problem.[11] The authors find that the dimensionally regulated version of the $^1S_0$ theory we describe above fails completely by energies of 10–20 MeV, well below the energy at which the cutoff theory fails. They conclude that the rules for applying effective field theories must be modified. In this I believe they are incorrect, especially given the obvious success of the cutoff version of the theory. They have uncovered a problem, but the problem is with dimensional regularization — or, more precisely, with minimally-subtracted dimensional regularization — and not with effective field theory.

The authors trace the problem in the dimensionally regulated theory to the unusually long scattering length a_s in the $^1S_0$ channel. As we have seen, an effective field theory generally has residual errors that are some power of the

[g] Much of the discussion in this section is based on a series of useful conversations I had with Martin Savage and David Kaplan at the INT in Seattle.

typical momentum q divided by the ultraviolet cutoff Λ. When the ${}^1S_0$ theory is dimensionally regulated, the effective Λ in the q/Λ errors is tied to the scattering length, and vanishes as $a_s \to \infty$. Therefore this version of the theory is accurate only for very small momenta when a_s is large.

Large scattering lengths arise when there is a bound state just below threshold (a_s positive) or a resonance just above threshold (a_s negative). We can see how Λ is related to the scattering length by analyzing a simplified problem: Consider a system with a barely bound S state caused by a short-range interaction that we model with a single contact term:

$$V(r) = \frac{4\pi}{\Lambda^2}\,\delta^3(\mathbf{r}) \tag{58}$$

We analyze this problem three ways. First we use first-order perturbation theory. In this approximation, the scattering amplitude is simply

$$T = \frac{4\pi}{\Lambda^2}, \tag{59}$$

and therefore the scattering length is given by

$$\frac{1}{a_s} \equiv \lim_{p\to 0} \frac{4\pi}{m}\,\mathrm{Re}\frac{1}{T} = \frac{\Lambda^2}{m}. \tag{60}$$

Thus $\Lambda = \sqrt{m/a_s}$, and large scattering lengths imply small Λ's. Generally bound states do not appear in first-order perturbation theory, since a bound state is characterized by particles interacting again and again, forever (not just once). Consequently, the presence of a large positive scattering length in this analysis does not indicate a bound state near threshold. This is confirmed by the fact that the potential here must be repulsive, not attractive, if a_s is to be positive.

Next we analyze the simplified problem nonperturbatively, using minimally-subtracted dimensional regularization. The contact potential, being separable, is easily iterated to all orders, and gives the amplitude

$$T = \frac{T_1}{1 - T_2/T_1} \tag{61}$$

where $T_1 = 4\pi/\Lambda^2$ is the amplitude in first order, and

$$T_2 = -(T_1)^2\,\frac{imp}{4\pi} \tag{62}$$

is the dimensionally-regulated second-order amplitude. Minimal subtraction implies that the second-order amplitude is purely imaginary. Consequently the

only effect of iterating the potential to all orders is to unitarize the amplitude. The scattering length obtained from Eq. (61) is identical to that obtained from first-order perturbation theory, and again $\Lambda = \sqrt{m/a_s}$ vanishes as a_s becomes large. When the analysis is extended to included higher-order contact terms, the small Λ restricts the model's utility to very low energies.[11]

Finally we redo the nonperturbative calculation using a cutoff on the loop-integral in T_2, which we implement in such a way that the interaction is still separable. While the potential in the previous two analyses was necessarily repulsive, here we flip the sign of V, making it attractive, in order to get a positive scattering length. Then

$$T_1 = -\frac{4\pi}{\Lambda^2}, \tag{63}$$

and

$$T_2 = -(T_1)^2 \left[\frac{\mathrm{i}mp}{4\pi} + \frac{m\Lambda}{2\pi^2} + \mathcal{O}(p^2 m/\Lambda) \right]. \tag{64}$$

Substituting into the formula for the nonperturbative amplitude, Eq. (61), we find that

$$\frac{1}{a_s} = \frac{2\Lambda}{\pi} - \frac{\Lambda^2}{m}. \tag{65}$$

Thus large a_s can coexist with large Λ if there is a delicate cancellation between the two terms on the right-hand side. In this case Λ goes to $2m/\pi$, which is large, as a_s goes to infinity.

Minimally-subtracted dimensional regularization is well known to drop power-law divergent terms that are physically significant, and that is what happens here. The dimensionally regulated calculation drops the linear divergence in T_2, and as a result captures only one of the two terms in Eq. (65). In the cutoff theory, the small binding energy results from a nearly exact cancellation between the kinetic and potential energies, each of which is separately quite large — of order $\Lambda/a_s m$. The dimensionally regulated calculation, however, cannot model this cancellation because, in effect, it drops the power-divergent kinetic-energy term.

The cutoff theory behaves similarly to the dimensionally regulated theory when cutoff Λ is small compared with $1/a_s$. To see this, we multiply the potential by a coupling constant c, thereby decoupling the range and strength of the potential. Then m is replaced by cm in Eq. (65) for the scattering length, and we can solve for the value of c corresponding to a particular a_s:

$$c = \frac{\Lambda^2}{m} \left(\frac{2\Lambda}{\pi} - \frac{1}{a_s} \right)^{-1} \tag{66}$$

for any Λ. When $\Lambda \ll 1/a_s$ the coupling constant becomes negative, that is the potential becomes repulsive, as in the dimensionally regulated analysis. The linear divergence in the loops is negligible for such small Λ's, and therefore first-order perturbation theory gives the correct scattering length, again as in the dimensionally regulated analysis. And the theory fails at relatively low energies since $\Lambda \ll 1/a_s$.

Large scattering lengths can arise in a range of situations that lie between two different physical extremes. One extreme is realized by adjusting the magnitude of the potential so that the first-order Born approximation gives the correct scattering length, while at the same time extending the potential's range so that corrections from second order and beyond are negligible (other than the imaginary parts always required by unitarity). The other extreme is realized by tuning a short-range potential to create a bound state or resonance near threshold; the presence of a bound state implies that contributions from all orders in the the potential are important. Minimal subtraction with dimensional regularization is well suited only to the first limit. Unfortunately the physics of low-energy nucleon-nucleon scattering corresponds to the second limit.

The problem encountered in the dimensionally regulated calculation can be avoided in perturbative analyses since high-order terms that cause trouble can be dropped by hand. In nonperturbative calculations, like that needed for S-wave nucleon-nucleon scattering, the problem can be easily repaired by modifying minimal subtraction; other, more baroque variations are also possible.[12] But the problem can be avoided altogether by not using dimensional regularization. Unlike in perturbative QCD, for example, there is no compelling reason for using dimensional regularization in the nucleon-nucleon problem. Simple cutoffs of the sort used in this paper are effective, robust, easy to use, and more physical.

4.7 Evaluation

Chiral perturbation theory, and, in particular, effective theories of the sort I have described provide a systematic framework for studying low-energy nuclear physics. This framework casts traditional low-energy nuclear theory as the rigorous low-energy limit of QCD. Where it works, the effective theory is far more efficient to use than QCD, and is preferred. Of course, the effective theory is only useful at momenta below 500-1000 MeV (or perhaps 300 MeV). Renormalization theory indicates that quarks and QCD are essential to a systematic and rigorous study of processes at higher momenta. One can, of course, add a parameter to a nucleon-nucleon potential model in order to fix a particular

piece of high-energy data; but usually that extra parameter will not improve anything else. By contrast, when we added the Λ^{-4} correction to our effective theory we improved the fit at 0.1 MeV, where we tuned, but also at all higher energies up to 100–150 MeV. And we improved the fit in a predictable manner.

We have seen that the effective theory is quantitatively very successful at explaining low-energy nucleon-nucleon scattering. This is particularly impressive given that the tuned parameters were not derived from a global fit, but rather from the lowest-energy data, to the left of all the curves plotted. What impresses me more than the size of the errors are their systematics: the predictable dependence on the energy and on the number of contact terms in the potential, the relationships between different channels, and so on. Thus plots of the errors, $\Delta\delta_l(E)$, are more interesting than plots of the phase shifts, $\delta_l(E)$, themselves. The systematics can only be observed if the theory is tuned to the lowest-energy data, or if energy-dependent truncation errors are included in the weights assigned to data in a global fit.

Using an effective potential model rather than a Bethe-Salpeter approach in the chiral field theory is a small technical advance. It simplifies the treatment of retardation effects, resulting in simpler potentials.

The effective theory is successful, but is it successful enough? A major application of nuclear theory in our energy region is to problems in nuclear structure. These typically involve nucleon energies of order the Fermi energy, roughly 50 MeV. Our analysis gives results that are accurate to within 5–15%, depending upon the value of Λ, in this region. This suggests that the $1/\Lambda^2$ expansion converges for these energies, and that a systematic analysis of nuclear matter is possible without recourse to QCD. Precision will improve substantially if two-pion exchange doubles the threshold cutoff. And it will also improve substantially when Λ^{-6} terms are added. In the not too distant future, lattice QCD will begin to generate numbers for the various coupling constants in the effective theory, and the connection to the fundamental theory will be complete. In the meantime we can continue to extract the couplings experimentally from the abundance of accurate nucleon-nucleon data.

Acknowledgements

I thank Ron Horgan, with whom I first started thinking about chiral models and nuclear physics almost ten years ago. I also thank my former student Lisa Angelos, who provided several key insights in the early stages of this work. I thank the Institute for Nuclear Theory, in Seattle, where the first version of these lectures was created. I owe much to the summer's worth of conversations I had with Martin Savage, David Kaplan, and others at the INT.

I also benefited from several conversations with Eric Braaten. Finally I am very grateful to the organizers of the *VIII Jorge Andre Swieca Summer School* for inviting me to speak on this subject.

References

[1] This introduction is a severe condensation of an earlier summer school lecture: G.P. Lepage, *What is Renormalization?*, in *From Actions to Answers,* edited by T. DeGrand and D. Toussaint, World Scientific Press (Singapore, 1990). It contains references to the general literature on renormalization theory; see especially K.G. Wilson's Nobel lecture: K.G. Wilson, Rev. Mod. Phys. **55**, 583 (1983).

[2] This point is emphasized in E. Angelos, Cornell University Ph.D. Thesis, 1996.

[3] I thank Martin Savage explaining the multiple solutions to me.

[4] See, for example, G.P. Lepage and S.J. Brodsky, *Phys. Rev.* D **22**, 2157 (1976).

[5] For recent examples see P. Labelle, G.P. Lepage and U. Magnea, *Phys. Rev. Lett.* **72**, 2006 (1994); and M. Nio and T. Kinoshita, *Phys. Rev.* D **55**, 7267 (1997).

[6] G.T. Bodwin, E. Braaten and G.P. Lepage, *Phys. Rev.* D **51**, 1125 (1995).

[7] A particularly useful reference is C. Ordonez, L. Ray and U. van Klock, *Phys. Rev.* D **53**, 2086 (1995). They reference much of the older material as well.

[8] Ordonez et al[7] summarize the essential ingredients needed for nucleon-nucleon scattering.

[9] Partial wave analysis of the Nijmegen University group, obtained from the World Wide Web page http::/nn-online.sci.kun.nl/.

[10] See, for example, U. van Klock, *Phys. Rev.* C **49**, 2932 (1994).

[11] D.B. Kaplan, M.J. Savage and M.B. Wise, *Nucl. Phys.* B **478**, 629 (1996); M. Luke and A.V. Manohar, *Phys. Rev.* D **55**, 4129 (1997).

[12] D.B. Kaplan, University of Washington preprint (1996), e-print nucl-th/9610052.