

B.7: RENORMALIZATION IN Q.F.T.

Now we come to a central feature of any QFT, and indeed of any physical theory at all, whether it be classical or quantum, and whether it involve fields or particles. As we've already discussed, all theories in physics are "effective theories", in which one focusses on some specific set of degrees of freedom. The idea of renormalization is simply one of transforming between these theories, at the same time changing the size and the meaning of the various parameters in the theories.

We have already seen this on many occasions in our previous discussions. The Landau Fermi Liquid Theory (FLT) is a very nice non-relativistic example, so is the non-linear sigma model (NLG). Non-relativistic physics itself, as we stressed in the last chapter, is simply a low-energy effective version of what becomes QED at higher energies (with the addition of nucleons). And QED is simply the low energy sector of the standard model. In QFT, only gravity stands outside this framework - it is "non-renormalizable" (we will see what this means), and in some ways fundamentally different from all other QFT's.

In classical physics, as we've already seen, the paradigm for a renormalized theory is hydrodynamics, which "coarse-grains" over short-range fast variables referring to the underlying "granular" structure of a fluid. It only makes sense to devise renormalized theories in classical mechanics when we are dealing with many degrees of freedom; but in QFT, and therefore in ordinary PM, all systems are ultimately many-body in nature. Even in classical mechanics, one can average over the rapid motion of a simple system (eg., over high- ω noise), to get an effective "slow" Hamiltonian.

We have also looked at quite a few techniques for dealing with and implementing this idea. These range from simple perturbation theory to "Born-Oppenheimer" style averaging over fast variables, as well as gradient expansions, and the use of background field techniques. Even the use of $1/N$ expansions, which derives an effective theory, can be thought of as a kind of renormalization.

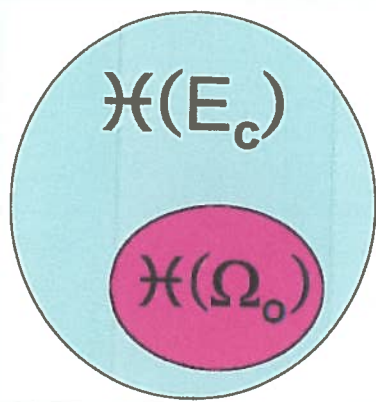
So what then does renormalization theory give us, in addition to what we already have? We will see that it gives us both new techniques and a new language with which to think about the way we move between different effective theories; and this then allows quite novel ways of thinking about what a physical theory is and what it can do. The techniques also allow us to extrapolate in a meaningful way from simple calculations (such as low-order perturbation theory, or "poor man's scaling", to deduce more general non-perturbative results.

In what follows I begin by discussing 2 naive ways of doing renormalization. One is perturbative, and its use depends on theorems which one has to take on trust in the early stages. The other is naive or "poor man's" scaling, which is easy to implement, and leads already to very important results.

After this, we start a more technical discussion which goes through the details of renormalization. This has a number of different formulations, and here I will focus on both perturbative calculations - used as an input to what is called BPHZ renormalization in relativistic QFT, and to similar schemes in non-relativistic systems - and on the

renormalization group technique (RG). These 2 different techniques involve various technical devices, which I will introduce as we go along - notable amongst these are various regularization devices (the most important of which is dimensional regularization), and ways of doing integrals.

In all of this, one common thread comes through, and is the key to the whole business of renormalization. It is that we start from some theory, which we can think of as the "bare theory", and which consists of a set of degrees of freedom defined in some Hilbert space $\mathcal{H}(E_c)$,

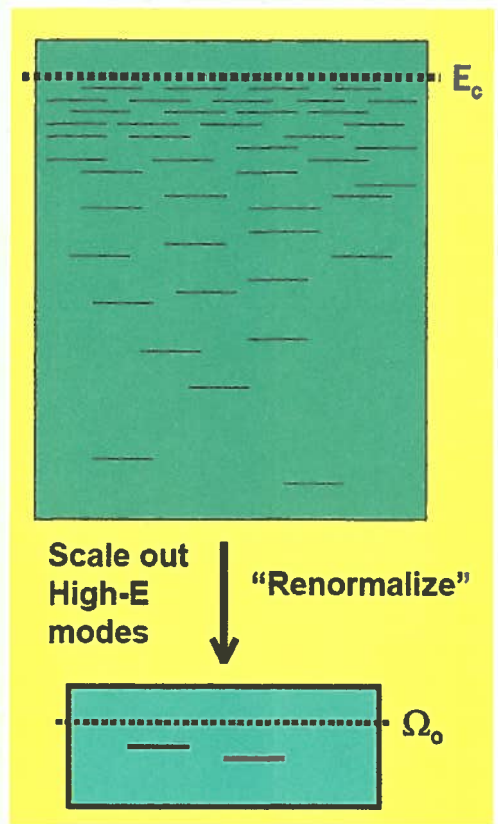


HILBERT SPACES (SCHEMATIC)

which conventionally we will define through its upper UV energy cut-off E_c . We then truncate this Hilbert space, down to a smaller one with UV cut-off Ω_0 .

The question then is how the theory changes under this change. We are reducing the number of degrees of freedom, and so the original Hamiltonian or Lagrangian or action is now defined for a smaller and different set of modes, with in general different eigenfunctions and different energy levels. The ideas of eigenmodes, and Hilbert spaces for these, may seem a little inappropriate to you if you prefer to think in path integral language - I will come presently to this point.

One of the first questions that one can ask here is - why should the Lagrangian change at all? After all, just because we've thrown away some degrees of freedom in our description doesn't mean that the real system, along with its real eigenfunctions & energy levels, has changed - it is not affected by something we write on paper! However this is to misunderstand what is being done here. To see



TRUNCATION/RENORMALIZATION

this, consider the scheme shown in the figure, in which a multi-level system with UV cut-off E_c to a 2-level system with cut-off Ω_0 . Then we go from $\mathcal{H}_{eff}(E_c)$ to $\mathcal{H}_{eff}(\Omega_0)$. The idea then is that these are different; in fact

$$\left. \begin{aligned} \mathcal{H}_{eff}^{(0)}(E_c) &= |\psi_i\rangle H_{ij}^{(0)} \langle \psi_j| \\ \rightarrow \mathcal{H}_{eff}^{(1)}(\Omega_0) &= |\varphi_\alpha\rangle H_{\alpha\beta}^{(1)} \langle \varphi_\beta| \end{aligned} \right\} (1)$$

and the point is that these are different, even for the 2 lowest levels, because in the original Hamiltonian $\mathcal{H}_{eff}^{(0)}(E_c)$, the 2 lowest levels interact with the higher levels, whereas in $\mathcal{H}_{eff}^{(1)}(\Omega_0)$ there are only 2 levels, and they are not allowed to interact with any other levels - the universe is exhausted by these 2 states in the effective 2-level Hamiltonian. By eliminating all interactions with higher levels, we are forced to change the eigenstates and energy levels. In the 1st section of this chapter we will look at toy models to see how this works.

Thinking in this way also allows us to imagine

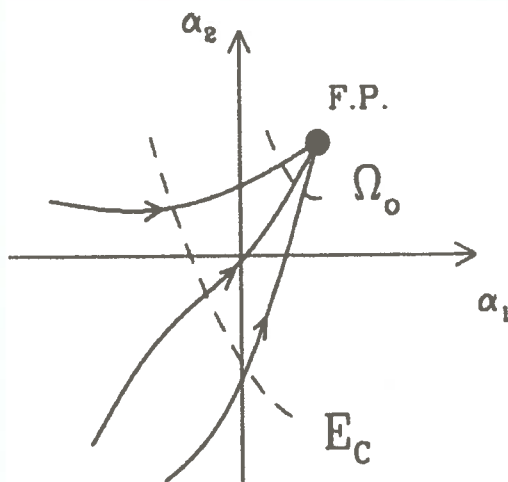
AMPAD

truncating out the degrees of freedom in incremental steps, rather than in one big jump. It is then clear that each infinitesimal change in the Hilbert space, usually done by making infinitesimal changes in the UV cut-off, will lead to an equally infinitesimal change in the effective Hamiltonian. Now suppose we write our effective Hamiltonians in the form

$$\mathcal{H}_{\text{eff}}(\Lambda) = - \sum_n g_n(\Lambda) \hat{O}_n(\{\varphi_j^\Lambda\}) \quad (2)$$

where the \hat{O}_n are operators acting on a set of functions $\{\varphi_j^\Lambda\}$, defined in a Hilbert space $\mathcal{H}(\Lambda)$, in which the UV cut-off Λ controls the dimension of the Hilbert space. Then as we change Λ , the relative importance of the different operators will change - the "coupling" $g_n(\Lambda)$ change with Λ as well.

If we then focus on some set of couplings which we consider to be relevant to the problem of interest. The figure shows this in the simple case where we have just 2 couplings, i.e., $g_n \rightarrow \alpha_1, \alpha_2$. As we change the UV cut-off, these couplings change - we can speak of the continuous "flow" or "RG flow" of the couplings as we change Λ . In the



FLOW TOWARD A FIXED POINT IN COUPLING CONSTANT SPACE. FOR UV CUTOFFS E_C OR Ω_0 , WE STOP ON LINES

figure this is shown for the case where for a large class of systems, each starting with different values for the initial high-energy couplings, the flow as Λ decreases is always towards a single point, known as the "fixed point" for this class of systems. These flow lines should be seen as representing the evolution of the $\{g_n(\Lambda)\}$ under changes of Λ , i.e., as showing the solution of a differential eqn of form

$$\frac{dg_n(\Lambda)}{d\Lambda} = \beta(g_n(\Lambda), \Lambda) \quad (3)$$

which describes the flow of the coupling constants - this differential eqn is the renormalization group eqn. In principle its behaviour and the solution can be very complicated, but we will see that for many physical systems, the flow takes us, in the low- E limit, to fixed point or fixed lines. The study of the RG structure of a class of effective Hamiltonians turns out to be enormously useful in physics.

One thing you should notice here is that the renormalization is typically being done in one direction, in which it is a UV cut-off that is being lowered, rather than an IR cut-off that is being increased (both have the effect of reducing the size of \mathcal{H} , and truncating out degrees of freedom). Why is this? In fact it is a way of expressing the reductionist worldview that has been prevalent in physics for centuries, according to which we build larger objects from smaller building blocks, and not vice-versa (we do not consider the microscopic structure of an object to be derivable or contingent upon its microstructure). Note, however, that in discussions of quantum measurement, one does make microscopic behaviour contingent upon macroscopic physics!

B.7.1. POOR MAN'S RENORMALIZATION

To do a proper job in renormalization theory, we sometimes have to introduce some fairly complex and sophisticated methods. To get a clear understanding of what is really going on it is then useful to begin by looking at some simple models which are amenable to simpler methods. So in this section we will study a selection of "toy models", and see how far we can get with these simpler methods.

We will be exploring, in these simple models, the 2 main ways in which renormalization is done. The first is perturbative - one tries to guess/infer/deduce the behaviour of the theory by calculating a few graphs, or maybe some class of graphs. This method is more powerful than it looks, for reasons which are not immediately obvious, but which will be explained in this section and in more detail later on. The 2nd method is called "poor man's scaling", and it consists in separating the modes into slow and fast variables, and then integrating out the fast variables to get an effective Hamiltonian/Lagrangian for the slow variables. We've already done this, of course - but now we will do it in a somewhat different way in which we derive an RG eqn of the form given in (3) above. This will be done with a certain kind of approximation, which is called the "poor man's" scaling method.

These 2 methods are related, as we will see. They are of course not the only ways one can think of to do renormalization, but they give us a good starting point.

AMPAD

B.7.1(a) PERTURBATION THEORY

Going back to our general description of renormalization as given in the intro to this chapter, we see that the simplest way to find the effect, in some low-energy Hilbert space, of the coupling to "virtual states" in a larger Hilbert space, is by doing perturbative expansions. We now see how this works on a variety of models.

(i) MULTI-LEVEL TOY MODELS : We already mentioned the idea of renormalizing a multi-level model in the introduction. This renormalization of a finite system is about the simplest thing we can imagine doing, and yet its study can be very illuminating. As noted above, the general idea is to truncate out the high-energy levels, to leave an effective Hamiltonian for some low-E states, i.e., we wish to implement in detail the manoeuvre described in eqn. (1).

2 COUPLED SPIN-1/2 SYSTEMS (COUPLED QUBITS) : Let's start with a pretty

simple system. We have a "high energy" effective Hamiltonian $\mathcal{H}(\tau_1, \tau_2)$ given by

$$\mathcal{H} = -\frac{1}{2}(\Delta_1 \hat{\tau}_1^x + \Delta_2 \hat{\tau}_2^x) - J_0 \hat{\tau}_1^z \hat{\tau}_2^z \tag{4}$$

where we have a transverse field acting on each "qubit", but a longitudinal interaction acting between them.

Before seeing how we would attack this problem perturbatively, let's note that we can solve it exactly with ease. One finds, for general values of Δ_1, Δ_2 , and J_0 , that the 4 eigenfunctions are

$$\left. \begin{aligned}
 |\psi_{++}\rangle &= \left(\frac{\Omega_+ + J_0}{2\Omega_+}\right)^{\frac{1}{2}} \left[\frac{\Delta_1 + \Delta_2}{2(\Omega_+ + J_0)} |\uparrow\uparrow\rangle + |\uparrow\downarrow\rangle + |\downarrow\uparrow\rangle + \frac{\Delta_1 + \Delta_2}{2(\Omega_+ + J_0)} |\downarrow\downarrow\rangle \right] \\
 |\psi_{+-}\rangle &= \left(\frac{\Omega_- + J_0}{2\Omega_-}\right)^{\frac{1}{2}} \left[\frac{\Delta_1 - \Delta_2}{2(\Omega_- + J_0)} |\uparrow\uparrow\rangle + |\uparrow\downarrow\rangle + |\downarrow\uparrow\rangle + \frac{\Delta_1 - \Delta_2}{2(\Omega_- + J_0)} |\downarrow\downarrow\rangle \right] \\
 |\psi_{-+}\rangle &= \left(\frac{\Omega_- - J_0}{2\Omega_-}\right)^{\frac{1}{2}} \left[\frac{\Delta_1 - \Delta_2}{2(\Omega_- - J_0)} |\uparrow\uparrow\rangle + |\uparrow\downarrow\rangle + |\downarrow\uparrow\rangle + \frac{\Delta_1 - \Delta_2}{2(\Omega_- - J_0)} |\downarrow\downarrow\rangle \right] \\
 |\psi_{--}\rangle &= \left(\frac{\Omega_+ - J_0}{2\Omega_+}\right)^{\frac{1}{2}} \left[\frac{\Delta_1 + \Delta_2}{2(\Omega_+ - J_0)} |\uparrow\uparrow\rangle + |\uparrow\downarrow\rangle + |\downarrow\uparrow\rangle + \frac{\Delta_1 + \Delta_2}{2(\Omega_+ - J_0)} |\downarrow\downarrow\rangle \right]
 \end{aligned} \right\} (5)$$

where we define

$$\Omega_{\pm}^2 = J_0^2 + \frac{1}{4}(\Delta_1 \pm \Delta_2)^2 \quad (6)$$

The eigenenergies of the system corresponding to the eigenstates in (5) take the simple form

$$\left. \begin{aligned}
 E_{++} &= \Omega_+ & E_{--} &= -\Omega_+ \\
 E_{+-} &= \Omega_- & E_{-+} &= -\Omega_-
 \end{aligned} \right\} (7)$$

and we plot these in the figure as a function of J_0 , the inter-qubit coupling.

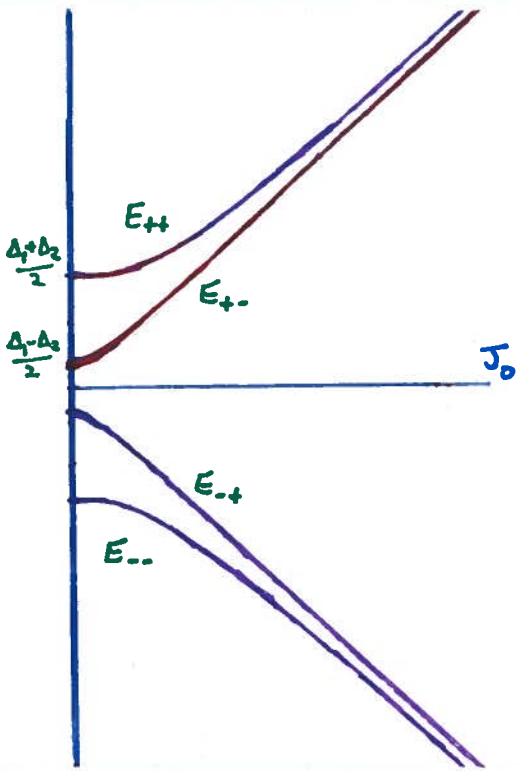
Now these results are fairly easy to understand, particularly in certain limits - one can then understand things in the whole parameter space by interpolating. We can consider the following limiting cases:

(i) FM limit (J_0 positive; $J_0 \gg \Delta_1, \Delta_2$): In this case we expect the ground state to closely resemble the state $|\psi_{--}\rangle$, which is very closely approximated in this limit by the simple triplet $\frac{1}{\sqrt{2}}(|\uparrow\downarrow\rangle + |\downarrow\uparrow\rangle)$, where the spins are forced to be parallel. The ground state energy is then $E_0 \sim E_{--} \sim -J_0$.

(ii) AFM limit (J_0 negative; $|J_0| \gg \Delta_1, \Delta_2$): Now we expect the ground state to be closely approximated by the singlet $\frac{1}{\sqrt{2}}(|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle)$, which is most closely mimicked by the state $|\psi_{++}\rangle$. The ground state energy E_0 is again given by $E_0 \sim -J_0$.

(iii) Decoupled limit ($|J_0| \ll \Delta_1, \Delta_2$): When the spins decouple, they are influenced only by the

transverse field acting on each spin; we therefore expect that the ground state will be closely approximated by $|\mathbb{E}\rangle \sim |\rightarrow\rangle\rangle \equiv \frac{1}{\sqrt{2}}[|\uparrow\uparrow\rangle, |\uparrow\downarrow\rangle]$, which is an equal weighting of all the states $|\uparrow\uparrow\rangle, |\uparrow\downarrow\rangle, |\downarrow\uparrow\rangle$ and $|\downarrow\downarrow\rangle$.



We get a much more detailed picture of what is going on if we look at the time dependence of the problem. This means calculating the time correlation functions - starting the system in some state at time $t=0$, and looking at it again after time t . Thus we can define the propagator

$$G_{\tau_1 \tau_2 \tau'_1 \tau'_2}(t) = \langle \tau_1 \tau_2 | e^{-i \hat{H} t} | \tau'_1 \tau'_2 \rangle \quad (8)$$

where $\tau_1, \tau_2, \tau'_1, \tau'_2 = \uparrow, \downarrow$; this defines the amplitude to go from state $|\tau'_1 \tau'_2\rangle$ to $|\tau_1 \tau_2\rangle$ after a time t . To get a look at this in a graph, without dealing with the complex variables involved, we look instead at

$$P_{\tau_1 \tau_2 \tau'_1 \tau'_2}(t) = |G_{\tau_1 \tau_2 \tau'_1 \tau'_2}(t)|^2 \quad (9)$$

and in the 1st of 2 figures here we

show the function $P_{\tau_1 \tau_2}(t) \equiv P_{\tau_1 \tau_2 \uparrow \uparrow}(t)$, i.e., we start the system in state $|\uparrow\uparrow\rangle$, and watch to see what happens; we do this in the strong coupling regime where $|J_0| \gg \Delta_1, \Delta_2$

(the sign of J_0 is not relevant here, but it is convenient to assume that $J_0 > 0$, i.e., we are in the strong FM limit).

The results show very clearly a slow oscillation of the system, in this FM limit, between states $|\uparrow\uparrow\rangle$ and $|\downarrow\downarrow\rangle$ at a frequency given by the splitting between the 2 lowest levels.

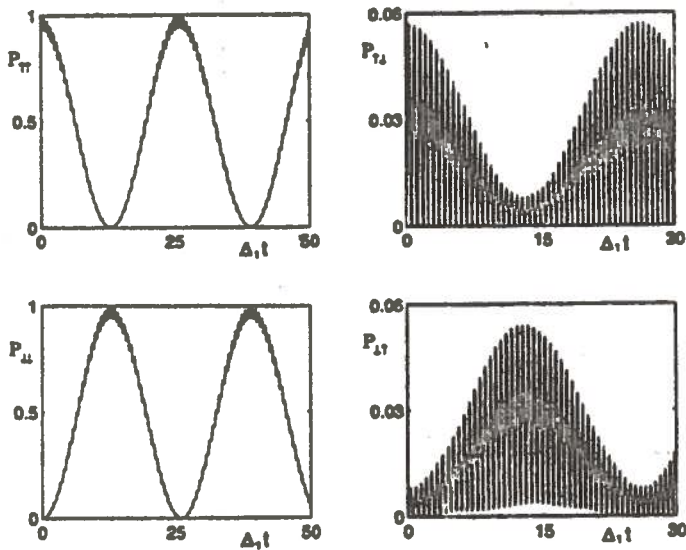


FIG.1
STRONG
COUPLING
REGIME
($J_0 \gg \Delta_1, \Delta_2$)

The probabilities $P_{\tau_1 \tau_2}(t)$ for a system of spins, coupled by J_0 , to start in the state $|\uparrow\uparrow\rangle$ at time $t=0$, and finish at time t in state $|\tau_1 \tau_2\rangle$. We assume $\Delta_2 = 2.5\Delta_1$, and a fairly strong coupling $|J_0| = 5\Delta_1$. $P_{\tau_1 \tau_2}(t)$ is plotted as a function of t/Δ_1 . Oscillations between the 2 states $|\uparrow\uparrow\rangle$ and $|\downarrow\downarrow\rangle$ occur at the slow "beat" frequency $\Delta_1 \Delta_2 / 2|J_0|$, with very weak high frequency oscillations superposed, coming from the weak mixing with the states $|\uparrow\downarrow\rangle$ and $|\downarrow\uparrow\rangle$. These high-frequency oscillations dominate $P_{\uparrow\downarrow}(t)$ and $P_{\downarrow\uparrow}(t)$, albeit with amplitude reduced by a factor $\sim (\Delta_1 + \Delta_2) / 2J_0$.

This frequency is equally found to be $\omega = \frac{1}{2} \Delta_1 \Delta_2 / J_0$, in this limit (it is just $\frac{1}{2} (E_{-+} - E_{--})$).

However we also notice that there is a weak small oscillation superposed on the slow oscillations, of much higher frequency (indeed of frequency $\propto J_0$). These are hardly visible in the probabilities $P_{nn}(t)$ and $P_{bb}(t)$, but if we look at the "off-diagonal" probabilities $P_{nb}(t)$ and $P_{bn}(t)$ (which tell us how much of the system is "leaking" out of the low-energy subspace into the 2 high-energy states) we see this more clearly (note however how much smaller is this probability, in comparison with $P_{nn}(t)$ and $P_{bb}(t)$).

Thus we see very directly in these results how the high frequency states affect the dynamics of the system, if we start it off in the low-energy subspace (and do not start it off in an eigenstate).

In the same way one can look at the case where $|J_0| \ll \Delta_1, \Delta_2$, so that the 2 spins are almost decoupled. As we see in the figure, the effect of J_0 is rather small, and the system spends its time cycling around between all 4 states in roughly equal measure.

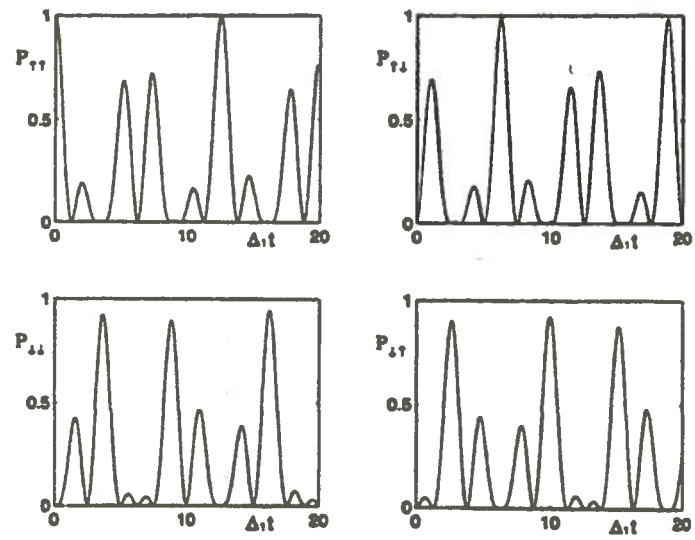


FIG. 2
WEAK
COUPLING
REGIME
($J_0 \ll \Delta_1, \Delta_2$)

The same probabilities as in ABOVE, but now for weak coupling; we have $\Delta_2 = 2.5\Delta_1$ again, but now $|J_0| = \Delta_1/10$.

The key difference between these 2 cases is of course the fact that there is a clear separation of energy scales in the strong coupling case. It then follows that the 2 subspaces (the high-energy doublet and the low-energy doublet) are largely independent of each other. This is not the case in the weak-coupling regime.

So now let's see how we might have attacked this problem using a perturbative approach. We will do this for the strong-coupling regime, where it makes sense - in fact, we will begin by taking the limit $\Delta_1/J_0, \Delta_2/J_0 \rightarrow 0$ and then perturbing about this limit. This is a "strong-coupling expansion", which is an expansion in the "kinetic" terms Δ_1, Δ_2 , about the strong-coupling limit. In the limit where $\Delta_1/J_0, \Delta_2/J_0 \rightarrow 0$, the ground state of the

system is a degenerate doublet, a linear combination of the states $|↑↑\rangle$ and $|↓↓\rangle$, and the first effect of the perturbation will be to break this degeneracy. The details of the perturbation calculation are standard. The low-energy subspace can be written in terms of the effective Hamiltonian

$$\begin{aligned}
 \mathcal{H}_{\text{eff}}(\mathcal{J}_0) &= |\varphi_\alpha\rangle H_{\text{eff}} \langle\varphi_\beta| \\
 &= (|↑↑\rangle \ |↓↓\rangle) \begin{pmatrix} 0 & \tilde{\Delta} \\ \tilde{\Delta} & 0 \end{pmatrix} \begin{pmatrix} |↑↑\rangle \\ |↓↓\rangle \end{pmatrix}
 \end{aligned} \tag{10}$$

where the renormalized splitting is just
$$\tilde{\Delta} = \frac{\Delta_1 \Delta_2}{2|\mathcal{J}_0|} \tag{11}$$

as we already know - here it is just obtained using 2nd-order perturbation theory, from the lowest degenerate doublet up to the higher doublet (with an energy gap of $2|\mathcal{J}_0|$) up to the higher intermediate state. We need the matrix elements Δ_1 and Δ_2 to flip each spin.

This example already shows the basic idea. Let's note again the key step is the truncation from a 4-level matrix to a much simpler 2-level one; in effect we have made the transition detailed in (1), with

$$H_y^{(0)} = \frac{1}{2} \begin{pmatrix} -2\mathcal{J}_0 & \Delta_2 & \Delta_1 & 0 \\ \Delta_2 & 2\mathcal{J}_0 & 0 & \Delta_1 \\ \Delta_1 & 0 & 2\mathcal{J}_0 & \Delta_2 \\ 0 & \Delta_1 & \Delta_2 & -2\mathcal{J}_0 \end{pmatrix} \begin{matrix} |↑↑\rangle \\ |↑↓\rangle \\ |↓↑\rangle \\ |↓↓\rangle \end{matrix} \tag{12}$$

$$H_{\text{eff}}^{(1)} = \frac{1}{2} \begin{pmatrix} 0 & \Delta\Delta_2/\mathcal{J}_0 \\ \Delta\Delta_2/\mathcal{J}_0 & 0 \end{pmatrix} \begin{matrix} |↑↑\rangle \\ |↓↓\rangle \end{matrix} \tag{13}$$

where on the right-hand side of the matrix we show the ordering of the states in the matrix

TUNNELING SPIN SYSTEM : We can easily extend the idea

discussed above to any multi-level system with a finite number of levels. Common examples include 3-level systems in which either one or two of the levels are at low energy, and the other(s) are truncated out - such models are common in optics.

The case of a tunneling spin S with $2S+1$ levels is of particular interest, because we can use both perturbation theory and WKB/Instanton methods on it, and compare the answers - we can even apply RG methods if we wish.

We have already met this system in the last chapter - the Hamiltonian

AMPAD

before truncation is of general form

$$H(\underline{S}) = \sum_{n=1}^{2S+1} K_{\alpha_1, \alpha_2, \dots, \alpha_n}^{(n)} \hat{S}^{\alpha_1} \hat{S}^{\alpha_2} \dots \hat{S}^{\alpha_n} \quad (14)$$

where $\alpha_j = x, y, z$. We will pick a particularly simple example, where

$$H = H_0 + V_0 = -K_0 \hat{S}_z^2 - B_0 \hat{S}_x = -\frac{1}{2} \left[\frac{K_0}{S} \hat{S}_z^2 + b_0 \hat{S}_x \right] \quad (15)$$

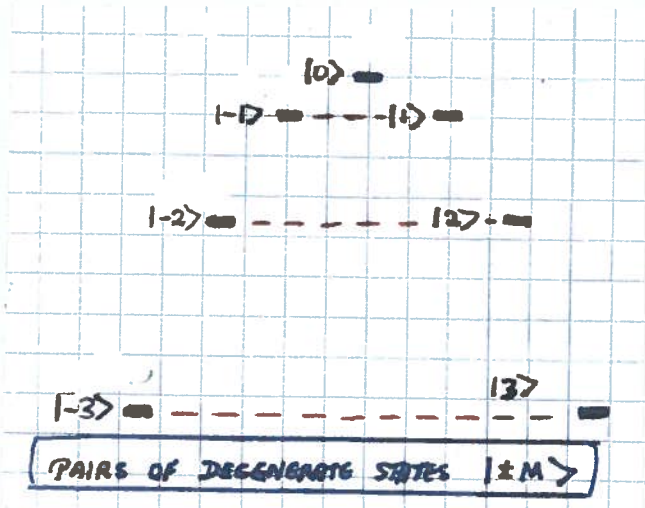
so that the transverse field B_0 breaks the symmetry w.r.t. rotations about \hat{z} .

A complete set of states for this system is provided by the $2S+1$ states $|m\rangle$, where

$$H_0 |m\rangle = -K_0 S_z^2 |m\rangle = -K_0 m^2 |m\rangle \quad (16)$$

and it is clear that if the transverse field is weak, we will only slightly mix up these states. If $|V_0| = B_0 S \ll |H_0| \approx K_0 S^2$, then we expect that the main effect of the perturbation will be to mix the degenerate pairs $|m\rangle$ and $|1-m\rangle$ into singlets and triplets, i.e., that the eigenstates of $H(S)$ will be accurately given by states

$$|\pm\rangle_m = \frac{1}{\sqrt{2}} (|m\rangle + |1-m\rangle) \quad (17)$$



with a splitting which can be calculated - see below. However, if the transverse field is strong, we expect that the levels will be severely scrambled by \hat{V}_0 , and a given level $|m\rangle$ will couple quite strongly to all the other levels. If we write the Hamiltonian (15) in matrix form, we see how this will work - we have the tridiagonal matrix

$$\begin{pmatrix} -S^2 K_0 & -\frac{1}{2} (2S)^{1/2} B_0 & 0 & 0 & 0 \\ -\frac{1}{2} (2S)^{1/2} B_0 & -(S-2)^2 K_0 & -\frac{1}{2} (4S-2)^{1/2} B_0 & 0 & \vdots \\ 0 & -\frac{1}{2} (4S-2)^{1/2} B_0 & -(S-2)^2 K_0 & \dots & 0 \\ \vdots & 0 & \dots & -(S-2)^2 K_0 & -\frac{1}{2} (2S)^{1/2} B_0 \\ 0 & 0 & \dots & -\frac{1}{2} (2S)^{1/2} B_0 & -S^2 K_0 \end{pmatrix} \quad (18)$$

where we use the standard result

$$\langle m | \hat{S}_x | m+1 \rangle = \frac{1}{2} [(S-m)(S+m+1)]^{1/2} \quad (19)$$

Now in what follows let's first see what perturbation theory does in general for this problem - and show a few simple examples - and then see how we can think of this in renormalization language.

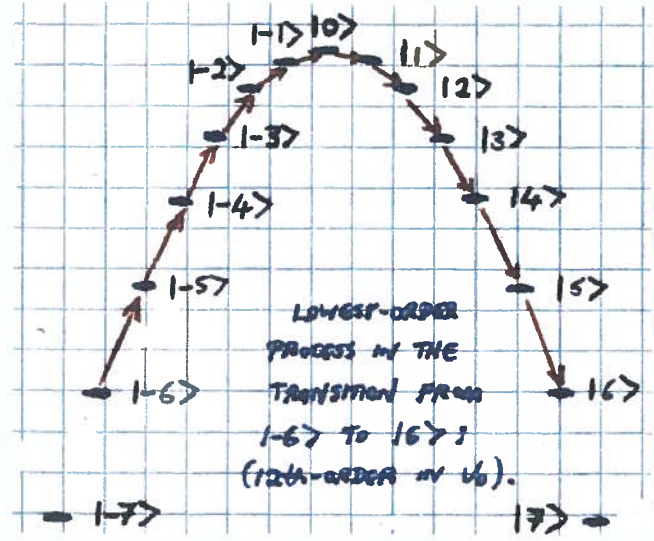
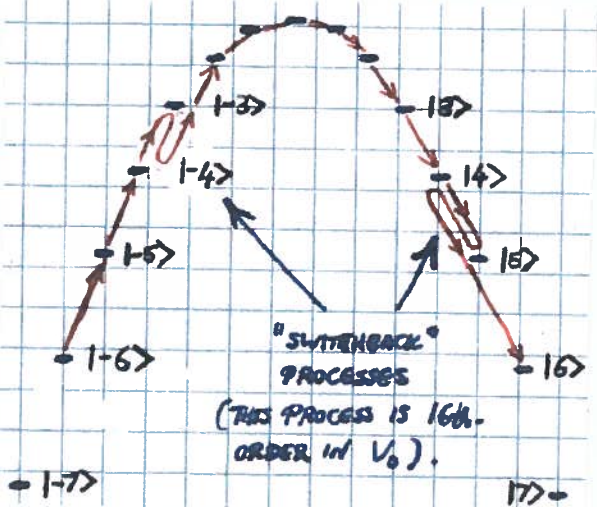
We begin by defining the matrix element, from (19), as

$$V_0^{m+1,m} = \langle m+1 | \hat{V}_0 | m \rangle = \frac{1}{2} B_0 [(S-m)(S+m+1)]^{1/2} \quad (20)$$

and then calculate the matrix element between the state $|m\rangle$ and $|m\rangle$ coming from this perturbation on H_0 . Applying $V_0 = -B_0 \hat{S}_x$ to the state $|m\rangle$ repeatedly, we find that this matrix element is

$$\Delta_m = 2 \left\{ \left[\frac{V_0^{-m,-(m-1)}}{\epsilon_{-(m-1)}^0 - \epsilon_m^0} \times \dots \times \frac{V_0^{m-2,m-1}}{\epsilon_{m-2}^0 - \epsilon_m^0} \frac{V_0^{m-1,m}}{\epsilon_{m-1}^0 - \epsilon_m^0} \right] + O(V_0^{2m+2}) \right\} \quad (21)$$

where the leading term involves a straight sequence of $2m$ transitions between $|m\rangle$ and $|m\rangle$, and the higher-order terms $\sim O(V_0^{2m+2})$ involve "switchback processes" in which the system retraces its perturbative path one or more times. All of this is shown in the figure, for a transition between states $|6\rangle$ and $|6\rangle$, in an $S=7$ system.



Now we may evaluate the dominant 1st term in (21) by carrying out the product - we have

$$\Delta_m = 2K_0 \left(\frac{B_0}{2K_0} \right)^{2m} \prod_{l=-m+1}^{m-1} \frac{1}{m^2 - l^2} \prod_{l'=-m}^{m-1} [(S-l')(S+l'+1)]^{1/2} \quad (22)$$

$$= A_m^S \left(\frac{B_0}{2K_0} \right)^{2m} K_0 + O \left(\frac{B_0}{2K_0} \right)^{2m+2} K_0$$

where the prefactor is

$$A_m^S = \frac{2}{[(2m-1)!]^2} \frac{(S+m)!}{(S-m)!} \quad (23)$$

AMPAD

From the point of view of renormalization theory, the most pertinent case here is that where $n = S$, i.e., where we look at the lowest doublet in the system. We then have

$$\Delta_S = \frac{4S(2S+1)}{(2S-1)!} \left(\frac{B_0}{2K_0}\right)^{2S} K_0 \quad \left. \begin{array}{l} \\ \\ \end{array} \right\} \quad (24)$$

$$\xrightarrow{S \gg 1} \frac{4}{\sqrt{\pi}} K_0 S^{3/2} \left(\frac{eB_0}{4SK_0}\right)^{2S}$$

which, for $S \gg 1$, clearly has a WKB form, i.e., we have


$$\Delta_S \xrightarrow{S \gg 1} A_S e^{-\Gamma_S} \quad (25)$$

with

$$\left. \begin{array}{l} A_S \sim K_0 S^{3/2} \\ \Gamma_S \sim 2S \ln\left(\frac{SK_0}{B_0}\right) \end{array} \right\} \quad (26)$$

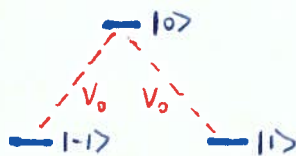
These results for large S show that the leading perturbative result for $S \gg 1$ just gives the leading WKB result; we glance below at the sub-dominant terms. For small S we can just write down the results:

- $S = 1/2$ (here $H_0 = 0$) Then perturbation theory gives the exact result:



$$\Delta_{1/2} = V_0^{1/2, -1/2} = B_0 \quad (27)$$

- $S = 1$ Then the transition amplitude Δ_1 between $|1-1\rangle$ and $|11\rangle$, from perturbation theory, is



$$\Delta_1 \sim \frac{|V_0^{01}|^2}{\epsilon_0 - \epsilon_1} = \frac{B_0^2}{K_0} \quad (28)$$

which is not the exact result - this latter is given by

$$\Delta_1 = \frac{1}{2} K_0 \left[\left(1 - 4B_0^2/K_0^2\right)^{1/2} - 1 \right] \quad (29)$$

which tends to the perturbative result (28) in the limit where $B_0/K_0 \ll 1$. This is because (28) neglects all switchback processes. We can continue in this vein:

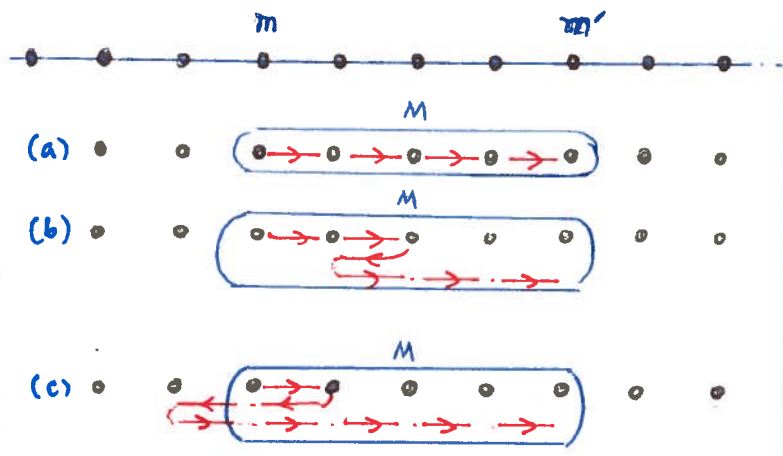
- $S = 3/2$ We now have

$$\left. \begin{array}{l} \Delta_{3/2} = \frac{3}{8} B_0^3/K_0^2 \\ \Delta_{1/2} = 2B_0 \end{array} \right\} S = 3/2 \quad (30)$$

• $S = 2$ Now we have
$$\left. \begin{aligned} \Delta_2 &= \frac{1}{12} \beta_0^4 / K_0^3 \\ \Delta_1 &= 3 \beta_0^2 / K_0 \end{aligned} \right\} S=2 \quad (31)$$

and we notice that Δ_1 is larger for $S=2$ than for $S=1$; this is because the larger Hilbert space allows more transition pathways.

Note that we don't have to think of this in terms of a spin problem. Suppose instead we just think of some line of sites, with a particle able to hop between them with some matrix element $V_{j,j+1}^0$, which in this case just happens to be given by (20). Then what we are doing here is calculating perturbatively the amplitude $\Delta_{mm'}$ to go between 2 sites m and m' . But there are different ways to make this transition. The first one is shown in the figure as process (a); it is the dominant one in perturbation theory. The process (b) is a switchback process, and contributes to higher order in the perturbation - but note that it never leaves the block of states, labelled M in the figure, which includes $|m\rangle$ and $|m'\rangle$, and all site states in between. There are also processes that go outside this block, and which still contribute to $\Delta_{mm'}$, the total transition amplitude between $|m\rangle$ and $|m'\rangle$. In our spin problem, we have to include such processes when we calculate



HOPPING PROCESSES BETWEEN SITES m & m' ON A LINE. THE PROCESSES (a) & (b) STAY WITHIN THE BLOCK M OF SITES INCLUDING m & m' . THE PROCESS (c) GOES OUTSIDE M .

Δ_m in (22), and they also contribute to the sub-dominant terms in (22).

Now let's think about all of this in renormalization terms. One way is obvious from what we did before - we can imagine truncating the system to the lowest doublet of states, i.e., truncate the Hamiltonian matrix in (18) right down to the qubit effective Hamiltonian

$$\mathcal{H}(\omega_0) = \begin{pmatrix} 0 & \Delta_S \\ \Delta_S & 0 \end{pmatrix} \quad (32)$$

with the matrix element Δ_S acting between states $|S\rangle$ and $|-S\rangle$, and the new UV cutoff energy ω_0 being $\propto K_0 S^2$. This is a standard renormalization to a low-energy effective Hamiltonian; the energy zero is rescaled in (32).

However, we notice that we could do things in quite a different way. Suppose that instead of truncating out all the high-energy levels, we restrict our

AMPAID

transition to the lowest levels - so that we keep only the high-energy levels, with small $|m|$. Then effectively we are truncating to an effective high-E Hamiltonian, which would make sense under some circumstances - eg., if we were looking primarily at thermal activation processes over the top of the barrier.

We can even go further - suppose we decide to keep only levels $|m\rangle$ and $|-m\rangle$, for some specific value of $|m\rangle$? This amounts to introducing both a UV and an IR cut-off in the problem. This would make sense if, for some reason, only these 2 levels were playing a direct role in the physics (eg., in some resonance process).

The effective Hamiltonians in these 2 cases are obvious from what we did before. We have, eg.,

$$H = \begin{pmatrix} -\tilde{E}_1 & \Delta_1 & 0 \\ \Delta_1 & 0 & \Delta_1 \\ 0 & \Delta_1 & -\tilde{E}_1 \end{pmatrix} \quad (33)$$

in the case where we just keep the 3 highest levels; and we have

$$H_0 = \begin{pmatrix} \tilde{E}_m & \Delta_m \\ \Delta_m & \tilde{E}_m \end{pmatrix} \quad (34)$$

in the case where we just keep the "m-pair" of levels (here the $\{\tilde{E}_m\}$ are the renormalized levels, shifted from E_m^0 ; we have not calculated these shifts here).

We can go further in this vein. Suppose we now go over to the site picture discussed above, and imagine isolating "blocks" of sites, like the "M-block" described above. Then we can imagine the following approach:

- (i) Calculate a renormalized effective Hamiltonian for the M-block sites only, in which no transitions to sites outside the M-block are allowed. This means both direct and switchback processes inside the block, but nothing else. We then get an effective H for the isolated block.
- (ii) Then introduce coupling between the block and all other sites - we can call these sites "the environment" if we wish. This will work fine if we can deal with this coupling perturbatively - it will give a new renormalized Hamiltonian for the M-block.
- (iii) Alternatively, we could calculate block Hamiltonians H_{M_1}, H_{M_2}, \dots for blocks all along the line, and then couple these together perturbatively to get a new renormalized H_{eff} for the whole line - but now it has been "coarse-grained", if in doing this we only incorporate the lowest states from each block. This technique is called "Kadanoff scaling" or "block scaling".

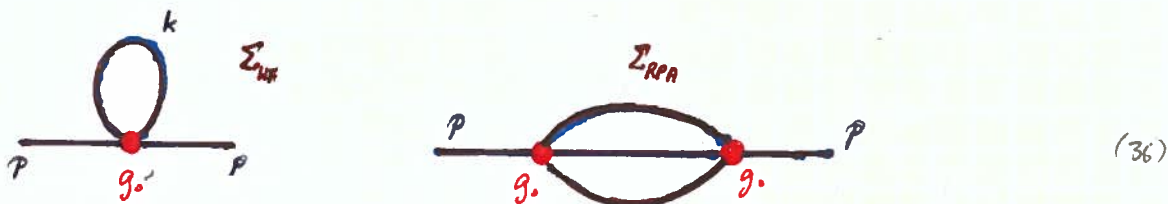
Simple ideas like this are at the basis of the whole RG technique, as introduced by Kadanoff & refined enormously by KG Wilson. We shall see how this works as we go along - its power becomes much more obvious when we go over to fields.

(ii) SCALAR FIELD THEORY : The original formulation of renormalization for QFT, before the idea of RG flow was really understood, was entirely based on perturbation theory, and in fact was technically quite complicated. To see how it works, & how one can try to extrapolate from low-order perturbation theory for a full quantum field, let's look at how perturbation theory is done for a simple scalar field. To be specific we pick the usual ϕ^4 theory, and see what the lowest-order diagrams look like. We will not calculate them (they are divergent), but this will not yet be necessary - we want instead to understand the relationships between them.

Thus we start again with our old favourite action, viz.,

$$S = \int d^4x \left\{ \frac{1}{2} [\partial_\mu \phi]^2 - m^2 \phi^2 - \frac{g_0}{4!} \phi^4(x) \right\} \quad (35)$$

Now, to prepare ourselves for what is coming, let's look at a few lower-order graphs. We first look at graphs for the self-energy - the 2 simplest ones are shown in the figure



for which we have the formal expressions

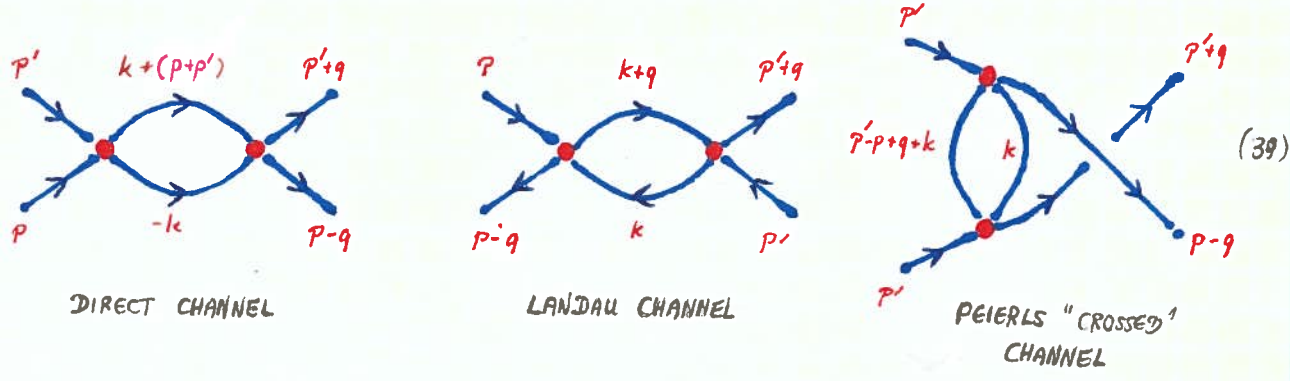
$$\Sigma_{HF}(p) = -\frac{ig_0}{\hbar} \int \frac{d^4k}{(2\pi)^4} \frac{i\hbar}{k^2 - m^2 + i\delta} \quad (37)$$

$$\Sigma_{RPA}(p) = \left(\frac{-ig_0}{\hbar} \right)^2 \int \frac{d^4k}{(2\pi)^4} \int \frac{d^4q}{(2\pi)^4} \frac{i\hbar}{k^2 - m^2 - i\delta} \frac{i\hbar}{(k+q)^2 - m^2 + i\delta} \frac{i\hbar}{(p-q)^2 - m^2 + i\delta} \quad (38)$$

and we notice that, in the absence of a UV cutoff, the graph for $\Sigma_{HF}(p)$ is divergent - indeed, with a UV cutoff Λ it is $\propto \Lambda^2$, and we say that it is "quadratically divergent". The same is true of the graph for $\Sigma_{RPA}(p)$ in (38). We see this by "power counting"; both graphs have 2 higher powers of 4-momentum in the numerator than in the denominator.

Suppose we now look at some graphs for the 4-point vertex. The lowest

graphs for the 4-point vertex are very familiar to you, and are shown in the figure.



where one should refer back to Chapter B5, eqns. (170)-(172), for more details. (NB: I have drawn the Peierls diagram a little differently, but one sees from the arguments of the internal lines that it is the same). One sees that the divergences involved are the same for all 3 diagrams. Thus the Landau graph just takes the usual form

$$\Gamma_L(p, p'; q) = \left(\frac{-ig_0}{\hbar}\right)^2 \int \frac{d^4k}{(2\pi)^4} \frac{i\hbar}{(k+q)^2 - m^2 + i\delta} \frac{i\hbar}{k^2 - m^2 + i\delta} \quad (40)$$

which we see is logarithmically divergent, i.e., with a cutoff Λ we have $\Gamma_L \sim O(\ln \Lambda)$.

We see that for a relativistic QFT with no UV cut-off, we are faced with a rather severe problem of UV divergences (there are no IR divergences here, because the mass $m > 0$; for massless propagators we have to face these as well). How can we deal with this?

DEGREE OF DIVERGENCE

To understand what to do, we begin by defining the "superficial degree of divergence" of a diagram. What this can be guessed from what we've just done - it refers to the dependence of the diagram on the UV cutoff momentum, and is determined by adding the powers coming into the graph from internal loop integration, say momentum dependence of the interaction itself, and the propagators in the diagram.

For the ϕ^4 theory just discussed above, the superficial degree of divergence is easy to analyze. We've seen that this is 2 for 2-point function, and 0 for 4-point function; and if we look at a 6-point graph we quickly find that the degree of divergence is -2, i.e., that the integral is quadratically convergent (check for yourself!). So for a ϕ^4 theory, we find the very simple result that the degree of divergence d is given by

$$d = D - n_B \quad (\phi^4 \text{ theory}) \quad (41)$$

where D is the spacetime dimension (here $D=4$) and n_B is the number of external legs to the graph.

We can easily generalize this expression to some arbitrary relativistic QFT

AMPAD

involving both fermion lines and boson lines, and with interactions that themselves have some non-trivial momentum dependence, because they involve derivatives ∂_μ of the fields to some power. In what follows, just to simplify things, I will ignore such derivative interactions (it is easy to put them back in) and consider some action of form

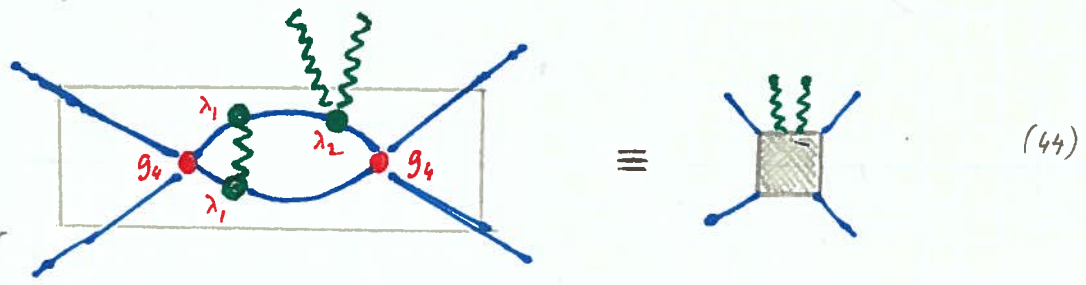
$$S[\phi_k, \psi_j] = S_0 + \sum_\alpha S_\alpha [g_\alpha; \phi_k, \psi_j] \tag{42}$$

where the $\{\phi_k\}$ denote different bosonic fields, and the $\{\psi_j\}$ different fermionic fields; the term S_0 is the free field term, with no couplings in it. Now each of the S_α will in principle involve different couplings, having different powers of the fields in it.

Let's now define the following quantities, for a given diagram

No. of external fermion legs :	n_F	}	(43)
No. of external boson legs :	n_B		
No. of internal fermion lines :	l_F		
No. of internal boson lines :	l_B		
No. of internal vertices of type α :	N_V^α		
fermionic power in vertex type α :	P_F^α		
bosonic power in vertex type α :	P_B^α		

To see what is meant by this, consider the graph shown in the figure, which describes a 6-point vertex $\Gamma_{42}^{\psi\psi\phi}$ for interacting fermion-boson theory.



Note that we have 3 different interaction terms in this theory. There is a 4-fermion interaction g_4 , plus an interaction between 2 fermions and either a single boson, or a pair of boson; we thus have

$$S_0 = S_0 = \int d^D x \left\{ g_4 (\bar{\psi}\psi)^2 + \lambda_1 (\bar{\psi}\psi\phi) + \lambda_2 (\bar{\psi}\psi\phi^2) \right\} \tag{45}$$

and these 3 interaction terms (with $\alpha = 1, 2, 3$) have $P_F^\alpha = 4, 2, 2$, and $P_B^\alpha = 0, 1, 2$, respectively. In the diagram in the figure, we also have $n_F = 4$, $n_B = 2$, $l_F = 2$, $l_B = 1$, and we have $N_V^\alpha = 2, 2, 1$ for $\alpha = 1, 2, 3$.

We can now establish some simple general results for the degree of divergence of an arbitrary graph for some theory. It is easy to see that

$$d = D(1 - \sum_\alpha N_V^\alpha) + (D-2)l_B + (D-1)l_F \tag{46}$$

We get (46) by noting the following: first, every internal vertex comes with a delta-function of momentum variables, in D spacetime dimensions; this would give a factor $-\sum_{\alpha} D N_V^{\alpha}$, but the δ -function integration which gives overall momentum conservation should not be included in the sum. Then for every internal boson line we have a D -dimensional integration, but with an integrand $\sim O(1/k^2)$, giving a factor $D-2$; and every fermion line has an integrand $\sim O(1/k)$, giving $D-1$. Hence we get (46).

Another way to write this formula is obtained by noting that

$$\left. \begin{aligned} \eta_B + 2L_B &= \sum_{\alpha} N_V^{\alpha} P_B^{\alpha} \\ \eta_F + 2L_F &= \sum_{\alpha} N_V^{\alpha} P_F^{\alpha} \end{aligned} \right\} \quad (47)$$

which arise because every external boson/fermion line is attached to one vertex, and every internal line to 2 vertices; summing over all lines we get (47). Substituting this into (46), we get

$$d = D - \frac{D-2}{2} \eta_B - \frac{D-1}{2} \eta_F + \sum_{\alpha} N_V^{\alpha} \Delta_{\alpha} \quad (48)$$

where Δ_{α} is the "index of divergence", given by

$$\Delta_{\alpha} = \frac{1}{2} \left[(D-2) P_{\alpha}^B + (D-1) P_{\alpha}^F \right] - D \quad (49)$$

To take two simple examples, we have

- For ϕ^4 theory, the index of divergence $\Delta_{\alpha} = 0$; the 4-boson interaction has $P_{\alpha}^B = 4$, and $P_{\alpha}^F = 0$ because there are no fermions; thus we get $\Delta_{\alpha} = \Delta_{\phi_4} = 0$ when $D = 4$. We then immediately recover the result given in (41) above for ϕ^4 theory in 4 dimensions - the degree of divergence depends only on the number of external legs. Note however that this is only true when $D = 4$; thus, e.g., when $D = 3$, then $\Delta_{\phi_4} = -1$, and the degree of divergence of any graph depends not only on η_B , but also on the number of internal vertices N_V .
- For the graph given in the figure (44), we have a degree of divergence $D-4$; this is easy to see using power counting, and one can also recover it using (48) and (49).

Thus we see that any diagrammatic expansion will be full of divergences, of different degrees.

SUBTRACTIONS & GRADIENT EXPANSIONS: Although the various vertices

in a relativistic QFT are, as we've seen, very often divergent, their behaviour is otherwise not so complicated to understand. We can see this if we again consider the 2-loop graph $\Sigma_{RPA}(p)$ in (36). Although this is quadratically divergent, there is nothing to stop us performing a gradient expansion (which here becomes an expansion in powers of p) around some specific value of p ; the simple

pole structure of the integrand makes such an expansion straightforward (using, e.g., a partial fraction expansion).

The general form that such an expansion will take is then

$$\Sigma(p^2) = \Sigma(p_0^2) + (p^2 - p_0^2) \frac{\partial \Sigma}{\partial p^2} \Big|_{p^2=p_0^2} + \Delta \Sigma_{p_0}(p^2) \quad (50)$$

where p_0 is some arbitrary momentum around which we expand, and we write everything in terms of p^2 because everything is an even function of p (and of m).

The basic idea now is to choose $p_0 = \tilde{m}$, the "physical mass", i.e., the low-energy limiting mass of the theory - assuming there is such a limit. We then see that the physical mass must be given by

$$\tilde{m}^2 = m^2 + \Sigma(\tilde{m}^2) \quad [\text{MASS RENORMALIZATION}] \quad (51)$$

i.e., that m is renormalized to \tilde{m} by Σ ; this is a self-consistent eqn. for \tilde{m} . We can then write

$$\begin{aligned} \mathcal{G}_2(p) &= \frac{i\hbar}{[1 - \partial \Sigma / \partial p^2 \Big|_{\tilde{m}^2}] (p^2 - \tilde{m}^2) - \Delta \Sigma_{\tilde{m}}(p^2) + i\delta} \\ &\equiv \frac{i\hbar Z_\phi}{(p^2 - \tilde{m}^2) - \tilde{\Sigma}(p^2) + i\delta} \end{aligned} \quad (52)$$

where the factor Z_ϕ is called the wave-function or field renormalization factor, and is

$$Z_\phi = \frac{1}{1 - \partial \Sigma / \partial p^2 \Big|_{p^2=\tilde{m}^2}} \quad [\text{FIELD RENORMALIZATION}] \quad (53)$$

and where it is also incorporated into the renormalized self-energy, i.e., we have

$$\tilde{\Sigma}(p^2) = Z_\phi \Delta \Sigma(p^2) \quad (54)$$

The name for mass renormalization is fairly obvious from its definition in (51). That Z_ϕ is a field renormalization is clear from the definition of $\mathcal{G}_2(p)$, which is quadratic in the field operators ϕ . Thus if we define a renormalized field

$$\tilde{\phi}(x) = \frac{1}{\sqrt{Z_\phi}} \phi(x) \quad (55)$$

then we have

$$\mathcal{G}_2(p) = \frac{1}{Z_\phi} G_0(p) \quad (56)$$

where the bare propagator $G_0(p)$ is, as usual, given by

$$G_0(p) = i\hbar \Delta_F(p) = \frac{i\hbar}{(p^2 - m^2) + i\delta} \quad \left. \vphantom{G_0(p)} \right\} (57)$$

$$= \int d^4x e^{-ip \cdot x} \langle 0 | T \{ \phi(x) \phi(0) \} | 0 \rangle_{\text{bare}}$$

Thus we establish that, given a well-behaved expansion for the 2-point vertex around some fixed momentum, we can define a renormalized vertex having the same form as the bare propagator, but with renormalized parameters. Two points are important here:

- In a relativistic QFT like ϕ^4 theory, these renormalizations are infinite, because of the divergences. Thus in (51), we have a finite physical mass m_0 ; but since $\Sigma(m_0^2)$ is in general infinite, this implies that m^2 is also (with the opposite sign). The same is true of Z_ϕ^{-1} . That these infinities exist is not surprising - they come because the original fields and masses refer to fields defined up to infinite energy.
- The basic form of the renormalization has nothing to do with the issue of divergence - at no point did the divergences enter the discussion from eqns (50) - (56). In fact we have seen all of this before, in the discussion of Fermi liquid theory in Chapter 5, where both the wave-function renormalization and the changes in $\Sigma(p)$ were finite. This finiteness arose simply because in any condensed matter effective Hamiltonian, a finite UV cut-off emerges naturally.

From this latter perspective it is clear that it is physically meaningless to be integrating up to infinite momenta in the internal states in a diagram. As we've already emphasized several times, any action like a ϕ^4 action has to be viewed as an effective action, only valid over a certain energy range - it then makes no sense to integrate outside this energy range.

COUNTERTERMS & BPHZ RENORMALIZATION: So far we have seen

how one might set up a renormalization scheme, even in the presence of divergences; but there is still no reason to assume that this is capable of dealing with all diagrams. Thus, in dealing with the ϕ^4 theory, we have seen how one might consistently renormalize the 2-point vertex, by introducing 2 (infinite) renormalization constants. But what about all the other vertices - do we have to now introduce a new set of renormalization constants for each new vertex (i.e., 4-point, 6-point, etc). A theory with an infinite set of renormalization constants is no theory at all.

This turns out not to be the case if we have a renormalizable theory. A good way to set up the whole structure was pioneered by

Bogolubov, with later refinements and generalization by Parasiak, Hepp, & Zimmermann. The idea of counterterms was first introduced in order to make the smooth link between the bare Lagrangian and the renormalized one.

The basic idea is fairly simple. Suppose that in doing a few lower order calculations of diagrams we find that the results for the lowest vertex parts can be rewritten in renormalized form as shown above, with a shifted mass, wave-function renormalization, and a renormalized coupling (these being the 3 relevant objects in the ϕ^4 theory). Now, the argument goes, let's make these shifts before we even start the calculation. In other words, let's add terms to the original action which already make these shifts.

For the ϕ^4 Lagrangian it is completely obvious how to do this. We simply write a new action

$$\tilde{S} = S - \int d^4x \left\{ \frac{1}{2} [A \partial_\mu \phi \partial^\mu \phi - B \phi^2] - \frac{C}{4!} \phi^4 \right\} \quad (58)$$

where A, B , and C are determined by low-order calculations of graphs. In the case of ϕ^4 theory, these constants are given by defining the new renormalized quantities \tilde{m} , $\tilde{\phi}$, and \tilde{g} , so that

$$\tilde{S}[\tilde{\phi}] = \int d^4x \left\{ \frac{1}{2} [\partial_\mu \tilde{\phi} \partial^\mu \tilde{\phi} - \tilde{m}^2 \tilde{\phi}^2] - \frac{\tilde{g}}{4!} \tilde{\phi}^4 \right\} \quad (59)$$

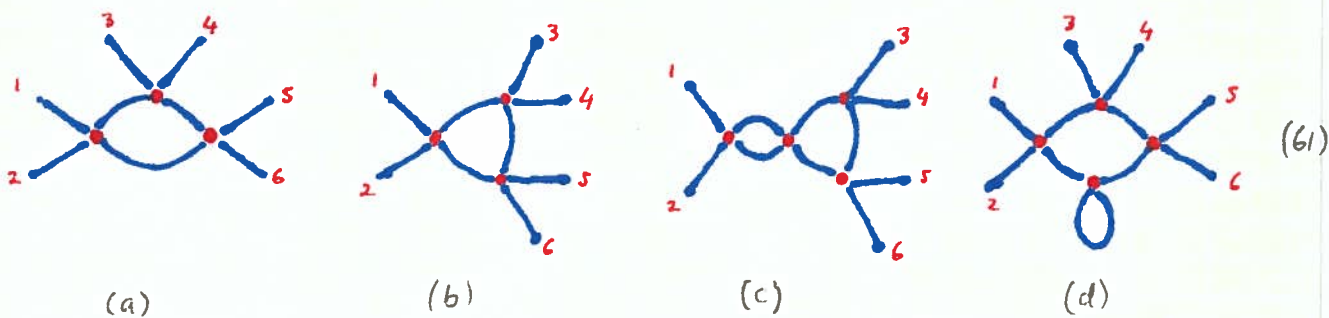
with the renormalized quantities adjusted to agree with the physical quantities. This then fixes A, B , and C ; we have

$$\left. \begin{aligned} \tilde{\phi}(x) &= \phi(x) \sqrt{1+A} \\ \tilde{m}^2 &= \frac{m^2+B}{1+A} \\ \tilde{g} &= \frac{g+C}{(1+A)^2} \end{aligned} \right\} \quad (60)$$

So far so good. But now the obvious question is - how do we know that this will work when we calculate any other graph, or when we do a graphical sum? There are a number of things that could easily go wrong; here are some of them:

- (1) The results for the renormalization in (60) seem natural enough; we have a field renormalization $\sqrt{1+A} \equiv Z_\phi^{-1/2}$, and the mass and coupling constant also contain this factor (squared for \tilde{m}^2 , and to the 4th power for \tilde{g} , because these vertices have 2 and 4 external legs respectively). Otherwise we simply shift \tilde{m}^2 and \tilde{g} by constants (which happen to be infinite). So what now happens when we calculate, say, the contributions to the 6-point or 8-point vertices? There are clearly 2 possible things that might ensue; either we generate new kinds of renormalization, and then we will need to include further terms in \tilde{S} of order ϕ^6 , ϕ^8 , etc; or we will find that the renormalized vertices for these

higher contributions can all be understood in terms of the three renormalizations that we already have in (60). Thus the question boils down to the consideration of the graphs shown in the figure, and others like them, for contributions to Γ_6 , and so on:

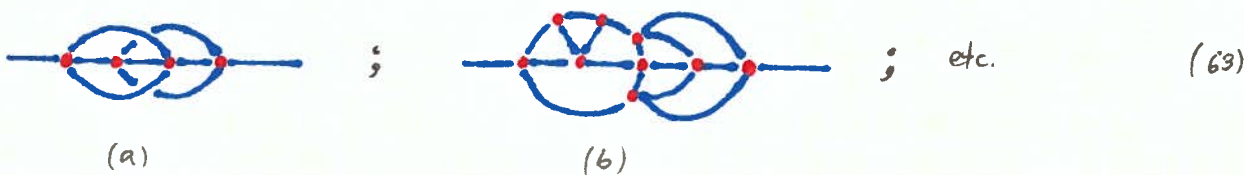


All of these graphs contribute to $\Gamma_6(1,2,\dots,6)$; but the real question we are trying to answer is - do they require us to add a term like

$$\delta \tilde{S}_6[\phi] \sim \int d^4x \mathcal{D}\phi^6(x) \tag{62}$$

to our effective action, or is the form (59) good enough? Note that all of the graphs in (61) are superficially convergent, from our original result (47); divergence is not the issue here. Note also that graph (d) in (61) just involves a simple self-energy insertion in (a), so we don't expect it to be independent of (a). However (b) and (c) obviously have a different structure from (a) and (d).

(ii) A separate question - what about divergences in higher order graphs? In ϕ^4 theory we have seen that higher vertices (6-point, 8-point, etc) are not divergent. But what about higher contributions to the 2-point and 4-point functions - how do we know that they will not contain different kinds of divergence from the ones we have already looked at in lowest-order perturbation theory? Thus, consider the graphs shown here for the 2-point vertex $\Sigma(p)$; what reason do we have to suppose that, even if these do not change the form of \tilde{S} , with 3 counterterms, that they will not considerably change the form of the counterterms? Notice that (a) involves an overlap of 2 graphs we've already seen (cf. (36)b); and (b) below involves insertions of 6-point vertex diagrams from (61) above.



These questions (and others) are not so easy to answer, and they require the full apparatus of renormalization theory. To give proper answers, therefore, lies beyond the scope of the "poor man's" theory we are doing here. However, we can summarize

the conclusion of all of this work by distinguishing between 3 possible classes of theory.

(a) SUPER-RENORMALIZABLE THEORY: In this case, by definition, the index of divergence Δ_α defined in eqn (49) is negative, i.e., we have

$$\Delta_\alpha < 0 \quad (\text{SUPER-RENORMALIZABLE}) \quad (63)$$

for all of the different terms in the action. In this case the total number of divergent graphs is actually finite - so we can catalogue these different divergences and after that everything is finite. An example of a super-renormalizable theory is the ϕ^3 scalar field theory in $D=4$ spacetime dimensions, where $\Delta = -1$. It is easy to see that there is a small number of graphs that diverge for this theory. Notice that for ϕ^3 theory, in 4 dimensions, we have a degree of divergence for any graph given by

$$\left. \begin{aligned} d &= D - n_B + \Delta \\ &= 3 - n_B \end{aligned} \right\} (\phi^3 \text{ theory, } D=4) \quad (64)$$

and so the only graphs having any divergences will be 2-point and 3-point graphs.

(b) RENORMALIZABLE THEORY: In this case we have a dividing line between 2 cases; by definition, a renormalizable interaction with coupling g_α has

$$\Delta_\alpha = 0 \quad (\text{RENORMALIZABLE}) \quad (65)$$

In this case, which we have already seen with ϕ^4 theory in 4-d, the degree of divergence d for graphs for some vertex part ~~is~~ is independent of the number of internal vertices. If the lower-order graphs are divergent, then so will the higher-order graphs be so. However, crucially, we only have a finite number of different kinds of divergence. This means that the "counterterm" procedure of Bogolubov's will work - if we now calculate with the renormalized Lagrangian, then everything will come out finite.

(c) NON-RENORMALIZABLE THEORY: In this case, the most problematic, we have a positive index of divergence, i.e., we have

$$\Delta_\alpha > 0 \quad (\text{NON-RENORMALIZABLE}) \quad (66)$$

for some interaction with coupling g_α . In this case, the degree of divergence d increases as we increase the number of g_α -vertices in our diagrams, and therefore any diagram, no matter how many external legs it has, will become divergent at sufficiently high order in g_α . An example is ϕ^5 or ϕ^6 theory. One then finds that we need an infinite number of counterterms to deal with these divergences - the theory becomes apparently useless.