Chapter I.1

Who Needs It?

Who needs quantum field theory?

Quantum field theory arose out of our need to describe the ephemeral nature of life.

No, seriously, quantum field theory is needed when we confront simultaneously the two great physics innovations of the last century of the previous millennium: special relativity and quantum mechanics. Consider a fast moving rocket ship close to light speed. You need special relativity but not quantum mechanics to study its motion. On the other hand, to study a slow moving electron scattering on a proton, you must invoke quantum mechanics, but you don't have to know a thing about special relativity.

It is in the peculiar confluence of special relativity and quantum mechanics that a new set of phenomena arises: Particles can be born and particles can die. It is this matter of birth, life, and death that requires the development of a new subject in physics, that of quantum field theory.

Let me give a heuristic discussion. In quantum mechanics the uncertainty principle tells us that the energy can fluctuate wildly over a small interval of time. According to special relativity, energy can be converted into mass and vice versa. With quantum mechanics and special relativity, the wildly fluctuating energy can metamorphose into mass, that is, into new particles not previously present.

Write down the Schrödinger equation for an electron scattering off a proton. The equation describes the wave function of one electron, and no matter how you shake and bake the mathematics of the partial differential equation, the electron you follow will remain one electron. But special relativity tells us that energy can be converted to matter: If the electron is energetic enough, an electron and a positron ("the antielectron") can be produced. The Schrödinger equation is simply incapable of describing such a phenomenon. Nonrelativistic quantum mechanics must break down.

You saw the need for quantum field theory at another point in your education. Toward the end of a good course on nonrelativistic quantum mechanics the interaction between radiation and atoms is often discussed. You would recall that the electromagnetic field is treated as a field; well, it is a field. Its Fourier components are quantized as a collection of harmonic oscillators, leading to creation and annihilation operators for photons. So there, the electromagnetic field is a quantum

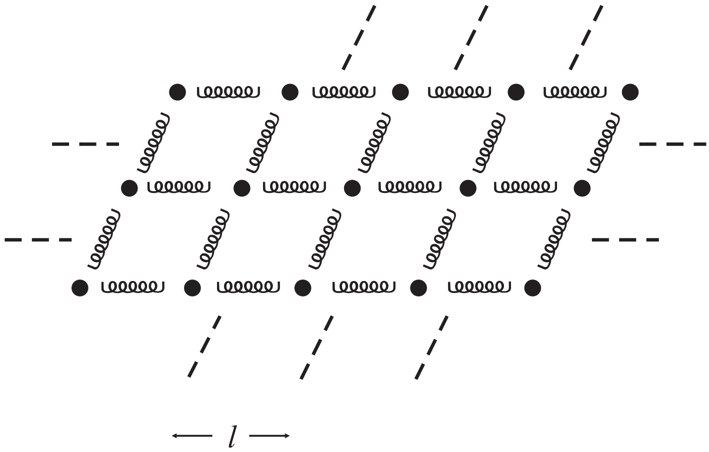


Figure I.1.1

field. Meanwhile, the electron is treated as a poor cousin, with a wave function $\Psi(x)$ governed by the good old Schrödinger equation. Photons can be created or annihilated, but not electrons. Quite aside from the experimental fact that electrons and positrons could be created in pairs, it would be intellectually more satisfying to treat electrons and photons, as they are both elementary particles, on the same footing.

So, I was more or less right: Quantum field theory is a response to the ephemeral nature of life.

All of this is rather vague, and one of the purposes of this book is to make these remarks more precise. For the moment, to make these thoughts somewhat more concrete, let us ask where in classical physics we might have encountered something vaguely resembling the birth and death of particles. Think of a mattress, which we idealize as a 2-dimensional lattice of point masses connected to each other by springs (Fig. I.1.1) For simplicity, let us focus on the vertical displacement [which we denote by $q_a(t)$] of the point masses and neglect the small horizontal movement. The index a simply tells us which mass we are talking about. The Lagrangian is then

$$L = \frac{1}{2}\left(\sum_a m\dot{q}_a^2 - \sum_{a,b} k_{ab}q_a q_b - \sum_{a,b,c} g_{abc}q_a q_b q_c - \cdots\right) \qquad (1)$$

Keeping only the terms quadratic in q (the "harmonic approximation") we have the equations of motion $m\ddot{q}_a = -\sum_b k_{ab}q_b$. Taking the q's as oscillating with frequency ω, we have $\sum_b k_{ab}q_b = m\omega^2 q_a$. The eigenfrequencies and eigenmodes are determined, respectively, by the eigenvalues and eigenvectors of the matrix k. As usual, we can form wave packets by superposing eigenmodes. When we quantize the theory, these wave packets behave like particles, in the same way that electromagnetic wave packets when quantized behave like particles called photons.

Since the theory is linear, two wave packets pass right through each other. But once we include the nonlinear terms, namely the terms cubic, quartic, and so forth in the q's in (1), the theory becomes anharmonic. Eigenmodes now couple to each other. A wave packet might decay into two wave packets. When two wave packets come near each other, they scatter and perhaps produce more wave packets. This naturally suggests that the physics of particles can be described in these terms.

Quantum field theory grew out of essentially these sorts of physical ideas.

It struck me as limiting that even after some 75 years, the whole subject of quantum field theory remains rooted in this harmonic paradigm, to use a dreadfully pretentious word. We have not been able to get away from the basic notions of oscillations and wave packets. Indeed, string theory, the heir to quantum field theory, is still firmly founded on this harmonic paradigm. Surely, a brilliant young physicist, perhaps a reader of this book, will take us beyond.

Condensed matter physics

In this book I will focus mainly on relativistic field theory, but let me mention here that one of the great advances in theoretical physics in the last 30 years or so is the increasingly sophisticated use of quantum field theory in condensed matter physics. At first sight this seems rather surprising. After all, a piece of "condensed matter" consists of an enormous swarm of electrons moving nonrelativistically, knocking about among various atomic ions and interacting via the electromagnetic force. Why can't we simply write down a gigantic wave function $\Psi(x_1, x_2, \cdots, x_N)$, where x_j denotes the position of the jth electron and N is a large but finite number? Okay, Ψ is a function of many variables but it is still governed by a nonrelativistic Schrödinger equation.

The answer is yes, we can, and indeed that was how solid state physics was first studied in its heroic early days, (and still is in many of its subbranches.)

Why then does a condensed matter theorist need quantum field theory? Again, let us first go for a heuristic discussion, giving an overall impression rather than all the details. In a typical solid, the ions vibrate around their equilibrium lattice positions. This vibrational dynamics is best described by so-called phonons, which correspond more or less to the wave packets in the mattress model described above.

This much you can read about in any standard text on solid state physics. Furthermore, if you have had a course on solid state physics, you would recall that the energy levels available to electrons form bands. When an electron is kicked (by a phonon field say) from a filled band to an empty band, a hole is left behind in the previously filled band. This hole can move about with its own identity as a particle, enjoying a perfectly comfortable existence until another electron comes into the band and annihilates it. Indeed, it was with a picture of this kind that Dirac first conceived of a hole in the "electron sea" as the antiparticle of the electron, the positron.

We will flesh out this heuristic discussion in subsequent chapters.

Marriages

To summarize, quantum field theory was born of the necessity of dealing with the marriage of special relativity and quantum mechanics, just as the new science of string theory is being born of the necessity of dealing with the marriage of general relativity and quantum mechanics.

Chapter I.2

Path Integral Formulation of Quantum Physics

The professor's nightmare: a wise guy in the class

As I noted in the preface, I know perfectly well that you are eager to dive into quantum field theory, but first we have to review the path integral formalism of quantum mechanics. This formalism is not universally taught in introductory courses on quantum mechanics, but even if you have been exposed to it, this chapter will serve as a useful review. The reason I start with the path integral formalism is that it offers a particularly convenient way of going from quantum mechanics to quantum field theory. I will first give a heuristic discussion, to be followed by a more formal mathematical treatment.

Perhaps the best way to introduce the path integral formalism is by telling a story, certainly apocryphal as many physics stories are. Long ago, in a quantum mechanics class, the professor droned on and on about the double-slit experiment, giving the standard treatment. A particle emitted from a source S (Fig. I.2.1) at time $t = 0$ passes through one or the other of two holes, A_1 and A_2, drilled in a screen and is detected at time $t = T$ by a detector located at O. The amplitude for detection is given by a fundamental postulate of quantum mechanics, the superposition principle, as the sum of the amplitude for the particle to propagate from the source S through the hole A_1 and then onward to the point O and the amplitude for the particle to propagate from the source S through the hole A_2 and then onward to the point O.

Suddenly, a very bright student, let us call him Feynman, asked, "Professor, what if we drill a third hole in the screen?" The professor replied, "Clearly, the amplitude for the particle to be detected at the point O is now given by the sum of three amplitudes, the amplitude for the particle to propagate from the source S through the hole A_1 and then onward to the point O, the amplitude for the particle to propagate from the source S through the hole A_2 and then onward to the point O, and the amplitude for the particle to propagate from the source S through the hole A_3 and then onward to the point O."

The professor was just about ready to continue when Feynman interjected again, "What if I drill a fourth and a fifth hole in the screen?" Now the professor is visibly

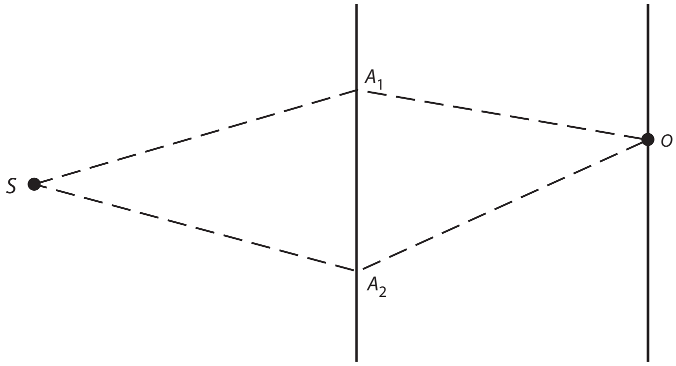


Figure I.2.1

losing his patience: "All right, wise guy, I think it is obvious to the whole class that we just sum over all the holes."

To make what the professor said precise, denote the amplitude for the particle to propagate from the source S through the hole A_i and then onward to the point O as $\mathcal{A}(S \to A_i \to O)$. Then the amplitude for the particle to be detected at the point O is

$$\mathcal{A}(\text{detected at } O) = \sum_i \mathcal{A}(S \to A_i \to O) \tag{1}$$

But Feynman persisted, "What if we now add another screen (Fig. I.2.2) with some holes drilled in it?" The professor was really losing his patience: "Look, can't you see that you just take the amplitude to go from the source S to the hole A_i in the first screen, then to the hole B_j in the second screen, then to the detector at O, and then sum over all i and j?"

Feynman continued to pester, "What if I put in a third screen, a fourth screen, eh? What if I put in a screen and drill an infinite number of holes in it so that the

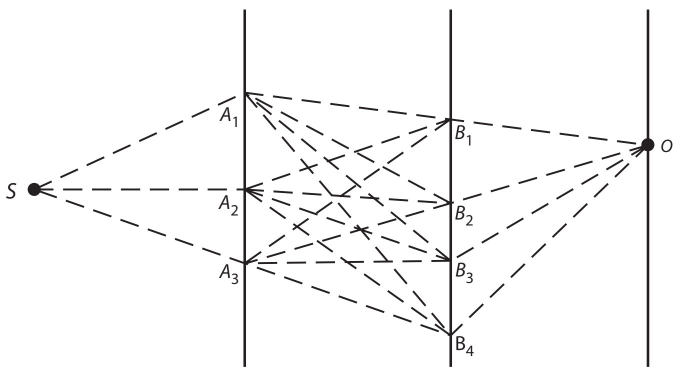


Figure I.2.2

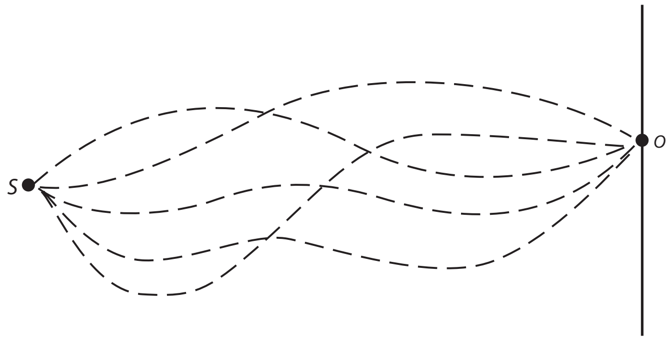


Figure I.2.3

screen is no longer there?" The professor sighed, "Let's move on; there is a lot of material to cover in this course."

But dear reader, surely you see what that wise guy Feynman was driving at. I especially enjoy his observation that if you put in a screen and drill an infinite number of holes in it, then that screen is not really there. Very Zen! What Feynman showed is that even if there were just empty space between the source and the detector, the amplitude for the particle to propagate from the source to the detector is the sum of the amplitudes for the particle to go through each one of the holes in each one of the (nonexistent) screens. In other words, we have to sum over the amplitude for the particle to propagate from the source to the detector following all possible paths between the source and the detector (Fig. I.2.3).

$$\mathcal{A}(\text{particle to go from } S \text{ to } O \text{ in time } T) =$$
$$\mathop{}_{\text{(paths)}} \mathcal{A}\left(\text{particle to go from } S \text{ to } O \text{ in time } T \text{ following a particular path}\right) \quad (2)$$

Now the mathematically rigorous will surely get anxious over how $\sum_{\text{(paths)}}$ is to be defined. Feynman followed Newton and Leibniz: Take a path (Fig. I.2.4), approximate it by straight line segments, and let the segments go to zero. You can see that this is just like filling up a space with screens spaced infinitesimally close to each other, with an infinite number of holes drilled in each screen.

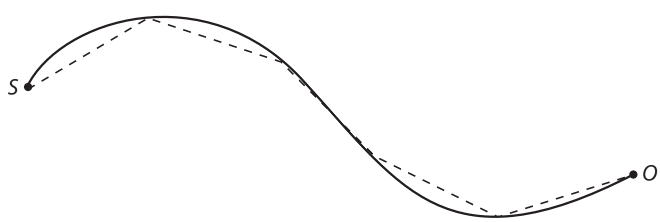


Figure I.2.4

Fine, but how to construct the amplitude $\mathcal{A}$(particle to go from S to O in time T following a particular path)? Well, we can use the unitarity of quantum mechanics: If we know the amplitude for each infinitesimal segment, then we just multiply them together to get the amplitude of the whole path.

In quantum mechanics, the amplitude to propagate from a point q_I to a point q_F in time T is governed by the unitary operator e^{-iHT}, where H is the Hamiltonian. More precisely, denoting by $|q\rangle$ the state in which the particle is at q, the amplitude in question is just $\langle q_F| e^{-iHT} |q_I\rangle$. Here we are using the Dirac bra and ket notation. Of course, philosophically, you can argue that to say the amplitude is $\langle q_F| e^{-iHT} |q_I\rangle$ amounts to a postulate and a definition of H. It is then up to experimentalists to discover that H is hermitean, has the form of the classical Hamiltonian, et cetera.

Indeed, the whole path integral formalism could be written down mathematically starting with the quantity $\langle q_F| e^{-iHT} |q_I\rangle$, without any of Feynman's jive about screens with an infinite number of holes. Many physicists would prefer a mathematical treatment without the talk. As a matter of fact, the path integral formalism was invented by Dirac precisely in this way, long before Feynman.

A necessary word about notation even though it interrupts the narrative flow: We denote the coordinates transverse to the axis connecting the source to the detector by q, rather than x, for a reason which will emerge in a later chapter. For notational simplicity, we will think of q as 1-dimensional and suppress the coordinate along the axis connecting the source to the detector.

Dirac's formulation

Let us divide the time T into N segments each lasting $\delta t = T/N$. Then we write

$$\langle q_F| e^{-iHT} |q_I\rangle = \langle q_F| e^{-iH\delta t} e^{-iH\delta t} \cdots e^{-iH\delta t} |q_I\rangle$$

Now use the fact that $|q\rangle$ forms a complete set of states so that $\int dq\, |q\rangle\langle q| = 1$. Insert 1 between all these factors of $e^{-iH\delta t}$ and write

$$\begin{aligned}&\langle q_F| e^{-iHT} |q_I\rangle \\ &= \left(\prod_{j=1}^{N-1} \int dq_j\right) \langle q_F| e^{-iH\delta t} |q_{N-1}\rangle\langle q_{N-1}| e^{-iH\delta t} |q_{N-2}\rangle \cdots \\ &\cdots \langle q_2| e^{-iH\delta t} |q_1\rangle\langle q_1| e^{-iH\delta t} |q_I\rangle \end{aligned} \tag{3}$$

Focus on an individual factor $\langle q_{j+1}| e^{-iH\delta t} |q_j\rangle$. Let us take the baby step of first evaluating it just for the free-particle case in which $H = \hat{p}^2/2m$. The hat on $\hat{p}$ reminds us that it is an operator. Denote by $|p\rangle$ the eigenstate of $\hat{p}$, namely $\hat{p}\,|p\rangle = p\,|p\rangle$. Do you remember from your course in quantum mechanics that $\langle q|p\rangle = e^{ipq}$? Sure you do. This just says that the momentum eigenstate is a plane wave in the coordinate representation. (The normalization is such that $\int (dp/2\pi)\, |p\rangle\langle p| = 1$.) So again inserting a complete set of states, we write

$$\begin{aligned}\langle q_{j+1}|\, e^{-i\delta t(\hat{p}^2/2m)}\,|q_j\rangle &= \int \frac{dp}{2\pi}\langle q_{j+1}|\, e^{-i\delta t(\hat{p}^2/2m)}\,|p\rangle\langle p|q_j\rangle \\ &= \int \frac{dp}{2\pi} e^{-i\delta t(p^2/2m)}\langle q_{j+1}|p\rangle\langle p|q_j\rangle \\ &= \int \frac{dp}{2\pi} e^{-i\delta t(p^2/2m)} e^{ip(q_{j+1}-q_j)}\end{aligned}$$

Note that we removed the hat from the momentum operator in the exponential: Since the momentum operator is acting on an eigenstate, it can be replaced by its eigenvalue.

The integral over p is known as a Gaussian integral, with which you may already be familiar. If not, turn to Appendix 1 to this chapter.

Doing the integral over p, we get

$$\begin{aligned}\langle q_{j+1}|\, e^{-i\delta t(\hat{p}^2/2m)}\,|q_j\rangle &= \left(\frac{-i2\pi m}{\delta t}\right)^{\frac{1}{2}} e^{[im(q_{j+1}-q_j)^2]/2\delta t} \\ &= \left(\frac{-i2\pi m}{\delta t}\right)^{\frac{1}{2}} e^{i\delta t(m/2)[(q_{j+1}-q_j)/\delta t]^2}\end{aligned}$$

Putting this into (3) yields

$$\langle q_F|\, e^{-iHT}\,|q_I\rangle = \left(\frac{-i2\pi m}{\delta t}\right)^{\frac{N}{2}} \int \prod_{j=0}^{N-1} dq_j e^{i\delta t(m/2)\Sigma_{j=0}^{N-1}[(q_{j+1}-q_j)/\delta t]^2}$$

with $q_0 \equiv q_I$ and $q_N \equiv q_F$.

We can now go to the continuum limit $\delta t \to 0$. Newton and Leibniz taught us to replace $[(q_{j+1}-q_j)/\delta t]^2$ by $\dot{q}^2$, and $\delta t \sum_{j=0}^{N-1}$ by $\int_0^T dt$. Finally, we define the integral over paths as

$$\int Dq(t) = \lim_{N\to\infty}\left(\frac{-i2\pi m}{\delta t}\right)^{\frac{N}{2}} \int \prod_{j=0}^{N-1} dq_j .$$

We thus obtain the path integral representation

$$\langle q_F|\, e^{-iHT}\,|q_I\rangle = \int Dq(t)\, e^{i\int_0^T dt \frac{1}{2}m\dot{q}^2} \tag{4}$$

This fundamental result tells us that to obtain $\langle q_F|\, e^{-iHT}\,|q_I\rangle$ we simply integrate over all possible paths $q(t)$ such that $q(0) = q_I$ and $q(T) = q_F$.

As an exercise you should convince yourself that had we started with the Hamiltonian for a particle in a potential $H = \hat{p}^2/2m + V(\hat{q})$ (again the hat on $\hat{q}$ indicates an operator) the final result would have been

$$\langle q_F|\, e^{-iHT}\,|q_I\rangle = \int Dq(t)\, e^{i\int_0^T dt[\frac{1}{2}m\dot{q}^2 - V(q)]} \tag{5}$$

We recognize the quantity $\frac{1}{2}m\dot{q}^2 - V(q)$ as just the Lagrangian $L(\dot{q}, q)$. The Lagrangian has emerged naturally from the Hamiltonian. In general, we have

$$\langle q_F | e^{-iHT} | q_I \rangle = \int Dq(t)\, e^{i\int_0^T dt L(\dot{q},q)} \tag{6}$$

To avoid potential confusion, let me be clear that t appears as an integration variable in the exponential on the right-hand side. The appearance of t in the path integral measure $Dq(t)$ is simply to remind us that q is a function of t (as if we need reminding). Indeed, this measure will often be abbreviated to Dq. You might recall that $\int_0^T dt L(\dot{q}, q)$ is called the action $S(q)$ in classical mechanics. The action S is a functional of the function $q(t)$.

Often, instead of specifying that the particle starts at an initial position q_I and ends at a final position q_F, we prefer to specify that the particle starts in some initial state I and ends in some final state F. Then we are interested in calculating $\langle F | e^{-iHT} | I \rangle$, which upon inserting complete sets of states can be written as

$$\int dq_F \int dq_I \langle F | q_F \rangle \langle q_F | e^{-iHT} | q_I \rangle \langle q_I | I \rangle,$$

which mixing Schrödinger and Dirac notation we can write as

$$\int dq_F \int dq_I \Psi_F(q_F)^* \langle q_F | e^{-iHT} | q_I \rangle \Psi_I(q_I).$$

In most cases we are interested in taking $|I\rangle$ and $|F\rangle$ as the ground state, which we will denote by $|0\rangle$. It is conventional to give the amplitude $\langle 0 | e^{-iHT} | 0 \rangle$ the name Z.

At the level of mathematical rigor we are working with, we count on the path integral $\int Dq(t)\, e^{i\int_0^T dt[\frac{1}{2}m\dot{q}^2 - V(q)]}$ to converge because the oscillatory phase factors from different paths tend to cancel out. It is somewhat more rigorous to perform a so-called Wick rotation to Euclidean time. This amounts to substituting $t \to -it$ and rotating the integration contour in the complex t plane so that the integral becomes

$$Z = \int Dq(t)\, e^{-\int_0^T dt[\frac{1}{2}m\dot{q}^2 + V(q)]}, \tag{7}$$

known as the Euclidean path integral. As is done in Appendix 1 to this chapter with ordinary integrals we will always assume that we can make this type of substitution with impunity.

One particularly nice feature of the path integral formalism is that the classical limit of quantum mechanics can be recovered easily. We simply restore Planck's constant $\hbar$ in (6):

$$\langle q_F | e^{-(i/\hbar)HT} | q_I \rangle = \int Dq(t)\, e^{(i/\hbar)\int_0^T dt L(\dot{q},q)}$$

and take the $\hbar \to 0$ limit. Applying the stationary phase or steepest descent method (if you don't know it see Appendix 2 to this chapter) we obtain $e^{(i/\hbar)\int_0^T dt L(\dot{q}_c, q_c)}$, where $q_c(t)$ is the "classical path" determined by solving the Euler-Lagrange equation $(d/dt)(\delta L/\delta \dot{q}) - (\delta L/\delta q) = 0$ with appropriate boundary conditions.

Appendix 1

I will now show you how to do the integral $G \equiv \int_{-\infty}^{+\infty} dx e^{-\frac{1}{2}x^2}$. The trick is to square the integral, call the dummy integration variable in one of the integrals y, and then pass to polar coordinates:

$$G^2 = \int_{-\infty}^{+\infty} dx\, e^{-\frac{1}{2}x^2} \int_{-\infty}^{+\infty} dy\, e^{-\frac{1}{2}y^2} = 2\pi \int_0^{+\infty} dr\, r e^{-\frac{1}{2}r^2}$$

$$= 2\pi \int_0^{+\infty} dw\, e^{-w} = 2\pi$$

Thus, we obtain

$$\int_{-\infty}^{+\infty} dx\, e^{-\frac{1}{2}x^2} = \sqrt{2\pi} \tag{8}$$

Believe it or not, a significant fraction of the theoretical physics literature consists of performing variations and elaborations of this basic Gaussian integral. The simplest extension is almost immediate:

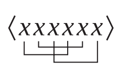
$$\int_{-\infty}^{+\infty} dx\, e^{-\frac{1}{2}ax^2} = \left(\frac{2\pi}{a}\right)^{\frac{1}{2}} \tag{9}$$

as can be seen by scaling $x \to x/\sqrt{a}$.

Acting on this repeatedly with $-2(d/da)$ we obtain

$$\langle x^{2n} \rangle \equiv \frac{\int_{-\infty}^{+\infty} dx\, e^{-\frac{1}{2}ax^2} x^{2n}}{\int_{-\infty}^{+\infty} dx\, e^{-\frac{1}{2}ax^2}} = \frac{1}{a^n}(2n-1)(2n-3)\cdots 5\cdot 3\cdot 1 \tag{10}$$

The factor $1/a^n$ follows from dimensional analysis. To remember the factor $(2n-1)!! \equiv (2n-1)(2n-3)\cdots 5\cdot 3\cdot 1$ imagine $2n$ points and connect them in pairs. The first point can be connected to one of $(2n-1)$ points, the second point can now be connected to one of the remaining $(2n-3)$ points, and so on. This clever observation, due to Gian Carlo Wick, is known as Wick's theorem in the field theory literature. Incidentally, field theorists use the following graphical mnemonic in calculating, for example, $\langle x^6 \rangle$: Write $\langle x^6 \rangle$ as $\langle xxxxxx \rangle$ and connect the x's, for example



The pattern of connection is known as a Wick contraction. In this simple example, since the six x's are identical, any one of the distinct Wick contractions gives the same value

a^{-3} and the final result for $\langle x^6 \rangle$ is just a^{-3} times the number of distinct Wick contractions, namely $5 \cdot 3 \cdot 1 = 15$. We will soon come to a less trivial example, in which we have distinct x's, in which case distinct Wick contraction gives distinct values.

An important variant is the integral

$$\int_{-\infty}^{+\infty} dx\, e^{-\frac{1}{2}ax^2+Jx} = \left(\frac{2\pi}{a}\right)^{\frac{1}{2}} e^{J^2/2a} \tag{11}$$

To see this, take the expression in the exponent and "complete the square": $-ax^2/2 + Jx = -(a/2)(x^2 - 2Jx/a) = -(a/2)(x - J/a)^2 + J^2/2a$. The x integral can now be done by shifting $x \to x + J/a$, giving the factor of $(2\pi/a)^{\frac{1}{2}}$. Check that we can also obtain (10) by differentiating with respect to J repeatedly and then setting $J = 0$.

Another important variant is obtained by replacing J by iJ:

$$\int_{-\infty}^{+\infty} dx\, e^{-\frac{1}{2}ax^2+iJx} = \left(\frac{2\pi}{a}\right)^{\frac{1}{2}} e^{-J^2/2a} \tag{12}$$

To get yet another variant, replace a by $-ia$:

$$\int_{-\infty}^{+\infty} dx\, e^{\frac{1}{2}iax^2+iJx} = \left(\frac{2\pi i}{a}\right)^{\frac{1}{2}} e^{-iJ^2/2a} \tag{13}$$

Let us promote a to a real symmetric N by N matrix A_{ij} and x to a vector x_i $(i, j = 1, \cdots, N)$. Then (11) generalizes to

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty} dx_1 dx_2 \cdots dx_N\, e^{-\frac{1}{2}x\cdot A\cdot x+J\cdot x} = \left(\frac{(2\pi)^N}{\det[A]}\right)^{\frac{1}{2}} e^{\frac{1}{2}J\cdot A^{-1}\cdot J} \tag{14}$$

where $x \cdot A \cdot x = x_i A_{ij} x_j$ and $J \cdot x = J_i x_i$ (with repeated indices summed.) To see this, diagonalize A by an orthogonal transformation O: $A = O^{-1} \cdot D \cdot O$ where D is a diagonal matrix. Call $y_i = O_{ij} x_j$. In other words, we rotate the coordinates in the N dimensional Euclidean space over which we are integrating. Using

$$\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty} dx_1 \cdots dx_N = \int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty} dy_1 \cdots dy_N$$

we factorize the left-hand side of (14) into a product of N integrals of the form in (11). The result can then be expressed in terms of D^{-1}, which we write as $O \cdot A^{-1} \cdot O^{-1}$. (To make sure you got it, try this explicitly for $N = 2$.)

Putting in some i's $(A \to -iA,\ J \to iJ)$, we find the generalization of (13)

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty} dx_1 dx_2 \cdots dx_N\, e^{(i/2)x\cdot A\cdot x+iJ\cdot x}$$
$$= \left(\frac{(2\pi i)^N}{\det[A]}\right)^{\frac{1}{2}} e^{-(i/2)J\cdot A^{-1}\cdot J} \tag{15}$$

The generalization of (10) is also easy to obtain. We differentiate (14) with respect to J repeatedly and then setting $J \to 0$. We find

$$\langle x_i x_j \cdots x_k x_l \rangle = \sum_{\text{Wick}} (A^{-1})_{ab} \cdots (A^{-1})_{cd} \tag{16}$$

where we have defined

$$\langle x_i x_j \cdots x_k x_l \rangle = \frac{\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} dx_1 dx_2 \cdots dx_N \, e^{-\frac{1}{2} x \cdot A \cdot x} x_i x_j \cdots x_k x_l}{\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} dx_1 dx_2 \cdots dx_N \, e^{-\frac{1}{2} x \cdot A \cdot x}} \tag{17}$$

and where the set of indices $\{a, b, \cdots, c, d\}$ represent a permutation of the set of indices $\{i, j, \cdots, k, l\}$. The sum in (16) is over all such permutations or Wick contractions. It is easiest to explain (16) for a simple example $\langle x_i x_j x_k x_l \rangle$. We connect the x's in pairs (Wick contraction) and write a factor $(A^{-1})_{ab}$ if we connect x_a to x_b. Thus,

$$\langle x_i x_j x_k x_l \rangle = (A^{-1})_{ij}(A^{-1})_{kl} + (A^{-1})_{il}(A^{-1})_{jk} + (A^{-1})_{ik}(A^{-1})_{jl} \tag{18}$$

(Recall that A and thus A^{-1} are symmetric.) Note that since $\langle x_i x_j \rangle = (A^{-1})_{ij}$, the right-hand side of (16) can also be written in terms of objects such as $\langle x_i x_j \rangle$. Please work out $\langle x_i x_j x_k x_l x_m x_n \rangle$; you will become an expert on Wick contractions. Of course, (16) reduces to (10) for $N = 1$.

Perhaps you are like me and do not like to memorize anything, but some of these formulas might be worth memorizing as they appear again and again in theoretical physics (and in this book).

Appendix 2

To do an exponential integral of the form $I = \int_{-\infty}^{+\infty} dq e^{-(1/\hbar) f(q)}$ we often have to resort to the steepest-descent approximation, which I will now review for your convenience. In the limit of $\hbar$ small, the integral is dominated by the minimum of $f(q)$. Expanding $f(q) = f(a) + \frac{1}{2} f''(a)(q-a)^2 + O[(q-a)^3]$ and applying (9) we obtain

$$I = e^{-(1/\hbar) f(a)} \left(\frac{2\pi\hbar}{f''(a)} \right)^{\frac{1}{2}} e^{-O(\hbar^{\frac{1}{2}})} \tag{19}$$

For $f(q)$ a function of many variables $q_1, \ldots, q_N$ and with a minimum at $q_j = a_j$, we generalize immediately to

$$I = e^{-(1/\hbar) f(a)} \left(\frac{2\pi\hbar}{\det f''(a)} \right)^{\frac{1}{2}} e^{-O(\hbar^{\frac{1}{2}})} \tag{20}$$

Here $f''(a)$ denotes the N by N matrix with entries $[f''(a)]_{ij} \equiv (\partial^2 f / \partial q_i \partial q_j)|_{q=a}$. In many situations, we do not even need the factor involving the determinant in (20). If you can derive (20) you are well on your way to becoming a quantum field theorist!

Exercises

I.2.1. Verify (5).

I.2.2. Derive (16).

Chapter I.3

From Mattress to Field

The mattress in the continuum limit

The path integral representation

$$Z \equiv \langle 0| e^{-iHT} |0\rangle = \int Dq(t)\, e^{i\int_0^T dt[\frac{1}{2}m\dot{q}^2 - V(q)]} \tag{1}$$

which we derived for the quantum mechanics of a single particle, can be generalized almost immediately to the case of N particles with the Hamiltonian

$$H = \sum_a \frac{1}{2m_a}\hat{p}_a^2 + V(\hat{q}_1, \hat{q}_2, \cdots, \hat{q}_N). \tag{2}$$

We simply keep track mentally of the position of the particles q_a with $a = 1, 2, \cdots, N$. Going through the same steps as before, we obtain

$$Z \equiv \langle 0| e^{-iHT} |0\rangle = \int Dq(t)\, e^{iS(q)} \tag{3}$$

with the action

$$S(q) = \int_0^T dt\Big(\sum_a \tfrac{1}{2}m_a\dot{q}_a^2 - V[q_1, q_2, \cdots, q_N]\Big).$$

The potential energy $V(q_1, q_2, \ldots, q_N)$ now includes interaction energy between particles, namely terms of the form $v(q_a - q_b)$, as well as the energy due to an external potential, namely terms of the form $w(q_a)$. In particular, let us now write the path integral description of the quantum dynamics of the mattress described in Chapter I.1, with the potential

$$V(q_1, q_2, \ldots, q_N) = \sum_{ab} \frac{1}{2}k_{ab}q_a q_b + \cdots$$

We are now just a short hop and skip away from a quantum field theory! Suppose we are only interested in phenomena on length scales much greater than the lattice spacing l (see Fig. I.1.1). Mathematically, we take the continuum limit $l \to 0$. In this limit, we can replace the label a on the particles by a two-dimensional position vector $\vec{x}$, and so we write $q(t, \vec{x})$ instead of $q_a(t)$. It is traditional to replace the Latin letter q by the Greek letter φ. The function $\varphi(t, \vec{x})$ is called a field.

The kinetic energy $\sum_a \frac{1}{2} m_a \dot{q}_a^2$ now becomes $\int d^2x \frac{1}{2}\sigma(\partial\varphi/\partial t)^2$. We replace $\sum_a$ by $1/l^2 \int d^2x$ and denote the mass per unit area m_a/l^2 by σ. We take all the m_a's to be equal; otherwise σ would be a function of $\vec{x}$, the system would be inhomogeneous, and we would have a hard time writing down a Lorentz-invariant action (see later).

We next focus on the first term in $V = \sum_{ab} \frac{1}{2} k_{ab} q_a q_b + \cdots$. Write $2q_a q_b = (q_a - q_b)^2 - q_a^2 - q_b^2$. Assume for simplicity that k_{ab} connect only nearest neighbors on the lattice. For nearest-neighbor pairs $(q_a - q_b)^2 \simeq l^2(\partial\varphi/\partial x)^2 + \cdots$ in the continuum limit; the derivative is obviously taken in the direction that joins the lattice sites a and b.

Putting it together then, we have

$$
\begin{aligned}
S(q) \to S(\varphi) &\equiv \int_0^T dt \int d^2x \mathcal{L}(\varphi) \\
&= \int_0^T dt \int d^2x \frac{1}{2}\left\{\sigma\left(\frac{\partial\varphi}{\partial t}\right)^2 - \rho\left[\left(\frac{\partial\varphi}{\partial x}\right)^2 + \left(\frac{\partial\varphi}{\partial y}\right)^2\right]\right. \\
&\quad \left. -\tau\varphi^2 - \varsigma\varphi^4 + \cdots\right\}
\end{aligned}
\tag{4}
$$

where the parameters ρ and τ are determined by k_{ab} and l. The precise relations do not concern us.

Henceforth in this book, we will take the $T \to \infty$ limit so that we can integrate over all of spacetime in (4).

We can clean up a bit by writing $\rho = \sigma c^2$ and scaling $\varphi \to \varphi/\sqrt{\sigma}$, so that the combination $(\partial\varphi/\partial t)^2 - c^2[(\partial\varphi/\partial x)^2 + (\partial\varphi/\partial y)^2]$ appears in the Lagrangian. The parameter c evidently has the dimension of a velocity and defines the phase velocity of the waves on our mattress. It is interesting that Lorentz invariance, with c playing the role of the speed of light, emerges naturally.

We started with a mattress for pedagogical reasons. Of course nobody believes that the fields observed in Nature, such as the meson field or the photon field, are actually constructed of point masses tied together with springs. The modern view, which I will call Landau-Ginzburg, is that we start with the desired symmetry, say Lorentz invariance if we want to do particle physics, decide on the fields we want by specifying how they transform under the symmetry (in this case we decided on a scalar field φ), and then write down the action involving no more than two time derivatives (because we don't know how to quantize actions with more than two time derivatives).

We end up with a Lorentz-invariant action (setting $c = 1$)

$$
S = \int d^dx \left[\frac{1}{2}(\partial\varphi)^2 - \frac{1}{2}m^2\varphi^2 - \frac{g}{3!}\varphi^3 - \frac{\lambda}{4!}\varphi^4 + \cdots\right] \tag{5}
$$

where various numerical factors are put in for later convenience. The relativistic notation $(\partial\varphi)^2 \equiv \partial_\mu\varphi\partial^\mu\varphi = (\partial\varphi/\partial t)^2 - (\partial\varphi/\partial x)^2 - (\partial\varphi/\partial y)^2$ was explained in the note on convention. The dimension of spacetime, d, clearly can be any integer, even though in our mattress model it was actually 3. We often write $d = D + 1$ and speak of a $(D + 1)$-dimensional spacetime.

We see here the power of imposing a symmetry. Lorentz invariance together with the insistence that the Lagrangian involve only at most two powers of $\partial/\partial t$ immediately tells us that the Lagrangian can only have the form[1] $\mathcal{L} = \frac{1}{2}(\partial\varphi)^2 - V(\varphi)$ with V a polynomial in φ. We will have a great deal more to say about symmetry later. Here we note that, for example, we could insist that physics is symmetric under $\varphi \to -\varphi$, in which case $V(\varphi)$ would have to be an even polynomial.

Now that you know what a quantum field theory is, you realize why I used the letter q to label the position of the particle in the previous chapter and not the more common $\vec{x}$. In quantum field theory, $\vec{x}$ is a label, not a dynamical variable. The $\vec{x}$ appearing in $\varphi(t, \vec{x})$ corresponds to the label a in $q_a(t)$ in quantum mechanics. The dynamical variable in field theory is not position, but the field φ. The variable $\vec{x}$ simply specifies which field variable we are talking about. I belabor this point because upon first exposure to quantum field theory some students, used to thinking of $\vec{x}$ as a dynamical operator in quantum mechanics, are confused by its role here.

In summary, we have the table

$q \to \varphi$
$a \to \vec{x}$
$q_a(t) \to \varphi(t, \vec{x}) = \varphi(x)$
$\sum_a \to \quad d^D x$

(6)

Thus we finally have the path integral defining a scalar field theory in $d = (D + 1)$ dimensional spacetime:

$$Z = \quad D\varphi e^{i \quad d^d x(\frac{1}{2}(\partial\varphi)^2 - V(\varphi))} \tag{7}$$

Note that a $(0 + 1)$-dimensional quantum field theory is just quantum mechanics.

---

[1] Strictly speaking, a term of the form $U(\varphi)(\partial\varphi)^2$ is also possible. In quantum mechanics, a term such as $U(q)(dq/dt)^2$ in the Lagrangian would describe a particle whose mass depends on position. We will not consider such "nasty" terms until much later.

The classical limit

As I have already remarked, the path integral formalism is particularly convenient for taking the classical limit. Remembering that Planck's constant $\hbar$ has the dimension of energy multiplied by time, we see that it appears in the unitary evolution operator $e^{(-i/\hbar)HT}$. Tracing through the derivation of the path integral, we see that we have to simply divide the overall factor i by $\hbar$ to get

$$Z = \int D\varphi e^{(i/\hbar)\int d^4x\mathcal{L}(\varphi)} \tag{8}$$

In the limit with $\hbar$ much smaller than the relevant action we are considering, we can evaluate the path integral using the stationary phase (or steepest descent) approximation, as I explained in the previous chapter in the context of quantum mechanics. We simply determine the extremum of $\int d^4x\mathcal{L}(\varphi)$. According to the usual Euler-Lagrange variational procedure, this leads to the equation

$$\partial_\mu \frac{\delta\mathcal{L}}{\delta(\partial_\mu\varphi)} - \frac{\delta\mathcal{L}}{\delta\varphi} = 0 \tag{9}$$

We thus recover the classical field equation, exactly as we should, which in our scalar field theory reads

$$(\partial^2 + m^2)\varphi(x) + \frac{g}{2}\varphi(x)^2 + \frac{\lambda}{6}\varphi(x)^3 + \cdots = 0 \tag{10}$$

The vacuum

In the point particle quantum mechanics discussed in Chapter I.2 we wrote the path integral for $\langle F|e^{iHT}|I\rangle$, with some initial and final state, which we can choose at our pleasure. A convenient and particularly natural choice would be to take $|I\rangle = |F\rangle$ to be the ground state. In quantum field theory what should we choose for the initial and final states? A standard choice for the initial and final states is the ground state or the vacuum state of the system, denoted by $|0\rangle$, in which, speaking colloquially, nothing is happening. In other words, we would calculate the quantum transition amplitude from the vacuum to the vacuum, which would enable us to determine the energy of the ground state. But this is not a particularly interesting quantity, because in quantum field theory we would like to measure all energies relative to the vacuum and so, by convention, would set the energy of the vacuum to zero (possibly by having to subtract an infinite constant from the Lagrangian). Incidentally, the vacuum in quantum field theory is a stormy sea of quantum fluctuations, but for this initial pass at quantum field theory, we will not examine it in any detail. We will certainly come back to the vacuum in later chapters.

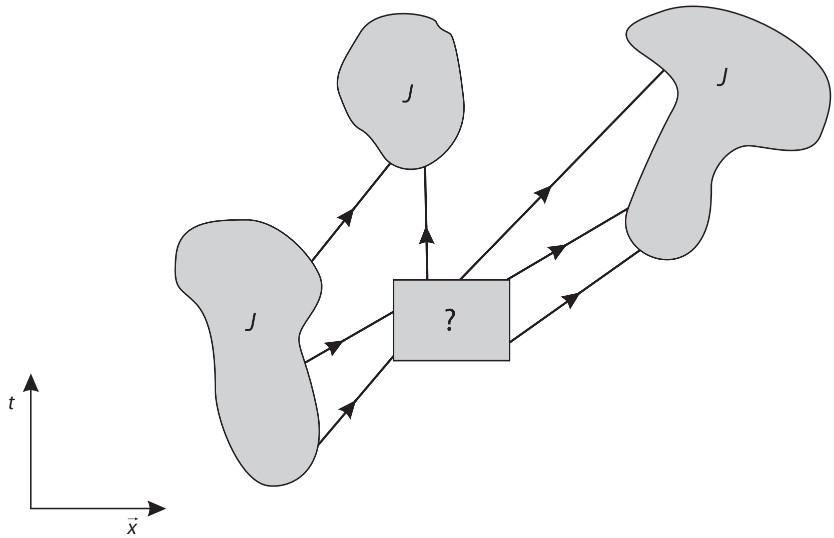


Figure I.3.1

Disturbing the vacuum

We would like to do something more exciting than watching a boiling sea of quantum fluctuations. We would like to disturb the vacuum. Somewhere in space, at some instant in time, we would like to create a particle, watch it propagate for a while, and then annihilate it somewhere else in space, at some later instant in time. In other words, we want to create a source and a sink (sometimes referred to collectively as sources) at which particles can be created and annihilated.

To see how to do this, let us go back to the mattress. Bounce up and down on it to create some excitations. Obviously, pushing on the mass labeled by a in the mattress corresponds to adding a term such as $J_a(t)q_a$ to the potential $V(q_1, q_2, \cdots, q_N)$. More generally, we can add $\sum_a J_a(t)q_a$. When we go to field theory this added term gets promoted to $J(x)\varphi(x)$ in the field theory Lagrangian, according to the promotion table (6).

This so-called source function $J(t, \vec{x})$ describes how the mattress is being disturbed. We can choose whatever function we like, corresponding to our freedom to push on the mattress wherever and whenever we like. In particular, $J(x)$ can vanish everywhere in spacetime except in some localized regions.

By bouncing up and down on the mattress we can get wave packets going off here and there (Fig. I.3.1). This corresponds precisely to sources (and sinks) for particles. Thus, we really want the path integral

$$Z = D\varphi e^{i \, d^4x[\frac{1}{2}(\partial\varphi)^2 - V(\varphi) + J(x)\varphi(x)]} \qquad (11)$$

Free field theory

The functional integral in (11) is impossible to do except when

$$\mathcal{L}(\varphi) = \tfrac{1}{2}[(\partial\varphi)^2 - m^2\varphi^2] \qquad (12)$$

The corresponding theory is called the free or Gaussian theory. The equation of motion (9) works out to be $(\partial^2 + m^2)\varphi = 0$, known as the Klein-Gordon equation.[2] Being linear, it can be solved immediately to give $\varphi(\vec{x}, t) = e^{i(\omega t - \vec{k}\cdot\vec{x})}$ with

$$\omega^2 = \vec{k}^2 + m^2 \qquad (13)$$

In the natural units we are using, $\hbar = 1$ and so frequency ω is the same as energy $\hbar\omega$ and wave vector $\vec{k}$ is the same as momentum $\hbar\vec{k}$. Thus, we recognize (13) as the energy-momentum relation for a particle of mass m, namely the sophisticate's version of the layperson's $E = mc^2$. We expect this field theory to describe a relativistic particle of mass m.

Let us now evaluate (11) in this special case:

$$Z = \quad D\varphi e^{i \quad d^4x\{\frac{1}{2}[(\partial\varphi)^2 - m^2\varphi^2] + J\varphi\}} \qquad (14)$$

Integrating by parts under the $\quad d^4x$ and not worrying about the possible contribution of boundary terms at infinity (we implicitly assume that the fields we are integrating over fall off sufficiently rapidly), we write

$$Z = \quad D\varphi e^{i \quad d^4x[-\frac{1}{2}\varphi(\partial^2 + m^2)\varphi + J\varphi]} \qquad (15)$$

You will encounter functional integrals like this again and again in your study of field theory. The trick is to imagine discretizing spacetime. You don't actually have to do it: Just imagine doing it. Let me sketch how this goes. Replace the function $\varphi(x)$ by the vector $\varphi_i = \varphi(ia)$ with i an integer and a the lattice spacing. (For simplicity, I am writing things as if we were in 1-dimensional spacetime. More generally, just let the index i enumerate the lattice points in some way.) Then differential operators become matrices. For example, $\partial\varphi(ia) \rightarrow (1/a)$ $(\varphi_{i+1} - \varphi_i) \equiv \sum_j M_{ij}\varphi_j$, with some appropriate matrix M. Integrals become sums. For example, $\quad d^4x\, J(x)\varphi(x) \rightarrow a^4 \sum_i J_i\varphi_i$.

Now, lo and behold, the integral (15) is just the integral we did in (I.2.15)

[2] The Klein-Gordon equation was actually discovered by Schrödinger before he found the equation that now bears his name. Later, in 1926, it was written down independently by Klein, Gordon, Fock, Kudar, de Donder, and Van Dungen.

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty} dq_1 dq_2 \cdots dq_N\, e^{(i/2)q\cdot A\cdot q + iJ\cdot q}$$

$$= \left(\frac{(2\pi i)^N}{\det[A]}\right)^{\frac{1}{2}} e^{-(i/2)J\cdot A^{-1}\cdot J} \tag{16}$$

The role of A in (16) is played in (15) by the differential operator $-(\partial^2 + m^2)$. The defining equation for the inverse $A \cdot A^{-1} = I$ or $A_{ij}A^{-1}_{jk} = \delta_{ik}$ becomes in the continuum limit

$$-(\partial^2 + m^2)D(x - y) = \delta^{(4)}(x - y) \tag{17}$$

We denote the continuum limit of A^{-1}_{jk} by $D(x - y)$ (which we know must be a function of $x - y$, and not of x and y separately, since no point in spacetime is special). Note that in going from the lattice to the continuum Kronecker is replaced by Dirac. It is very useful to be able to go back and forth mentally between the lattice and the continuum.

Our final result is

$$Z(J) = \mathcal{C}e^{-(i/2)\int\int d^4x d^4y J(x)D(x-y)J(y)} \equiv \mathcal{C}e^{iW(J)} \tag{18}$$

with $D(x)$ determined by solving (17). The overall factor $\mathcal{C}$, which corresponds to the overall factor with the determinant in (16), does not depend on J and, as will become clear in the discussion to follow, is often of no interest to us. As a rule I will omit writing $\mathcal{C}$ altogether. Clearly, $\mathcal{C} = Z(J = 0)$ so that $W(J)$ is defined by

$$Z(J) \equiv Z(J = 0)e^{iW(J)} \tag{19}$$

Observe that

$$W(J) = -\frac{1}{2}\int\int d^4x d^4y J(x)D(x - y)J(y) \tag{20}$$

is a simple quadratic functional of J. In contrast, $Z(J)$ depends on arbitrarily high powers of J. This fact will be of importance in Chapter I.7.

Free propagator

The function $D(x)$, known as the propagator, plays an essential role in quantum field theory. As the inverse of a differential operator it is clearly closely related to the Green's function you encountered in a course on electromagnetism.

Physicists are sloppy about mathematical rigor, but even so, they have to be careful once in a while to make sure that what they are doing actually makes sense. For the integral in (15) to converge for large φ we replace $m^2 \to m^2 - i\varepsilon$ so that the integrand contains a factor $e^{-\varepsilon \quad d^4 x \varphi^2}$, where ε is a positive infinitesimal[3] we will let tend to zero later.

We can solve (17) easily by going to momentum space and recalling the representation of the Dirac delta function

$$\delta^{(4)}(x-y) = \quad \frac{d^4 k}{(2\pi)^4} e^{ik(x-y)} \tag{21}$$

The solution is

$$D(x-y) = \quad \frac{d^4 k}{(2\pi)^4} \frac{e^{ik(x-y)}}{k^2 - m^2 + i\varepsilon} \tag{22}$$

which you can check by plugging into (17). Note that the so-called $i\varepsilon$ prescription we just mentioned is essential; otherwise the k integral would hit a pole.

To evaluate $D(x)$ we first integrate over k^0 by the method of contours. Define $\omega_k = +\sqrt{\vec{k}^2 + m^2}$. The integrand has two poles in the complex k^0 plane, at $\pm\sqrt{\omega_k^2 - i\varepsilon}$, which in the $\varepsilon \to 0$ limit is equal to $+\omega_k - i\varepsilon$ and $-\omega_k + i\varepsilon$. For x^0 positive we can extend the integration contour that goes from $-\infty$ to $+\infty$ on the real axis to include the infinite semicircle in the upper half-plane, thus enclosing the pole at $-\omega_k + i\varepsilon$ and giving $-i \quad [d^3k/(2\pi)^3 2\omega_k] e^{i(\omega_k t - \vec{k}\cdot\vec{x})}$. For x^0 negative we close the contour in the lower half-plane. Thus

$$D(x) = -i \quad \frac{d^3 k}{(2\pi)^3 2\omega_k} [e^{-i(\omega_k t - \vec{k}\cdot\vec{x})}\theta(x^0) + e^{i(\omega_k t - \vec{k}\cdot\vec{x})}\theta(-x^0)] \tag{23}$$

Physically, $D(x)$ describes the amplitude for a disturbance in the field to propagate from the origin to x. We expect drastically different behavior depending on whether x is inside or outside the lightcone. Without evaluating the integral we can see roughly how things go. For $x = (t, 0)$ with, say, $t > 0$, $D(x) = -i \quad [d^3k/(2\pi)^3 2\omega_k] e^{-i\omega_k t}$ is a superposition of plane waves and thus we should have oscillatory behavior. In contrast, for $x^0 = 0$, we have, upon interpreting $\theta(0) = \frac{1}{2}$, $D(x) = -i \quad \left[d^3k/(2\pi)^3 2\sqrt{\vec{k}^2 + m^2}\right] e^{-i\vec{k}\cdot\vec{x}}$ and the square root cut starting at $\pm im$ leads to an exponential decay $\sim e^{-m|\vec{x}|}$, as we would expect. Classically, a particle cannot get outside the lightcone, but a quantum field can "leak" out over a distance of the order m^{-1}.

Exercises

I.3.1. Verify that $D(x)$ decays exponentially for spacelike separation.

I.3.2. Work out the propagator $D(x)$ for a free-field theory in $(1+1)$-dimensional spacetime and study the large x^1 behavior for $x^0 = 0$.

---

[3] As is customary, ε is treated as generic, so that ε multiplied by any positive number is still ε.

Chapter I.4

From Field to Particle to Force

From field to particle

In the previous chapter we obtained for the free theory

$$W(J) = -\frac{1}{2} \qquad d^4x d^4y J(x) D(x-y) J(y) \tag{1}$$

which we now write in terms of the Fourier transform $J(k) \equiv \quad d^4x e^{-ikx} J(x)$:

$$W(J) = -\frac{1}{2} \qquad \frac{d^4k}{(2\pi)^4} J(k)^* \frac{1}{k^2 - m^2 + i\varepsilon} J(k) \tag{2}$$

[Note that $J(k)^* = J(-k)$ for $J(x)$ real.]

We can jump up and down on the mattress any way we like. In other words, we can choose any $J(x)$ we want, and by exploiting this freedom of choice, we can extract a remarkable amount of physics.

Consider $J(x) = J_1(x) + J_2(x)$, where $J_1(x)$ and $J_2(x)$ are concentrated in two local regions 1 and 2 in spacetime (Fig. I.4.1). Then $W(J)$ contains four terms, of the form $J_1^* J_1$, $J_2^* J_2$, $J_1^* J_2$, and $J_2^* J_1$. Let us focus on the last two of these terms, one of which reads

$$W(J) = -\frac{1}{2} \qquad \frac{d^4k}{(2\pi)^4} J_2(k)^* \frac{1}{k^2 - m^2 + i\varepsilon} J_1(k) \tag{3}$$

We see that $W(J)$ is large only if $J_1(x)$ and $J_2(x)$ overlap significantly in their Fourier transform and if in the region of overlap in momentum space $k^2 - m^2$ almost vanishes. There is a "resonance type" spike at $k^2 = m^2$, that is, if the energy-momentum relation of a particle of mass m is satisfied. (We will use the language of the relativistic physicist, writing "momentum space" for energy-momentum space, and lapse into nonrelativistic language only when the context demands it, such as in "energy-momentum relation.")

We thus interpret the physics contained in our simple field theory as follows: In region 1 in spacetime there exists a source that sends out a "disturbance in the field," which is later absorbed by a sink in region 2 in spacetime. Experimentalists choose to call this disturbance in the field a particle of mass m. Our expectation

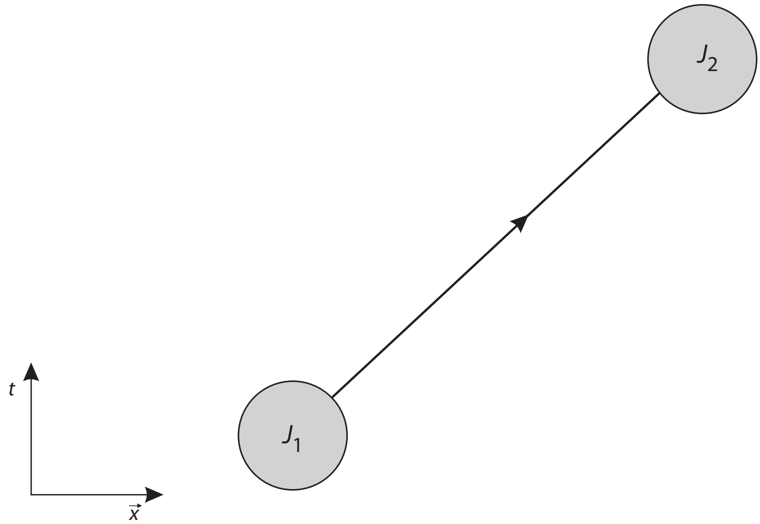


Figure I.4.1

based on the equation of motion that the theory contains a particle of mass m is fulfilled.

A bit of jargon: When $k^2 = m^2$, k is said to be on mass shell. Note, however, that in (3) we integrate over all k, including values of k far from the mass shell. For arbitrary k, it is a linguistic convenience to say that a "virtual particle" of momentum k propagates from the source to the sink.

From particle to force

We can now go on to consider other possibilities for $J(x)$ (which we will refer to generically as sources), for example, $J(x) = J_1(x) + J_2(x)$, where $J_a(x) = \delta^{(3)}(\vec{x} - \vec{x}_a)$. In other words, $J(x)$ is a sum of sources that are time-independent infinitely sharp spikes located at $\vec{x}_1$ and $\vec{x}_2$ in space. (If you like more mathematical rigor than is offered here, you are welcome to replace the delta function by lumpy functions peaking at $\vec{x}_a$. You would simply clutter up the formulas without gaining much.) More picturesquely, we are describing two massive lumps sitting at $\vec{x}_1$ and $\vec{x}_2$ on the mattress and not moving at all [no time dependence in $J(x)$].

What do the quantum fluctuations in the field φ, that is, the vibrations in the mattress, do to the two lumps sitting on the mattress? If you expect an attraction between the two lumps, you are quite right.

As before, $W(J)$ contains four terms. We neglect the "self-interaction" term $J_1 J_1$ since this contribution would be present in W regardless of whether J_2 is present or not. We want to study the interaction between the two "massive lumps" represented by J_1 and J_2. Similarly we neglect $J_2 J_2$.

Plugging into (1) and doing the integral over d^3x and d^3y we immediately obtain

$$W(J) = - \qquad dx^0 dy^0 \qquad \frac{dk^0}{2\pi} e^{ik^0(x-y)^0} \qquad \frac{d^3k}{(2\pi)^3} \frac{e^{i\vec{k}\cdot(\vec{x}_1-\vec{x}_2)}}{k^2 - m^2 + i\varepsilon} \tag{4}$$

(The factor 2 comes from the two terms $J_2 J_1$ and $J_1 J_2$.) Integrating over y^0 we get a delta function setting k^0 to zero (so that k is certainly not on mass shell, to throw the jargon around a bit). Thus we are left with

$$W(J) = \qquad dx^0 \qquad \frac{d^3k}{(2\pi)^3} \frac{e^{i\vec{k}\cdot(\vec{x}_1-\vec{x}_2)}}{\vec{k}^2 + m^2} \tag{5}$$

Note that the infinitesimal $i\varepsilon$ can be dropped since the denominator $\vec{k}^2 + m^2$ is always positive.

The factor $(\quad dx^0)$ should have filled us with fear and trepidation: an integral over time, it seems to be infinite. Fear not! Recall that in the path integral formalism $Z = \mathcal{C}\, e^{iW(J)}$ represents $\langle 0|\, e^{-iHT}\, |0\rangle = e^{-iET}$, where E is the energy due to the presence of the two sources acting on each other. The factor $(\quad dx^0)$ produces precisely the time interval T. All is well. Setting $iW = iET$ we obtain from (5)

$$E = - \qquad \frac{d^3k}{(2\pi)^3} \frac{e^{i\vec{k}\cdot(\vec{x}_1-\vec{x}_2)}}{\vec{k}^2 + m^2} \tag{6}$$

This energy is negative! The presence of two delta function sources, at $\vec{x}_1$ and $\vec{x}_2$, has lowered the energy. In other words, the two sources attract each other by virtue of their coupling to the field φ . We have derived our first physical result in quantum field theory!

We identify E as the potential energy between two static sources. Even without doing the integral we see that as the separation $\vec{x}_1 - \vec{x}_2$ between the two sources becomes large, the oscillatory exponential cuts off the integral. The characteristic distance is the inverse of the characteristic value of k, which is m. Thus, we expect the attraction between the two sources to decrease rapidly to zero over the distance $1/m$.

The range of the attractive force generated by the field φ is determined inversely by the mass m of the particle described by the field. Got that?

The integral is done in the appendix to this chapter and gives

$$E = -\frac{1}{4\pi r} e^{-mr} \tag{7}$$

The result is as we expected: The potential drops off exponentially over the distance scale $1/m$. Obviously, $dE/dr > 0$: The two massive lumps sitting on the mattress can lower the energy by getting closer to each other.

What we have derived was one of the most celebrated results in twentieth-century physics. Yukawa proposed that the attraction between nucleons in the atomic nucleus is due to their coupling to a field like the φ field described here. The known range of the nuclear force enabled him to predict not only the existence of the particle associated with this field, now called the π meson[1] or the pion, but its mass as well. As you probably know, the pion was eventually discovered with essentially the properties predicted by Yukawa.

Origin of force

That the exchange of a particle can produce a force was one of the most profound conceptual advances in physics. We now associate a particle with each of the known forces: for example, the photon with the electromagnetic force and the graviton with the gravitational force; the former is experimentally well established and the latter while it has not yet been detected experimentally hardly anyone doubts its existence. We will discuss the photon and the graviton in the next chapter, but we can already answer a question smart high school students often ask: Why do Newton's gravitational force and Coulomb's electric force both obey the $1/r^2$ law?

We see from (7) that if the mass m of the mediating particle vanishes, the force produced will obey the $1/r^2$ law. If you trace back over our derivation, you will see that this comes about from the fact that the Lagrangian density for the simplest field theory involves two powers of the spacetime derivative ∂ (since any term involving one derivative such as $\varphi\,\partial\varphi$ is not Lorentz invariant). Indeed, the power dependence of the potential follows simply from dimensional analysis: $d^3k(e^{i\vec{k}\cdot\vec{x}}/k^2) \sim 1/r$.

Connected versus disconnected

We end with a couple of formal remarks of importance to us only in Chapter I.7. First, note that we might want to draw a small picture Fig.(I.4.2) to represent the integrand $J(x)D(x-y)J(y)$ in $W(J)$: A disturbance propagates from y to x (or vice versa). In fact, this is the beginning of Feynman diagrams! Second, recall that

$$Z(J) = Z(J=0) \sum_{n=0}^{\infty} \frac{[iW(J))^n]}{n!}$$

For instance, the $n=2$ term in $Z(J)/Z(J=0)$ is given by

$$\frac{1}{2!} \quad -\frac{i}{2} \quad ^2 \qquad d^4x_1 d^4x_2 d^4x_3 d^4x_4 D(x_1-x_2)$$
$$D(x_3-x_4)J(x_1)J(x_2)J(x_3)J(x_4)$$

The integrand is graphically described in Figure I.4.3. The process is said to be disconnected: The propagation from x_1 to x_2 and the propagation from x_3 to x_4

[1] The etymology behind this word is quite interesting (A. Zee, *Fearful Symmetry*: see pp. 169 and 335 to learn, among other things, the French objection and the connection between meson and illusion).

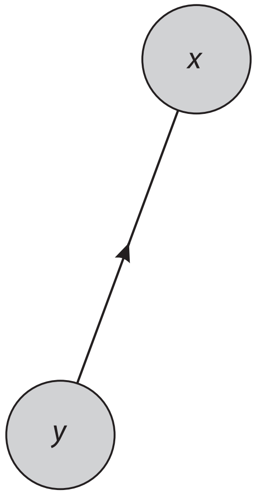


Figure I.4.2

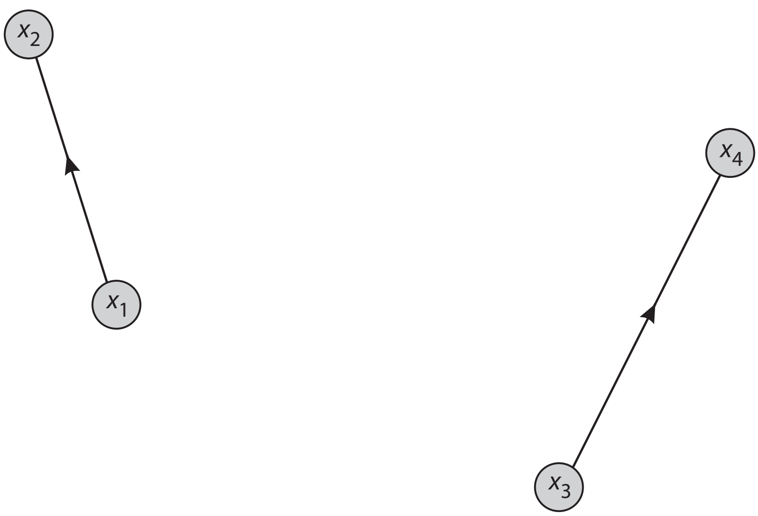


Figure I.4.3

proceed independently. We will come back to the difference between connected and disconnected in Chapter I.7.

Appendix

Writing $\vec{x} \equiv (\vec{x}_1 - \vec{x}_2)$ and $u \equiv \cos\theta$ with θ the angle between $\vec{k}$ and $\vec{x}$, we evaluate the integral in (6) spherical coordinates (with $k = |\vec{k}|$ and $r = |\vec{x}|$) to be

$$I \equiv \frac{1}{(2\pi)^2} \int_0^\infty dk\, k^2 \int_{-1}^{+1} du \frac{e^{ikru}}{k^2 + m^2} = \frac{2i}{(2\pi)^2 ir} \int_0^\infty dk\, k \frac{\sin kr}{k^2 + m^2} \tag{8}$$

Since the integrand is even, we can extend the integral and write it as

$$\frac{1}{2} \int_{-\infty}^{\infty} dk\, k \frac{\sin kr}{k^2 + m^2} = \frac{1}{2i} \int_{-\infty}^{\infty} dk\, k \frac{1}{k^2 + m^2} e^{ikr}.$$

Since r is positive, we can close the contour in the upper half-plane and pick up the pole at $+im$, obtaining $(1/2i)(2\pi i)(im/2im)e^{-mr} = (\pi/2)e^{-mr}$. Thus, $I = (1/4\pi r)e^{-mr}$.

Exercise

I.4.1. Calculate the analog of the inverse square law in a $(2 + 1)$-dimensional universe, and more generally in a $(D + 1)$-dimensional universe.

Chapter I.5

Coulomb and Newton: Repulsion and Attraction

Why like charges repel

We suggested that quantum field theory can explain both Newton's gravitational force and Coulomb's electric force naturally. Between like objects Newton's force is attractive while Coulomb's force is repulsive. Is quantum field theory "smart enough" to produce this observational fact, one of the most basic in our understanding of the physical universe? You bet!

We will first treat the quantum field theory of the electromagnetic field, known as quantum electrodynamics or QED for short. In order to avoid complications at this stage associated with gauge invariance (about which much more later) I will consider instead the field theory of a massive spin 1 meson, or vector meson. After all, experimentally all we know is an upper bound on the photon mass, which although tiny is not mathematically zero. We can adopt a pragmatic attitude: Calculate with a photon mass m and set $m = 0$ at the end, and if the result does not blow up in our faces, we will presume that it is OK.[1]

Recall Maxwell's Lagrangian for electromagnetism $\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu}$, where $F_{\mu\nu} \equiv \partial_\mu A_\nu - \partial_\nu A_\mu$ with $A_\mu(x)$ the vector potential. You can see the reason for the important overall minus sign in the Lagrangian by looking at the coefficient of $(\partial_0 A_i)^2$, which has to be positive, just like the coefficient of $(\partial_0\varphi)^2$ in the Lagrangian for the scalar field. This says simply that time variation should cost a positive amount of action.

I will now give the photon a small mass by changing the Lagrangian to $\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \frac{1}{2}m^2 A_\mu A^\mu + A_\mu J^\mu$. (The mass term is written in analogy to the mass term $m^2\varphi^2$ in the scalar field Lagrangian; we will see later that it is indeed the photon mass.) I have also added a source $J^\mu(x)$,which in this context is more familiarly known as a current. We will assume that the current is conserved so that $\partial_\mu J^\mu = 0$.

[1] When I took a field theory course as a student with Sidney Coleman this was how he treated QED in order to avoid discussing gauge invariance.

Well, you know that the field theory of our vector meson is defined by the path integral $Z = \quad DA\, e^{iS(A)} \equiv e^{iW(J)}$ with the action

$$S(A) = \quad d^4x\mathcal{L} = \quad d^4x\{\tfrac{1}{2}A_\mu[(\partial^2 + m^2)g^{\mu\nu} - \partial^\mu\partial^\nu]A_\nu + A_\mu J^\mu\} \quad (1)$$

The second equality follows upon integrating by parts [compare (I.3.15)].

By now you have learned that we simply apply (I.3.16). We merely have to find the inverse of the differential operator in the square bracket; in other words, we have to solve

$$[(\partial^2 + m^2)g^{\mu\nu} - \partial^\mu\partial^\nu]D_{\nu\lambda}(x) = \delta^\mu_\lambda \delta^{(4)}(x) \quad (2)$$

As before [compare (I.3.17)] we go to momentum space by defining

$$D_{\nu\lambda}(x) = \quad \frac{d^4k}{(2\pi)^4} D_{\nu\lambda}(k)e^{ikx}$$

Plugging in, we find that $[-(k^2 - m^2)g^{\mu\nu} + k^\mu k^\nu]D_{\nu\lambda}(k) = \delta^\mu_\lambda$, giving

$$D_{\nu\lambda}(k) = \frac{-g_{\nu\lambda} + k_\nu k_\lambda/m^2}{k^2 - m^2} \quad (3)$$

This is the photon, or more accurately the massive vector meson, propagator. Thus

$$W(J) = -\frac{1}{2} \quad \frac{d^4k}{(2\pi)^4} J^\mu(k)^* \frac{-g_{\mu\nu} + k_\mu k_\nu/m^2}{k^2 - m^2 + i\varepsilon} J^\nu(k) \quad (4)$$

Since current conservation $\partial_\mu J^\mu(x) = 0$ gets translated into momentum space as $k_\mu J^\mu(k) = 0$, we can throw away the $k_\mu k_\nu$ term in the photon propagator. The effective action simplifies to

$$W(J) = \frac{1}{2} \quad \frac{d^4k}{(2\pi)^4} J^\mu(k)^* \frac{1}{k^2 - m^2 + i\varepsilon} J_\mu(k) \quad (5)$$

No further computation is needed to obtain a profound result. Just compare this result to (I.4.2). The field theory has produced an extra sign. The potential energy between two lumps of charge density $J^0(x)$ is positive. The electromagnetic force between like charges is repulsive!

We can now safely let the photon mass m go to zero thanks to current conservation, [Note that we could not have done that in (3).] Indeed, referring to (I.4.7) we see that the potential energy between like charges is

$$E = \frac{1}{4\pi r}e^{-mr} \to \frac{1}{4\pi r} \quad (6)$$

To accommodate positive and negative charges we can simply write $J^\mu = J^\mu_p - J^\mu_n$. We see that a lump with charge density J^0_p is attracted to a lump with charge density J^0_n.

Bypassing Maxwell

Having done electromagnetism in two minutes flat let me now do gravity. Let us move on to the massive spin 2 meson field. In my treatment of the massive spin 1 meson field I took a short cut. Assuming that you are familiar with the Maxwell Lagrangian, I simply added a mass term to it and took off. But I do not feel comfortable assuming that you are equally familiar with the corresponding Lagrangian for the massless spin 2 field (the so-called linearized Einstein Lagrangian, which I will discuss in a later chapter). So here I will follow another strategy.

I invite you to think physically, and together we will arrive at the propagator for a massive spin 2 field. First, we will warm up with the massive spin 1 case.

In fact, start with something even easier: the propagator $D(k) = 1/(k^2 - m^2)$ for a massive spin 0 field. It tells us that the amplitude for the propagation of a spin 0 disturbance blows up when the disturbance is almost a real particle. The residue of the pole is a property of the particle. The propagator for a spin 1 field $D_{\nu\lambda}$ carries a pair of Lorentz indices and in fact we know what it is from (3):

$$D_{\nu\lambda}(k) = \frac{-G_{\nu\lambda}}{k^2 - m^2} \tag{7}$$

where for later convenience we have defined

$$G_{\nu\lambda}(k) \equiv g_{\nu\lambda} - \frac{k_\nu k_\lambda}{m^2} \tag{8}$$

Let us now understand the physics behind $G_{\nu\lambda}$. I expect you to remember the concept of polarization from your course on electromagnetism. A massive spin 1 particle has three degrees of polarization for the obvious reason that in its rest frame its spin vector can point in three different directions. The three polarization vectors $\varepsilon_\mu^{(a)}$ are simply the three unit vectors pointing along the x, y, and z axes, respectively ($a = 1, 2, 3$): $\varepsilon_\mu^{(1)} = (0, 1, 0, 0)$, $\varepsilon_\mu^{(2)} = (0, 0, 1, 0)$, $\varepsilon_\mu^{(3)} = (0, 0, 0, 1)$. In the rest frame $k^\mu = (m, 0, 0, 0)$ and so

$$k^\mu \varepsilon_\mu^{(a)} = 0 \tag{9}$$

Since this is a Lorentz invariant equation, it holds for a moving spin 1 particle as well. Indeed, with a suitable normalization condition this fixes the three polarization vectors $\varepsilon_\mu^{(a)}(k)$ for a particle with momentum k.

The amplitude for a particle with momentum k and polarization a to be created at the source is proportional to $\varepsilon_\lambda^{(a)}(k)$, and the amplitude for it to be absorbed at the sink is proportional to $\varepsilon_\nu^{(a)}(k)$. We multiply the amplitudes together to get the amplitude for propagation from source to sink, and then sum over the three possible polarizations. Now we understand the residue of the pole in the spin 1 propagator $D_{\nu\lambda}(k)$: It represents $\sum_a \varepsilon_\nu^{(a)}(k)\, \varepsilon_\lambda^{(a)}(k)$. To calculate this quantity, note that by Lorentz invariance it can only be a linear combination of $g_{\nu\lambda}$ and $k_\nu k_\lambda$. The condition $k^\mu \varepsilon_\mu^{(a)} = 0$ fixes the combination as $g_{\nu\lambda} - k_\nu k_\lambda / m^2$. We evaluate the left-hand side for k at rest with $\nu = \lambda = 1$, for instance, and fix the overall and all-crucial sign to be -1. Thus

$$\sum_a \varepsilon_\nu^{(a)}(k)\varepsilon_\lambda^{(a)}(k) = -G_{\nu\lambda}(k) \equiv -\left(g_{\nu\lambda} - \frac{k_\nu k_\lambda}{m^2}\right) \tag{10}$$

We have thus constructed the propagator $D_{\nu\lambda}(k)$ for a massive spin 1 particle, bypassing Maxwell. Onward to spin 2! We want to similarly bypass Einstein.

Bypassing Einstein

A massive spin 2 particle has 5 ($2 \cdot 2 + 1 = 5$, remember?) degrees of polarization, characterized by the five polarization tensors $\varepsilon_{\mu\nu}^{(a)}$ ($a = 1, 2, \cdots, 5$) symmetric in the indices μ and ν satisfying

$$k^\mu \varepsilon_{\mu\nu}^{(a)} = 0 \tag{11}$$

and the tracelessness condition

$$g^{\mu\nu}\varepsilon_{\mu\nu}^{(a)} = 0 \tag{12}$$

Let's count as a check. A symmetric Lorentz tensor has $4 \cdot 5/2 = 10$ components. The four conditions in (11) and the single condition in (12) cut the number of components down to $10 - 4 - 1 = 5$, precisely the right number. (Just to throw some jargon around, remember how to construct irreducible group representations? If not, read Appendix C.) We fix the normalization of $\varepsilon_{\mu\nu}$ by setting the positive quantity $\sum_a \varepsilon_{12}^{(a)}(k)\varepsilon_{12}^{(a)}(k) = 1$.

So, in analogy with the spin 1 case we now determine $\sum_a \varepsilon_{\mu\nu}^{(a)}(k)\varepsilon_{\lambda\sigma}^{(a)}(k)$. We have to construct this object out of $g_{\mu\nu}$ and k_μ, or equivalently $G_{\mu\nu}$ and k_μ. This quantity must be a linear combination of terms such as $G_{\mu\nu}G_{\lambda\sigma}$, $G_{\mu\nu}k_\lambda k_\sigma$, and so forth. Using (11) and (12) repeatedly (Exercise I.5.1) you will easily find that

$$\sum_a \varepsilon_{\mu\nu}^{(a)}(k)\varepsilon_{\lambda\sigma}^{(a)}(k) = (G_{\mu\lambda}G_{\nu\sigma} + G_{\mu\sigma}G_{\nu\lambda}) - \tfrac{2}{3}G_{\mu\nu}G_{\lambda\sigma} \tag{13}$$

The overall sign and proportionality constant are determined by evaluating both sides for $\mu = \lambda = 1$ and $\nu = \sigma = 2$, for instance.

Thus, we have determined the propagator for a massive spin 2 particle

$$D_{\mu\nu,\,\lambda\sigma}(k) = \frac{(G_{\mu\lambda}G_{\nu\sigma} + G_{\mu\sigma}G_{\nu\lambda}) - \tfrac{2}{3}G_{\mu\nu}G_{\lambda\sigma}}{k^2 - m^2} \tag{14}$$

Why we fall

We are now ready to understand one of the fundamental mysteries of the universe: Why masses attract.

Recall from your courses on electromagnetism and special relativity that the energy or mass density out of which mass is composed is part of a stress-energy tensor $T^{\mu\nu}$. For our purposes, in fact, all you need to remember is that it is a symmetric tensor and that the component T^{00} is the energy density.

To couple to the stress-energy tensor, we need a tensor field $\varphi_{\mu\nu}$ symmetric in its two indices. In other words, the Lagrangian of the world should contain a term like $\varphi_{\mu\nu}T^{\mu\nu}$. This is in fact how we know that the graviton, the particle responsible for gravity, has spin 2, just as we know that the photon, the particle responsible for electromagnetism and hence coupled to the current J^{μ}, has spin 1. In Einstein's theory, which we will discuss in a later chapter, $\varphi_{\mu\nu}$ is of course part of the metric tensor.

Just as we pretended that the photon has a small mass to avoid having to discuss gauge invariance, we will pretend that the graviton has a small mass to avoid having to discuss general coordinate invariance.[2] Aha, we just found the propagator for a massive spin 2 particle. So let's put it to work.

In precise analogy to (4)

$$W(J) = -\frac{1}{2}\int \frac{d^4k}{(2\pi)^4} J^{\mu}(k)^* \frac{-g_{\mu\nu} + k_\mu k_\nu/m^2}{k^2 - m^2 + i\varepsilon} J^{\nu}(k) \tag{15}$$

describing the interaction between two electromagnetic currents, the interaction between two lumps of stress energy is described by

$$W(T) = -\frac{1}{2}\int \frac{d^4k}{(2\pi)^4} T^{\mu\nu}(k)^* \frac{(G_{\mu\lambda}G_{\nu\sigma} + G_{\mu\sigma}G_{\nu\lambda}) - \frac{2}{3}G_{\mu\nu}G_{\lambda\sigma}}{k^2 - m^2 + i\varepsilon} T^{\lambda\sigma}(k) \tag{16}$$

From the conservation of energy and momentum $\partial_\mu T^{\mu\nu}(x) = 0$ and hence $k_\mu T^{\mu\nu}(k) = 0$, we can replace $G_{\mu\nu}$ in (16) by $g_{\mu\nu}$.

Now comes the punchline. Look at the interaction between two lumps of energy density T^{00}. We have from (16) that

$$W(T) = -\frac{1}{2}\int \frac{d^4k}{(2\pi)^4} T^{00}(k)^* \frac{1 + 1 - \frac{2}{3}}{k^2 - m^2 + i\varepsilon} T^{00}(k) \tag{17}$$

Comparing with (5) and using the well-known fact that $(1 + 1 - \frac{2}{3}) > 0$, we see that while like charges repel, masses attract. Trumpets, please!

[2] For the moment, I ask you to ignore all subtleties and simply assume that in order to understand gravity it is kosher to let $m \to 0$. I will give a precise discussion of Einstein's theory of gravity in Chapter VIII.1.

The universe

It is difficult to overstate the importance (not to speak of the beauty) of what we have learned: The exchange of a spin 0 particle produces an attractive force, of a spin 1 particle a repulsive force, and of a spin 2 particle an attractive force, realized in the hadronic strong interaction, the electromagnetic interaction, and the gravitational interaction, respectively. The universal attraction of gravity produces an instability that drives the formation of structure in the early universe.[3] Denser regions become denser yet. The attractive nuclear force mediated by the spin 0 particle eventually ignites the stars. Furthermore, the attractive force between protons and neutrons mediated by the spin 0 particle is able to overcome the repulsive electric force between protons mediated by the spin 1 particle to form a variety of nuclei without which the world would certainly be rather boring. The repulsion between likes and hence attraction between opposites generated by the spin 1 particle allow electrically neutral atoms to form.

The world results from a subtle interplay among spin 0, 1, and 2.

In this lightning tour of the universe, we did not mention the weak interaction. In fact, the weak interaction plays a crucial role in keeping stars such as our sun burning at a steady rate.

Degrees of freedom

Now for a bit of cold water: Logically and mathematically the physics of a particle with mass $m \neq 0$ could well be different from the physics with $m = 0$. Indeed, we know from classical electromagnetism that an electromagnetic wave has 2 polarizations, that is, 2 degrees of freedom. For a massive spin 1 particle we can go to its rest frame, where the rotation group tells us that there are $2 \cdot 1 + 1 = 3$ degrees of freedom. The crucial piece of physics is that we can never bring the massless photon to its rest frame. Mathematically, the rotation group $SO(3)$ degenerates into $SO(2)$, the group of 2-dimensional rotations around the direction of the photon's momentum.

We will see in Chapter II.7 that the longitudinal degree of freedom of a massive spin 1 meson decouples as we take the mass to zero. The treatment given here for the interaction between charges (6) is correct. However, in the case of gravity, the $\frac{2}{3}$ in (17) is replaced by 1 in Einstein's theory, as we will see Chapter VIII.1. Fortunately, the sign of the interaction given in (17) does not change. Mute the trumpets a bit.

---

[3] A good place to read about gravitational instability and the formation of structure in the universe along the line sketched here is in A. Zee, *Einstein's Universe* (formerly known as *An Old Man's Toy*).

Appendix

Pretend that we never heard of the Maxwell Lagrangian. We want to construct a relativistic Lagrangian for a massive spin 1 meson field. Together we will discover Maxwell. Spin 1 means that the field transforms as a vector under the 3-dimensional rotation group. The simplest Lorentz object that contains the 3-dimensional vector is obviously the 4-dimensional vector. Thus, we start with a vector field $A_\mu(x)$.

That the vector field carries mass m means that it satisfies the field equation

$$(\partial^2 + m^2)A_\mu = 0 \tag{18}$$

A spin 1 particle has 3 degrees of freedom [remember, in fancy language, the representation j of the rotation group has dimension $(2j+1)$; here $j=1$.] On the other hand, the field $A_\mu(x)$ contains 4 components. Thus, we must impose a constraint to cut down the number of degrees from 4 to 3. The only Lorentz covariant possibility (linear in A_μ) is

$$\partial_\mu A^\mu = 0 \tag{19}$$

It may also be helpful to look at (18) and (19) in momentum space, where they read $(k^2 - m^2)A_\mu(k) = 0$ and $k_\mu A^\mu(k) = 0$. The first equation tells us that $k^2 = m^2$ and the second that if we go to the rest frame $k^\mu = (m, \vec{0})$ then A^0 vanishes, leaving us with 3 nonzero components A^i with $i = 1, 2, 3$.

The remarkable observation is that we can combine (18) and (19) into a single equation, namely

$$(g^{\mu\nu}\partial^2 - \partial^\mu\partial^\nu)A_\nu + m^2 A^\mu = 0 \tag{20}$$

Verify that (20) contains both (18) and (19). Act with ∂_μ on (20). We obtain $m^2\partial_\mu A^\mu = 0$, which implies that $\partial_\mu A^\mu = 0$. (At this step it is crucial that $m \neq 0$ and that we are not talking about the strictly massless photon.) We have thus obtained (19); using (19) in (20) we recover (18).

We can now construct a Lagrangian by multiplying the left-hand side of (20) by $+\frac{1}{2}A_\mu$ (the $\frac{1}{2}$ is conventional but the plus sign is fixed by physics, namely the requirement of positive kinetic energy); thus

$$\mathcal{L} = \tfrac{1}{2}A_\mu[(\partial^2 + m^2)g^{\mu\nu} - \partial^\mu\partial^\nu]A_\nu \tag{21}$$

Integrating by parts, we recognize this as the massive version of the Maxwell Lagrangian. In the limit $m \to 0$ we recover Maxwell.

A word about terminology: Some people insist on calling only $F_{\mu\nu}$ a field and A_μ a potential. Conforming to common usage, we will not make this fine distinction. For us, any dynamical function of spacetime is a field.

Exercises

I.5.1. Write down the most general form for $\sum_a \varepsilon^{(a)}_{\mu\nu}(k)\varepsilon^{(a)}_{\lambda\sigma}(k)$ using symmetry repeatedly. For example, it must be invariant under the exchange $\{\mu\nu \leftrightarrow \lambda\sigma\}$. You might end up with something like

$$AG_{\mu\nu}G_{\lambda\sigma} + B(G_{\mu\lambda}G_{\nu\sigma} + G_{\mu\sigma}G_{\nu\lambda}) + C(G_{\mu\nu}k_\lambda k_\sigma + k_\mu k_\nu G_{\lambda\sigma})$$
$$+ D(k_\mu k_\lambda G_{\nu\sigma} + k_\mu k_\sigma G_{\nu\lambda} + k_\nu k_\sigma G_{\mu\lambda} + k_\nu k_\lambda G_{\mu\sigma}) + Ek_\mu k_\nu k_\lambda k_\sigma \quad (22)$$

with various unknown $A, \cdots, E$. Apply $k^\mu \sum_a \varepsilon^{(a)}_{\mu\nu}(k)\varepsilon^{(a)}_{\lambda\sigma}(k) = 0$ and find out what that implies for the constants. Proceeding in this way, derive (13).

Chapter I.6

Inverse Square Law and the Floating 3-Brane

Why inverse square?

In your first encounter with physics, didn't you wonder why an inverse square force law and not, say, an inverse cube law? You now have the deep answer. When a massless particle is exchanged between two particles, the potential energy between the two particles goes as

$$V(r) \propto \quad d^3k\, e^{i\vec{k}\cdot\vec{x}} \frac{1}{\vec{k}^2} \propto \frac{1}{r} \tag{1}$$

The spin of the exchanged particle controls the overall sign, but the $1/r$ follows just from dimensional analysis, as I remarked earlier. Basically, $V(r)$ is the Fourier transform of the propagator. The $\vec{k}^2$ in the propagator comes from the $(\partial_i\varphi \cdot \partial_i\varphi)$ term in the action, where φ denotes generically the field associated with the massless particle being exchanged, and the $(\partial_i\varphi \cdot \partial_i\varphi)$ form is required by rotational invariance. It couldn't be $\vec{k}$ or $\vec{k}^3$ in (1); $\vec{k}^2$ is the simplest possibility. So you can say that in some sense ultimately the inverse square law comes from rotational invariance!

Physically, the inverse square law goes back to Faraday's flux picture. Consider a sphere of radius r surrounding a charge. The electric flux per unit area going through the sphere varies as $1/4\pi r^2$. This geometric fact is reflected in the factor d^3k in (1).

Brane world

Remarkably, with the tiny bit of quantum field theory I have exposed you to, I can already take you to the frontier of current research, current as of the writing of this book. In string theory, our $(3+1)$-dimensional world could well be embedded in a larger universe, the way a $(2+1)$-dimensional sheet of paper is embedded in our everyday $(3+1)$-dimensional world. We are said to be living on a 3 brane.

So suppose there are n extra dimensions, with coordinates $x^4, x^5, \cdots, x^{n+3}$. Let the characteristic scales associated with these extra coordinates be R. I can't go into the different detailed scenarios describing what R is precisely. For some reason I can't go into either, we are stuck on the 3 brane. In contrast, the graviton is associated intrinsically with the structure of spacetime and so roams throughout the $(n+3+1)$-dimensional universe.

All right, what is the gravitational force law between two particles? It is surely not your grandfather's gravitational force law: We Fourier transform

$$V(r) \propto \quad d^{3+n}k \, e^{i\vec{k}\cdot\vec{x}} \frac{1}{\vec{k}^2} \propto \frac{1}{r^{1+n}} \tag{2}$$

to obtain a $1/r^{1+n}$ law.

Doesn't this immediately contradict observation?

Well, no, because Newton's law continues to hold for $r \gg R$. In this regime, the extra coordinates are effectively zero compared to the characteristic length scale r we are interested in. The flux cannot spread far in the direction of the n extra coordinates. Think of the flux being forced to spread in only the three spatial directions we know, just as the electromagnetic field in a wave guide is forced to propagate down the tube. Effectively we are back in $(3+1)$-dimensional spacetime and $V(r)$ reverts to a $1/r$ dependence.

The new law of gravity (2) holds only in the opposite regime $r \ll R$. Heuristically, when R is much larger than the separation between the two particles, the flux does not know that the extra coordinates are finite in extent and thinks that it lives in an $(n+3+1)$-dimensional universe.

Because of the weakness of gravity, Newton's force law has not been tested to much accuracy at laboratory distance scales, and so there is plenty of room for theorists to speculate in: R could easily be much larger than the scale of elementary particles and yet much smaller than the scale of everyday phenomena. Incredibly, the universe could have "large extra dimensions"! (The word "large" means large on the scale of particle physics.)

Planck mass

To be quantitative, let us define the Planck mass $M_{\rm Pl}$ by writing Newton's law more rationally as $V(r) = G_N m_1 m_2 (1/r) = (m_1 m_2 / M_{Pl}^2)(1/r)$. Numerically, $M_{\rm Pl}$ 10^{19} Gev. This enormous value obviously reflects the weakness of gravity.

In fundamental units in which $\hbar$ and c are set to unity, gravity defines an intrinsic mass or energy scale much higher than any scale we have yet explored experimentally. Indeed, one of the fundamental mysteries of contemporary particle physics is why this mass scale is so high compared to anything else we know of. I will come back to this so-called hierarchy problem in due time. For the moment, let us ask if this new picture of gravity, new in the waning moments of the last century, can alleviate the hierarchy problem by lowering the intrinsic mass scale of gravity.

Denote the mass scale characteristic of gravity in the $(n+3+1)$-dimensional universe by $M_{{\rm Pl}(n+3+1)}$ so that the gravitational potential between two objects of masses m_1 and m_2 separated by a distance $r \ll R$ is given by

$$V(r) = \frac{m_1 m_2}{[M_{\text{Pl}(n+3+1)}]^{2+n}} \frac{1}{r^{1+n}}$$

Note that the dependence on $M_{\text{Pl}(n+3+1)}$ follows from dimensional analysis: two powers to cancel $m_1 m_2$ and n powers to match the n extra powers of $1/r$. For $r \gg R$, as we have argued, the geometric spread of the gravitational flux is cut off by R so that the potential becomes

$$V(r) = \frac{m_1 m_2}{[M_{\text{Pl}(n+3+1)}]^{2+n}} \frac{1}{R^n} \frac{1}{r}$$

Comparing with the observed law $V(r) = (m_1 m_2 / M_{\text{Pl}}^2)(1/r)$ we obtain

$$M_{\text{Pl}(n+3+1)}^2 = \frac{M_{\text{Pl}}^2}{[M_{\text{Pl}(n+3+1)} R]^n} \qquad (3)$$

If $M_{\text{Pl}(n+3+1)} R$ can be make large enough, we have the intriguing possibility that the fundamental scale of gravity $M_{\text{Pl}(n+3+1)}$ is much lower than what we have always thought.

Thus, R is bounded on one side by our desire to lower the fundamental scale of gravity and on the other by experiments.

Exercise

I.6.1. Putting in the numbers show that the case $n = 1$ is already ruled out. For help, see S. Nussinov and R. Schrock, *Phys. Rev.* D59: 105002, 1999.

Chapter I.7

Feynman Diagrams

Feynman brought quantum field theory to the masses.

—J. Schwinger

Anharmonicity in field theory

The free field theory we studied in the last few chapters was easy to solve because the defining path integral (I.3.14) is Gaussian, so we could simply apply (I.2.15). (This corresponds to solving the harmonic oscillator in quantum mechanics.) As I noted in Chapter I.3, within the harmonic approximation the vibrational modes on the mattress can be linearly superposed and thus they simply pass through each other. The particles represented by wave packets constructed out of these modes do not interact:[1] hence the term free field theory. To have the modes scatter off each other we have to include anharmonic terms in the Lagrangian so that the equation of motion is no longer linear. For the sake of simplicity let us add only one anharmonic term $-\frac{\lambda}{4!}\varphi^4$ to our free field theory and, recalling (I.3.11), try to evaluate

$$Z(J) = \int D\varphi \, e^{i \int d^4x \{\frac{1}{2}[(\partial\varphi)^2 - m^2\varphi^2] - \frac{\lambda}{4!}\varphi^4 + J\varphi\}} \tag{1}$$

(We suppress the dependence of Z on λ.)

Doing quantum field theory is no sweat, you say, it just amounts to doing the functional integral (1). But the integral is not easy! If you could do it, it would be big news.

---

[1] A potential source of confusion: Thanks to the propagation of φ, the sources coupled to φ interact, as was seen in Chapter I.4, but the particles associated with φ do not interact with each other. This is like saying that charged particles coupled to the photon interact, but (to leading approximation) photons do not interact with each other.

Feynman diagrams made easy

As an undergraduate, I heard of these mysterious little pictures called Feynman diagrams and really wanted to learn about them. I am sure that you too have wondered about those funny diagrams. Well, I want to show you that Feynman diagrams are not such a big deal: Indeed we have already drawn little spacetime pictures in Chapters I.3 and I.4 showing how particles can appear, propagate, and disappear.

Feynman diagrams have long posed somewhat of an obstacle for first-time learners of quantum field theory. To derive Feynman diagrams, traditional texts typically adopt the canonical formalism (which I will introduce in the next chapter) instead of the path integral formalism used here. As we will see, in the canonical formalism fields appear as quantum operators. To derive Feynman diagrams, we would have to solve the equation of motion of the field operators perturbatively in λ. A formidable amount of machinery has to be developed.

In the opinion of those who prefer the path integral, the path integral formalism derivation is considerably simpler (naturally!). Nevertheless, the derivation can still get rather involved and the student could easily lose sight of the forest for the trees. There is no getting around the fact that you would have to put in some effort.

I will try to make it as easy as possible for you. I have hit upon the great pedagogical device of letting you discover the Feynman diagrams for yourself. My strategy is to let you tackle two problems of increasing difficulty, what I call the baby problem and the child problem. By the time you get through these, the problem of evaluating (1) will seem much more tractable.

A baby problem

The baby problem is to evaluate the ordinary integral

$$Z(J) = \int_{-\infty}^{+\infty} dq\, e^{-\frac{1}{2}m^2q^2 - \frac{\lambda}{4!}q^4 + Jq} \tag{2}$$

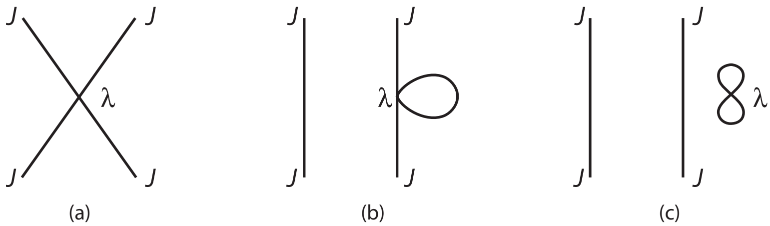
evidently a much simpler version of (1).

First, a trivial point: we can always scale $q \to q/m$ so that $Z = m^{-1}\mathcal{F}(\frac{\lambda}{m^4}, \frac{J}{m})$, but we won't.

For $\lambda = 0$ this is just one of the Gaussian integrals done in the appendix of chapter I.2. Well, you say, it is easy enough to calculate $Z(J)$ as a series in λ : expand

$$Z(J) = \int_{-\infty}^{+\infty} dq\, e^{-\frac{1}{2}m^2q^2 + Jq} \left(1 - \frac{\lambda}{4!}q^4 + \frac{1}{2}(\frac{\lambda}{4!})^2 q^8 + \cdots\right)$$

and integrate term by term. You probably even know one of several tricks for computing $\int_{-\infty}^{+\infty} dq\, e^{-\frac{1}{2}m^2q^2+Jq}q^{4n}$: you write it as $(\frac{d}{dJ})^{4n} \int_{-\infty}^{+\infty} dq\, e^{-\frac{1}{2}m^2q^2+Jq}$ and refer to (I.2.11). So



$(\frac{1}{m^2})^4 \lambda J^4$

Figure I.7.1

$$Z(J) = (1 - \frac{\lambda}{4!}(\frac{d}{dJ})^4 + \frac{1}{2}(\frac{\lambda}{4!})^2(\frac{d}{dJ})^8 + \cdots) \int_{-\infty}^{+\infty} dq e^{-\frac{1}{2}m^2q^2 + Jq} \quad (3)$$

$$= e^{-\frac{\lambda}{4!}(\frac{d}{dJ})^4} \int_{-\infty}^{+\infty} dq e^{-\frac{1}{2}m^2q^2 + Jq} = (\frac{2\pi}{m^2})^{\frac{1}{2}} e^{-\frac{\lambda}{4!}(\frac{d}{dJ})^4} e^{\frac{1}{2m^2}J^2} \quad (4)$$

(There are other tricks, such as differentiating $\int_{-\infty}^{+\infty} dq e^{-\frac{1}{2}m^2q^2+Jq}$ with respect to m^2 repeatedly, but I want to discuss a trick that will also work for field theory.) By expanding the two exponentials we can obtain any term in a double series expansion of $Z(J)$ in λ and J. [We will often suppress the overall factor $(2\pi/m^2)^{\frac{1}{2}} = Z(J = 0, \lambda = 0) \equiv Z(0, 0)$ since it will be common to all terms. When we want to be precise, we will define $\tilde{Z} = Z(J)/Z(0, 0)$.]

For example, suppose we want the term of order λ and J^4 in $\tilde{Z}$. We extract the order J^8 term in $e^{J^2/2m^2}$, namely, $[1/4!(2m^2)^4]J^8$, replace $e^{-(\lambda/4!)(d/dJ)^4}$ by $-(\lambda/4!)(d/dJ)^4$, and differentiate to get $[8!(-\lambda)/(4!)^3(2m^2)^4]J^4$. Another example: the term of order λ^2 and J^6 is $\frac{1}{2}(\lambda/4!)^2(d/dJ)^8[1/7!(2m^2)^7]J^{14} = [14!(-\lambda)^2/(4!)^2 6!7!2(2m^2)^7]J^6$. A third example: The term of order λ^2 and J^4 is $[12!(-\lambda)^2/(4!)^3 3!(2m^2)^6]J^4$. Finally, the term of order λ and J^0 is $[1/2(2m^2)^2]$ $(-\lambda)$.

You can do this as well as I can! Do a few more and you will soon see a pattern. In fact, you will eventually realize that you can associate diagrams with each term and codify some rules. Our four examples are associated with the diagrams in Figures I.7.1–I.7.4, respectively. You can see, for a reason you will soon understand, that each term can be associated with several diagrams. I leave you to work out the rules carefully to get the numerical factors right (but trust me, the "future of democracy" is not going to depend on them). The rules go something like this: (1) diagrams are made of lines and vertices at which four lines meet; (2) for each vertex assign a factor of $(-\lambda)$; (3) for each line assign $1/m^2$; and (4) for each external end assign J (e.g., Figure I.7.2 has seven lines, two vertices, and six ends, giving $\sim [(-\lambda)^2/(m^2)^7]J^6$.) (Did you notice that twice the number of lines is equal to four times the number of vertices plus the number of ends? We will meet relations like that in Chapter III.2.)

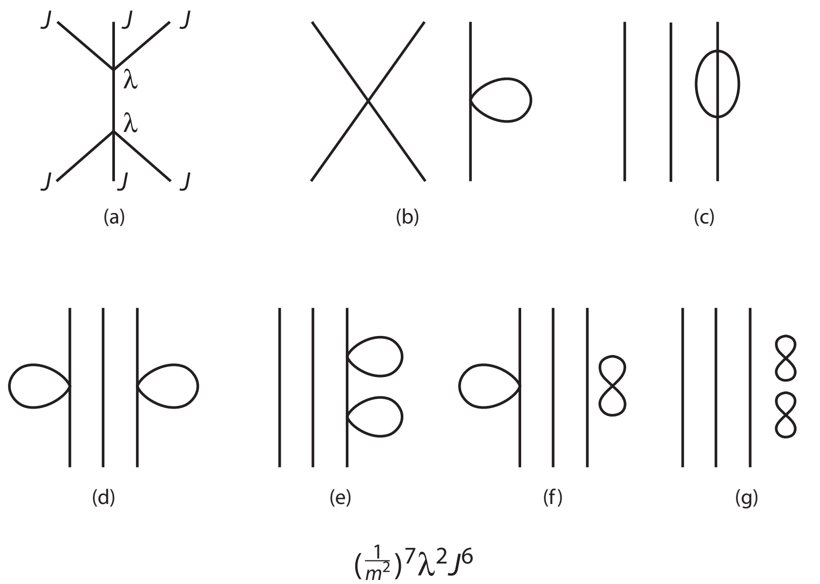


Figure I.7.2

For obvious reasons, some diagrams (e.g., Figure I.7.1a, I.7.2a) are known as tree[2] diagrams and others (e.g., Figs. I.7.1b and I.7.3a) as loop diagrams.

Do as many examples as you need until you feel thoroughly familiar with what is going on, because we are going to do exactly the same thing in quantum field theory. It will look much messier, but only superficially. Be sure you understand how to use diagrams to represent the double series expansion of $\tilde{Z}(J)$ before reading on. Please. In my experience teaching, students who have not thoroughly understood the expansion of $\tilde{Z}(J)$ have no hope of understanding what we are going to do in the field theory context.

Wick contraction

It is more obvious than obvious that we can expand $Z(J)$ in powers of J, if we please, instead of in powers of λ. As you will see, particle physicists like to classify in power of J. In our baby problem, we can write

$$Z(J) = \sum_{s=0}^{\infty} \frac{1}{s!} J^s \int_{-\infty}^{+\infty} dq e^{-\frac{1}{2}m^2q^2-(\lambda/4!)q^4} q^s \equiv Z(0,0) \sum_{s=0}^{\infty} \frac{1}{s!} J^s G^{(s)} \quad (5)$$

[2] The Chinese character for tree (A. Zee, *Swallowing Clouds*) is shown in Fig. I.7.5. I leave it to you to figure out why this diagram does not appear in our $Z(J)$.

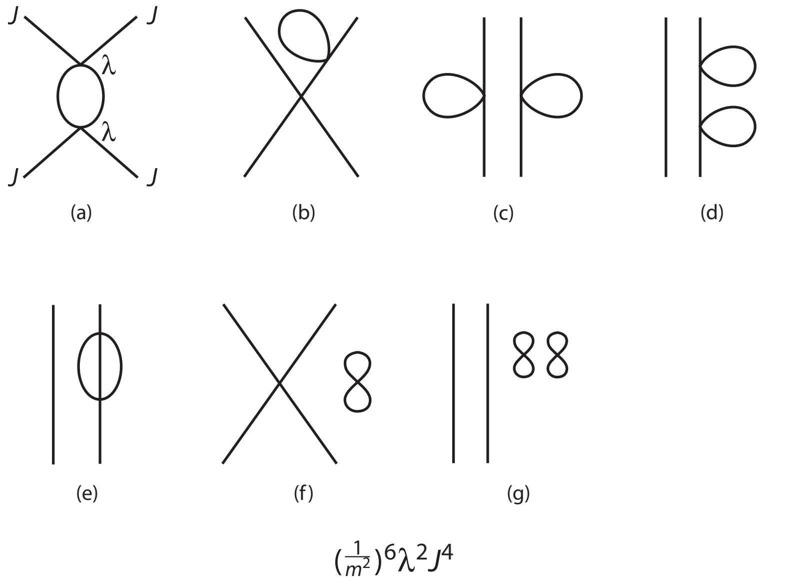


Figure I.7.3

The coefficient $G^{(s)}$, whose analogs are known as "Green's functions" in field theory, can be evaluated as a series in λ with each term determined by Wick contraction (I.2.10). For instance, the $O(\lambda)$ term in $G^{(4)}$ is

$$\frac{1}{(4!)^2}(-\lambda)\int_{-\infty}^{+\infty} dq e^{-\frac{1}{2}m^2q^2} q^8 = \frac{7!!}{(4!)^2}\frac{1}{m^8}$$

which of course better be equal[3] to what we obtained above for the λJ^4 term in $\tilde{Z}$. Thus, there are two ways of computing Z : you expand in λ first or you expand in J first.

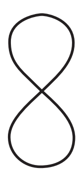


Figure I.7.4

[3] As a check on the laws of arithmetic we verify that indeed $7!!/(4!)^2 = 8!/(4!)^3 2^4$.

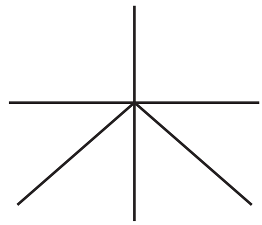


Figure I.7.5

Connected versus disconnected

You will have noticed that some Feynman diagrams are connected and others are not. Thus, Figure I.7.2a is connected while 2b is not. I presaged this at the end of Chapter I.4 and in Figures I.4.2 and I.4.3. Write

$$Z(J,\lambda) = Z(J=0,\lambda)e^{W(J,\lambda)} = Z(J=0,\lambda)\sum_{N=0}^{\infty}\frac{1}{N!}[W(J,\lambda)]^N \qquad (6)$$

By definition, $Z(J=0,\lambda)$ consists of those diagrams with no external source J, such as the one in Figure I.7.4. The statement is that W is a sum of connected diagrams while Z contains connected as well as disconnected diagrams. Thus, Figure I.7.2b consists of two disconnected pieces and comes from the term $(1/2!)[W(J,\lambda)]^2$ in (6), the 2! taking into account that it does not matter which of the two pieces you put "on the left or on the right." Similarly, Figure I.7.2c comes from $(1/3!)[W(J,\lambda)]^3$. Thus, it is W that we want to calculate, not Z. If you've had a good course on statistical mechanics, you will recognize that this business of connected graphs versus disconnected graphs is just what underlies the relation between free energy and the partition function.

Propagation: from here to there

All these features of the baby problem are structurally the same as the corresponding features of field theory and we can take over the discussion almost immediately. But before we graduate to field theory, let us consider what I call a child problem, the evaluation of a multiple integral instead of a single integral:

$$Z(J) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty} dq_1 dq_2 \cdots dq_N \, e^{-\frac{1}{2}q\cdot A\cdot q-(\lambda/4!)q^4+J\cdot q} \qquad (7)$$

with $q^4 \equiv \sum_i q_i^4$. Generalizing the steps leading to (3) we obtain

$$Z(J) = {\frac{(2\pi)^N}{\det[A]}}^{\frac{1}{2}} e^{-(\lambda/4!)\sum_i(\partial/\partial J_i)^4} e^{\frac{1}{2}J\cdot A^{-1}\cdot J} \tag{8}$$

Alternatively, just as in (5) we can expand in powers of J

$$Z(J) = {}_{s=0}^{\infty} \frac{1}{s!} J_{i_1}\cdots J_{i_s} {}_{-\infty}^{+\infty} \left({}_{l} \, dq_l \right) e^{-\frac{1}{2}q\cdot A\cdot q-(\lambda/4!)q^4} q_{i_1}\cdots q_{i_s}$$

$$= Z(0,0) {}_{s=0}^{\infty} \frac{1}{s!} J_{i_1}\cdots J_{i_s} G^{(s)}_{i_1\cdots i_s} \tag{9}$$

which again we can expand in powers of λ and evaluate by Wick contracting.

The one feature the child problem has that the baby problem doesn't is propagation "from here to there". Recall the discussion of the propagator in Chapter I.3. Just as in (I.3.16) we can think of the index i as labeling the sites on a lattice. Indeed, in (I.3.16) we had in effect evaluated the "2-point Green's function" $G^{(2)}_{ij}$ to zeroth order in λ (differentiate (I.3.16) with respect to J twice):

$$G^{(2)}_{ij}(\lambda=0) = {}_{-\infty}^{+\infty} \left({}_{l} \, dq_l \right) e^{-\frac{1}{2}q\cdot A\cdot q} q_i q_j \quad /Z(0,0) = (A^{-1})_{ij}$$

(see also the appendix to Chapter I.2). The matrix element $(A^{-1})_{ij}$ describes propagation from i to j. In the baby problem, each term in the expansion of $Z(J)$ can be associated with several diagrams but that is no longer true with propagation.

Let us now evaluate the "4-point Green's function" $G^{(4)}_{ijkl}$ to order λ :

$$G^{(4)}_{ijkl} = {}_{-\infty}^{+\infty} \left({}_{m} \, dq_m \right) e^{-\frac{1}{2}q\cdot A\cdot q} q_i q_j q_k q_l \quad 1 - \frac{\lambda}{4!} {}_{n} \, q_n^4 + O(\lambda^2) \quad /Z(0,0)$$

$$= (A^{-1})_{ij}(A^{-1})_{kl} + (A^{-1})_{ik}(A^{-1})_{jl} + (A^{-1})_{il}(A^{-1})_{jk}$$

$$- \lambda {}_{n} (A^{-1})_{in}(A^{-1})_{jn}(A^{-1})_{kn}(A^{-1})_{ln} + O(\lambda^2) \tag{10}$$

The first three terms describe one excitation propagating from i to j and another propagating from k to l, plus the two possible permutations on this "history." The order λ term tells us that four excitations, propagating from i to n, from j to n, from k to n, and from l to n, meet at n and interact with an amplitude proportional to λ, where n is anywhere on the lattice or mattress. By the way, you also see why it is convenient to define the interaction $(\lambda/4!)\varphi^4$ with a $1/4!$: q_i has a choice of four q_n's to contract with, q_j has three q_n's to contract with, and so on, producing a factor of 4! to cancel the $(1/4!)$.

Perturbative field theory

You should now be ready for field theory!

Indeed, the functional integral in (1) (which I repeat here)

$$Z(J) = \quad D\varphi\, e^{i \quad d^4x\{\frac{1}{2}[(\partial\varphi)^2 - m^2\varphi^2] - (\lambda/4!)\varphi^4 + J\varphi\}} \tag{11}$$

has the same form as the ordinary integral in (2) and the multiple integral in (7). There is one minor difference: there is no i in (2) and (7), but as I noted in Chapter I.2 we can Wick rotate (11) and get rid of the i, but we won't. The significant difference is that J and φ in (11) are functions of a continuous variable x, while J and q in (2) are not functions of anything and in (7) are functions of a discrete variable. Aside from that, everything goes through the same way.

As in (3) and (8) we have

$$Z(J) = Z(0,0)e^{-(i/4!)\lambda \quad d^4w[\delta/i\delta J(w)]^4} \quad D\varphi e^{i \quad d^4x\{\frac{1}{2}[(\partial\varphi)^2 - m^2\varphi^2] + J\varphi\}}$$

$$= Z(0,0)e^{-(i/4!)\lambda \quad d^4w[\delta/i\delta J(w)]^4} e^{-(i/2) \quad d^4x d^4y J(x) D(x-y) J(y)} \tag{12}$$

The structural similarity is total.

The role of $1/m^2$ in (3) and of A^{-1} (8) is now played by the propagator

$$D(x-y) = \quad \frac{d^4k}{(2\pi)^4} \frac{e^{ik\cdot(x-y)}}{k^2 - m^2 + i\varepsilon}$$

Incidentally, if you go back to Chapter I.3 you will see that if we were in d-dimensional spacetime, $D(x-y)$ would be given by the same expression with $d^4k/(2\pi)^4$ replaced by $d^dk/(2\pi)^d$. The ordinary integral (2) is like a field theory in 0-dimensional spacetime: if we set $d = 0$, there is no propagating around and $D(x-y)$ collapses to $-1/m^2$. You see that it all makes sense.

We also know that $J(x)$ corresponds to sources and sinks. Thus, if we expand $Z(J)$ as a series in J, the powers of J would indicate the number of particles involved in the process. (Note that in this nomenclature the scattering process $\varphi + \varphi \to \varphi + \varphi$ counts as a 4-particle process: we count the total number of incoming and outgoing particles.) Thus, in particle physics it often makes sense to specify the power of J. Exactly as in the baby and child problems, we can expand in J first:

$$Z(J) = Z(0,0) \quad {}^{\infty}_{s=0} \frac{1}{s!} J(x_1)\cdots J(x_s) G^{(s)}(x_1, \cdots, x_s)$$

$$= \quad {}^{\infty}_{s=0} \frac{1}{s!} J(x_1)\cdots J(x_s) \quad D\varphi\, e^{i \quad d^4x\{\frac{1}{2}[(\partial\varphi)^2 - m^2\varphi^2] - (\lambda/4!)\varphi^4\}}$$

$$\varphi(x_1)\cdots\varphi(x_s) \tag{13}$$

In particular, we have the 2-point Green's function

$$G(x_1, x_2) \equiv \frac{1}{Z(0,0)} \quad D\varphi \, e^{i \quad d^4x\{\frac{1}{2}[(\partial\varphi)^2 - m^2\varphi^2] - (\lambda/4!)\varphi^4\}} \varphi(x_1)\varphi(x_2) \quad (14)$$

the 4-point Green's function,

$$G(x_1, x_2, x_3, x_4) \equiv \frac{1}{Z(0,0)} \quad D\varphi \, e^{i \quad d^4x\{\frac{1}{2}[(\partial\varphi)^2 - m^2\varphi^2] - (\lambda/4!)\varphi^4\}}$$
$$\varphi(x_1)\varphi(x_2)\varphi(x_3)\varphi(x_4) \quad (15)$$

and so on. [Sometimes $Z(J)$ is called the generating functional as it generates the Green's functions.] Obviously, by translation invariance, $G(x_1, x_2)$ does not depend on x_1 and x_2 separately, but only on $x_1 - x_2$. Similarly, $G(x_1, x_2, x_3, x_4)$ only depends on $x_1 - x_4$, $x_2 - x_4$, and $x_3 - x_4$. For $\lambda = 0$, $G(x_1, x_2)$ reduces to $iD(x_1 - x_2)$, the propagator introduced in Chapter I.3. While $D(x_1 - x_2)$ describes the propagation of a particle between x_1 and x_2 in the absence of interaction, $G(x_1 - x_2)$ describes the propagation of a particle between x_1 and x_2 in the presence of interaction. If you understood our discussion of $G^{(4)}_{ijkl}$, you would know that $G(x_1, x_2, x_3, x_4)$ describes the scattering of particles.

In some sense, there are two ways of doing field theory, what I might call the Schwinger way (12) or the Wick way (13).

Thus, to summarize, Feynman diagrams are just an extremely convenient way of representing the terms in a double series expansion of $Z(J)$ in λ and J.

As I said in the preface, I have no intention of turning you into a whiz at calculating Feynman diagrams. In any case, that can only come with practice. Besides, there are excellent texts devoted to the evaluation of diagrams. Instead, I tried to give you as clear an account as I can muster of the concept behind this marvellous invention of Feynman's, which as Schwinger noted rather bitterly, enables almost anybody to become a field theorist. For the moment, don't worry too much about factors of 4! and 2!

Collision between particles

As I already mentioned, I described in chapter I.4 the strategy of setting up sources and sinks to watch the propagation of a particle (which I will call a meson) associated with the field φ. Let us now set up two sources and two sinks to watch two mesons scatter off each other. The setup is shown in Figure I.7.6. The sources localized in regions 1 and 2 both produce a meson, and the two mesons eventually disappear into the sinks localized in regions 3 and 4. It clearly suffices to find in Z a term containing $J(x_1)J(x_2)J(x_3)J(x_4)$. But this is just $G(x_1, x_2, x_3, x_4)$.

Let us be content with first order in λ. Going the Wick way we have to evaluate

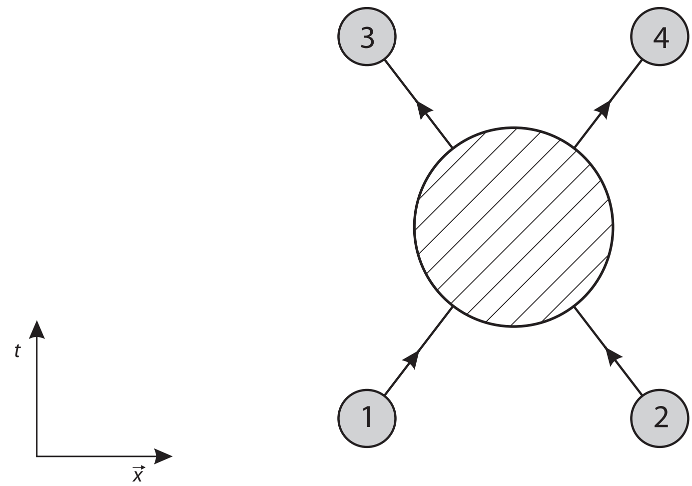


Figure I.7.6

$$\frac{1}{Z(0,0)}\left(-\frac{i\lambda}{4!}\right) d^4w \quad D\varphi\, e^{i \quad d^4x\{\frac{1}{2}[(\partial\varphi)^2 - m^2\varphi^2]\}}$$

$$\varphi(x_1)\varphi(x_2)\varphi(x_3)\varphi(x_4)\varphi(w)^4 \tag{16}$$

Just as in (10) we Wick contract and obtain

$$(-i\lambda) \quad d^4w D(x_1 - w)D(x_2 - w)D(x_3 - w)D(x_4 - w) \tag{17}$$

As a check, let us also derive this the Schwinger way. Replace $e^{-(i/4!)\lambda \quad d^4w(\delta/\delta J(w))^4}$ by $(i/4!)\lambda \quad d^4w(\delta/\delta J(w))^4$ and $e^{-(i/2) \quad d^4xd^4yJ(x)D(x-y)J(y)}$ by

$$\frac{i^4}{4!2^4}\left(d^4xd^4yJ(x)D(x-y)J(y)\right)^4 .$$

To save writing, it would be sagacious to introduce the abbreviations J_a for $J(x_a)$, $_a$ for d^4x_a, and D_{ab} for $D(x_a - x_b)$. Dropping overall numerical factors, which I invite you to fill in, we obtain

$$\sim i\lambda \quad _w\left(\frac{\delta}{\delta J_w}\right)^4 \qquad D_{ae}D_{bf}D_{cg}D_{dh}J_aJ_bJ_cJ_dJ_eJ_fJ_gJ_h \tag{18}$$

The four $(\delta/\delta J_w)$'s hit the eight J's in all possible combinations producing many terms, which again I invite you to write out. Two of the three terms are disconnected. The connected term is