

The quantum theory of fields

David Wallace

May 19, 2018

Abstract

I give an introduction to the conceptual structure of quantum field theory as it is used in mainstream theoretical physics today, aimed at non-specialists (though presuming some familiarity with quantum mechanics). My main focuses in the article are (a) the common structure of quantum field theory as it is applied in solid-state physics and as it is applied in high-energy physics; (b) the modern theory of renormalisation.

1 Introduction

Quantum theory, like Hamiltonian or Lagrangian classical mechanics,¹ is not a concrete theory like general relativity or electromagnetism, but a *framework theory* in which a great many concrete theories, from qubits and harmonic oscillators to proposed quantum theories of gravity, may be formulated. Quantum field theory, too, is a framework theory: a sub-framework of quantum mechanics, suitable to express the physics of spatially extended bodies, and of systems which can be approximated as extended bodies. Fairly obviously, this includes the solids and liquids that are studied in condensed-matter physics, as well as quantised versions of the classical electromagnetic and (more controversially) gravitational fields. Less obviously, it also includes pretty much any theory of relativistic matter: the physics of the 1930s fairly clearly established that a quantum theory of relativistic *particles* pretty much has to be reexpressed as a quantum theory of relativistic *fields* once interactions are included. So quantum field theory includes within its framework the Standard Model of particle physics, the various low energy limits of the Standard Model that describe different aspects of particle physics, gravitational physics below the Planck scale, and almost everything we know about many-body quantum physics. No more need be said, I hope, to support its significance for naturalistic philosophy.

In this article I aim to give a self-contained introduction to quantum field theory (QFT), presupposing (for the most part) only some prior exposure to classical and quantum mechanics (parts of sections 9–10 also assume a little familiarity with general relativity and particle physics). Since the normal form

¹See North, this volume, for discussion of these formulations of classical mechanics.

of that ‘introduction’ in a physicist’s education is two semesters of graduate-level courses, *this* introduction is inevitably pretty incomplete. I give virtually no calculations and require my reader to take almost everything on trust; my focus is on the conceptual structure of QFT and its philosophically interesting features. This article will not, realistically, equip its readers to carry out research in philosophy of QFT, but I hope it will help make sense of references to QFT in the philosophical and physics literature, serve to complement more technical introductions, and to give some sense of just how important and interesting this field of physics has been in the last fifty years.

My account is in logical rather than historical order, and the physics described is all standard; I make no attempt to reference primary sources, though I give some suggested introductions to the literature at the end. I use units in which \hbar and c are both set to 1; note that this means that mass and energy have the same dimensions, and that length has the dimensions of inverse mass, so that we can talk interchangeably of something happening at large energies/masses and at short lengthscales.

2 Warm-up: classical continuum mechanics

Classical mechanics represents the instantaneous state of a fluid by (inter alia) a pair of functions on space: a *mass density* $\rho(x)$ and a *velocity field* $\mathbf{v}(x)$. Their interpretations are straightforward and standard: the integral of $\rho(x)$ over some region R gives the total mass of the part of the fluid in R ; the integral of $\rho(x)\mathbf{v}(x)$ over R gives the total momentum of that same part. ρ and \mathbf{v} jointly satisfy well-known equations, which can be derived from first principles on phenomenological grounds and which characterise the fluid’s dynamics in terms of a small number of parameters, such as its viscosity and compressibility. The equations are notoriously difficult to solve, but when they are solved (analytically, numerically, or under certain local approximations) they do an excellent job of describing the physics of fluids in terms of their continuously varying density and velocity fields.

Remarkably so: because real fluids don’t have continuously varying densities and velocity fields. Whatever the ‘fundamental’ story about the constituents of matter, it’s beyond serious doubt that ordinary fluids like water or treacle are composed of discretely many atoms, which in turn have considerable substructure, and so — insofar as the language of classical physics is applicable at all — the ‘real’ density of the fluid is varying wildly on lengthscales of $\sim 10^{-10}\text{m}$ and even more wildly on shorter lengthscales.

The issue is usually addressed early in a kinetic-theory or fluid-dynamics course. There, one is typically² told that these functions are defined by averaging over some region large enough to contain very many atoms, but small enough that the macroscopic parameters do not differ substantially across the width of that region. For a fluid that phenomenologically has fairly constant density and momentum over lengthscales of $\sim 10^{-4}\text{m}$, for instance, that region needs

²See, e.g. , Kenyon (1960, p.3) or Kambe (2007, pp.2-3).

to have a width L satisfying $10^{-4}\text{m} \gg L \gg 10^{-10}\text{m}$. But the merely kinematic task of defining these quantities does not capture the remarkable feature of fluid dynamics (and of emergence in physics in general³): that there *exist* closed-form dynamical equations for these quantities, so that we can actually reason from current bulk features of the fluid to future bulk features without any need for additional information about the microphysics.

That is not to say that the microphysics is irrelevant to the fluid's behaviour. When the assumptions under which the density is defined fail — that is, when there is *no* lengthscale both short relative to the scales on which the fluid varies, and long compared to atomic lengthscales — then the fluid-dynamics description breaks down entirely and the underlying physics must be considered directly. Droplet formation has this feature, for instance: droplets break off from a stream of water when the width of the tube of water connecting the proto-droplet to the rest of the fluid becomes only a few atoms wide. So do shock waves: the width of the shock front depends on microphysics and not just on the phenomenological parameters. But outside these special cases, the relevance of the microphysics is purely that the viscosity and other parameters that characterise the fluid are determined — and can in some cases be calculated — from that microphysics.

So consider what we could learn about a given fluid if all we knew was fluid dynamics. We would have determined the coefficients in the equations directly by experiment, not by deduction from the microphysics, for we do not know what it is. We can predict from features *internal* to fluid dynamics that it will break down at some short lengthscales: droplet formation, and shock waves, lead to unphysical singularities if the continuum physics is exact. And we might (depending on how the details of the thought experiment are filled in) have reason external to the theory to expect such a breakdown. Furthermore, on the assumption that there is an underlying theory and that it reduces to fluid dynamics in some limit, we can reverse-engineer a few facts about that underlying theory — but the latter will be grossly underdetermined by the macro-level facts. Unless we have empirical access to droplet or shock-wave phenomena, or other means to probe the underlying theory directly, we will simply have to remain ignorant of that underlying theory. And if — with many real-life metaphysicians — we wish to set aside the emergent description of the world at large scales that fluid dynamics gives us and look for information about *fundamental metaphysics* contained within our physics, we will look in vain.⁴

This is pretty much the empirical situation that modern quantum field theory leaves us in.

3 Formal quantum theory of the continuum

Suppose we set out to construct a *quantum* theory appropriate to describe a continuous entity — whether a field, or a solid object, or a fluid. Perhaps we know that the entity is not continuous at shorter scales; perhaps we even possess

³On emergence, see Wilson, this volume.

⁴See Wallace (2018a) for further discussion of this point.

the physics applicable at those scales; perhaps we only suspect a breakdown of continuity; perhaps we believe there is no breakdown; but in any case we carry out our construction without using detailed information about the short-distance physics. (For simplicity, let's assume the continuum is spatially finite for now.)

As a concrete model, suppose that the continuum has one large-scale degree of freedom per spatial point: suppose, for instance, that it is a scalar field theory where the degree of freedom is field strength, or (less realistically) a solid body in one dimension where the degree of freedom is displacement from equilibrium in some fixed direction.⁵ (In the rest of this article, I will use this model extensively for quantitative examples, but the qualitative features apply to pretty much all quantum field theories.)

Observable quantities will be (or at least: will include) averages of this quantity over some region, so to any spatial region X we would expect to be able to assign an operator $\widehat{\varphi}_X$ representing that average. Furthermore, if X and Y are disjoint, we should expect $\widehat{\varphi}_X$ and $\widehat{\varphi}_Y$ to commute (recall that in quantum mechanics, operators commute iff the physical quantities they represent can simultaneously have definite values — something we would expect of physical quantities assigned to distinct subsystems).

If we partition space into, say, a grid of equal-size volumes X_i , each of length L , centred on points x_i , we can define simultaneous eigenstates of all the \widehat{X}_i ; any such state $|\chi\rangle$ will define a real function $[\chi]_L$, given by

$$[\chi]_L(x_i) = \langle \chi | \widehat{X}_i | \chi \rangle \quad (1)$$

and interpolating between the x_i in some arbitrary smooth way. There will be many $|\chi\rangle$ corresponding to a given function $[\chi]_L$, corresponding to the many possibilities of short-distance physics on scales below L ; we can think of each such $|\chi\rangle$ as representing a state of the continuum whose structure on scales large compared to L is given by $[\chi]_L$. Wave-packets around such states are appropriate candidates to represent quasi-classical states of the continuum on such scales.

Formally speaking, we can try to take the continuum limit of this theory, in which:

- To each point x of space is assigned an operator $\widehat{\varphi}(x)$, with any two such operators commuting;
- To each such $\widehat{\varphi}(x)$ can be assigned a conjugate momentum operator $\widehat{\pi}(x)$, with $[\widehat{\varphi}(x), \widehat{\pi}(y)] = i\delta(x - y)$;
- A simultaneous eigenstate $|\chi\rangle$ of all of the $\widehat{\varphi}(x)$ is represented by a function χ , with

$$\chi(x) = \langle \chi | \widehat{\varphi}(x) | \chi \rangle. \quad (2)$$

⁵The most realistic solid-state equivalent would include *three* degrees of freedom per spatial point, since the solid can be displaced in any direction; I stick to one degree of freedom for expository simplicity.

In this limit the degeneracy vanishes and each such χ picks out a *unique* state $|\chi\rangle$; continuing to proceed formally, we can represent an arbitrary state $|\Psi\rangle$ by

$$|\Psi\rangle = \int \mathcal{D}\chi \Psi[\chi] |\chi\rangle, \quad (3)$$

where $\int \mathcal{D}\chi$ is the *path integral* over all functions χ and $\Psi[\chi] \equiv \langle \chi | \Psi \rangle$ is a complex functional assigning a complex number to each real function.

- With the state space thus represented, we can write down a dynamics by means of a Hamiltonian like

$$\hat{H} = \int dx^3 \left(\frac{1}{2} \hat{\pi}(x)^2 + (\nabla \hat{\varphi}(x))^2 + \frac{1}{2} m^2 \hat{\varphi}(x)^2 + V(\hat{\varphi}(x)) \right) \quad (4)$$

or (in practice usually more useful) via a Feynman path integral

$$\langle \chi_1 | \hat{U}(t-t') | \chi_2 \rangle = \int_{\chi(t',x)=\chi_1(x)}^{\chi(t,x)=\chi_2(x)} \mathcal{D}\chi \exp(-iS[\chi]) \quad (5)$$

where

$$S[\chi] = \int_{t'}^t dt \int dx^3 \left(\frac{1}{2} \dot{\chi}(x)^2 - \frac{1}{2} (\nabla \chi(x))^2 - \frac{1}{2} m^2 \chi(x)^2 - V(\chi(x)) \right). \quad (6)$$

In either (4) or (6), the first three terms represent a linear (i. e. free) continuum theory, and the V term encodes self-interaction.

4 Particles

The formal limit discussed above is mathematically pathological, and understanding and resolving that pathology is key to understanding modern quantum field theory. But since the necessary discussion will be somewhat abstract, let's start off by proceeding formally and extracting some of the physical content of the continuum theory.

To begin with, let's consider free field theories, where $V = 0$ and so the Hamiltonian is quadratic. The paradigm of a quadratic Hamiltonian is the simple harmonic oscillator

$$\hat{H} = \frac{1}{2} (\hat{P}^2 + \omega^2 \hat{Q}^2), \quad (7)$$

which can be solved exactly by introducing 'annihilation' and 'creation' operators \hat{a} , \hat{a}^\dagger defined by

$$\hat{a} = (2\omega)^{-1/2} \hat{Q} + i(\omega/2)^{1/2} \hat{P}; \quad \hat{a}^\dagger = (2\omega)^{-1/2} \hat{Q} - i(\omega/2)^{1/2} \hat{P} \quad (8)$$

and satisfying $[\hat{a}, \hat{a}^\dagger] = 1$. The ground state $|\Omega\rangle$ of the theory satisfies $\hat{a}|\Omega\rangle = 0$ and has energy $\omega/2$, and the eigenstates of the theory are given by successive actions of \hat{a}^\dagger on the ground state:

$$|n\rangle \propto (\hat{a}^\dagger)^n |\Omega\rangle; \quad \hat{H} |n\rangle = (n + 1/2)\omega |n\rangle. \quad (9)$$

The Hamiltonian can be rewritten as

$$\widehat{H} = \omega(\widehat{a}^\dagger \widehat{a} + 1/2) \quad (10)$$

and the concrete mathematical form of the ground state, as a wavefunction in ‘position’ space, is

$$\langle x|\Omega\rangle \propto \exp(-\omega x^2/2). \quad (11)$$

(I put ‘position’ in quotes because in this abstracted form of the simple harmonic oscillator there is no reason to require x to be interpreted as position in physical space.)

A general quadratic Hamiltonian with form

$$\widehat{H} = \frac{1}{2} \sum_{mn} J_{mn} \widehat{P}_n^2 + \frac{1}{2} \sum_{n,m} K_{nm} \widehat{Q}_n \widehat{Q}_m \quad (12)$$

can then be understood as a collection of *coupled* harmonic oscillators, and can always be diagonalised, by a linear change of variables, into a sum of *uncoupled* harmonic oscillators – *modes* — of different frequencies $\omega(k)$, with creation and annihilation operators $\widehat{a}^\dagger(k)$ and $\widehat{a}(k)$:

$$\widehat{H} = \sum_k \omega(k) (\widehat{a}^\dagger(k) \widehat{a}(k) + 1/2) \equiv \sum_k \omega(k) \widehat{a}^\dagger(k) \widehat{a}(k) + \text{constant}. \quad (13)$$

These creation and annihilation operators satisfy

$$[\widehat{a}(k), \widehat{a}^\dagger(k)] = 1; \quad [a(k), a(l)] = [\widehat{a}(k), \widehat{a}^\dagger(l)] = 0 \text{ for } k \neq l. \quad (14)$$

We can formally treat a free field theory as the continuum limit of this oscillator sum. Indeed, the coordinate transform can be concretely calculated; the modes are labelled by spatial vectors k and given by

$$\widehat{a}^\dagger(k) = \mathcal{N} \int dx^3 \left(2\omega(k) \right)^{-1/2} e^{-ik \cdot x} \widehat{\varphi}(x) - i(\omega(k)/2)^{1/2} e^{+ik \cdot x} \widehat{\pi}(x) \quad (15)$$

where $\omega(k) = \sqrt{m^2 + k^2}$, and where the normalisation constant \mathcal{N} and the allowed values of k depend on the spatial extent of the system. The Hamiltonian of this system is formally infinite because of the constant term in (13) — the first of the infinities that occur because of the mathematical pathologies of the theory and which we will shortly have to tame — but formally speaking we can just remove the constant term by subtracting an infinite correction from the Hamiltonian, without affecting the physics. (If this makes you uncomfortable, good! — but bear with me a little longer.)

We expect the ground state of a set of coupled harmonic oscillators to be highly entangled with respect to the original variables, which in the case of our continuum theory is to say that the degrees of freedom at distinct points of space ought to be entangled. And so it turns out to be, as can most easily be seen by calculating $\langle \Omega | \widehat{\varphi}(x) \widehat{\varphi}(y) | \Omega \rangle$; we find that

$$\langle \Omega | \widehat{\varphi}(x) \widehat{\varphi}(y) | \Omega \rangle = \frac{K}{|x - y|} \exp(-|x - y|m). \quad (16)$$

(A similar expression holds for $\hat{\pi}(x)$.) So if two points are separated by a distance $\ll 1/m$, Bell-type measurements of the fields will pick up significant Bell-inequality violation; on scales $\gg 1/m$, this will be negligible.⁶ (The distance $1/m$ — or \hbar/mc , in more familiar units — is called the *Compton wavelength*; it is typically much larger than the scale on which we expect the continuum approximation to fail.) Unsurprisingly, this entanglement persists for interacting theories, and is a general feature of quantum field theories. It is a clue that even in particle physics, $|\Omega\rangle$ cannot be thought of simply as “empty space”: it has a considerable amount of quantum structure. (A further clue comes from the formally infinite energy density of the vacuum). Starting with the ground state, arbitrary states of the field can be created by superposing states of form

$$|n_k\rangle = \Pi_k(\hat{a}^\dagger(k))^{n_k} |\Omega\rangle \quad (17)$$

— that is, states like this form a basis for the full Hilbert space of the theory.

To see the full physical significance of the harmonic-oscillator analysis of the continuum theory, let’s define a subspace \mathcal{H}_{1P} as spanned by the states

$$|k\rangle = \hat{a}^\dagger(k) |\Omega\rangle; \quad (18)$$

that is, states in \mathcal{H}_{1P} are arbitrary superpositions of singly-excited modes. Since each of these states is an eigenstate of energy, \mathcal{H}_{1P} is conserved under the Schrodinger equation, so that the physics of singly-excited states is a self-contained dynamics in its own right.

Given the expression (15) for $\hat{a}(k)$, any such state can be written (non-uniquely) as

$$|\psi\rangle = \int dx^3 (f(x)\hat{\varphi}(x) |\Omega\rangle + g(x)\hat{\pi}(x) |\Omega\rangle) \quad (19)$$

for complex functions f, g , and in fact it is easy to show that the converse is also true. But relations like (16) tell us that the states $\hat{\varphi}(x) |\Omega\rangle$ and $\hat{\pi}(x) |\Omega\rangle$ are localised excitations of the continuum that are negligible at distances from x much larger than the Compton wavelength. So we can think of the states of \mathcal{H}_{1P} as superpositions of singly-localised excitations.

But that is exactly the concept of ‘particle’ in quantum *mechanics*: quantum particles are not in general localised, but they can be expressed as superpositions of states that are localised (wave-packets, say, or — formally — position eigenstates). This suggests that \mathcal{H}_{1P} can be understood as a space of one-particle states — the *one-particle subspace* (hence, “1P”) of the quantum field theory. This naturally suggests identifying multiply-excited states like (17) as multi-particle states, and indeed reinterpreting any state of the theory as a superposition of multi-particle states. (Because the operators $\hat{a}^\dagger(k)$ and $\hat{a}^\dagger(l)$ commute, it is easy to show that these particles obey Bose statistics.)

As long as we continue to consider *non-interacting* field theories, this reinterpretation of a continuum theory as a multi-particle theory is exact, and indeed serves as an alternative construction of a quantum field theory: start with a

⁶See this volume, section 4, for discussion of the Bell inequality.

one-particle quantum theory and construct from it the direct sum of the symmetrised N -fold tensor product for each N :

$$\mathcal{F}(\mathcal{H}_{1P}) = \mathcal{H}_0 \oplus \mathcal{H}_{1P} + \mathcal{S}\mathcal{H}_{1P} \otimes \mathcal{H}_{1P} \oplus \dots \quad (20)$$

(where \mathcal{H}_0 is a one-dimensional Hilbert space and \mathcal{S} is the symmetrisation operator, imposing Bose statistics), and define the field operators by inverting (15). The space thus constructed is called the (symmetric) *Fock space* of \mathcal{H}_{1P} and the construction process itself is called *second quantisation*; see, e. g. , Saunders (1992) or Wald (1994) for contemporary presentations of it.

Given these equivalent ways of thinking of the theory, one sometimes hears talk of *field-particle duality*, of the idea that ‘field’ and ‘particle’ are equally valid ways of interpreting the underlying physics. But this talk of duality only really applies in the (ultimately physically uninteresting) case of theories without interactions.

If a *small* interaction term is introduced to the free Hamiltonian, we expect that the particle analysis of the theory remains approximately valid. The interaction term can then be naturally interpreted as introducing transitions between excited modes of the harmonic oscillators, which under the particle interpretation can be understood as scattering effects between particles, and its effects can be studied by means of perturbation theory. But this analysis will only ever be approximate: as the interaction strength increases, the particle description of the theory becomes less and less valid, and eventually will need to be abandoned altogether as a useful description of the theory. For this reason we would (in my view) do better here to speak of ‘emergence’ of particles from the continuum theory, rather than of duality. From this perspective, ‘particles’ are certain excitations of the ground state of the continuum which, to a varying degree of accuracy, approximately instantiate the physics of an interacting-particle theory.⁷

It’s worth stressing that this picture of particles plays out pretty much identically whether the underlying continuum quantum theory is the quantum theory of a solid or liquid, or a quantum field. In the former, particles are often referred to as *quasi-particles* (such as the phonon, the quantum of vibration) but as far as modern field theory is concerned, all particles are quasi-particles.

To discuss the field-particle relation any further, though, we need to address the fact that any talk of ‘small’ or ‘large’ interactions is completely undefined mathematically, as long as the pathologies of the continuum theory are unaddressed.

5 Effective field theories

These pathologies are trackable to the continuum theory’s infinitely many degrees of freedom per spacetime volume. For a start, a Hilbert space like this

⁷See Wallace (2012, ch.2) and Rosaler (2015) for further discussion of this notion of instantiation.

is *non-separable*: it has no countable basis, and hence naturally factors into uncountably many superselection sectors, which differ only by the (ex hypothesi unphysical) features of the theory on arbitrarily short lengthscales. More seriously, any attempt to calculate physical quantities by the normal methods of theoretical physics gives an infinite answer. This occurs even in the case of a non-interacting system: we have seen that the expected value of the Hamiltonian in (4) is formally infinite, and so is the path integral in (5), though in both cases we can formally treat this as an unobservable (infinite) correction to the energy or action. This rather unsatisfactory situation becomes intolerable as soon as interactions are introduced (in other words: as soon as the dynamics stops being trivial), at which point the formal machinery of quantum physics delivers infinities for pretty much any question we choose to ask.

Although there is an honorable tradition in mathematical physics (see the Further Reading) of trying to formulate a fully mathematically rigorous theory of the continuum that avoids these pathologies, while remaining well-defined on all lengthscales, the current consensus in mainstream physics⁸ is that the problems should be resolved by taking seriously the idea that we were in the first place only looking for a theory describing the continuum down to some lengthscale, and that the pathologies are caused by going to the continuum limit in the first place rather than remaining finite. To see why this might be, let's consider in more detail the path integral (3) that formally defines the inner product. This integral is supposed to be over *all* functions χ , or at the least over all square-integrable functions, and (restricting, for simplicity, to one dimension) an arbitrary such function can be written as

$$\chi(x) = \sum_{n=-\infty}^{n=\infty} \alpha_n \exp(-2\pi ni/R) \quad (21)$$

where R is the spatial extent of the system and $\alpha_{-n}^* = \alpha_n$. So the integral can be decomposed as

$$\int \mathcal{D}\chi = \prod_{n=0}^{\infty} \int d\alpha_n = \int d\alpha_0 \int d\alpha_1 \cdots \quad (22)$$

And now it's fairly unsurprising that carrying out infinitely many such integrals gives an infinite answer. But if, say, $R = 10\text{m}$, then the integrals over α_n for $n \gg 10^{11}$ integrate over functions varying rapidly on scales $\ll 10^{-10}\text{m}$. So if we were only trying to construct a continuum theory describing features of the continuum on scales longer than that (if, for instance, we were studying a "continuum" which is actually made of discrete atoms of size $\sim 10^{-10}\text{m}$, as in condensed-matter physics) then it becomes reasonable to consider cutting off the integral, by discarding the integration over functions varying on those scales (that is, by discarding the integrals over α_n for $n > 10^{11}$). Doing so removes the infinities from the theory.

⁸For presentations, see (e. g.) Weinberg (1995b), Zee (2003), or Banks (2008).

If you're not suspicious of this process, you should be. It is one thing to set out to construct a theory applicable only above certain lengthscales; it is quite another to suppose that it can be done simply by discarding any influence of shorter-lengthscale physics. After all, I could just as easily have chosen the cutoff length at 10^{-9}m , or imposed it by some other means than Fourier modes. The actual physics at the cutoff lengthscale is presumably extremely complicated, and there doesn't seem any reason not to expect those complexities to affect the form of the larger-scale physics. We might try to avoid this by supposing that the cutoff occurs at precisely the *physical* lengthscale at which the theory fails — at the atomic lengthscale in condensed-matter physics, say) so that those degrees of freedom are unphysical anyway — but we still face the problem that our crude imposition of the cutoff is presumably far removed from the actual way in which a microphysical description fails.

It is one of the most remarkable features of quantum field theory — and one of the key discoveries of theoretical physics in the postwar period — that on the contrary, the details of physics below the cutoff have *almost no empirical consequences* for large-scale physics. The details are mostly too technical for an article at this level, but the general idea can be understood as follows.

Consider again the dynamics (5). Expanding V as a power series in its argument,

$$V(x) = V_0 + \frac{1}{4!}\lambda_4 x^4 + \frac{1}{6!}\lambda_6 x^6 + \dots \quad (23)$$

(where we assume V is symmetric under $x \rightarrow -x$, and ignore the x^2 term since it is already included in the Hamiltonian), we can see that the theory is specified by m^2 (corresponding, where a particle interpretation is valid, to the squared particle mass), V_0 , the infinite number of coefficients $\lambda_4, \lambda_6, \dots$ and, tacitly, by the method used to cut off the theory at short lengthscales, which we can schematically write as Λ . (We can think of Λ as denoting the lengthscale at which the cutoff is imposed as well as the details of the method by which it is imposed; by abuse of notation, I will also use Λ to denote the lengthscale alone.) So we can consider an infinite family of theories parametrised by these variables. I will write α to denote, collectively, all the variables except Λ ; (α, Λ) then denotes a particular theory in this family. It will be helpful to think of α as a set of coordinates in a *theory space* \mathcal{A} ; a theory is specified by a point in \mathcal{A} together with a choice of cutoff.

In general, any two such theories will be physically distinguishable by some in-principle-measurable features. (For instance, any two theories with different cutoff procedures can be distinguished by probing the physics around the cutoff scale.) But recall that we are only really interested in using these theories to describe physics at scales large compared to the cutoff; close to the cutoff scale, our arbitrary assumptions about the nature of the cutoff make the theory untrustworthy. So trustworthy physical predictions arise only from the large-lengthscale features of these theories. With this in mind, define two theories T_1, T_2 as *IR-equivalent* (IR for 'infra-red', physics jargon for long-distance), $T_1 \sim_{IR} T_2$, if they make the same prediction values for the evolution of dynamical variables on lengthscales large compared to the cutoff, at least for quantum

states which do not themselves vary sharply on lengthscales close to the cutoff. There are various ways of making this precise; for our purposes, a heuristic understanding will suffice.

(It is important to appreciate that IR-equivalence is a much finer notion than empirical equivalence. The “long-distance” features of a theory can be defined on scales of tens of nanometers (in condensed-matter physics) or many orders of magnitude smaller (in particle physics); they do not correspond simply to directly observable features of the continuum.)

Now consider two cutoffs Λ, Λ' (with, for definiteness, $\Lambda' > \Lambda$). For any given $\mathcal{T}(\alpha, \Lambda)$, it turns out that changing the cutoff from Λ to Λ' can be compensated for by changing the other variables from α to some α' , without in any way affecting the physical predictions of the theory at lengthscales large compared to Λ and Λ' . That is: there is some transformation $\mathcal{R}(\Lambda \rightarrow \Lambda')$ acting on theory space \mathcal{A} such that

$$(\alpha, \Lambda) \sim_{IR} (\mathcal{R}(\Lambda \rightarrow \Lambda') \cdot \alpha, \Lambda'). \quad (24)$$

(Again, this can be proved with various degrees of generality and rigor; again, here I just assert it without proof.) The transformation $U(\Lambda \rightarrow \Lambda')$ is known as the *renormalisation group*; it has a central role in modern quantum field theory and is discussed in more detail by Williams (this volume).

So there is in a sense redundancy in our family of theories: two apparently different theories may correspond to the same large-scale phenomena and so be interchangeable, as long as we regard the theories as in any case trustworthy only as regards large-scale phenomena. In fact the redundancy is considerably more dramatic than this: we can (in this particular example, and in fact in general) identify a “relevant” subset of the coordinates of \mathcal{A} such that only the relevant coordinates have any significant effect on the physics. And this ‘relevant’ subset is in general finite-dimensional. In the case of the three-dimensional scalar theory we have been considering, for instance, it is three-dimensional (the relevant coordinates can *loosely* be thought of as the zero of the potential V_0 , the free-Hamiltonian parameter m^2 and the first coefficient λ_4 in the expansion of V , though as we will see in the next section, it’s a bit subtler than that suggests.)

So: for given cutoff Λ , we can specify a theory — at least as far as the long-distance content of the theory is concerned — just by giving the finitely many values of the relevant coordinates. And Λ itself can be chosen according to convenience or whim, for if we wish to replace it with Λ' we need only make a compensating change in those relevant coordinates.

This approach to “continuum” quantum physics (which, to repeat, is absolutely standard in modern physics, both in particle physics and in condensed-matter physics) is known as *effective field theory*. The term also applies to individual theories: an effective field theory is a quantum field theory understood as applying only on lengthscales larger than some short-distance cutoff, and identified with a equivalence class of IR-equivalent theories defined by various specific cutoff schemes. All empirically-relevant quantum field theories in physics, at present, are effective field theories. To quote the author of one

popular textbook,

I emphasize that Λ should be thought of as physical, parametrizing our threshold of ignorance, and not as a mathematical construct. Indeed, physically sensible quantum field theories should all come with an implicit Λ . If anyone tries to sell you a field theory claiming that it holds up to arbitrarily high energies, you should check to see if he sold used cars for a living. (Zee 2003, p.162).

Equally, though, the precise (or even approximate) value of Λ is irrelevant to the large-scale physics; all that matters is that it is much smaller than the lengthscales on which we deploy the field theory to model the world.

6 Renormalisation

To get a better sense of how effective field theory works, let's look (schematically) at how we might calculate some physical quantity — say, the “four-point function” of the theory G , which gives the expectation value of quadruples of field operators (possibly at different times) with respect to the system's ground state. (G corresponds, loosely speaking, to the scattering amplitude between pairs of particles.) A standard approach to doing so in ordinary quantum mechanics is to decompose the Hamiltonian into a sum of a ‘free’ and ‘interaction’ term:

$$\hat{H} = \hat{H}_0 + \hat{H}_{int}. \quad (25)$$

The idea is that we can solve the physics of \hat{H}_0 exactly, and that \hat{H}_{int} can be treated as a small correction to \hat{H}_0 , whose effects can be estimated by the methods of perturbation theory.

If we try this for the Hamiltonian (4), it would be natural to take

$$\hat{H}_0 = \int dx^3 \frac{1}{2} (\hat{\pi}(x)^2 + \nabla \hat{\varphi}(x)^2 + m^2 \hat{\varphi}(x)^2) \quad (26)$$

and

$$\hat{H}_{int} = \int dx^3 \hat{V}(x). \quad (27)$$

After introducing a cutoff, both are well-defined and finite; the former can be solved exactly to give a well-behaved theory of a non-self-interacting approximately-continuous quantum system, along the lines of section 3 (but with the formal moves of that section legitimated by the cutoff).

Suppose for now that V is purely quartic: that is, $V(x) = \lambda_4 x^4/4!$. Then what perturbation theory is *supposed* to deliver for us is an expression for G in powers of λ_4 : something like

$$G = G_0 + G_1 \lambda_4 + G_2 (\lambda_4)^2 + \dots \quad (28)$$

If we calculate this power series to first order in λ_4 (known in physics as “tree-order”, a reference to the Feynman-diagram notation used in practice to work

out the power series) we get sensible, well-behaved answers (and answers independent of the cutoff scale Λ). But when we come to calculate the λ_4^2 term (the “one-loop correction”, in physics terminology) we get a quite large result, a sum of terms the largest of which is proportional to the logarithm of the inverse cutoff length, $\log(1/\Lambda)$. (If we had tried to calculate this quantity formally in the continuum theory, without any cutoff, the answer would have been not just *large* but *infinite*). Terms this large invalidate the perturbative expansion and call into question the validity even of the ‘sensible’ tree-order result. Evaluating subsequent terms in the power series likewise gives very large results.

It turns out, however, that these large results can largely be removed by absorbing them into the definitions of the parameters λ_4 and m^2 . What this amounts to (roughly speaking⁹) is that we can rewrite the Hamiltonian as

$$\widehat{H} = \widehat{H}_0 + \widehat{H}_{int} = (\widehat{H}_0 + \widehat{\Delta}) + (\widehat{H}_{int} - \widehat{\Delta}) \equiv \widehat{H}'_0 + \widehat{H}'_{int} \quad (29)$$

for some operator $\widehat{\Delta}$, such that

1. \widehat{H}'_0 has the same functional form as \widehat{H}_0 , but with a new value $(m^2)^{ren}$ (‘ren’ for ‘renormalised’) for the quadratic parameter, related to m^2 by an expression like

$$(m^2)^{ren} = m^2 + \alpha/\Lambda^2 + (\text{smaller terms}) \quad (30)$$

for some dimensionless quantity α .

2. \widehat{H}'_{int} is small enough to treat as a perturbation of \widehat{H}'_0 , at least for states in a certain energy range (which range is determined by the choice of $\widehat{\Delta}$);
3. That perturbative expansion is a power series expansion not in λ_4 , but in a new parameter λ_4^{ren} , related to λ_4 by an expression like

$$\lambda_4^{ren} = \lambda_4 + \beta \log(1/m\Lambda) + (\text{smaller terms}) \quad (31)$$

for some dimensionless quantity β . The leading-order term in that expansion is the tree-order term from before (but using λ_4^{ren} and $(m^2)^{ren}$, not λ_4 and m^2).

This might seem a block to the applicability of the theory: to make calculations we need to know $(m^2)^{ren}$ and λ_4^{ren} , and we can only calculate them from m^2 and λ_4 via detailed knowledge of the cutoff mechanism and scale. And indeed this would be a block if we were presented with the theory by giving the original parameters m^2 and λ_4 (the so-called “bare parameters”) as a gift from God. But in fact, we determine the parameters through experiment — and what the experiment gives us is the renormalised parameters, not the bare parameters. The latter are related to the measured parameters through a cutoff-dependent expression, but we don’t in any case need them for calculations.

Furthermore, if we now include the higher-order terms $\lambda_6 x^6$ (etc) in $V(x)$, the result is that these terms:

⁹“Roughly speaking” because I have omitted, for expository simplicity, the fact that the field strength also gets renormalised.

1. Further renormalise m^2 and λ_4 , adding corrections that are functions of Λ ;
2. Make tree-order contributions to the calculation that are proportional to powers of L/Λ , where L is the lengthscale on which G is evaluated (and thus are extremely small if $L \gg \Lambda$);
3. Make loop-order contributions suppressed by successively larger powers of L/Λ ;
4. Have no effect on the dynamics in the long-distance limit *except* to renormalise m^2 and λ_4 .

We can now identify the renormalised parameters as two of the three parameters that (I claimed in the previous section) suffice to determine the theory up to IR-equivalence. The third parameter can be identified as the energy density of the ground state, related to the zero V_0 of the function V by an expression like

$$V_0^{ren} = V_0 + \gamma(1/\Lambda)^4 \tag{32}$$

but irrelevant to the physics except in the presence of gravity.

7 Scale-dependence, and particles again

I should stress one crucial feature of this renormalisation process: it is *scale-dependent*. There is a certain amount of freedom in how to divide out the contribution of higher-order terms in the perturbative expansion between (i) renormalising the bare parameters, and (ii) contributing corrections to the tree-order calculations. In practice this is usually done by picking some scale at which the tree-order calculation (expressed in terms of the renormalised parameters) is *exact*. Calculations made at lengths close to this scale are well-approximated by the methods of perturbation theory, but these methods become successively less effective at too-large or too-small lengths. The theory can still be understood as being *specified* by the parameters even at lengths far from the chosen scale — but the meaning of those parameters at that scale will be far from transparent.

To illustrate this, consider two examples from particle physics: quantum electrodynamics (QED) and quantum chromodynamics (QCD). In popular accounts, the former is the theory of electrons and photons, the latter the theory of quarks and gluons, but we will do better to think of the former as a theory of an electron *field* interacting with the photon field (i. e., quantized electromagnetic field) and the former as a theory of a quark and gluon field interacting.¹⁰ In the absence of an interaction term, though, the usual particle account can be recovered: the electron is the particle associated with the electron field, etc.

¹⁰Even this is an imperfect way to think, due to gauge freedom: the electron and photon fields, or the quark and gluon fields, jointly represent the underlying physics, and the split between them is to some degree gauge dependent; see Wallace (2014) for further discussion.

In both theories (and simplifying slightly), one of the relevant parameters (relevant in the sense of our previous discussion, that is) is the strength of the interaction between the fields (electron-photon, or quark-gluon); another, in the free-field limit, may be interpreted as the mass of the ‘matter’ particle, i. e. , the electron or quark. Schematically, let’s call these parameters λ and m , respectively.

In both cases, the values of λ and m are scale-dependent. For QED, λ decreases at larger lengthscales. For lengthscales very large compared to the cutoff length, it is small enough that the interactions may be treated as a small perturbation of the free-field theory. In this regime, the particle interpretation of the theory remains approximately valid, and we can meaningfully interpret the theory as a theory of electrons scattering off one another with their interactions mediated by photons. At shorter and shorter lengthscales (corresponding to higher and higher energies) the electron-photon interaction becomes stronger (the mass also changes, though the exact form of that change is not relevant here).

Let’s pause to reconsider the emergent status of electrons (originally discussed in section 3) in this context. Recall that the particles of a free-field theory are created from the ground state of the free-field Hamiltonian by applying creation operators. But the definitions of the creation operators, and of the free-field ground state, depend on the parameters of the theory: on m directly, and on λ indirectly via its role in the renormalisation process. So the one-particle Hilbert space constructed to analyse QED at high energies is a *different Hilbert space* from the one constructed at low energies. This ought to drive home the point that electrons cannot be thought of as fundamental building blocks of nature; they are simply a useful, but scale-relative, emergent feature of the underlying theory. But recall that this theory, too, should not be thought of as fundamental, given the tacit presence of the cutoff. In fact, QED actually imposes a minimum value for the cutoff length: the interaction strength λ continues to increase at smaller and smaller scales and eventually becomes infinite (the so-called “Landau pole”), indicating that QED ceases to be well-defined for cutoffs smaller than this length. To be sure, the Landau pole occurs at a length far shorter than the point at which external reasons (like gravity) would lead QED to fail as a physical description of the world, but it is strong reason to think that there is no genuine continuum version of QED.

In QCD, conversely, the interaction strength λ *decreases* at shorter lengthscales (i. e. , higher energies). This means that in high-energy physics, the field can be viewed as describing a collection of weakly-interacting quarks. At low energies this description breaks down entirely. We can still describe the field in terms of particles, but now they are different particles: protons, neutrons, and various mesons and other hadrons. These particles can be thought of *very* loosely as bound states of the quarks; a somewhat more accurate statement would be that they are excitations associated not with the quark field but with various symmetrised products of that field.

8 Symmetry and universality

Let's return again to the scalar field. In classical field theory, saying “this is a theory of an interacting scalar field” is nothing like enough to pick out a unique theory. Fundamentally different Lagrangians could be written down for a scalar field: some with φ^4 interactions, some with φ^6 interactions, some with combinations of both, and so forth. This is not merely a matter of specifying the coupling constants, but the actual functional form of the Lagrangian. Of course, any term written down will have to satisfy the symmetries of the scalar field (by definition; otherwise it would not be a scalar field), but specifying the symmetry is only the beginning of specifying the dynamics.

But we have seen that the situation is very different in quantum field theory. Firstly, the entire class of *renormalisable* interactions is finite and small: indeed, it basically contains the (renormalised) λ_4 parameter. So as far as interactions on scales large compared to the cutoff are concerned, there is only really one way to write down a nontrivial dynamics. Secondly, and more profoundly, even if we exclude from the Lagrangian some (renormalisable or non-renormalisable) interaction terms by setting the associated parameter λ_n to zero, the parameter will move away from zero again if we shift the cutoff. To say that, for instance, the φ^6 term in the QFT Lagrangian is absent is to say something *cutoff-dependent*, and hence something that is not really of physical significance given the effective-field-theory approach to understanding quantum field theories.

This is not confined to scalar field theories, of course. In full generality, writing down the symmetries of a quantum field theory pretty much fixes the form of its dynamics, up to a very small number of parameters. This phenomenon — often called *universality* — goes some way to explaining the central role of symmetry in contemporary theoretical physics: once the symmetries of a quantum field theory are known, its dynamics are pretty much specified. (See Batterman, this volume, and Williams, this volume, for more on this subject and its connection to the renormalisation group.)

9 Other features of quantum field theory

Pretty much everything I have discussed so far applies to *any* quantum field theory. In this section, I will mention some conceptually important features of QFT that apply in more specific theories. My focus will be on examples from particle physics; of necessity, the discussions will be brief, and I concentrate on examples which rely on specifically quantum-mechanical features of QFT (as opposed to, say, gauge theory, which plays a central role in the Standard Model but is to a large degree classical).

9.1 Lorentz covariance and the classification of particles

In particle physics (but not in condensed-matter physics), the imposition of Lorentz covariance places strong constraints on the form of a quantum field theory. Examples include:

Wigner’s classification of particles: If we ask, independent of their origin as excitations of a field, what quantum states deserve the name “particles”, we can argue — following Wigner (1939) — that they should correspond to irreducible representations of the Poincaré group. (A heuristic rationale: the one-particle subspace must transform under the Poincaré group; if it transforms reducibly, we can decompose it into a superposition of components that each transform irreducibly. I don’t know of a really careful conceptual analysis, though I make some suggestions in Wallace (2009).) And if we ask what linear field theories can be written down, we get the same result. (If we write down a field that transforms reducibly under the Poincaré group, the irreducible components get renormalised differently, becoming in effect different fields.) Either way, it seems at least highly plausible that the particle-like excitations of a field theory can be group-theoretically classified.

The result (skipping over some representations that seem unphysical) is:

- Particles are classified completely by their mass (which can be zero or positive) and by their spin, which can have any positive integer or integer-plus-half value. This classification coincides, so far, with the nonrelativistic version of the same approach, which classifies particles via irreducible representations of the Galilei group (Bargmann 1954). There are longstanding if somewhat contested arguments that particles of spin > 2 cannot consistently be associated to an interacting quantum field, but that lies beyond the group-theoretic analysis; see Bekaert, Boulanger, and Sundell (2012) and references therein.
- Particles of nonzero mass and spin s have a $2s+1$ -dimensional internal space, again as in nonrelativistic mechanics.
- Particles of mass zero and spin > 0 have a two-dimensional internal space (so that photons, for instance, are spin-one particles but have only two orthogonal spin states for a given momentum). There are subtle connections between gauge symmetry and this reduction of a massless particle’s internal degrees of freedom; again, they go beyond the group-theoretic analysis.

Antimatter: If a quantum field transforms under a representation of its internal symmetry group that has a natural complex structure (such as the standard representations of $U(1)$ or $SU(N)$), its one-particle Hilbert space separates naturally into matter and antimatter components. A consequence is that to all charged particles (and some uncharged particles, like the neutrino) is associated an antiparticle of the same mass but opposite charge and other quantum numbers. This is a purely relativistic

effect with no nonrelativistic analog; see Wallace (2009) or (for a discussion from an algebraic-quantum-field-theory viewpoint) Baker and Halvorson (2009).

Discrete symmetries and the CPT theorem: A relativistic quantum field theory might in principle have three discrete symmetries in addition to any continuous internal symmetries and the continuous Poincaré symmetries: Parity (reflection in space), Time reversal (reflection in time) and Charge conjugation (exchange of matter for antimatter). The *CPT theorem* (also known as the TCP, CTP, PCT, and PTC theorem — I don't recall ever seeing TPC) establishes that any quantum field theory has a symmetry which can be identified as the product of all three transformations: that is, the transformed field at x, t is a function of the untransformed field at $-x, -t$ and the symmetry exchanges matter and antimatter. There is no requirement that the individual transformations are symmetries or even that they are well-defined transformations on the theory's Hilbert space.

The Spin-Statistics theorem: See below.

9.2 The fermion/boson distinction

The scalar field theory I used above as an example is specified (formally) by a function from points of space to operators such that pairs of spatially separated operators commute; insofar as it can be treated as weakly interacting, it can be interpreted as a theory of bosonic particles. But it is also possible to construct a quantum field (either in solid-state physics or in particle physics) where spatially separated operators *anti*-commute. The resultant theory, in the weakly-interacting regime, can be analysed in terms of fermionic particles. Such field theories are called *fermionic*, by contrast with the *bosonic* fields we have focussed on so far.

The celebrated *spin-statistics theorem* establishes that a quantum field is bosonic if its associated particles have integer spin, and fermionic if it has (integer-plus-half) spin. In the Standard Model of particle physics, in particular, the fermionic fields are the quarks and leptons (electrons, neutrinos and their heavier variants), which have spin $1/2$; the bosonic fields are the force-carriers (gluons, photons, the W and Z bosons) with spin 1, and the Higgs field, with spin zero (and, depending how the Standard Model is defined, possibly also the graviton, with spin 2). The theorem holds for (relativistic) spacetimes of dimension 3 or higher; conversely in two spacetime dimensions, there are examples where the same quantum field can be analysed in terms of fermions in one regime and bosons in another Coleman (1985, ch.6).

Formally speaking, fermionic fields can be treated very much like bosonic fields (other than a large number of minus signs that have to be kept track of in calculations). Conceptually, they seem dissimilar in important respects: for instance, the quantum state of a fermionic field cannot in any straightforward sense be understood as a wavefunctional on a space of classical field configura-

tions. To the best of my knowledge there has been rather little discussion of fermionic fields in the philosophy literature on QFT.

9.3 Infrared divergences and the large-volume limit

In my presentation of QFT so far I have assumed a spatially finite system. That seems reasonable enough in most applications of QFT, whether in solid-state physics (where the physical system to which QFT is applied is manifestly finite) and in particle physics (where the physical processes of interest are normally confined to a region of finite — and usually pretty small — extent). In each case, the physically reasonable implication is that the size R of the finite region will have little or no effect on the physics, provided it is much larger than the scales which we are studying.

However, in some physical applications — notably those involving zero-mass particles — this is not the case: loop-order calculations of physical quantities include terms that depend on R , and that diverge as $R \rightarrow \infty$. These *infra-red* divergences, analogously with the ultra-violet (short-range) divergences we have already considered, can be handled in one of two ways: either by developing, rigorously, a quantum theory of genuinely infinite systems, or by keeping R fixed but large (or otherwise regularising the divergences, e. g. by adding a small mass term) and absorbing R -dependent terms into a renormalisation of the physical parameters. The latter is the route taken in mainstream physics, and delivers useful physical insight into the origin of the infrared divergences: they occur because zero-mass particles can be created with arbitrarily low energy, and so a physical particle is surrounded by a cloud of very many — in the $R \rightarrow \infty$, infinitely many — such particles.

The infrared and ultraviolet divergences are *dis*analogous in one important way, though: while the short-distance cutoff is normally taken to describe a *physical* cutoff above which the theory cannot be trusted, the long-distance cutoff just reflects the finite extent of the part of physical reality we are trying to model, together with a physical assumption that the details of the boundary conditions on that region aren't physically significant for shorter-distance physics. Not unrelatedly, while (in my biased opinion) not much of conceptual value has been gained by trying to define continuum QFT on arbitrarily short lengthscales, mathematically rigorous considerations of spatially infinite QFTs have been conceptually very informative (see Ruetsche (2011) and references therein).

9.4 Symmetries and symmetry breaking

The duality between field and particle (in the noninteracting limit) suggests that a symmetry of a quantum field theory will be represented as a symmetry of the particle(s) associated with that theory, and so it often turns out. For instance, the quark field has the group $SU(3)$ as a dynamical symmetry, and transforms as the three-dimensional complex representation of that symmetry; correspondingly, quarks as particles have a three-complex-dimensional space of

internal degrees of freedom. (This is usually described in popular accounts as there being three sorts of quarks — red, green and blue — but in fact ‘red’, ‘green’ and ‘blue’ are just arbitrary bases in the quark’s internal space, and $(1/\sqrt{2})(\text{red} + \text{blue})$ or $(1/\sqrt{2})(\text{blue} - i \text{green})$ are just as valid as quark states.) Similarly, there is an approximate symmetry called ‘isospin’ in the effective field theory of the hadrons (the low-energy states of quantum chromodynamics) that shows up in the approximately equal masses of the proton and neutron.

However, this straightforward relation between field and particle symmetries only works if the theory’s ground state $|\Omega\rangle$ is invariant under the symmetry group. If this is not the case — which entails that ‘the’ ground state is degenerate — then we can construct a different set of particles by acting with creation operators on each ground state. A symmetry transformation under which the ground state is not invariant will transform between different (though indiscernible) sets of particles, rather than transforming a set of particle states among themselves. As a consequence, the symmetry is not visible in the dynamics of the particles, and so is said to be “spontaneously broken” (“hidden” might be a better term).

A spontaneously broken symmetry shows up in the phenomenology via constraints between the measured coupling constants. In the case of a global symmetry (one that acts the same way at every point of spacetime) it also shows up via the presence of a ‘Goldstone boson’, a massless particle which always occurs in the particle spectrum of such a system. (In the long-wavelength limit it corresponds to the symmetry that maps from one ground state to another.) For instance, the spatial translation symmetry is spontaneously broken in condensed-matter physics by the lattice structure of the ground state of a solid body; the associated Goldstone boson is the phonon, the quantum of vibration. The pion (a two-quark low-energy excitation of the quark field) can be understood as a Goldstone boson of a spontaneously broken approximate symmetry of the hadrons; as such, its mass is low but not zero.

The physics is somewhat subtler when the spontaneously broken symmetry is local. In that situation, the *Higgs mechanism* causes the gauge fields associated with the local symmetry (which are normally massless) to acquire mass. The mechanism is sometimes heuristically described as the vector boson “eating” the Goldstone boson; there is some controversy in foundations of physics as to what a better description would be (Earman 2004; Struyve 2011; Friederich 2013).

There are conceptually interesting mathematical subtleties involved with spontaneous symmetry breaking (global or local). In finite systems, it is known that the ground state is not genuinely degenerate because of the possibility of tunnelling between symmetry-related states. The infinite-volume limit is required for true degeneracy, and in that limit some of the other assumptions of the theory break down. For discussion, see Ruetsche (1998) and references therein.

9.5 Non-renormalisable interactions in physics

My account of renormalisation theory might give the impression that *non*-renormalisable interactions have no part to play in physics: either we are at energy levels very low compared to the cutoff (in which case their influence is swamped by the renormalisable interactions) or we are relatively close to the cutoff (in which case the theory is not reliable in any case). This is not quite correct. In particular, suppose we have a field theory with *no* renormalisable interactions. Then we would predict that:

1. The interactions will be dominated by that nonrenormalisable term which drops off least rapidly at greater lengthscales;
2. The interaction strength will be suppressed by some power of L/Λ , where L is the characteristic lengthscale of the interaction and Λ is the cutoff;
3. In particular, the interactions will be very weak at large lengths.

I give two important examples of this in practice. The first is the so-called *four-fermion theory*, where the only field is an uncharged spin-half fermionic field (taken to represent neutrinos, say). In the Standard Model, neutrino interactions are mediated by the W and Z bosons, but in situations where the energy levels of interactions are much lower than the W and Z masses, we can regard those masses as a cutoff on an effective field theory where the neutrinos interact directly. Any direct interaction between fermion fields is nonrenormalisable; the lowest-order such term is a two-particle scattering term that allows pairs of neutrinos to scatter off one another. We would predict that this interaction is very small at interaction scales large compared to the Compton wavelength of the W and Z . And indeed this is what we find: the force that mediates neutrino interactions is called the Weak Interaction in particle-physics phenomenology.

The second example occurs in quantum gravity. There is no renormalisable interaction between the metric field and matter (or between the metric field and itself) but the lowest-order nonrenormalisable interaction term is the Einstein-Hilbert action term of general relativity (together with a cosmological-constant term, of which more below). The extreme weakness of the gravitational field is then explained by the fact that gravitational phenomena are studied on scales extremely large compared to the Planck length at which we expect full quantum gravity to impose a cutoff on field theory.

9.6 Quantum field theory on curved spacetime

Most of the theoretical development (and almost all the experimental data) in QFT assumes flat spacetime, either Newtonian (for condensed-matter physics) or Minkowski (for particle physics). But quantum field theory can be formulated in at least some nonflat spacetimes, and doing so is the basis of important work in cosmology and in the physics of black holes, in particular in one of the most celebrated and surprising discoveries of the past forty years: Hawking's discovery

of black hole radiation (see Wallace (2017a) and references therein for further details).

10 Outstanding questions of particle physics

I have tried to indicate in this article just how successful and powerful quantum field theory is. But — optimistically from the point of view of exciting new physics — there remain some deep puzzles in the theory, in particular in its applications in particle physics. Here I identify three such puzzles; there are many others, but these are perhaps the most visible in contemporary physics. The first two are discussed in more detail by Barnes, this volume, and Williams, this volume.

10.1 Fine-tuning of the Higgs mass

My discussion of the scalar field glossed over one subtlety. I noted that the mass of the field is renormalised, so that the empirically-accessible mass is related to the ‘bare’ mass by a cutoff-dependent term. In general in quantum field theory, renormalisations like this are logarithmic in the cutoff, as in expression (31): the mass rescaling in a spin-half field, for instance, has abstract form

$$m^{ren} = m_0(1 + A \log(1/\Lambda)), \quad (33)$$

for some dimensionless A not usually too far from unity. Because of the slow scaling of the logarithm function, this means that if the bare mass is much less than the cutoff energy, so will the renormalised mass. But in the case of the scalar particle, the mass rescales according to equation (30), with additive corrections to the bare mass proportional to $1/\Lambda^2$. This means that we would expect the renormalised mass of a scalar particle to be of the same order as the cutoff energy, whatever the bare mass might be.

However, the Higgs boson — which, in the simplest versions of the Standard Model, is the particle associated with a scalar field — has a mass far below whatever the Standard Model’s cutoff energy is. This is not a contradiction in the theory — a sufficiently careful choice of the bare mass can yield whatever value we like for the renormalised mass — but it seems to involve rather unattractive fine-tuning of the theory’s parameters.

10.2 Fine-tuning of the cosmological constant

We saw in section 3 that the formal energy density of the vacuum of a free field theory is infinite. Adding a cutoff tames the infinity but still leaves a very large finite term, of order $(1/\Lambda)^4$. Interactions add further contributions, also of order $(1/\Lambda)^4$. So the expression for the total energy density ρ_{vac} , schematically, is

$$\rho_{vac} = V(0) + \text{free-field contribution} + \text{interaction renormalisation} \quad (34)$$

where $V(0)$, the classical vacuum energy density, is the value of the Lagrangian at zero field.

In nongravitational physics, none of this matters: the energy density has no effect on the physics and can be set to whatever value we find convenient (usually zero) by an appropriate choice of $V(0)$ without empirical consequence. But the stress-energy tensor of the field has form

$$T_{\mu\nu} = T_{\mu\nu}^{ren} + g_{\mu\nu}\rho_{vac} \quad (35)$$

where the ‘renormalised’ stress-energy tensor $T_{\mu\nu}^{ren}$ is defined by the requirement that it vanishes for the ground state. And the Einstein field equation is

$$G_{\mu\nu} + g_{\mu\nu}C = 8\pi GT_{\mu\nu}, \quad (36)$$

where C is the cosmological constant. So it looks as if the vacuum expectation value should make an enormous contribution to the cosmological constant, albeit it’s not fully clear how to interpret the field equation without a quantum theory of gravity (the simplest approach would be to take the right hand side of the equation to be the *expectation value* of the stress-energy tensor; see Wald (1994) for further discussion of this hybrid theory). That contribution is more than 100 orders of magnitude larger than the observed value of the cosmological constant. Again, this is not a contradiction, as we can tune the classical vacuum energy density (or, equivalently, the bare cosmological constant¹¹) to whatever value we like; as with the Higgs mass, the problem is the extreme fine-tuning of the parameters that seems to be required.

10.3 Quantum gravity

A common claim about quantum gravity is that it is a puzzle that arises from the incompatibility of quantum mechanics with general relativity, and so would arise (in principle) whenever quantum effects and gravity apply simultaneously. This is not the perspective of most quantum field theorists: to them, the metric field is at least perturbatively perfectly well-behaved — albeit non-renormalisable — and can be handled in the effective-field theory framework. (We have already seen that the extreme weakness of the gravitational field can be understood in effective-field-theory terms as a consequence of the non-renormalisability of the Einstein-Hilbert action.) Indeed, exactly this formalism is applied in the quantum-fluctuation calculations that underpin our theoretical models of the cosmic microwave background radiation, and so ‘quantum gravity’, in the sense of a quantum-field-theoretic understanding of general relativity, has already passed at least a crude experimental test. (For the formalism, see Weinberg (2008, ch.10); for conceptual discussion, see Wallace (2016)).

¹¹There is an odd tendency (which I observe mostly in conversation) for philosophers of physics to draw a sharp distinction between a bare constant on the left-hand side of the Einstein field equation (where it is taken to pertain to spacetime) and the negative of the same constant on the right-hand side of the equation (where it is taken to pertain to matter) But (as I once heard Sean Carroll remark in response to one such comment) it is permissible to move terms from one side of an equation to the other!

What most quantum field theorists mean by ‘quantum gravity’ is the breakdown of effective-field-theory general relativity — and, it is usually assumed, the rest of the Standard Model of particle physics — around the Planck length, at $\sim 10^{-34}$ metres. A quantum theory of gravity, in this sense, would be a genuinely finite theory from which particle physics, and general relativity, would emerge as effective field theories in appropriate long-distance regimes. (It is generally assumed that this theory would also tame the formal infinities that occur due to singularities in classical general relativity.)

Unfortunately, the great insensitivity of an effective field theory to the physical details of its high-energy cutoff, and the sheer energy scale of that cutoff, makes it very hard to gain evidence about the details of that theory: it “has imprinted few traces on physics below the Planck energy” (Bousso 2002, p.2). There is not the least hope that particle accelerators will ever probe the Planck scale; so far, only early-universe cosmology seems to hold out any hope of giving us observational access to quantum gravity. One reason why string theory, loop quantum gravity and other would-be quantum gravity programs have paid so much attention to the thermodynamics of black holes is that black hole radiation can be understood (via different choices of foliation) *either* as an effective-field theory result occurring at tolerably low energies, and so reasonably well understood, *or* as a fully quantum-gravitational effect; as such, black holes are a highly non-trivial consistency check on a putative quantum theory of gravity, above and beyond that given by ordinary effective-field-theory methods.¹²

11 Philosophical morals

Quantum field theory, as the language in which a huge part of modern physics is written, is a natural setting for all manner of detailed questions in philosophy and foundations of physics, from the search for relativistic versions of dynamical-collapse and hidden-variable theories¹³ to the correct understanding of the gauge principle. But in this last section I want to draw a more general moral. Contemporary philosophy of physics is for the most part focussed on so-called *fundamental* physics: that is, on those parts of physics which describe the world in full detail and at every scale, not simply in some emergent, approximate way in some regimes.¹⁴ We currently have no fully worked-through fundamental physical theory. Indeed, we never have: there has never been a time when physicists had plausible ground to believe that they possessed any such theory. (Perhaps the closest point was at the turn of the twentieth century, after the development of electromagnetism and thermodynamics, and before the twin revolutions of relativity and quantum theory.) So in practice philosophy of physics proceeds by taking a theory like classical or quantum particle mechanics, or classical general relativity, and studying it under the fiction that it is fundamental. It is tempting to imagine studying QFT on that basis too.

¹²See Curiel (this volume), Wallace (2017a, Wallace (2017b) and references therein.

¹³See Lewis (this volume) and Tumulka (this volume) for further discussion.

¹⁴For more on this point, see French, this volume, and Wallace (2018b).

But quantum field theory is not that kind of theory. For all that our best and deepest physics is cast in its framework, it is *by its own nature* non-fundamental. An effective field theory – and, recall, all empirically-successful quantum field theories are effective field theories — is defined through the methods of cutoff and renormalisation, and does not even purport to fully describe the world. It is a remarkable irony that the Standard Model at one and the same time is the nearest we have ever come to a Theory of Everything, and is uninterpretable even in fiction as an exact description of the world. It is further irony that the theory itself tells us that it is compatible with an indefinitely large range of ways in which the deeper-level physics might be specified. The fact that such a theory is (most physicists assume) quantum-mechanical allows us to say *something* about it, but quantum mechanics, like classical mechanics, is a framework theory and all manner of different theories fit within it.

Quantum field theory is a reminder to philosophers that physics, like other sciences, is hardly ever in the business of formulating theories that purport to describe the world on all scales. They who wish to learn ontology from our best science, in the era of effective field theories, have two choices: recognise that deep and interesting metaphysical questions come up at all length-scales in physics and are not confined to the ‘fundamental’, or remain silent and wait, and hope, for a truly fundamental theory in the physics that is to come.

12 Further reading

There are many textbooks on quantum field theory. Probably the best conceptually-focused book-length account is Duncan (2012); other books that I have found helpful include Zee (2003), Banks (2008) (insightful but very terse), Peskin and Schroeder (1995) (the standard graduate-level textbook), and Weinberg (1995a) (not recommended as a first introduction). But tastes vary; get hold of several and see what suits your learning style. Coleman (1985) is not a textbook, exactly, but is highly insightful on a number of conceptual issues in QFT. All of these textbooks focus primarily on particle-physics applications of QFT.

For discussions of QFT more focussed on solid-state physics, see Abrikosov *et al* (1963) (old but classic) or Altland and Simons (2010) (much more up to date). For a more detailed account of the Standard Model, see Cheng and Li (1984) or Donoghue *et al* (2014).

The methods of the renormalisation group extend beyond QFT as understood as a quantum theory of the continuum and also have deep significance in classical statistical mechanics; for a very clear presentation of the renormalisation group in this context, see Binney *et al* (1992). (One word of caution here: do not confuse the *formal* analogy between classical statistical mechanics and QFT, with the *physical* similarities between QFT as applied to condensed-matter systems and to particle physics.)

My favorite reference on quantum field theory in *curved* spacetime is Jacobson (2005); Wald (1994) is also excellent. Wald also provides an introduction to the broader issues of black hole thermodynamics, albeit now a little out of date;

for more recent reviews from various perspectives, see Harlow (2016), Hartman (2015) and Grumiller *et al* (2015). For philosophical considerations (going rather beyond QFT in each case), see Curiel, this volume, Belot *et al* (1999) and Wallace (2017a, 2017b).

For a philosophical defense (as opposed to simply an exposition, as here) of ‘mainstream’ effective-field-theory QFT against the more rigorous, but empirically less successful, approach of trying to define continuum quantum mechanics exactly (nowadays usually called ‘algebraic quantum field theory’, or ‘AQFT’), see Wallace (2006, 2011); for a response from the AQFT perspective, see Fraser (2009, 2011) (see also Baker (2016) for observations on the debate). Ruetsche (2011) provides a general introduction to AQFT methods and a route into the broader literature on philosophy of AQFT; other (more advanced) introductions are Haag (1996) and Halvorson (2007).

The question of particles in QFT has been extensively discussed in the philosophy literature (albeit mostly disjoint from the ‘emergent’ attitude to particles I advocated in section 3; see Fraser, this volume, and references therein).

Acknowledgements

Thanks to Doreen Fraser and Alastair Wilson for detailed and helpful comments on earlier drafts of this paper.

References

- Abrikosov, A. A., L. P. Gorkov, and I. E. Dzyalohinski (1963). *Methods of Quantum Field Theory in Statistical Physics*. New York: Dover. Revised English edition; translated and edited by R. A. Silverman.
- Altland, A. and B. D. Simons (2010). *Condensed Matter Field Theory* (2nd ed.). Cambridge: Cambridge University Press.
- Baker, D. (2016). The philosophy of quantum field theory. Oxford Handbooks Online; DOI: 10.1093/oxfordhb/9780199935314.013.33.
- Baker, D. and H. Halvorson (2009). Antimatter. *British Journal for the Philosophy of Science* 60, 1–29.
- Banks, T. (2008). *Modern Quantum Field Theory: a Concise Introduction*. Cambridge: Cambridge University Press.
- Bargmann, V. (1954). On unitary ray representations of continuous groups. *Annals of Mathematics* 59, 1–46.
- Bekaert, X., N. Boulanger, and P. A. Sundell (2012). How higher-spin gravity surpasses the spin-two barrier. *Reviews of Modern Physics* 84, 987.
- Belot, G., J. Earman, and L. Ruetsche (1999). The Hawking information loss paradox: the anatomy of a controversy. *British Journal for the Philosophy of Science* 50, 189–229.

- Binney, J. J., N. J. Dowrick, A. J. Fisher, and M. E. J. Newman (1992). *The Theory of Critical Phenomena : an introduction to the renormalisation group*. Oxford: Oxford University Press.
- Bousso, R. (2002). The holographic principle. *Reviews of Modern Physics* 74, 825.
- Cheng, T.-P. and L.-F. Li (1984). *Gauge Theory of Elementary Particle Physics*. Oxford, UK: Oxford University Press.
- Coleman, S. (1985). *Aspects of Symmetry: selected Erice lectures*. Cambridge, UK: Cambridge University Press.
- Donoghue, J. F., E. Golowich, and B. R. Holstein (2014). *Dynamics of the Standard Model (Second Edition)*. Cambridge: Cambridge University Press.
- Duncan, T. (2012). *The Conceptual Framework of Quantum Field Theory*. Oxford: Oxford University Press.
- Earman, J. (2004). Laws, symmetry, and symmetry breaking: Invariance, conservation principles, and objectivity. *Philosophy of Science* 71, 1227–1241.
- Fraser, D. (2009). Quantum field theory: Underdetermination, inconsistency, and idealization. *Philosophy of Science* 76, 536–567. Forthcoming in *Philosophy of Science*; available online at philsci-archival.pitt.edu .
- Fraser, D. (2011). How to take particle physics seriously: A further defence of axiomatic quantum field theory. *Studies in the History and Philosophy of Modern Physics* 42, 126–135.
- Friederich, S. (2013). Gauge symmetry breaking in gauge theories — in search of clarification. *European Journal for Philosophy of Science* 3, 157–182.
- Grumiller, D., R. McNees, and J. Salzer (2015). Black holes and thermodynamics: the first half century. In X. Calmet (Ed.), *Quantum Aspects of Black Holes: Fundamental Theories of Physics 178*, pp. 27–70. Springer.
- Haag, R. (1996). *Local Quantum Theory: Fields, Particles, Algebras*. Berlin: Springer-Verlag.
- Halvorson, H. (2007). Algebraic quantum field theory. In J. Butterfield and J. Earman (Eds.), *Handbook of the Philosophy of Science: Philosophy of Physics, Part A*. Boston: Elsevier.
- Harlow, D. (2016). Jerusalem lectures on black holes and quantum information. *Reviews of Modern Physics* 88, 015002.
- Hartman, T. (2015). Lectures on quantum gravity and black holes. <http://www.hartmanhep.net/topics2015/gravity-lectures.pdf>.
- Jacobson, T. (2005). Introduction to quantum fields in curved spacetime and the Hawking effect. In A. Gomberoff and D. Marolf (Eds.), *Lectures on Quantum Gravity*, pp. 39–89. Springer. Available online at <https://arxiv.org/abs/gr-qc/0308048>.

- Kambe, T. (2007). *Elementary Fluid Mechanics*. Singapore: World Scientific.
- Kenyon, R. A. (1960). *Principles of Fluid Mechanics*. New York: The Ronald Press Company.
- Peskin, M. E. and D. V. Schroeder (1995). *An introduction to Quantum Field Theory*. Reading, Massachusetts: Addison-Wesley.
- Rosaler, J. (2015). “formal” versus “empirical” approaches to quantum-classical reduction. *Topoi* 34, 325–328.
- Ruetsche, L. (1998). How Close is ‘Close Enough’? In D. Dieks and P. E. Vermaas (Eds.), *The Modal Interpretation of Quantum Mechanics*, pp. 223–240. Dordrecht: Kluwer Academic Publishers.
- Ruetsche, L. (2011). *Interpreting Quantum Theories*. Oxford: Oxford University Press.
- Saunders, S. (1992). Locality, Complex Numbers and Relativistic Quantum Theory. *Philosophy of Science Association 1992* 1, 365–380.
- Struyve (2011). Gauge-invariant accounts of the Higgs mechanism. *Studies in the History and Philosophy of Modern Physics* 42, 226–236.
- Wald, R. M. (1994). *Quantum Field Theory in Curved Spacetime and Black Hole Thermodynamics*. Chicago: University of Chicago Press.
- Wallace, D. (2006). In defence of naiveté: The conceptual status of Lagrangian quantum field theory. *Synthese* 151, 33–80.
- Wallace, D. (2009). QFT, antimatter, and symmetry. *Studies in the History and Philosophy of Modern Physics* 40, 209–222.
- Wallace, D. (2011). Taking particle physics seriously: a critique of the algebraic approach to quantum field theory. *Studies in the History and Philosophy of Modern Physics* 42, 116–125.
- Wallace, D. (2012). *The Emergent Multiverse: Quantum Theory according to the Everett Interpretation*. Oxford: Oxford University Press.
- Wallace, D. (2014). Deflating the Aharonov-Bohm effect. Forthcoming; available online at <https://arxiv.org/abs/1407.5073>.
- Wallace, D. (2016). Interpreting the quantum mechanics of cosmology. Forthcoming in “Guide to the Philosophy of Cosmology”, Ijjas and Loewer (ed.).
- Wallace, D. (2017a). The case for black hole thermodynamics, part I: phenomenological thermodynamics. <https://arxiv.org/abs/1710.02724>.
- Wallace, D. (2017b). The case for black hole thermodynamics, part II: statistical mechanics. <https://arxiv.org/abs/1710.02725>.
- Wallace, D. (2018a). Lessons from realistic physics for the metaphysics of quantum theory. *Synthese*.

- Wallace, D. (2018b). On the plurality of quantum theories: Quantum theory as a framework, and its implications for the quantum measurement problem. Forthcoming in S. French and J. Saatsi (eds.), *Scientific Metaphysics and the Quantum* (Oxford: OUP, forthcoming).
- Weinberg, S. (1995a). *The Quantum Theory of Fields, Volume I: Foundations*. Cambridge: Cambridge University Press.
- Weinberg, S. (1995b). *The Quantum Theory of Fields, Volume II: Modern Applications*. Cambridge: Cambridge University Press.
- Weinberg, S. (2008). *Cosmology*. Oxford: Oxford University Press.
- Wigner, E. (1939). On unitary representations of the inhomogenous Lorentz group. *Annals of Mathematics* 40, 149.
- Zee, A. (2003). *Quantum Field Theory in a Nutshell*. Princeton: Princeton University Press.