Contents

1

Introduction to Field Theory

The purpose of this book is twofold. Here I will introduce field theory as a framework for the study of systems with a very large number of degrees of freedom, $N \to \infty$. And I will also introduce and develop the tools that will allow us to treat such systems. Systems that involve a large (in fact, infinite) number of *coupled* degrees of freedom arise in many areas of physics, notably in high-energy and in condensed matter physics, among others. Although the physical meanings of these systems and their symmetries are quite different, they actually have much more in common than it may seem at first glance. Thus, we will discuss, on the same footing, the properties of relativistic quantum field theories, classical statistical mechanical systems, and condensed matter systems at finite temperature. This is a very broad field of study, and we will not be able to cover each area in great depth. Nevertheless, we will learn that it is often the case that what is clear in one context can be used to expand our knowledge in a different physical setting. We will focus on a few unifying themes, such as the construction of the ground state (the "vacuum"), the role of quantum fluctuations, collective behavior, and the response of these systems to weak external perturbations.

1.1 Examples of fields in physics

1.1.1 The electromagnetic field

Let us consider a very large box of linear size $L \to \infty$ and the electromagnetic field enclosed inside it. At each point in space $\boldsymbol{x}$, we can define a *vector* (which is a function of time as well) $\boldsymbol{A}(\boldsymbol{x}, t)$ and a *scalar* $A_0(\boldsymbol{x}, t)$. These are the vector and scalar potentials. The physically observable electric field $\boldsymbol{E}(\boldsymbol{x}, t)$ and the magnetic field $\boldsymbol{B}(\boldsymbol{x}, t)$ are defined in the usual way:

$$\boldsymbol{B}(\boldsymbol{x}, t) = \boldsymbol{\nabla} \times \boldsymbol{A}(\boldsymbol{x}, t), \qquad \boldsymbol{E}(\boldsymbol{x}, t) = -\frac{1}{c}\frac{\partial \boldsymbol{A}}{\partial t}(\boldsymbol{x}, t) - \boldsymbol{\nabla} A_0(\boldsymbol{x}, t) \tag{1.1}$$

The time evolution of this dynamical system is determined by a *local* Lagrangian density (which we will consider in section 2.6). The equations of motion are just the Maxwell equations. Let us define the 4-vector field

$$A^\mu(x) = \left(A^0(x), \boldsymbol{A}(x)\right), \qquad A^0 \equiv A_0 \tag{1.2}$$

where $\mu = 0, 1, 2, 3$ are the time and space components. Here x stands for the 4-vector

$$x^\mu = (ct, \boldsymbol{x}) \tag{1.3}$$

To every point x^μ of Minkowski spacetime $\mathcal{M}$ we associate a value of the vector potential A^μ. The vector potentials are ordered sets of four real numbers and hence are elements of $\mathbb{R}^4$. Thus a field configuration can be viewed as a mapping of the Minkowski spacetime $\mathcal{M}$ onto $\mathbb{R}^4$,

$$A^\mu : \mathcal{M} \mapsto \mathbb{R}^4 \tag{1.4}$$

Since spacetime is continuous, we need an infinite number of 4-vectors to specify a configuration of the electromagnetic field, even if the box were finite (which it is not). Thus we have a infinite number of degrees of freedom for two reasons: spacetime is both continuous and infinite.

1.1.2 The elastic field of a solid

Consider a three-dimensional crystal. A configuration of the system can be described by the set of positions of its atoms relative to their equilibrium state (i.e., the set of deformation vectors $\boldsymbol{d}$ at every time t). Lattices are labeled by ordered sets of three integers and are equivalent to the set

$$\mathbb{Z}^3 = \mathbb{Z} \times \mathbb{Z} \times \mathbb{Z} \tag{1.5}$$

whereas deformations are given by sets of three real numbers and are elements of $\mathbb{R}^3$. Hence a *crystal configuration* is a mapping

$$\boldsymbol{d} : \mathbb{Z}^3 \times \mathbb{R} \mapsto \mathbb{R}^3 \tag{1.6}$$

At length scales ℓ, which are large compared to the lattice spacing a but small compared to the linear size L of the system, we can replace the lattice $\mathbb{Z}^3$ by a *continuum* description, in which the crystal is replaced by a continuum three-dimensional Euclidean space $\mathbb{R}^3$. Thus the dynamics of the crystal requires a four-dimensional spacetime $\mathbb{R}^3 \times \mathbb{R} = \mathbb{R}^4$. Hence the configuration space becomes the set of continuous mappings

$$\boldsymbol{d} : \mathbb{R}^4 \mapsto \mathbb{R}^3 \tag{1.7}$$

In this continuum description, the dynamics of the crystal is specified in terms of the displacement vector field $\boldsymbol{d}(\boldsymbol{x}, t)$ and its time derivatives, the velocities $\frac{\partial \boldsymbol{d}}{\partial t}(\boldsymbol{x}, t)$, which define the mechanical state of the system. This is the starting point of the theory of elasticity. The displacement field $\boldsymbol{d}$ is the elastic field of the crystal.

1.1.3 The order-parameter field of a ferromagnet

Let us now consider a ferromagnet. This is a physical system, usually a solid, in which there is a local average magnetization field $\boldsymbol{M}(\boldsymbol{x})$ in the vicinity of a point $\boldsymbol{x}$. The local magnetization is simply the sum of the local magnetic moments of each atom in the neighborhood of $\boldsymbol{x}$. At scales long compared to microscopic distances (the interatomic spacing a), $\boldsymbol{M}(\boldsymbol{x})$ is a continuous real vector field. In some situations of interest, the magnitude of the local moment does not fluctuate, but its local orientation does. Hence, the local state of the system is specified locally by a three-component unit vector $\boldsymbol{n}$. Since the set of unit vectors is in one-to-one correspondence with the points on a sphere S^2, the configuration space

is equivalent (isomorphic) to the sets of mappings of Euclidean three-dimensional space onto S^2,

$$\boldsymbol{n} : \mathbb{R}^3 \mapsto S^2 \tag{1.8}$$

In an ordered state, the individual magnetic moments become spontaneously oriented along some direction. For this reason, the field $\boldsymbol{n}$ is usually said to be an *order-parameter field*. In the theory of phase transitions, the order-parameter field represents the important degrees of freedom of the physical system (i.e., the degrees of freedom that drive the phase transition).

1.1.4 Hydrodynamics of a charged fluid

Charged fluids can be described in terms of hydrodynamics. In hydrodynamics, one specifies the charge density $\rho(\boldsymbol{x}, t)$ and the current density $\boldsymbol{j}(\boldsymbol{x}, t)$ near a spacetime point x^μ. The charge and current densities can be represented in terms of the 4-vector

$$j^\mu(x) = \big(c\rho(\boldsymbol{x}, t), \boldsymbol{j}(\boldsymbol{x}, t)\big) \tag{1.9}$$

where c is a suitably chosen speed (generally *not* the speed of light!). Clearly, the configuration space is the set of maps

$$j^\mu : \mathbb{R}^4 \mapsto \mathbb{R}^4 \tag{1.10}$$

In general, we will be interested both in the dynamical evolution of such systems and in their large-scale (thermodynamic) properties. Thus, we will need to determine how a system that, at some time t_0, is in some initial state, manages to evolve to some other state after time T. In classical mechanics, the dynamics of any physical system can be described in terms of a *Lagrangian*. The Lagrangian is a *local functional* of the field and of its space and time derivatives. "Local" here means that the equations of motion can be expressed in terms of partial differential equations. In other words, we do not allow for "action-at-a-distance," but only for local evolution. Similarly, the thermodynamic properties of these systems are governed by a local energy functional, the *Hamiltonian*. That the dynamics is determined by a Lagrangian means that the field itself is regarded as a mechanical system, to which the standard laws of classical mechanics apply. Here the wave equations of the fluid are the equations of motion of the field. This point of view will also tell us how to quantize a field theory.

1.2 Why quantum field theory?

From a historical point of view, quantum field theory (QFT) arose as an outgrowth of research in the fields of nuclear and particle physics. In particular, Dirac's theory of electrons and positrons was, perhaps, the first QFT. Nowadays, QFT is used, both as a picture and as a tool, in a wide range of areas of physics. In this book, I will not follow the historical path of the way QFT was developed. By and large, it was a process of trial and error in which the results had to be reinterpreted a posteriori. The introduction of QFT as the general framework of particle physics implied that the concept of *particle* had to be understood as an *excitation* of a *field*. Thus, *photons* become the quantized excitations of the electromagnetic field with particle-like properties (such as momentum), as anticipated by Einstein's 1905

paper on the photoelectric effect. Dirac's theory of the electron implied that even such "conventional" particles should also be understood as the *excitations of a field*.

The main motivation of these developments was the need to reconcile, or *unify*, quantum mechanics with special relativity. In addition, the experimental discoveries of the spin of the electron and of electron-positron creation by photons showed that not only was the Schrödinger equation inadequate to describe such physical phenomena, but the very notion of a particle itself had to be revised.

Indeed, let us consider the Schrödinger equation

$$H\Psi = i\hbar\frac{\partial\Psi}{\partial t} \tag{1.11}$$

where H is the Hamiltonian

$$H = \frac{\hat{\boldsymbol{p}}^2}{2m} + V(\boldsymbol{x}) \tag{1.12}$$

and $\hat{\boldsymbol{p}}$ is the momentum represented as a differential operator

$$\hat{\boldsymbol{p}} = \frac{\hbar}{i}\boldsymbol{\nabla} \tag{1.13}$$

acting on the Hilbert space of wave functions $\Psi(\boldsymbol{x})$.

The Schrödinger equation is invariant under Galilean transformations, provided the potential $V(\boldsymbol{x})$ is constant, but not under general Lorentz transformations. Hence, quantum mechanics, as described by the Schrödinger equation, is not compatible with the requirement that the description of physical phenomena must be identical for all inertial observers. In addition, it cannot describe pair-creation processes, since in the nonrelativistic Schrödinger equation, the number of particles is strictly conserved.

Back in the late 1920s, two apparently opposite approaches were proposed to solve these problems. We will see that these approaches actually do not exclude each other. The first approach was to stick to the basic structure of "particle" quantum mechanics and to write down a relativistically invariant version of the Schrödinger equation. Since in special relativity, the natural Lorentz scalar involving the energy E of a particle of mass m is $E^2 - (\boldsymbol{p}^2c^2 + m^2c^4)$, it was proposed that the "wave functions" should be solutions of the equation (the "square" of the energy)

$$\left[\left(i\hbar\frac{\partial}{\partial t}\right)^2 - \left(\left(\frac{\hbar c}{i}\boldsymbol{\nabla}\right)^2 + m^2c^4\right)\right]\Psi(\boldsymbol{x},t) = 0 \tag{1.14}$$

This is the Klein-Gordon equation. This equation is invariant under the Lorentz transformations

$$x^\mu = \Lambda^{\mu,\nu}x'_\nu \qquad x^\mu = (x_0, \boldsymbol{x}) \tag{1.15}$$

provided that the "wave function" $\Psi(x)$ is also a scalar (i.e., invariant) under Lorentz transformations

$$\Psi(x) = \Psi'(x') \tag{1.16}$$

However, it soon became clear that the Klein-Gordon equation was not compatible with a particle interpretation. In addition, it cannot describe particles with spin. In particular, the

solutions of the Klein-Gordon equation have the (expected) dispersion law

$$E^2 = \boldsymbol{p}^2 c^2 + m^2 c^4 \tag{1.17}$$

which implies that there are positive and negative energy solutions

$$E = \pm\sqrt{\boldsymbol{p}^2 c^2 + m^2 c^4} \tag{1.18}$$

From a "particle" point of view, negative energy states are unacceptable, since they would imply that there is no ground state. We will see in chapter 4 that in QFT, there is a natural and simple interpretation of these solutions that in no way make the system unstable. However, the meaning of the negative-energy solutions was unclear in the early 1930s.

To satisfy the requirement from special relativity that energy and momentum must be treated equally, and to avoid the negative-energy solutions that came from working with the square of the Hamiltonian H, Dirac proposed to look for an equation that was linear in derivatives (Dirac, 1928). To be compatible with special relativity, the equation must be *covariant* under Lorentz transformations (i.e., it should have the same form in all reference frames). Dirac proposed a *matrix* equation that is linear in derivatives with a "wave function" $\Psi(x)$ in the form of a four-component vector, a 4-spinor $\Psi_a(x)$ (with $a = 1, \ldots, 4$):

$$i\hbar \frac{\partial \Psi_a}{\partial t}(x) + \frac{\hbar c}{i} \sum_{j=1}^{3} \alpha_j^{ab} \partial_j \Psi_b(x) + mc^2 \beta_{ab} \Psi_b(x) = 0 \tag{1.19}$$

where α_j and β are four 4×4 matrices. For this equation to be *covariant*, it is necessary that the 4-spinor field Ψ should transform as a *spinor* under Lorentz transformations

$$\Psi'_a(\Lambda x) = S_{ab}(\Lambda) \Psi_b(x) \tag{1.20}$$

where $S(\Lambda)$ is a suitable matrix. The matrix elements of the matrices α_j and β have to be pure numbers that are independent of the reference frame. By further requiring that the iterated form of this equation (i.e., the "square") satisfies the Klein-Gordon equation for each component separately, Dirac found that the matrices obey the (Clifford) algebra

$$\{\alpha_j, \alpha_k\} = 2\delta_{jk} \mathbf{1}, \qquad \{\alpha_j, \beta\} = 0, \qquad \alpha_j^2 = \beta^2 = \mathbf{1} \tag{1.21}$$

where $\mathbf{1}$ is the 4×4 identity matrix. The solutions are easily found to have the energy eigenvalues $E = \pm\sqrt{\boldsymbol{p}^2 c^2 + m^2 c^4}$. (We will come back to this in chapter 2.) It is also possible to show that the solutions are spin-1/2 particles and antiparticles (we will discuss this later on).

However, the particle interpretation of both the Klein-Gordon and the Dirac equations was problematic. Although spin 1/2 appeared now in a natural way, the meaning of the negative energy states remained unclear.

The resolution of all of these difficulties was the fundamental idea that these equations should *not* be regarded as the generalization of Schrödinger's equation for relativistic particles but, instead, as the *equations of motion of a field*, whose excitations are the particles, much in the same way as photons are the excitations of the electromagnetic field. In this picture, particle number is not conserved, but charge is. Thus, photons interacting with matter can create electron-positron pairs. Such processes do not violate charge conservation,

but the notion is lost of a particle as an object that is a fundamental entity and has a distinct physical identity. Instead, the field becomes the fundamental object, and the particles become the excitations of the field.

Thus, the relativistic generalization of quantum mechanics is QFT. This concept is the starting point of QFT. The basic strategy is to seek a field theory with specific symmetry properties and whose equations of motion are Maxwell, Klein-Gordon, and Dirac equations, respectively. Notice that if the particles are to be regarded as the excitations of a field, there can be as many particles as we wish. Thus, the Hilbert space of a QFT has an arbitrary (and indefinite) number of particles. Such a Hilbert space is called a *Fock space*.

Therefore, in QFT, the field is *not* the wave function of anything. Instead the field represents an infinite number of degrees of freedom. In fact, the wave function in a QFT is a *functional* of the field configurations, which themselves specify the state of the system. We will see in chapter 4 that the states in Fock space are given either by specifying the number of particles and their quantum numbers or, alternatively, in terms of the amplitudes (or configurations) of some properly chosen fields.

Different fields transform differently under Lorentz transformations and constitute different representations of the Lorentz group. Consequently, their excitations are particles with different quantum numbers that label the representation. Thus,

1) The Klein-Gordon field $\phi(x)$ represents charge-neutral scalar spin-0 particles. Its configuration space is the set of mappings of Minkowski space onto the real numbers $\phi : \mathcal{M} \mapsto \mathbb{R}$, or complex numbers for charged spin-0 particles $\phi : \mathcal{M} \mapsto \mathbb{C}$.
2) The Dirac field represents charged spin-$1/2$ particles. It is a complex 4-spinor $\Psi_\alpha(x)$ ($\alpha = 1, \ldots, 4$), and its configuration space is the set of maps $\Psi_\alpha : \mathcal{M} \mapsto \mathbb{C}^4$, while it is real for neutral spin-$1/2$ particles (such as neutrinos).
3) The gauge field $A^\mu(x)$ represents the electromagnetic field, and its non-abelian generalizations for gluons (and so forth).

The description of relativistic quantum mechanics in terms of relativistic quantum fields solved essentially all problems that originated in its initial development. Moreover, QFT gives exceedingly accurate predictions of the behavior of quantized electromagnetic fields and charged particles, as described by quantum electrodynamics (QED). QFT also gives a detailed description of both the strong and weak interactions in terms of field theories known as quantum chromodynamics (QCD), based on Yang-Mills gauge field theories, and unified and grand unified gauge theories.

However, along with its successes, QFT also brought with it a completely new set of physical problems and questions. Essentially, any QFT of physical interest is necessarily a nonlinear theory, as it has to describe interactions. So even though the quantum numbers of the excitations (i.e., the "particle" spectrum) may be quite straightforward in the absence of interactions, the intrinsic nonlinearities of the theory may actually unravel much of this structure. Note that the equations of motion of QFT are nonlinear, as they also are in quantum mechanics. However, the wave functional of a QFT obeys a linear Schrödinger equation, just as the wave function does in nonrelativistic quantum mechanics.

In the early days of QFT, and indeed for some time thereafter, it was assumed that perturbation theory could be used in all cases to determine the actual spectrum. It was soon found out that while there are several cases of great physical interest in which some sort of perturbation theory yields an accurate description of the physics, in many more situations this is not the case. Early on it was found that, at every order in perturbation theory, there are singular contributions to many physical quantities. These singularities reflected the existence of an infinite number of degrees of freedom, both at short distances,

since spacetime is a continuum (the *ultraviolet* (UV) domain), and at long distances, since spacetime is (essentially) infinite (the *infrared* (IR) domain). Qualitatively, divergent contributions in perturbation theory come about because degrees of freedom from a wide range of length scales (or wavelengths) and energy scales (or frequencies) contribute to the expectation values of physical observables.

Historically, the way these problems were dealt with was through the process of regularization (i.e., making the divergent contributions finite), and renormalization (i.e., defining a set of effective parameters which are functions of the energy and/or momentum scale at which the system is probed). Regularization required that the integrals be cut off at some high-energy scale (in the UV). Renormalization was then thought of as the process by which these arbitrarily introduced cutoffs were removed from the expressions for physical quantities. This was a physically obscure procedure, but it worked brilliantly in QED and, to a lesser extent, in QCD. Theories for which such a procedure can be implemented with the definition of only a *finite* number of renormalized parameters (the actual input parameters to be taken from experiment) are said to be *renormalizable* QFTs. QED and QCD are the most important examples of renormalizable QFTs, although there are many others.

Renormalization implies that the connection between the physical observables and the parameters in the Lagrangian of a QFT is highly nontrivial, and that the spectrum of the theory may have little to do with the predictions of perturbation theory. This is the case for QCD, whose "fundamental fields" involve quarks and gluons but the actual physical spectrum consists only of bound states whose quantum numbers are not those of either quarks or gluons. Renormalization also implies that the behavior of the physical observables depends on the scale at which the theory is probed. Moreover, a closer examination of these theories also revealed that they may exist in different *phases*, in which the observables have different behaviors with a specific particle spectrum in each phase. In this way, to understand what a given QFT predicted became very similar to the study of phases in problems in statistical physics. We will explore these connections in detail later in this book when we develop the machinery of the renormalization group in chapter 15. In this picture, the *vacuum* (or ground state) of a QFT corresponds to a *phase*, much in the same way as in statistical (or condensed matter) physics.

While the requirement of renormalizability works for the Standard Model of particle physics, it fails for gravity. The problem of unifying gravity with the rest of the forces of Nature remains a major problem in contemporary physics. A major program to solve this problem is string theory. String theory is the only known viable candidate to quantize gravity in a consistent manner. However, in string theory, QFT is seen as an effective low-energy (hydrodynamic) description of nature, and the QFT singularities are "regulated" by string theory in a natural way (but at the price of locality).

Index