1 Quantum field theory and the renormalization group

Without a minimal understanding of quantum or statistical field theories (formally related by continuation to imaginary time), the theoretical basis of a notable part of twentieth century physics remains incomprehensible.

Indeed, field theory, in its various incarnations, describes fundamental interactions at the microscopic scale, singular properties of phase transitions (like liquid–vapour, ferromagnetic, superfluid, separation of binary mixture,...) at the transition point, properties of diluted quantum gases beyond the model of Bose–Einstein condensation, statistical properties of long polymeric chains (as well as self-avoiding random walks), or percolation, and so on.

In fact, quantum field theory offers at present the most comprehensive framework to discuss physical systems that are characterized by a large number of strongly interacting local degrees of freedom.

However, at its birth, quantum field theory was confronted with a somewhat unexpected problem, the problem of *infinities*. The calculation of most physical processes led to infinite results. An empirical recipe, *renormalization*, was eventually discovered that allowed extracting from divergent expressions finite predictions. The procedure would hardly have been convincing if the predictions were not confirmed with increasing precision by experiment. A new concept, the *renormalization group* related in some way to the renormalization procedure, but whose meaning was only fully appreciated in the more general framework of the theory of phase transitions, has led, later, to a satisfactory interpretation of the origin and role of renormalizable quantum field theories and of the renormalization process.

This first chapter tries to present a brief history of the origin and the development of quantum field theory, and of the evolution of our interpretation of renormalization and the renormalization group, which has led to our present understanding.

This history has two aspects, one directly related to the theory of fundamental interactions that describes physics at the microscopic scale, and another one related to the theory of phase transitions in macroscopic physics and their universal properties. That two so vastly different domains of physics have required the development of the same theoretical framework, is extremely surprising. It is one of the attractions of theoretical physics that such relations can sometimes be found.

A few useful dates:

1925 Heisenberg proposes a quantum mechanics, under the form of a mechanics of matrices.

1926 Schrödinger publishes his famous equation that bases quantum mechanics on the solution of a non-relativistic wave equation. Since relativity theory was already well established when quantum mechanics was formulated, this may surprise.

In fact, for accidental reasons, the spectrum of the hydrogen atom is better described by a non-relativistic wave equation than by a relativistic equation without spin,* the Klein–Gordon equation (1926).

1928 Dirac introduces a relativistic wave equation that incorporates the spin 1/2 of the electron, which describes much better the spectrum of the hydrogen atom, and opens the way for the construction of a relativistic quantum theory. In the two following years, Heisenberg and Pauli lay out, in a series of articles, the general principles of quantum field theory.

1934 First correct calculation in quantum electrodynamics (Weisskopf) and confirmation of the existence of divergences, called ultraviolet (UV) since they are due, in this calculation, to the short-wavelength photons.

1937 Landau publishes his general theory of phase transitions.

1944 Exact solution of the two-dimensional Ising model by Onsager.

1947 Measurement of the so-called Lamb shift by Lamb and Retherford, which agrees well with the prediction of quantum electrodynamics (QED) after cancellation between infinities.

1947–1949 Construction of an empirical general method to eliminate divergences called *renormalization* (Feynman, Schwinger, Tomonaga, Dyson, *et al*).

1954 Yang and Mills propose a non-Abelian generalization of Maxwell's equations based on non-Abelian gauge symmetries (associated to non-commutative groups).

1954–1956 Discovery of a formal property of quantum field theory characterized by the existence of a *renormalization group* whose deep meaning is not fully appreciated (Peterman–Stückelberg, Gell-Mann–Low, Bogoliubov–Shirkov).

1967–1975 The Standard Model, a renormalizable quantum field theory based on the notions of non-Abelian gauge symmetry and spontaneous symmetry breaking, is proposed, which provides a complete description of all fundamental interactions, but gravitation.

1971–1972 After the initial work of Kadanoff (1966), Wilson, Wegner, *et al*, develop a more general concept of renormalization group, which includes the field theory renormalization group as a limit, and which explains universality properties of continuous phase transitions (liquid–vapour, superfluidity, ferromagnetism) and later of geometrical models like self-avoiding random walks or percolation.

1972–1975 Several groups, in particular Brézin, Le Guillou and Zinn-Justin, develop powerful quantum field theory techniques that allow a proof of universality properties of critical phenomena and calculating universal quantities.

1973 Using renormalization group arguments, Politzer and Gross–Wilczek establish the property of *asymptotic freedom* of a class of non-Abelian gauge theories, which allows explaining the free-particle behaviour of quarks within nucleons.

1975–1976 Additional information about universal properties of phase transitions are derived from the study of the non-linear σ model and the corresponding $d-2$ expansion (Polyakov, Brézin–Zinn-Justin).

* intrinsic angular momentum of particles, that takes half-integer (fermions) or integer (bosons) values in units of $\hbar$.

1977–1980 Following a suggestion of Parisi, Nickel calculates several successive terms of the series expansion of renormalization group functions, by field theory methods. Applying summation methods to these series, Le Guillou and Zinn-Justin obtain the first precise estimates of critical exponents in three-dimensional phase transitions from renormalization group methods.

1.1 Quantum electrodynamics: A quantum field theory

Quantum electrodynamics (QED) describes, in a quantum and relativistic framework, interactions between all electrically charged particles and the electromagnetic field. QED is not a theory of individual particles, as in non-relativistic quantum mechanics, but a *quantum field theory* (QFT). Indeed, it is also a quantum extension of a classical relativistic field theory: Maxwell electromagnetism in which the dynamical quantities are fields, the electric and magnetic fields. Moreover, the discovery that to the electromagnetic field were associated quanta, the photons, which are massless spin one particles, has naturally led to the postulate that all particles were also manifestations of quantum fields.

But such a theory differs radically from a particle theory in the sense that fields have an *infinite number of degrees of freedom.* Indeed, a point particle in classical mechanics has three degrees of freedom; it is characterized by its three Cartesian coordinates. By contrast, a field is characterized by its values at all space points, which thus constitutes an infinite number of data. The non-conservation of the number of particles in high-energy collisions is a manifestation of such a property.

Moreover, the field theories that describe microscopic physics have a *locality* property, a notion that generalizes the notion of point-like particle: they display no short-distance structure.

The combination of an infinite number of degrees of freedom and locality explain why QFT has somewhat 'exotic' properties.

Gauge symmetries. In what follows, we mention gauge symmetry and gauge theories, the simplest example being provided by QED. In non-relativistic quantum mechanics, in the presence of a magnetic field, gauge invariance corresponds simply to the possibility of adding a gradient term to the vector potential without affecting the equations of motion. In non-relativistic quantum mechanics, physics is not changed if one multiplies the wave function by a phase factor $\mathrm{e}^{i\theta}$ (corresponding to a transformation of the Abelian, or commutative, group $U(1)$). In the case of a charged particle, in the presence of a magnetic field, one discovers a much larger symmetry, a gauge symmetry: it is possible to change the phase of the wave function at each point in space independently,

$$\psi(x) \mapsto \mathrm{e}^{i\theta(x)}\,\psi(x),$$

by modifying in a correlated way the vector potential.

Unlike an ordinary symmetry that corresponds to transforming in a global way all dynamical variables, a gauge symmetry corresponds to independent transformations at each point in space (or space-time). Gauge symmetry is a dynamical principle: it generates interactions instead of simply relating them as an ordinary or global symmetry. To render a theory gauge invariant, it is necessary to introduce a vector potential coupled in a universal way to all charged particles. In a relativistic quantum theory, this vector potential takes the form of a gauge field corresponding to a spin-one particle, the photon in the case of QED.

Units in relativistic quantum theory. The phenomena that we evoke below are characteristic of a relativistic and very quantum limit. It is thus physically meaningful to take the speed of light, c, and Planck's constant, $\hbar$, as units. As a consequence, in a relativistic theory mass-scales M, momenta p and energies E can be related by the speed of light c:

$$E = pc = Mc^2$$

and, thus, expressed in a common unit like the electron volt (eV). It is then equivalent to talk about large momenta or large energies.

Moreover, in a quantum theory momentum-scales p can be related to length-scales ℓ by Planck's constant,

$$p\ell = \hbar\,.$$

As a consequence, high-energy experiments probe properties of matter at short distance.

1.2 Quantum electrodynamics: The problem of infinities

The discovery of the Dirac equation in 1928 opened the way to the construction of a quantum and relativistic theory, allowing a more precise description of the electromagnetic interactions between protons and electrons. This theory, whose principles were established by Heisenberg and Pauli (1929–1930), was a QFT and not a theory of individual particles, because the discovery that the electromagnetic field was associated to quanta suggested that, conversely, all particles could be manifestations of the existence of underlying fields.

After the articles of Heisenberg and Pauli, Oppenheimer and Waller (1930) published independent calculations of the effect of the electromagnetic field on the electron propagation, at first-order in the fine structure constant, a *dimensionless* constant that characterizes the intensity of the electromagnetic interactions,

$$\alpha = \frac{e^2}{4\pi\hbar c} \approx 1/137\,, \tag{1.1}$$

where e is the electron charge, defined in terms of the Coulomb potential parametrized as a function of the separation R as $e^2/4\pi R$. Since this constant α is numerically small (a meaningful statement because it is a dimensionless quantity), a first-order calculation is reasonable.

The physical process responsible of this contribution is a process typical of a QFT, the emission and absorption by an electron of energy-momentum (or quadri-momentum) q of a virtual photon of quadri-momentum k, as represented by the

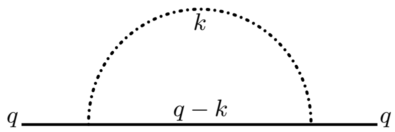

Fig. 1.1 Feynman diagram: A contribution to electron propagation, dotted line for photon, full line for electron.

Feynman diagram in Figure 1.1 (a representation proposed by Feynman several years later).

One possible motivation for undertaking such a calculation was a determination of the first corrections to the electron mass in QED and the solution of the puzzle of the 'classical model' of the electron. In a relativistic theory, the mass of a particle is proportional to its rest energy. This thus includes the potential self-energy. But it was well-known that the classical model of the electron as a charged sphere of radius R led to a result that diverged as e^2/R in the zero R limit. It could have been hoped that quantum mechanics, which is a theory of wave functions, would solve the problem generated by the point-like nature of the electron.

However, the first results were paradoxical. Not only was the contribution to the mass still infinite, but it diverged even more strongly than in the classical model: introducing a bound Λc^2 on the photon energy (this is equivalent to modifying the theory at a short-distance $R = \hbar/c\Lambda$), one found a quadratic divergence $\Lambda^2 \propto 1/R^2$. In fact, it was soon discovered that these results were wrong. Indeed, perturbative calculations with the technical tools of the time were laborious. The formalism was not explicitly relativistic; the role of the 'holes' in Dirac's theory (predicted to be anti-electrons or positrons in 1931 and experimentally discovered in 1932) was unclear, and gauge invariance was at the origin of additional difficulties. Only in 1934 was the correct result published by Weisskopf (after a last error had been pointed out by Furry). It confirmed that the contribution was definitively infinite, even though it diverged less strongly than in the classical model. The quadratic divergence was replaced by a less severe logarithmic divergence and the contribution to the electron mass at order α was found to be given by

$$\delta m_{\text{QED}} = -3\frac{\alpha}{2\pi} m \ln(mRc/\hbar),$$

where m is the electron mass for $\alpha = 0$.

The divergence is generated by the summation over virtual photons of arbitrary high momenta (due to the absence of a short-distance structure), which explains the denomination of ultraviolet (UV) divergences. Moreover, conservation of probabilities implies that all processes contribute additively.

The conclusion was that QFT was less singular than the classical model. Nevertheless, the problem of infinities was not solved and no straightforward modification could be found to save QFT.

These divergences were understood to have a profound meaning, seeming to be an unavoidable consequence of unitarity (conservation of probabilities) and locality (point-like particles with contact interactions). Moreover, it appeared extremely difficult to conceive of a consistent relativistic theory with extended particles.

The problem thus was very deep and touched at the essence of the theory itself. QED was an incomplete theory, but it seemed difficult to modify it without sacrificing some fundamental physical principle. It was possible to render the theory finite by abandoning unitarity and, thus, conservation of probabilities (as proposed by Dirac (1942)), but the physics consequences seemed hardly acceptable. What is often called Pauli–Villars' regularization, a somewhat *ad hoc* and temporary procedure to render the perturbative expansion finite before *renormalization* (see below), has this nature. It seemed even more difficult to incorporate the theory into a relativistic, non-local extension (which would correspond to giving an internal structure to particles), though Heisenberg proposed in 1938 the introduction of a fundamental length. In fact, it is only in the 1980s that possible candidates for a non-local extension of QFT were proposed in the form of superstring theories.

The crisis was so severe that Wheeler (1937) and Heisenberg (1943) proposed abandoning QFT altogether in favour of a theory of physical observables, in fact scattering data between particles: the S-matrix theory, an idea that became quite fashionable in the 1960s in the theory of *strong interactions* (those that generate nuclear forces).

Infinities and the problem of charged scalar bosons. After the first QED calculations, some pragmatic physicists started calculating other physical quantities, exploring the form and nature of infinities. Let me mention here another important work of Weisskopf (1939) in which the author shows that logarithmic divergences persist to all orders in the perturbative expansion, that is, in the expansion in powers of α. But he also notes that in the case of scalar (i.e., spinless) charged particles the situation is much worse: the divergences are quadratic, which is disastrous. Indeed, if the divergences are suppressed by some momentum cut-off $\Lambda = \hbar/Rc$ related to some new, unknown, physics and if Λ/m is not too large (and for some time 100 MeV, which is the range of nuclear forces, seemed a plausible candidate), then the product $\alpha \ln(\Lambda/m)$ remains small: a logarithmic divergence produces undetermined but nevertheless small corrections, but this is no longer the case for quadratic divergences. This result could have been understood as an indication that scalar particles cannot be considered as fundamental.

Note that the problem is more relevant than ever since the Standard Model that describes all experimental results up to the highest available energies in colliders, contains a scalar particle, the Higgs boson, and is now called the *fine tuning* or *hierarchy* problem. Indeed, to cancel infinities, it is necessary to adjust one parameter of the initial theory with a precision related to the ratio of the physical mass and the large-momentum cut-off, something that is not *natural*. The problem has now become specially severe since physicists have realized that mass-scales as large as 10^{15} (the so-called unification mass) or 10^{19} GeV (Planck's mass) can be involved. It is one of the main motivations for the introduction of *supersymmetry* (a symmetry that, surprisingly, relates bosons to fermions). The experimental discovery of the Higgs boson and the understanding of its properties are among the main goals of the new proton accelerator, the Large Hadron Collider or LHC, presently under construction at CERN near Geneva (first beam expected in late 2007).

1.3 Renormalization

Calculating a number of different physical quantities, physicists noticed that, although many physical quantities were divergent, it was always the same kind of divergent contributions that appeared. One could thus find combinations that were finite (Weisskopf 1936). However, the physical meaning of such a property, the cancellation of infinities, was somewhat obscure. In fact, in the absence of any deeper understanding of the problem, little progress could be expected.

Each time physicists are confronted with such conceptual difficulties, some clue must eventually come from experiment.

Indeed, in 1947 Lamb and Retheford measured precisely the separation between the $2s_{1/2}2p_{1/2}$ levels of the hydrogen atom, while Rabi's group at Columbia measured the anomalous moment of the electron. Remarkably enough, it was possible to organize the calculation of the Lamb shift in such the way that infinities cancel (first approximate calculation by Bethe) and the result was found to be in very good agreement with experiment. Shortly after, Schwinger obtained the leading contribution to the anomalous magnetic moment of the electron.

These results initiated extraordinary theoretical developments (earlier work of Kramers concerning the mass renormalization of the extended classical electron proved to be important to generalize the idea of cancellation of infinities by subtraction, to the idea of renormalization), and in 1949 Dyson, relying in particular on the work of Feynman, Schwinger and Tomonaga, gave a proof of the cancellation of infinities to all orders in the perturbative expansion. What became known as the *renormalization theory* led in QED to finite results for all physical observables.

The general idea is the following: one begins with an initial theory called *bare*, which depends on parameters like the *bare mass* m_0 and the *bare charge* e_0 of the electron (mass and charge in the absence of interactions), or equivalently bare fine structure constant $\alpha_0 = e_0^2/4\pi\hbar c$. Moreover, one introduces a large-momentum cut-off $c\Lambda$ (which corresponds to modifying in a somewhat arbitrary and unphysical way the theory at a very short distance of order $\hbar/c\Lambda$). One then calculates the physical values (i.e., those one measures), called *renormalized*, of the same quantities (as the observed charge e and, thus, the fine structure constant α, and the physical mass m) as functions of the bare parameters and the cut-off:

$$\begin{aligned}\alpha &= \alpha_0 - \beta_2\alpha_0^2 \ln(\Lambda/m_0) + \cdots,\\ m &= m_0 - \gamma_1\; m_0\alpha \ln(\Lambda C_1/m_0) + \cdots.\end{aligned}$$

(β_2, γ_1 and C_1 are three numerical constants.) One inverts these relations, now expressing bare quantities as functions of the renormalized one. In this substitution, one exchanges, for example, the bare constant α_0 with the physical or renormalized constant α as the expansion parameter:

$$\begin{aligned}\alpha_0 &= \alpha + \beta_2\alpha^2 \ln(\Lambda/m) + \cdots,\\ m_0 &= m + \gamma_1\; m\alpha \ln(\Lambda C_1/m) + \cdots.\end{aligned}$$

One then expresses any other observable (i.e., any other measurable quantity), initially calculated in terms of bare parameters, in terms of these physical or renormalized quantities. Most surprisingly, in the limit of infinite cut-off Λ, all physical observables then have a finite limit.

This *a priori* somewhat strange procedure, renormalization, has allowed and still allows calculations of increasing precision in QED. The remarkable agreement between predictions and experiment convincingly demonstrates that renormalized QFT provides a suitable formalism to describe electrodynamics in the quantum regime (it is even the domain in physics where agreement between theory and experiment is verified with the highest precision).

Moreover, renormalization theory has led to the very important concept of *renormalizable theories*. Only a limited number of field theories lead to finite results by this procedure. This severely restricts the structure of possible theories.

Finally, let us point out that for more than 15 years theoretical progress had been stopped by the problem of divergences in QFT. However, once experiment started producing decisive information, in two years a complete and consistent framework for perturbative calculations was set up.

The mystery of renormalization. Though it was now obvious that QED was the correct theory to describe electromagnetic interactions, the renormalization procedure itself, allowing the extraction of finite results from initial infinite quantities, had remained a matter of some concern for theorists: the meaning of the renormalization 'recipe' and, thus, of the bare parameters remained obscure. Much effort was devoted to try to overcome this initial conceptual weakness of the theory. Several types of solutions were proposed:

(i) The problem came from the use of an unjustified perturbative expansion and a correct summation of the expansion in powers of α would solve it. Somewhat related, in spirit, was the development of the so-called *axiomatic QFT*, which tried to derive rigorous, non-perturbative, results from the general principles on which QFT was based.

(ii) The principle of QFT had to be modified: only renormalized perturbation theory was meaningful. The initial bare theory with its bare parameters had no physical meaning. This line of thought led to the BPHZ (Bogoliubov, Parasiuk, Hepp, Zimmerman) formalism and, finally, to the work of Epstein and Glaser, where the problem of divergences in position space (instead of momentum space) was reduced to a mathematical problem of a correct definition of singular products of distributions. The corresponding efforts much clarified the principles of perturbation theory, but disguised the problem of divergences in such a way that it seemed never having existed in the first place.

(iii) Finally, the cut-off had a physical meaning and was generated by additional interactions, non-describable by QFT. In the 1960s some physicists thought that *strong interactions* could play this role (the cut-off then being provided by the range of nuclear forces). Renormalizable theories could then be thought as theories somewhat insensitive to this additional unknown short-distance structure, a property that obviously required some better understanding.

This latter point of view is in some sense close to our modern understanding, even though the cut-off is no longer provided by strong interactions.

1.4 Quantum field theory and the renormalization group

In the mid 1950s, several groups, most notably Peterman and Stückelberg (1953), Gell-Mann and Low (1954) and Bogoliubov and Shirkov (1955–1956), noticed that in the limit of a QED of photons and massless electrons, the renormalized perturbative expansion has a peculiar formal property, a direct consequence of the renormalization process itself.

In a massive theory, the renormalized charge can be defined through the electric interaction of particles at rest (Coulomb force). This definition is no longer applicable to massless particles, which always travel at the speed of light. It becomes then necessary to introduce some arbitrary mass (or energy or momentum) scale μ to define the renormalized charge e: it is related to the observed strength of the electromagnetic interaction in particle collisions at momenta of order μ. One can then call the renormalized charge the *effective* charge at scale μ. However, since this mass-scale is arbitrary, one can find other couples $\{e', \mu'\}$ which give the same physical results. The set of transformations of the physical parameters associated with the change in scale μ and necessary to keep the physics constant was called the *renormalization group* (RG). Making an infinitesimal change of scale, one can describe the variation of the effective charge by a differential (flow) equation

$$\mu \frac{\mathrm{d}\alpha(\mu)}{\mathrm{d}\mu} = \beta\big(\alpha(\mu)\big), \quad \beta(\alpha) = \beta_2 \alpha^2 + O(\alpha^3), \tag{1.2}$$

where the function $\beta(\alpha)$ can be calculated as a series expansion in powers of α.

Actually, even in a massive theory one can introduce this definition of the renormalized charge. This effective charge then has the following physical interpretation. At large distance, the intensity of the electromagnetic force does not vary and the charge has the value measured through the Coulomb force. However, at distances much shorter than the wavelength $\hbar/mc$ associated with a particle (one explores in some sense the 'interior' of the particle), one observes screening effects. What is remarkable is that these short-distance effects have a direct relation with renormalization.

Since one main concern was the large-momentum divergences in QFT, Gell-Mann and Low tried to use the RG to study the large-momentum behaviour of the electron propagator, beyond perturbation theory, in relation with the large cut-off behaviour of the bare charge. The bare charge can indeed be considered as an effective charge at the cut-off scale. If the function $\beta(\alpha)$ were to have a zero with a negative slope, the zero would have been the finite limit at infinite cut-off of the bare charge, beyond perturbation theory.

Unfortunately QED is a so-called IR-free theory ($\beta_2 > 0$), which means that the effective charge decreases at low momentum, and conversely at large momentum increases until the perturbative expansion of the β-function is no longer reliable.

(This variation is observed in experiments since the effective value of α measured near the Z vector boson, i.e., at about 100 GeV, is about 4% larger than its low--energy value.)

It is quite striking that if they were to have turned the argument around, they would have found that, at fixed bare charge, the effective charge goes to zero as $1/\ln(\Lambda/m_{\text{el.}})$, which is acceptable for any reasonable value of cut-off and may even account for the small value of α, but their hope of course was to get rid of the cut-off.

Note some related speculations: Landau and Pomeranchuk (1955) noticed that if, in the calculation of the electron propagation in an electromagnetic field, one sums the leading terms at large momentum at each order, one predicts the existence of a particle of mass $M \propto m\, e^{1/\beta_2\alpha}$. This could have corresponded to a boundstate, but unfortunately this particle has unphysical properties leading to non-conservation of probabilities, and thus was called the Landau 'ghost'. For Landau, this was obviously the sign of some inconsistency of QED, though of no immediate physical consequence, because α is so small that this mass is of the order of 10^{30} GeV. Bogoliubov and Shirkov correctly pointed out that this result amounted to solving the RG equation (1.2) at leading order, that is, for small effective charge. Since the effective charge becomes large at large momenta, perturbation theory can eventually no longer be trusted. It is amusing to note that, in the modern point of view, we believe that Landau's intuition was basically correct, even though the argument, as initially formulated, was somewhat too naive.

1.5 A triumph of QFT: The Standard Model

QFT in the 1960s. After the triumph of QED, the 1960s were a time of doubt and uncertainty for QFT. Three outstanding problems remained, related to the three other known interactions:

(i) *Weak interactions* (related to the weak nuclear force) were described by the non-renormalizable four-fermion Fermi (–Feynman–Gell-Mann) interaction. Since the coupling was weak and the interaction was of a current–current type, as in QED when the photon is not quantized, it was conceivable that the theory was, in some way, the leading approximation to a QED-like theory, but with at least two very heavy (of the order of 100 GeV) photons, because the interaction was essential point-like. Classical field theories containing several photons, called non-Abelian gauge theories, that is, theories in which interactions are generated by a generalized symmetry principle called gauge symmetry, had been proposed by Yang and Mills (1954). However, their quantization led to new and difficult technical problems. Moreover, gauge theories, like QED, have a strong tendency to produce massless vector fields. So a few theorists were trying both to quantize the so-called Yang–Mills fields and to find ways to generate mass terms for them, within the framework of renormalizable QFTs.

(ii) On the other hand, many thought that, in the theory of strong interactions, the case for QFT was desperate: because the interactions were so strong, no perturbative expansion could make sense. Only a theory of physical observables, called

S-matrix theory, could provide the right theoretical framework, and strict locality had to be abandoned. One can note the first appearance of string models in this context.

(iii) Finally, since *gravitational forces* are extremely small at short distance, there was no immediate urgency to deal with quantum gravity, and the solution to this problem of uncertain experimental relevance could be postponed.

The triumph of renormalizable QFT. Toward the end of the 1960s, the situation changed quite rapidly. At last, methods to quantize non-Abelian gauge theories were found (Faddeev–Popov, De Witt 1967). These new theories could be proved to be renormalizable ('t Hooft, 't Hooft–Veltman, Slavnov, Taylor, Lee–Zinn-Justin, Becchi–Rouet–Stora, Zinn-Justin, 1971–1975) even in the broken symmetry phase in which masses could be generated for vector bosons (the Higgs mechanism, Higgs, Brout–Englert, Guralnik–Hagen–Kibble 1964) and fermions. These developments allowed constructing a quantum version of a model for combined weak and electromagnetic interactions proposed earlier by Weinberg (1967) and Salam (1968). Its predictions were soon confirmed by experiment.

In the rather confusing situation of strong interactions, the solution came as often in such circumstances from experiments: *deep inelastic scattering* experiments at SLAC, probing the interior of protons or neutrons, revealed that hadrons were composed of almost free point-like objects, initially called *partons* and eventually identified with the quarks that had been used as mathematical entities to provide a simple description of the symmetries of the hadron spectrum.

To understand this peculiar phenomenon, RG ideas were recalled in the modernized version of Callan and Symanzik (1970), valid also for massive theories, but the phenomenon remained a puzzle for some time until field theories could be found in which interactions became weak at short distance (unlike QED) such as to explain SLAC results. Finally, the same theoretical advances in the quantization of non-Abelian gauge theories which had provided a solution to the problem of weak interactions, allowed constructing a theory of strong interactions: *quantum chromodynamics* (QCD). Indeed, it was found that non-Abelian gauge theories, with a limited number of fermions, were *asymptotically free* (Gross–Wilczek, Politzer 1973). Unlike QED, the first coefficient β_2 of the β-function there is negative. The weakness of the interactions between quarks at short distance becomes a consequence of the decrease of the effective strong charge.

Therefore, around 1973–1974, a complete QFT model for all fundamental interactions but gravity was proposed, now called the Standard Model, which has successfully survived all experimental tests up to now, more than 30 years later, except for some minor recent modifications to take into account the non-vanishing neutrino masses. This was the triumph of all ideas based on renormalizable QFT.

At this point it was tempting to conclude that some kind of new law of nature has been discovered: all interactions can be described by a *renormalizable* QFT and renormalized perturbation theory. The divergence problem had by then been so well hidden that many physicists no longer worried about it.

A remaining potential issue was what Weinberg called the *asymptotic safety* condition: the consistency of QFT on all length-scales seemed to require the existence of a UV fixed point, in the formalism of equation (1.2) a solution of

$$\beta(\alpha) = 0 \quad \text{with} \quad \beta'(\alpha) < 0\,,$$

(one of the options already considered by Gell-Mann and Low). Asymptotically free field theories share of course this property, but scalar fields (as required by the Higgs mechanism) have a tendency to destroy asymptotic freedom. Finally, it still remained to cast quantum gravity into the renormalizable framework and this became the goal of many theorists in the following years. The failure of this programme eventually led to the introduction of string theories.

1.6 Critical phenomena: Other infinities

The theory of *critical phenomena* deals with continuous, or second-order, phase transitions in macroscopic systems. Simple examples are liquid–vapour, binary mixtures, superfluid He and magnetic transitions. The simplest lattice model that exhibits such a phase transition is the famous Ising model.

These transitions are characterized by a collective behaviour on large scales near the transition temperature (the critical temperature T_{c}). For example, the correlation length, which characterizes the scale of distance on which a collective behaviour is observed, becomes infinite at the transition. Near T_{c}, these systems thus depend on two very different length-scales, a microscopic scale given by the size of atoms, the lattice spacing or the range of forces, and another scale dynamically generated, given by the correlation length. To the latter scale are associated non-trivial large-distance or macroscopic phenomena.

It could then have been expected that physics near the critical temperature could be described, in some leading approximation, by a few effective macroscopic parameters, without explicit reference to the initial degrees of freedom. This idea leads to *mean field theory* (MFT) and in its more general form to Landau's theory of critical phenomena (1937). Such a theory can be called *quasi-Gaussian*, in the sense that it assumes implicitly that the remaining correlations between stochastic variables at the microscopic scale can be treated perturbatively and, thus, that macroscopic expectation values are given by quasi-Gaussian distributions, in the spirit of the central limit theorem.

Among the simplest and most robust predictions of such a theory, one finds the *universality* of the singular behaviour of thermodynamical quantities at the critical temperature T_{c}: for instance, the correlation length ξ always diverges as $(T - T_{\mathrm{c}})^{-1/2}$, the spontaneous magnetization vanishes like $(T_{\mathrm{c}} - T)^{1/2}$, and so on, these properties being independent of the dimension of space, the symmetry of the system, and of course the detailed microscopic dynamics.

Therefore, physicists were surprised when some experiments as well as numerical calculations in simple lattice models started questioning MFT predictions. An additional blow to MFT came from Onsager's (1944) exact solution of the 2D Ising model which confirmed the corresponding lattice calculations. In the following years, empirical evidence accumulated that critical phenomena in two and three space dimensions could not be described quantitatively by MFT. In fact, the critical behaviour was found to depend on space dimensions, symmetries and some general properties of models. Nevertheless, there were also some indications that some universality survived, but in a more limited sense. Some specific properties were important, but not all details of the microscopic dynamics.

The non-decoupling of scales. To understand how deep the problem was, one has to realize that such a situation had never been met before (except, perhaps, in turbulence): indeed, the main ingredient in Landau's theory is the hypothesis that, as usual, physical phenomena on too different length-scales decouple. Let us illustrate this idea with a simple example. Naively, one derives the period τ of the pendulum, up to a numerical factor, from dimensional analysis,

$$\tau \propto \sqrt{\ell/g}\,,$$

where ℓ is the length of the pendulum and g is the gravitational acceleration. But in this argument is hidden a deeper and essential hypothesis: the internal atomic structure of the pendulum, the size of the earth or the distance between earth and sun are irrelevant because these length-scales are either much too small or much too large compared to the size of the pendulum. One expects that they lead to corrections of order the ratio $\lambda =$ (small scale/large scale), which, thus, are totally negligible. Of course, one can find mathematical functions of the ratio λ that decrease only slowly with λ like, for example, $1/\ln\lambda$. But these functions are singular and we have reasons to believe that, in general, nature is not perverse and does not introduce singularities where they are not absolutely needed.

In the same way, in Newtonian mechanics, to describe the motion of planets one can forget, to a very good approximation, the existence of other stars, the size of the sun and the planets, which can be replaced by point-like objects. Again in the same way, in non-relativistic quantum mechanics, one can ignore the internal structure of the proton to calculate with a very good precision the energy levels of the hydrogen atom.

The failure of MFT has demonstrated, on the contrary, that the decoupling of scales is not always true in the theory of critical phenomena, a new and totally unexpected situation. In fact, if one tries to calculate corrections to MFT, one finds divergences at the critical temperature. These divergences are generated by contributions that depend on the ratio of the correlation length and the microscopic scale. This situation is reminiscent of particle physics, except that in the early interpretation of QFT, it is the microscopic scale that goes to zero, while here it is the macroscopic scale (the correlation length) that becomes infinite.

The divergences met in QFT, as describing microscopic physics, and in the theory of critical phenomena, have actually a common origin: the non-decoupling of very different physical scales. The infinities appear when one tries to ignore, as one does usually and is generally justified, the existence of other length-scales where relevant physical laws may be very different.

Therefore, one could have feared that macroscopic physics would be sensitive to the details of the short-distance structure, that large-scale phenomena would depend on the detailed microscopic dynamics and, thus, would essentially be unpredictable. The emergence of a surviving universality, even more limited, was therefore even more surprising. To understand these observations, a new conceptual framework obviously had to be invented.

1.7 Kadanoff and Wilson's renormalization group

In 1966, Kadanoff proposed a strategy to deal with the problem: calculate physical observables by summing recursively over short-distance degrees of freedom. One then obtains a sequence of effective models that have all the same large-distance properties. Following Kadanoff, we use the Ising model to illustrate the idea, but take a more general viewpoint.

Example: The Ising model. The Ising model is a statistical lattice model. To each lattice site i is associated a random variable S_i that takes only two values ± 1, a 'classical' spin. Thermodynamic quantities are calculated by averaging over all spin configurations with a Boltzmann weight $\mathrm{e}^{-\mathcal{H}_a(S)/T}/\mathcal{Z}$, where T is the temperature, a the lattice spacing, $\mathcal{H}_a(S)$ a configuration energy corresponding to some short--range interactions (for example, only nearest-neighbour spins on the lattice are coupled) and $\mathcal{Z}$ is a normalization factor of the probability distribution called the *partition function.* Note that phase transitions can only occur in the infinite-volume limit, called the thermodynamic limit.

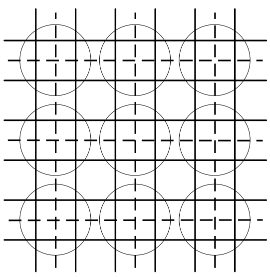

Fig. 1.2 Initial lattice (full line) and lattice with doubled spacing $---$.

Of course, in general expectation values cannot be calculated exactly. But, one would like, at least, to understand the origin of universality. The idea then is to sum over spins S_i at fixed average on a lattice of spacing $2a$. For example, on a square lattice one groups spins on a set of disjoint squares (inside the circles in Figure 1.2), and fixes the average over each square. After summation, the statistical sum is given by a summation over configurations of these average spins (which now take more than two values) belonging to a lattice of double spacing. To these spins corresponds a new configuration energy $\mathcal{H}_{2a}(S)$ called the *effective* interaction at scale $2a$.

This transformation can then be iterated,

$$\mathcal{H}_{2^n a}(S) = \mathcal{T}\left[\mathcal{H}_{2^{n-1}a}(S)\right], \tag{1.3}$$

as long as the lattice spacing remains small compared to the correlation length, that is, the scale of the macroscopic phenomena of interest. If the repeated application of the transformation produces an effective interaction whose asymptotic form is, to a large extent, independent of the initial interaction, one has found a mechanism that explains the remaining universality. Such asymptotic interactions will be *fixed points* or belong to *fixed surfaces* of the transformation $\mathcal{T}$:

$$\mathcal{H}^*(S) \underset{n\to\infty}{\to} \mathcal{H}_{2^n a}(S)\,, \quad \mathcal{H}^*(S) = \mathcal{T}\left[\mathcal{H}^*(S)\right].$$

Eventually, Wilson (1971) transformed this initial, somewhat vague, idea into a precise operational scheme, unifying finally Kadanoff's renormalization group (RG) idea with the RG of QFT. This led to an understanding of universality, as being a consequence of the existence of large-distance (IR) fixed points of a general RG. It even became possible to develop systematic methods to calculate universal quantities, with the help of partially preexisting QFT techniques (Brézin–Le Guillou–Zinn-Justin 1973).

Continuum limit and QFT. A first step consists in understanding that the iteration (1.3) leads asymptotically to a field theory in continuum space, even if the initial model is a lattice model and the dynamical variables take only discrete values.

For example, in the Ising model it is clear that, after many iterations, the effective spin variable, which is a local average of a large number of spins, takes a dense discrete set of values and can be replaced by a continuum variable. Similarly, the spacing of the initial lattice becomes arbitrarily smaller than the spacing of the iterated lattice. One can thus replace the effective spin variable by a field $S(x)$ in continuum space. The sum over spins becomes an integral over fields (generalization of Feynman's path integral), formally analogous to the field integrals that allow calculating physical observables in QFT.

The Gaussian fixed point. On can verify that one RG fixed point has the form of a Gaussian distribution (a property in direct relation with the central limit theorem of probabilities). At the critical temperature, in the large-distance limit, the Gaussian fixed point takes the form of a free (scalar) QFT (particles do not interact). Moreover, the theory is massless (the correlation length plays the role of an inverse mass) in the language of microscopic physics. The weakly perturbed Gaussian model (quasi-Gaussian approximation) reproduces all universal results predicted by MFT.

A finer analysis shows, however, that a small perturbation of the Gaussian fixed point generates infinite contributions, at least in space dimensions smaller than, or equal to 4. Below four dimensions, the Gaussian fixed point then corresponds to an unstable point fixed of the RG.

Moreover, the most singular terms are generated by a renormalizable QFT whose large-distance properties can be studied. This is a very striking result since it indicates that renormalizable QFTs can emerge as effective theories describing the large-distance properties of critical phenomena. The RG of QFT then appears as an asymptotic form of the general Wilson–Kadanoff RG.

Conversely, one is then strongly tempted to apply the ideas that have emerged in the theory of phase transitions to the QFT that describes the physics of fundamental microscopic interactions.

1.8 Effective quantum field theories

The condition that fundamental microscopic interactions should be described by a renormalizable QFT has been used as a basic principle for constructing the Standard Model. From the success of the programme, it could have been concluded that the principle of renormalizability was a new law of nature. This would have implied that all interactions including gravity should be describable by such theories. The failure, so far, to exhibit a renormalizable version of quantum gravity has shed doubts on the whole programme itself. Indeed, if the Standard Model and its natural possible extensions are only approximate theories, it becomes difficult to understand why it should obey such an abstract principle.

The theory of critical phenomena, with the natural appearance of renormalizable QFTs, has led to another simpler and more plausible explanation. One can now imagine that fundamental interactions are described at distances much shorter than those presently accessible to experiment (unification, Planck's scales?), and thus at much higher energies, by a finite theory that does not have the form of a local QFT. Although this theory is entirely defined at the microscopic scale, for reasons that can be better understood only when we get a more precise idea about this more fundamental theory, it generates, by a cooperative effect of a large number of degrees of freedom, a non-trivial large-distance physics with effective interactions between very light particles. In phase transitions, it is the experimentalist that adjusts the temperature at its critical value to make the correlation length diverge (i.e., the mass vanish). In particle physics, this must happen automatically, otherwise one is confronted with the famous *fine tuning* problem. Since Planck's mass, for example, is at least of the order of 10^{13} times larger than the mass of the Higgs particle, whose existence is conjectured by the Standard Model, this would imply that one parameter of the theory is accidentally close to some critical value with a precision of 10^{-13}.

A few possible mechanisms are known, that generate massless particles, spontaneously broken continuous symmetries that generate massless scalar bosons (Goldstone bosons), gauge symmetries that generate massless vector bosons like the photon, chiral symmetry that produces massless fermions. But none of these mechanisms solves the problem of the Higgs particle.

Assuming this problem has been solved, then one can imagine that, as a consequence of the existence of a large-distance fixed point, low-energy or large-distance physics is described by an effective QFT. This field theory comes naturally equipped with a cut-off, a reflection of the initial microscopic structure, and contains all local interactions allowed by the field content and symmetries. If the free or Gaussian field theory is not too bad an approximation (at least in some energy range), which implies that the fixed point is close enough to the Gaussian fixed point, the interactions can be classified according to the dimensions of their coefficients. Then, interactions of non-renormalizable type, which have to be excluded in the traditional viewpoint, are automatically suppressed by powers of the cut-off (Einstein's gravitation corresponds presumably already to this class). Renormalizable interactions, which are dimensionless, evolve only slowly with the scale (logarithmically) and, thus, survive at large distance. They determine low-energy physics. The super-renormalizable interactions (this include possible mass terms), which are considered innocuous in the traditional presentation of QFT because they generate only a finite number of divergences, must be naturally absent or much suppressed because they grow as a power of the cut-off. The bare theory is then also a version of the effective theory in which all non-renormalizable interactions have already been omitted. QFT is not required to be physically consistent at very short distance where it is no longer a valid approximation and where it can be rendered finite by a modification that is, to a large extent, arbitrary.

Of course, such an interpretation has no immediate influence on perturbative calculations and one could thus consider the whole issue as somewhat philosophical. But this is not completely true!

We have mentioned above that taking the bare theory seriously, leads, in particular, to a confrontation with the fine tuning problem in the case of masses of scalar particles (and thus the Higgs boson) and, thus, forces us to look for solutions (supersymmetry, bound state of more fundamental fermions?).

This interpretation also solves the problem of *triviality*: renormalized interactions decreasing logarithmically with the cut-off are acceptable because the physical cut-off is finite. For example, let us consider QED and reverse Gell-Mann and Low's argument. At fixed bare charge, the effective charge at a mass-scale μ decreases like $1/\ln(\Lambda/\mu)$, which is acceptable for any sensible value of the cut-off Λ if, for instance, μ is of the order of the electron mass. This decrease may even explain the small value of the fine structure constant.

This interpretation also suggests that quantized Einstein gravitation is a surviving non-renormalizable interaction. Moreover, other non-renormalizable interactions could also be detected through very small symmetry violations. Indeed, the theory of critical phenomena provides examples of a possible mechanism. There, one finds situations in which the theory reduced to renormalizable interactions has more symmetry that the complete initial theory (cubic symmetry on the lattice leads to rotation symmetry at large distance).

This modern viewpoint, deeply based on RG ideas and the notion of scale-dependent effective interactions, not only provides a more consistent picture of QFT, but also a framework in which new physics phenomena can be discussed.

It implies that QFTs are somewhat temporary constructions. Due to an essential coupling of very different physical scales, *renormalizable* QFTs have a consistency limited to low-energy (or large-distance) physics. One uses the terminology of *effective QFT*, approximations of an as yet unknown more fundamental theory of a radically different nature.