Introduction

Presenting a brief review of the history of the subject. — The modern perspective.

Quantum field theory (QFT) was born as a consistent theory for a unified description of physical phenomena in which both quantum-mechanical aspects and relativistic aspects are important. In historical reviews it is always difficult to draw a line that would separate "before" and "after."[1] Nevertheless, it would be fair to say that QFT began to emerge when theorists first posed the question of how to describe the electromagnetic radiation in atoms in the framework of quantum mechanics. The pioneers in this subject were Max Born and Pascual Jordan, in 1925. In 1926 Max Born, Werner Heisenberg, and Pascual Jordan formulated a quantum theory of the electromagnetic field, neglecting polarization and sources to obtain what today would be called a free field theory. In order to quantize this theory they used the canonical quantization procedure. In 1927 Paul Dirac published his fundamental paper "The quantum theory of the emission and absorption of radiation." In this paper (which was communicated to the *Proceedings of the Royal Society* by Niels Bohr), Dirac gave the first complete and consistent treatment of the problem. Thus quantum field theory emerged inevitably, from the quantum treatment of the only known classical field, i.e. the electromagnetic field.

Dirac's paper in 1927 heralded a revolution in theoretical physics which he himself continued in 1928, extending relativistic theory to electrons. The Dirac equation replaced Schrödinger's equation for cases where electron energies and momenta were too high for a nonrelativistic treatment. The coupling of the quantized radiation field with the Dirac equation made it possible to calculate the interaction of light with relativistic electrons, paving the way to quantum electrodynamics (QED).

For a while the existence of the negative energy states in the Dirac equation seemed to be mysterious. At that time – it is hard to imagine – antiparticles were not yet known! It was Dirac himself who found a way out: he constructed a "Dirac sea" of negative-energy electron states and predicted antiparticles (positrons), which were seen as "holes" in this sea.

The hole theory enabled QFT to explore the notion of antiparticles and its consequences, which ensued shortly. In 1927 Jordan studied the canonical quantization of fields, coining the name "second quantization" for this procedure. In 1928 Jordan and Eugene Wigner found that the Pauli exclusion principle required the electron field to be expanded in plane waves with anticommuting creation and destruction operators.

[1] For a more detailed account of the first 50 years of quantum field theory see e.g. Victor Weisskopf's article [1] or the "Historical Introduction" in [2] and vivid personal recollections [3].

In the mid-1930s the struggle against infinities in QFT started and lasted for two decades, with a five-year interruption during World War II. While the infinities of the Dirac sea and the zero-point energy of the vacuum turned out to be relatively harmless, seemingly insurmountable difficulties appeared in QED when the coupling between the charged particles and the radiation field was considered at the level of quantum corrections. Robert Oppenheimer was the first to note that logarithmic infinities were a generic feature of quantum corrections. The best minds in theoretical physics at that time addressed the question how to interpret these infinities and how to get meaningful predictions in QFT beyond the lowest order. Now, when we know that every QFT requires an ultraviolet completion and, in fact, represents an effective theory, it is hard to imagine the degree of desperation among the theoretical physicists of that time. It is also hard to understand why the solution of the problem was evasive for so long. Landau used to say that this problem was beyond his comprehension and he had no hope of solving it [4]. Well . . . times change. Today's students familiar with Kenneth Wilson's ideas will immediately answer that there are no actual infinities: all QFTs are formulated at a fixed short distance (corresponding to large Euclidean momenta) and then evolved to large distances (corresponding to small Euclidean momenta); the only difference between renormalizable and nonrenormalizable field theories is that the former are insensitive to ultraviolet data (which can be absorbed in a few low-energy parameters) while the latter depend on the details of the ultraviolet completion. But at that time theorists roamed in the dark. The discovery of the renormalization procedure by Richard Feynman, Julian Schwinger, and Sin-Itiro Tomonaga, which came around 1950, was a breakthrough, a ray of light. Crucial developments (in particular, due to Freeman Dyson) followed immediately. The triumph of quantum field theory became complete with the emergence of invariant perturbation theory, Feynman graphs, and the path integral representation for amplitudes,

$$\mathcal{A} = \int \prod_i \mathcal{D}\varphi_i e^{iS/\hbar}, \tag{0.1}$$

where the subscript i labels all relevant fields while S is the classical action of the theory calculated with appropriate boundary conditions.

In the mid-1950s Lev Landau, Alexei Abrikosov, and Isaac Khalatnikov discovered a feature of QED, the only respectable field theory of that time, that had a strong impact on all further developments in QFT. They found the phenomenon of zero charge (now usually referred to as infrared freedom): independently of the value of the bare coupling at the ultraviolet cut-off, the observed (renormalized) interaction between electric charges at "our" energies *must vanish* in the infinite cut-off limit. All other field theories known at that time were shown to have the same behavior. On the basis of this result, Landau pronounced quantum field theory dead [5] and called for theorists to seek alternative ways of dealing with relativistic quantum phenomena.[2] When I went to the theory department of ITEP[3] in 1970 to work on my Master's thesis, this attitude was still very much alive and studies of QFT were

[2] Of course, people "secretly" continued using field theory for orientation, e.g. for extracting analytic properties of the S-matrix amplitudes, but they did it with apologies, emphasizing that that was merely an auxiliary tool rather than the basic framework.
[3] The Institute of Theoretical and Experimental Physics in Moscow.

strongly discouraged, to put it mildly. Curiously, this was just a couple of years before the next QFT revolution.

The renaissance of quantum field theory, its second début, occurred in the early 1970s, when Gerhard't Hooft realized that non-Abelian gauge theories are renormalizable (including those in the Higgs regime) and, then, shortly after, David Gross, Frank Wilczek, and David Politzer discovered asymptotic freedom in such theories. Quantum chromodynamics (QCD) was born as *the* theory of strong interactions. Almost simultaneously, the standard model of fundamental interactions (SM) started taking shape. In the subsequent decade it was fully developed and was demonstrated, with triumph, to describe all known phenomenology to a record degree of precision. All fundamental interactions in nature fit into the framework of the standard model (with the exception of quantum gravity, of which I will say a few words later).

Thus, the gloomy prediction of the imminent *demise* of QFT – a wide spread opinion in the 1960s – turned out to be completely false. In the 1970s QFT underwent a conceptual revolution of the scale comparable with the development of renormalizable invariant perturbation theory in QED in the late 1940s and early 1950s. It became clear that the Lagrangian approach based on Eq. (0.1), while ideally suited for perturbation theory, is not necessarily the only (and sometimes, not even the best) way of describing relativistic quantum phenomena. For instance, the most efficient way of dealing with two-dimensional conformal field theories is algebraic. In fact, many different Lagrangians can lead to the same theory (according to Alexander Belavin, Alexander Polyakov, and Alexander Zamolodchikov, in 1981). This is an example of the QFT dualities, which occur not only in conformal theories and not only in two dimensions. Suffice it to mention that the sine-Gordon theory was shown long ago to be dual to the Thiring model. Even more striking were the extensions of duality to four dimensions. In 1994 Nathan Seiberg reported a remarkable finding: supersymmetric Yang–Mills theories with distinct gauge groups can be dual, leading to one and the same physics in the infrared limit!

Some QFTs were found to be integrable. Topological field theories were invented which led mathematical physicists to new horizons in mathematics, namely, in knot theory, Donaldson theory, and Morse theory.

Look through Introduction to Part II, Section 9.6.

The discovery of supersymmetric field theories in the early 1970s (which we will discuss later) was a milestone of enormous proportions, a gateway to a new world, described by QFTs of a novel type and with novel – and, quite often, – counterintuitive properties. In its impact on QFT, I can compare this discovery to that of the New World in 1492. People who ventured on a journey inside the new territory found treasures and exotic, and previously unknown, fruits: a richness of dynamical regimes in super-Yang–Mills theories, including a broad class of superconformal theories in four dimensions; exact results at strong coupling; hidden symmetries and cancellations; unexpected geometries and more.

Supersymmetric theories proved to be a powerful tool, allowing one to reveal intriguing aspects of gauge (color) dynamics at strong coupling. Continuing my analogy with Columbus's discovery of America in 1492, I can say that the expansion of QFT in the four decades that have elapsed, since 1970 has advanced us to the interior of a new continent. Our task is to reach, explore, and understand this continent and to try to open the ways to yet other continents. The reader should be warned that the very nature of the frontier explorations in QFT has changed considerably in comparison with what is found in older textbooks. A nice characterization of this change is given by an outstanding mathematical physicist, Andrey Losev, who writes [6]:

> In the good old days, theorizing was like sailing between islands of experimental evidence. And, if the trip was not in the vicinity of the shoreline (which was strongly recommended for safety reasons) sailors were continuously looking forward, hoping to see land – the sooner the better ...
>
> Nowadays, some theoretical physicists (let us call them sailors) [have] found a way to survive and navigate in the open sea of pure theoretical construction. Instead of the horizon they look at the stars,[4] which tell them exactly where they are. Sailors are aware of the fact that the stars will never tell them where the new land is, but they may tell them their position on the globe. In this way sailors – all together – are making a map that will at the end facilitate navigation in the sea and will help to discover new lands.
>
> Theoreticians become *sailors* simply because they just like it. Young people seduced by captains forming crews to go to a Nuevo El Dorado of Unified Quantum Field Theory or Quantum Gravity soon realize that they will spend all their life at sea. Those who do not like sailing desert the voyage, but for true potential sailors the sea becomes their passion. They will probably tell the alluring and frightening truth to their students – and the proper people will join their ranks.

Approximately at the same time as supersymmetry was born in the early-to-mid-1970s, a number of remarkable achievements occurred in uncovering the nonperturbative side of non-Abelian Yang–Mills theories: the discovery of extended objects such as monopoles (G. 'tHooft; A. Polyakov), domain walls, and flux tubes (H. Nielsen and P. Olesen) and, finally, tunneling trajectories (currently known as instantons) in Euclidean space–time (A. Polyakov and collaborators). A microscopic theory of magnetic monopoles was developed. It took people a few years to learn how to quantize magnetic monopoles and similar extended objects. The quasiclassical quantization of solitons was developed by Ludwig Faddeev and his school in St. Petersburg and, independently, by R. F. Dashen, B. Hasslacher, and A. Neveu. Then Y. Nambu, S. Mandelstam, and G. 't Hooft put forward (practically simultaneously but independently) the dual Meissner effect conjecture as the mechanism responsible for color confinement in QCD. It became absolutely clear that, unlike in QED, crucial physical phenomena go beyond perturbation theory and field theory *is* capable of describing them.

The phenomenon of color confinement can be summarized as follows. The spectrum of asymptotic states in QCD has no resemblance to the set of fields in the Lagrangian; at the Lagrangian level one deals with quarks and gluons while experimentalists detect pions, protons, glueballs, and other color singlet states – never quarks and gluons. Color confinement makes colored degrees of freedom inseparable. In a bid to understand this phenomenon Nambu, 't Hooft, and Mandelstam suggested a non-Abelian dual Meissner effect. According to their vision, non-Abelian monopoles condense in the vacuum, resulting in the formation of non-Abelian chromoelectric flux tubes between color charges, e.g. between a probe

[4] Here by "stars" he means aspects of the internal logic organizing the mathematical world rather than outstanding members of the community.

heavy quark and antiquark pair. Attempts to separate these probe quarks would lead to stretching of the flux tubes, so that the energy of the system grows linearly with separation. That is how linear confinement was visualized.

One may ask: where did these theorists get their inspiration? The Meissner effect, known for a long time and well understood theoretically, yielded a rather analogous picture. It answered the question: what happens if one immerses a magnetic charge and anticharge in a type-II superconductor?

If we place a probe magnetic charge and anticharge in empty space, the magnetic field they induce will spread throughout space, while the energy of the magnetic charge–anticharge configuration will obey the Coulomb $1/r$ law. The force will die off as $1/r^2$. Inside the superconductor, however, Cooper pairs condense, all electric charges are screened, and the photon acquires a mass; i.e., according to modern terminology the electromagnetic $U(1)$ gauge symmetry is Higgsed. The magnetic field cannot be screened in this way; in fact, the magnetic flux is conserved. At the same time the superconducting medium cannot tolerate a magnetic field. This clash of contradictory requirements is solved through a compromise. A thin tube (known as an Abrikosov vortex) is formed between the magnetic charge and anticharge immersed in the superconducting medium. Within this tube superconductivity is destroyed – which allows the magnetic field to spread from the charge to the anticharge through the tube. The tube's transverse size is proportional to the inverse photon mass while its tension is proportional to the Cooper pair condensate. Increasing the distance between the probe magnetic charges (as long as they are within the superconductor) does not lead to their decoupling; rather, the magnetic flux tubes become longer, leading to linear growth in the energy of the system.

This physical phenomenon inspired Nambu, 't Hooft, and Mandelstam's idea of non-Abelian confinement as a dual Meissner effect. Many people tried to quantify this idea. The first breakthrough, instrumental in all later developments, came only 20 years later, in the form of the Seiberg–Witten solution of $\mathcal{N} = 2$ supersymmetric Yang–Mills theory. This theory has eight supercharges, which makes the dynamics quite "rigid" and helps one to find the full analytic solution at low energies. The theory bears a resemblance to quantum chromodynamics, sharing common family traits. By and large, one can characterize it as QCD's second cousin.

The problem of confinement in QCD *per se* (and in nonsupersymmetric theories in four dimensions in general) is not yet solved. Since this problem is of such paramount importance for the theory of strong interactions we will discuss at length instructive models of confinement in lower dimensions.

The topics listed above have become part of "operational" knowledge in the community of field theory practitioners. In fact, they transcend this community since many aspects reach out to string theorists, cosmologists, astroparticle physicists, and solid state theorists. My task is to present a coherent pedagogical introduction covering the basics of the above subjects in order to help prepare readers to undertake research of their own.

We will start from the Higgs effect in non-Abelian gauge theories. Then we will study the basic phases in which non-Abelian gauge theories can exist – Coulomb, conformal, Higgs, and so on. Some "exotic" phases discovered in the context of supersymmetric theories will not be discussed.

A significant part of this book will be devoted to topological solitons, that is, the topological defects occurring in various field theories. The term “soliton” was introduced in the 1960s, but scientific research on solitons had started much earlier, in the nineteenth century, when a Scottish engineer, John Scott-Russell, observed a large solitary wave in a canal near Edinburgh. Condensed matter systems in which topological defects play a crucial role have been well known for a long time: suffice it to mention the magnetic flux tubes in type II superconductors and the structure of ferromagnetic materials, with domain walls at the domain boundaries.

In 1961 Skyrme [7] was the first to introduce in particle physics a three-dimensional topological defect solution arising in a nonlinear field theory. Currently such solitons are known as Skyrmions. They provide a useful framework for the description of nucleons and other baryons in multicolor QCD (in the so-called ’t Hooft limit, i.e. at $N_c \to \infty$ with $g^2 N_c$ fixed, where N_c is the number of colors and g^2 is the gauge coupling constant).

In general, in this book we will pay much attention to the broader aspects of multicolor gauge theories and the ’t Hooft limit. We will see that a large-N expansion is equivalent to a topological expansion. Each term in a $1/N$ series is in one-to-one correspondence with a particular topology of Feynman graphs, e.g. planar graphs, those with one handle, and so on. Large-N analysis presents a very fruitful line of thought, allowing one to address and answer a number of the deepest questions in gauge theories.

As early as in 1965 Nambu anticipated the cosmological significance of topological defects [8]. He conjectured that the universe could have a kind of domain structure. Subsequently Weinberg noted the possibility of domain-wall formation at a phase transition in the early universe [9].

From the general theory of solitons we pass to a specific class of supersymmetric critical (or Bogomol’nyi–Prasad–Sommerfield-saturated) solitons.

I will present a systematic and rather complete introduction to supersymmetry that is (almost) sufficient for bringing students to the cutting edge in this area.

Readers should be warned that nothing will be said on the quantum theory of gravity. There is no consistent theory of quantum gravity. Attempts to develop such a theory led people to the inception of critical string theory in the late 1970s. This theory builds on quantum field theory and, it is hoped, goes beyond it. It is believed that, after its completion, string theory will describe all fundamental interactions in nature, including quantum gravity. However, the completion of superstring theory seems to be in the distant future. Today neither is its mathematical structure clear nor its relevance to real-world phenomena established. A number of encouraging indications remain in disassociated fragments. If there is a definite lesson for us from string theory today, it is that the class of relativistic quantum phenomena to be considered must be expanded as far as possible and that we must explore, to the fullest extent, nonperturbative aspects in the hope of finding a path to quantum geometry, when the time is ripe, probably with many other interesting findings en route.

Finally, a few words on the history of supersymmetry are in order.[5] The history of supersymmetry is exceptional. All other major conceptual developments in physics

[5] For more details see [10].

have occurred because physicists were trying to understand or study some established aspect of nature or to solve some puzzle arising from data. The discovery in the early 1970s of supersymmetry, that is, invariance under the interchange of fermions and bosons, was a purely intellectual achievement, driven by the logic of theoretical development rather than by the pressure of existing data.

The discovery of supersymmetry presents a dramatic story. In 1970 Yuri Golfand and Evgeny Likhtman in Moscow found a superextension of Poincaré algebra and constructed the first four-dimensional field theory with supersymmetry, the (massive) quantum electrodynamics of spinors and scalars.[6] Within a year Dmitry Volkov and Vladimir Akulov in Kharkov suggested nonlinear realizations of supersymmetry and then Volkov and Soroka started developing the foundations of supergravity. Because of the Iron Curtain which existed between the then USSR and the rest of the world, these papers were hardly noticed. Supersymmetry took off after the breakthrough work of Julius Wess and Bruno Zumino in 1973. Their discovery opened to the rest of the community the gates to the Superworld. Their work on supersymmetry has become tightly woven into the fabric of contemporary theoretical physics.

Often students ask where the name "supersymmetry" comes from. The first paper of Wess and Zumino [11] was entitled "Supergauge transformations in four dimensions." A reference to supersymmetry (without any mention the word "gauge") appeared in one of Bruno Zumino's early talks [12]. In the published literature Salam and Strathdee were the first to coin the term *supersymmetry*. In the paper [13], in which these authors constructed supersymmetric Yang–Mills theory, super-symmetry (with a hyphen) was in the title, while in the body of the paper Salam and Strathdee used both the old terminology due to Wess and Zumino, "super-gauge symmetry," and the new one. This paper was received by the editorial office of *Physical Letters* on June 6, 1974, exactly eight months after that of Wess and Zumino [11]. An earlier paper, of Ferrara and Zumino [14] (received by the editorial office of *Nuclear Physics* on 27 May 1974),[7] where the same problem of super-Yang–Mills theory was addressed, mentions only supergauge invariance and supergauge transformations.

References for the Introduction

[1] V. Weisskopf, The development of field theory in the last 50 years, *Physics Today* **34**, 69 (1981).
[2] S. Weinberg, *The Quantum Theory of Fields* (Cambridge University Press, 1995), Vol. 1.
[3] S. Weinberg, Living with infinities [arXiv:0903.0568 [hep-th]].
[4] B. L. Ioffe, private communication.

[6] At approximately the same time, supersymmetry was observed as a world-sheet *two-dimensional* symmetry by string theory pioneers (Ramond, Neveu, Schwarz, Gervais, and Sakita). The realization that the very same superstring theory gave rise to supersymmetry in the target space came much later.

[7] The editorial note says it was received on May 27, 1973. This is certainly a misprint, otherwise the event would be acausal.

[5] L. Landau, in *Niels Bohr and the Development of Physics* (Pergamon Press, New York, 1955), p. 52.

[6] A. Losev, From Berezin integral to Batalin–Vilkovisky formalizm: a mathematical physicist's point of view, in M. Shifman (ed.), *Felix Berezin: Life and Death of the Mastermind of Supermathematics* (World Scientific, Singapore, 2007), p. 3.

[7] T. H. R. Skyrme, *Proc. Roy. Soc. A* **262**, 237 (1961) [reprinted in E. Brown (ed.), *Selected Papers, with Commentary, of Tony Hilton Royle Skyrme* (World Scientific, Singapore, 1994)].

[8] Y. Nambu, General discussion, in Y. Tanikawa (ed.), *Proc. Int. Conf. on Elementary Particles: In Commemoration of the Thirtieth Anniversary of Meson Theory*, Kyoto, September 1965, 327–333.

[9] S. Weinberg, *Phys. Rev. D* **9**, 3357 (1974) [reprinted in R. N. Mohapatra and C. H. Lai, (eds.), *Gauge Theories of Fundamental Interactions* (World Scientific, Singapore, 1981), pp. 581–602].

[10] G. Kane and M. Shifman (eds.), *The Supersymmetric World: The Beginnings of the Theory* (World Scientific, Singapore, 2000); S. Duplij, W. Siegel, and J. Bagger (eds.), *Concise Encyclopedia of Supersymmetry* (Kluwer Academic Publishers, 2004), pp. 1–28.

[11] J. Wess and B. Zumino, Supergauge transformations in four dimensions, *Nucl. Phys. B* **70**, 39 (1974).

[12] B. Zumino, Fermi–Bose supersymmetry (supergauge symmetry in four dimensions), in J. R. Smith (ed.), *Proc. 17th Int. Conf. High Energy Physics*, London, 1974, (Rutherford Lab., 1974).

[13] A. Salam and J. Strathdee, Super-symmetry and non-Abelian gauges, *Phys. Lett. B* **51**, 353 (1974).

[14] S. Ferrara and B. Zumino, Supergauge invariant Yang–Mills theories, *Nucl. Phys. B* **79**, 413 (1974).