The Lesson of Quantum Theory, edited by J. de Boer, E. Dal and O. Ulfbeck

Some Lessons of Renormalization Theory

Arthur S. Wightman

Princeton University,
Princeton, New Jersey, USA

Contents

Introduction 201
1. Zero-point energy 202
2. The off-diagonal density matrix 206
3. The dimension of coupling constants 208
4. Perturbative renormalization theory 209
5. Summing the perturbation series; renormalons and instantons 212
6. Renormalizable versus non-renormalizable–non-perturbatively 219
7. Coupling constant bounds; the ultraviolet phase transition and beyond 222
References 223
Discussion 225

Introduction

In their fundamental paper on quantum mechanics, Born, Heisenberg and Jordan (1926) gave the first quantum mechanical treatment of a system of an infinite number of degrees of freedom: the vibrating string. They encountered the first divergence difficulty of the quantum theory of fields: the zero-point energy of the ground state. They dropped it, thereby performing the first infinite subtraction in the history of renormalization theory. In the course of the sixty years that have since elapsed, the quantum theory of systems of an infinite number of degrees of freedom has been extended greatly in scope and depth and renormalization theory along with it.

By now the term renormalization has a variety of associations both mathematical and physical. On the one hand, renormalization in one broad sense has often come to include any procedure by which infinite or ambiguous expressions in quantum field theory are replaced by well defined mathematical objects. A more precise definition would here distinguish *renormalization* and *regularization*, the latter being any rule that produces finite answers while the former is reserved for the special case in which the rule gives answers associated with a self-consistent field theory. On the other hand, renormalization is often used as a catch word for a family of

methods of analyzing the significant parameters labeling the states of a theory and of their relations to the parameters actually appearing in the Hamiltonian. In the following we shall have the occasion to use renormalization in both these senses.

We are far from having a complete renormalization theory for the quantum theory of fields. The main point of the present chapter is to examine some of the significant developments in the history of renormalization theory for the light they can throw on our present unsolved problems. The text is arranged in the form of seven lessons with commentary.

1. Zero-point energy

Described more explicitly, what Born, Heisenberg and Jordan did was to write a formal expansion of the real valued function u, which specifies the transverse displacement of an unquantized string with ends fixed at 0 and l:

$$u(x, t) = \sum_{k=1}^{\infty} q_k(t) \sqrt{\frac{2}{l}} \sin \frac{\pi k}{l} x, \tag{1.1}$$

where

$$q_k(t) = \int_0^l \mathrm{d}x \sqrt{\frac{2}{l}} \sin \frac{\pi k}{l} x \, u(x, t).$$

If the tension in the string is τ and the mass per unit length is ρ, the equation of motion is

$$\rho \frac{\partial^2 u}{\partial t^2} = \tau \frac{\partial^2 u}{\partial x^2},$$

so insertion of expression (1.1) gives

$$\ddot{q}_k(t) + \omega_k^2 q_k(t) = 0,$$

with $\omega_k = \sqrt{\frac{\tau}{\rho}} \frac{\pi k}{l}$. The total energy is

$$\begin{aligned} H &= \tfrac{1}{2} \int_0^l \mathrm{d}x \left\{ \rho \left(\frac{\partial u}{\partial t} \right)^2 + \tau \left(\frac{\partial u}{\partial x} \right)^2 \right\} \\ &= \tfrac{1}{2} \sum_{k=1}^{\infty} \left[\rho (\dot{q}_k(t))^2 + \left(\frac{k\pi}{l} \right)^2 \tau q_k(t)^2 \right] \\ &= \tfrac{1}{2} \sum_{k=1}^{\infty} \left[P_k^2 + \omega_k^2 Q_k^2 \right], \end{aligned}$$

if we define $P_k = \rho\dot{q}_k$, $Q_k = \rho q_k$. The quantization procedure was (and is today) to quantize each oscillator independently, replacing P_k by the operator $P_k^{\mathrm{op}} = -\mathrm{i}\hbar(\partial/\partial Q_k)$ and Q_k by the multiplication operator Q_k^{op}. Then

$$\tfrac{1}{2}\left[\left(P_k^{\mathrm{op}}\right)^2 + \omega_k^2\left(Q_k^{\mathrm{op}}\right)^2\right] = \hbar\omega_k\left[N_k + \tfrac{1}{2}\right],$$

where N_k is the number operator for the kth oscillator; it has the eigenvalues 0, 1, 2, ... The infinite subtraction of Born, Heisenberg and Jordan referred to above is the omission of the sum $\frac{1}{2}\sum_{k=1}^{\infty}\hbar\omega_k$ from H.

Of course, if this sum were finite its omission would have no effect whatever on the time evolution of observables defined by H:

$$q(0) \to q(t) = \mathrm{e}^{\mathrm{i}Ht}\, q(0)\, \mathrm{e}^{-\mathrm{i}Ht},$$

since such a constant cancels out. Nevertheless, this infinite subtraction raises a question. Does the omission of the zero-point energy mean that the zero-point vibrations are without physical significance? This question has a very unambiguous negative answer in a number of other contexts. For example, the low-lying vibrational states of a molecule can be approximately described in terms of a quantum-mechanical harmonic oscillator and measurements of the ground-state energy relative to the energy of the dissociated atoms can be compared to the minimum potential energy; the result is an ambiguous support for the reality of the zero-point energy.

Born, Heisenberg and Jordan's discussion of the problem was directed toward a resolution of longstanding difficulties in the theory of energy fluctuations in black body radiation. In his famous report on quanta at the 1911 Solvay Conference, Einstein (1911) had displayed a calculation of the energy fluctuations of electromagnetic radiation in equilibrium with atoms, modeled as a set of oscillators. He showed that there were two contributions to the mean square energy fluctuation, only one of which would be present according to classical wave theory. Born, Heisenberg and Jordan found agreement with Einstein's result.

It was some years later that Heisenberg (1931) showed that the three-man derivation is incorrect; the energy fluctuation calculated is in fact infinite. Heisenberg convinced himself that the infinity in question had nothing to do with the zero-point energy by calculating the analogous (infinite) quantity for a free Schrödinger particle for which there is no zero-point energy. He attributed the infinity of the fluctuation to the fact that the fluctuation was calculated for a sharply defined region. If the energy in a region R of space averaged over a time interval $\{t,\ t+T\}$,

$$\frac{1}{T}\int_t^{t+T}\mathrm{d}t\int_R \mathrm{d}^3x\ T_{00}(\boldsymbol{x},\tau),$$

is replaced by

$$\int\int f(x)\,\mathrm{d}^4x\ T_{00}(x),$$

where f is a smooth function, equal to 1 inside R during the time interval $(t, t+T)$, and dropping rapidly to zero outside, then the fluctuation turns out to be finite, provided that the zero-point contribution to $T_{00}(x)$ is dropped. This fact that, in general, quantized fields like $T_{00}(x)$ only make sense as operators on states when they are smeared with sufficiently smooth functions f, was a key idea in the paper of N. Bohr and L. Rosenfeld (1933) on the measurability of the electromagnetic field. For the electromagnetic field, it was often possible to permit f to be discontinuous, say 1 inside a region and zero outside. However, Heisenberg (1934) showed that the smeared charge and current of a Dirac spin $\frac{1}{2}$ field has infinite fluctuations in the vacuum, unless function f has bounded first derivatives. That the energy density and charge density require somewhat smoother test functions is, roughly speaking, a consequence of the fact that they are quadratic in the annihilation and creation operators for photons, whereas the electromagnetic fields are linear. (All these calculations are for free fields, the coupling constant having been set equal to zero.)

Nowadays, following the general mathematical ideas developed by L. Schwartz (1952) in his great treatise of 1951–1952, one regards quantized fields as generalized functions i.e. linear functionals $T(f)$ of a test function f; in Schwartz's terminology $T_{00}(x)$ is an operator-valued distribution. So far, this mathematical concept has proved itself adequate for the purposes of quantum field theory.

While all this discussion of fields as operator-valued distributions is indispensible for understanding field fluctuations, it does not address directly the significance of the zero-point energy, except that it provides the rule: drop the terms in T_{00} which make its expectation value in the vacuum different from zero. This rule still provoked some uneasiness and efforts were made by Rosenfeld and Solomon (1931) and by Pauli (1933) in his "Handbuch" article to write the energy density as a function of fields in such a fashion that the zero-point contribution automatically drops out. For example for the electromagnetic field

$$T_{00}(x) = \tfrac{1}{2}\left[\boldsymbol{E}^2(x) + \boldsymbol{B}^2(x)\right] + \mathrm{i}\left[\boldsymbol{E}(x),\ \frac{1}{\sqrt{-\Delta}}\nabla x\boldsymbol{B}\right],$$

where $\boldsymbol{E}$(x) and $\boldsymbol{B}$(x) are the electric and magnetic field strengths, respectively. (See Pauli's article p. 256.) Exactly what the significance of such an expression is was not clear at the time, apart from the fact that it leads to an energy with no contributions from zero-point vibration.

It was in 1950, long after, that G.C. Wick (1950) introduced the operation that we now call Wick ordering, which permits us to write this expression as

$$T_{00}(x) = \tfrac{1}{2}\left[:\boldsymbol{E}^2:(x) + :\boldsymbol{B}^2:(\boldsymbol{x})\right], \tag{1.2}$$

with the Wick ordering : : defined by

$$:\boldsymbol{E}^2(x):(x) = \lim_{x\to y}\left[\boldsymbol{E}(x)\cdot\boldsymbol{E}(y) - \left(\Phi_0,\ \boldsymbol{E}(x)\cdot\boldsymbol{E}(y)\Phi_0\right)\right]. \tag{1.3}$$

Here Φ_0 is the vacuum state. The effectiveness of this formula can be described in words as follows: The quantity $\boldsymbol{E}(x)\cdot\boldsymbol{E}(y)$ has a singularity at $x=y$; if the

singularity is cancelled in the vacuum expectation value of the quantity by the subtraction of $(\Phi_0, \boldsymbol{E}(x)\cdot\boldsymbol{E}(y)\Phi_0)$, it is cancelled in all physically significant matrix elements.

In retrospect, the occurrence of such singularities in the vacuum expectation value of the product of two fields should not have been a great surprise. The vacuum expectation values are wave-equation solutions intimately related to fundamental solutions and such fundamental solutions are well known to have singularities on the light cone. These singularities had already been calculated explicitly by Jordan and Pauli in 1928. With the wisdom of hindsight, one can say that what should have been surprising is that cancelling singularities in the vacuum expectation value should suffice to cancel all singularities. That fact was the beginning of the general subject of operator-product expansions, of which more will be said later.

That the above procedure for defining the energy density does not eliminate all the physical consequences of electromagnetic zero point vibrations is clear from the Casimir effect. As Casimir (1948) pointed out, the introduction of a pair of metal plates into the vacuum of the electromagnetic field alters the zero-point vibrations of the field and thereby produces an attraction between the plates. The answer would be infinite if calculated naively, but when the zero-point contribution for the vacuum without plates is subtracted, the remainder is finite, depends on the distance between the plates, and agrees with experiment. This result gives a precise meaning to the statement that after the infinite contribution of the zero-point vibration to the energy has been dropped, there remain physically significant finite contributions which should be regarded as consequences of zero-point vibrations.

The question of the proper definition of the energy–momentum density rose again in the 1970s in the context of quantum field theory on curved space–time. The problem there is that no unique vacuum state exists, in general, so that a more refined definition of Wick ordering is necessary. It has been shown that for a class of states Ψ, and a free field which for simplicity will be assumed scalar, the expectation value of the "point-split" energy momentum tensor $T_{\mu\nu}(x, y)$ has singularities of the form $A(x, y)/\mathrm{d}(x, y)$, $B(x, y)\ln \mathrm{d}(x, y)$, where A and B are smooth functions depending in the limit $x \to y$ only on the curvature of the space–time manifold and its derivatives. Here $\mathrm{d}(x, y)$ is the geodesic distance between x and y. Cancellation of these terms provides the required generalization of the cancellations implicit in the definition (1.3) of Wick ordering. This procedure also provides a first step toward a semi-classical theory of gravitation coupled to a quantized field. In such a theory, the expectation value of the energy–momentum tensor is treated as the source in the classical Einstein equations

$$R_{\mu\nu} - \tfrac{1}{2}g_{\mu\nu}R = \lim_{x\to y}\left[\left(\Psi, T_{\mu\nu}(x, y)\Psi\right) - \mathrm{sing}\right]. \tag{1.4}$$

It is instructive to regard this equation together with the operator equation for the scalar field as describing a self-consistent analogue of the Casimir effect. Here, for each metric $g_{\mu\nu}$, one can solve the operator equation for the scalar field and, given Ψ, compute the right-hand side of eq. (1.4). Then the self-consistent $g_{\mu\nu}$ is one for which the left- and right-hand sides coincide.

2. The off-diagonal density matrix

The definition of Wick ordering in eq. (1.3) as

$$\lim_{x \to y} \left[\boldsymbol{E}(x) \cdot \boldsymbol{E}(y) - (\Phi_0, \boldsymbol{E}(x)) \cdot \boldsymbol{E}(y) \Phi_0) \right]$$

is often referred to as "point splitting". In fact, point splitting was first used in another context, the problem of vacuum polarization. In his report to the Solvay Conference of 1933, P.A.M. Dirac (1934a) computed the polarization of the vacuum by a given external electric field according to his hole theory of electrons and positrons and found that, in lowest approximation in an expansion in the electric charge e, an external field produces a polarization charge density ρ, proportional to the external charge density, ρ_{ext}, with an infinite proportionality constant as well as a polarization charge density proportional to $\Delta\rho_{\text{ext}}$:

$$\rho(x) = -\frac{e^2}{\hbar c}\left(\frac{2}{3\pi}\right)\left[\ln\left(\frac{2P}{mc}\right) - \frac{5}{6}\right]\rho_{\text{ext}}(x) - \frac{1}{15\pi}\left(\frac{e^2}{\hbar c}\right)\left(\frac{\hbar}{mc}\right)^2 \Delta\rho_{\text{ext}}(x). \tag{2.1}$$

Here P is a cutoff which should be taken as infinity but which Dirac took as 137 mc for purposes of discussion. He speculated on the possibility that this vacuum polarization might lead to deviations from the Klein–Nishina formula for the Compton effect and the Rutherford formula for Coulomb scattering.

To derive this formula, Dirac introduced a quantity $\langle \boldsymbol{x}, t \mid R \mid \boldsymbol{x}'t' \rangle$ which may be described as the subtracted off-diagonal density matrix. This terminology is partly justified by the fact that, in the absence of an electromagnetic field and before the subtraction, $\langle \boldsymbol{x}, t \mid R \mid \boldsymbol{x}', t' \rangle$ reduces to

$$\sum_{\text{occupied}} \psi_n(\boldsymbol{x}, t)\, \overline{\psi_n(\boldsymbol{x}', t')}, \tag{2.2}$$

where the ψ_n are the eigenfunctions of the Dirac equation in a large box chosen for convenience. The summation is over occupied states. According to the original ideas of Dirac's hole theory of the positron, in the unperturbed vacuum all negative energy states are occupied, but there is a subtraction prescription which causes their contributions to charge, current, and energy density to vanish. A state with a finite number of free positive-energy electrons and positrons is then to be described by a density matrix with holes in the negative-energy sea describing the positrons—the subtraction prescription leads to positive-energy contributions from the holes appropriate to positrons.

In the presence of an external electromagnetic field there are ambiguities in this procedure, because one cannot in general separate unambiguously into positive and negative energy states. Dirac (1934b) found a way to resolve these ambiguities by displaying the characteristic singularities of $\langle \boldsymbol{x}t \mid R \mid \boldsymbol{x}'t' \rangle$ on the light cone $(\boldsymbol{x} - \boldsymbol{x}')^2 - c^2(t - t')^2 = 0$. The cancellation of the infinities to which these singularities give

rise in the charge density etc. then provided the required definition of subtraction. Applied to the vacuum polarization eq. (2.1), it would mean that the first (infinite) term should be dropped—that is, charge renormalization. Dirac's definition of these subtractions had the reputation of being deep and dark. At least that was the way it appeared to me when I heard him lecture on it in the spring of 1947.

Dirac's work on $\langle x \mid R \mid x' \rangle$ was further developed, clarified and generalized by Heisenberg (1934) who introduced the operator

$$R(x, x') = \psi(x)\ \psi(x')^*, \tag{2.3}$$

which for a free Dirac field ψ has an expectation value in the vacuum equal to eq. (2.2). In the presence of an external field incapable of creating pairs, the expectation value of $R(x, x')$ in the vacuum reduces to the $\langle x \mid R \mid x' \rangle$ considered by Dirac. However, Heisenberg was using the charge-symmetric formalism which treats positrons and electrons on an equal footing, so what was a mysterious subtraction of a perturbed negative-energy sea in the hole theory became simply the removal of a singularity on the diagonal $x = x'$ of the operator $R(x, x') = 0$.

Nowadays, for conceptual clarity, Heisenberg's discussion of the quantized Dirac field in an external electromagnetic field would be divided into two parts. In the first, one would solve the Dirac equation for the quantized field ψ in a given external field. It is a linear equation and for a reasonable class of external fields gives rise to no divergences whatsoever. The second part would be to define the electric current and the energy–momentum tensor in terms of the operator R, e.g. the current

$$j^{\mu}(x) = -e \lim_{x' \to x} \mathrm{Tr}\big[(R(x, x') - \text{subtraction terms})\gamma^0\gamma^{\mu}\big], \tag{2.4}$$

where the trace is over the Dirac spinor indices. This arrangement of the argument makes it clear that the choice of the subtraction terms is a matter of definition. If one wants $j^{\mu}(x)$ to exist, and to be conserved $\partial^{\mu} j_{\mu}(x) = 0$, that imposes a constraint on the subtraction terms, a constraint Heisenberg showed could be satisfied. What he did not notice was that with the same subtraction terms the axial current defined by

$$j^{5\mu}(x) = \lim_{x' \to x} \mathrm{Tr}\big[(R(x, x') - \text{subtraction terms})\gamma^0\gamma^5\gamma^{\mu}\big] \tag{2.5}$$

has an anomaly in the sense that

$$\partial^{\mu} j^5_{\mu}(x) = 2m\ \mathrm{Tr}\big[(R(x, x') - \text{subtraction})\gamma^5\big] + \frac{e^2}{12\pi}\boldsymbol{E}(x) \cdot \boldsymbol{B}(x), \tag{2.6}$$

where $\boldsymbol{E}$ and $\boldsymbol{B}$ are the external electric and magnetic fields. In the unquantized theory only the first term would appear on the right-hand side of eq. (2.6). This anomaly was discovered in 1968 by Adler (1969), although in the context of π^0 decay into two photons it had been known since the work of Steinberger (1949) and Schwinger (1951).

Heisenberg also considered the problem of defining $R(x, x')$ in the full quantum electrodynamics in which both the Dirac spinor field and the electromagnetic field are quantized. He noted that the subtraction terms for $R(x, x')$ can then be determined order by order in perturbation theory but that high-order terms are afflicted with divergences similar to that occurring in the electron self-energy and therefore beyond the control of the then existing theory. Progress on that front had to await the development of perturbative renormalization theory.

3. *The dimension of coupling constants*

It was Heisenberg (1936) who first attempted to distinguish theories according to the dimension of their coupling constants. He compared quantum electrodynamics with Fermi's theory of nuclear β decay and noted that while a cross-section for the production of n pairs calculated in lowest-order perturbation theory in quantum electrodynamics (QED) is proportional to α^{2n} where α is the fine-structure constant $\alpha = e^2/\hbar c$, in Fermi's theory it is proportional to $[G_{\text{Fermi}}/\hbar c)k^2]^{2n}$ where k is some typical wave number associated with the reaction. The conclusion is that when k becomes as large as $(\sqrt{G_{\text{Fermi}}/\hbar c})^{-1}$ so the expansion parameter is near 1, perturbation theory should break down and multiple production processes should become frequent. On the other hand, in QED the probability of multiple processes gets smaller as n increases whatever the value of the energy. Heisenberg hoped that this fundamental distinction between the Fermi interaction and the electromagnetic interaction could account for the phenomenon of bursts in cosmic rays. He regarded these arguments and the experimental evidence on bursts as evidence that there is a fundamental length in nature which has to be incorporated into the description of elementary particles. In fact, as Heisenberg noted, the critical value

$$\hbar ck \sim \frac{\hbar c}{\sqrt{G_{\text{Fermi}}/\hbar c}} \sim 600 \text{ GeV}$$

is too high. At first, he tried to appeal to the Konopinski–Uhlenbeck theory of nuclear β-decay to lower this critical value, but later, after the success of the Yukawa theory of mesons, and the development of the cascade theory of electron–photon showers, he abandoned the idea but nevertheless pursued the notion of a fundamental length. He also argued (Heisenberg 1938a) that the Yukawa theory of mesons, because its dimensionless coupling constant $g^2/\hbar c$ is so much larger than the fine-structure constant, should give rise to multiple production of mesons.

In a paper in the Planck Festschrift (Heisenberg 1938b), he gave semi-philosophical arguments for the existence of a fundamental length. He noted that $\hbar$ and c are universal constants in the sense that they correspond to restrictions on physical laws applicable under all circumstances: the condition $0 < c < \infty$ is associated with the requirement that all laws be consistent with the special theory of relativity; the condition $0 < \hbar < \infty$ is associated with the requirement that all physical laws be consistent with quantum mechanics. He proposed that there be a fundamental

length λ of similarly universal character, and suggested that if it was to be associated with multiple meson production, it would be reasonable if $\lambda \sim e^2/mc^2 = 2.8 \times 10^{-13}$ cm. As he noted, the existence of λ would bring a certain tidiness to elementary particle theory: every physical quantity could be given as a dimensionless number times appropriate powers of $\hbar$, c and λ.

Heisenberg despaired of being able to incorporate a fundamental length into conventional field theory. So he was led to develop elementary particle theories based on the S-matrix (Heisenberg 1943). In a sequel (Heisenberg 1944) he gave examples of S-matrix theories with a fundamental length and multiple meson production.

Nevertheless, the distinction between dimensional and non-dimensional coupling constants persisted, and later, in 1952 it was given a deep interpretation in terms of perturbative renormalization theory by Sakata, Umezawa and Kamefuchi (1952). They pointed out that there are three categories of theories according to the dimension of their coupling constants

$$g = [M]^n \quad \text{super-renormalizable,}$$
$$g = [M]^0 \quad \text{renormalizable,}$$
$$g = [M]^{-n} \quad \text{non-renormalizable.}$$

Here $n > 0$. I will discuss the meaning of this distinction in the next section.

4. Perturbative renormalization theory

The field theory of the 1930s was largely based on formal perturbation theory so that solutions, if they could be constructed at all, were expected to be formal power series in the coupling constant. The expansions typically had well defined first terms, but all higher orders were infinite or ambiguous — the so-called divergences. There was no systematic theory giving control of the divergences, until a theory was created in the 1940s by Tomonaga (1948), Feynman (1949), Dyson (1949), Schwinger (1958) and others.

The now familiar methods involve the expansion of the quantity in question, typically an S-matrix element or Green function, into terms corresponding to Feynman diagrams, their regularization to obtain well-defined expressions and finally their renormalization by the isolation of a finite part. The divergent part which has to be separated is located by the method of power counting. That method attributes to an integral of a rational function P/Q, where P and Q are polynomials,

$$\int \frac{P(k)}{Q(k)} \, \mathrm{d}^n k$$

a degree of divergence $D = \deg P - \deg Q + n$. If $D \geq 0$ the integral, in general, turns out to be divergent, while if $D < 0$ it is convergent. The subtle part of the

analysis arises when one deals with the divergence and convergence of subintegrals and shows how to separate contributions which would arise from counter terms in the Lagrangean. If there is at most a finite number of divergent graphs occurring in a theory, the theory is said to be super renormalizable; if an infinite number of divergent graphs occur but they all can be cancelled by a finite number of counter terms, the theory is renormalizable, if the theory has an infinite number of divergent graphs and requires an infinite number of counter terms it is non-renormalizable.

These definitions give a precise meaning to the result of Sakata, Umezawa and Kamefuchi (1952) quoted above: theories with coupling constants which have a dimension of a positive power of mass are super-renormalizable, the ones with dimensionless coupling constants are renormalizable, and the ones with negative power of mass are non-renormalizable. For example:

Super-renormalizable:	$P(\phi)_2$, ϕ_3^4, Y_2, $\mathrm{QED}_{\leq 3}$,
Renormalizable:	QED_4, ϕ_4^4,
Non-renormalizable:	$\mathrm{Fermi}_{\geq 3}, \phi^4_{\geq 5}$.

The success of perturbative renormalization theory was twofold. First, in QED of electrons and muons it yielded theoretical predictions which have been verified to an accuracy greater than that in any other measurement made by mankind. Second, it revolutionized the prevailing attitude toward quantum field theory as a language for describing Nature—perhaps quantum field theory is more consistent than the founding fathers Heisenberg and Pauli believed in 1929.

However, even in perturbative renormalization theory there were loose ends. The problem of overlapping divergences remained somewhat elusive. It was not until Bogoliubov and Shirkov invented the R-operation in 1955 (Bogoliubov and Shirkov 1959), and Hepp (1966), completing the work of Bogoliubov and Parasiuk (1957), gave a proof that the R-operation really works, that perturbative renormalization theory was put on a sound mathematical footing. [Here, I risk offending my friend Abdus Salam; I have never understood whether his papers of 1951 (Salam 1951a, b) really solved the problem of overlapping divergences in QED or not. Maybe I should study them again.] This was followed by Zimmermann's formula of 1969 (Zimmermann 1969) which gave an explicit expression for the R-operation. The resulting BPHZ renormalization scheme is now standard. There were others, analytic renormalization developed by Speer in 1967 (Speer 1968) and dimensional renormalization with numerous parents shortly thereafter. All were shown to be equivalent up to finite renormalizations. I will say more about dimensional renormalization because it has played an important role in recent developments.

The simplest treatment of dimensional renormalization takes as its starting point the so-called α-parametrization of the contribution of a Feynman diagram. It is an integral over a bounded domain in an α-space of which the number of dimensions is determined by the number of lines in the Feynman diagram. The dimension of space–time appears only in the power to which a polynomial in the denominator is raised. The integral turns out to be a meromorphic function of the dimension with poles only on the real axis at the dimensions for which the integral is divergent. Some typical diagrams and their poles for $\lambda\phi_d^4$ theory are shown in fig. 1.

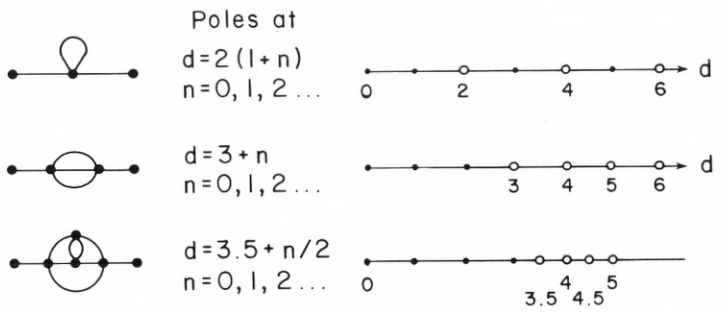

Fig. 1. Some typical diagrams and their dimensional poles for the two-point function in $\lambda\phi_d^4$.

When the poles for all the diagrams contributing to the two-point function are plotted on the same line, one finds that all are positive and ≥ 2. Below $d = 4$ the poles are isolated but have 4 as a limit point, while above $d = 4$, all rational numbers give poles. Thus, viewed from the point of view of dimensional renormalization, the distinction between super-renormalizable ($d < 4$), renormalizable ($d = 4$), and non-renormalizable ($d > 4$) is vividly displayed for $\lambda\phi_d^4$ theory in this distribution of poles.

I would now like to make a digression to describe an analogy between this distribution of dimensional poles and a situation occurring in Poincaré's thesis (1882). This analogy is probably at best poetic, but it may possibly give us some solace in our struggle to understand quantum field theory.

Poincaré's thesis treated, among other things, the normal-form problem for an ordinary differential equation at an equilibrium point. It is formulated as follows. Let x be a real n-dimensional vector and the differential equation be

$$\frac{\mathrm{d}x}{\mathrm{d}t} = F(x),$$

with $F(0) = 0$ and $\mathrm{F}(x) = Ax + O(x^2)$, where A is some real linear transformation. We ask: when does there exist a change of coordinates near the origin $x = h(y)$ so that the equation for y is

$$\frac{\mathrm{d}y}{\mathrm{d}t} = Ay,$$

i.e. so that the differential equation can be reduced by change of coordinates to its linear part? To state Poincaré's result we use a definition: let the eigenvalues of A be $\lambda_1 \ldots \lambda_n$, then A is *resonant* if for some non-negative integers $m_1 \ldots m_n$ and s with $1 \leq s \leq n$ and $\sum_{j=1}^n m_j \geq 2$

$$\lambda_s = \sum_{j=1}^{n} m_j \lambda_j.$$

A belongs to the *Poincaré region* if the origin does not belong to the convex hull of the set of complex numbers $\{\lambda_1, \ldots, \lambda_n\}$.

Theorem (Poincaré) *If F is a formal power series, then h exists as a formal power series if, and only if, A is not resonant.*

If F is analytic in a neighborhood of the origin the series for h converges and defines an analytic change of coordinates if A belongs to the Poincaré region.

In the terminology of V.I. Arnold (1980), whose book I recommend for an account of these matters, the exterior of the Poincaré region is the Siegel region. The normal-form problem for A in the Siegel region remained open for more than half a century until the work of Siegel (1952). The qualitative difference between the two regions is that resonant A are isolated in the Poincaré region and dense in the Siegel region. Below follows the result of Siegel which accounts for his name appearing in the terminology.

Theorem (Siegel) *Let the hypotheses of the preceding theorem be satisfied except that A lies in the Siegel region. Suppose further that for all positive integers $m_1 \dots m_n$ and s such that $\sum m_j = |m| \geq 2$ and $1 \leq s \leq j$ there are positive constants c and ν such that*

$$\left| \lambda_s - \sum_{j=1}^{n} m_j \lambda_j \right| \geq \frac{c}{|m|^{\nu}},$$

then the series for h converges and defines an analytic change of coordinates.

My analogy juxtaposes the Siegel region for A in the normal-form problem and the range of dimensions, $d \geq 4$, for which $\lambda\phi_d^4$ theory is non-renormalizable. Let us hope that it does not take as long for us to understand non-renormalizable field theory as the six decades from Poincaré to Siegel.

5. *Summing the perturbation series; renormalons and instantons*

The output of perturbative renormalization theory is a set of well-defined formal power series for the S matrix elements and Green function of a field theory. A natural question is whether the series converge, if they converge whether the resulting sums define solutions of the theory, and if they do not converge whether some summability method can be used to recover functions which are solutions.

It was argued by Hurst (1952), Thirring (1953) and by Petermann (1953) that the renormalized perturbation series for the two-point function in $\lambda\phi_4^4$ theory is divergent. Their work still left some uncertainties about the possible effect of renormalization in causing cancellations and therefore making convergence possible. (Renormalization changes the sign of some terms which would be positive in the unrenormalized expansion.) It was pointed out by Jaffe (1965) that in $P(\phi)_2$ theory no such difficulty occurs; he gave a complete proof of the divergence of the perturbation series for an arbitrary connected Green function. For $\lambda\phi_3^4$ the difficulty does arise, but de Calan and Rivasseau (1982) showed that by rearrangement of the terms they could obtain a divergent series of positive terms. Although it is

widely believed that the renormalized perturbation series of most field theories diverge, there is still no complete proof in most cases. For example, for spinor QED_4, Dyson (1952) already gave an ingenious and suggestive argument supporting divergence, but there is no complete proof to this day. It is worth noting that there is a counter example: The Euclidean Green functions of the massive sine-Gordon theory in two-dimensional space–time are analytic in the mass for sufficiently small masses, as was shown by Fröhlich in 1975 (Fröhlich 1976).

If the perturbation series for a Green function is divergent, it may still be asymptotic to a solution of the theory. In the mid 1970s as a consequence of the deep results of constructive field theory, the existence of solutions was established for $P(\phi)_2$, $\lambda\phi_3^4$, Y_2 and a number of other super-renormalizable theories. The question whether the perturbation series was asymptotic was then open to direct attack. Dimock (1974) answered it affirmatively for the $\lambda P(\phi)_2$ theory. He proved that the Schwinger functions and Green functions of $\lambda P(\phi)_2$ are infinitely differentiable in λ for $0 \leq \lambda < \infty$ and derivatives from the right of all orders exist at $\lambda = 0$. If the Schwinger function in question is denoted $S(\lambda)$, Taylor's theorem with remainder then gives

$$\left| S(\lambda) - \sum_{n=0}^{N} c_n \lambda^n \right| = 0(\lambda^{N+1}) \quad \text{near } \lambda = 0, \tag{5.1}$$

where

$$c_n = \frac{1}{n!} \frac{\mathrm{d}^n S(\lambda)}{\mathrm{d}\lambda^n}\bigg|_{\lambda=0+}. \tag{5.2}$$

The estimate (5.1) is by definition what is meant by the formal power series $\Sigma_{n=0}^{n} c_n \lambda^n$ being asymptotic to $S(\lambda)$: $f(\lambda \sim \Sigma_{n=0}^{\infty} c_n \lambda^n$. Here, at last, was proof that the traditional expansions in terms of the contributions from Feynman diagrams, when renormalized, give an asymptotic series for the solution of the theory.

The knowledge of an asymptotic series for a function does not permit one to determine a function uniquely, but with further information it can happen that not only is the function uniquely determined but that a more or less explicit method exists for recovering it from the coefficients of the asymptotic series. Such methods are called summability methods. I will describe briefly two such methods and then state some of the results obtained by applying them in field theory.

The Padé method gives for each formal power series a table of rational functions labeled by a pair of positive integers

$$f^{[M,N]}(\lambda) = \frac{P^{[M]}(\lambda)}{Q^{[N]}(\lambda)}, \tag{5.3}$$

where P is a polynomial of degree M, and Q a polynomial of degree N. P and Q are so determined that $f^{[M,N]}$ has a Taylor series agreeing with the given formal series up to terms of degree $M + N + 1$, which is the number of free parameters in

the ratio. A sufficient condition that f be recoverable from diagonal sequence of Padé approximants is given by the following theorem of Stieltjes.

Theorem *Let f be a function of the form*

$$f(\lambda) = \int_0^\infty \frac{d\rho(t)}{1+\lambda t},$$

where ρ is a positive measure with finite moments of all orders

$$\int_0^\infty t^n \, d\,\rho(t) < \infty, \quad n = 0, 1, 2, \ldots .$$

Then $f(\lambda)$ has an asymptotic expansion

$$f(\lambda) \approx \sum_{n=0}^{\infty} c_n \lambda^n, \quad \textit{with} \quad c_n = (-1)^n \int_0^\infty t^n \, d\rho(t),$$

and

$$f(\lambda) = \lim_{N \to \infty} f^{[N, N+j]}(\lambda),$$

for $j = 0, 1, 2, \ldots$, and the convergence is uniform in any compact set of the complex plane cut along the negative real axis.

In the late sixties the Padé method was tried on a test problem, the eigenvalues of the quartic anharmonic oscillator, a system which can be regarded as $\lambda\phi_1^4$, a theory of a Hermitean scalar field in space–time of one dimension. The Hamiltonian is taken as

$$H = \tfrac{1}{2}(\pi^2 + m^2\phi^2) + \lambda\phi^4, \quad \pi = -i\frac{d}{d\phi}, \tag{5.4}$$

with m, $\lambda > 0$. Regarded as acting in the space of square integrable functions on the real line, H has a pure discrete spectrum bounded below. By a scaling argument one has for the ground state energy

$$E_0(m^2, \lambda) = mE_0\left(1, \frac{\lambda}{m^3}\right). \tag{5.5}$$

Bender and Wu in 1968 showed that the perturbation series for $E_0(1, \lambda/m^3)$ in the parameter (λ/m^3) is divergent (Bender and Wu 1969). Loeffel et al. (1969) showed that the eigenvalues are functions satisfying the hypotheses of Stieltjes theorem. Thus, the Padé method works for the eigenvalues of the quartic anharmonic oscillator with $m^2 > 0$. Unfortunately, it turns out that, if the quartic anharmonicity $\lambda\phi^4$ is replaced by $\lambda\phi^{2n}$ with $n > 3$, one has to modify the Padé method as a

function of n so one cannot use the above theorem to prove the convergence of the diagonal Padé approximants. Fortunately, as was shown by Graffi, Grecchi and Simon (1970), there is a more powerful summability method, the Borel summability, which does cover these cases. To this Borel summability I will now turn.

If f has the asymptotic expansion $f \sim \sum_{n=0}^{\infty} c_n \lambda^n$, its Borel transform f_B is defined as the asymptotic series

$$f_B \sim \sum_{n=0}^{\infty} \frac{c_n}{n!} \lambda^n,$$

f_B may converge even if the series for f diverges. Then under conditions first obtained by Nevanlinna in 1919, one can recover f. [See also the paper by A. Sokal (1980).]

Theorem *If f is analytic in the interior of the circle $C_R: \mathrm{Re}(1/z) = 1/R$ and its derivatives satisfy*

$$\left| \frac{\mathrm{d}^n f}{\mathrm{d}\lambda^n}(\lambda) \right| \leq A\sigma^n (n!)^2$$

in the closure of the region then the Borel transform, f_B, of the Taylor series of f at the origin

$$f_B(t) = \sum_{n=0}^{\infty} \frac{t^n}{(n!)^2} \left(\frac{\mathrm{d}^n f}{\mathrm{d}x^n} \right) \Bigg|_{\lambda=0},$$

is analytic in the half strip

$$\mathrm{d}(t, \boldsymbol{R}^+) \leq \frac{1}{\sigma},$$

and satisfies the inequality

$$|f_B(t)| \leq C\, \mathrm{e}^{t/R}, \quad 0 \leq t < \infty.$$

Furthermore, f can be recovered from f_B by the inverse Borel transform

$$f(\lambda) = \frac{1}{\lambda} \int_0^{\infty} \mathrm{e}^{-t/\lambda} f_B(t)\, \mathrm{d}t, \tag{5.6}$$

valid for λ in the interior of C_R.

One of the pleasant consequences of the constructive field theory treatment of the $\lambda P(\phi)_2$ and $\lambda\phi_3^4$ field theory is the proof that, for degree $P = 4$, the perturbation series for the Euclidean Green functions are Borel summable. This was shown by Eckmann, Magnen and Sénéor (1975) for $P(\phi)_2$ and by Magnen and Sénéor

(1977) for $\lambda\phi_3^4$. One could speculate that Borel summability might provide a general method for obtaining solutions of quantum field theories from their perturbation series. However, this illusion did not last long. It was soon realized that things are likely to be much more complicated. Two of the additional complications are indicated by the catchwords *instantons* and *renormalons*. Both give rise to singularities in the Borel transform of perturbation series and these singularities sometimes appear on the positive real axis, making the Borel inversion (5.6) inapplicable as it stands. Thus to understand the physical meaning of these singularities one must understand the physical mechanisms for producing instantons and renormalons. Of course, assuming that one can understand their occurrence, and prove that no other singularities occur, one still has the problem of recovering the non-perturbative solutions from the Borel transform in the presence of such singularities. Such questions are at or beyond the limits of our present understanding of the perturbation series, so what follows is an impressionistic account of recent developments with an emphasis which reflects my personal taste.

Instantons first appeared in semi-classical approximations of functional–integral solutions of quantum field theories. They are solutions of the classical field equations which give stationary points of the action appearing in the quantum mechanical expressions. They played a significant role in answering the following basic question: how can a field theory of bosons imply the existence of fermions. It was Skyrme (1961) who studied what is now known as the sine-Gordon equation and attempted to construct a fermion theory which would contain it. To say that Skyrme's work was not widely understood at the time is surely an understatement. A second contribution concerning this question, but with an entirely different flavor, was the work of Haag and Kastler (1964) on algebras of observables. In their general framework, the Hilbert space of states decomposes into sectors in each of which there is a representation of the algebra of operators generated by local observables. The representations are unitary inequivalent in different sectors, but the corresponding states look like the vacuum state in the limit in which all observations take place far from one another and at approximately the same time. If one is given any one of the representations all the others are uniquely determined. If this general theory is applied to the sine-Gordon equation with a vacuum sector determined by the sine-Gordon vacuum, it turns out that there is an infinity of other sectors labelled by the (non-vanishing) fermion number of Skyrme's fermions. However, Haag and Kastler did not apply their general theory to Lagrangean field theory models and several years passed before Streater and Wilde (1970) worked out the details for the special case in which the sine-Gordon equation reduces to the massless free wave equation. Even then it took some time before it was realized by Streater and Dell'Antonio that the fermion theory in question is the massless Thirring model. Finally, Coleman (1975) gave the connection between the general sine-Gordon model and the massive Thirring model.

The second important role played by instantons involves tunneling phenomena. When the potential occurring in a classical action has several degenerate minima, quantum mechanical tunneling between them has an important effect on the structure of low-lying states. The non-degeneracy of the ground state depends on the existence of instantons connecting the degenerate minima. This non-degeneracy

is a non-perturbative effect; the splitting of the ground and first excited state as a function of the coupling constant has a zero-perturbative expansion. Thus, here the presence of instantons invalidates Borel summability. The simplest case in which this phenomenon can be studied is the anharmonic oscillator, the $\lambda\phi_1^4$ theory whose Hamiltonian is displayed in eq. (5.4). By the same scaling argument that produced eq. (5.5), we have

$$E_n(m^2, \lambda) = \lambda^{1/3} E_n(m^2 \lambda^{-2/3}, 1).$$

For a field with $m^2 > 0$, the case discussed in eq. (5.5), the right-hand side is analytic in λ for λ on the three sheeted Riemann surface of the cube root, except at a set of singularities which have zero as a limit point and lie inside two horn-shaped regions on the second and third sheets. When $m^2 = 0$, these singularities all disappear into the origin only to reappear on the first sheet when $m^2 < 0$, but where? Parisi (1977) argued that there would be two horns tangent to the real axis containing the singularities. Crutchfield (1979) argued that they would be tangent to the imaginary axis and on circles tangent at the origin. He based his statement on a semi-classical approximation. If Parisi is right the violation of the hypotheses of analyticity in Nevanlinna's theorem on Borel summability is blatant. On the other hand, Crutchfield has to appeal to a failure in the remainder estimate to account for the failure of Borel summability. To my knowledge, there has as yet been no rigorous discussion, which is a bit of a scandal, especially since Zinn-Justin (1982) has conjectured an intriguing and numerically very successful asymptotic expansion involving logarithms as well as powers of the coupling constant.

In his Erice lectures of 1977, 't Hooft (1979) analyzed the singularities of the two-point function of various theories using renormalization group methods to relate singularities in the coupling constant to singularities in momentum–space. He concluded that for asymptotically free theories there is a horn of singularities in the right-half plane like that described by Parisi (1977).

Instantons are intimately related to the asymptotics of the perturbation series. Such a connection was suggested in 1968 by Bender and Wu (1969) for the anharmonic oscillator and generalized to a finite number of dimensions, but it was Lipatov (1977) who gave a general steepest-descent method usable in quantum field theory. Brezin, Le Guillou and Zinn-Justin (1977) perfected the method and applied it to a wide variety of examples. The result is that the nth coefficient of a

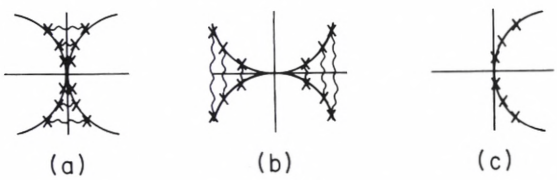

Fig. 2. The singularities in λ of $E_n(m^2, \lambda)$ for (a) $m^2 > 0$, according to Bender and Wu (1969) and Simon (1970); (b) $m^2 < 0$, according to Parisi (1977); (c) $m^2 < 0$, according to Crutchfield (1979).

perturbation series is typically of the form

$$(-a)^n n^b n! A\left[1 + 0\left(\frac{1}{n}\right)\right],$$

where a turns out to be the reciprocal of the radius of convergence of the Taylor series for the Borel transform at the origin. Since a can also be expressed in terms of the instanton solution it is natural to call the corresponding singularity an instanton singularity. The Lipatov method for ϕ_2^4 was put on a rigorous footing in work by Breen (1983). Magnen and Rivasseau (1986) recently showed that the Lipatov method correctly gives the radius of convergence of the Taylor series for the Borel transform in the $\lambda\phi_3^4$ model.

Now let me turn to renormalons. They turned up in various connections in the work of Gross and Neveu (1974), of Lautrup (1977) and of 't Hooft (1977). For the massitive $\lambda\phi_d^4$ model, they are absent for $d = 1, 2, 3$ and first appear for $d = 4$, when renormalization produces logarithmic corrections to the momentum dependence of the propagator. An $(n+k)$th order Feynman diagram containing a chain of n bubbles then has an amplitude proportional to $n!$ If all $(n+k)$th order Feynman diagrams were of this magnitude, the $(n+k)$th series coefficient for the Borel transform would be of order $Ac^{n+k}\lambda^{n+k}(n+k)!$ instead of $Ac^{n+k}\lambda^{n+k}$ and the series for the Borel transform would diverge for all non-vanishing values of the coupling constant. Thus, the result of de Calan and Rivasseau (1981), that the total nth order contribution to the series for the Borel transform is bounded by $Ac^n\lambda^n$, is decidedly non-trivial. It gives a toehold for a possible non-perturbative approach by analytic continuation to the definition of renormalon singularities. So far no one seems to have succeeded along this line.

From each renormalon singularity trails a cut of the Borel transform. It was Parisi (1979a) who first attempted to calculate the jump of the Borel transform across these cuts and recognized that they are related to matrix elements of composite operators of arbitrarily high dimension. Parisi's conjectures were established to all orders in a $1/N$ expansion for the $\lambda(\boldsymbol{\phi}^2)_4^2$ model by Bergère and David (1986a, b). In 1979, Parisi (1979b) gave an analogous discussion of the infrared singularities in massless theories, showing that multilocal counter-terms can cancel the divergences. It had already been known since the work of Symanzik (1973) that these infrared singularities require counter terms which are nonanalytic in the coupling constant. Parisi offered evidence that the resulting finite perturbation expansion has a Borel transform with characteristic singularities which it is natural to call infrared renormalons. In 1985, Bergère and David (1986a, b) were able to verify Parisi's conjectures to all orders in a $1/N$ expansion.

This information on ultraviolet and infrared renormalons together with the knowledge of instantons represents qualitatively new information about the nature of the solutions of renormalizable and super-renormalizable field theories which has not been given a complete nonperturbative treatment in the sense of constructive quantum field theory. If past experience is any guide, a full understanding will require a nonperturbative construction of solutions without reference to the Borel transform of the perturbation series. Then a posteriori, as was the case for $P(\phi)_2$

and $\lambda\phi_3^4$, one may verify that the solution is uniquely determined from its perturbation expansion. Current work using rigorous renormalization group methods by Feldman et al. (1986), Gawedzki and Kupiainen (1985), Balaban (1984, 1985), King (1986) and by Gallavotti and Nicolo (1985) give constructions of the Gross–Neveu model in two- and three-dimensional space–time (the first non-trivial solutions of renormalizable and nonrenormalizable theories, respectively), as well as the first steps toward a constructive field theory treatment of gauge field theories in four dimensions. These developments are taking place rapidly, at least by the standards of speed of constructive field theory, so it is perhaps not too optimistic to hope that in a few years we may finally understand in what sense the perturbation series of a quantum field theory determine solutions.

6. *Renormalizable versus non-renormalizable–non-perturbatively*

It is characteristic of the axiomatic field theory that was invented in the early 1950s that the fields are operator-valued distributions whose definition involves a choice of test functions. From the beginning, the customary assumption was that the test functions are infinitely differentiable functions of fast decrease at infinity. More precisely, the test functions are assumed to be in Schwartz's space $\mathscr{S}(\boldsymbol{R}^4)$. The space $\mathscr{S}(\boldsymbol{R}^n)$ consists of all infinitely differentiable functions, f, on $\boldsymbol{R}^n$ for which the following semi-norms are finite:

$$\| f \|_{k,l} = \sup_{x \in \boldsymbol{R}^n} \sup_{|\alpha| \leq k} \sup_{|\beta| \leq l} | x^\alpha D^\beta f(x) |, \tag{6.0}$$

where

$$x^\alpha = x_1^{\alpha_1} x_2^{\alpha_2} \dots x_n^{\alpha_n}, \quad D^\beta = \left(\frac{\partial}{\partial x_1} \right)^{\beta_1} \dots \left(\frac{\partial}{\partial x_n} \right)^{\beta_n},$$

$$|\alpha| = \sum_{j=1}^{n} \alpha_j, \qquad |\beta| = \sum_{j=1}^{n} \beta_j.$$

Then the quantized fields are tempered operator-valued distributions, by assumption, i.e. the matrix elements, $(\Phi, \phi(f)\Psi)$, regarded as linear functionals in f are elements of Schwartz's space $\mathscr{S}'(\boldsymbol{R}^n)$.

Although this assumption of temperedness was supported by evidence of perturbation theory, it was clear that alternatives are possible, and might be necessary for the construction of some theories. For example, if the matrix elements $(\Phi, \phi(x)\Psi)$ of the field increase too rapidly as $x \to \infty$ one could restrict the test functions to have compact support. With an appropriate definition of convergence, these functions constitute Schwartz's space $\mathscr{D}(\boldsymbol{R}^n)$. The corresponding continuous linear functionals form the space $\mathscr{D}'(\boldsymbol{R}^n)$ of distributions, and one would be led to assume $(\Phi, \phi(f)\Psi) \in \mathscr{D}'(\boldsymbol{R}^n)$.

Replacement of $\mathscr{S}'(\boldsymbol{R}^n)$ by $\mathscr{D}'(\boldsymbol{R}^n)$ enables one to treat fields which are singular at $x=\infty$. To treat fields which are more singular than distributions in neighborhoods of a finite point, requires quite a different set of test functions. The standard example arises if one attempts to define the exponential of a free scalar field. [This example was already studied in the 1950s; see Wightman (1981b) for a brief review with references.] Formally,

$$:\exp g\phi:(x)=\sum_{h=0}^{\infty}\frac{g^n}{n!}:\phi^n:(x) \tag{6.1}$$

and a ready calculation shows

$$(\Phi_0, :\exp g\phi:(x) :\exp g\phi:(y)\Phi_0)=\exp\bigl[g^2(\Phi_0, \phi(x)\ \phi(y)\Phi_0)\bigr]. \tag{6.2}$$

Whatever the choice of operator convergence in eq. (6.1) it is reasonable to expect eq. (6.2) to hold. Now in space–time of four dimensions, the worst singularity of $F(x-y)=(\Phi_0, \phi(x)\ \phi(y)\Phi_0)$ is const. $[(x-y)^2]^{-1}$. Furthermore, a distribution in $\mathscr{S}'(\boldsymbol{R}^4)$ or $\mathscr{D}'(\boldsymbol{R}^4)$ can have singularities which are at worst bounded by a power, so whatever one does to define $:\exp g\phi:$ it cannot be a distribution in $\mathscr{S}'(\boldsymbol{R}^4)$ or $\mathscr{D}'(\boldsymbol{R}^4)$. [A rigorous argument to back up this somewhat loose statement goes as follows. If the two point function is in $\mathscr{D}'(\boldsymbol{R}^n)$ and satisfies the positivity condition

$$\int\bar{f}(x)\ F(x-y)\ f(y)\ \mathrm{d}^4x\ \mathrm{d}^4y\geq 0 \quad \text{for } f\in\mathscr{D}(\boldsymbol{R}^4),$$

then F must be tempered and the Fourier transform must be of a positive measure on momentum space. By the spectral condition this measure must vanish outside the positive cone $E^2-\boldsymbol{p}^2\geq 0$, $E\geq 0$ and therefore the two-point function has an analytic continuation $F(y-x+\mathrm{i}\eta)$, where η lies in the future light cone. This analytic function is Lorentz invariant and so a function of $(y-x+\mathrm{i}\eta)^2$. It can grow no faster than const. $[((y-x)+\mathrm{i}\eta)^2]^{-k}$ for some k, as $\eta\to 0$, unlike eq. (6.2).]

There is a wide variety of other classes of generalized functions which are candidates for the treatment of such an example. I mention first hyperfunctions in the sense of Sato, because Nagamachi and Mugibayasi have developed a general theory of hyperfunction fields and proved that the exponential function $:\exp g\phi:$ belongs to that class. (For the general theory see their papers of 1976; for the exponential function their paper of 1986.) I also mention the Jaffe (1967) classes which have the conceptual advantage that their test functions include many functions of compact support. Fields evaluated on test function of compact support should commute (or anti-commute) when the supports are space-like separated. Since hyperfunctions use analytic test functions, for them there are no test functions of compact support. It is necessary to introduce the notion of the carrier of a hyperfunction to formulate local commutativity for hyperfunction fields.

Most of the preceding distinctions and some others were discussed at length, but in a somewhat different terminology in a paper by Schroer (1964), who argued that

field theories that are non-renormalizable in the sense of perturbation theory should require generalized functions worse than distributions. This would give a non-perturbative definition of non-renormalizability.

There is evidence that it would be wiser to make the distinction without trying to force the definition. There may well be theories with tempered fields which are non-renormalizable in the sense of perturbation theory. Gawedzki and Kupiainen (1985) have announced an existence theorem for solutions of (Gross–Neveu)$_3$, just such a theory. On the other hand in 1984 Gallavotti and Nicolo (1985) have shown that an infinite number of distinct renormalization counter-terms is necessary to extend the range of the coupling constant over the interval, $4\pi \leq \beta^2 < 8\pi$ for the sine-Gordon equation. The situation for $\beta^2 > 8\pi$ is not yet clear. Schroer and Truong (1977) conjectured that there would be non-renormalizable solutions in the sense of solutions which are not distributions.

Furthermore, there are other distinctions within the class of perturbatively non-renormalizable theories. The Green functions of a theory may fail to be C^∞ in the coupling constant because all sufficiently high derivatives blow up. On the other hand, for theories with instantons there are contributions to the Green functions which are C^∞ but have all zero derivatives. All these distinctions appear in a major reexamination of the non-renormalizable field theories undertaken by Symanzik (1975) and by Parisi (1975). On the basis of this exploratory work, there is now major effort in constructive field theory to understand the structure of non-renormalizable and renormalizable theories. The work of Felder and Gallavotti (1985) deserves mentioning in this connection.

The problem is particularly acute with $\lambda\phi_4^4$. There are no-go theorems of Fröhlich (1982) and of Aizenman and Graham (1983) which assert that the traditional ferromagnetic lattice approximation converges to a trivial solution, if the fields have anomalous dimension as everyone expects. The puzzle is: if the only non-perturbative solution is trivial, how can it have, as it does, a non-trivial renormalized perturbation series? Gallavotti in his review of 1985 concludes that there could be other approximation procedures which would converge to a non-trivial solution. The evidence he displays does not say much about the necessity of solutions which would be generalized functions more singular than distributions. Rigorous methods have not yet penetrated deeply enough so that one can see the effects of the renormalon singularities described above. As was mentioned there, the results of Parisi (1979a) and Bergère and David (1986a, b) indicate that to control the solution of $\lambda\phi_4^4$ one will have to control the powers ϕ^{2n} for all integers $n \geq 2$, and would seem to say that the solution of $\lambda\phi_4^4$ would have more to do with the non-renormalizable theories $\lambda\phi_{4+\epsilon}^4$ than the super-renormalizable theories $\lambda\phi_{4-\epsilon}^4$. This is a point of view that I have been urging for some time, the cogency of which has not been clear, for lack of rigorous results on non-renormalizable theories, (Wightman 1977, 1979, 1981a, b).

From a physical point of view, the striking feature of the recent work on non-renormalizable theories is that the solutions have only a finite number of arbitrary parameters. This is contrary to their description in terms of the perturbation series, but not unreasonable; an illegitimate expansion of a function can easily give rise to an infinite number of parameters to be "renormalized".

7. Coupling constant bounds; the ultraviolet phase transition and beyond

One of the discoveries of constructive quantum field theory in the 1970s was the existence of coupling constant bounds in $\lambda\phi^4_\nu$ theories. Originally found by Glimm and Jaffe (1975) in a form valid for all space–time dimensions, ν, these bounds were improved to be optimal for $\nu = 0, 1$ by Newman (1981). Newman considered the dimensionless scale-invariant coupling constant

$$g = -m^\nu \frac{\bar{u}_4}{(\bar{u}_2)^2},$$

where m is the mass gap between the vacuum Ψ_0, and the first excited state, $\bar{u}_4$ is the integrated four-point Ursell function

$$\bar{u}_4 = \int\int\int \mathrm{d}^\nu x_2\, \mathrm{d}^\nu x_3\, \mathrm{d}^\nu x_4\, u_4(x_1, x_2, x_3, x_4),$$

where $u_4(x_1, x_2, x_3, x_4)$ is the analytic continuation to imaginary time of

$$\begin{aligned}
&(\Psi_0, \phi(x_1)\,\phi(x_2)\,\phi(x_3)\,\phi(x_4)\Psi_0)\\
&\quad -(\Psi_0, \phi(x_1)\,\phi(x_2)\Psi_0)(\Psi_0, \phi(x_3)\,\phi(x_4)\Psi_0)\\
&\quad -(\Psi_0, \phi(x_1)\,\phi(x_3)\Psi_0)(\Psi_0, \phi(x_2)\,\phi(x_4)\Psi_0)\\
&\quad -(\psi_0, \phi(x_1)\,\phi(x_4)\Psi_0)(\Psi_0, \phi(x_2)\,\phi(x_3)\Psi_0)
\end{aligned}$$

and $\bar{u}_2$ is the integrated two-point Ursell function

$$\bar{u}_2 = \int \mathrm{d}^\nu x_2\, u_2(x_1, x_2),$$

where $u_2(x_1, x_2)$ is the analytic continuation to imaginary time of $(\Psi_0, \phi(x_1)\,\phi(x_2)\Psi_0)$. He proved

$$\nu = 0, \qquad 0 \le g \le 2,$$

$$\nu = 1, \qquad 0 \le g \le 6,$$

and showed that the upper bound is reached only for the Ising model in which $[\phi(x)]^2$ is equal to a constant. Analogous statements for $\nu \ge 4$ alluded to above can be interpreted as upper bounds of zero analogous to those above. It is a striking feature of Newman's results that he has to make no assumptions about his Hamiltonian except that is has a mass gap $m > 0$.

An obvious question suggested by these bounds: is there any solution for g above or below the bounds? This question was investigated in 1985 by Baker and Wightman (1986). They showed that a one-parameter family of solutions of the

equations of motion for $\nu = 0$, found by Sokal (1982) gave theories violating the bounds above and below. However, an analogous procedure for $\nu = 1$, gave no satisfactory example. Baker and Wightman attributed this to the fact that the form of the equations of motion they assumed is not appropriate for a solution beyond the bounds, when $\nu = 1$.

It is natural to seek guidance on the above question from exactly soluble models. Here the work of Ruijsenaars (1983) on the Federbush model deserves mentioning. Recall that the Federbush model is a parity violating Fermi interaction between two massive fermion fields $\psi^{(1)}$ and $\psi^{(2)}$ of the form $\lambda j_\mu^{(1)} \epsilon^{\mu\nu} j_\nu^{(2)}$. Ruijsenaars constructed operator solutions for λ in the restricted range $|\lambda| \leq 2\pi$ and conjectured that the theory makes sense as a non-renormalizable field theory for $|\lambda| > 2\pi$.

For $-\lambda\phi_4^4$, i.e. $\lambda\phi_4^4$ with the "wrong" sign for the coupling constant, it was pointed out by Symanzik (1973a, b) that asymptotic freedom holds, so one might hope for an easier existence proof for solutions. First 't Hooft (1982) and then Rivasseau (1984) gave a complete construction for planar $-\lambda\phi_4^4$, and then Feldman et al. (1986) constructed the Green functions for $-\lambda\phi_4^4$ without the planar restriction. Presumably, neither of these solutions gives rise to an honest quantum dynamics. The powerful rigorous renormalization group methods used for the latter and the ingenious special methods for the former do not prove reflection positivity, which is necessary to give a positive probability interpretation to a field theory formulated in the Euclidean world.

I draw the conclusion from all this that one should take the possibility of a phase beyond the ultraviolet phase transition seriously and should not exclude a priori the possibility of non-renormalizable solutions or solutions in indefinite metric.

In these seven little lessons I hope to have offered evidence that there is still a lot of life left in quantum field theory.

References

Adler, S., 1969, Phys. Rev. **177**, 2426.

Aizenman, M., and R. Graham, 1983, Nucl. Phys. **B225**, 261.

Arnold, V.I., 1980, Chapitres Supplémentaires de la Théorie des Équations Différentielles Ordinaires (Mir, Moscow).

Baker, G., and A. Wightman, 1986, Trying to violate the coupling constant bound, to appear in the Umezawa Festschrift.

Balaban, T., 1984, Commun. Math. Phys. **96**, 223.

Balaban, T., 1985, Commun. Math. Phys. **99**, 75.

Bender, C., and T.T. Wu, 1969, Phys. Rev. **184**, 1231.

Bell, J., and R. Jackiw, 1969, Nuovo Cimento **60A**, 47.

Bergère, M., and F. David, 1986a, Nonanalyticity of the Perturbative Expansion of Super-renormalizable Massless Field Theories, Ann. Phys., to appear.

Bergère, M. and F. David, 1986b, Ambiguities of renormalized ϕ_4^4 field theory and the singularities of its Borel transform, to appear.

Bogoliubov, N., and O. Parasiuk, 1957, Acta Math. **97**, 227.

Bogoliubov, N., and D. Shirkov, 1959, Introduction to the Theory of Quantized Fields (Interscience, New York).

Bohr, N., and L. Rosenfeld, 1933, Mat.-Fys. Medd. Dan. Vidensk. Selsk. **XII-8**, 1.

Born, M., W. Heisenberg and P. Jordan, 1926, Z. Phys. **35**, 557.

Breen, S., 1983, Commun. Math. Phys. **92**, 179.
Brezin, E., 1977, Phys. Rev. **D15**, 1544, 1558.
Brezin, E., J. Le Gillou and J. Zinn-Justin, 1977, Phys. Rev. **D15**, 1544, 1558.
Brydges, D., J. Fröhlich, T. Spencer, C. de Calan and V. Rivasseau, 1981, Commun. Math. Phys. **82**, 69.
Casimir, H.B.G., 1948, Proc. K. Ned. Akad. Wet. Amsterdam **5**, 793.
Coleman, S., 1975, Phys. Rev. **D11**, 2088.
Crutchfield II, W.Y., 1979, Phys. Rev. **D19**, 2370.
de Calan, C., and V. Rivasseau, 1981, Commun. Math. Phys. **82**, 69.
de Calan, C., and V. Rivasseau, 1982, Commun. Math. Phys. **83**, 77.
Dimock, J., 1974, Commun. Math. Phys. 35, 347;
Dimock, J., 1976, J. Funct. Anal. **21**, 340.
Dirac, P.A.M., 1934a, Institut Solvay Septième Conseil de Physique, Bruxelles, 22–29 Octobre 1933 (Gauthier-Villars, Paris) pp. 203–219.
Dirac, P.A.M., 1934b, Proc. Comb. Phil. Soc. **30**, 150.
Dyson, F.J., 1949, Phys. Rev. **75**, 1736.
Dyson, F.J., 1952, Phys. Rev. **85**, 631.
Eckmann, J.P., J. Magnen and R. Sénéor, 1975, Commun. Math. Phys. **39**, 251.
Einstein, A., 1911, Institut Solvay, 1911 Conseil de Physique (Gauthier-Villars, Paris) pp. 407–415.
Felder, G., and G. Gallavotti, 1986, Perturbation Theory and Non-Renormalizable Scalar Fields, to appear.
Feldman, J., J. Magnen, R. Sénéor and V. Rivasseau, 1986, A Renormalizable Field Theory: The Massive Gross–Neveu Model in Two Dimensions, to appear.
Feynman, R., 1949, Phys. Rev. **76**, 769.
Fröhlich, J., 1976, in: Renormalization Theory, ed. G. Velo, A. Wightman (Reidel, Dordrecht) pp. 371–414.
Fröhlich, J., 1982, Nucl. Phys. **B200**, 281.
Gallavotti, G., 1985, Rev. Mod. Phys. **57**, 471.
Gallavotti, G., and F. Nicolo, 1985, Commun. Math. Phys. **100**, 545.
Gawedzki, K., and A. Kupiainen, 1985, Phys. Rev. Lett. **54**, 2192.
Glimm, J., and A. Jaffe, 1975, Ann. Inst. H. Poincaré, **XXII**, 97.
Graffi, S., V. Grecchi and B. Simon, 1970, Phys. Lett. **32B**, 631.
Gross, D., and A. Neveu, 1974, Phys. Rev. **D10**, 3235.
Haag, R., and D. Kastler, 1964, J. Math. Phys. **5**, 747.
Heisenberg, W., 1931, Sitzber. Sächs. Akad. Leipzig **83**, 3.
Heisenberg, W., 1934, Z. Phys. **90**, 209.
Heisenberg, W., 1936, Z. Phys. **101**, 533.
Heisenberg, W., 1938a, Z. Phys. **110**, 251.
Heisenberg, W., 1938b, Ann. Phys. **32**, 20.
Heisenberg, W., 1943, Z. Phys. **120**, 513, 673.
Heisenberg, W., 1944, Z. Phys. **123**, 93.
Hepp, K., 1966, Commun. Math. Phys. **2**, 301.
Hurst, C.A., 1952, Proc. Camb. Phil. Soc. **48**, 625.
Jaffe, A., 1965, Commun. Math. Phys. **1**, 127.
Jaffe, A., 1967, Phys. Rev. **158**, 1454.
Jordan, P., and W. Pauli, 1928, Z. Phys. **47**, 151.
King, C., 1986, Commun. Math. Phys., **103**, 323.
Lautrup, B., 1977, Phys. Lett. **69B**, 109.
Lipatov, L.N., 1977, Sov. Phys. JETP **45**, 216.
Loeffel, J., A. Martin, B. Simon and A. Wightman, 1969, Phys. Lett. **30B**, 656.
Magnen, J., and V. Rivasseau, 1986, The Lipator Argument for ϕ_3^4 Perturbation Theory, to appear.
Magnen, J., and R. Sénéor, 1977, Commun. Math. Phys. **56**, 237.
Nagamachi, S., and N. Mugibayashi, 1976a Commun. Math. Phys. **46**, 119.
Nagamachi, S., and N. Mugibayashi, 1976b, Commun. Math. Phys. **49**, 257.
Nagamachi, S., and N. Mugibayashi, 1986, J. Math. Phys. **27**, 832.
Newman, C., 1981, Commun. Math. Phys. **79**, 133.

Parisi, G., 1975, Nucl. Phys. **B10**, 368.
Parisi, G., 1977, Cargese Lectures, Hadron Structure and Lepton–Hadron Interactions, eds M. Levy, J.-L. Basdevant, D. Speiser, J. Weyers, R. Gastmans and J. Zinn-Justin (Plenum, New York) pp. 665–686.
Parisi, G., 1979a, Phys. Rev. **49**, 215.
Parisi, G., 1979b, Nucl. Phys. **B150**, 163.
Pauli, W., 1933, Die allgemeinen Prinzipien der Wellenmechanik, Handbuch der Physik, Zweite Aufl. 24-1 (Springer, Berlin).
Petermann, A., 1953, Helv. Phys. Acta **26**, 291.
Poincaré, H., 1882, Oeuvres, **I**, XXXVI.
Rivasseau, V., 1984, Commun. Math. Phys. **95**, 445.
Rosenfeld, L., and J. Solomon, 1931, J. de Physique (7) **2**, 129.
Ruijsenaars, S., 1983, J. Math. Phys. **24**, 922.
Sakato, S., H. Umezawa and S. Kamefuchi 1952, Prog. Theor. Phys. **7**, 377, 551.
Salam, A., 1951a, Phys. Rev. **82**, 217.
Salam, A., 1951b, Phys. Rev. **84**, 426.
Schroer, B., 1964, J. Math. Phys. **5**, 1361.
Schroer, B., and T. Truong, 1977, Phys. Rev. **D15**, 1684.
Schwartz, L., 1952, Théorie des Distributions I and II (Hermann, Paris).
Schwinger, J., 1951, Phys. Rev. **82**, 664.
Schwinger, J. (ed.), 1958, Selected Papers on Quantum Electrodynamics (Dover, New York).
Siegel, C.L., 1952, Nach. Akad. Wiss. Göttingen Math. Phys. Klasse, p. 21.
Simon, B., 1970, Ann. Phys. (USA) **58**, 76.
Skyrme, T.H.R., 1961, Proc. Roy. Soc. London **A262**, 237.
Sokal, A., 1980, J. Math. Phys. **21**, 1583.
Sokal, A., 1982, Ann. Inst. H. Poincaré **A37**, 317.
Speer, E., 1968, J. Math. Phys. **9**, 1404.
Steinberger, J., 1949, Phys. Rev. **76**, 1180.
Streater, R., and I.F. Wilde, 1970, Nucl. Phys. **B24**, 561.
Symanzik, K., 1973a, Lett. Nuovo Cimento **6**, 77.
Symanzik, K., 1973b, Lett. Nuovo Cimento **8**, 771.
Symanzik, K., 1975, Commun. Math. Phys. **45**, 79.
't Hooft, G., 1979, Can we make sense out of quantum chromodynamics, in: The Whys of Subnuclear Physics, Lectures at "Ettore Majorana", Int. School of Subnuclear Physics Erice 1977, ed. A. Zichichi (Plenum, New York).
't Hooft, G., 1982, Phys. Lett. **119B**, 369.
Thirring, W., 1953, Helv. Phys. Acta **26**, 33.
Tomonaga, S., 1948, Phys. Rev. **74**, 224.
Wick, G.C., 1950, Phys. Rev. **80**, 268.
Wightman, A., 1977, in: Cargese Lectures 1976, eds M. Levy and P. Mitter (Plenum, New York) pp. 67–77.
Wightman, A., 1979, in: The Whys of Subnuclear Physics, Lectures at "Ettore Majorana", Int. School of Subnuclear Physics, Erice 1977, ed. A. Zichichi (Plenum, New York). pp. 983–1026.
Wightman, A., 1981a, Physica Scripta **24**, 813.
Wightman, A., 1981b, Adv. Math. Suppl. Studies **7B**, 769.
Zimmermann, W., 1969. Commun. Math. Phys. **15**, 208.
Zinn-Justin, J., 1982, The Principles of Instanton Calculus: A Few Applications, Les Houches 1982, Session XXXIX, eds J.B. Zuber and R. Stora (North-Holland, Amsterdam, 1984) p. 39.

Discussion, session chairman A. Salam

Salam: This extension of field theory to non-renormalizable cases is surprising. I remind you that the success of the standard model was due to the fact that it *was* renormalizable.

Johnson: What has worried people about non-renormalizable theories is not that they don't exist, but rather that there are infinitely many of them for a given Lagrangean. So my question is: what about the uniqueness of these theories?

Wightman: Even with conventional methods it is difficult to specify a theory uniquely. Suppose you solve a theory defined by a functional integral with a cut-off and find that the appropriate axioms are satisfied. How can you answer the question, "What problem did you solve?" The answer would be "I solved the problem of showing that certain limits existed and that they had certain properties". But you never write down any condition which fixed the theory you were talking about.

We argue, of course, that conventional renormalized theories are fixed by choosing coupling constants and masses. In a non-renormalizable model these parameters are usually infinite in number. But in the procedure I have discussed you only have to fix the same number of constants that you would fix in a lower number of dimensions. In this way you decide which terms in the Lagrangean are fixed by the renormalization procedure and which are fixed by the normalization.

Lehmann: You mentioned the work of Gawedzki and Kupiainen. I think that they have made real progress, by showing that the $2+\epsilon$ dimensional Gross–Neveu model is well defined with Green functions bounded polynomially by non-integer powers. This is the first time that any non-renormalizable theory has been treated rigorously. In addition this theory is non-asymptotically free, so we have a direct proof that such theories can exist.

Wightman: Yes, this is very interesting, but this theory is not unitary since the dimensionality is non-integer. The $d=3$ case would be much more interesting.

Lee: There seems to be two types of theories, those which confine and those which do not. From the axiomatic point of view, how does one make the distinction?

Wightman: There are some general theorems proved by the group in Zürich in the early 1960s which say you cannot get an increasing potential out of a relativistic positive metric theory. Gauge theories have positive metric only for non-covariant gauges, so you can easily see the possibility of an increasing potential.

It has now been shown that the Higgs model has a continuum limit in three space–time dimensions. This theory [with a compact Lie group broken to U(1)] is supposed to confine. Perhaps confinement can be proved in this concrete continuum model.

Nielsen: Is it possible in an abstract manner, that is, without constructing examples, to tell whether there exist many non-renormalizable theories?

Wightman: I don't know, but there is the grand program of Symanzik to classify all non-trivial theories (which is by no means completed).