ULTRAVIOLET STRESS COMPENSATION IN THE STRAINED VACUUM — AN EFFECTIVE RECIPE FOR CONVERGENCE

By G. R. Allcock

Department of Applied Mathematics and Theoretical Physics, University of Liverpool*

(Received April 15, 1980; revised version received July 1, 1980)

The physical role of zero-point oscillations and zero-point energies in relativistic local quantum field theory is discussed. A mathematically rigorous distribution-theoretic approach to free fields and their divergent zero-point stress integrals is formulated. A chronic malady of Lorentz non-invariance on the part of the vacuum stress is diagnosed and cured, leading to a privileged family of free-field systems for which the vacuum stresses are already finite and relativistic, without need for subtractions. Building upon this, a heuristic principle is introduced for interacting fields. The principle asserts that the zero-point energy should change only finitely under external perturbations. Its implementation leads to an ultraviolet convergence condition, which expresses many relations between masses and coupling constants through a single formula. The propagators and vacuum bubbles of two-dimensional theories obeying the formula are entirely finite, wholly unambiguous, and unitary; in dimensions higher than two the divergences are softened. Global supersymmetry appears as a special mass-degenerate case, and further convergence conditions appropriate to three and four dimensions can be foreseen.

PACS numbers: 11.10,-z; 11.10.Gh

Introduction

In this paper we show that the ultraviolet divergences of relativistic quantum field theories in two space-time dimensions can be eliminated, and those in higher dimensions softened, by arranging a balance between the Bose and Fermi zero-point contributions to the stress tensor.

The main new concepts are presented in Sections 1 and 4. Section 1 is concerned with free fields. In this section we show that the zero-point stresses of non-interacting relativistic quantum field systems can be rendered unambiguously finite, by having equal numbers of Bose and Fermi modes, with suitable relations between the masses. In Section 4 we extend this principle of Fermi–Bose balance to interacting systems, obtaining thereby

* Address: Department of Applied Mathematics and Theoretical Physics, Chadwick Laboratory, University of Liverpool, P.O.Box 147, Liverpool L69 3BX, United Kingdom.

certain further conditions between the bare coupling constants and the bare masses. The graphical implications of these conditions are examined in Section 5, with particular reference to fields in two space-time dimensions. It is proved that the divergences of various graphs of differing topologies mutually cancel, leaving behind a finite and completely unambiguous mathematical system.

The physical reasoning lying behind our new approach is expounded in Section 2. We discuss there some relevant aspects of the cancellation mechanisms of quantum electrodynamics, looking at this topic from an unconventional point of view, in which the inner structure of the vacuum is no longer ignored. A proper study of vacuum structure considerably increases one's understanding of the ultraviolet problem. Arguing from commonplace experimental facts, we thus adduce some compulsive arguments, suggesting that cancellation must be Nature's unique way to get a consistent system of relativistic quantal laws.

The full extension of the Fermi–Bose cancellation mechanism to four space-time dimensions will require further basic principles, as yet unconceived. Some of the problems presently visible are mentioned in Section 6.

Various technicalities, including the two-dimensional Dirac equation and some necessary power-counting criteria, are reviewed in the third section.

1. The unperturbed vacuum and its zero-point energy

Practitioners of quantum field theory have always recognized that the zero-point fluctuations of the fields are an inherent part of the quantum formalism, and that they are physically necessary. For instance, excited atoms would stay for ever in their excited states, and would not emit radiation, were it not for the benign influence of the electromagnetic zero-point field. Nevertheless, the zero-point energy density associated with the sum over an infinite number of fluctuating field modes is infinite, and this has been an ever-present source of embarrassment. In the free-field case it has often been advocated that one should cut out the zero-point energy by some mathematical artifice, such as Wick's normal-ordering prescription [1]. Alternatively, people sometimes argued that one might tactfully ignore an infinite constant in the energy density, on the grounds that a constant term is physically irrelevant [2]. This second argument, however, took no account of the universal belief that the stress tensor is in some way or other the direct source of an observable physical effect, namely gravitation, and as such must be properly defined. There is also the experimentally verified (cf. Ref. [3]) Casimir force between macroscopic electrically neutral bodies. This force can be obtained as the summed effect of all the retarded interatomic Van der Waals forces [3], but a study of the field energy in the space between the bodies gives the results more quickly. The field energy treatments [4–7] clearly show that the zero-point fluctuation energies of the electromagnetic field are as genuine and physically necessary an effect as the field fluctuations themselves, and that they can neither be tactfully ignored nor be removed wholesale by undiscriminating mathematical legislation. Granted therefore that zero-point energies are physically meaningful per se, one must perforce seek some mathematical procedure appropriate to the physics of the situation, to bring them under control. The procedure sought must be manifestly covariant and local, so that it can be extended to interacting systems. Both this and the Casimir effect rule out Wick normal ordering as the ultimate solution. However, the study of quantum field theories in curved space-time has led in recent years to the development of other subtraction or regularization procedures better adapted to the interacting situation, so that in some sense at least the problem just outlined can be said to be on the way towards resolution [8–12]. But there is another sense in which the problem remains with us. For it is characteristic of all the said subtraction or regularization prescriptions that the leading infinities in the stress (which appear already in flat space-time) are not argued away in any very convincing manner, but are simply dismissed as being mathematically meaningless and without physical import. This formalized way of disposing of infinities is quite different from that to be proposed in the present work. Here, paying full regard to well-known quantum principles, we shall enquire into the physical origins of the infinite terms. We shall also examine their mathematical structure in a relevant way, by making use of limiting processes compatible with quantum mechanics. Through this we shall discover how to make the leading infinities cancel out. This is not at all the same as merely throwing them away. It will therefore be logically clear to the reader that the work of the present paper goes outside the confines of formal subtraction theory or formal regularization theory. This being so, it can neither be bound by nor supplanted by the formal postulates and methods of these theories.

One of the methodological features of the said formal theories is that they approach the infinities through some Lorentz covariant limit process, for which several widely different candidates are currently available. Relativity is thus manifestly maintained at all stages, but, because the mathematical situation is extremely delicate, some quantum principles tend to get obscured. In the present discussion we propose therefore to work the other way round, and to invoke an old-fashioned non-covariant limit, which in the free-field case can be formulated just by cutting out all the contributions of all plane waves whose wave vectors $\boldsymbol{p}$ exceed in magnitude some finite momentum K. Thus, given a relativistic Lagrange function $\mathscr{L}$ dependent upon a field $\phi(t, \boldsymbol{x})$, our free-field prescription is simply to insert into $\mathscr{L}$ in place of $\phi(t, \boldsymbol{x})$ the formula

$$\phi(t, \boldsymbol{x}, K) \equiv \int \theta(K - |\boldsymbol{p}|)\tilde{\phi}(t, \boldsymbol{p})e^{i\boldsymbol{p}\cdot\boldsymbol{x}}d^3\boldsymbol{p} \tag{1}$$

before varying the action. The advantage of this archaic procedure is that all the principles of canonical Lagrangian quantum mechanics survive intact when K is finite. At the same time, old-fashioned folk-lore would have us believe that if we calculate an observable quantity, then this quantity should find itself endowed with relativistic properties in the relativistic limit $K \to \infty$. Our attitude in the present paper will be that this ought to be a permissible way to set up the relativistic quantum mechanics of fields, and that if in some respect relativity fails to emerge in the limit $K \to \infty$, then something is wrong with the original Lagrangian itself, and must be put right.

Surprisingly, the literal implementation of this simple-minded philosophy leads to new restrictions on the set of admissible relativistic quantum field theories, even at the free-field level. The new restrictions produce improved ultraviolet behaviour. Thus, by demanding the full validity of the old-fashioned approach to the relativistic limit, we shall learn something new and of basic importance about the structure and potentialities of relativistic quantum field theory. This new information could not easily be obtained by working actually "at the limit", bacause "at the limit" some mathematical concepts relevant to our argument cease to exist. In particular, the simple square $\phi^2(t, x)$ of the operator $\phi(t, x)$ of Eq. (1) exists only while $K < \infty$, and even the Wick-ordered square $:\phi^2(t, x):$ becomes rather nasty "at the limit".

All local bilinears exhibit this bad mathematical misbehaviour. Its cause may be traced through a well-known result of axiomatic field theory, according to which any relativistic field obeying the positivity axiom of quantum mechanics must necessarily be distribution--valued in its dependence on the space-time coordinates. It is this which makes it impossible to deal with $\phi^2(t, x)$ "at the limit", and with the other bilinears contributing to the stress.

We cannot therefore start our mathematical discussion without taking cognizance of certain concepts from the theory of distributions. There are however two quite distinct approaches to distribution theory, and so we must first decide which one is the more appropriate to the problem on hand. Let us begin by considering the method of Mikusiński, Temple and Lighthill, in which every distribution is represented as a limiting sequence of ordinary functions [13, 14]. Applied in the quantum field context this concept will evidently require the introduction of a limiting sequence of function-valued operator fields (necessarily Lorentz non-invariant in view of the axiomatic result just mentioned). The sequence $K \to \infty$ in Eq. (1) provides a suitable example of such a limiting sequence of quantum fields, and we shall use it. This particular sequence commends itself to us for three reasons: (*i*) each member of the sequence is a bona-fide translation-invariant Lagrangian quantum system obeying the positivity axiom and possessing a conserved stress tensor, (*ii*) local products such as $\phi^2(t, x, K)$ are bona-fide Hilbert-space operators within each member of the sequence, (*iii*) the sequence is formulated in a simple and tractable way.

Both of the theories of distributions make extensive use of the concept of test function, and so far as these are concerned we shall tacitly take for granted all the usual properties, as laid out by Temple [13] and Lighthill [14]. However, since we have four space-time coordinates x^μ to play with, the question at once arises as to which of these four are to feature as active arguments within the test functions. It is not appropriate to prejudge this issue by suggesting that all four should appear. The truth of the matter is that the Green functions of hyperbolic and parabolic partial differential equations (and hence the associated quantum fields) show simultaneously several distinct and different distribution-theoretic properties, depending upon how many and which of the space-time coordinates are involved. Thus in the case of the Schrödinger Green function $G(t, x)$ in one space dimension no fewer than three distinct distribution properties arise, referring to test functions $S(x)$, $S(t)$, and $S(t, x)$. All these conjoint properties can be effectively explored with the help of one limiting sequence [15], and the same can be demonstrated in the relativistic context. But the great beauty of the limiting sequence approach is not so much that it provides a ready resolution of such matters as that it permits the consideration of the same to be postponed until a definite need for it arises, whereas the other approach requires some prejudgement for its very formulation.

Let us now enquire briefly into the other approach in which, going back to Schwartz [16, 17], one regards distributions as continuous linear functionals on some postulated space of test functions (typically $S(t, x, y, z)$ in our case). The Schwartz approach appeals at once to the tidy mathematical mind, for it is formulated ab initio "at the limit". In the present problem this means that it accommodates both relativity and positivity from the very beginning. It is within this manifestly relativistic distribution context that the currently popular subtractions and regularizations may be most appropriately understood. But clearly a Schwartz formulation cannot possibly help us to understand the divergent aspects of the stress tensor, since it enforces an unhealthy agnosticism at the very point where one requires information, ϕ^2 and other local bilinears being undefinable "at the limit".

The mathematical situation is therefore that there are two rigorous approaches to distribution theory, both equally valid in the contexts for which distributions were originally devised. But of these two only one is sufficiently flexible to adapt to the problem which we now wish to treat, while the other is powerless to say anything at all about it.

Working therefore in the light of the method of limiting sequences, let us reexamine the stress tensor in the spirit which Schwinger first expounded thirty-one years ago [18]. At that time Schwinger argued formally from Lorentz invariance that the divergent vacuum expectation value $\langle\Theta_{\mu\nu}\rangle_0$ of the fully interacting stress tensor $\Theta_{\mu\nu}$ in a Minkowski-space theory must be a constant scalar multiple of the isotropic tensor $g_{\mu\nu}$. He therefore proposed an invariant subtraction of the form $\Theta_{\mu\nu} \to \Theta_{\mu\nu} - \Lambda g_{\mu\nu}$, with the cosmological multiplier $\Lambda = \langle\Theta_{00}\rangle_0$. Such a subtraction would reduce $\langle\Theta_{\mu\nu}\rangle_0$ to zero, in accord both with common sense and with the requirements of gravitation theory. And it is indeed obvious that this invariant subtraction would be uniquely correct for any genuinely relativistic theory of $\Theta_{\mu\nu}$. But when we actually look in more detail at a typical field theory, even a theory of a free field, we find that the divergences in its vacuum stresses are so strong that the stated relativistic stress properties fail to emerge in the K-limit, other relativistic properties notwithstanding. In other words, we find that no subtraction of a K-dependent multiple of $g_{\mu\nu}$ can keep the stress finite as the limit $K \to \infty$ is approached. We have therefore a contradiction between the formal and the actual properties of the theory.

To see how this arises, let us consider as an example the quantized free Maxwell field, with the contributions of the short waves removed by the cut-off formula (1). The cut-off fields $\boldsymbol{E}(x, K)$ and $\boldsymbol{B}(x, K)$ obey Maxwell's equations exactly, since we deal with a linear system. Also they are finite, through the cut-off, and simple local products such as $\boldsymbol{E}^2$ and $\boldsymbol{B}^2$ are meaningful. The Maxwell stresses, computed now by using symmetrized simple local products, obey exactly the conservation laws $\partial^\nu\Theta_{\mu\nu}(x, K) = 0$, and the space integrals of $\Theta_{\mu 0}(x, K)$ generate space-time translations correctly. So far as quantum mechanics goes, we have a perfectly good theory. The diagonal elements of the cut-off stress tensor are $\frac{1}{2}(\boldsymbol{E}^2+\boldsymbol{B}^2)$, $\frac{1}{2}(\boldsymbol{E}^2+\boldsymbol{B}^2)-E_x^2-B_x^2$, ..., where $\boldsymbol{E}$ stands for $\boldsymbol{E}(x, K)$ etc. Obviously there is no factor ordering problem for these diagonal elements and, by inspection, the sum of all four vanishes unambiguously and exactly:

$$\Theta^\mu{}_\mu(x, K) = 0. \tag{2}$$

At the same time we see that $\Theta_{00}(x, K)$ is a sum of perfect squares of finite non-zero Hermitian operators ($\frac{1}{2}E^2+\frac{1}{2}B^2$), and as such is positive definite. Consequently, for all finite K, and still more so as K approaches infinity, we have

$$\langle \Theta_{00}(x, K)\rangle_0 \neq 0. \tag{3}$$

Bringing this and the trace condition (2) and spatial isotropy into play, we are forced to conclude that for all finite K (and hence also in the limit $K \to \infty$) we have

$$\frac{\langle \Theta_{\mu\nu}(x, K)\rangle_0}{\langle \Theta_{00}(x, K)\rangle_0} = \mathrm{diag}\,[1, \tfrac{1}{3}, \tfrac{1}{3}, \tfrac{1}{3}] \neq \mathrm{diag}\,[1, -1, -1, -1]. \tag{4}$$

The same contradictory property may of course be verified by a more detailed calculation, using creation and annihilation operators. The factors $\frac{1}{3}$ remind us of the kinetic theory of gases, and the stresses are indeed precisely those of an isotropic gas of zero--point photons. This feature becomes explicitly apparent in the more detailed calculation. There is no doubt that (4) is both mathematically correct and intuitively acceptable within the cut-off theory. Schwinger's argument is thus rendered inapplicable and to keep the stress finite we have to make a formally non-invariant subtraction. This seems most unsatisfactory. Also, the picture of the gas seems so obviously plausible and right. But how can it be right and yet in manifest contradiction with relativity? We shall resolve this paradox within the present section, and we shall extend the same considerations to interacting theories in the third section.

A similar situation arises with respect to the simplest of all fermion field systems, the field theory of the massless Majorana neutrino. Here there is a factor-ordering problem in the Belinfante stress, which we resolve in the standard way [18], by using an antisymmetrical average over the two possible orderings (we shall see presently that the vacuum stress itself is not affected by this adjustment). The antisymmetrized expression for the stress is

$$\Theta_{\mu\nu} = \tfrac{1}{8}\, i\hat{\psi}(\gamma_\mu\vec{\partial}_\nu+\gamma_\nu\vec{\partial}_\mu)\psi-\tfrac{1}{8}\, i\hat{\psi}(\gamma_\mu\overleftarrow{\partial}_\nu+\gamma_\nu\overleftarrow{\partial}_\mu)\psi. \tag{5}$$

[We use here real Dirac matrices, such that $\gamma_0\gamma_0 = -g_{00} = -1$, $\gamma_1\gamma_1 = -g_{11} = +1$, etc., with $\gamma_0{}^{\mathrm{T}} = -\gamma_0$ and $\gamma_j{}^{\mathrm{T}} = +\gamma_j\,(j = 1, 2, 3)$. All four matrices $\gamma_0\gamma_\mu$ are symmetric and the Majorana adjoint field is $\hat{\psi} \equiv -\psi\gamma_0$. The Majorana–Dirac equation is $\gamma_\mu\vec{\partial}^\mu\psi = 0$, which can also be written in the adjoint form $\hat{\psi}\gamma_\mu\overleftarrow{\partial}^\mu = 0$. ψ is Hermitian, since it is a Majorana field. Its scale is set by the equal-time anticommutator $\{\psi_r, \psi'_s\} = \delta_{rs}\delta^3\,(\boldsymbol{x}-\boldsymbol{x}')$].

As with the Maxwell field, we now cut out the short wave Fourier components. This does not upset the Majorana–Dirac equation, and once again Eq. (2) holds good with all rigour. Also, using the implied normal-mode decomposition of ψ, and working within the vacuum, we find that we can make substitutions of the type

$$i\vec{\partial}_0 \to -\omega(\boldsymbol{p}), \qquad -i\overleftarrow{\partial}_0 \to -\omega(\boldsymbol{p}), \tag{6}$$

where $\omega(\boldsymbol{p})$ represents the energy quantum of a mode, and is positive. The energy density thus acquires the signature of $-\hat{\psi}\gamma_0\psi \equiv -\psi\psi$, the negative of a sum of squares of non--zero Hermitian operators. Thus for the Majorana field we find that $\langle \Theta_{00}(x, K)\rangle_0$ is *negative* definite. Therefore, as with the Maxwell field, we are led once again through Eqs. (2) and (3) to the contradictory property (4). Factor ordering does not affect this issue at all. There are only two possible locally defined factor orderings. They are both displayed in the stress formula (5), and both exhibit tracelessness with vacuum energy negativity.

From these two simple examples we see that the stresses of relativistic field systems of the sorts usually considered are not properly relativistic under the provisions of our proposed K-limit. Also we see that Schwinger's vacuum subtraction formula cannot be consistently applied to them to cancel the divergent stress within the context of the K-limit, since the stress at finite K is in no way a multiple of $g_{\mu\nu}$.

A newcomer might argue at this juncture that we have demonstrated merely that the K-limit is inappropriate to the niceties of relativistic quantum theory. But the reader will recall that our attitude is that the K-limit ought to work, and that if it does not then the fault should be sought elsewhere.

How then shall we put matters right? Obviously no reasonable adjustment of the limit process is ever going to change the factors $+\frac{1}{3}$ of a zero-point Bose or Fermi gas into the factors -1 of the metric tensor (Eq. (4)). But what we can do is to exploit the two signs of $\langle\Theta_{00}\rangle_0$ which we have encountered. These signs are indeed of fundamental importance. It is very easy to see where they come from. For the normal mode analysis of a free and symmetrized Bose field theory leads necessarily to a collection of symmetrized oscillator Hamiltonians, each of the well-known form

$$\mathscr{H}_\mathrm{B} = \tfrac{1}{2}\,\omega(b^\dagger b + bb^\dagger) = \omega(b^\dagger b + \tfrac{1}{2}), \tag{7}$$

while a free and antisymmetrized Fermi field theory leads necessarily to a collection of antisymmetrized oscillator Hamiltonians showing the canonical structure [19]

$$\mathscr{H}_\mathrm{F} = \tfrac{1}{2}\,\omega(b^\dagger b - bb^\dagger) = \omega(b^\dagger b - \tfrac{1}{2}). \tag{8}$$

From this it is fully evident that the oscillator zero-point energies for the two symmetry types are respectively $+\frac{1}{2}\omega(\boldsymbol{p})$ and $-\frac{1}{2}\omega(\boldsymbol{p})$ per mode $\boldsymbol{p}$, in agreement with the signs found above. It follows that we can account quantitatively for many free-field vacuum properties by supposing every bosonic mode of excitation to contain $\frac{1}{2}$ of a quantum, and every fermionic mode of excitation to contain $-\frac{1}{2}$ of a quantum. This simple accounting rule fully embodies the universal symmetrization/antisymmetrization prescription, and it applies to the vacuum expectation value of any expression bilinear in the fields, such as, e.g, a current or a component of stress. The zero-point energy density in a free field system with cut-off will therefore be represented in all cases by the potentially divergent integral

$$\langle\Theta_{00}(x,K)\rangle_0 = \int \frac{d^3p}{(2\pi)^3}\,\theta(K-|\boldsymbol{p}|)\left\{\frac{1}{2}\sum_\alpha (\boldsymbol{p}^2+m_\alpha{}^2)^{1/2} - \frac{1}{2}\sum_a (\boldsymbol{p}^2+m_a{}^2)^{1/2}\right\}, \tag{9}$$

wherein $\sum_\alpha$ covers all bosonic excitation modes having the 3-momentum $\boldsymbol{p}$, m_α is the mass associated with the αth boson mode, and $\sum_a$ and m_a relate similarly to the fermionic modes. A Maxwell field will contribute two terms to $\sum_\alpha$ and a massive neutral vector field three terms, while a Majorana field will contribute two terms to $\sum_a$, a complex Dirac field four terms, and so forth.

The contradictions and inconsistencies that we have noted in the stress arise because of the bad behaviour of the integrand of (9) for large $\boldsymbol{p}$. We deal with

$$|\boldsymbol{p}|\sum_\alpha (1+\tfrac{1}{2}\, m_\alpha{}^2/\boldsymbol{p}^2-\tfrac{1}{8}\, m_\alpha{}^4/\boldsymbol{p}^4+\ \ldots)-|\boldsymbol{p}|\sum_a (\ldots). \tag{10}$$

From this we see that the integral (9) diverges in the limit $K \to \infty$ unless all of the following three convergence conditions hold:

$$\sum_\alpha 1-\sum_a 1 = 0, \tag{11}$$

$$\sum_\alpha m_\alpha{}^2-\sum_a m_a{}^2 = 0, \tag{12}$$

$$\sum_\alpha m_\alpha{}^4-\sum_a m_a{}^4 = 0. \tag{13}$$

In space-times with only two or three dimensions the last condition (13) can be dropped, but the other two must be maintained if convergence is required.

Under the three conditions (11)–(13) the integral (9) stays finite in the limit. The way in which the limit appears in (9) makes clear that in effect we are proposing to cancel infinities against each other at fixed $\boldsymbol{p}$. This does not look very propitious for Lorentz invariance. But we shall prove nevertheless that the cancellation is fully relativistic in the limit, the explanation for this being that fixed $\boldsymbol{p}$ tends to fixed p^μ as $\boldsymbol{p}$ tends to infinity along the mass hyperboloids.

Before embarking on a study of this rather tricky point it will be helpful to consider a general free field theory, and to work out $\langle\Theta_{\mu\nu}\rangle_0$ for it using from the beginning the standard relativistic methods of calculation. In view of what we have found we must expect that a relativistic calculation honestly performed must somehow produce a relativity--violating answer, and indeed this turns out quite unambiguously to be the case. Relativity--breaking terms arise because of the easily-verified operator properties

$$\begin{aligned} T\{\partial_\mu\phi(x, K)\partial'_\nu\phi(x', K)\} &= \partial_\mu\partial'_\nu T\{\phi(x, K)\phi(x', K)\} \\ &\quad -i\delta_{\mu 0}\delta_{\nu 0}\delta(t-t')\delta^3(\boldsymbol{x}-\boldsymbol{x}', K), \\ T\{\psi_r(x, K)\partial'_\nu\psi_s(x', K)\} &= \partial'_\nu T\{\psi_r(x, K)\psi_s(x', K)\} \\ &\quad +\delta_{\nu 0}\delta_{rs}\delta(t-t')\delta^3(\boldsymbol{x}-\boldsymbol{x}', K), \end{aligned} \tag{14a}$$

where

$$\delta^3(\boldsymbol{x}-\boldsymbol{x}', K) \equiv (2\pi)^{-3}\int d^3\boldsymbol{p}\,\theta(K-|\boldsymbol{p}|)e^{i\boldsymbol{p}\cdot(\boldsymbol{x}-\boldsymbol{x}')}. \tag{14b}$$

When we use these formulae, and allow x' to approach x from a purely spatial direction, so that the T products on the left hand sides rigorously become the symmetrized/anti-symmetrized products required for the stress, we find that

$$\langle \Theta_{\mu\nu}(x, K)\rangle_0 = i \int \frac{d^4p}{(2\pi)^4}\, \theta(K-|\boldsymbol{p}|) \left\{ \sum_{\alpha} \left(\frac{p_\mu p_\nu}{p^2 - m_\alpha{}^2 + i\varepsilon} - \delta_{\mu 0}\delta_{\nu 0} \right) - \sum_{a} (\ldots) \right\}, \quad (15)$$

which contains imaginary non-tensorial contributions! (In the case of zero spin we may verify (15) using either the canonical Klein–Gordon stress or the improved stress of Callan, Coleman and Jackiw [13]. The two differ only by a gradient, and are therefore equivalent in the vacuum.)

The divergent and non-covariant integral (15) has several remarkable properties, which we now enumerate and discuss:

(*i*) The presence of non-tensorial terms shows again that ordinary relativistic quantum field theory fails to give a relativistic vacuum stress in the limit $K \to \infty$!

(*ii*) In the case of the vacuum energy density $\langle \Theta_{00}\rangle_0$ the imaginary non-tensorial terms act to reduce the power of p_0 at infinity in the p_0-plane, thereby permitting the p_0 contour to be closed by a semicircle at infinity for each α-contribution or a-contribution separately. By proceeding in this way and shrinking the contour we obtain complete agreement with the energy formula (9). This shows that the relativistic calculation methods are strictly in accord with quantum principles, in that they reproduce exactly the non-covariant but unequivocally correct result of a normal-mode zero-point sum.

(*iii*) In the case of the vacuum momentum density $\langle \Theta_{j0}\rangle_0$ we encounter p_0-integrals of the form

$$\int_{-\infty}^{\infty} \frac{p_0 dp_0}{p_0^2 - \boldsymbol{p}^2 - m^2 + i\varepsilon} = 0. \quad (16)$$

These integrals are only conditionally convergent. However, a detailed analysis of the T-product used in deriving (15) indicates unambiguously that (16) should be approached in the spirit of a symmetrical p_0 cut-off, and since the integrand is odd it follows that a correct evaluation gives zero, as indicated above. The momentum density is therefore zero, in agreement with any normal mode summation carried out under the symmetry $\boldsymbol{p} \to -\boldsymbol{p}$.

(The thoughtful reader may be puzzled that the zero property of the vacuum momentum density arises already in the p_0-integration, and not in the $\boldsymbol{p}$-integration. The reason for this may be found by looking at the contribution which a single normal mode of momentum quantum $\boldsymbol{p}$ and energy quantum $\omega = (\boldsymbol{p}^2 + m^2)^{\frac{1}{2}}$ makes to the Feynman propagator $\Delta(x)$. Its contribution is proportional to $\theta(t)\exp(i\boldsymbol{p}\cdot\boldsymbol{x} - i\omega t) + \theta(-t)\exp(-i\boldsymbol{p}\cdot\boldsymbol{x} + i\omega t)$. From this it follows that the single point $\boldsymbol{p}$ in the Fourier integral representation of $\Delta(x)$ carries contributions both from the mode $\boldsymbol{p}$ and from the mode $-\boldsymbol{p}$. These two modal contributions are represented with equal weight at the point $\boldsymbol{p}$ in the stress integral (15), because this integral derives from an equal-time configuration. Therefore they have to cancel exactly in the momentum density. Equation (16) expresses precisely that this cancellation has occurred.)

(*iv*) The p_0-contour in (15) can obviously be closed separately in each α-contribution or a-contribution to the vacuum momentum flow tensor $\langle\Theta_{jk}\rangle_0$. We obtain at once

$$\langle\Theta_{jk}(x, K)\rangle_0 = \int \frac{d^3p}{(2\pi)^3}\,\theta(K-|\boldsymbol{p}|)\left\{\tfrac{1}{2}\sum_\alpha \frac{p_j p_k}{(\boldsymbol{p}^2+m_\alpha{}^2)^{1/2}} - \tfrac{1}{2}\sum_a (\ldots)\right\}, \tag{17}$$

in complete agreement with the normal mode prediction for the k-directed flux of j-directed momentum.

(*v*) Each bosonic or fermionic contribution in (15) separately and unambiguously shows a vanishing trace if the corresponding mass m is zero, and at the same time exhibits a real non-vanishing energy density, in full accord with Eq. (3).

(*vi*) Suppose now that the leading convergence criterion (11) is obeyed (equal numbers of Bose and Fermi modes). In this case the non-tensorial terms in (15) cancel between bosons and fermions, and play no part in the overall stress. The stress as a whole becomes formally relativistic, and it is only the physically unobservable division into separate boson and fermion contributions that brings trouble. At the same time the convergence properties of (15) improve sufficiently to allow the p_0 contour in the overall expression to be rotated about the origin through an anticlockwise angle of $\frac{1}{2}\pi$ for all values of μ and ν. Thus, setting $p_0 = ip_4$, we get

$$\langle\Theta_{00}(x, K), \Theta_{j0}(x, K), \Theta_{jk}(x, K)\rangle_0$$

$$= \int_{-\infty}^{\infty} \frac{dp_4}{2\pi}\int \frac{\theta(K-|\boldsymbol{p}|)d^3p}{(2\pi)^3}\left\{\sum_\alpha \frac{-p_4^2,\ ip_jp_4,\ p_jp_k}{p_4^2+\boldsymbol{p}^2+m_\alpha{}^2} - \sum_a (\ldots)\right\} \quad \text{(subject to (11))}, \tag{18}$$

whence Schwinger's relativistic relation $\langle\Theta_{\mu\nu}\rangle_0 = \Lambda g_{\mu\nu}$ now emerges *formally* "at the limit" $K = \infty$. The constant Λ is still divergent at this stage, since (18) is still divergent in the limit. However, if all the masses vanish then (18) is strictly zero, because the Bose and Fermi terms in the integrand annul each other identically, before integration. Then Λ is zero too, and the trace condition (2) is obeyed. Thus the single convergence condition (11) has already sufficed to restore relativistic invariance and to remove the previously noted contradictions between the trace, the definiteness of the energy, and relativity.

(*vii*) Suppose now that all three convergence criteria (11)–(13) are obeyed. In this case the integral (18) manifestly converges to a finite multiple of $g_{\mu\nu}$ in the limit $K \to \infty$, and the same property holds therefore for the modal integrals (9) and (17) and for the relativistic formula (15), since these and (18) are equivalent through Cauchy's theorems.

(*viii*) Our theory of the free-field vacuum stress is thus finite, unambiguous and relativistic, and it has the remarkable property that the boson–fermion cancellation on which it depends gives quite the same answers whether one formulates it at fixed 4-momentum, as in (15), or at fixed 3-momentum, as in (9) and (17). Moreover, everything is in strict accord both with the principles of quantum mechanics and with the principles of local field theory. The cancellation at fixed 4-momentum expresses in Lagrangian language the concept of a *local* cancellation between boson and fermion bilinears; by contour closure the very same cancellation concept reappears in Hamiltonian language as a cancellation at fixed 3-momentum. One could think of other relativistic recipes for cancellation. For example, one could cancel at fixed 4-velocity. But a cancellation at fixed 4-velocity would not be local, and it would not carry over to interacting systems. Nor could it be carried over to massless fields.

To evaluate the covariant vacuum stress integral (15) "at the limit" under the convergence conditions (11)–(13) it suffices now to look at the trace, since we have proved by (18) that Schwinger's property is obeyed in the limit. The calculation can be facilitated by introducing a fixed mass M, the same for each term, and making thereby a series of subtractions according to the scheme

$$\frac{p^2}{p^2-m_\alpha{}^2+i\varepsilon}=1+\frac{m_\alpha{}^2}{p^2-m_\alpha{}^2+i\varepsilon}\rightarrow\frac{m_\alpha{}^2}{p^2-m_\alpha{}^2+i\varepsilon}\rightarrow\frac{m_\alpha{}^2}{p^2-m_\alpha{}^2+i\varepsilon}$$

$$-\frac{m_\alpha{}^2}{p^2-M^2+i\varepsilon}+\frac{m_\alpha{}^2(M^2-m_\alpha{}^2)}{(p^2-M^2+i\varepsilon)^2}=\frac{m_\alpha{}^2(M^2-m_\alpha{}^2)^2}{(p^2-m_\alpha{}^2+i\varepsilon)\,(p^2-M^2+i\varepsilon)^2}. \tag{19}$$

This makes each α-term or a-term separately finite. The procedure is reminiscent of a Pauli–Villars regularization [21], but now none of the three subtractions makes any difference to the total, because of the three conditions (11)–(13). Proceeding in this way, we find ourselves with the standard convergent integral [22]

$$\int\frac{d^4p}{(p^2-m^2+i\varepsilon)\,(p^2-M^2+i\varepsilon)^2}=\frac{-i\pi^2}{(M^2-m^2)^2}\left\{M^2-m^2-m^2\ln\frac{M^2}{m^2}\right\}. \tag{20}$$

Finally, we put everything together, and use (11)–(13) again to cancel out the subtraction mass M. We are left with

$$\langle\Theta_{\mu\nu}\rangle_0=(1/64\pi^2)g_{\mu\nu}\Big(\sum_\alpha m_\alpha{}^4\ln m_\alpha{}^2-\sum_a m_a{}^4\ln m_a{}^2\Big)\quad(D=4). \tag{21}$$

Analogous calculations in three and two space-time dimensions, using just the first two convergence conditions (11) and (12), yield

$$\langle\Theta_{\mu\nu}\rangle_0=-(1/12\pi)g_{\mu\nu}\Big(\sum_\alpha|m_\alpha|^3-\sum_a|m_a|^3\Big)\quad(D=3), \tag{22}$$

$$\langle\Theta_{\mu\nu}\rangle_0=-(1/8\pi)g_{\mu\nu}\Big(\sum_\alpha m_\alpha{}^2\ln m_\alpha{}^2-\sum_a m_a{}^2\ln m_a{}^2\Big)\quad(D=2). \tag{23}$$

The logarithms which arise here when D is even do not upset the physical dimension of $\Theta_{\mu\nu}$, because of conditions (13) or (12). If all masses vanish then so does $\langle\Theta_{\mu\nu}\rangle_0$, and we obtain a strictly scale-invariant theory. In cases with non-zero masses a finite and consistent Schwinger subtraction can be made if required, to bring $\langle\Theta_{\mu\nu}\rangle_0$ to zero.

2. Macrophysics and ultraviolet divergences, real scattering and closed loops

Emboldened by our successful mathematical resolution of the paradoxical limiting aspects of free fields, we now introduce some further considerations of a physico-mathematical nature, in order to see how we might reasonably try to extend our ideas to the interacting case. For it is of course the interaction divergences which have usually been reckoned to be the more worrying.

In the early days Heisenberg [23] argued on dimensional grounds from the uncertainty principle, and associated the divergent intermediate state summations of the perturbation expansion with waves that are travelling over short distances. From this he inferred the existence of a fundamental length, below which local field concepts would fail. This rationalization of the ultraviolet problem used to be generally regarded as acceptable, if only because no good mathematical sense can be made of the local field products that are involved. Thus Feynman [22] used convergence factors extensively as a model for the presumed short distance modifications in the interaction and spoke elsewhere of "ignorance at short distances", Umezawa [24] referred to "the structure of the elementary particles hidden behind the renormalization procedure", Schwinger [25] suggested that "a convergent theory cannot be formulated consistently within the framework of present space-time concepts", and so forth. Pauli [26] and others studied integro-differential field equations, in which relativistically smeared field products replaced the singular local products of standard formulations.

Works such as those mentioned embodied in their philosophy the basic assumption that there is a necessary and inevitable two-way connection between high momentum physics and short distance physics. But we shall argue here that this very assumption is itself a mistake, and that it has prevented cognition of the true nature of the problem. We must also caution the reader against a possible confusion between the mathematical utilization of cut-offs to formulate limiting sequences, as in Section 1, and the use of cut-offs to simulate the presumed physical manifestations of Heisenberg's fundamental length. For the logical distinction between these two different functions does not seem to have been drawn heretofore.

That something is very much amiss with the high momentum–short distance philosophy could already have been suspected by reference to the unperturbed zero-point energy problem, which we have resolved in Section 1. It is clear that this problem arises solely because of the relativistic nature of the energy–momentum spectrum, and that it provides no point of entry for short-distance modifications. Another simple argument comes by consideration of diffraction scattering at very high energies. The energy dependence of the angular width of the diffraction peak is satisfactorily explained in terms of the interference of Huygens wavelets, a concept which clearly requires that the carrier space of the wavelets be definable, even on the length scale set by the extremely short wavelengths involved. But to see and learn from the stark realities of the situation, as it concerns divergent integrals, we cannot do better than look at the phenomenon of static vacuum polarization, in which a weak static macroscopic electromagnetic field induces currents in the electron–positron vacuum. We shall examine this from the point of view afforded by a normal mode analysis. Each electron mode of excitation is associated with a positive energy eigen-solution $\psi_n(x)$ of the coupled Dirac equation, and contributes to the vacuum a current density $-\frac{1}{2}(-e)\hat{\psi}_n^*(x)\gamma_\mu\psi_n(x)$, and each positron mode contributes a term of opposite signature. Here $\hat{\psi}^*\gamma_0\psi$ (cf. Section 1) and e are both positive and $-e$ is the signed charge of the electron, and the additional signed factor $-\frac{1}{2}$ comes from the Fermi statistics, as explained below equation (8). Thus electron modes contribute positive charge to the vacuum, and positron modes contribute negative charge. In the absence of the applied electromagnetic field these cubically divergent zero-point charges and currents mutually cancel, much as the energies cancel in Section 1, and thus the vacuum is left electrically neutral. When the field is applied it perturbs the electron and positron modal functions $\psi_n(x)$ oppositely, depending on their charge signature. The electron–positron cancellation is no longer complete, and a quadratic divergence in the resultant vacuum polarization current arises on summing over all the modes. No short distances are discernible anywhere, for the applied field is macroscopic, while the modal functions themselves are running waves, which propagate unhindered over macroscopic regions of space and time.

This example provides a context in which a divergent integral arises, even though the short-distance concept is quite clearly irrelevant. How then can we explain the universal acceptance of the said concept? We shall argue that the answer to this lies not in the physics, but in the deceptive mathematical subtleties of perturbation theory. The customary perturbative formulations of the vacuum polarization phenomenon pay indeed no direct attention to the inner structure of the vacuum, nor to the composition of its currents. They are concerned only with the overall change induced in the total vacuum current by the applied field, and this they represent in terms of the creation and annihilation of virtual electron–positron pairs. The associated closed loops contain within themselves various unperturbed modal summations of the general oscillatory type $\sum_n \psi_n(x)\hat{\psi}_n^*(y)$, corresponding to the Green function $S(x-y)$ of the electron-positron field. When such summations are performed with a large separation $x-y$ there is an obvious tendency for self-annulment through destructive interference, while at short distances the sum builds up disastrously. This is where the notion of short distances comes from. But the modal concepts of the present paper permit an alternative and startlingly different viewpoint. They permit one to see that the vacuum polarization can also be regarded as due to the scattering of zero-point electrons and positrons, each already present with weight $-\frac{1}{2}$ in the unperturbed vacuum. Scattering is a gentle process compared with pair creation, and the higher the energy of the particle the more gentle its scattering becomes. Our intuitive understanding of gentle scattering is much more reliable than our intuitive understanding of pair creation. We can link it directly to visible macrophysical realities, such as the curvatures of cloud-chamber tracks, a consideration to which we shall return presently. But let us first fully clarify our thinking on short distances! The scattering picture which we advocate leads to perturbed modal summations with the structure $\sum_n \hat{\psi}_n^*(x)\gamma_\mu\psi_n(x)$. Here there is no visible interference, since the summand is not oscillatory, and since only one world point is involved there is nothing in sight that one could identify as a distance. The sum diverges simply because so many modes contribute, and not because of any bad high energy behaviour of the individual contributions. In this way of looking at things we see each vacuum particle influenced by the whole of its past history. But there is nevertheless one question about times and distances that we can properly ask. We can ask how the perturbed overall sum $\sum_n \psi_n^*(x)\gamma_\mu\psi_n(x)$ depends on the presence of applied fields at points distant from x. To this the answer is that it is insensitive. In other words, the short-distance effects found in the conventional perturbation treatments of the phenomenon express merely a statistical property of cancellation arising over long distances in the summation over all perturbed normal modes. They do not express some lack of long-distance dependence on the part of each individual and separate vacuum particle, and they do not express some short distance enhancement in the dynamics of these individual particles.

To get a good understanding of the properties of the perturbed vacuum we should now look at the perturbed modal functions $\psi_n(x)$ in their own right, not at complicated bilinear sums! We see that each perturbed $\psi_n(x)$ has two distinct roles in the static external field, depending on whether the associated Fermi–Dirac oscillator is unexcited (vacuum polarization) or excited (real electron or positron). The customary distinction between off-shell and on-shell processes does not arise in this sort of analysis, and therefore we see that any attempt to modify the one will inevitably affect also the other. Thus, following historical precedent, we might try to make the divergent vacuum polarization sum finite by smearing the electron–positron field in the interaction, so as to reduce the perturbing effect on the high-energy modes (the electromagnetic potential is supposed macroscopic, and therefore already smooth, so there is no use in smearing it!). But it now becomes evident that this would unavoidably upset the behaviour of real particles too, since these are described by the very same modal functions. Let us consider the motion of these real particles. At high energies a semiclassical approximation to the Dirac equation will suffice, and so we find real electrons moving under the Lorentz force law, with their spins precessing accordingly. To upset this is unthinkable! The Dirac equation is known to give the right high energy behaviour for real particles in a macroscopic field! Therefore if field theory is ever to be put right, it will have to be done by arranging cancellations of infinities between different fields, much as we have already done in Section 1. So long as the normal mode concept of quantum mechanics survives, it will not be possible to make field theory both finite and physically acceptable by mutilating the local structure of its interactions, nor by introducing strange new notions about the nature of space-time. These overiding considerations prompt the present paper, and motivate the exploratory two-dimensional work reported in the remaining pages.

3. *The ultraviolet problem in two dimensions*

In most of this section we study quantum field theories in a space-time with dimension $D = 2$. There is no room for spin concepts in this low dimension, and so we need only two basic field equations, the real Klein–Gordon equation for bosons and the real Majorana–Dirac equation for fermions. The two real Dirac matrices γ_0 and γ_1 can be identified as Pauli matrices; $\gamma_0 = i\sigma_2, \gamma_1 = \sigma_1$. They obey $\{\gamma_\mu, \gamma_\nu\} = -2g_{\mu\nu}$, as usual. Both the matrices $\gamma_0\gamma_\mu$ are symmetric, and the Majorana adjoint is therefore $\hat{\psi} = -\psi\gamma_0$, as in four dimensions. The Majorana–Dirac equation $(\gamma_\mu\partial^\mu + m)\psi = 0$ appears in matrix form as

$$\begin{pmatrix} m & \partial_0 - \partial_1 \\ -\partial_0 - \partial_1 & m \end{pmatrix} \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix} = 0. \tag{24}$$

σ_3 serves as a mass-reversal matrix, but mass reversal changes the parity signature of ψ. It seems simpler to retain a uniform parity operation (e.g. $\psi \to \gamma_0\psi$) and to allow masses of both signs. Thus we write the general positive energy rest frame solution in a form that allows for this:

$$\psi = \begin{pmatrix} 1 \\ -i \operatorname{sgn} m \end{pmatrix} e^{-i|m|x^0}. \tag{25}$$

The extreme relativistic limiting forms of the solutions for motion in the positive and negative directions of the space coordinate are respectively

$$\psi = \begin{pmatrix} 1 \\ 0 \end{pmatrix} e^{-i\omega(x^0 - x^1)} \quad \text{and} \quad \psi = \begin{pmatrix} 0 \\ 1 \end{pmatrix} e^{-i\omega(x^0 + x^1)}. \tag{26}$$

The zero-point energy density from one Majorana–Dirac field is

$$\langle \Theta_{00} \rangle_0 = -\tfrac{1}{2} \int_{-\infty}^{\infty} \frac{dp}{2\pi} (p^2 + m^2)^{1/2}. \tag{27}$$

We now consider field theories depending on several commutative real Klein–Gordon fields ϕ_α ($\alpha = 1, 2, \ldots$) and several anticommutative real Majorana–Dirac fields ψ_a ($a = 1, 2, \ldots$). We assume for simplicity that all the Klein–Gordon fields transform as scalars under the parity operation $x^1 \to -x^1$, and we use only such parity-conserving interactions as would be renormalizable in $D = 4$. Subject to these provisoes, the most general Lagrangian density is

$$\mathscr{L} = \tfrac{1}{2} \sum_\alpha (\partial_\mu\phi_\alpha\partial^\mu\phi_\alpha - m_\alpha{}^2\phi_\alpha{}^2) + \tfrac{1}{2} i \sum_a \hat{\psi}_a(\gamma_\mu\partial^\mu + m_a)\psi_a - \sum_{\alpha\beta\gamma} g_{\alpha\beta\gamma}\phi_\alpha\phi_\beta\phi_\gamma$$
$$- \sum_{\alpha\beta\gamma\delta} g_{\alpha\beta\gamma\delta}\phi_\alpha\phi_\beta\phi_\gamma\phi_\delta - i \sum_{\alpha ab} g_{\alpha ab}\varphi_\alpha\hat{\psi}_a\psi_b. \tag{28}$$

The coupling constants $g_{\alpha\beta\gamma}$ and $g_{\alpha\beta\gamma\delta}$ are real and totally symmetric, the coupling constants $g_{\alpha ab}$ are real and symmetric under exchange of a with b. The Lagrangian contains no non-diagonal mass terms, since these could always be removed by performing suitable linear transformations among the fields. In the same way it contains no linear terms, since these could always be removed by adding constants to the fields (except in the case of zero mass, where consistency arguments can be used to the same end).

The Feynman graphs associated with (28) are built from fermion lines ——, boson lines - - - - - -, and vertices. There are three different vertices:

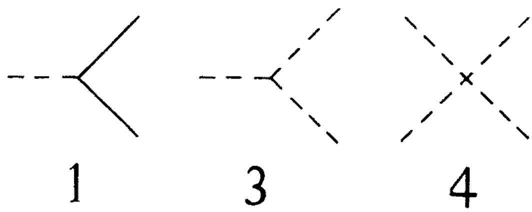

Fig. 1. Identification of the three elementary vertices

We distinguish the vertices according to the number of boson lines attached, this number being either 1, 3 or 4.

Consider now a connected graph having I_F internal fermion lines, E_F external fermion lines, I_B internal boson lines, E_B external boson lines, V_1 vertices of type 1, V_3 of type 3, and V_4 of type 4. Following the familiar Dyson–Salam convergence criteria [27, 28] we pick up the vertices one at a time and count the ends of the lines to which they attach. Every internal line gets counted twice, and every external line once, and therefore

$$2V_1 = 2I_F + E_F,$$
$$V_1 + 3V_3 + 4V_4 = 2I_B + E_B. \tag{29}$$

Each internal fermion line gives a factor $-i\gamma p + m$ in the denominator, and each internal boson line a factor $p^2 - m^2$, the number of momenta p to be integrated over is $I_B + I_F$, and the number of momentum δ-functions from the space-time integrations at the vertices is $V_1 + V_3 + V_4 - 1$. Thus, working for the moment in D space-time dimensions, we estimate the convergence properties of the graph through the convergence index

$$C \equiv D(I_B + I_F) - D(V_1 + V_3 + V_4 - 1) - I_F - 2I_B. \tag{30}$$

The graph is said to be primitively divergent if $C \geqslant 0$, and it is said to be superficially convergent if $C < 0$. According to Nakanishi [29] and Weinberg [30] a graph is genuinely convergent if it and all the subgraphs within it are superficially convergent. Using (29) in (30) to eliminate I_B and I_F, we see that the condition for superficial convergence can be written as

$$C \equiv \tfrac{1}{2}(D-6)V_3 + \tfrac{1}{2}(D-4)(V_1 + 2V_4) - \tfrac{1}{2}(D-2)E_B - \tfrac{1}{2}(D-1)E_F + D < 0. \tag{31}$$

$D = 6$ is the critical dimension, above which there is no hope of renormalizability. $D = 4$ is the critical dimension for renormalizability when fermions are present. When $D < 4$ every graph with a fixed set of external lines becomes superficially convergent if enough extra vertices are put in. Theories with this property are said to be superenormalizable, since they display only a finite number of primitively divergent graphs. Theories which are merely renormalizable have an infinite number of primitively divergent graphs, since for them one can put in more and more vertices without ever satisfying (31).

The ultraviolet problem is particularly mild in the case $D = 2$. There are only nine primitively divergent graphs, and they all diverge only logarithmically. They come in three categories; tadpoles ($E_F = 0$, $E_B = 1$), boson self energy graphs ($E_F = 0$, $E_B = 2$), and vacuum bubbles ($E_F = E_B = 0$). The nine primitively divergent graphs are drawn in Figs. 2–5. All graphs other than these are superficially convergent. A graph that is superficially convergent can of course still diverge, but only if it contains primitive divergents inside itself. The theory will therefore be finite to all orders if we can achieve mutual cancellations of divergences between the two tadpoles, and between the two boson self energy graphs, and between the five bubbles. We shall show that this can all be arranged, by a natural extension of the boson–fermion compensation principle of Section 1.

4. *A boson-fermion compensation principle*

In the lowest approximation the quantum system (28) vibrates harmonically around the classical solution $\phi = 0$, $\psi = 0$, and so we shall certainly need to apply the convergence conditions (11) and (12). But the problem now is to extend (12), to take account of the perturbing effects of the interaction. We can approach this problem by considering the response of the system to externally applied forces. Suppose we force the system by adding to $\mathscr{L}$ various terms of the type $m_\alpha^2 f_\alpha \phi_\alpha$, where the forces $m_\alpha^2 f_\alpha$ are constants. In lowest approximation the system will then vibrate harmonically around the displaced configuration $\phi_\alpha = f_\alpha$. We demand, as a reasonable heuristic criterion, that the zero-point energy calculated in the naive quadratic approximation should stay finite under any such displacement. We have in mind that a satisfactory field system should respond finitely to harmless-looking perturbations such as that suggested. Quantum electrodynamics provides an effective example of the idea; the work done in lifting the electron–positron vacuum through a fixed electrostatic potential is finite only because there are equal numbers of electron and positron modes, which contribute oppositely.

Consider first the boson oscillators in (28). These are governed by the boson potential energy

$$\mathscr{V}_B = \tfrac{1}{2} \sum_\alpha m_\alpha^2 \phi_\alpha^2 + \sum_{\alpha\beta\gamma} g_{\alpha\beta\gamma} \phi_\alpha \phi_\beta \phi_\gamma + \sum_{\alpha\beta\gamma\delta} g_{\alpha\beta\gamma\delta} \phi_\alpha \phi_\beta \phi_\gamma \phi_\delta. \tag{32}$$

In the quadratic approximation this yields the following symmetric mass-squared matrix for the displaced system;

$$\begin{aligned}(\mathscr{M}^2_B)_{\alpha\beta} &\equiv [\partial^2 \mathscr{V}_B / \partial\phi_\alpha \partial\phi_\beta]_{\phi=f} \\ &= (m_\alpha^2)\delta_{\alpha\beta} + 6 \sum_\gamma g_{\alpha\beta\gamma} f_\gamma + 12 \sum_{\gamma\delta} g_{\alpha\beta\gamma\delta} f_\gamma f_\delta.\end{aligned} \tag{33}$$

Similarly the fermion oscillators acquire a symmetric mass matrix:

$$(\mathscr{M}_F)_{ab} = (m_a)\delta_{ab} - 2 \sum_\alpha g_{\alpha ab} f_\alpha. \tag{34}$$

Now equation (10) shows that the zero-point energies of the individual high-momentum modes are only slightly displaced, by amounts conditioned directly by the squared masses.

Therefore the zero-point energy cannot possibly stay finite for all displacements f unless

$$\mathrm{Tr}\,(\mathscr{M}^2{}_\mathrm{B}) - \mathrm{Tr}\,(\mathscr{M}_\mathrm{F})^2 = 0, \quad \text{identically in } f. \tag{35}$$

Clearly the imposition of (35) will give some improvement in the ultraviolet properties of the system, and since the divergences are only logarithmic it might, with good fortune, be enough to make all three sets of primitive divergents simultaneously finite. We shall show by direct computation of graphs that this is precisely what happens.

The equation (35) expresses within itself three distinct convergence conditions. The terms without f yield (12) again. Those linear in f produce the further conditions

$$3\sum_\beta g_{\alpha\beta\beta} + 2\sum_a m_a g_{\alpha aa} = 0, \tag{36}$$

and those quadratic in f give

$$3\sum_\gamma g_{\alpha\beta\gamma\gamma} - \sum_{ab} g_{\alpha ab} g_{\beta ab} = 0. \tag{37}$$

In section 5.6 we shall demonstrate that (36) and (37) have many acceptable solutions.

5. *Graphical proof of finiteness*

The Lagrangian (28) leads to the following non-zero equal-time commutators and anticommutators:

$$[\phi_\alpha(\boldsymbol{x}), \dot{\phi}_\beta(\boldsymbol{x}')] = i\delta(\boldsymbol{x}-\boldsymbol{x}')\delta_{\alpha\beta},$$
$$\{\psi_{ar}(\boldsymbol{x}), \psi_{bs}(\boldsymbol{x}')\} = \delta(\boldsymbol{x}-\boldsymbol{x}')\delta_{ab}\delta_{rs}. \tag{38}$$

From the first of these it follows that the factors Δ to be inserted in the boson lines of the graphs for a pair of fields $\phi_\alpha(x)$ and $\phi_\alpha(x')$ are given by

$$\Delta_\alpha(x-x') \equiv \frac{i\delta^2(x-x')}{-(\partial\partial+m_\alpha{}^2)} = \frac{i}{(2\pi)^2}\int \frac{d^2p e^{-ip(x-x')}}{p^2-m_\alpha{}^2+i\varepsilon} = \Delta_\alpha(x'-x). \tag{39}$$

For a pair of Majorana fields $\psi_{ar}(x)$ and $\psi_{as}(x')$ written under the T symbol in the order $\psi_{ar}(x)\psi_{as}(x')$ the factor for insertion is

$$[S_a(x-x')\gamma_0]_{rs} \equiv \left[\frac{\delta^2(x-x')}{\gamma\partial+m_a}\gamma_0\right]_{rs}$$
$$= \left[\frac{1}{(2\pi)^2}\int \frac{d^2pe^{-ip(x-x')}}{-i\gamma p+m_a-i\varepsilon}\gamma_0\right]_{rs} = -[S_a(x'-x)\gamma_0]_{sr}. \tag{40}$$

The factors γ_0 arise here because we refer always to $\psi(x')$, rather than to $\bar{\psi}(x')$ (recall that for a Majorana field we have simply $\bar{\psi} = -\psi\gamma_0$, so that separate consideration of $\bar{\psi}$ serves no purpose). The vertex factors, duly adjusted for these Majorana field conventions, are

$$g_{\alpha ab}\gamma_{0rs}, \qquad -ig_{\alpha\beta\gamma} \quad \text{and} \quad -ig_{\alpha\beta\gamma\delta} \tag{41}$$

for the vertices of types 1, 3 and 4 respectively (cf. Fig. 1).

5.1. Finiteness of tadpoles

Consider the primitively divergent tadpole graphs of Fig. 2. If we were to calculate the energy shift of the vacuum under the perturbation $\sum_{\alpha} m_\alpha^2 f_\alpha \phi_\alpha$ then these very same tadpoles would appear, but multiplied overall by the force constants f_α. The linear part (36)

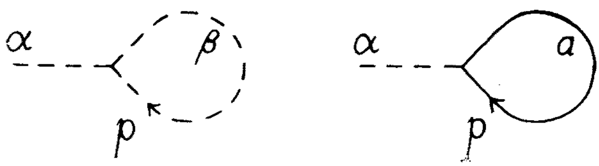

Fig. 2. The cancellation scheme for the primitively divergent tadpole graphs

of the convergence condition for the energy shift of the forced vacuum must therefore bear upon them directly. An improvement in their convergence properties is thus guaranteed in advance. We now check this by straight evaluation.

Omitting the external α-line factor Δ_α, the two α-tailed primitively divergent tadpoles in Fig. 2 correspond to the expression

$$-3i \sum_{\beta} g_{\alpha\beta\beta} \Delta_\beta(0) + \sum_{a} g_{\alpha aa} \operatorname{Tr} S_a(0), \tag{42}$$

which can be interpreted on our picture by saying that the zero-point particles act as a source for the field ϕ_α. The second term here is nominally linearly divergent, but in actuality it diverges only logarithmically, since Tr $\gamma_\mu = 0$. Putting in the Green functions (39) and (40) and using Tr $1 = 2$ we write (42) as

$$\int \frac{d^2 p}{(2\pi)^2} \left\{ 3 \sum_{\beta} \frac{g_{\alpha\beta\beta}}{p^2 - m_\beta{}^2 + i\varepsilon} + 2 \sum_{a} \frac{m_a g_{\alpha aa}}{p^2 - m_a{}^2 + i\varepsilon} \right\}, \tag{43}$$

which converges under the condition (36), and is equal to

$$(i/4\pi)\,(3 \sum_{\beta} g_{\alpha\beta\beta} \ln m_\beta{}^2 + 2 \sum_{a} m_a g_{\alpha aa} \ln m_a{}^2). \tag{44}$$

This expression is dimensionally correct, by virtue of (36), notwithstanding the logarithmic factors.

We notice that (43) continues to converge when $D = 3$, but becomes logarithmically divergent when $D = 4$.

5.2. Finiteness of self energies

Consider next the contribution which the two primitively divergent boson self energy graphs of Fig. 3 make to the boson propagator $\langle T\phi_\alpha(x)\phi_\beta(y)\rangle_0$. These graphs would appear linked to quadratic factors $f_\alpha f_\beta$ in the calculation of the displaced vacuum energy, and therefore it is guaranteed in advance that the quadratic part (37) of the vacuum energy convergence condition (35) will improve their convergence properties. Omitting again the external line factors, and multiplying also by i, we obtain the mass-squared operator

$$12 \sum_{\gamma} g_{\alpha\beta\gamma\gamma} \Delta_\gamma(0) \delta^2(x-y) - 2i \sum_{ab} g_{\alpha ab} g_{\beta ab} \operatorname{Tr} S_a(x-y) S_b(y-x). \tag{45}$$

The customary interpretation of Feynman graphs makes use of Fock-space particle concepts and the perturbative concept of virtual transition. In this interpretation the first term in (45) gives some conceptual difficulty, but the second is more amenable. The second term brings together the contributions of intermediate states containing either two fermions, or two fermions with the two bosons, depending on the relative time order of x and y. But in the alternative picture which we advocate here we interpret the first term of (45) as describing forward scattering of the boson on the vacuum bosons which are

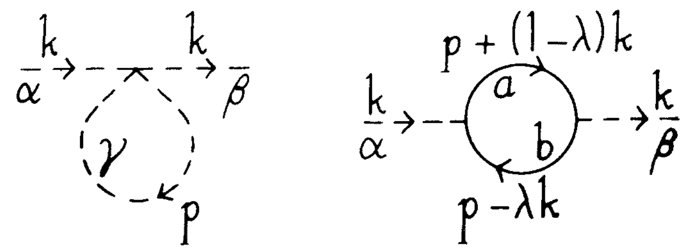

Fig. 3. The cancellation scheme for the primitively divergent boson self energy graphs, showing the role of the momentum sharing factor λ

already there, and the second term as describing forward scattering of the boson on the vacuum fermions which are already there. The scattering has to be forward since the vacuum carries no momentum.

With the two terms of (45) thus interpreted in a uniform way it now becomes more plausible than before that we shall encounter a cancellation of divergences between them. Nevertheless, the cancellation is between two graphs with different topologies. Because of this we inevitably must anticipate some ambiguity as to what should be cancelled against what. To avoid prejudging this matter we may introduce a sharing factor λ for the momenta in the lines of the fermion loop. λ is just a real constant, the precise role of which is indicated in Fig. 3. The Fourier transform of (45) at two-momentum k thus assumes the λ-dependent form

$$\delta m^2(k)_{\alpha\beta} = 4i \int \frac{d^2p}{(2\pi)^2} \Bigg\{ 3 \sum_{\gamma} \frac{g_{\alpha\beta\gamma\gamma}}{p^2 - m_\gamma{}^2 + i\varepsilon}$$

$$- \sum_{ab} \frac{g_{\alpha ab} g_{\beta ab} \{ m_a m_b + [p + (1-\lambda)k]\,(p - \lambda k)\}}{\{[p+(1-\lambda)k]^2 - m_a{}^2 + i\varepsilon\}\,\{(p-\lambda k)^2 - m_b{}^2 + i\varepsilon\}} \Bigg\}, \tag{46}$$

which expresses sufficiently well the nature of the said ambiguity. We see at once that (46) is finite under the condition (37), as anticipated, and that it is so for any value of λ. We must now look at the λ-dependence. The integral in which λ appears, namely that arising from the second term in (46), is by itself divergent, but only logarithmically. It follows that the shift theorem of Jauch and Rohrlich [31] may be applied. According to this theorem the value of (46) is in fact entirely independent of λ. Thus (46) is both finite and unambiguous. The same applies in three dimensions. To establish the λ-independence in the case $D = 3$ one needs an extension of the Jauch–Rohrlich shift theorem to linearly divergent integrals, namely the theorem that

$$\int d^3p \left\{ \frac{1}{(p-k)^2 - m^2 + i\varepsilon} + \frac{1}{(p+k)^2 - m^2 + i\varepsilon} - \frac{2}{p^2 - m^2 + i\varepsilon} \right\} = 0. \tag{47}$$

In evaluating (46) it is very useful to introduce a subtraction mass, which then cancels out, as in the derivation of Eq. (21). We find

$$\begin{aligned}\delta m^2(k)_{\alpha\beta} &= \frac{1}{\pi}\left\{ \sum_{ab} g_{\alpha ab} g_{\beta ab} \ln |m_a m_b| - 3 \sum_{\gamma} g_{\alpha\beta\gamma\gamma} \ln m_\gamma^2 \right\} \\ &+ \frac{1}{\pi} \sum_{ab} \frac{g_{\alpha ab} g_{\beta ab}[k^2 - (m_a + m_b)^2]}{B_-(m_a^2\ m_b^2 : k^2) - B_-(m_a^2, m_b^2 : k^2)} \ln \left\{ \frac{B_+(k^2, m_a^2 : m_b^2) B_+(k^2, m_b^2 : m_a^2)}{B_-(k^2, m_a^2 : m_b^2) B_-(k^2, m_b^2 : m_a^2)} \right\}\end{aligned} \tag{48}$$

where, for any three variables m^2, m'^2, m''^2 the functions $B_\pm$ are defined by

$$\begin{aligned} B_\pm(m^2, m'^2 : m''^2) &\equiv m^2 + m'^2 - m''^2 \\ &\pm [(m + m' + m'')(m - m' - m'')(m' - m'' - m)(m'' - m - m')]^{1/2}. \end{aligned} \tag{49}$$

The physical Riemann sheet of the k^2-dependent logarithm in (48) is that on which the logarithm is real for negative real k^2. The logarithm shows a branch point at the physical threshold $k^2 = (|m_a| + |m_b|)^2$, and in the usual way we may run a branch cut from here along the positive real axis of k^2. The logarithm possesses imaginary parts equal respectively to $\mp 2\pi$ all along the upper/lower (retarded/advanced) sides of the cut, and the $i\varepsilon$ prescription indicates that one should work on the upper side of the cut when k^2 is real.

The behaviour of (48) for large k^2 may be of interest. We find

$$\delta m^2(k^2)_{\alpha\beta} \sim \pi^{-1} \sum_{ab} g_{\alpha ab} g_{\beta ab} \ln(-k^2), \qquad |k^2| \to \infty. \tag{50}$$

The sign of this is consistent with the idea that we deal with an unbounded but real spectrum of intermediate state masses.

Further mass contributions arise from other convergent or superficially convergent self energy graphs. Therefore the physical masses with which the particles ultimately propagate in this theory depend in a complicated way upon the bare masses and the bare coupling constants. This is what one would expect in a finite theory. It is essential here to understand that the convergence conditions (11), (12), (36) and (37) all refer to the bare parameters.

We remark finally that the functions $B_\pm$ defined by Eqs (49) turn up in many quantum field integrals, and that they have an interesting kinematical interpretation. We treat this in the Appendix, since it is not vital to our main theme.

5.3. Finiteness of tadpole bubbles

Using the tadpole formula (44) we readily find that the three bubble graphs of Fig. 4 contribute to the vacuum stress $\langle\Theta_{\mu\nu}\rangle_0$ a finite extra term $\langle\delta\Theta_{\mu\nu}\rangle_0$ given by

$$\langle\delta\Theta_{\mu\nu}\rangle_0 = -g_{\mu\nu}/(32\pi^2)\sum_\alpha m_\alpha^{-2}(3\sum_\beta g_{\alpha\beta\beta}\ln m_\beta^2+2\sum_a m_a g_{\alpha aa}\ln m_a^2)^2. \tag{51}$$

The virtual particle transition concepts of perturbation theory enable some partial understanding of the significance of this term to be gained, in so far as it obviously does refer to the spontaneous emission and subsequent absorption of a virtual boson α by the

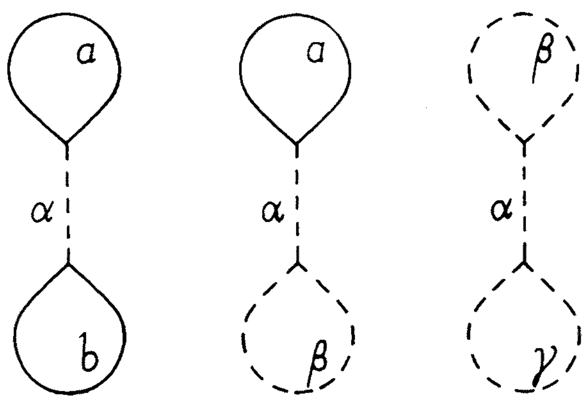

Fig. 4. The three primitively divergent tadpole bubble graphs

vacuum. But, as with the elementary tadpole itself (section 5.1), the two tadpole heads elude description in terms of virtual transitions, and definitely require the recognition of zero-point concepts as well. Perhaps the most illuminating and helpful interpretation comes by saying that the zero-point particles already present in the vacuum are acting on each other by exchanging one α-boson in the t-channel. The overall sign is in this view explained, for it is consistent with the well known fact that the Yukawa force between like particles is attractive.

5.4. Finiteness of other bubbles

We come now to the two primitively divergent vacuum bubbles of Fig. 5. The first of these accounts for some of the self energy corrections to the zero-point bosons — a process which can in turn be viewed or conceptualized in terms of zero-point boson–boson forward scattering proceeding through the pointlike four-boson interaction. In a similar way the second bubble of Fig. 5 folds together no fewer than six imaginable vacuum effects, namely s- and u-channel one-boson exchange forces acting between zero-point fermions,

s- and u-channel one-fermion exchange forces acting between zero-point fermions and zero-point bosons, and self energy corrections both to the zero-point bosons and to the zero-point fermions. From this point of view the Feynman integral represents a truly remarkable unification of many different processes. The conventional Feynman interpretation does not permit such kaleidoscopic variety — one speaks there of a single process in which an intermediate or admixed state contains one virtual boson and two virtual fermions, relative to vacuum, and one does not try to picture the vacuum itself.

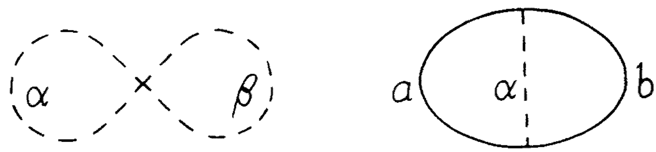

Fig. 5. The remaining primitively divergent bubble graphs

The two bubbles of Fig. 5 have different topologies, and the second one contains overlapping loops. We shall prove nevertheless that their divergences cancel, subject to (37), and that the finite remainder is unambiguous. At the present time I find myself unable to explain on general grounds why this cancellation occurs. The bubbles have no external lines by which one might hook them onto the applied forces, and so we must infer that our heuristic forcing principle is here working for us better than we had any right to expect. In other words, (35) is the right convergence condition, but there should be a more illuminating way to get to it.

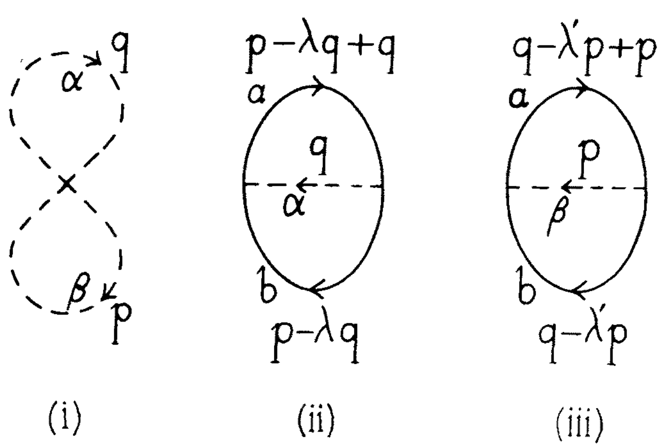

Fig. 6. Cancellation scheme for the bubble graphs, showing the resolution of the overlap, and the role of the momentum sharing factors λ and λ'. The cancellation in the p-subintegration is between (*i*) and (*ii*), and that in the q-subintegration is between (*i*) and (*iii*)

Let us now formulate this miraculous cancellation! The first graph of Fig. 5 contains two distinct and non-overlapping divergent loops, to which momenta p and q can be assigned, as shown in Fig. 6(i). Following the general criteria discovered and laid down by Salam [28] we must first ensure convergence in the p- and q-subintegrations. To compensate the divergence in the p-subintegration we must of necessity use the fermion loops of the second graph of Fig. 5, since this is the only divergent loop available. This leads to the momentum assignment of Fig. 6(ii), in which the boson momentum q is shared between the two fermion lines through a real constant sharing factor λ, as yet undetermined. We find that the divergence in the p-subintegration actually does cancel if Fig. 6(ii) is counted with weight $\frac{1}{2}$. This very fortunately leaves us with the alternative assignment of Fig. 6(iii), also with weight $\frac{1}{2}$, to compensate the divergence of the q-subintegration. Proceeding thus, we manipulate the raw contribution to $\langle\Theta_{00}\rangle_0$, namely

$$\langle\delta\Theta_{00}\rangle_0 \equiv 3\sum_{\alpha\beta} g_{\alpha\alpha\beta\beta}\Delta_\alpha(0)\Delta_\beta(0) - i\sum_{\alpha ab}(g_{\alpha ab})^2\int d^2x\,\Delta_\alpha(x)\,\mathrm{Tr}\,S_b(x)S_a(-x), \tag{52}$$

into the form

$$\begin{aligned}\langle\delta\Theta_{00}\rangle_0 = &-\frac{1}{(2\pi)^4}\iint d^2p\,d^2q\left\{3\sum_{\alpha\beta} g_{\alpha\alpha\beta\beta}\frac{1}{(p^2-m_\beta{}^2)(q^2-m_\alpha{}^2)}\right.\\ &-\sum_{\alpha ab}(g_{\alpha ab})^2\frac{m_a m_b+(p-\lambda q+q)(p-\lambda q)}{(q^2-m_\alpha{}^2)[(p-\lambda q+q)^2-m_a{}^2][(p-\lambda q)^2-m_b{}^2]}\\ &\left.-\sum_{\beta ab}(g_{\beta ab})^2\frac{m_a m_b+(q-\lambda' p+p)(q-\lambda' p)}{(p^2-m_\beta{}^2)[(q-\lambda' p+p)^2-m_a{}^2][(q-\lambda' p)^2-m_b{}^2]}\right\},\end{aligned} \tag{53}$$

in which the convergence of each of the two 2-dimensional subintegrations is clearly seen (we suppress the $i\varepsilon$ instruction). It is now helpful to simplify the integrand by making in it the substitution

$$(p-\lambda q+q)(p-\lambda q) = \tfrac{1}{2}(p-\lambda q+q)^2+\tfrac{1}{2}(p-\lambda q)^2-\tfrac{1}{2}q^2. \tag{54}$$

This brings (53) to the more easily managed form

$$\begin{aligned}\langle\delta\Theta_{00}\rangle_0 = &-\frac{1}{(2\pi)^4}\iint d^2p\,d^2q\left\{3\sum_{\alpha\beta} g_{\alpha\alpha\beta\beta}\frac{1}{(p^2-m_\beta{}^2)(q^2-m_\alpha{}^2)}\right.\\ &-\sum_{\alpha ab}(g_{\alpha ab})^2\left[\frac{\frac{1}{2}}{(q^2-m_\alpha{}^2)[(p-\lambda q)^2-m_b{}^2]}\right.\\ &+\frac{\frac{1}{2}}{(q^2-m_\alpha{}^2)[(p-\lambda q+q)^2-m_a{}^2]}-\frac{\frac{1}{2}}{[(p-\lambda q+q)^2-m_a{}^2][(p-\lambda q)^2-m_b{}^2]}\end{aligned}$$

$$+\frac{\frac{1}{2}(m_a+m_b)^2-\frac{1}{2}m_\alpha^{\,2}}{(q^2-m_\alpha^{\,2})\,[(p-\lambda q+q)^2-m_a^{\,2}]\,[(p-\lambda q)^2-m_b^{\,2}]}\Bigg]$$

$$-\sum_{\beta ab}(g_{\beta ab})^2\Big[\alpha\to\beta,\,p\to q,\,q\to p,\,\lambda\to\lambda'\Big]\Bigg\}. \tag{55}$$

The terms here with the triple denominators evidently obey Salam's power counting criteria for both of the subintegrations (at least four powers underneath, depending on how the constants λ and λ' are assigned), and also for the final integration (six powers underneath), and therefore they converge. The divergences in the final integration are thus isolated in the simpler terms with the double denominators. Let us take first the case where λ and λ' are both zero. Discarding all the masses, and invoking (37), we obtain a high momentum structure ($p^2\to\infty$, $q^2\to\infty$, $(p+q)^2\to\infty$) comprising various inverse fourth powers

$$\frac{1}{p^2q^2}-\frac{\frac{1}{2}}{q^2p^2}-\frac{\frac{1}{2}}{q^2(p+q)^2}+\frac{\frac{1}{2}}{(p+q)^2p^2}-\frac{\frac{1}{2}}{p^2q^2}-\frac{\frac{1}{2}}{p^2(q+p)^2}+\frac{\frac{1}{2}}{(q+p)^2q^2}. \tag{56}$$

It will be seen that these leading powers cancel completely, term against term. We are thus left with six powers underneath, so that Salam's convergence criterion is fulfilled. A similar cancellation occurs if we choose λ and λ' to be both equal to unity. The two cases differ from each other only in whether the boson momentum flows with ($\lambda=\lambda'=0$) or against ($\lambda=\lambda'=1$) the fermion loop momentum in the overlap. The respective integrands are thus related by the overall change of variable $p\to -p$. Since the integrals are convergent, this implies equality.

Suppose now that λ and λ' are given values other than those just suggested. The integrand then depends on p and q in many different combinations, and a cancellation of the leading powers term against term no longer occurs. The behaviour of the integral can however be investigated directly by rotating the p^0 and q^0 contours through $\frac{1}{2}\pi$ and looking at the asymptotic behaviour of the integrand along various radius vectors of Euclidean 4-space. By an analysis too tedious to reproduce here we find in this way that the integral diverges logarithmically for all assignments of λ and λ' other than

$$\lambda=\lambda'=0,\qquad \lambda=\lambda'=1. \tag{57}$$

What happens is that the Euclidean integrand dies away only like its individual terms, assignments (57) excepted. These special and mutually equivalent assignments produce a Euclidean integrand which dies away in all Euclidean directions faster than the individual terms within it, and this is here sufficiently to ensure convergence. In three dimensions the integral is still divergent, but only logarithmically. Without the cancellations it would diverge quadratically.

We can now calculate the contribution (55) of Fig. 5 to the vacuum stress. Again it is helpful to use cancelling subtractions, as in the derivation of (21). The cancellation of overlapping logarithmic divergences under (37) leaves behind an interesting doubly logarithmic structure. We get

$$\langle \delta\Theta_{00}\rangle_0 = (1/16\pi^2)\{3\sum_{\alpha\beta} g_{\alpha\alpha\beta\beta} \ln m_\alpha{}^2 \ln m_\beta{}^2$$

$$-\sum_{\alpha ab} (g_{\alpha ab})^2 (2\ln m_a{}^2 \ln m_\alpha{}^2 - \ln m_a{}^2 \ln m_b{}^2)$$

$$+\sum_{\alpha ab} (g_{\alpha ab})^2 [(m_a+m_b)^2 - m_\alpha{}^2] I(m_\alpha{}^2, m_a{}^2, m_b{}^2)\}. \tag{58}$$

Here I is a convergent integral, which arises from the triple denominator in (55). It is defined by

$$I(m^2, m'^2, m''^2) \equiv \frac{1}{\pi^2}\iint \frac{d^2p\, d^2q}{(p^2-m^2)[(p+q)^2-m'^2](q^2-m''^2)}$$

$$= \int_0^1 d\xi \int_0^{1-\xi} d\eta [m^2\xi\eta + m'^2\xi(1-\xi-\eta) + m''^2\eta(1-\xi-\eta)]^{-1}. \tag{59}$$

The finite integral I is analytic in the three mass-squared variables, and it is fully symmetric under permutations thereof. It can be evaluated in terms of Spence functions. Its permutation symmetry properties link up with those of the Spence function in a most fascinating and unusual way, but since this has nothing to do with our main theme we relegate further consideration of its properties to the Appendix.

5.5. Subprocesses and cancellations

We have now shown that the two-dimensional theory is finite and free of all ambiguity, since in those places where the sharing factors λ have any influence the convergence requirements fix their values. The theory is thus fully predictive. We may ask nevertheless whether our λ assignments and other momentum assignments (Fig. 6) can be supported by arguments based on general mathematical structure, rather than upon the purely pragmatic need to cancel infinities. We shall now indicate and suggest that the required alternative arguments will be found by considering vacuum subprocesses.

Take first the λ problem. Our findings make it seem probable that $\lambda = 0$ and $\lambda = 1$ will be the correct sharing factors to use for any momentum arriving at a vertex in a graph with overlapping loops. We can now argue uniquely to this end by demanding that each of the subintegrations within a multiloop Feynman integral should already produce a Feynman integral as subintegral. No other λ choice shows this desirable property. We remark that in his beautiful and powerful general investigations Salam [28] took this for granted, since he did not consider any other sharings.

The further restriction of the λ choices in Fig. 6 to the correlated pairs (57), as opposed to such pairs as $\lambda = 0$, $\lambda' = 1$, may now be regarded as a non-trivial extension of the principle of cancellation at fixed momentum, and as such can be justified by the locality argument mentioned in Section 1. For if $\lambda = \lambda' = 0$ then the only momenta involved in Fig. 6 are p, q and $p+q$, and similarly, if $\lambda = \lambda' = 1$, then the only momenta involved are p, q and $p-q$, whereas the forbidden combination $\lambda = 0$, $\lambda' = 1$ would give p, q and $p \pm q$ all together.

Take next the assignments of p and q in Fig. 6. We did not use any assignment with $p+q$ on the boson line of figures 6(*ii*) and 6(*iii*), because this would have produced an uncancellable divergence. Our action at this point was again in accord with the general rules laid down by Salam [28] – an assignment with $p+q$ on the boson line would be "incorrect" because neither subintegration would "display" the divergence of the fermion loop. But let us see whether we can find a less pragmatic argument. Consider Fig. 6(*ii*) with $\lambda = 0$. We see that the p-subintegration produces the self energy integral of a boson of momentum q, while the q-subintegration produces the self energy integral for a fermion of momentum p. This is fair to fermions and bosons. If however we would put $p+q$ on the boson line then both of the subintegrations would produce fermion self energy parts, and this would be unfair to the zero-point bosons. Another way to see this comes by looking at the s- and u-channel scattering subprocesses mentioned at the beginning of this subsection, from which the self energies of the zero-point particles arise. The Born diagrams for forward fermion–fermion scattering by s- and u-channel boson exchange would put p and q onto the fermions and $p \pm q$ onto the boson, and thus give the momentum assignments we have not used. But when we close together the external fermion lines and integrate on d^2p and d^2q these diagrams mutually cancel, since there is a relative Fermi sign between them. The two remaining scattering diagrams relevant to Figs. 6(*ii*) and 6(*iii*) are those for forward fermion–boson scattering by fermion exchange in the s- and u--channels, and these both put $p \pm q$ onto the exchanged fermion, again in magical agreement with the momentum assignments which we have actually used. We thus begin to suspect an amazing correspondence between our purely pragmatic convergence prescriptions and some tentative new concept of Feynman integral as a sum over all possible contributions of the zero-point particles. It does not seem likely that such delicately contrived correspondences could arise by chance. Rather, we surmise that we are dealing with a genuinely finite and properly structured mathematical system, as indeed the closed nature of the formula (35) might also suggest. A problem for the future will be to learn how to reexpress this mathematical system in fully finite and comprehensive terms, instead of assembling it from mutually cancelling bits and pieces.

Meanwhile, since these vacuum subprocess concepts are as yet only tentative and unsubstantial, let us see whether we can rigorize our finite theory in some better way, namely by exhibiting it as the end point of a limiting sequence of theories of function--valued operator fields, as in the non-interacting case. To this end introduce again the cut--off (1), and apply it inside every Dirac and Klein–Gordon field factor in the Lagrangian (28) before varying the action. Then the tadpoles and self energies do indeed converge to the required relativistic and finite limits as $K \to \infty$, but the vacuum bubbles diverge. This divergence arises because the integration domains in the overlap graphs of Figs. 6(*ii*) and 6(*iii*) are regulated by factors of the type $\theta(K-|\boldsymbol{p}|)\ \theta(K-|\boldsymbol{q}|)\theta(K-|\boldsymbol{p}+\boldsymbol{q}|)$ while that of the compensating graph of Fig. 6(*i*) is regulated only by $\theta(K-|\boldsymbol{p}|)\theta(K-|\boldsymbol{q}|)$. But this little snag is readily dealt with (consistently with all positivity requirements — see next subsection) by cutting off some constant fraction of the quartic interaction term at a suitably chosen constant fraction of the principal cut-off momentum K. For the mismatch is only singly logarithmic, while the compensating terms are doubly logarithmic in K. Our relativistic theory is therefore attainable as the end point of a completely acceptable limiting sequence of good quantum theories.

5.6. Energy positivity

Since the system (28) is perturbatively finite, it should also be possible to establish non-perturbative existence theorems for it, subject to certain energy positivity conditions. The standard positivity requirement comprises two simultaneous statements about the bosonic coupling constants, which dominate the mass terms in the energy density when the boson fields are sufficiently large. These statements are

$$(i) \qquad \sum_{\alpha\beta\gamma\delta} g_{\alpha\beta\gamma\delta}\phi_\alpha\phi_\beta\phi_\gamma\phi_\delta \geqslant 0, \quad \forall\, \phi, \qquad (60)$$

$$(ii) \qquad \sum_{\alpha\beta\gamma\delta} g_{\alpha\beta\gamma\delta}\phi_\alpha\phi_\beta\phi_\gamma\phi_\delta > 0 \quad \text{whenever} \quad \sum_{\alpha\beta\gamma} g_{\alpha\beta\gamma}\phi_\alpha\phi_\beta\phi_\gamma \neq 0. \qquad (61)$$

These are the only conditions required for full consistency, apart from the convergence conditions (36) and (37). It might be thought, mistakenly, that one would also require positivity of the forced boson and fermion mass-squared matrices in (35). For we derived our convergence conditions by representing the forced fields as a relativistic set of harmonic oscillators, and considering the zero point energies of these oscillators. This procedure would be invalid and meaningless if it involved imaginary oscillator frequencies. It is important therefore to realise that the convergence conditions are in principle concerned only with the oscillators with very large $\boldsymbol{p}$. For these the relevant positivity condition is that $\boldsymbol{p}^2\delta_{\alpha\beta}+\mathcal{M}^2_{\alpha\beta}$ should be positive, which is always the case if $\boldsymbol{p}$ is large enough, since $\mathcal{M}^2$ (Eqs. (33) and (34)) is independent of $\boldsymbol{p}$. The ultraviolet finiteness of the theory is thus unaffected by the signature properties of the $\mathcal{M}^2$ matrices.

This is not to say that $\mathcal{M}^2$-signature is irrelevant in other respects. We would expect that an indefinite signature for some field configurations would make itself felt in some non-perturbative aspects. In this connection we notice that the fermionic mass squared matrix $(\mathcal{M}_\mathrm{F})^2$ is the square of the symmetric real matrix (34), and as such is always non-negative. So it is the bosonic mass squared matrix (33) that we should think about.

We must now establish that the convergence conditions (36) and (37) and the positivity conditions (60) and (61) can be simultaneously satisfied. The general situation can be illuminated with the help of two lemmas.

Lemma 1. If the fermions do not interact, then nothing interacts.

Proof. If there are no fermion couplings then equation (37) gives $\sum_\gamma g_{\alpha\beta\gamma\gamma} = 0$, which is equivalent to saying that $\Delta Q = 0$, where Q denotes the quartic polynomial $\sum_{\alpha\beta\gamma\delta} g_{\alpha\beta\gamma\delta}\phi_\alpha\phi_\beta\phi_\gamma\phi_\delta$

and Δ denotes the Laplacian in ϕ-space. But Q is zero at the point $\phi = 0$. Therefore, applying at this point the mean value theorem [32] of the multidimensional Laplace equation, we see that Q will be negative in some regions of ϕ-space unless it is zero everywhere. The first positivity condition (60) thus demands that $Q \equiv 0$. The second positivity condition (61) then kills all the trilinear boson couplings.

Lemma 2. If the fermions do interact, then quartic boson couplings must be present.

Proof. Set $\alpha = \beta$ and sum over β in the convergence condition (37). The second term becomes a sum of squares, some of which are non-zero. Therefore the coupling coefficients $g_{\beta\beta\gamma\gamma}$ cannot all vanish.

These two lemmas highlight the most relevant feature of our theory, which is that positivity, unitarity and finiteness are simultaneously achieved by virtue of Fermi–Bose cancellations in the interactions.

The simplest solution of all uses just one boson field ϕ and one fermion field ψ with the coupling term

$$\mathscr{L}_{\text{int.}} = ig\varphi\hat{\psi}\psi - \tfrac{2}{3}\, gm_{\text{F}}\phi^3 - \tfrac{1}{3}\, g^2\phi^4. \tag{62}$$

Here g is a coupling constant and m_{F} is the bare mass of the fermion, as defined by its role in the free part of the Lagrangian (28). In principle m_{F} can take either sign, but this avails nothing in this particular case, since we have the mass reversal transformation $\psi \to \sigma_3\psi$, $\phi \to -\phi$, and have only one fermion mass to consider. The interaction (62) satisfies the two cancellation conditions (36) and (37), but we should also take account of the cancellation condition (12) for the unperturbed part of the vacuum stress. In the present example the condition (12) fixes $m^2{}_{\text{B}} = (m_{\text{F}})^2$, as in the supersymmetric theories (see next section), but obviously this is merely because we have only one field of each type. Mass degeneracy is in no way a general feature of our approach.

Finally, we notice that with (62) we already obey not only the necessary positivity conditions (60) and (61), but also the optional positivity condition mentioned earlier in this section, to the effect that (33) should be non-negative for all f. The viability of the two-dimensional theory is thus fully established.

6. *Discussion*

6.1. Global supersymmetry as a special case

In 1974 Wess and Zumino [33] introduced an elegant new geometrical transformation concept, whereby the conformal Lie algebra of four-dimensional space-time is embedded as the even part of a graded Lie algebra, or Lie superalgebra. The supergroup idea was significantly developed by Salam and Strathdee [34], and has since been extended in many ways, particularly in connection with various concepts of local or x-dependent supersymmetry [35] and spontaneous symmetry breakdown by a superfield Higgs mechanism [36, 37]. Our interest here, however, lies with the earlier globally supersymmetric four-dimensional quantum field theory of Wess and Zumino [38], which is invariant only under the x-independent part of the original Lie supergroup. This theory is a special mass--degenerate case of the general class of theories presented here. It is similar to our example (62) except that it contains two bosons (as is appropriate to match the doubling of Majorana spin states at $D = 4$), and that one of the bosons is pseudoscalar (a possibility omitted herein, but purely for the sake of simplicity). It uses just one irreducible supermultiplet, and thus depends on one mass, one coupling constant, and one Z-factor. Correspondingly, the use of pure supersymmetry arguments clearly suggests that three mutually distinct and infinite renormalizations will be required [39]. But it was already noticed in Ref. [38] that the mass and coupling constant renormalization infinities gratuitously cancel out at the one-loop level, and in Ref. [39] it was shown that this amazing extra cancellation occurs in all orders. The present work now provides a satisfactory explanation for these extra cancellations, for it may be verified that the Lagrangian given in Ref. [38] obeys our universal mass-squared convergence condition (35). At the same time, the fact that the extra cancellations miraculously persist in all orders augurs well for the eventual extension of the present non-supersymmetric ideas beyond the limitations of superenormalizable systems and on into the multiloop integrals of four-dimensional field theories.

Responses to the mass degeneracy problem in supersymmetry have followed three main channels: (*i*) to have a non-positive boson mass-squared matrix (33), and in this way to obtain spontaneous non-Higgsian supersymmetry breaking [40, 41], (*ii*) to use a Higgs mechanism [35, 36, 37] and (*iii*) to insert supersymmetry-breaking mass terms by hand [39]. The last approach fits in with the simple ideas of the present paper, and the established fact that it can be done without upsetting the cancellations [39] again clearly suggests that for finiteness supersymmetry itself is not the main issue, and that we should look forward to finite non-supersymmetric theories in four-dimensional Minkowski space. It is also extremely interesting to note that when supersymmetry is broken without upsetting the cancellations then the mass relation (12) emerges [42, 43]! We must mention finally some recent work of Nath and Arnowitt [44, 45]. This work demonstrates the existence of a class of fully finite supersymmetric local gauge theories in four dimensions. This is very encouraging too, for although the basic concepts of such theories do not include that of a rigid Minkowski-space background, the relevant renormalization problems are of the same general nature as those encountered in the latter more familiar context.

6.2. The essential role of charged vector particles

The notion that interaction divergences might cancel between several fields is by no means a new one. Already in 1945 Pais [46] observed that the electron self energy could be rendered finite in order e^2 by a scalar–vector cancellation, and the same idea occurred to Sakata and his collaborators [47]. Shortly afterwards various authors [48–51] discovered that the notorious gauge-violating quadratically divergent photon self energy could be killed by a Fermi–Bose cancellation.

It so happened that this first Fermi–Bose cancellation used our convergence conditions (11) and (12), which were however inferred through the assumption that the various Bose and Fermi fields contributing to the photon self energy integral all carried the same charge e. There was thus no logical connection with the zero-point stress or the theoretical considerations of Sections 1 and 4. Nevertheless, it was promptly noticed [52] that the newly discovered cancellation conditions would indeed also serve to remove the ever--vexatious zero-point energy. The connection in principle between the shift of the zero--point energy of the vacuum under a static external potential and the divergent photon self energy integral was also recognised at this early stage [52]. Thus it needed only a correct understanding of fermion zero-point phenomena, and one would have had all that was necessary to suggest the convergence condition (35). But most thinking about fermion zero-point effects at that time was still guided by Dirac's asymmetric picture of empty positive energy states and a filled negative energy sea. This familiar picture gives an energetically incorrect and unphysical response to external potentials, since all the charges present have the same sign (cf. Section 2; notice also that the *unperturbed* zero-point energy is the same on either system of reckoning!). Unfortunately the widespread discrepancy occasioned on this account was not identified as such, and therefore the requisite confluence of ideas did not and could not take place. The historical forces at play may be plainly discerned. With the photon self energy cancelled, attention had naturally switched to the logarithmic vacuum polarization divergence, and there it had been found that charged scalar particles and charged Dirac particles contribute with the same sign. This thwarted further implementation of the cancellation programme, for the inclusion of charged vector particles introduced a quadratic divergence into the vacuum polarization, which made the matter worse rather than better [49–51, 53]. This impasse, together with the simultaneous and highly successful developments in the renormalization method, effectively removed the impetus for further investigation of cancellation mechanisms. A 1950 review by Sakata and Umezawa [53] provides a comprehensive contemporary account of this fascinating chapter in the history of field theory.

In 1971 t'Hooft [54], working with non-Abelian gauge theories, demonstrated the possibility of renormalizable theories of charged vector particles. In the gauge theories there is another cancellation mechanism, of a type quite different to those discussed in this paper, by virtue of which the divergence in the vacuum polarization is only logarithmic, and no longer quadratic. Moreover, explicit graph evaluation [55] shows that the sign of this divergent logarithm is opposite to the sign pertaining in the scalar and Dirac cases. The problem is thus wide open once again, and seems to hold considerable promise.

The essential thing in the charged gauge theories, which distinguishes them from all earlier spin-1 theories, is that the sources of the electromagnetic field emit and absorb quanta of the charged vector field. It is also vitally important that the vector particles have the "natural" Landé factor $g = 2$, like the electron and muon. It is apparently these particular features of the coupling, rather than the gauge transformations as such, which impart acceptable high energy behaviour. That there is great virtue in having a coupling with $g = 2$ was in fact already discovered in 1940, by Corben and Schwinger [56], and was then rediscovered with the gauge theories. But in their source aspect the gauge theories are entirely novel, and indeed revolutionary.

The sign of the logarithmically infinite term in the gauge theory vacuum polarization can be understood in a simple semiclassical way, by using the concepts of vacuum particle explained in Section 2. One knows from scattering theory that a positive charge inserted into the vacuum will push away positive vacuum particles and attract negative ones towards itself. Therefore it is readily understood that a negative polarization charge is induced in the scalar vacuum. Now if this were the only effect involved, the Dirac vacuum would obviously respond by producing a positive polarization charge, twice as large. The opposite sign would come because the Dirac vacuum particles are present with negative weight (Section 2), and the factor two would come by counting the spin states. There is however a second effect — the Dirac particles carry magnetic moments. There are thus two Dirac contributions, corresponding to the presence of both convection currents and spin currents. They can be unambiguously separated, and it can be shown that the spin currents yield a divergent vacuum polarization with the same sign as that for a scalar field (namely negative), and larger by a factor of six. The net effect is thus to give the known [49, 53, 57] factor of plus four in the overall comparison. Now the vector particles have $g = 2$, like the Dirac particles, and they behave well at high energy, and therefore their contribution will also be dominated by the magnetic term. And, obviously, this magnetic term must be positive, just because it is negative for the fermion case. Thus a positive charge inserted into the spin-1 vacuum induces around itself a logarithmically divergent and positive polarization charge (antidielectric property of the vector vacuum).

6.3. Conclusions

A new way of looking at the relativistic vacuum has been described in this paper. Its effectiveness has been demonstrated in various ways, most particularly in that it has led to the discovery of a large class of non-supersymmetric fully finite relativistic quantum field theories in two-dimensional Minkowski space. These theories require further investigation with regard to non-perturbative aspects, alternative formulations, Euclidean extensions, pseudo-Riemannian generalizations, and so forth. The extension to three dimensions must also be studied.

The ultimate aim must be to see what can be done in four dimensions. The contributions of pseudoscalar and vector fields will be needed, and one also expects that the mathematical concept of an anticommutative spinor force will play a decisive part. Convergence conditions involving fourth powers of the mass matrices will come into play, generalizing (13) as (35) generalizes (12). These conditions will produce cancellations of divergences in the four-point function, etc. Also, with the loss of superenormalizability, it will become appropriate to study the possibility of cancellations to all orders. Improved techniques will therefore be required, and of course there is no present guarantee of success. But the physical arguments of Section 2, the formal arguments of subsection 5.5, and the four-dimensional discoveries discussed in subsections 6.1 and 6.2 certainly justify considerable optimism.

I wish to thank Dr. C. J. Hamer, Professor C. Michael and other colleagues for stimulation, timely information, and helpful advice. I wish also to specially acknowledge Professor B. K. P. Scaife, who enthusiastically described the mysteries of the Casimir effect and its experimental verification to me several years ago, and thus prompted the first slow development of the present train of thought. For all that I am now able to tell, my comprehension of the vacuum might otherwise never have reached the zero-point level!

APPENDIX

In this appendix we first study the kinematic significance of the functions $B_\pm$ of equations (49), and thereby we gain further understanding of ultraviolet processes. We then turn to some closely related matters concerning the evaluation of the convergent two-loop vacuum integral (59) and the symmetries of the Spence function.

The functions $B_\pm(m^2, m'^2 : m''^2)$ depend upon three mass-squared variables, which may be permuted in six ways, giving twelve functions in all. They are however even with respect to exchange of the first two arguments, so that we deal with only six values. Suppose now that one of the three masses $|m|$, $|m'|$, $|m''|$ exceeds the sum of the other two. Some on-shell three-prong vertex can then arise as a real physical process, and the six values are all real. Otherwise they are all complex. Let $v_{mm'}$ denote the proper relative speed between m and m' in the said real process. Two distinct cases have to be considered, depending upon whether m and m' appear one in the final state and one in the initial state, or both in the same state. If m'' is not the heaviest particle then m and m' appear in opposite states. We then find that

$$\frac{1}{2|mm'|} B_\pm(m^2, m'^2 : m''^2) = \left(\frac{1 \pm v_{mm'}}{1 \mp v_{mm'}}\right)^{1/2}, \quad m'' \text{ not the heaviest.} \tag{63}$$

On the right hand side of this equation we recognize respectively the familiar blue and red Doppler factors, by which the energy of a photon would be boosted under a collinear Lorentz transformation with speed $v_{mm'}$. In the other case, where the heaviest particle is m'', the particles m and m' appear in the same state, and we then have respectively red and blue Doppler boost factors, prefixed by a sign:

$$\frac{1}{2|mm'|} B_\pm(m^2, m'^2 : m''^2) = -\left(\frac{1 \mp v_{mm'}}{1 \pm v_{mm'}}\right)^{1/2}, \quad m'' \text{ the heaviest.} \tag{64}$$

The physical meaning of the functions $B_\pm$ is thus rendered apparent. By the same token we infer several symmetry properties, which may be checked directly:

$$B_\pm(m^2, m'^2 : m''^2) = B_\pm(m'^2, m^2 : m''^2), \tag{65}$$

$$B_+(m^2, m'^2 : m''^2)B_-(m^2, m'^2 : m''^2) = 4m^2m'^2, \tag{66}$$

$$B_\pm(m^2, m'^2 : m''^2)B_\pm(m'^2, m''^2 : m^2)B_\pm(m''^2, m^2 : m'^2) = -8m^2m'^2m''^2. \tag{67}$$

Furthermore, there is a symmetry property which expresses the conservation laws of energy and momentum:

$$B_\pm(m^2, m'^2 : m''^2)+B_\mp(m^2, m''^2 : m'^2) = 2m^2. \tag{68}$$

(For example, we can substitute (63) here if m is the heaviest, in which case we find that the two equations (68) express energy and momentum conservation in the processes $m \leftrightarrow m$

$+m''$). The six $B_\pm$ also satisfy respectively three quadratic equations, of which that for $B_\pm(m^2, m'^2 : m''^2)$ is

$$B_\pm^2(m^2, m'^2 : m''^2)+2(m''^2-m^2-m'^2)B_\pm(m^2, m'^2 : m''^2)+4m^2m'^2 = 0. \qquad (69)$$

The pretty thing about the $B_\pm$ is that they and the related formulae (65)–(69) express the kinematic properties of the on-shell three-prong vertex in terms of mass-squared variables, thus harmonizing with the substitution law [31] and the analytic properties of quantum field theory. It seems only natural therefore that they should appear in the integrated form (48) of the self energy integral (46). Their presence there reveals the influence of processes in which a finite energy vacuum fermion scattered at one point in space-time acts as source for a boson field at another and different point (we use the vacuum particle scattering concepts of Section 2). For if the space-time separation between the two vertices were zero then the velocity concepts embodied in B could not possibly have relevance. On the other hand the simpler B-independent logarithmic terms in (48) contain no velocity parameters, and produce a space-time δ-function. These sharply localized terms are precisely those left over from the ultraviolet cancellation, and they owe their localization to the contributions of vacuum fermions travelling with infinitely short wavelengths. The ultrarelativistic vacuum fermions effectively maintain the limiting speed under collision with the boson fields in Fig. 3, and relative speeds become dynamically irrelevant. The infinite coefficient which marks their contribution in ordinary quantum field systems is rendered finite by cancellation in the present theory, but the characteristic strict localization persists, clearly indicating an ultrarelativistic origin.

The variables B emerge in Feynman integration during the process of taking partial fractions. However, the quadratic equation there involved is

$$m^2x_\pm^2+(m''^2-m^2-m'^2)x_\pm+m'^2 = 0, \qquad (70)$$

which is unsymmetrical under every mass permutation, unlike (69). Thus in practice we find ourselves dealing with twelve distinct values of x, rather than six B. The permutation pattern among these twelve entangles with the sign of the square root. We may conveniently express it by defining

$$\begin{aligned} x_\pm &\equiv (1/2m^2)B_\pm(m^2, m'^2 : m''^2), & X_\pm &\equiv (1/2m'^2)B_\mp(m'^2, m^2 : m''^2), \\ y_\pm &\equiv (1/2m'^2)B_\pm(m'^2, m''^2 : m^2), & Y_\pm &\equiv (1/2m''^2)B_\mp(m''^2, m'^2 : m^2), \\ z_\pm &\equiv (1/2m''^2)B_\pm(m''^2, m^2 : m'^2), & Z_\pm &\equiv (1/2m^2)B_\mp(m^2, m''^2 : m'^2). \end{aligned} \qquad (71)$$

The symmetry relation (65) and its permuted variants then pass over to

$$m^2x_\pm = m'^2X_\mp, \text{ etc.} \qquad (72)$$

The remaining symmetry relations (66)–(68) split into two mass-independent sets, depending on whether the upper or the lower sign is taken:

$$\begin{aligned} X_\pm &= 1/x_\pm, \quad Y_\pm = x_\pm/(x_\pm-1), \quad Z_\pm = 1-x_\pm, \\ y_\pm &= (x_\pm-1)/x_\pm, \quad z_\pm = 1/(1-x_\pm). \end{aligned} \qquad (73)$$

Including also the identity transformation $x_\pm = x_\pm$, we perceive in (73) a group of six non-linear transformations on the (real or complex) one-dimensional space of points x. This point transformation group is a faithful realization of the permutation group on three symbols. It is also the point group which arises in the theory of the Spence function [58]. The two groups come together and identify with each other in the mathematics of the permutation-invariant integral I of equation (59), to which we now direct our attention.

The curious thing about I is that although it is perfectly symmetrical between the three masses, there is no way to reduce it further except by traversing the integration triangle in an unsymmetrical way. This hides the permutation symmetry.

It is easiest to integrate along rays drawn through one corner of the triangle. Choosing the corner at the origin we substitute

$$\xi = u(1-x), \qquad \eta = ux \tag{74}$$

and so obtain the unsymmetrical looking one-dimensional form

$$I = \int_0^1 dx \, \frac{\ln\left[m''^2 x + m'^2(1-x)\right] - \ln\left[m^2 x(1-x)\right]}{m''^2 x + m'^2(1-x) - m^2 x(1-x)} \, . \tag{75}$$

Using (70) and splitting the denominator into partial fractions we further reduce this to the form

$$I = \int_0^1 dx \, \frac{\ln\left[x(1-x_+-x_-)+x_+x_-\right] - \ln x - \ln(1-x)}{m^2(x_+-x_-)} \left\{\frac{1}{x-x_+} - \frac{1}{x-x_-}\right\}, \tag{76}$$

which exploits the Doppler boost factor $B_\pm$ again, through (71).

We now bring in the complex-valued Spence function $f(z)$ defined by Mitchell [58]. Mitchell's integral representation

$$f_{\mathrm{P}}(z) = z \int_1^\infty \frac{\ln t\,dt}{t(t-z)} \tag{77}$$

displays its analytic properties in the principal sheet of the cut plane of z, and suggests the improved integration formula

$$\int \frac{\ln(az+b)}{cz+d}\,dz = \frac{1}{c} f_{\mathrm{P}}\left(\frac{c(az+b)}{bc-ad}\right) + \frac{1}{c}\ln(az+b)\ln_{\mathrm{P}}\left(1-\frac{c(az+b)}{bc-ad}\right), \tag{78}$$

wherein both the Spence function f_{P} and the last logarithm $\ln_{\mathrm{P}}$ are to be taken on their principal sheet. That is, $\ln_{\mathrm{P}}(1-z)$ is real when $1-z$ is positive, and has a cut for $1-z$ negative, or

$$\ln_{\mathrm{P}}(1-z) = -z \int_1^\infty \frac{dt}{t(t-z)} \, . \tag{79}$$

The improved integration formula (78) is valid for arbitrary complex parameters a, b, c, d, z (subject only to not crossing the cut during the integration). The integration formula given by Mitchell can also be extended to the complex domain, but is then only valid subject to careful choice of Riemann sheets in its various logarithms. The difficulty with the Mitchell formula is indeed that it depends upon the property $\ln(z_1 z_2) = \ln z_1 + \ln z_2$, which is not always true when one deals with complex logarithms on their principal sheets.

The points $x = x_\pm$ lie within the integration interval if $|m| > |m'| + |m''|$. However the integrand in (75) is perfectly regular at these points, and so the situation can be brought under control by giving small imaginary parts to x_+ and x_- and proceeding to the limit. On this understanding the integral I reduces in all cases to

$$I = \frac{1}{m^2(x_- - x_+)}\left\{\left[f_P\left(\frac{1}{x_+}\right) + \tfrac{1}{4}\ln_P{}^2\left(\frac{x_- - 1}{x_+}\right) + \ln(x_- x_+)\ln_P\left(\frac{1 - x_- - x_+}{1 - x_+}\right)\right]\right.$$
$$\left. -[x \leftrightarrow 1-x] - [+ \leftrightarrow -] + [x \leftrightarrow 1-x,\ + \leftrightarrow -]\right\}. \qquad (80)$$

In this formula four Spence functions f_P appear, four others present after the initial use of (78) having collapsed into $\ln_P{}^2$ terms through application of one of Mitchell's symmetry relations [58]. Further application of the Mitchell relations can reduce the total number of Spence functions to two.

The permutation symmetry of I is completely hidden in formula (80), and cannot be rendered explicit except by using three times as many terms. But it is nevertheless there, and can be shown to be there by using the transformations (73) with the Mitchell relations.

Note added in proof

The discussion of section 6.2 regarding the sign of the ultraviolet logarithm in the polarization current of the vector vacuum may be rendered fully quantitative by using some recent results of Hughes [59]. Hughes has calculated the relevant Feynman graphs of massless Corben-Schwinger electrodynamics, and has compared them with the analogous graphs of Dirac electrodynamics. He finds that the spin contributions to the logarithm stand in the ratio $-(s'/s)^2$, where $s' = 1$ and $s = \frac{1}{2}$. That the ratio is this and not $-s'(s'+1)/s(s+1)$ may be understood by reference to the helicity structure of a massless theory. However, the same ratio applies even in the massive case, because with $g = 2$ the spin moment effectively disappears from the dynamics of the longitudinal modes in the extreme relativistic limit. The convective contributions of the longitudinal modes persist at high energies and therefore we expect that the total coefficients of the logarithm stand in the ratios —21/4 (massive case) and —22/4 (massless case).

I wish to thank Dr. Hughes for informing me of this work.

REFERENCES

[1] S. S. Schweber, *An Introduction to Relativistic Quantum Field Theory*, Row, Peterson and Co., New York 1961, p. 185.
[2] G. Wentzel, *Quantum Theory of Fields* (trans. C. Houtermans and J. M. Jauch), Interscience Publ., Inc., New York 1949, p. 35.
[3] J. Schwinger, L. L. De Raad, K. A. Milton, *Ann. Phys.* (N.Y.) **115**, 1 (1978).

[4] H. B. G. Casimir, *Proc. K. Ned. Akad. Wet.* **51**, 793 (1948).
[5] L. D. Landau, E. M. Lifshitz, *Electrodynamics of Continuous Media* (trans. J. B. Sykes and J. S. Bell), Pergamon Press, London 1960, p. 368.
[6] T. H. Boyer, *Ann. Phys.* (N.Y.) **56**, 474 (1970).
[7] B. K. P. Scaife, T. Ambrose, *Proc. R. Ir. Acad.* **72A**, 155 (1972).
[8] L. S. Brown, *Phys. Rev.* **D15**, 1469 (1977).
[9] L. S. Brown, J. P. Cassidy, *Phys. Rev.* **D16**, 1712 (1977).
[10] P. C. W. Davies, S. A. Fulling, S. M. Christensen, T. S. Bunch, *Ann. Phys.* (N.Y.) **109**, 108 (1977).
[11] T. S. Bunch, *J. Phys. A* **11**, 603 (1978); **12**, 517 (1979).
[12] T. S. Bunch, L. Parker, *Phys. Rev.* **D20**, 2499 (1979).
[13] G. Temple, *Proc. R. Soc.* **A228**, 175 (1955).
[14] M. J. Lighthill, *Introduction to Fourier Analysis and Generalized Functions*, Cambridge University Press, London 1955.
[15] G. R. Allcock, *Ann. Phys.* (N.Y.) **53**, 253 (1969), esp. sect. III.
[16] J.-P. Marchand, *Distributions — an Outline*, North Holland, Amsterdam 1962.
[17] I. M. Gel'fand, G. E. Shilov, *Generalized Functions* (trans. E. Saletan), Academic Press, New York 1964, Vol. I.
[18] J. Schwinger, *Phys. Rev.* **75**, 651 (1949).
[19] G. R. Allcock, *Contribution* to *Cooperative Phenomena*, ed. H. Haken, M. Wagner, Springer Verlag, Heidelberg 1973, p. 350.
[20] C. G. Callan, S. Coleman, R. Jackiw, *Ann. Phys.* (N.Y.) **59**, 42 (1970).
[21] W. Pauli, F. Villars, *Rev. Mod. Phys.* **21**, 434 (1949).
[22] R. P. Feynman, *Phys. Rev.* **76**, 769 (1949).
[23] W. Heisenberg, *Z. Phys.* **110**, 251 (1938).
[24] H. Umezawa, *Prog. Theor. Phys.* **7**, 551 (1952).
[25] J. Schwinger, *Quantum Electrodynamics*, Dover Publ., New York 1958, p. XVI.
[26] W. Pauli, *Nuovo Cimento* **10**, 648 (1953).
[27] F. J. Dyson, *Phys. Rev.* **75**, 1736 (1949).
[28] A. Salam, *Phys. Rev.* **82**, 217 (1951); *Phys. Rev.* **84**, 426 (1951).
[29] N. Nakanishi, *Prog. Theor. Phys.* **17**, 401 (1957).
[30] S. Weinberg, *Phys. Rev.* **118**, 838 (1960).
[31] J. M. Jauch, F. Rohrlich, *The Theory of Photons and Electrons*, Addison-Wesley Publ. Co., Inc., New York 1955, Appendix A5-2 and p. 161.
[32] R. Courant, D. Hilbert, *Methoden der Mathematischen Physik*, Springer Verlag, Berlin 1937 Vol. II, pp. 249, 419.
[33] J. Wess, B. Zumino, *Nucl. Phys.* **B70**, 39 (1974).
[34] A. Salam, J. Strathdee, *Nucl. Phys.* **B76**, 477 (1974).
[35] P. Fayet, S. Ferrara, *Phys. Rep.* **32C**, 249 (1977).
[36] P. Fayet, J. Iliopoulos, *Phys. Lett.* **51B**, 461 (1974).
[37] S. Deser, B. Zumino, *Phys. Rev. Lett.* **38**, 1433 (1977).
[38] J. Wess, B. Zumino, *Phys. Lett.* **B49**, 52 (1974).
[39] J. Iliopoulos, B. Zumino, *Nucl. Phys.* **B76**, 310 (1974).
[40] L. O'Raifeartaigh, *Nucl. Phys.* **B96**, 331 (1975).
[41] A. Salam, J. Strathdee, *Phys. Lett.* **49B**, 465 (1974).
[42] S. Ferrara, L. Girardello, F. Palumbo, *Phys. Rev.* **20D**, 403 (1979).
[43] S. Ferrara, B. Zumino, *Phys. Lett.* **86B**, 279 (1979).
[44] Pran Nath, R. Arnowitt, *Phys. Rev. Lett.* **42**, 138 (1979).
[45] R. Arnowitt, Pran Nath, *Nucl. Phys.* **B161**, 321 (1979).
[46] A. Pais, *Phys. Rev.* **68**, 227 (1945).
[47] S. Sakata, O. Hara, *Prog. Theor. Phys.* **2**, 30 (1947).
[48] J. Rayski, *Acta Phys. Pol.* **9**, 129 (1948); *Phys. Rev.* **75**, 1961 (1949).

[49] D. Feldman, *Phys. Rev.* **76**, 1369 (1949).
[50] H. Umezawa, J. Yukawa, E. Yamada, *Prog. Theor. Phys.* **4**, 25, 113 (1949).
[51] H. Umezawa, R. Kawabe, *Prog. Theor. Phys.* **4**, 443 (1949).
[52] H. Umezawa, R. Kawabe, *Prog. Theor. Phys.* **5**, 153 (1950); also Ref. [51], p. 459.
[53] S. Sakata, H. Umezawa, *Prog. Theor. Phys.* **5**, 682 (1950).
[54] G. t'Hooft, *Nucl. Phys.* **B35**, 167 (1971).
[55] D. J. Gross, F. Wilczek, *Phys. Rev.* **D8**, 3633 (1973).
[56] H. C. Corben, J. Schwinger, *Phys. Rev.* **58**, 953 (1940).
[57] J. Schwinger, *Phys. Rev.* **75**, 651 (1949); **76**, 790 (1949); **82**, 664 (1951).
[58] K. Mitchell, *Phil. Mag.* **40**, 351 (1949).
[59] R. J. Hughes, Some Comments on Asymptotic Freedom, *Phys. Lett.* (to appear).