Is the Zero-Point Energy Real?

Simon Saunders
Faculty of Philosophy, 10 Merton St., Oxford.

To appear in *Ontological Aspects of Quantum Field Theory*, M. Kuhlmann, H. Lyre, and A. Wayne, (eds)., Singapore: World Scientific, 2002.

Abstract

I consider the arguments to show that the vacuum energy density should receive a large contribution from the zero-point energy. This is the cosmological constant problem, as it was originally framed.

1 Introduction

The nature of the vacuum state has proved to be of enduring interest in field theory. Time and again it has been refashioned. Evidently we need the right physical concepts, even to understand the state in which nothing exists.

When it comes to the vacuum of quantum field theory it is increasingly clear that we do not have the correct physical concepts. Of course, concerning the Planck scale, there are plenty of reasons to question the basic principles of field theory. But the problem I have in mind, whilst it does concern gravity, arises at all scales. It can be posed as a problem of elementary quantum theory. The difficulty is this: it appears that there must remain a very large energy in the vacuum state, and that this should contribute massively to the source terms of general relativity. It has traditionally been called the cosmological constant problem, since this source term is proportional to the metric tensor - yielding a value wildly inconsistent with observation. This problem has received a great deal of attention in recent years, but there is very little consensus as to how it may be solved.

Here I am concerned with the original statement of the problem, in terms of the zero-point energy of QFT. Is the zero-point energy real? The usual argument given for it is the Casimir effect; I shall consider this is detail in due course. But prior to that, I will attempt to gain some historical perspective on the problem. The vacuum of field theory has seen some radical changes, first, with the eliination of ether in classical electromagnetism, and second, with the elimination of the Dirac negative energy sea, in quantum electrodynamics. The latter is particularly instructive; the negative energy of the Dirac vacuum can be viewed as the fermion zero-point energy by another name.

Both examples are cases where the vacuum turned out not to have the problematic feature it was thought to have. The question arises as to whether a similar fate awaits the zero-point energy - whether in fact the cosmological constant problem (as traditionally formulated) is a spurious one. But if so, and if the previous historical examples are anything to go by, not just a change in philosophy is needed; there will be a change in physics as well. I will at the close suggest a change in philosophy, but not yet a change in the physics. My suggestion is doubly limited, in that it has no bearing on the other ways in which the vacuum state can pick up energy - in particular the energy shifts, possibly large, that are expected to arise on spontaneous symmetry breaking.

2 The Cosmological Constant Problem

A common statement of the problem is as follows:

> The cosmological constant problem is one of the most serious puzzles confronting particle physics and cosmology. No symmetries or principles of General Relativity prohibit a cosmological constant term from appearing in the Einstein equations. Moreover, any vacuum energy such as that predicted by quantum field theory must - according to the equivalence principle - gravitate, and will act as a cosmological constant. However, the upper bound on the present day cosmological constant is extremely small in particle physics units: $\frac{\lambda}{m^4_{Planck}} < 10^{-122}$ (Brandenburger 1999).

A slightly different formulation of the CCP is as a fine-tuning problem, with a bound that would be even smaller if we did not introduce a cut-off at the Planck scale (on the optimistic assumption that whatever physics comes into play in the Planck regime, it will not add to the vacuum energy). For a cutoff Λ, the zero-point energy of a field of mass $m \ll \Lambda$ is (with $h = c = 1$) :

$$\int_0^\Lambda \frac{1}{2}\sqrt{k^2+m^2}\frac{4\pi k^2 dk}{(2\pi)^3} \sim \frac{\Lambda^4}{16\pi^2}$$

This quantity is just the sum of the zero-point energy over the normal modes of the field up to the cut-off Λ. If this is set at the Planck mass, $\Lambda \sim m_{Planck} \sim 10^{19} GeV$, then given the current upper bound on the cosmological constant $\lambda < 10^{-29} g/cm^3 \sim (10^{-11} GeV)^4$, the observed value is more than 120 orders of magnitude smaller than we expect. If the contribution from the zero-point energy is to be cancelled by the true cosmological constant, the latter will have to be equal to it and of opposite sign to one part in 120 - making it the most accurately known number in physics.

The latter observation is often greeted with irritation. It seems odd to some that the value of a constant that need not even be introduced into Einstein's field equations - a move which Einstein himself later repudiated - should be a problem to physics. But evidently the problem arises whether or not there is a true cosmological constant term in the field equations. If there is no such constant, then something else must be found to cancel the vacuum energy.

Brandenberger goes on to suggest a mechanism whereby scalar gravitational fluctuations, with wavelength greater than the Hubble radius, are formed as a back-reaction to the presence of cosmological perturbations, which act as a negative cosmological constant in the de Sitter background. He suggests this mechanism may in fact be self-regulating, leading, more or less independent of the original value of the cosmological constant, to an effective value to it of order unity (on the Planck scale), which cancels the stress-energy tensor due to the zero-point energy.

This proposed solution is typical of the genre. Coleman's well-known proposal is the same (Coleman 1988): the cosmological constant becomes a dynamical variable in a certain Euclidean path-integral formulation of quantum gravity, whereby the amplitude is shown to be greatly peaked at the net value of zero. Cancellation of the cosmological constant, with the source term due to the zero-point energy, is the name of the game. On Coleman's proposal, wormholes, connecting geometries, make the Euclidean action very large for geometries with net non-zero cosmological constants. They therefore make vanishingly small contribution to the path integral.

More recent attempts have considered quintessence, anthropic, k-essence, braneworld, and holographic approaches. Indeed, the literature is by now enormous, and still growing rapidly. Very little of it consides the original motivation for the problem critically - whether because the arguments are so clear-cut as to be unanswerable, or because here is a problem worth taking seriously because there are lots of speculative things that can be said about it. And, it is worth saying again, it is at least a problem of well-accepted theory, which for many faces no obvious conceptual challenges:

> Physics thrives on crisis....Unfortunately, we have run short of crises lately. The 'standard model' of electroweak and strong interactions currently faces neither internal inconsistencies nor conflicts with experiment. It has plenty of loose ends; we know no reason why the quarks and leptons should have the masses they have, but then we know no reason why they should not.
>
> Perhaps it is for want of other crises to worry about that interest is increasinly centered on one veritable crisis: theoretical expectations for the cosmological constant exceed observational limits by some 120 orders of magnitude. (Weinberg 1989, p.1).

No doubt many would be disappointed if there turns out to be no good reason, after all, to take the zero-point energy seriously.

Here is the argument to show why the vacuum, if it contributes any energy at all, will yield an effective cosmological constant. By the Equivalence Principle, the local physics is Lorentz invariant, so in the absence of any local matter or radiation, we should see the same physics as in the vacuum state. But the vacuum expectation value of the stress-energy tensor must be a multiple of the Minkowski metric. Going to a freely falling frame, therefore, we expect to find a term $\lambda g^{\mu\nu}$ on the RHS of the Einstein field equations. Such a term characterizes a perfect fluid with equation of state:

$$P_{vac} = -\rho_{vac}. \tag{1}$$

Under an adiabatic expansion from V to $V+dV$, an amount of work PdV is done, which provides exactly the mass-energy to fill the new volume $V + dV$ with the same energy-density ρ_{vac}. Expanding or compressing nothing changes nothing, as one would expect. Since locally we expect the equation

$$R^{\mu\upsilon} - \frac{1}{2}g^{\mu\nu}R = \frac{8\pi G}{c^4} < T^{\mu\nu} >_{\omega} . \tag{2}$$

to hold for the observed metric and curvature (the c-number quantities on the LHS), then in a local vacuum state ω we expect the RHS to contribute a term $\lambda g^{\mu\nu}$ - and, indeed, if ω is anything like the Fock space vacuum, a formally divergent value for λ, given a zero-point energy $\frac{1}{2}h\nu$ for each normal mode of a quantum field.

Of course the notion of a local vacuum state is not easily defined in QFT. We cannot use number density operators to define it, for these are non-local quantities. But one would expect that the Fock space vacuum would give a *smaller* expectation value for the components of the stress-energy tensor, than any more realistic state. There is also good reason to suppose that there is no non-Fock vacuum in which the local stress-energy can be zero (we shall consider this shortly). And the physical picture of the vacuum as a fluctuating field of zero-point energy is an extremely general one, that surely does not depend on the presence of exact symmetries. The objection that the argument cannot be made mathematically rigorous is pedeantic. The zero-point energy is present in elementary quantum mechanics; it has widespread applications in semiclassical theories of electromagnetism; it is not going to go away because the vacuum can only be defined rigorously in the case of global symmetries.

That is not to say that the zero-point energy is in fact real, and has the huge value that it seems it must have. But it is widely assumed that it is real, and it is widely assumed that there must be a mechanism in play which tends to cancel it out. But the difficulties here are severe. It is a conspiracy that is needed. All of the fields in the standard model will contribute to it, so in terms of the Planck length, the cancellation will have to be fine-tuned across all these fields. It is true that fermion fields contribute with opposite sign to boson fields; were every fermion field accompanied by a bosonic partner, the cancellation would be exact. But supersymmetry, if it a symmetry at all, is a broken symmetry. The lower bound on the mass differences of fermions and their supersymmetric partners is of the order of few hundred GeV, so whilst it is true that for energy scales much larger than this, we would expect to find cancellation of the zero-point energies, we will still have a vacuum energy with cut-off $\Lambda \sim 100GeV$, contributing to the effective cosmological constant a term $\lambda \sim (100GeV)^4$ - much smaller than before, but still more than 50 orders of magnitude greater than that observed.

I have said that there have been plenty of specualtive solutions to the problem; I should add that some of them are more philosophical in outlook, particularly those based on the anthropic principle. But among these only those using the weak anthropic principle (WAP) appear to me to have any credence, where it is assumed that there exist many universes, or parts of one universe, in which the cosmological constant takes on different values.[1] In general, then, putting to one side the

[1]It is not clear that there is any real consensus on this matter. Weinberg, for example, makes use of what he calls the weak anthropic principle, but according to him in this "one explains which of the various possible eras of the universe we inhabit,

Everett interpretation of the state - wherupon the WAP can be applied in the context of standard quantum mechanics - these also stand in need of new physical principles. The question here is whether the search for such principles is well-motivated; of whether the zero-point fluctuations are real. I shall begin with previous pictures of vacuum which did turn out to be mistaken.

3 The Classical Ether

The history of the classical ether is familiar, so here I shall be brief. The classical models of ether gave rise to severe problems. It appeared to need contradictory mechanical properties, and, independent of those, it led to the prediction of observable effects that were not in fact detected.

Why was it thought to exist at all? The answer one often meets is that since light and electromagnetic forces propagate as waves, their must be a substratum which is in motion; something has to wave. But this answer is unedifying. It does not connect with the question. The point is to explain the belief in a *mechanical* ether, not in something or other which waves (the electromagnetic field is something or other which waves). A better response is that as a matter of historical fact, wave equations were first formulated as applications of Newtonian mechanical principles to a mechanical medium, with movable parts - and that it remained a problem to reconcile Newtonian principles with the subsequent modifications of those equations, as the theory of the luminiferous ether progressed. In extending this theory to electromagnetic phenomena, Maxwell's use of mechanical models, as a heuristic device, is well-known. And of course a great deal of late nineteenth century work on the structure of electromagnetic media was directed to the investigation of *material* dielectrics, again using Newtonian principles.

In Lorentz's work the treatment of material dielectrics was initially continuous with his treatment of ether, but the ether was progressively shorn of its mechanical properties. Its principal role, at the end, was to define the resting frame, to which all the electrodynamical equations were referred - and in terms of which the properties of moving dielectrics were analyzed. It is common to view the disappearance of the ether theory, following Einstein's intervention, as abrupt, but it would be more ac-

by calculating which eras or parts of the universe we could inhabit" (Weinberg 1989 p.7), leaving it ambiguous whether these different eras are merely possible or whether they are actual. In particular he appears to allow, both here and in his most recent use of the principle, that it is sufficient if the state is a superposition of components in each of which the cosmological constant has a different value. This would permit the use of the weak principle, as I understand it, only on the Everett or many-worlds interpretation.

curate to say that the ether was progressively whittled away. In Lorentz's hands it remained little more than a frame of reference, and a dwindling catalog of methods for reducing differential to difference equations. It was the grin of the Cheshire Cat.

Bell's defence of Lorentz's approach to special relativity illustrates how little has changed: Lorentz would have been at home with it. Bell defends the view that contraction and dilation effects should be understood as dynamical, rather than kinematic, effects; they reflect real and objective changes in the electromagnetic structure of measuring rods and clocks (referred to a fixed resting frame throughout (Bell 1989)). Lorentz would have been at home with it.

So when, precisely, did the ether disappear? Perhaps it was when McCullogh's ether, providing as it did constitutive equations for a continuum mechanical system, giving the right boundary conditions for Fresnel's ether theory; failed to be adopted; or when Maxwell, having derived the equations for the displacement current from his system of cogs and wheels, derived them instead (in the *Treatise*) abstractly, from a Lagrange function; or perhaps it was when Lorentz, struggling to derive the Fresnel ether drag coefficient from his study of moving dielectrics, exempted his molecules of ether from Newton's third law. Einstein's 1905 intervention, which removed not only the special status that Lorentz thought attached to the ether frame, but Newton's force laws as well, delivered the death blow; but it was the *coup de grâce*, not the *coup mortel*.

The slow realignment in our understanding of the vacuum of classical electromagnetism, spread over more than half a century, at the same time eroded the paradoxical properties that it was thought to require: stiff enough to sustain transverse vibrations, whilst offering no resistance to the motion of ordinary matter; able to support torsion, but with a velocity field with non-zero curl; yielding no drag coefficient for a moving dielectric, when any other pair of dielectrics in relative motion have non-zero coefficient. All of these problems were solved, or dissolved, in Lorentz's theory of ether; the only problem outstanding was how to account for the null result of the Michelson-Morely experiment, given the absence of ether drag. And that too, along Bell's lines, it could eventually do.

4 The Dirac Negative Energy Sea

A second and much closer precursor to the zero-point energy problem is the Dirac negative energy sea. This vacuum was explicitly introduced to solve the negative-energy difficultly. That in turn had plagued every previous attempt to unify quantum mechanics and special relativity, beginning with the relativistic scalar wave equation introduced by Schrödinger, Gordon, and Klein in 1926.

Dirac was dissatisfied with this equation. It admitted negative-energy solutions, but further, as a second-order equation, it appeared to him inconsistent with the basic structure of quantum mechanics. In particular the charge-current equation yielded a conserved quantity that was not positive definite, so one could not use it to define a positive-definite norm. Dirac concluded, rightly, that it could not itself be interpreted in terms of probability. In 1928 he found the Lorentz covariant first-order equation that bears his name, for spin-half particles, which did yield a positive-definite norm. With that he was in a position to define a Hilbert space. But the equation still admitted negative-energy solutions, and it was clear that the Klein paradox could be formulated here as in the scalar case. In general one could not simply exclude the negative-energy solutions; the obvious candidates for interaction terms all led to transitions from positive to negative energy states.

Dirac's remarkable solution was to suppose that in the vacuum state all the negative energy states were already filled. It then followed, from the Pauli exclusion principle, that electrons could not make any transitions from positive to negative energies. The exception, of course, was if some of the negative energy states were *not* filled. And this could come about easily enough, if only a negative energy electron is given sufficient positive energy. It would then be ejected from the vacuum, leaving behind it a *hole* in the negative-energy sea. And this hole, Dirac argued, would behave just like a particle of positive energy, but with opposite charge. In this way the concept of *antimatter* was born.

With it, automatically, came the concept of pair annihilation and creation: a negative-energy electron, given sufficient energy, would appear as a positive-energy electron, leaving behind it a hole, i.e. a positron. A particle and its antimatter partner would both come into being -. *pair creation*. Equally, the hole could subsequently be filled; a positive-energy electron would then disappear, as the electron enters the hole, and the hole itself disappear - *pair annihilation*. All will be in order so long as only energies *relative* to the vacuum have any observable effect.

These ideas translate readily into a reinterpretation of the canonical second quantization formalism, that Dirac had earlier developed to treat the many-body problem in NRQM, and in his (non-relativistic) treatment of radiation as a boson ensemble. In this formal framework, many-body operators $d\Gamma(X)$ are defined, for any 1-particle operator X, by replacing what in NRQM would be the expectation values of X by the corresponding expression in which the state (ket) ψ is replaced by a q-number field Ψ (the annihilation field), and the conjugate (bra) replaced by the adjoint field (the creation field). Thus, in the case of the Hamiltonian H, using the position representation:

$$< H >= \int \psi^*(x,t)H\psi(x,t)d^3x \to d\Gamma(H) = \int \Psi^*(x,t)H\Psi(x,t)d^3x. \tag{3}$$

Note that the RHS is a q-number, whereas the LHS is a c-number. Note further that the annihilation field always stands to the right, so such expressions always annihilate the vacuum; the vacuum is always the zero eigenstate of the total energy.

Even more so than in elementary quantum mechanics, the momentum representation has a special status. Let $b_r^*(\overrightarrow{p})$, $b_r(\overrightarrow{p})$ be, respectively, the creation and annihilation fields for a positive-energy solution to the free Dirac equation, with 3-momentum $\overrightarrow{p}$ and bispinor w_r, $r = 1, 2$. For each p the set of such bispinors is a 2-dimensional vector space, so any such solution can be written as a linear combination of these creation operators applied to the vacuum. Let $b_{r+2}^*(\overrightarrow{p})$, $b_{r+2}(\overrightarrow{p})$ be the corresponding operators for the negative energy bispinor w_{r+2}, $r = 1, 2$. Let $p_0 = +\sqrt{\overrightarrow{p}^2 + m^2}$. Then the annihilation field $\Psi(x,t)$ has the Fourier decomposition:

$$\Psi(x,t) = \int \sum_{r=1,2} [w_r(\overrightarrow{p})b_r(\overrightarrow{p})e^{-i(p_o t - \overrightarrow{p}.\overrightarrow{x})/\hbar} + w_{r+2}(\overrightarrow{p})b_{r+2}(\overrightarrow{p})e^{i(p_o t + \overrightarrow{p}.\overrightarrow{x})/\hbar}]\frac{d^3p}{p_0}. \tag{4}$$

It involves annihilation operators only. The Fourier expansion for the adjoint field only involves creation operators. We now follow the recipe of second quantization, Eq(4), using (5) and its adjoint, for the free-field Hamiltonian $H = \pm p_0$ (in the momentum representation) and the charge $-e$ (a multiple of the identity). The total energy and charge are then:

$$d\Gamma(H) = \int \sum_{r=1,2} p_0[b_r^*(\overrightarrow{p})b_r(\overrightarrow{p}) - b_{r+2}^*(-\overrightarrow{p})b_{r+2}(-\overrightarrow{p})]\frac{d^3p}{p_0} \tag{5}$$

$$d\Gamma(-e) = -e\int \sum_{r=1,2} [b_r^*(p)b_r(p) + b_{r+2}^*(-\overrightarrow{p})b_{r+2}(-\overrightarrow{p})]\frac{d^3p}{p_0}. \tag{6}$$

Necessarily, in accordance with (3), these operators still annihilate the vacuum, since the annihilation operator automatically appears on the right.

So far everything has been done in exact (in fact in rigorous) correspondence to NRQM. Now for the change in the physical picture, and corresponding to that, a change of notation. We are going to consider that all the negative energy states are filled. Since the absence of a negative energy electron with bispinor w_{r+2} and 4-momentum $(-p_0, \overrightarrow{p})$ behaves just like the presence of a positive energy particle of opposite charge, the annihilation of the former is equivalent to the creation of the latter, and vice versa. Furthermore, the operators (5), (6) and (7) do not annihilate this new vacuum state. To reflect these facts, denote $b_3(-\overrightarrow{p})$by $d_1^*(\overrightarrow{p})$, and $b_4^*(-\overrightarrow{p})$ by $d_2(\overrightarrow{p})$. Also introduce new notation for the bispinors, denoting w_r by u_r and w_{r+2} by v_r . In terms of these notational changes, the Fourier expansion for the field Ψ (before a pure annihilation field, as given by Eq.(5)) becomes:

$$\Psi(x,t) = \int \sum_{r=1,2} [u_r(p)b_r(p)e^{-ipx/\hbar} + v_r(p)d_r^*(p)e^{ipx/\hbar}]\frac{d^3p}{p_0} \qquad (7)$$

Evidently it does not annihilate the new vacuum; “effectively”, it is a sum of (hole) creation and (electron) annihilation fields. Concerning (6), (7), here we want to end up with second quantized expressions which do annihilate the new vacuum, so let us re-order terms in these expressions so that the effective annihilation operators always stand to the right. These operators must obey *anticommutation* relations, so as to preserve the antisymmetrization of the states they act on, so this introduces a change in sign. It also introduces the c-number value of the anticommutators, which we must integrate over (a divergent integral). We thus obtain:

$$d\Gamma(H) = \int \sum_{r=1,2} p_0[b_r^*(\overrightarrow{p})b_r(\overrightarrow{p}) + d_r^*(\overrightarrow{p})d_r(\overrightarrow{p})]\frac{d^3p}{p_0} - \text{ infinite constant} \qquad (8)$$

$$d\Gamma(-e) = -e\int \sum_{r=1,2} [b_r^*(\overrightarrow{p})b_r(\overrightarrow{p}) - d_r^*(\overrightarrow{p})d_r(\overrightarrow{p})]\frac{d^3p}{p_0} - \text{ infinite constant.} \qquad (9)$$

The infinite constants are readily interpreted as the energy and charge (both negative) of the Dirac vacuum. The change in sign makes the q-number part of the total energy non-negative, that of the total charge indefinite. Each involves number operators for the electrons out of the sea, and the holes, the positrons, in the sea. In both cases they have only positive energies.

Evidently this theory brings with it a problem exactly like the zero–point energy difficulty - a vacuum energy twice the value of the latter for each p, r, but only for the negative-energy states. Indeed, the cancellation of the zero-point energy for fermion fields (negative) and boson fields (positive), given unbroken supersymmetry, is replicated using the hole theory instead. The negative-energy fermion sea cancels the zero-point energy of the associated boson and antiboson fields.

A comparison with the classical ether is also instructive. The Dirac vacuum was formulated in terms of *traditional* mechanical principles (traditional both in classical and quantum mechanical terms). In particular it was based on the principle that particle number is conserved. In the hole theory pair creation and annihilation processes are always transitions between states of the same number of particles. The functor $d\Gamma$ of the second-quantization process yields operators which always preserve particle number. This was not of course a metaphysical principle, on a par with the principle that there must be a bearer of the motion of waves - but then neither was the latter a very plausible basis for the commitment of classical ether theorists to the ether. Its roots, like the roots of the ether, were pragmatic: there was no known method for introducing particle interactions, nor for so much as defining a quantum or classical particle system, other than in terms of a phase space of fixed dimensionality. It was the same in the case of Dirac's quantization of the electromagnetic field four years previously. There too, although it was clear that photon number should be subject to change, Dirac modelled such processes in terms of transitions between states which preserved particle number. In this case the transitions were to and from a sea of zero-energy photons. There too photon number was preserved.

The negative energy sea was effective in other ways as well. Like the classical ether, it was a fertile source of heuristics. Dirac was quickly led to the concepts of vacuum polarization, and of contributions from the sea to the effective charge and mass of the electrons and holes. But equally, and again in parallel to the classical ether, the new vacuum did not really make physical sense. It was hard to take the theory as literally true - it was "learned trash", in Heisenberg's words. But equally, it was not quickly or easily dispensed with, not even after the negative energy difficulty was dealt with by field quantization, where no explicit mention of the negative energy sea need ever be made.

The field-theory which replaced the hole theory was introduced by Heisenberg and Pauli in 1932. There was to be no correspondence with n-particle quantum mechanics. There was no canonical second quantization. There was to be a field (and an adjoint field) now taken to be fundamental, obeying anticommutation relations which were understood as *quantization* rules. Each was written down as before as a Fourier expansion in normal modes, but now the coefficients of these expansions were interpreted *ab initio* as a combination of antiparticle creation and particle annihilation fields, exactly as in Eq.(8). The global operators for the field could be obtained in formally the same way as with the second-quantized theory (as in Eq.(4)), but they were to be understood in terms of the generators of the symmetries of a field Lagrangian. Of course, the q-number fields were to have the very particular action on the vacuum - no longer a negative-energy sea! - as given in Eq.(8), acting on the Dirac vacuum. For this no explanation was given. The re-ordering process was still to be used, because with the field (8), entering into expressions of the form (4), one obtains creation fields to the right; but now the c-number values of the anticommutators were simply discarded. It was called normal-ordering. It was viewed as part of the increasingly elaborate procedure for isolating finite expressions in perturbation theory. Only one infinite negative term was allowed an occasional physical explanation: a (negative) zero-point energy.

Much more common was to reserve physical interpretation only for normal-ordered quantifies. In terms of the normal-ordered energy, the vacuum of the field theory has zero energy. For two generations physicists have shifted back and forth between the view that the subtractions of renormalization theory are no more than formal, and the view that they reflect real physical quantities. The Dirac vacuum provided a clear physical picture of *all* these subtractions, based at was on the canonical formalism of NRQM, and, apart from certain notable exceptions, for at least for one generation of physicists - Dirac's generation - this picture remained the fundamental one. Witness Wightman:

> It is difficult for one who, like me, learned quantum electrodynamics in the mid 1940s to assess fairly the impact of Dirac's proposal. I have the impression that many in the profession were thunderstruck at the audacity of his ideas. This impression was received partly from listening to the old-timers talking about quantum-electrodynamics a decade-and-a-half after the creation of hole theory; they still seemed shell-shocked. (Wightman, 1972 p.99)

One might add that Wightman never had to *accept* the hole theory; he never had to work with it. Fiction is never shell-schocking.

A further blow for the Dirac vacuum came with Pauli and Weiskopf's treatment of the complex *scalar* field, using *commutator* relations, soon after - with a similar interpretation of the Fourier expansion as the electron-positron field, with similar normal ordering prescriptions, and with its application to the new field of meson physics. With that antiparticles were seen as ubiquitous; they were the appropriate field-theoretic account of the negative energy terms, likewise ubiquitous in relativistic quantum theory, with no special connection to the Pauli exclusion principle. Obviously with scalar fields there could be no question of a filled negative energy sea.

There is a last chapter to this story, but a contentious one. If we consider again the classical ether, the situation *vis a vis* the Dirac vacuum is closer to Lorentz's position on the ether than to Einstein's. (Quantum) mechanical principles are no longer applied to the QED vacuum; the explanations that it offers of pair creation and annihilation events, in NRQM terms, are no longer valued; but still there lurks a story of sorts to be told, to account for the plane-wave expansion Eq.(8), as there lurked a story to be told to explain the length contraction and time-dillation effects in classical electrromagnetism. There is an analog to the Lorentz pedagogy; call it the Dirac pedagogy. Even very recent introductions to the subject make use of it. The Dirac vacuum is not taken realistically - indeed, it is sometimes introduced without even mentioning the hole theory - no more than the Lorentz pedagogy takes seriously the resting frame, or makes mention of ether. But what is missing is an alternative account of the plane wave expansion, and of the details of the relationship of negative energy states to antiparticles, and of the meaning of normal ordering. There is as yet no good analog to the Einstein-Minkowski geometric account of contraction and dilation phenomena, in terms of invariant intervals between events. One explains these phenomena very simply, as the consequence of taking different events as the simultaneous end-points of rods, and as the simultaneous ticks of clocks, depending on one's choice of simultaneity.

Is there an alternative explanation for the plane-wave expansion - for how antiparticles get into the theory? It can certainly be shown that only fields built up out of creation and annihilation fields for two kinds of particles, as given by Eq.(8) and its adjoint, can be Lorentz-covariant, satisfy microcausality, and transform simply under a U(1) gauge symmetry (Weinberg 1964). The two kinds of particles have to be identical in all respects, but of opposite charge. On this approach one starts from the free one-particle Hilbert space theory, using the Wigner classification of the irreducible representations of the Poincaré group. Creation and annihilation operators can be defined in these terms over the associated Fock spaces, just as in NRQM.. But one never in this way makes any mention of negative-energy states, and the normal ordering process is unexplained.

Weinberg's account of the structure of free-field theory is on the right lines, but it can certainly be improved on. For this we need Segal's method of quantization, itself a fragment of what is nowadays called geometric quantization. Given a complex structure J on the classical solution manifold V of a linear system of equations - a linear map such that J^2 acts as minus the identity - and given a non-degenerate bilinear form S on V, one can always define a Hilbert space V_J, and from this construct a Fock space $F(V_J)$ over V_J . If these equations and the bilinear form are covariant, this Hilbert space will inherit their covariance group. In the symmetric case, define the *Segal field* abstractly, as a linear map Φ from V to self-adjoint operators on a suitable Hilbert space:

$$\{\Phi(f), \Phi(g)\} = \hbar S(f, g). \tag{10}$$

A field with these properties can be represented concretely, given the creation and annihilation fields on $F(V_J)$, by the relations:

$$\Psi_J(f) = \frac{1}{\sqrt{2\hbar}}(\Phi(f) + i\Phi(Jf)) \tag{11}$$

$$\Psi_J^*(f) = \frac{1}{\sqrt{2\hbar}}(\Phi(f) - i\Phi(Jf)) \tag{12}$$

From the anticommutator (11), Ψ_J and Ψ_J^* obey the anticommutators characteristic of annihilation and creation operators on the antisymmetric Fock space, and *vice versa*. These anticommutators vanish for f, g with spacelike separated supports, if S does.

If one begins with a complex system of Poincaré-covariant equations, there are always two possible choices of J. One of them is just multiplication by i, denote J_N. It is local, but the energy it gives rise to (when used in Stone's theorem, for the generator of translations in time) is indefinite. The other choice, denote J_P, makes use of the decomposition of the classical solutions into positive and negative frequency parts (positive and negative "energies"). This is non-local, but gives a positive energy. The two are related by the Segal field Φ, which is independent of the complex structure. One then finds that (with $f^+(f^-)$ the positive (negative) frequency part of f):

$$\Psi_{J_N}(f) = \Psi_{J_P}(f^+) + \Psi_{J_P}^*(f^-) \tag{13}$$

$$\Psi_{J_N}^*(f) = \Psi_{J_P}(f^-) + \Psi_{J_P}^*(f^+). \tag{14}$$

The quantities on the left, defined with respect to the local complex structure J_N, are the local, causal fields; they have exactly the interpretation we have been seeking, in terms of the creation and annihilation operators defined using the non-local complex structure. Eq.(12) is the abstract analogue of Eq.(8). Dirac was canonically second-quantizing using the local complex structure, so using the Ψ_{J_N}'s in Eq.(4). By linearity, this ensures the invariance of such terms under rotations in the local complex structure. This in turn forces charge conservation. But he used the same complex structure to define their particle interpretation, their Fock-space action. In this way he forced number conservation too, but at the price of introducing negative energies. These same interaction terms are not invariant under rotations in the non-local complex structure, the one which ought to be used to define the particle interpretation. Particle number is correspondingly not preserved by them. Introducing the negative energy sea is in fact a way to switch between one complex structure and the other. If one normally-orders canonically second quantized operators, defined on $F(V_{J_N})$, one obtains the canonically quantized operators, defined on $F(V_{J_P})$. In this sense the normal ordering process likewise switches between the two complex structures (Saunders 1991). But the canonical second quantization cannot be used to define local bilinear quantities, in particular couplings to other fields, so this correspondence is not of much use outside of free-field theory, for the global quantities.

This same framework applies to boson fields, save that there one has an *antisymmetric* bilinear form (a symplectic form). The Segal field in this case satisfies commutation relations, but otherwise the same analysis goes through.The fact that the non-local complex structure is the one that is used at the level of the Hilbert-space theory, in both cases, also explains why there is no local, covariant position operator in relativistic quantum theory (Saunders 1992). The analysis applies equally to NRQM, save that there the two complex structures coincide.

With that the Dirac vacuum can, I think, finally be laid to rest. We have not replaced it with a better physical picture, but we have shown why any linear, Lorentz covariant system of equations, used to define a one-particle Hilbert space and a Fock space over that, will lead to the two systems of fields, related in a way which can be explained using the hole theory (in the case of a symmetric bilinear form). We see why, both for bosons and fermions, the negative energy states are associated with antiparticles; we see why interactions built up from local, covariant, gauge invariant quantities, automatically lead to pair creation and annihilation processes. And we have some insight into the meaning of the normal ordering process.

I do not believe there is a negative energy sea. But whilst we understand the normal ordering in the free case, and can see that it yields the true vacuum energy using the non-local complex structure - the correct one for defining a positive energy particle interpretation - with value zero (i..e. $< d\Gamma(H)_{J_P} >_\omega$, where ω is the vacuum of $F(V_{J_P})$), to conclude from this that there is no vacuum energy is to suppose that really the theory concerns particles; that the fields are only devices for introducing local interactions; that if there are no particles present, there is no field. To solve the cosmological constant problem in this way, one will have to say the same for both fermion and boson fields. And even if one is prepared to do this, there will remain the problem of vacuum expectation values of local bilinear self-adjoint expressions, using the local fields (14), (15), which cannot be normal-ordered (normal ordering is a global procedure). What of these?

5 The Reality of Zero-Point Fluctuations

If one takes quantum fields as fundamental, rather than as devices for introducing local particle interactions, then there is a clear intuitive argument for supposing the ground state has non-zero energy, deriving from the elementary theory of the harmonic oscillator.

This is again non-local, depending as it does on the Fourier decomposition of the field into normal modes, and with that the particle interpretation. A more direct argument is possible, which shows that the Minkowski vacuum cannot be a zero eigenstate of the local fields. I have said that these cannot be normal-ordered, as normal-ordering is a global operation. A much stronger claim can be made: the vacuum cannot be a zero eigenstate of the stress-energy tensor, whatever subtractions are made to it.

A first-point to make is that no positive-definite operator, like the square of electric field, can have zero expectation value in any state. Now consider the CCR's for the components of the stress-energy tensor $T^{\mu\nu}$:

$$[T^{0k}(x), T^{00}(x')] = -i(T^{00}(x)\partial^k\delta(\mathbf{x}-\mathbf{x}') - i(T^{kl}(x')\partial_l\delta(\mathbf{x}-\mathbf{x}') \quad (15)$$

(see Schwinger 1973 p.26). If we replace T by $\widetilde{T} = T + kM$, for some matrix of real numbers M, then from the commutator it follows that M will have to be a multiple of the Minkowski metric g. But in that case, if we require that all the components of $\widetilde{T}$ vanish in the vacuum:

$$< \widetilde{T}^{00} > = < \widetilde{T}^{0k} > = < \widetilde{T}^{kl} > = 0 \quad (16)$$

then certain constraints follow:

$$< T^{00} + \frac{1}{3}T^{kk} >= 0. \tag{17}$$

(Here the Einstein summation convention applies.) In the case of the electromagnetic field, $T^{kk} = T^{00}$, so:

$$\frac{4}{3} < T^{00} >= 0. \tag{18}$$

Since T^{00} is $\frac{1}{2}(\overrightarrow{E}^2 + \overrightarrow{B}^2)$, a positive definite operator, Eq.(18) cannot be satisfied.

The vacuum state is therefore not an eigenstate of the components of the stress-energy tensor, whatever subtractions are introduced. In consequence it is pictured as fluctuating, a picture which is clearly in evidence in vacuum polarization effects in perturbation theory, where one interprets bubble diagrams for internal lines as virtual pair creation and annihilation events.

Of course in perturbation theory such expressions are formally divergent. One does eventually obtain finite expressions, small in QED, reflecting such contributions. One might grant that the vacuum in QFT is a non-zero eigenstate of the energy, but yet hope the quantities involved are small. It may be on a naive approach one ends up with divergent expressions, as for the zero-point energy; a more careful analysis may lead to different conclusions.

This point of view appears to me to be perfectly sound, but it is hard to evaluate it in the absence of a quantum theory of gravity. The point about renormalization theory - and this is especially evident in renormalized perturbation theory, where one introduces counterterms at each order, and in this way obtains controlled and finite contributions from the perturbation considered - is that one is always interested in shifts in physical quantities - be it the energy or coupling constant or charge - relative to a baseline. The theory is always defined at a certain scale M; the conventions there adopted remove all ultraviolet divergences by fiat. One then imagines a shift in the scale M, with a corresponding shift in the renormalized coupling constants, field strengths, and Green's function. The renormalization group equation tells us how these shifts are interrelated. One never considers the absolute values - but just these are what are relevant, uniquely, to gravity.

On the older approach to renormalization theory, where one does not introduce counterterms, the procedure for extracting finite parts of formally divergent expressions assumed that the bare values of the coupling constants and so on have whatever (infinite!) value are necessary to yield the phenomenologically observed values. It is hard to see how this sort of procedure could be applied to eliminate the zero-point energy, however, unless it amounts to the view already considered, where one simply introduces a cosmological constant to cancel it - to one part in 10^{120}. Of course the cancellations needed between formally infinite quantities, to yield the observed phenomenological parameters, are infinitely more exact. One cannot take these cancellations seriously. It is surely for this reason that Schwinger, one of the greatest practitioners of the older methods of renormalization theory, was increasingly driven to the more radical step, of viewing fields as wholly tied to their sources, so that there is no field in the absence of sources - and of backing this up with a very different perspective on field theory, which essentially does without q-number fields altogether (the so-called source theory).

Renormalization group methods, and the many applications of them to condensed matter physics as well as to RQFT, have considerably improved the situation. And whilst there are surely vacuum polarization effects leading to energy shifts in observed quantities, and good arguments to show that the vacuum cannot be an exact eigenstate of the local stress-energy tensor, there seems to be no good reason to suppose the vacuum energy will be excessively large - except for the zero-point energy.

I have already indicated one solution: one views the fields as auxiliary devices, used to introduce local interactions among the real objects of the theory, the particles. Schwinger's program, of eschewing q-number fields altogether, is a further and more radical step, motivated as much by the unsatisfactory status of renormalization theory as Schwinger knew it as by the problem of the zero-point energy. But even on the former approach is revisionary; it is not simply a shift in interpretation; one has yet to give a principled distinction between field quantities which are real, reflecting the properties of particles, and those which are fictitious. There is moreover the further argument for the reality of the zero-point fluctuations, considered by many as direct evidence for them: the Casimir effect. It is high time we considered it in detail.

Casimir's discovery, recall, was that an attractive force acts between parallel, conducting plates, separated by a distance a, of magnitude (per unit area):

$$P = \hbar c \frac{\pi^2}{240} \frac{1}{a^4}. \tag{19}$$

His explanation for it was beautifully simple. It is that the presence of the plates imposes boundary conditions (Dirichlet condition) on the stationary modes of the electromagnetic field. In particular, wavelengths $\lambda > a$ are excluded, even in the case of the vacuum fluctuations; therefore, if plates are brought further apart the energy density between the plates increases, as normal modes of longer wavelength can then come into play; so work has to be done to separate the plates.

There is an another explanation. There are fluctuations in the polarization fields associated with the electrons in the conducting plates (realistically, no metal is a perfect conductor at arbitrarily high frequencies), and these couple with each other, giving rise to van der Waals forces. The coupling can be modelled either in terms of distance forces, or as mediated by electromagnetic fields with these fluctuating polarization fields as sources.

This explanation is due to Lifshitz. He introduced a random polarization field for a material of dielectric constant ϵ_0; the expression he derived for the resulting pressure, from their mutual attraction, is:

$$P = \frac{\hbar c \pi^2}{240 a^4} (\frac{\epsilon_0 - 1}{\epsilon_0 + 1})^2 \varphi(\epsilon_0) \tag{20}$$

where φ is a function with the limiting behavior, for $\epsilon_0 \to \infty$:

$$\varphi(x) \to 1 - (\frac{1.11}{\sqrt{\epsilon_0}}) \ln(\frac{\epsilon_0}{7.6}). \tag{21}$$

In the limit Eq.(22) and (20) coincide. Of course one should not take this limit seriously, as then one has a perfect conductor, and there will be no polarization field, but the point is that one has approximate agreement with the observed Casimir force. Casimir himself had to introduce a high-frequency cut-off to obtain a finite result; he too made the same assumption.

But Lifshitz's methods are perfectly consistent with the interpretation of the effect in terms of vacuum fluctuations. Whence the random fluctuations in the dielectric medium, in the zero-temperature limit? On general physical principles, such fluctuations should be dissipative, and equilibrium can only be reached if there are likewise fluctuations in the electromagnetic field in the cavity. That is just the picture invoked by Casimir: even at zero temperature, there is a residual energy density in the vacuum. Lifshitz's argument can therefore be seen as strengthening Casimir's.

It is worth remarking that the Casimir effect does not always give rise to an attractive force. In fact it is repulsive in the case of spherical shells, a discovery which killed off a speculative attempt to calculate the fine-structure constant $e^2/\hbar c$ from first principles. The old Abraham-Lorentz model, recall, supposed the electron, in its rest frame, to be a charged conducting spherical shell of radius a. Its electrostatic energy is

$$E_e = e^2/2a \tag{22}$$

with a corresponding tension $e^2/8\pi a^4$, tending to expand the shell. In view of his results for parallel plates, Casimir suggested that there would be an energy E_c associated with the sphere which would increase with its radius, so giving rise to a compensating attractive force. On dimensional grounds this will be of the form:

$$E_c = -C(\hbar c/2a) \tag{23}$$

So that the resultant tension in the surface will vanish if C equals the fine structure constant. Using the parallel plates result, approximating a sphere of radius a as two plates of area πa^2 a distance a apart, one finds

$$C \approx 0.09 \tag{24}$$

which is only a factor 10 off from the fine-structure constant. Actually, this rough calculation is numerically correct; C *is* of the order 0.09 (Boyer 1968). But the energy *decreases* with increasing radius; the Casimir force is repulsive, in the case of a sphere, not attractive. More generally, the Casimir effect turns out to be extremely sensitive to the geometry of the boundaries involved, as well as their composition (and to the space-time dimension, curvature, and type of field).

The fact that repulsive forces can be obtained has been cited as a reason for rejecting the interpretation of the effect in terms of van der Waals forces (Elizalde and Romeo 1991), on the grounds that the latter are always attractive. But this claim is a dubious one; dipole and higher moments can give rise to repulsive forces as well as attractive ones; neither can the overall, collective behavior, of polarization fields in dielectrics, be obtained by simple additive effects of constituent molecules, no more than from additive effects of vacuum fluctuations. In fact neither explanation - neither in terms of van der Waals forces, nor in terms of zero-point fluctuations - gives a simple account of the dependence of the sign of the force on the geometry of the conductors.

I have already remarked on Schwinger's source theory. It was first applied by Schwinger to the Casimir effect in 1975; more refined studies were presented in 1977 and 1978. It was granted that "the Casimir effect poses a challenge for source theory, where the vacuum is regard as truly a state with all physical properties equal to zero". The approach that was taken was similar to that of Lifschitz:

> The Casimir effect is a manifestation of van der Waals forces under macroscopic circumstances. When material, un-

> charged bodies are present, whether dielectrics or conductors, nonvanishing fluctuating electric and magnetic fields are generated, which give rise to stresses on the objects. These fields may be most easily expressed in terms of Green's functions evaluated in a source theory formulation of quantum electrodynamics. (Milton et al 1977).

To give a flavour of the analysis, one starts from the electric and magnetic fields defined by the source field $\overrightarrow{P}$, just as for a polarization field:

$$\nabla \times \overrightarrow{H} = \epsilon \dot{\overrightarrow{E}} + \dot{\overrightarrow{P}}, \ \nabla \cdot (\epsilon \overrightarrow{E} + \overrightarrow{P}) = 0 \tag{25}$$

The fundamental objects in the theory are the Green's functions $\overrightarrow{\Gamma}(x, x')$ by means of which the fields and sources are related:

$$\overrightarrow{E}(x) = \int \overrightarrow{\Gamma}(x, x') \cdot \overrightarrow{P}(x') d^3x. \tag{26}$$

The general method is to consider the change in the action (or the energy), expressed as integrals over the sources, on variation of the parameters determining the geometry of the dielectrics: by subtracting the vacuum Greens function (spherical case), and by variation of the dielectric constant (parallel conductors, corresponding to a change in the distance between the plates). In the latter case Schwinger discards a term which he identifies as the change in vacuum energy. The change in energy defined by the volume integral of the fields (including the polarization fields), due to variation of the dielectric constant $\delta\epsilon$ is:

$$\delta E = \frac{i}{2} \int \delta\epsilon(\mathbf{r}, \omega) \mathbf{\Gamma}_{kk}(\mathbf{r}, \mathbf{r}, \omega) \frac{d\omega}{2\pi} (d\mathbf{r}). \tag{27}$$

The polarization field $\overrightarrow{P}$ has dropped out of the analysis. What remains is the computation of the Green's function; its components Γ_{kk} are given by the expression:

$$\Gamma_{kk} = [\omega^2 g^E + \frac{k^2}{\epsilon\epsilon'} g^H + \frac{1}{\epsilon} \frac{\partial}{\partial z} \frac{1}{\epsilon'} \frac{\partial}{\partial z'} g^H] \mid_{z=z'} \tag{28}$$

where the g's are the Green functions for the electric and magnetic field satisfying, for the electric field:

$$[-\frac{\partial^2}{\partial z^2} + k^2 - \omega^2 \epsilon] g^E(z, z') = \delta(z - z') \tag{29}$$

(the equation for the magnetic case is similar). Of the term contributing to δE, which Schwinger interprets as the change in volume energy, he says: “Since this term in the energy is already phenomenologically described, it must be cancelled by an appropriate contact term”. But it is not clear that this is correct; the analysis throughout has been in terms of the phenomenological quantities; the maneuver appears to be *ad hoc*.The remaining method used by Schwinger, whereby the radial component of the stress-energy tensor for the electromagnetic field is calculated, likewise involves an infinite subtraction. It is justified with the words: “No physical meaning can be ascribed to such a term, however, since no stress can arise from a homogeneous dielectric (as such it can be canceled by a contact term)”. This term too has the same form as the expression for the vacuum energy arising in Casimir’s calculation.

Whatever the virtues of Schwinger’s source theory, transparency, and statements of clear and systematic principles, are not among them. I do not believe his methods deliver an unambiguous verdict on this matter.

The essential question remains: are the conductors the source of the vacuum energy, if any, in the region bounded by the conductors? If so there is no evidence, coming from this quarter, for the zero-point energy. The argument that it is not is as stated: no equilibrium would then seem to be possible. By the fluctuation-dissipation theorem, the system will be dissipative. The principle is an extremely general one.

Against this Rugh, Zinkernagel and Cao (1999), in their review of the various treatments of the Casimir effect, have suggested:

> The answer seems to be that there is no place for the dissipated energy (from the dipole) to go. If there is no vacuum field (in standard QED, an ensemble of harmonic oscillators) where the energy can be dissipated into, then there is no vacuum-dipole fluctuation-dissipation theorem.

The proposal will have to more radical than merely repudiating the vacuum of QED, however. They will have to impugn the reality of classical fields as well. This is not a claim that can be underwritten by the source theory, for example; it is not as though Schwinger considered the c-number fields as unreal. Rugh *et al* also make another suggestion: why not adopt the Lifschitz’s view, accepting that equilibrium is maintained by appeal to vacuum fluctuations between the plates, but suppose that such fluctuations are brought into existence by the plates? But that will hardly do, unless such fluctuations are brought into existence not only between the plates, but everywhere in space; for the combined system will in that case be dissipative, with nothing beyond it to restore equilibrium.

6 Conclusion

We cannot lightly abandon general arguments on the nature of thermodynamic equilibrium. Wherever the vacuum state is probed, we see evidence of stochastic activity. I do not think we can do without appeal to the zero-point energy in explaining the Casimir effect. It is also clear that vacuum fluctuations have similar applications in other instances. For example, they can also be used to explain the thermal background predicted for accelerating detectors (the Unruh effect). As Sciama (1991) and others have argued, this background would be expected if the detector is sampling the zero-point spectrum along non-inertial trajectories. In a semiclassical treatment, if one evaluates the 2-point function along world-lines of constant acceleration, one obtains a thermal distribution at the Unruh temperature. And the picture of the vacuum that arises makes it perfectly clear that it contains energy:

> How can a unit volume of empty space contain energy? The answer in part lies in the fact that, according to quantum mechanics, physical quantities tend to fluctuate unavoidably. Even in the apparent quiet of the vacuum state pairs of particles are continuously appearing and disappearing. Such fluctuations contribute energy to the vacuum. (Abbott 1988).

But what comes to the fore in these explanations is that they are semiclassical ones. They are explanations which could equally be made on the assumption of a classical stochastic background. That assumption is in fact enough to derive a surprisingly large fragment of quantum electromagnetic phenomenology, on a purely classical basis. By Sudarshan's theorem, one has to go to effects involving 4-point correlations to find an application that cannot be treated in semiclassical radiation theory.

There are evident parallels with the situation regarding the classical ether and the Dirac hole theory. In both cases a presumed structure to the vacuum extended the reach of familiar physical concepts into a novel terrain. The zero-point energy extends the reach of classical stochastic theories. This presumed structure to the vacuum led to internal tensions; they were resolved not by simply jettisoning the medium, but be reformulating it in terms of different dynamical principles - purely electromagnetic ones, in the classical electromagnetic case (giving up Newtonian force laws, and ultimately Newtonian space-time), and purely local quantum principles, in the case of the hole theory (giving up the framework of NRQM). Is something similar possible in the case of the zero-point energy?

It seems we can go at least part way on this strategy. The familiar concept in the case of QED vacuum is the idea of a stochastic background. I do not doubt that there is evidence for it, but it is another matter to suppose the vacuum is a state of fluctuation fields even when there is nothing to probe these fluctuations. It is a natural assumption that the fluctuations are there even when nobody looks, but one only has to put the matter in these terms for it to appear immediately in doubt: it is too closely linked to the general interpretational problem of quantum mechanics. It is the uncritical view, that underlying laboratory phenomena, at the microscopic scale, is a world of chancy events. It is the world it is the aim of stochastic hidden variable theories to describe.

In point of fact, on every other of the major schools of thought on the interpretation of quantum mechanics - the Copenhagen interpretation, the pilot-wave theory, and the Everett interpretation - there is no reason to suppose that the observed properties of the vacuum, when correlations are set up between fields in vacuuo and macroscopic systems, are present in the *absence* of such correlations. This is clear in the Copenhagen interpretation, if for no other reason than on this approach no statement at all can be made which is not tied to a definite experimental context. It is also true on the Everett interpretation, at least on those versions of it in which the preferred basis is not viewed as fundamental; where it is significant only because it the basis to be used to define stable, complex dynamical systems, of the sort that *we* are made of.[2] On this approach there are no probabilistic events underlying our macroscopic environment, unless and insofar as correlations have been established with them. And it is true on the pilot-wave theory as well. In this theory nothing is in motion in the quantum vacuum. Take, for simplicity, the harmonic oscillator. The phase of the ground-state is independent of the position. It can be chosen as a constant. In this circumstance the particle - I am talking of the pilot-wave theory of NRQM - is stationary. Nothing at the level of the beables is in motion. It is the same with the c-number fields in the pilot-wave theory of QFT: nothing is moving in the vacuum. (This is true even though the uncertainty relations show there is statistical dispersion.) And as for the "effective" collapse of the wave-function in the pilot-wave theory (where one gets rid of terms entering into the overall superposition which no longer significantly contribute to the quantum potential), this is governed by decoherence theory just as in the Everett approach. Nothing like this goes on in the ultraviolet limit.

[2]Equivalently, the preferred basis problem is explained by the WAP. For the argument that the basis must be consistent, see Saunders (1993), Halliwell (1994); of course one wants quasiclassicality, not just consistency.

It seems that the suggestion of Rugh *et al* may be on the right lines after all, save that it is not that there is no field in the vacuum, ready to act as a sink for the fluctuating fields established between the plates; it is that fluctuations themselves only extend so far into the microworld, as there are mechanisms in play to establish correlations with macroscopic phenomena.

But if the earlier problems with vacuum are any guide, this can only be a part of the story. It is one thing to shift the physical picture of the vacuum, but quite another to fill it out quantitatively. If the stochastic background is real, but only insofar as it is correlated with matter, then how is its contribution to be calculated? It may be that here is a basis to explain the observed approximate parity between the energy density of vacuum and matter,[3] And there remain the unkown, but possibly large shifts in the vacuum energy density due to spontaneous symmetry breaking: evidently gravity had better couple to the net stress-energy tensor with the shifted vacuum as baseline; but why should be this be so, and how this coupling is to be regularized, poses entirely different problems.

References

Abbot, , L. (1988), "The Mystery of the Cosmological Constant", *Scientific American*, May p.82-88.

Bell, J. (1989), "How to Teach Special Relativity", in *Speakable and Unspeakable in Qauntum Mechanics*,

Boyer, T. (1968), "Quantum Electromagnetic Zero-Point Energy of a Conducting Spherical Shell and the Casimir Model for a Charged Particle", *Physical Review* **174**, 1764-76.

Brandenberger R. (1999), "Back Reaction of Cosmological Perturbations and the Cosmological Constant Problem", Brown-HET-1180.

Callen, H. and T. Welton (1951), "Irreversibility and Generalized Noise", *Physical Review* **83**, 34-40.

Coleman S., (1988), *Nuclear Physics B* **307**, 867.

Elizalde E., and A. Romeo (1991), "Essentials of the Casimir Effect and its Computation", *American Journal of Physics* **59**, 711-19.

Halliwell, J. (1994), 'Aspects of the Decoherent Histories Approach to Quantum Mechanics', in *Stochastic Evolution of Quantum States in Open Systems and in Measurement Processes*, L. Di¢si, ed., p.54-68, World Scientific, Singapore.

[3] Also called the "cosmic coincidence problem". For a recent review see Vilenkin (2001).

Lifschitz, E. (1956), "The Theory of Molecular Attractive Forces Between Solids", *Soviet Physics JETP* **2**, 73-83.
Milton, K., L. DeRaad, and J. Schwinger (1978), "Casimir Self-Stress on a Perfectly Conducting Spherical Shell", *Annals of Physics* **115**, 388-403.
Rugh, S., H. Zinkernagel, and T. Cao, 1999: "The Casimir Effect and the Interpretation of the Vacuum", *Studies in the History and Philosophy of Modern Physics* **30**, 111-39.
Saunders, S. (1991), “The Negative Energy Sea”, in *Philosophy of Vacuum*, ed. S. Saunders and H. Brown, Oxford: Clarendon Press.
Saunders, S. (1992), “Locality, Complex Numbers, and Relativistic Quantum Theory”, *Proceedings of the Philosophy of Science Association , Vol.1*, p.365-380.
Saunders S. (1993), “Decoherence and Evolutionary Adaptation”, *Physics Letters A* **184**, 1-5.
Sciama, D. (1991), "The Physical Significance of the Vacuum State of a Quantized Field", in *Philosophy of Vacuum*, ed. S. Saunders and H. Brown, Oxford: Clarendon Press.
Schwinger J. (1975), "Casimir Effect in Source Theory", *Letters in Mathematical Physics* **1,** 43-47.
Schwinger, J. (1973), *Particles, Sources, and Fields*, Vol.1, Reading, Mass.: Addison Wesley.
Schwinger, J., L. DeRaad, and K. Milton (1977), "Casimir Effect in Dielectrics", *Annals of Physics* **115,** 1-23.
Vilenkin, A. (2001), “Cosmological Constant Problems and their Solutions”, *arXiv:hep-th/0106083 v2.*
Weinberg, S. (1964), “Feynman Rules for Any Spin”
Weinberg, S. (1989): "The Cosmological Constant Problem", *Reviews of Modern Physics* **61**, 1-23.
Weinberg, S. (2000), “The Cosmological Constant Problems”, *Physical Review D* **61** 103505.
Wightman, A., (1972), "The Dirac Equation", in *Aspects of Quantum Theory*, Cambridge: Cambridge University Press.