PHYSICAL REVIEW D VOLUME 20, NUMBER 12 15 DECEMBER 1979

Casimir effect in quantum field theory*

Bernard S. Kay†
Institute for Theoretical Physics, University of Berne, Sidlerstrasse 5, 3012 Berne, Switzerland
(Received 13 December 1978)

A new conceptual foundation for renormalizing $T_{\mu\nu}$ on locally flat space-times—to obtain the so-called Casimir effect—is presented. The Casimir ground state is viewed locally as a (nonvacuum) state on Minkowski space-time and the expectation value of the *normal-ordered* $T_{\mu\nu}$ is taken. The same ideas allow us to treat, for the first time, self-interacting fields for arbitrary mass in perturbation theory—using traditional flat-space-time renormalization theory. First-order results for zero-mass $\lambda\phi^4$ theory agree with those recently announced by Ford. We point out the crucial role played by the simple *renormalization condition* that the vacuum expectation value of $T_{\mu\nu}$ must vanish in Minkowski space-time, and in a critical discussion of other approaches, we clarify the question of renormalization ambiguities for $T_{\mu\nu}$ in curved space-times. In an Appendix, we show how the Casimir effect arises in the C^*-algebra approach to quantum field theory.

I. INTRODUCTION

There has recently been a lot of interest[1] in defining the energy-momentum tensor for a quantum field in a fixed curved background. It was realized by DeWitt[2] that interesting problems arise already if we modify the global topology of *flat* space-time by identifying surfaces, introducing boundaries, etc. DeWitt pointed out that the physically measurable Casimir effect[3]—the attraction of two conducting plates in a vacuum—is in fact a problem of this nature. We follow DeWitt in believing that a clear understanding of these "Casimir-effect" situations should throw light on the more difficult generic case where there is the added complication of local curvature effects.

By now there is a good measure of agreement on calculations of the Casimir effect (and indeed the energy-momentum tensor in the generic case). However, there is a wide diversity of points of view on the *origin* and *nature* of the effect. Thus some approaches[4] make essential use of the concept of zero-point energy and suggest that the usual procedure of normal ordering is inadequate for dealing with such subtle phenomena. Other approaches[5] invoke the gravitational interaction and relate the effect to a renormalization of the cosmological constant. Yet other approaches involve a method,[6] unique in quantum field theory, in which infinities are handled and finite numbers extracted solely by regularization and without the need for any renormalization. The implication is that somehow the regularization procedure itself contains some deep physics.

Now each of these points of view may be valuable in inspiring some future theory. Nevertheless, from the point of view of the systematic development of flat-space-time quantum field theory, none of them is completely satisfactory. Thus:

(1) Zero-point energy (like "bare" coupling constants, etc.) is not usually felt to be a *necessary* concept. It would be preferable if it were relegated (again like "bare" coupling constants, etc.) to the status of providing some optional (if albeit valuable) intuition.

(2) Again, one would hope that a consideration of the gravitational interaction would not be *necessary* for the Casimir effect which can be regarded as an essentially flat-space-time phenomenon. (See also Sec. IV A.)

(3) Finally, we shall show (in Sec. IV B) that the "regularization without renormalization" referred to above does not always work.

The purpose of this paper is to show that Casmir-effect calculations *can* be performed within the usual framework of flat-space-time physics, using only conservative ideas about normal ordering and renormalization. In the second section we explain how, illustrating our ideas with free massless and massive scalar fields in a two-dimensional cylindrical universe. In the third section we treat self-interacting fields and show how they can be handled in perturbation theory. For certain cases of zero mass we obtain agreement with results recently announced by Ford,[8] and thus put these results on a firmer foundation. Unlike Ford's our method applies to fields of any mass and should, in principle, generalize to any order in perturbation theory.

Although the present work strictly applies only to locally flat space-times, it is our hope that the methods developed in this paper will suggest new approaches to quantum field theory in general curved space-times. We thus review in Sec. IV several current approaches to the quantum energy-momentum tensor in the light of their application to the Casimir effect. Included in the discussion are Wald's "axiomatic" approach,[9] Dowker and Critchley's and Hawking's "ζ-function" approach,[10] and the so-called "point-separation" approach.[11] Finally, we sketch in an Appendix the relationship with axiomatic quantum field theory.

II. CASIMIR EFFECT FOR FREE FIELDS

In this section we explain our approach to the Casimir effect in a simple intuitive way. A more careful, mathematical treatment is given in the Appendix.

Let us take, for sake of illustration, a massless Klein-Gordon field in a two-dimensional flat cylindrical universe of radius[12] R (see Fig. 1.):

$$\frac{\partial^2\varphi}{\partial t^2}-\frac{\partial^2\varphi}{\partial x^2}=0\,. \tag{1}$$

We want to calculate the expectation value of the energy density $H(p)=T_{00}(p)$ at a point p.

Now everything will follow from the basic principle that measuring $H(p)$ is a *local*-operation: To measure $H(p)$, our measuring device need only examine the properties of a small neighborhood N of p; the only quantum-mechanical observables involved are the field operators for points in the region N. But the structure of observables in the region N is precisely the same as it would be if N was embedded in a globally flat space-time. A measuring device—only making measurements in N—has no information about the global topology and must behave exactly as it would in a globally flat space-time.[13]

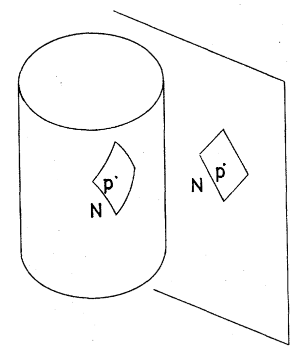

FIG. 1. The structure of field operators on the cylinder space-time is locally identical with that of ordinary flat space-time.

It follows that our *calculation* must proceed exactly as in a (globally) flat space-time. But, in a globally flat space-time, we know how to proceed. We must take the expectation value of the *normal-ordered* energy density.

Now what is different of course is the *state of the system*. The ground state s of the field propagating on the cylinder will *not* look like the vacuum state ω of globally flat space-time, even when we restrict our observations to observables in the region N. This is easily demonstrated by examining the two-point functions:

$$\begin{aligned} s(\varphi(x_2)\varphi(x_1)) &= -\frac{1}{4\pi}\ln\left\{\left[1-\exp\left(-\frac{i}{R}\left[(t_2-t_1)-(\chi_2-\chi_1)-i\epsilon\right]\right)\right]\left[1-\exp\left(-\frac{i}{R}\left[(t_2-t_1)+(\chi_2-\chi_1)-i\epsilon\right]\right)\right]\right\}+C\,,\\ \omega(\varphi(x_2)\varphi(x_1)) &= -\frac{1}{4\pi}\ln\left[(t_2-t_1)^2-(\chi_2-\chi_1)^2+i\epsilon t\right]+C' \end{aligned} \tag{2}$$

for $x_1=(t_1,\chi_1)$, $x_2=(t_2,\chi_2)$ in our region N, say.[14] It is important not to think of the state s as a "vacuum." Indeed, much of the confusion in the past has been caused by using the same symbol $\langle 0|X|0\rangle$ to indicate both $s(X)$ and $\omega(X)$.

We shall be careful in what follows to reserve the word "vacuum" for the "true" global Minkowski-space vacuum ω. Likewise, normal ordering will always be with respect to this "true" vacuum. We want, then, to calculate the expectation value $s(:H:(p))$ of the normal-ordered energy density $:H:(p)$ in the state s.

We shall find it advantageous to view this calculation as a simple type of renormalization, the physical "renormalization condition" being the demand that the *vacuum* expectation value of the energy density $\omega(:H:(p))$ is zero. $s(:H:(p))$, then, is calculated as follows: We take the formal expression

$$H(p)=\frac{1}{2}\left[\left(\frac{\partial\varphi}{\partial t}\right)^2(p)+\left(\frac{\partial\varphi}{\partial\chi}\right)^2(p)\right] \tag{3}$$

and regularize it in some way, obtaining $H_\Lambda(p)$ where Λ symbolizes the regularization parameter. (Experience in quantum field theory leads us to expect that, after renormalization, results will be regularization independent for a wide class of regularization procedures.)

Then

$$s(:H:(p)) = \lim_{\Lambda\to\infty} [s(H_\Lambda(p)) - \omega(H_\Lambda(p))] . \tag{4}$$

In words, we first subtract the vacuum expectation value, then we remove the regularization.

We choose here a point-separation-type regularization, obtaining

$$s(:H:(p)) = \lim_{x_1, x_2 \to p} \{(s-\omega)[\tfrac{1}{2}(\partial_t\varphi(x_1)\partial_t\varphi(x_2) + \partial_\chi\varphi(x_1)\partial_\chi\varphi(x_2))]\} \tag{5}$$

$$= \lim_{x_1, x_2\to p} \tfrac{1}{2}(\partial_{t_1}\partial_{t_2} + \partial_{\chi_1}\partial_{\chi_2})[s(\varphi(x_1)\varphi(x_2)) - \omega(\varphi(x_1)\varphi(x_2))] . \tag{6}$$

For the skeptical reader, we sketch an independent and rigorous derivation of Eq. (6) in the Appendix.

From translational invariance, we obtain

$$\lim_{x\to 0} -\frac{1}{2}\left(\frac{\partial^2}{\partial t^2} + \frac{\partial^2}{\partial\chi^2}\right)[s(\varphi(x)\varphi(0)) - \omega(\varphi(x)\varphi(0))] . \tag{7}$$

Changing to null coordinates $u = t-\chi$, $v = t+\chi$, and using Eq. (2)

$$s(:H:(p)) = \lim_{u,v\to 0} \frac{1}{4\pi}\left\{\frac{1}{R^2}\left[\frac{e^{-iu/R}}{(1-e^{-i(u/R-i\epsilon)})^2} - \frac{1}{(u-i\epsilon)^2}\right] + (u \leftrightarrow v)\right\} \tag{8}$$

$$= 2\lim_{u\to 0} \frac{1}{4\pi}\left\{\frac{1}{R^2}\left[\frac{1}{(u/R-i\epsilon)^2} - \tfrac{1}{12} + O\left(\frac{u}{R}\right)\right] - \frac{1}{(u-i\epsilon)^2}\right\} . \tag{9}$$

So,

$$s(:H:(p)) = -\frac{1}{24\pi R^2} \tag{10}$$

We thus recover the usual result[15] for this model.

Now, it might be objected that the renormalization ideology we propose in this paper is rather a luxury for this simple zero-mass case. We could simply have thrown away the $1/u^2$, $1/v^2$ divergences in $s(H(p))$ and got the right answer.[16]

Indeed, it is sometimes said[17] that renormalization is only really needed to deal with *logarithmic* divergences. For this reason we provide an example of such a case: We sketch the Casimir effect for a *massive* field in a two-dimensional cylindrical universe and postpone further discussion to Sec. IV.

Casimir effect in massive case. Now our energy density is formally

$$H(p) = \frac{1}{2}\left[\left(\frac{\partial\varphi}{\partial t}\right)^2 + \left(\frac{\partial\varphi}{\partial\chi}\right)^2 + m^2\varphi^2\right]. \tag{11}$$

So we replace Eq. (6) by

$$s(:H:(p)) = \lim_{x_1, x_2\to p} \tfrac{1}{2}(\partial_{t_1}\partial_{t_2} + \partial_{\chi_1}\partial_{\chi_2} + m^2)[s(\varphi(x_1)\varphi(x_2)) - \omega(\varphi(x_1)\varphi(x_2))] , \tag{12}$$

where now Eq. (2) is replaced by

$$s(\varphi(x_2)\varphi(x_1)) = \frac{1}{2\pi R}\sum_{-\infty}^{\infty} \frac{\exp\{-i[\omega_n(t_2-t_1) - (n/R)(\chi_2-\chi_1)] - \epsilon|n|\}}{2\omega_n} ,$$
$$\omega(\varphi(x_2)\varphi(x_1)) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{dk}{2\omega_k} \exp\{-i[\omega_k(t_2-t_1) - k(\chi_2-\chi_1)] - \epsilon|k|\} , \tag{13}$$

where

$$\omega_k = (k^2+m^2)^{1/2}, \quad \omega_n = \left(\frac{n^2}{R^2} + m^2\right)^{1/2} .$$

If we leave the points separated only in the spacelike direction, we easily have

$$s(:H:(p))=\lim_{x\to 0}\left[\frac{1}{4\pi R^2}\,\Sigma\left(mR,\frac{\chi}{R}\right)-\frac{1}{4\pi}\int_{-\infty}^{\infty}dk(k^2+m^2)^{1/2}e^{ik\chi-\epsilon|k|}\right] \tag{14}$$

where

$$\Sigma(a,z)=\sum_{-\infty}^{\infty}(n^2+a^2)^{1/2}e^{inz-\epsilon|n|}$$

To get the short-distance behavior of $\Sigma(a,z)$, we write it as (dropping the ϵ, which is understood)

$$\Sigma(a,z)=\sum_{-\infty}^{\infty}(n^2+a^2)^{1/2}e^{inz}=a+2\left[\sum_{1}^{\infty}n\cos nz+\tfrac{1}{2}a^2\sum_{1}^{\infty}\frac{\cos nz}{n}+S(a)+O(z)\right], \tag{15}$$

where

$$S(a)=\sum_{1}^{\infty}\left[(n^2+a^2)^{1/2}-n-a^2/2n\right].$$

Using

$$\sum_{1}^{\infty}n\cos nz=-\tfrac{1}{4}\csc^2(\tfrac{1}{2}z)\quad\text{and}\quad\sum_{1}^{\infty}\frac{\cos nz}{n}=-\ln(2\sin\tfrac{1}{2}z),$$

this is

$$a-\frac{2}{z^2}-\tfrac{1}{6}-a^2\ln z+2S(a)+O(z). \tag{16}$$

Also, we have (Erdélyi,[18] p. 17, No. 27, and differentiating twice)

$$\int_{-\infty}^{\infty}dk(k^2+m^2)^{1/2}e^{ik\chi-\epsilon|k|}=\frac{2m}{x}K_0'(m\chi), \tag{17}$$

$$=-2/\chi^2-m^2\ln(\tfrac{1}{2}m\chi)-m^2(\gamma-\tfrac{1}{2})+O(\chi), \tag{18}$$

where K_0' is the derivative of a modified Bessel function of zero order and γ is Euler's constant.

With these results Eq. (14) gives

$$s(:H:(p))=\frac{1}{4\pi}\left[-\frac{1}{6R^2}+\frac{2}{R^2}S(mR)-m^2\ln\left(\frac{2}{mR}\right)+m^2(\gamma-\tfrac{1}{2})+\frac{m}{R}\right], \tag{19}$$

where $S(a)$ is defined after Eq. (15).

III. INTERACTING FIELDS

We take for our model a $\frac{1}{4}\lambda\varphi^4$ interaction

$$H(p)=\tfrac{1}{2}[(\partial_t\varphi)^2+(\nabla\varphi)^2+m^2\varphi^2+\tfrac{1}{2}\lambda\varphi^4], \tag{20}$$

where m may be zero or nonzero. We take a flat four-dimensional space-time with topology $\mathsf{R}\times\mathsf{R}^2\times\mathsf{S}$, the last space dimension being a circle of radius R.[19] We aim to calculate the Casimir effect to first order in perturbation theory. Naively one might try to simply apply Eq. (4) to Eq. (20), leading to a perturbation in the energy density of magnitude

$$\tfrac{1}{4}\lambda\lim_{\Lambda\to\infty}(s-\omega)(\varphi_\Lambda{}^4).$$

However, one soon sees that this quantity diverges.

The resolution of this apparent difficulty lies in a straightforward and simple application of renormalization theory.[20] As it stands, the term $\frac{1}{4}\lambda\varphi^4$ cannot be treated as a small perturbation, for, no matter how small λ is, φ^4 will have divergent expectation values.

In fact, we have [e.g., from point separation, see Eq. (A3)]

$$\varphi_\Lambda{}^4=:\varphi_\Lambda{}^4:+6\omega(\varphi_\Lambda{}^2)\varphi_\Lambda{}^2-3[\omega(\varphi_\Lambda{}^2)]^2. \tag{21}$$

We must therefore add to our regularized Hamiltonian density a counterterm

$$-\tfrac{1}{4}\lambda[6\omega(\varphi_\Lambda{}^2)\varphi_\Lambda{}^2]$$

to cancel the divergent-operator part of $\frac{1}{4}\lambda\varphi_\Lambda{}^4$.

In systematic Feynman-diagram language, the only one-particle-irreducible (1PI) divergent graph to first order is the tadpole graph (Fig. 2) giving rise to the same counterterm. m^2 is then the square of the "physical mass" and $m^2 - 3\lambda\omega(\varphi_\Lambda{}^2)$ is the square of the "bare" or unrenormalized mass to first order.

So the correct generalization of Eq. (4) to first order in λ is

$$s(:H:(p)) = \lim_{\Lambda\to\infty} (s-\omega)[H_\Lambda - \tfrac{3}{2}\lambda\omega(\varphi_\Lambda{}^2)\varphi_\Lambda{}^2]\,. \tag{22}$$

The reader may find it instructive to compare the above discussion with the closely parallel treatment of the "$\lambda\varphi^4$ kink mass" given in Rajaraman,[21] Sec. 3.4.

Applying formula (22) we get the usual Casimir effect for a free field $s(:H_0:(p))$ plus a term of order λ:

$$\tfrac{1}{4}\lambda \lim_{\Lambda\to\infty}(s-\omega)[-6\omega(\varphi_\Lambda{}^2)\varphi_\Lambda{}^2 + \varphi_\Lambda{}^4] \tag{23}$$

$$= \tfrac{1}{4}\lambda \lim_{\Lambda\to\infty}[-6\omega(\varphi_\Lambda{}^2)s(\varphi_\Lambda{}^2) + 6\omega(\varphi_\Lambda{}^2)\omega(\varphi_\Lambda{}^2) + s(\varphi_\Lambda{}^4) - \omega(\varphi_\Lambda{}^4)]\,. \tag{24}$$

Now, e.g., from point separation, see Eq. (A4) and (A5), we have

$$\omega(\varphi_\Lambda{}^4) = 3[\omega(\varphi_\Lambda{}^2)]^2, \quad s(\varphi_\Lambda{}^4) = 3[s(\varphi_\Lambda{}^2)]^2\,, \tag{25}$$

giving finally[22]

$$s(:H:(p)) = s(:H_0:(p)) + \frac{3\lambda}{4}\lim_{\Lambda\to\infty}[s(\varphi_\Lambda{}^2) - \omega(\varphi_\Lambda{}^2)]^2 + O(\lambda^2)\,. \tag{26}$$

Note that this formula is valid for arbitrary mass. We calculate this in the case of zero physical mass. Using the methods of Sec. II, we have

$$s(:H_0:(p)) = \lim_{t\to 0^+}\left(\frac{1}{2(2\pi)^3 R}\sum_{-\infty}^{\infty}\int d^2k\,\omega_{\underline{k},n}\exp[i\omega_{\underline{k},n}(t+i\epsilon)] - \frac{1}{2(2\pi)^3}\int d^3k\,\omega_{\underline{k}}^{+}\exp[i\omega_{\underline{k}}^{+}(t+i\epsilon)]\right), \tag{27}$$

where $\omega_{\underline{k},n} = (k_x{}^2 + k_y{}^2 + n^2/R^2)^{1/2}$, $\omega_{\underline{k}}^{+} = (k_x{}^2 + k_y{}^2 + k_z{}^2)^{1/2}$, where our regularization consists of separating points in a timelike direction.[23] This is easily calculated using the method given in Fierz[24] to be

$$s(:H_0:(p)) = -1/1440\pi^2 R^4\,. \tag{28}$$

To calculate the first-order perturbation, we use the same regularization, obtaining

$$s(\varphi_t)^2 - \omega(\varphi_t)^2 = \frac{1}{2(2\pi)^3 R}\sum_{-\infty}^{\infty}\frac{d^2k}{\omega_{\underline{k},n}}\exp[i\omega_{\underline{k},n}(t+i\epsilon)] - \frac{1}{2(2\pi)^3}\int\frac{d^3k}{\omega_{\underline{k}}^{+}}\exp[i\omega_{\underline{k}}^{+}(t+i\epsilon)]\,, \tag{29}$$

which, again by methods in Fierz,[24] gives

$$\left(-\frac{1}{4\pi^2 t} + \frac{1}{48\pi^2 R^2} + O(t)\right) - \left(-\frac{1}{4\pi^2 t}\right) = \frac{1}{48\pi^2 R^2} + O(t)\,, \tag{30}$$

thus yielding a first-order correction [Eq. (26)]

$$\frac{3\lambda}{4}\left(\frac{1}{48\pi^2 R^2}\right)^2 = \frac{\lambda}{3072\pi^4 R^4}\,. \tag{31}$$

These results are in agreement with those given for this model by Ford.[25] However, in contradiction to Ford, we have seen that point separation or mode-sum regularization give perfectly unambiguous results, provided we use them as part of a proper renormalization scheme.[26]

Finally, to illustrate the massive case, we return to our two-dimensional cylinder model (see Sec. II) and add a $\lambda\varphi^4$ self-interaction. Clearly, Eq. (26) still holds. Using methods which are by now familiar, we obtain a first-order correction:

$$\tfrac{1}{4}\lambda s(:\varphi^4:) = \frac{3\lambda}{4}\lim_{\Lambda\to\infty}[s(\varphi_\Lambda{}^2) - \omega(\varphi_\Lambda{}^2)]^2 \tag{32}$$

$$= \frac{3\lambda}{4}\lim_{\chi\to 0}\left[\frac{1}{4\pi R}\sum_{-\infty}^{\infty}\left(\frac{n^2}{R^2} + m^2\right)^{-1/2} e^{in\chi/R} - \frac{1}{4\pi}\int_{-\infty}^{\infty} dk(k^2+m^2)^{-1/2}e^{ik\chi}\right]^2 \tag{33}$$

$$= \frac{3\lambda}{4}\lim_{\chi\to 0}\left\{\frac{1}{4\pi}\left[\frac{1}{mR} + 2\sum_{1}^{\infty}\frac{\cos n\chi/R}{n} + 2C(mR) + O(\chi)\right] - \frac{2}{4\pi}K_0(m\chi)\right\}^2\,, \tag{34}$$

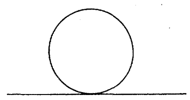

FIG. 2. The $\frac{1}{4}\varphi^4$ tadpole.

where

$$C(a)=\sum_1^\infty \left[(n^2+a^2)^{-1/2}-n^{-1}\right].$$

Using

$$\sum_1^\infty \frac{\cos nz}{n}=-\ln(2\sin\tfrac{1}{2}z)=-\ln z+O(z)$$

and

$$K_0(m\chi)=[-\ln(\tfrac{1}{2}m\chi)+\gamma]+O(\chi)\,,$$

we have finally

$$\tfrac{1}{4}\lambda s(:\varphi^4:)=\frac{3\lambda}{4(4\pi)^2}\left[\frac{1}{mR}+2\ln\left(\frac{mR}{2}\right)+2C(mR)-2\gamma\right]^2, \qquad (35)$$

where $C(a)$ is defined after Eq. (34). Note that this diverges as $m\to 0$, one expects this to be a special pathology of two dimensions (cf. Ref. 14).

IV. DISCUSSION

A. Relation with current formulations for renormalizing $T_{\mu\nu}$

A central concept throughout Secs. II and III was our "locality" (or "equivalence") principle that small regions in locally flat space-times may be identified with small regions in globally flat space-time. A second important concept was the physical renormalization condition that the vacuum expectation value of the energy density (and, more generally, of the energy-momentum tensor) should vanish in globally flat space-time. We saw how these two concepts, taken together, led to our basic equations (4) and (22). Now our Casimir-effect situations may be considered as special cases of the more general problem of renormalizing the energy-momentum tensor in a curved space-time. One therefore expects that these principles (or some replacement for them[27]) must play a role in that wider context. We discuss the situation as it is presently understood.

Wald's axiomatic approach.[28] Wald correctly states our renormalization condition as his Axiom 2.[29] However, his work lacks a replacement for our locality principle. Indeed, a careful examination of Wald's arguments[29] shows that he implicitly chooses, to begin with, a fixed space-time manifold and then considers $T_{\mu\nu}$ as a functional of all metrics on that manifold. For this reason, his Axiom 2 (validity of normal ordering in flat space-time) is ineffective for manifolds with nontrivial topology such as our cylinder.

The axioms of Ref. 9, then, only define $T_{\mu\nu}$ up to a constant multiple of the metric[30] and are thus incapable of fixing the Casimir effect. We should point out however, that the *point separation procedure* of Ref. 9 yields results on the Casimir effect in agreement with our own. Also, more recent developments of the axiomatic approach contain a possible approach to the Casimir effect.[31]

The point-separation approach.[32] To fix ideas, consider a massive field in a two-dimensional cylinder which we treated in Sec. II. If we simply separate points in the expression for $s(H(p))$ we would obtain the expression [see Eq. (16)]

$$\frac{1}{4\pi}\left[-\frac{2}{\chi^2}-m^2\ln\left(\frac{\chi}{R}\right)+\frac{m}{R}+\frac{2S(mR)}{R^2}-\frac{1}{6R^2}+O(\chi)\right]. \qquad (36)$$

In order to obtain a finite expression, we must subtract from this another expression with the same singularity structure before taking the limit $\chi\to 0$. In fact, we are told to subtract the Christensen-DeWitt expression. This Christensen-DeWitt expression is the universal state-independent singular part of $\langle T_{\mu\nu}\rangle$ and is only a function of local geometrical terms. However, as is well known, whenever it contains a logarithmically divergent piece, it is ambiguous because of the freedom to write

$$\ln\left(\frac{\chi}{\lambda}\right)=\ln\left(\frac{\chi}{\lambda'}\right)+\ln\left(\frac{\lambda'}{\lambda}\right), \qquad (37)$$

where λ,λ' are arbitrary length scales. If we refer back to Sec. II, we see that our renormalization condition essentially fixed the scale, telling us to subtract precisely [Eq. (18)]

$$\frac{1}{4\pi}\left[-\frac{2}{\chi^2}-m^2\ln(\tfrac{1}{2}m\chi)-m^2(\gamma-\tfrac{1}{2})\right]. \qquad (38)$$

Without this renormalization condition, $\langle T_{\mu\nu}\rangle$ would only have been fixed up to an arbitrary multiple of $g_{\mu\nu}$ and we would not know the Casimir effect.

This situation arises whenever we have logarithmic divergences. Now for massive fields, we have logarithmic divergences both in flat and curved space-times. In fact, we can take advantage of the flat situation described above to fix the scale once and for all, i.e., the Christensen-DeWitt series[11] is completely fixed by demanding that it gives zero when used to renormalize $\langle T_{\mu\nu}\rangle$ for the vacuum in Minkowski space-time. This explains why Christensen chooses precisely the term

$$\gamma+\tfrac{1}{2}\ln|\tfrac{1}{4}m^2(\sigma^\rho\sigma_\rho)|$$

in Eq. (6.4) of Ref. 11. Thus, we obtain unique values for $\langle T_{\mu\nu}\rangle$ for massive fields in curved space-times.

On the other hand, for massless fields we only have logarithmic divergences when there is non-vanishing local curvature. Thus, our flat-space-time renormalization condition cannot fix the scale, and we are forced to leave an arbitrary length scale; thus in the massless case $\langle T_{\mu\nu}\rangle$ can be defined only up to local geometrical terms.[33] The above discussion describes what is actually done in point-separation calculations. However, these things are not always made clear in the literature.[34] Indeed, the (erroneous) impression is sometimes given[34] that these issues are settled by renormalizing the cosmological constant.[35] Now it may well be that, in a fully quantized theory, the correct solution to these problems will come from renormalizing the cosmological constant[35] (and that the methods described here will fail to generalize). But in the semiclassical approximation where we write

$$G_{\mu\nu} + \Lambda g_{\mu\nu} = \kappa\langle T_{\mu\nu}\rangle , \tag{39}$$

we can never settle the ambiguity

$$\langle T_{\mu\nu}\rangle \to \langle T_{\mu\nu}\rangle + c g_{\mu\nu} \tag{40}$$

by measuring Λ.

The ζ-function approach.[36] The situation here is quite analogous to that for point separation. When there are no logarithmic divergences in the point-separation approach, the ζ-function method automatically throws away the divergent part. However, whenever point separation would give logarithmic divergences, the ζ-function result includes an unknown normalization constant μ with dimensions of inverse length.[37] Once again, we can fix this μ in the massive case by appealing to our renormalization condition (that the vacuum expectation value of $T_{\mu\nu}$ vanishes in globally flat space-time). But the ambiguity will remain for massless fields in curved space-times.

In conclusion to Sec. IV A, we wish to underline that the situation as it is presently understood is not a satisfactory one. There does not at present exist a comparable understanding of the generic case to that which we have given here for the special Casimir-effect cases of local flatness. First, as we pointed out above, we have no way in general of fixing the renormalization ambiguities for massless fields. Second, the procedure outlined above for fixing the ambiguity in the massive case can probably not be taken seriously either, when there is genuine local curvature. If it was a sensible procedure, one would expect to be able to fix the ambiguity in the massless case also by taking the massless limit of the result in the massive case. However, recent calculations suggest[38] that (for space-times which yield logarithmic divergences in the massless case) this limit can diverge

B. Other analytic regularization methods

One popular method[39] for obtaining the Casimir effect in the zero-mass case is to take (e.g., for the two-dimensional cylinder) the formal expression

$$s(H(p)) = \frac{1}{4\pi R}\sum_{-\infty}^{\infty}\frac{|n|}{R} = \frac{1}{2\pi R^2}\sum_{1}^{\infty} n . \tag{41}$$

One replaces this by

$$\frac{1}{2\pi R^2}\sum_{1}^{\infty} n^{-s} .$$

This will converge for sufficiently large s. Analytically continuing back to $s=-1$, one obtains[40]

$$\frac{1}{2\pi R^2}\zeta(-1) = -\frac{1}{24\pi R^2} . \tag{42}$$

It is then claimed that the Casimir effect is given by regularizing without the need to renormalize. However, in general, the method fails. Take, for example, a massive field in two-dimensional globally flat space-time (where the vacuum energy density should vanish). We write

$$s(H(p)) = \frac{1}{4\pi}\int_{-\infty}^{\infty}(m^2+k^2)^{1/2}dk .$$

Now

$$\int_0^\infty (m^2+k^2)^{-s/2}dk = m^{-s+1}\int_0^\infty \cosh^{-s+1}y\,dy = \frac{m^{-s+1}}{2}\,\frac{\Gamma((s-1)/2)\Gamma(1/2)}{\Gamma(s/2)} , \tag{43}$$

which diverges at $s\to -1$.

Another counterexample is obtained for the massive field on the two-dimensional cylinder. Here, one must deal with the discrete sum

$$\sum_{n=-\infty}^{\infty}(a^2+n^2)^{-s/2} .$$

The resulting generalized ζ function is more difficult to treat analytically but it appears also to have a pole at $s=-1$.[49] This failure of analytic regularization seems to occur whenever we would have logarithmic divergences in the point-separation case, though this relation seems to be not yet fully understood.[50]

C. Historical precedents

In Refs. 12, 41, and 42 will be found points of view on the Casimir effect which are related to—but not the same as—that given in this paper. It is interesting to note that in the Appendix to his important 1973 paper,[43] Fulling attempted to calculate essentially

$$\lim_{\Lambda\to\infty}(s-\omega)(H_\Lambda(p))$$

with a point-separation regularization for a massive field in a two-dimensional cylinder. As we showed in Sec. II, this yields the Casimir effect for this model. Unfortunately, Fulling's method failed to yield a clear-cut result, and the consensus of opinion was that this calculation has nothing to do with the Casimir effect.

V. CONCLUSIONS

We have developed a consistent, paradox-free, and indeed infinity-free framework for Casimir-effect calculations. We have seen, on the one hand, that the concepts of "zero-point energy" of "vacuum polarization," "virtual particles," etc., are not necessary. We have also seen that regularization alone fails, in general, to yield the effect.

Rather, we showed that the effect can be completely understood in the framework of ordinary flat-space-time quantum field theory with its unique vacuum state ω and its unique normal-ordering procedure. All the physical effects of nontrivial topology or conducting plates are coded into the state s of the field which is then studied locally. We also showed how this viewpoint can be consistently applied to treat self-interacting fields in perturbation theory.

Finally, our clearer understanding of the Casimir effect helped us to clarify some issues about renormalizing the energy-momentum tensor in a curved space-time.

Note added in proof.

1. The calculation of the Casimir effect for interacting fields presented in Sec. III is correct but depends on rather special properties of our model: In particular, Eq. (22) is justified only because $s(:H:(p))$ is proportional to the expectation value of the Hamiltonian, for which we have adapted the familiar perturbation theory result

$$\text{“}E_1 = \langle\psi_0|H_1|\psi_0\rangle\text{”}.$$

In general (for nontranslationally invariant situations and for other components of $T_{\mu\nu}$) there will be, in addition, a term involving the expectation value of the zero-order part of $T_{\mu\nu}$ in the first-order correction to the vacuum. A systematic method for calculating interacting-field Casimir effects for arbitrary components of $T_{\mu\nu}$ and to arbitrary order in perturbation theory will be given in a future paper. For a brief resume, see B. S. Kay, in Proceedings of 2nd Marcel Grossman Conference on General Relativity, Trieste 1979 (unpublished).

2. The conjecture at the end of Sec. IV (ii) is easy to resolve: Consider for example T_{00} expectation values. In mode-sum cutoff (or point-separation) schemes we have a formula such as

$$\langle|T_{00}|\rangle \sim \sum_n \omega_n e^{-\omega_n\lambda},$$

whereas the ζ-function definition depends on

$$\zeta_A(S) = \sum_n (\omega_n)^{-s}.$$

(A denotes the operator with eigenvalues ω_n.) We then have

$$\langle|T_{00}|\rangle = -\frac{d}{d\lambda}\sum_n e^{-\omega_n\lambda}$$

and

$$\zeta_A(s) = \frac{1}{\Gamma(s)}\int_0^\infty \lambda^{s-1}\sum_n e^{-\omega_n\lambda}$$

(for sufficiently large s). Now writing, e.g.,

$$\sum_n e^{-\omega_n\lambda} = \frac{A}{\lambda^2} + \frac{B}{\lambda} + X\ln\lambda + C + D\lambda + Y\lambda\ln\lambda + \cdots$$

and writing

$$\int_0^\infty = \int_0^1 + \int_1^\infty,$$

we obtain

$$\zeta(s) = \frac{1}{\Gamma(s)}\left[\frac{A}{s-2} + \frac{B}{s-1} - \frac{X}{s^2} + \frac{C}{s} + \frac{D}{s+1} - \frac{Y}{(s+1)^2} + \cdots\right],$$

whereupon if $Y=0$, $\zeta(-1)=D$ which is the same as throwing away the divergent part of $\langle|T_{00}|\rangle$. But if $Y\neq 0$, $\zeta(-1)=\infty$.

ACKNOWLEDGMENTS

I thank the Royal Society for a research fellowship and Professor Abdus Salam for hospitality at the International Centre for Theoretical Physics, Trieste, where this work was begun. I thank P. C. W. Davies, L. H. Ford, J. Gasser, P. Hajicek, S. W. Hawking, C. J. Isham, P. Minkowski, and R. F. Streater for conversations and encouragement. I should also like to thank S. A. Fulling who introduced me to this interesting topic.

APPENDIX

We give an alternative discussion of the model in Sec. II, in the language of axiomatic quantum field theory. Especially relevant is the more general viewpoint on what constitutes a *quantum state* made available by the C^* *algebra* approach to quantum field theory.[44] One of the fruits of the C^*-algebra approach has been clearer understanding of the origin of the ultraviolet divergences in quantum field theory. These are understood as being due to inequivalent representations of the field algebra. However, it has also long been conjectured[44] that these inequivalent representations are only of importance for representing *global* observables (e.g., total energy). As long as we restrict to *local* observables, all representations are "physically equivalent"[44] and the distinction between inequivalent representations should be unimportant.

We sketch, then, how the Casimir effect arises in this framework, without the appearance of infinities at any stage.

(a) The *only* representation for the field algebra we ever consider is the usual Poincaré-invariant Minkowksi-space representation.

(b) At the point p, in our cylinder space-time, we choose some sufficiently small space-time neighborhood N—with compact closure—and identify it in an obvious way with a neighborhood (also to be called N) of the origin in Minkowski space (see Fig. 1).

(c) The cylinder ground state s [see Eq. (2)] will restrict to a "partial state" $s|_N$ (in the sense of local algebras) on the algebra $\mathfrak{A}(N)$ of the region N (considered to be a region in two-dimensional Minkowski space, under the identification given above).

(d) This state $s|_N$ will *not* coincide with the Minkowski-space vacuum $\omega|_N$ on $\mathfrak{A}(N)$ [see Eq. (2)]. However, we conjecture that it will arise as a density matrix state ρ_s in the flat-space-time representation. Equivalently, we expect the flat-Minkowski- and cylindrical-space-time representations to be "locally quasi-equivalent."[44]

Once we have ρ_s, we can take the expectation value of the usual Minkowski-space *normal-ordered* energy-momentum tensor at p (suitably smeared as $:H:(f)$ supp $f \in N$)—there is no problem in defining this as a Wick polynomial (see below)—and we expect $:H:(f)$ to be in the domain of ρ_s and to be sufficiently regular for $\rho_s(:H:(p))$ to exist. We sketch how this may be calculated without needing to construct ρ_s. Wightman and Gårding[45] give the following formulas:

$$:\varphi^2:(x) = \lim_{x_1, x_2 \to x} [\varphi(x_1)\varphi(x_2) - \omega(\varphi(x_1)\varphi(x_2))], \tag{A1}$$

$$:\frac{\partial\varphi}{\partial x^\alpha}\frac{\partial\varphi}{\partial x^\beta}: = \lim_{x_1, x_2 \to x}\left[\frac{\partial\varphi}{\partial x^\alpha}(x_1)\frac{\partial\varphi}{\partial x^\alpha}(x_2) - \omega\left(\frac{\partial\varphi}{\partial x^\alpha}(x_1)\frac{\partial\varphi}{\partial x^\alpha}(x_2)\right)\right], \tag{A2}$$

whereupon we define $s(:H:(p))$ to be

$$\rho_s(:H:(p)) = \rho_s\{\lim_{x_1, x_2 \to x} \tfrac{1}{2}(\partial_{t_1}\partial_{t_2} + \partial_{\chi_1}\partial_{\chi_2})[\varphi(x_1)\varphi(x_2) - \omega(\varphi(x_1)\varphi(x_2))]\}$$
$$= \lim_{x_1, x_2 \to x} \tfrac{1}{2}(\partial_{t_1}\partial_{t_2} + \partial_{\chi_1}\partial_{\chi_2})[\rho_s(\varphi(x_1)\varphi(x_2)) - \omega(\varphi(x_1)\varphi(x_2))].$$

But since the points are now separated, we can replace ρ_s by s itself and recover Eq. (6).

Notes

(1) That $\rho_s(:H:(p))$ may be negative [Eqs. (10) and (19)] is nothing paradoxical, in fact, it is well known[46] that $:H:(p)$ cannot be a positive operator-valued distribution.

(2) The appearance of s as a density matrix state in flat space-time has led Ford to conjecture that s may be considered as an "imaginary-temperature state" on Minkowski space-time.[47] We hope to explore this idea in a later publication.[48]

Finally, we note here some more formulas from Wightman and Gårding[45] which we will need to refer to in Sec. III:

$$:\varphi(x_1)\varphi(x_2)\varphi(x_3)\varphi(x_4): = \lim_{x_1x_2x_3x_4 \to x} (\varphi(x_1)\varphi(x_2)\varphi(x_3)\varphi(x_4)$$
$$- \{[x_1x_2]\varphi(x_3)\varphi(x_4) + [x_1x_3]\varphi(x_2)\varphi(x_4) + [x_1x_4]\varphi(x_2)\varphi(x_3)$$
$$+ [x_2x_3]\varphi(x_1)\varphi(x_4) + [x_2x_4]\varphi(x_1)\varphi(x_2) + [x_3x_4]\varphi(x_1)\varphi(x_2) + [x_1x_2x_3x_4]\}), \tag{A3}$$

where $[x_1x_2] = \omega(\varphi(x_1)\varphi(x_2))$,

$$[x_1x_2\,x_3x_4]=\omega(\varphi(x_1)\varphi(x_2)\varphi(x_3)\varphi(x_4))$$
$$=[x_1x_2][x_3x_4]+[x_1x_3][x_2x_4]+[x_1x_4][x_2x_3]\,, \qquad \text{(A4)}$$

Also in analogy with Eq. (A4), we have for the state s

$$s(\varphi(x_1)\varphi(x_2)\varphi(x_3)\varphi(x_4))=\langle x_1x_2\rangle\langle x_3x_4\rangle+\langle x_1x_3\rangle\langle x_2x_4\rangle+\langle x_1x_4\rangle\langle x_2x_3\rangle\,, \qquad \text{(A5)}$$

where $\langle x_1x_2\rangle=s(\varphi(x_1)\varphi(x_2))$.

---

*Work supported in part by the Schweizerische Nationalfonds.

†Present address: Blackett Laboratory, Imperial College, London SW7 2BZ, England.

[1]For reviews, see L. Parker, in *Gravitation: Recent Developments*, 1978 Cargèse Lectures, edited by M. Lévy and S. Deser (Plenum, New York, to be published) (Milwaukee Report No. UWM-4867-78-9); and P. C. W. Davies, in *GRG Einstein Centennial Volume* (Plenum, New York, to be published) (King's College, London report).

[2]B. S. DeWitt, Phys. Rep. 19C, 297 (1975).

[3]H. B. G. Casimir, Proc. K. Ned. Akad. Wet. 51, 793 (1948); D. Tabor and R. H. S. Winterton, Proc. R. Soc. London, A312, 435 (1969).

[4]See, for example, Ref. 2 and T. H. Boyer, Ann. Phys. (N. Y.) 56, 474 (1970).

[5]See Sec. IV A.

[6]See Sec. IV B.

[7]To be fair, we should stress that these claims are made only for the locally flat case, and not when there is genuine curvature. For further discussion, see Sec. IV.

[8]L. H. Ford, Proc. R. Soc. London (to be published).

[9]R. M. Wald, Commun. Math. Phys. 54, 1 (1977).

[10]J. S. Dowker and R. Critchley, Phys. Rev. D 13, 3224 (1976); S. W. Hawking, Commun. Math. Phys. 55, 133 (1977); and L. Parker (see Ref. 1).

[11]B. S. DeWitt (Ref. 2); S. M. Christensen, Phys. Rev. D 14, 2490 (1976); and references in Ref. 1.

[12]Cf. L. H. Ford, Phys. Rev. D 11, 3370 (1975).

[13]We may think of this as a version of the equivalence principle—valid at least for locally flat space-times.

[14]Note: (1) These two-point functions may be easily derived heuristically by expanding the field in appropriate creation and annihilation operators: For s

$$\varphi(x)=(2\pi R)^{-1/2}\sum_{-\infty}^{\infty}\left(\frac{2|n|}{R}\right)^{-1}\times[a(x)\,e^{-i[(|n|/R)t-(n/R)\chi]}+a^\dagger(x)\,e^{i[(|n|/R)t-(n/R)\chi]}]\,.$$

For ω

$$\varphi(x)=(2\pi)^{-1/2}\int_{-\infty}^{\infty}dk\,(2|k|)^{-1/2}\,[b(x)\,e^{-i(|k|t-k\chi)}+b^\dagger(x)\,e^{i(|k|t-k\chi)}]\,,$$

and then forming $\langle 0|\,\varphi(x_2)\varphi(x_1)|\,0\rangle$ for $a(x)|\,0\rangle=0$, $b(x)|\,0\rangle=0$, respectively, remembering to put in an $i\epsilon$ factor in the resulting sum/integral [set $m=0$ in Eq. (13)] so that the resulting two-point function is defined correctly as a distribution which is the boundary value of an analytic function. (2) The unknown constants C, C' are a special feature of two dimensions and zero mass related to the ill-defined zero mode in the above equations. [See A. S. Wightman, in *Cargèse Lectures in Theoretical Physics (1964)*, edited by M. M. Levy (Gordon and Breach, New York, 1967)]. We can safely ignore this complication for the calculations in this section.

[15]See Refs. 2 and 16.

[16]P. C. W. Davies and S. A. Fulling, Proc. R. Soc. London A354, 59 (1977).

[17]S. W. Hawking, private communication.

[18]*Tables of Integral Transforms* (Bateman Manuscript Project), edited by A. Erdélyi (McGraw-Hill, New York, 1954), Vol. I.

[19]We avoid discussing here the separate issue of which is the "correct" energy-momentum tensor [see Ref. 2 and C. G. Callan, S. Coleman, and R. Jackiw, Ann. Phys. (N. Y.) 59, 42 (1970)]. For the translationally invariant states arising with our "boundary conditions" it will not make any difference.

[20]See, e.g., S. Coleman, in *Properties of the Fundamental Interactions*, edited by A. Zichichi (Editrice Compositori, Bologna, Italy, 1973), Vol. 9, Part B.

[21]R. Rajaraman, Phys. Rep. 21C, 227 (1975).

[22]We could also have obtained this result directly using Eqs. (21) and (26). Thus

$$s(:\varphi^4:)=3[(s(\varphi^2))^2-6\omega(\varphi^2)s(\varphi^2)+3(\omega(\varphi^2))^2]$$
$$=3[s(\varphi^2)-\omega(\varphi^2)]^2\,.$$

[23]This is easily seen to be essentially the same as a "mode-sum cutoff," cf. Refs. 4 and 12.

[24]M. Fierz, Helv. Phys. Acta 33, 855 (1960).

[25]L. H. Ford (see Ref. 8).

[26]Indeed the situation is rather the contrary to that claimed in Ref. 8. One expects the renormalization method given here to work quite generally, whereas Ford's "analytic regularization" method sometimes completely fails (see Sec. IV B).

[27]Whether our equivalance principle idea (cf. Ref. 13) can be generalized *directly* to curved space-times is an intriguing speculation, but one that we shall not pursue here.

[28]R. M. Wald, Ref. 9. See also Phys. Rev. D 17, 1477 (1978); and Ann. Phys. (N. Y.) 110, 472 (1978).

[29]See Ref. 9.

[30]Indeed, if we drop Wald's controversial "fifth axiom" we only know $T_{\mu\nu}$ up to a multiple of a conserved geometrical term.

[31]R. M. Wald, Ann. Phys. (N. Y.) 110, 472 (1978).

[32]See Ref. 11. Wald's work (Ref. 28) also contains an interesting variant of point separation.

[33]See, for example, T. S. Bunch and P. C. W. Davies, J. Phys. A 11, 1315 (1978) [quoted in Sec. 3.4 of P. C. W. Davies (Ref. 1)].

[34]See, e.g., Sec. 4.1 of P. C. W. Davies (Ref. 1).

[35]And in the general case, other coupling constants in some generalized Einstein's equations.

[36]See Ref. 10.

[37]See, e.g., Eq. 7.8 in S. W. Hawking (Ref. 10).

[38]P. C. W. Davies and W. Unruh, private communication.

[39]See, for example, L. H. Ford in Ref. 8; C. J. Isham, Proc. R. Soc. London A362, 383 (1978); J. R. Ruggiero, A. H. Zimerman, and A. Villani, Rev. Bras. Fís. 7, 663 (1977).

[40]Beware that this method is not the same as the ζ-function method discussed in Sec. IV A.

[41]L. S. Brown and G. J. Maclay, Phys. Rev. 184, 1272 (1969).

[42]J. S. Dowker and R. Critchley, J. Phys. A 9, 535 (1976).

[43]S. A. Fulling, Phys. Rev. D 7, 2850 (1973).

[44]R. Haag and D. Kastler, J. Math. Phys. 5, 848 (1964).

[45]A. S. Wightman and L. Gårding, Ark. Fys. 28, 129 (1964).

[46]H. Epstein, V. Glaser, and A. Jaffe, Nuovo Cimento 36, 1016 (1965).

[47]L. H. Ford, private communication.

[48]L. H. Ford and B. S. Kay (unpublished).

[49]L. H. Ford, private communication.

[50]For the connection between point separation and analytic regularization *when it does work*, see J. R. Ruggiero, A. Villani, and A. H. Zimerman, São Paulo Report No. IFT P-08/79 (unpublished).