

Conceptualization of the Casimir effect

D L Andrews and L C Dávila Romero

School of Chemical Sciences, University of East Anglia, Norwich NR4 7TJ, UK

E-mail: david.andrews@physics.org

Received 10 April 2001

Abstract

The origins and physical significance of the Casimir effect are reviewed, linking the zero-point energy of the vacuum in quantum electrodynamics with a force between conducting plates. It is shown how, by the use of dimensional and other simple physical arguments, the major features of the phenomenon can be derived.

1. Introduction

Hendrik Brugt Gerhard Casimir, whose name is known throughout the physics world, died on 4 May 2000. This brief article written in his honour concerns the discovery, formulation, physical significance and impact of one of the phenomena that bears his name, the eponymous *Casimir effect*. The discovery of this fundamental and very general effect is remarkable in the history of science for a host of reasons, and there is an extensive literature on the subject. To quote his original publication, ‘there exists an attractive force between two metal plates which is independent of the material of the plates. . .’, qualified by the condition that the intervening distance is sufficiently large ‘. . .that for the wavelengths comparable with that distance the penetration depth is small compared with the distance’. Such an attraction between two metal plates in vacuum has nothing to do with the forces of gravity or electrostatics. ‘This force may be interpreted as a zero point pressure of electromagnetic waves’ [1]. The existence of such a phenomenon, which even now continues to surprise those who encounter it for the first time, was at the time of its initial recognition by Casimir not at all an obvious possibility. Nonetheless, since the effect operates to some extent even in systems that only very loosely approximate to a pair of conducting plates, it involves a principle of wide relevance.

As a young man Casimir had been highly active in the developing field of quantum theory, benefiting from a training in physics successively under the tutelage of Ehrenfest, Bohr and Pauli. Before the Second World War, he worked at the University of Leiden on a much wider variety of topics, including the theory of Lie groups, thermodynamics, paramagnetic relaxation and electrical conduction. After moving to Philips Laboratories Casimir published his paper on the attraction of two plates in the wake of another landmark piece of science with which it has connections (*vide infra*), his work with Dik Polder on the long-range influence of retardation on the London–van der Waals dispersion forces between neutral atoms [2]. In the development of that theory, the principles of which are described elsewhere in this issue by Edwin Power

[3], the first step was the derivation of the potential energy of interaction between a neutral atom and an infinite perfectly conducting plate, which in SI units can be expressed as

$$\Delta E = -\frac{3\hbar c}{32\pi^2\epsilon_0} \frac{\alpha}{R^4} \quad (1)$$

where α is the atomic polarizability and R the distance between the atom and the plate. (In older systems of units where the factor of $4\pi\epsilon_0$ is absent, the polarizability has the dimensions of volume). It was the consideration of this primary relationship that subsequently led Casimir to focus directly on the attraction between two such plates, formulating the result for a force per unit area of magnitude:

$$P = \frac{\pi^2\hbar c}{240a^4} \quad (2)$$

where a is the separation between the plates. In passing we celebrate the fact that (2) is a rare and happy instance, in this branch of physics, of an equation that needs no conversion between any of the major systems of units. More significantly, however, it is a result that invokes no electrical constants.

The physical explanation that underpins the derivation of the above formula invokes the concept of quantization of the electromagnetic field. Quantum electrodynamics establishes that for each mode of the radiation field the associated energy levels are equally spaced (as is true for any harmonic oscillator) and each level is associated with an integer number of photons. Again, from the quantum treatment of harmonic motion, the lowest level for each such mode has a finite zero-point energy, a physical manifestation of quantum uncertainty. For the electromagnetic field in free space, there exist an infinite number of radiation modes, each contributing to an infinite zero-point energy. However in the presence of two parallel, perfectly conducting plates, only longitudinal waves that have nodes at each surface are supported (at the microscopic level much the same principle is at work in the modern theory of photonic crystals). There is no effective constraint on shorter wavelengths, since each plate ‘. . . is hardly an obstacle at all and therefore the zero point energy of these waves will not be influenced by the position. . .’ [1]. The net (still infinite) zero-point energy between the plates is nonetheless countably modified through the presence of the plates. The energy shift has to be interpreted in terms of a lowering in the potential energy as the plates are moved towards each other from infinite separation. Relating this to the force required to separate the plates results in an expression for the pressure exerted.

2. Didactic approaches to the formula

In Casimir’s derivation, which in common with many subsequent treatments of the interaction is relatively short but mathematically quite intricate, the material properties of the plates are fully encompassed by the twin assumptions of perfect conductivity and planarity; neither charge nor electric polarization is involved. The familiar approximation methods of perturbation theory based on expansions in powers of the electron charge are not well suited to dealing with such a task, and in the calculational procedure some relatively unusual procedures had to be brought into play. In fact it is notable that the *basis* for the calculation entails no further assumptions at all, although calculational approximations are involved in securing a simple form for the final result. A large number of more or less equally demanding theoretical treatments is now to be found in the extensive literature of the subject, and it is not the intention to reproduce any such derivation here—the reader is instead referred to the excellent accounts by Power [4] and by Milonni [5], for example. Instead it is our purpose as a didactic exercise to see how far one can get using simple physical and dimensional arguments—especially since some purposely over-simplified treatments have been shown to give results of similar form but having slightly different numerical factors. Certainly it is easy enough, given the premise of a mechanism for such an interaction, first to determine its sign.

It is considered that the presence of the two parallel conducting plates imposes conditions on the vacuum radiation modes that can be supported in their midst, the region of space commonly called the *cavity*. The premise is that, because the cavity cannot support modes whose half-wavelength will not fit an integer number of times inside it, the longest supportable wavelength is of half the plate separation, and all other allowed modes are correspondingly higher order harmonics—the detailed calculation has also to take account of the transverse mode structure. For the excluded modes the corresponding zero-point energies associated with vacuum fluctuations are therefore absent, whereas in the absence of the plates such modes would make their usual free-space contributions to the vacuum energy. So the presence of the plates engenders an energy loss $-\Delta E$ in the cavity, signifying a negative potential energy U . On physical grounds this must reduce with distance, so that U has to be a negatively sloped monotonic function of the plate separation a . As is verified below, this means that the distance dependence of the potential energy has the functional form of an inverse power law, and consequently the corresponding force $F = -\int U da$ itself results in a negative quantity, signifying attraction between the plates. (More precisely $F = -\int U da$, where a is the displacement vector of one plate with respect to the other and F is the force the former plate experiences).

On seeking the detailed form of the force equation, dimensional arguments can profitably be brought into play. Such arguments are of unusual transparency in this case, precisely *because* the phenomenon entails no electrical properties (such as the electric charge or permittivity that arise in connection with dispersion forces, for example). This is why the Casimir formula is cast in exactly the same form irrespective of the system of units employed. Suppose then that the functional form is sought for the quantum force between the parallel plates. For fairly obvious physical reasons the net force must be dependent on the plate area, and it can be anticipated that the corresponding pressure is area independent—although this is an assumption we return to subsequently. So, the only variable in terms of which this pressure P can be cast is the separation a . As a quantum effect the result must invoke Planck's constant \hbar and, also, conceivably the speed of light c , since the latter is the only other independent universal constant involved with electromagnetic radiation. If we write $P = a^x c^y \hbar^z$ and use dimensional analysis, then since $[P] = [M][L]^{-1}[T]^{-2}$, $[a] = [L]$, $[c] = [L][T]^{-1}$, $[\hbar] = [M][L]^2[T]^{-1}$, we obtain the following simultaneous equations for $[M]$, $1 = z$; for $[L]$, $-1 = (x + y + 2z)$; and for $[T]$, $-2 = -(y + z)$. Hence the result is of the form $P \sim \hbar c a^{-4}$.

If the plate area l^2 were to be also included as a possible variable, dimensional analysis would yield the incompletely determined form $P \sim \hbar c a^x l^{2w}$, with $x + 2w = -4$. For symmetry reasons the area of not just one plate but *both* would have to play the same role—and with two plates of the same area, w would have to be even. It is inconceivable that w could be positive, since then the effect could easily yield immense forces with even moderately large plates, and any such effect operative over macroscopic scales would have long been well known. However with w even and negative, then unless x was zero, which would signal the unreasonable possibility of a distance-independent result, x would have to be positive, giving the equally unpalatable prospect of a force increasing with distance. Since these alternative scenarios are so clearly unphysical there is no possibility for doubt that a force per unit plate area is the requisite form for the result as given above. It then follows that the force on each plate is given by $F \sim \hbar c l^2 a^{-4}$ and the corresponding potential energy of interaction is $U \sim \hbar c l^2 a^{-3}$. As shown above, it ensues that both F and U must carry negative signs.

Although a multiplicity of other physical grounds for the Casimir–Polder interaction can be argued, the basic premise for the Casimir effect is remarkably robust, namely the zero-point energy of the vacuum. The only significantly different route to the result is one based instead on the notion of a vacuum photon momentum, interpreting the force on each plate in terms of the reduction (owing to the cavity exclusion of insupportable modes) of a vacuum Maxwell–Bartoli radiation pressure. The premise here is one that is less generally useful or indeed measurable—since the vector character of the envisaged $\frac{1}{2}\hbar\mathbf{k}$ zero-point momentum per mode of wave-vector \mathbf{k} necessarily sums to zero [6]. Nonetheless the concept again affords

a transparent link to the functional form of equation (2). Consider a cubic cavity of side a . Any particles of momentum p bouncing back and forth between opposite walls would at each wall impact impart a momentum change $2p$, with a repeat interval given by the round-trip time $2a/c$. From Newton's Second Law the associated time-averaged force is pc/a and the force per unit area is pc/a^3 . Elementary quantum considerations dictate that, since the system has a localization uncertainty a , then even with no particles present there is a residual linear momentum uncertainty $\sim\hbar/a$, with an ensuing uncertainty in the force per unit area $\sim\hbar c/a^4$. As with the notion of zero-point energy itself, such an uncertainty for a quantity which classically should be zero reflects a finite physical value, and the result gives a clear indication of its form.

3. Physical significance and impact

The links that exist between the theory of the Casimir effect and the Casimir–Polder dispersion interaction between atoms are less transparent than might at first be suspected. The latter exhibits retardation effects that modify the form of the distance dependence at large distances, owing to the finite time of signal propagation; the former does not. In its quantum mechanical formulation the latter generally invokes perturbation theory, whereas the former does not. (Although classical formulations of the Casimir–Polder interaction appear to circumvent such approximations, the concept of atomic polarizability thereby entailed is itself an approximation based on precisely linear optical response.) A *perfectly* polarizable atom, whose polarizability in old-fashioned units would equate to its physical volume, corresponds to a classical object with a continuum of energy levels. As such the onset of retardation effects associated with inter-atomic interactions would become significant only at an essentially infinite range. Indeed, if the non-retarded (short-range) London form of the interatomic potential is employed as a basis for pairwise summing of interactions between atomic components of the plates, the result does indeed run with a^{-4} . Nonetheless, even then the summing of interactions between atomic components of the two plates fails to deliver the correct numerical factor because it is based only on pairwise interactions [7].

It is interesting to reflect on the fact that the original prediction of the Casimir effect should come from the hand of an industrial scientist—perhaps a little less surprising at the time than would be the case today, when direct profitability is the universal byword. However, although it may have been an unlikely provenance for this particular work, Philips Laboratories in Eindhoven has continued to nurture a great deal of other fundamental research, a later example being the landmark series of publications on radiation trapping in atomic vapours by van Trigt—see for example [8] and references therein. It is certainly notable that Casimir developed his theory at a time when the development of quantum electrodynamics was still at a very early stage. Moreover his result antedated by nearly a decade the first (not entirely convincing) attempts at experimental detection [9–11]—see also [12]. Casimir had himself appeared somewhat reticent on the likelihood of experimental verification, and it is clear that attempted measurements were fraught with a host of experimental difficulties, not least the obviation of electrostatic and gravitational interference. As such the phenomenon appears to have acquired the connotation of being academically interesting but fundamentally esoteric, and it is perhaps not so surprising that it was not until 1996 that the first definitive measurements verified the existence of a force exactly as Casimir had predicted [13, 14]¹.

Admitting the reality of the zero-point electromagnetic energy of the vacuum is conceptually, mathematically and inextricably linked with acceptance of the notion of *vacuum fluctuations* [15], which play a key role in the quantum electrodynamical explanation of *spontaneous emission* [16, 17]. The fact that numerous alternative means of explicating that much more familiar process have been sought over many years, since semiclassical theory

¹ To avoid the problem of keeping two flat surfaces parallel, experimental measurements have usually been made on the forces between one plane and a sphere, introducing a different functional form to the Casimir force.

fails at this point, is a powerful indicator of a general and deeply entrenched reluctance to accept the reality of zero-point energies. Such a position is of course no longer tenable. The Casimir effect has implications far wider than the somewhat artificial system it addresses: it is now known to have very wide implications in elementary particle physics and cosmology [18]. However, it is already clear that accepting the reality has opened a Pandora's box containing a host of other entertaining diversions, including some highly fanciful propositions for exploiting the zero point energy for free-energy devices, the spontaneous generation of energy, perpetual motion and the like [19].

Perhaps more realistic is recent work suggesting the utilization of the Casimir effect in nanotechnology where, as miniaturization accelerates, it may prove fundamentally significant for device design and operation. Thus it has been recognised that the effect should have a noticeable influence on the performance of molecular sized machines [20], where the scale of distance between components is such that the pressures exerted can run into atmospheres. For example simple application of equation (2) to a pair of plates separated by 10 nm shows that each plate experiences a negative pressure of 1.3 atm. At such a distance a 1 nm displacement would nonetheless change the pressure by about 40%. Even the atomic irregularity of a 'perfect' metal plate needs to be brought into consideration [20]. Recent work in Bell Laboratories has shown that, exploiting such effects, it may for example prove necessary to modify the performance characteristics of ultrasensitive switches designed to respond to subtle background variations. Recalling again the provenance of the Casimir effect it is heartening to see that that the phenomenon is now being brought forward into practical utility, with commercial relevance finally in sight.

References

- [1] Casimir H B G 1948 *Proc. Kom. Ned. Akad. Wetensch. B* **51** 793
- [2] Casimir H B G and Polder D 1948 *Phys. Rev.* **73** 360
- [3] Power E A *Eur. J. Phys.* **22** 453
- [4] Power E A 1964 *Introductory Quantum Electrodynamics* (London: Longman) p 34 and p 131
- [5] Milonni P W 1994 *The Quantum Vacuum: An Introduction to Quantum Electrodynamics* (San Diego, CA: Academic) p 54
- [6] Milonni P W and Eberly J H 1988 *Lasers* (New York: Wiley)
- [7] Lifshitz E M 1956 *Sov. Phys.-JETP* **2** 73
- [8] van Trigt C 1976 *Phys. Rev. A* **13** 726
- [9] Derjaguin B V and Abrikossova J J 1957 *Sov. Phys.-JETP* **3** 819
- [10] Derjaguin B V, Abrikossova J J and Lifshitz E M 1956 *Q. Rev. Chem. Soc.* **10** 295
- [11] Sparnaay M J 1958 *Physica* **24** 751
- [12] Elizalde E and Romeo A 1991 *Am. J. Phys.* **59** 711
- [13] Lamoreaux S K 1997 *Phys. Rev. Lett.* **78** 5
Lamoreaux S K 1998 *Phys. Rev. Lett.* **81** 5475
- [14] Mohideen U and Roy A 1998 *Phys. Rev. Lett.* **81** 4549
- [15] Jaekel M-T and Reynaud S 1997 *Rept. Prog. Phys.* **60** 863
- [16] Mandel L and Wolf E 1995 *Optical Coherence and Quantum Optics* (Cambridge: Cambridge University Press) p 764
- [17] Loudon R 2000 *The Quantum Theory of Light* 3rd edn (Oxford: Oxford University Press) p 286
- [18] Milonni P W and Shih M-L 1992 *Contemp. Phys.* **33** 313
- [19] Yam P 1997 *Sci. Amer.* **277** 82
- [20] Chan H B, Aksyuk V A, Kleiman R N, Bishop P J and Capasso F 2001 *Science* **291** 1941