PHYSICAL REVIEW LETTERS

VOLUME 63 | 20 NOVEMBER 1989 | NUMBER 21

Aharonov-Bohm Effect and the Mass of the Photon

David G. Boulware
Physics Department FM-15, University of Washington, Seattle, Washington 98195

S. Deser
Physics Department, Brandeis University, Waltham, Massachusetts 02254
(Received 16 May 1989; revised manuscript received 25 July 1989)

We show that the Aharonov-Bohm effect requires that the vector potential couple minimally to matter, not that it be a gauge field: The effect is present in massive (finite-range) electrodynamics, reducing smoothly to the original result in the limit of infinite range. Indeed, it may be used to provide an experimental bound on the range which is much larger than the "table-top" apparatus, namely a lower limit of order 10^2 km.

PACS numbers: 03.65.Bz, 12.20.Ds, 12.20.Fv, 14.80.Am

The vast literature on the Aharonov-Bohm effect (see, for example, Ref. 1) is explicitly based on the gauge nature of the electromagnetic field. To our knowledge, there have been no discussions of whether it might exist for a nongauge field such as finite-range electrodynamics (in quantum language, for finite photon mass). There is also a literature on massive electrodynamics (e.g., Refs. 2 and 3) which likewise does not include any analyses of the Aharonov-Bohm effect, let alone whether its infinite-range limit is smooth.[4]

We will show that the effect indeed exists for finite mass, has a smooth limit, and may even be exploited to provide a bound on the photon's mass: Precision experiments could observe a range as high as 10^2 km.

The effect.— We formulate the Aharonov-Bohm effect in the following terms. In the semiclassical limit, the wave function of a charged particle contains the path-dependent phase $\exp(ie\int \mathbf{A}\cdot d\mathbf{x})$, which means that when a beam is split and recombined, there will be a phase shift of the form ($\hbar = 1 = c$)

$$\exp\left[ie\oint_c \mathbf{A}\cdot d\mathbf{x}\right] \equiv \exp(i\Phi)\,. \qquad (1)$$

The flux

$$\Phi \equiv e\oint_c \mathbf{A}\cdot d\mathbf{s} \equiv e\int \nabla\times\mathbf{A}\cdot d\mathbf{S} \equiv e\int \mathbf{B}\cdot d\mathbf{S} \qquad (2)$$

is through any surface bounded by the closed curve c defined by the two paths. Although this phase factor is usually invoked from gauge invariance of the theory, it is simply a consequence of minimal ($\partial_\mu \to \partial_\mu - ieA_\mu$) coupling to the vector field A_μ rather than of the specific action for A_μ. We will stay within the same minimal-coupling framework here (since Pauli terms are in any case irrelevant to the effect) so that the currents remain conserved. Thus, the value of Φ (mod2π) will be observable in the interference pattern produced by the recombined beams. The transverse (nonlocal and gauge-invariant) vector potential $\mathbf{A}^T$ which contributes to Φ is uniquely determined by $\mathbf{B}$ according to $-\nabla^2\mathbf{A}^T = \nabla\times\mathbf{B}$. Consequently, if there is a nonvanishing magnetic field $\mathbf{B}\equiv\nabla\times\mathbf{A}$ anywhere in space (even one confined to a bounded region), its contribution to $\mathbf{A}^T = -\nabla^{-2}\nabla\times\mathbf{B}$ is nonvanishing almost everywhere. Note that the above discussion is completely independent of the action governing A_μ.

We turn next to the equations which determine $(\mathbf{A}^T,\mathbf{B})$. If the field is of finite range, it obeys the Proca equation

$$\partial_\nu F^{\mu\nu} + m^2 A^\mu = J^\mu\,, \qquad (3)$$

where $F_{\mu\nu} \equiv \partial_\mu A_\nu - \partial_\nu A_\mu$, and our metric is $(-+++)$.

For stationary currents and magnetic fields, this means

$$\nabla\times\mathbf{B}+m^2\mathbf{A}=\mathbf{J}\,, \quad (4)$$

$$(m^2-\nabla^2)A_0=0\,. \quad (5)$$

Since $\nabla\cdot\mathbf{J}=0$, this implies vanishing longitudinal vector (and scalar) potentials. (Physical longitudinal photons, which are the hallmark of the $m\neq 0$ theory, are well known to decouple from matter in the $m\rightarrow 0$ limit, provided that the current is conserved. If the current is not conserved, then the coupling to the longitudinal photons becomes strong as $m\rightarrow 0$ and only *extremely* weak coupling is allowed experimentally.[2]) The relevant equation is then

$$(-\nabla^2+m^2)\mathbf{A}^T=\mathbf{J}^T\,; \quad (6)$$

for the usual solenoidal configuration, it reads

$$(-\nabla^2+m^2)\mathbf{A}^T=\hat{\boldsymbol{\phi}}j\delta(\rho-a)=\nabla\times j\hat{\mathbf{z}}\theta(a-\rho)\,, \quad (7)$$

where j is the current per unit length in the surface of an infinite cylinder of radius a, $(\hat{\boldsymbol{\phi}},\hat{\mathbf{z}})$ are unit vectors, θ is the step function, and ρ is the radius of the x-y plane. Writing $\mathbf{A}=\hat{\mathbf{z}}\times\nabla\Pi(\rho)$, we obtain the scalar equation

$$(\partial_\rho^2+\rho^{-1}\partial_\rho-m^2)\Pi(\rho)=j\theta(a-\rho)\,. \quad (8)$$

The relevant homogeneous solutions are the Bessel functions I_0, K_0 which are regular at the origin and at infinity, respectively:

$$x\rightarrow 0:\; I_0(x)\sim 1+x^2/4,\; K_0\sim -\ln(x/2)\,,$$
$$x\rightarrow \infty:\; I_0\sim(2\pi x)^{-1/2}e^{x},\; K_0\sim(\pi/2)^{1/2}e^{-x}\,. \quad (9)$$

Consequently, the Green's function is just

$$G(\rho,\rho')=I_0(m\rho_<)K_0(m\rho_>) \quad (10)$$

and Π is given by

$$\Pi(\rho)=-j\left[\theta(\rho-a)K_0(m\rho)\int_0^a\rho'd\rho'I_0(m\rho')+\theta(a-\rho)\left(K_0(m\rho)\int_0^\rho\rho'd\rho'I_0(m\rho')+I_0(m\rho)\int_\rho^a\rho'd\rho'K_0(m\rho')\right)\right]\,. \quad (11)$$

The magnetic field therefore reads

$$\mathbf{B}=\nabla\times\mathbf{A}=\hat{\mathbf{z}}\nabla^2\Pi=\hat{\mathbf{z}}j\theta(a-\rho)+\hat{\mathbf{z}}m^2\Pi(\rho)\,. \quad (12)$$

The first term on the right-hand side of (12) is, of course, just $\mathbf{B}_0$, the usual Maxwell field. It is clear from (11) that $m^2\Pi$ vanishes in the $m\rightarrow 0$ limit since I_0 is regular and K_0 is logarithmic at vanishing argument (for fixed ρ).

We have thus established continuity of the Aharonov-Bohm effect as the range increases (for fixed ρ,a): The magnetic field reduces to its Maxwell value B_0 in the region $\rho<a$ and to 0 outside. What about the corrections? Clearly, since there is now a length m^{-1} in the problem, there are three possible cases (we assume as usual that the wave function is excluded from the solenoid, i.e., that $\rho>a$), depending on whether the range is shorter than a, between a and ρ, or greater than ρ. The first two possibilities are physically irrelevant since we know from the absence of an observed Yukawa decay of the terrestrial magnetic field[3] (let alone astrophysical considerations) that the range is much larger than the laboratory scales involved in ρ. Nevertheless, we record $m^2\Pi$ in these two cases:

$$m\rho\gg 1\gg ma:\; m^2\Pi\sim(\pi/8m\rho)^{1/2}(ma)^2e^{-m\rho}\,,$$
$$m\rho\gg ma\gg 1:\; m^2\Pi\sim(a/\rho)^{1/2}e^{-m(\rho-a)}\,. \quad (13)$$

We now concentrate on the interesting case of long range: $1\gg m\rho>ma$. There are now corrections to $\mathbf{B}_0$ both inside and outside the solenoid:

$$\Delta\mathbf{B}=\hat{\mathbf{z}}m^2\Pi\sim\hat{\mathbf{z}}[\theta(\rho-a)c+\theta(a-\rho)D],\; 1\gg m\rho\gg ma\,,$$

$$C\equiv-(j/2)(ma)^2\ln(m\rho/2)\,, \quad (14)$$

$$D\equiv(j/2)[-\tfrac{1}{2}(ma)^2+\tfrac{1}{2}(m\rho)^2-(ma)^2\ln(ma/2)]\,.$$

The magnetic field "leakage" C outside the solenoid is an extremely small fraction of B_0, namely of order $\frac{1}{2}(ma)^2\ln(m\rho/2)$; it represents the closing of the field lines at finite ($\sim m^{-1}$) distance. The fractional change in the internal field is also small, of fractional order $\frac{1}{2}(m\rho)^2$. Of course the coupling is through $\nabla-ie\mathbf{A}$ and the Hamiltonian is a function of $\mathbf{A}$ rather than $\mathbf{B}$. Although the local exterior $\mathbf{B}$ is in any case extremely small, it is the nonlocal $\mathbf{A}^T$ which carries the interaction effect and which is *not* small.

The contributions $\Delta\Phi$ to the flux Φ inside the observation circle ρ come from both regions. The interior contribution is given by

$$\Delta\Phi_i=2\pi\int_0^a\rho'd\rho'D(\rho')\sim-(\pi/2)ja^2(ma)^2\ln(ma/2)\,, \quad (15)$$

where we have only kept the leading, logarithmic, part. The leading exterior correction,

$$\Delta\Phi_e=2\pi\int_a^\rho\rho'd\rho'C(\rho')$$
$$\sim(\pi/2)j(ma)^2[\rho^2\ln(m\rho/2)-a^2\ln(ma/2)]\,, \quad (16)$$

is the dominant new effect $[\Delta\Phi_i\sim(a/\rho)^2\Delta\Phi_e]$. Now the Maxwell flux Φ_0 is, of course, $\Phi_0\equiv\int\mathbf{B}_0\cdot d\mathbf{S}=\pi a^2 j$, so that the fractional correction is

$$\Delta\Phi/\Phi_0\sim-\tfrac{1}{2}(m\rho)^2\ln(m\rho/2)=(m\rho/2)^2\ln(2/m\rho)^2\,, \quad (17)$$

which is essentially quadratic in the small parameter $m\rho$. Note that varying ρ is equivalent to "varying" m. Equation (17) is our basic result for the correction to the usual effect due to finite (but large) range.

In the case of a finite apparatus, this result is modified. The logarithmic dependence upon ρ remains, but it is scaled by some factor of order of the size of the apparatus. In general, the correction must be evaluated numerically; however, for a circular flux line of radius R, the correction to the flux through a small circle of radius $\rho \ll R$ enclosing the flux line may be written in terms of elliptic integrals. The approximate result is

$$\Delta\Phi/\Phi_0 \sim (m\rho/2)^2[\ln(R^2/\rho^2)+O(1)]\,. \tag{18}$$

We have seen that the nonlocal effects of a magnetic field in normal electrodynamics are simply ascribed to the fact that the physical gauge-invariant transverse vector potential $\mathbf{A}^T$ is a nonlocal function of $\mathbf{B}$, and that it is $\mathbf{A}^T$ which couples to the wave function. These facts are independent of the dynamics governing the A field, and persist whatever its range. For the standard solenoid configuration, finite range does lead to a small $\mathbf{B}$ field leakage outside the solenoid, as seen in (14); for the physical situation in which $1 \gg m\rho > ma$, the additional flux is dominated by the contribution from the leakage field, although the interior field corrections also contribute. This does not mean that the leakage field is the cause of the effect here; indeed, one could even arrange in principle for the leakage flux addition to be an integer, in which case only the interior field would contribute, making the picture just as nonlocal as in the massless case. Thus, the standard "paradoxical" features of the Aharonov-Bohm effect are reproduced by the massive, nongauge, field. We also noted that, as might be expected, the $m \to 0$ limit yields the $m=0$ results.

Experimental bounds on the photon mass.— We now show that some nontrivial limits on the range of the transverse photon can be obtained from a table-top experiment. (Experimentally, limits on the electric field range from electrostatics are a quite different story, based on deviations from the Gauss law. The best of these limits based on the longitudinal field is $m^{-1} \geq 3.1\times 10^4$ km.[5]) In the typical experiment, the magnetic field is constrained to a small "whisker." If the whisker consists of a type-II superconductor, then the flux quantization of the individual vortex lines will assure that Φ_0 will be an integer multiple of 2π, so that there is a "null" effect in Maxwell theory. A naive application of the Ginzburg-Landau theory modified to include finite-range electrodynamics still yields flux quantization and we conjecture that a more sophisticated treatment will yield a fractional change in the shape of the free energy as a function of B of order $(ml)^2$, where l is the size of the flux vortex (l^{-1} is the effective London photon "mass").

To obtain an experimental limit on the range m^{-1}, one must detect a phase shift which is some fraction ϵ of 2π. Fortunately, since Φ_0 does not contribute when the "solenoid" is a finite superconducting whisker, this means that the correction $\Delta\Phi$ gives the whole effect, and so we must have

$$\Delta\Phi \geq 2\pi\epsilon\,. \tag{19}$$

By (17), this means that the value of $m\rho$ at which the effect is just observable is

$$(m\rho/2)^2 = 2\pi\epsilon/\Phi_0 A\,, \tag{20}$$

where A depends upon the detailed experimental configuration but is of order 1. We consider a whisker of radius $a \sim 0.1$ cm, a magnetic field of 10 T, and a resolution $\epsilon \sim 10^{-3}$. This yields, as the longest range which could be observed,

$$m^{-1} \simeq 14\rho \text{ km}\,, \tag{21}$$

where ρ is measured in cm. For a coherent beam of $\rho \sim 10$ cm, the observable ranges are then $\lesssim 10^2$ km. Naturally there will be uncertainties due to, e.g., stray fields which vary with the flux in the whisker. Even if we have been overly optimistic in choosing our parameters, it is nonetheless interesting that a reasonably large lower bound on the range of electrodynamics can be obtained from a precision experiment of small dimensions. An analogous experiment has already been performed by Tonomura.[1] However, the geometry and field strengths obtained are not optimal for our purposes.

It should be emphasized that the proposed experiment probes the deviation of the magnetic field from its classical value at short distances. No measurements are proposed at a distance scale of m^{-1}. The precision is obtained only because it is a null experiment for $m=0$ combined with the use of large magnetic fields and the sensitivity obtainable by detecting phase shifts.

We thank Ya. Smorodinskii for asking the right questions, R. Schrieffer for discussions on flux quantization in massive electrodynamics, and B. Heckel for providing estimates on some of the parameters. This work was supported by the U.S. Department of Energy under Contract No. DE-AS06-88ER-40423 and by NSF Grant No. PHY88-04561. S.D. thanks Boston University for its hospitality while this work was begun.

[1]Y. Aharonov and D. Bohm, Phys. Rev. **115**, 485 (1959); W. Ehrenberg and R. E. Siday, Proc. Phys. Soc. London, Sect. B **62**, 8 (1949); C. N. Yang, in *Proceedings of the International Symposium on the Foundations of Quantum Mechanics, Tokyo, 1983,* edited by S. Kamefuchi *et al.* (Physical Society of Japan, Tokyo, 1984); S. Olariu and I. Popescu, Rev. Mod. Phys. **57**, 339 (1985); A. Tonomura *et al.*, Phys. Rev. Lett. **56**, 792 (1986).

[2]L. Bass and E. Schrödinger, Proc. Roy. Soc. London A **232**, 1 (1955); E. C. G. Stueckelberg, Helv. Phys. Acta **30**, 209 (1957); D. G. Boulware and W. Gilbert, Phys. Rev. **126**, 1563 (1962); S. Deser, Ann. Inst. Henri Poincare **16**, 79 (1972).

[3]A. S. Goldhaber and M. M. Nieto, Rev. Mod. Phys. **43**, 277 (1971).

[4]H. Van Dam and M. Veltman, Nucl. Phys. **B22**, 297 (1970); Y. Iwasaki, Phys. Rev. D **2**, 2225 (1970); V. I. Zaharov, Pis'ma Zh. Eksp. Teor. Fiz. **12**, 447 (1970) [JETP Lett. **12**, 312 (1970)]; D. G. Boulware and S. Deser, Phys. Lett. **40B**, 227 (1972); Phys. Rev. D **6**, 3368 (1972).

[5]E. R. Williams, J. E. Faller, and H. A. Hill, Phys. Rev. Lett. **26**, 721 (1972).