

Origin of the Bohm-Aharonov Effect with Half Flux Quanta

John P. Carini, K. A. Muttalib, and Sidney R. Nagel

*The James Franck Institute and The Department of Physics, The University of Chicago,
Chicago, Illinois 60637*

(Received 2 March 1984)

We show that the calculation of the Bohm-Aharonov effect in disordered systems by Al'tshuler, Aronov, and Spivak can be simply explained in terms of first-order degenerate perturbation theory. We find that the conductivity varies periodically with half a flux quantum in both one- and two-dimensional geometries.

PACS numbers: 71.30.+h, 71.50.+t, 71.55.Jv, 72.15.Gd

The prediction by Al'tshuler, Aronov, and Spivak¹ of a Bohm-Aharonov effect in highly disordered systems has generated much interest because of its two surprising differences from what is expected to occur in clean metals.² First, it predicts an interference effect that *increases* with increasing disorder. This is believed to be the result of some kind of "coherent backscattering" associated with localization. Second, the period of oscillation in the conductivity as a function of the magnetic flux turns out to be given by $\phi = \phi_0/2 = hc/2e$ (where h is Planck's constant, c is the velocity of light, and e is the electronic charge) in contrast to the usual periodicity of one flux quantum, ϕ_0 . These predictions are derived from consideration of the maximally crossed diagrams. However, a simple physical picture of these surprising results has not yet emerged although the periodicity with half a flux quantum has been observed experimentally by Sharvin and Sharvin³ and by Ladan and Maurer,⁴ and more complex behavior has recently been seen by several other groups.^{5,6} The maximally crossed diagrams in zero magnetic field have a striking resemblance to the Cooper-pair propagator in superconductivity (it has been suggested that they be called "Cooperons"⁷). The magnetoresistance contribution calculated⁸ from these diagrams also involves Landau orbits of electrons with charge $2e$. Thus these diagrams seem to suggest a kind of "pairing" whose origin remains mysterious.⁹ It is not apparent how the size, geometry, and dimensionality of the system influence the effect or to what extent disorder plays a role in finding a charge

$2e$. What is the physical interpretation of the Al'tshuler, Aronov, and Spivak calculation?

In this paper we will give a very simple and transparent explanation of the effect including the role of dimensionality and disorder. We show that the oscillations in the conductivity as a function of flux originate from the degeneracies of several eigenstates in the ordered system. In two dimensions these degeneracies occur not only at $\phi/\phi_0 = 0$ and 1 but also at other values of ϕ including $\phi/\phi_0 = \frac{1}{2}$. The degeneracies at $\phi/\phi_0 = 0, \frac{1}{2}$, and 1 are special cases in the sense that an increase in disorder destroys the oscillation at those values more slowly than at all other values of ϕ . Thus an oscillation with period $\phi_0/2$ would show up with increasing disorder.

An easy way to see the basic physics of the phenomenon is to look at a simple example in one dimension. We will then extend our calculations to the more complicated case of two dimensions. We start with a chain of atoms in a circle with a magnetic field penetrating the loop. The Schrödinger equation is

$$\left[\frac{1}{2m} \left(\hbar \frac{d}{dx} - \frac{e}{c} A(x) \right)^2 + V(x) \right] \Psi(x) = E \Psi(x),$$

where m is the electron mass, $A(x)$ is the vector potential associated with the field, $V(x)$ is the potential, and E is the eigenvalue. In order to simplify our calculations we have worked with a discrete gauge-invariant Schrödinger equation on a lattice of N points:

$$\frac{\hbar^2}{2ma^2} \left[2\Psi_n - \exp\left(-i\frac{ea}{\hbar c} A_{n+1,n}\right) \Psi_{n+1} - \exp\left(i\frac{ea}{\hbar c} A_{n,n-1}\right) \Psi_{n-1} \right] + V_n \Psi_n = E \Psi_n,$$

where Ψ_n and V_n are the wave function and potential on site n , respectively, a is the distance between sites, and $A_{n,n-1}$ is the vector potential between sites n and $n-1$. It is clear that this equation has the two required properties. As one takes the continuum limit ($a \rightarrow 0$) it reduces to the original Schrödinger equation and if one replaces $A_{n+1,n}$ by $A_{n+1,n} + (\lambda_{n+1} - \lambda_n)/a$ the wave function is changed only by a phase factor. We

have chosen to work in the singular gauge where all $A_{n+1,n} = 0$ except $A_{1,N}$. This corresponds to placing the entire vector potential on the link between the first site and the N th site (which is next to the first one because it is in a ring). The problem thus reduces to diagonalizing the matrix

$$\begin{pmatrix} 2+V_1 & -1 & 0 & \cdots & -\exp(i2\pi\phi/\phi_0) \\ -1 & 2+V_2 & -1 & \cdots & 0 \\ 0 & -1 & 2+V_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -\exp(-i2\pi\phi/\phi_0) & 0 & 0 & \cdots & 2+V_N \end{pmatrix}$$

where we have used the units $\hbar^2/2ma^2 = 1$. In order to calculate a quantity related to the conductivity of the system we have chosen to study the participation ratio, $P = [\sum_{n=1}^N |\Psi_n|^4]^{-1}$. This has physical relevance for telling how extended or how localized the wave function is. It is also approximately related to the quantum correction to the conductivity,⁸ $\delta\sigma$, in highly disordered samples¹⁰: $\delta\sigma \propto -P^{-1}$.

We first examine the solutions of this equation on a completely ordered system ($V_n = 0$ for all n). We can see immediately what will happen. For the ordered case all states can be chosen to be extended (traveling) waves with $P = N$. However, at $\phi/\phi_0 = 0, \frac{1}{2}$, and 1 there are degeneracies. This is the key to the effect. At those values, because of the degeneracy, we can also choose the wave functions to be standing waves which will have a much smaller value of the participation ratio. Varying ϕ infinitesimally away from 0, $\phi_0/2$, or ϕ_0 will give a discontinuous jump in the value of P . Clearly for the perfectly ordered case we could also have chosen states such that the P was a continuous function of ϕ . However, when one considers how even a small amount of disorder will perturb the system and break the degeneracies we see that the choice of standing waves is the natural one for this problem. One way of seeing this is to look at the matrix that we are diagonalizing. When the flux is 0, $\phi_0/2$, or ϕ_0 the matrix is completely real and the nondegenerate states of such a Hermitian matrix can likewise always be chosen to be real (i.e., standing waves). Thus as soon as the disorder lifts the degeneracy we get standing-wave solutions for these values of ϕ and therefore a low value for P . As the disorder increases, the structure in P at these flux values become smoother. For very large values of disorder where the localization length of a wave function is much smaller than the size of the ring, the effect will completely disappear. This can be seen since we could then have chosen a gauge where all the vector potential is placed along a link where the wave function does not exist. Changing the flux would then not perturb the wave function

appreciably. The participation ratio for each individual level is not strictly periodic with flux $\phi_0/2$. However, it can be easily shown that in large systems, when the ensemble average of the participation ratio is taken, this function does indeed have the periodicity of a half flux quantum. We have shown this both numerically by averaging over different systems with the same statistical disorder and analytically from arguments based on symmetry considerations. These results will be presented elsewhere.¹¹

The astonishing thing about our calculation is that we should observe the charge- $2e$ Bohm-Aharonov effect in an arbitrarily small system (for example three sites) with an arbitrarily small amount of disorder. In Fig. 1 we show what happens for a three-site system. Figure 1(a) shows the three eigenvalues as a function of ϕ in the case of zero disorder. Figure 1(b) shows the participation ratio for the three states when there is disorder (V_n

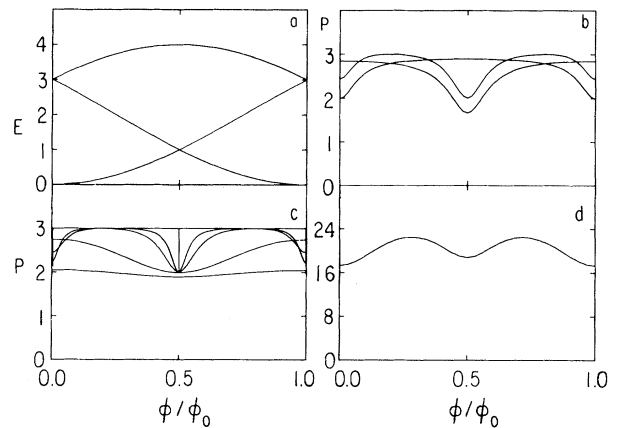


FIG. 1. (a) The eigenvalues vs ϕ/ϕ_0 for a three-site ring with no disorder. (b) The participation ratio vs ϕ/ϕ_0 for the three states of a three-site system with $-2 < V_n < 0$. (c) The participation ratio vs ϕ/ϕ_0 for the intermediate state of the three-site system with varying amounts of disorder: $\Delta V = 0, 1, 2, 8$, and 16. (d) The participation ratio vs ϕ/ϕ_0 for a state in a 32-site system with $\Delta V = 1$.

random with $-2 < V_n < 0$). Figure 1(c) shows how P of the intermediate state varies as the amount of disorder increases. For zero disorder there are delta function spikes which gradually broaden and then disappear as disorder is increased. We have seen the same effects for all the other one-dimensional systems studied. Figure 1(d) shows the participation ratio versus ϕ/ϕ_0 for a 32-site problem with $-1 < V_n < 0$ (i.e., $\Delta V = 1$). Clearly for the same amount of disorder a small system will show a larger effect than one with more sites since the localization length of the wave function can be smaller than the size of the ring for the larger system. Although we have only shown that this effect occurs numerically for relatively small systems, it is clear that our arguments are applicable to a ring of any size as long as the localization length is sufficiently large. In arbitrarily large rings, the energy bands are qualitatively similar to those of the three-site system. In particular for the pure case, they are degenerate at $\phi/\phi_0 = 0, \frac{1}{2}$, and 1. As in the three-site system only the highest and lowest levels will not have this feature.

If we now turn to a two-dimensional calculation, it might seem, at first glance, that the effect will get completely destroyed. In a two-dimensional surface there are many more degeneracies than in a one-dimensional ring. This can be seen in Fig. 2(a) where we show the eigenstates for a cylinder made up of four circular rings each with four atoms (for a total of sixteen sites). Each site is connected to its four neighbors (two on the same ring and one each for the two neighboring rings). We have used here periodic boundary conditions so that the top ring is also connected to the bottom one in a torus geometry. All the results we show here, however, are independent of the boundary conditions we used. There are many degeneracies occurring at points other than $\phi/\phi_0 = 0, \frac{1}{2}$, and 1. If we compute the participation ratio when there is a very small amount of disorder there is no longer only structure at flux 0 and $\phi_0/2$ but rather there is a much more complicated variation with ϕ . This is seen in Fig. 2(b) where we show P for two different states with $\Delta V = 1$. (These are the two curves with the larger amount of structure.) However, if we increase the amount of disorder the dips at 0 and $\phi_0/2$ remain while the other structures disappear! This is also shown in Fig. 2(b) for $\Delta V = 4$. (These are the smooth curves.) Again after averaging over many different ensembles we see that we recover the Bohm-Aharonov effect with charge $2e$.

This, too, can be easily understood. As we mentioned above, by appropriately choosing the gauge

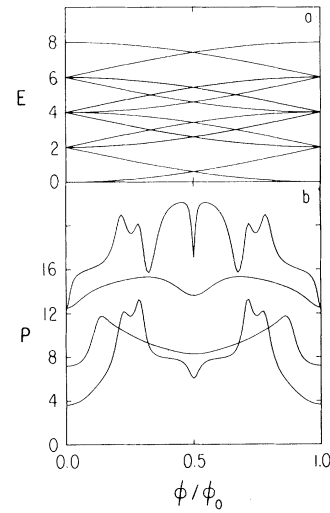


FIG. 2. (a) The eigenvalues vs ϕ/ϕ_0 for a two-dimensional cylinder made up of four rings with four sites each and with no disorder. (b) The participation ratio for two states in the four-by-four cylinder with $\Delta V = 1$ and $\Delta V = 4$. The upper two curves, which correspond to the same state with different amounts of disorder, have been displaced vertically by eight units for clarity.

the matrix can be made totally real at $\phi/\phi_0 = 0, \frac{1}{2}$, or 1 and thus the wave function will be standing waves. Even though there are degeneracies at other values of ϕ , the matrix cannot be made totally real at those points and the nondegenerate wave functions will have some traveling-wave component. In particular, as we increase the disorder, the degenerate states will not only split from each other but will have wave functions which are at different energies mixed into them. At $\phi/\phi_0 = 0, \frac{1}{2}$, or 1 these other wave functions are also standing waves and they do not increase the participation ratio greatly. However, at other values of ϕ , the wave functions which are mixed into the formerly degenerate ones are extended traveling waves and will increase the value of P compared to its value at $\phi/\phi_0 = 0, \frac{1}{2}$, and 1. Thus we see that the effect is regained in the two-dimensional case because a gauge transformation can always be performed to make the wave functions into standing waves at $\phi/\phi_0 = 0, \frac{1}{2}$, and 1 but not otherwise.

Thus we conclude that in a one-dimensional ring the amplitude of the charge- $2e$ Bohm-Aharonov effect seen in the participation ratio becomes stronger as the disorder is decreased. In two dimensions the cleanest signal occurs at some intermediate value of disorder and is weaker when the system is too close to localization and much more complicated when

the system is too well ordered.

In conclusion, we have given a transparent explanation of the physics of the Al'tshuler, Aronov, and Spivak prediction and shown that the essence of their diagrammatic calculation can be reduced to first-order degenerate-state perturbation theory. The degenerate states are particularly sensitive to disorder—the relative shift of their energy levels with ΔV varies linearly rather than quadratically. The presence of a flux can already break this degeneracy even without disorder. Thus there is a particularly strong influence of disorder at $\phi/\phi_0 = 0, \frac{1}{2}$, or 1 where the pure system has degeneracies. The maximally crossed diagrams can be interpreted in terms of splitting this degeneracy between the original time-reversed states. That is why their contribution is important every half flux quantum. It does not escape our attention that the mixing of degenerate states which appears in our calculation can be described in terms of backscattering of the electrons by the impurities. We also note that the condition that our matrix is real when the flux is 0, $\phi_0/2$, or ϕ_0 is the same condition as is necessary for the Hamiltonian to be time-reversal invariant. It is not sufficient only to have a degeneracy to see the effect nor is it sufficient to have the matrix be real [see Figs. 1(a) and 1(b)]. Both conditions are necessary to observe the Bohm-Aharonov effect with a half flux quantum. It is possible that this exceedingly simple way of examining the problem may be useful in studying other phenomena, such as magnetoresistance, where the same types of diagrams appear as were used in the Al'tshuler, Aronov, and Spivak calculations.

We wish to thank Albert Libchaber for his many discussions and for first bringing this problem to our attention. We also wish to thank Norman Birge, Dana Browne, Yoon Jeong, Leo Kadanoff, and Steven Shenker for many helpful discussions. This work was supported in part by the National Science Foundation under Grants No. DMR 82-01182 and No. NSF-MRL 82-16892.

¹B. L. Al'tshuler, A. G. Aronov, and B. Z. Spivak, *Pis'ma Zh. Eksp. Teor. Fiz.* **33**, 101 (1981) [*JETP Lett.* **33**, 94 (1981)].

²I. O. Kulik, *Pis'ma Zh. Eksp. Teor. Fiz.* **11**, 407 (1970) [*JETP Lett.* **11**, 275 (1970)].

³D. Yu. Sharvin and Yu. V. Sharvin, *Pis'ma Zh. Eksp. Teor. Fiz.* **34**, 285 (1981) [*JETP Lett.* **34**, 272 (1981)]; B. L. Al'tshuler, A. G. Aronov, B. Z. Spivak, D. Yu. Sharvin, and Yu. V. Sharvin, *Pis'ma Zh. Eksp. Teor. Fiz.* **35**, 476 (1982) [*JETP Lett.* **35**, 588 (1982)].

⁴F. Ladan and J. Maurer, *C. R. Acad. Sci., Ser. II* **296**, 1753 (1983).

⁵G. Blonder, *Bull. Am. Phys. Soc.* **29**, 535 (1984).

⁶C. Umbach, *Bull. Am. Phys. Soc.* **29**, 535 (1984).

⁷P. W. Anderson, *Physica (Utrecht)* **109&110B**, 1830 (1982).

⁸B. L. Al'tshuler, D. Khmel'nitzkii, A. I. Larkin, and P. A. Lee, *Phys. Rev. B* **22**, 5142 (1980).

⁹Y. Gefen, Y. Imry, and M. Ya Azbel, *Phys. Rev. Lett.* **52**, 129 (1984).

¹⁰This means that since the less-extended states contribute less to the conductivity they are subtracted from the normal conductivity in proportion to the extent to which they are localized.

¹¹D. A. Browne, J. P. Carini, K. A. Muttalib, and S. R. Nagel, to be published.