Aharonov-Bohm Effect in Semiconductor Microstructures

S. Datta and S. Bandyopadhyay
School of Electrical Engineering, Purdue University, West Lafayette, Indiana 47907
(Received 13 May 1986)

In this Letter we present a simple theory for the Aharonov-Bohm effect in semiconductor microstructures, assuming ballistic transport. It is shown that in well-designed symmetric structures it may be possible to attain large (approaching 100% in principle) conductance modulation in a magnetic field even if the transverse dimension of the structure is large, the aspect ratio is poor, and k_BT exceeds the correlation energy. The theoretical formulation can also be used to describe the electrostatic effect which has not yet been observed.

PACS numbers: 72.20.My, 73.60.Br, 73.60.Cs

Oscillatory magnetoresistance due to the Aharonov-Bohm effect[1] has been observed in bulk Mg,[2] cylinders,[3] wire arrays,[4] small metallic rings,[5] and most recently in a molecular-beam-epitaxy-(MBE) grown double quantum well.[6] The origin of h/e and $h/2e$ oscillations in metals is now well understood following the extensive theoretical work of a number of researchers.[7-9] The theoretical work, however, has emphasized *diffusive* transport through an array of elastic scatterers assuming negligible inelastic scattering. The purpose of this Letter is to draw attention to a totally different regime, namely *ballistic* transport, where elastic as well as inelastic scattering is negligible. Advances in semiconductor microtechnology have made it possible to fabricate extremely high-mobility conductive channels isolated from the surface by epitaxial insulators. In GaAs, an electron with a velocity of 10^7 cm/sec has an elastic (plus inelastic) mean free path as long as ~ 4 μm, if the mobility is 10^6 cm^2/V-sec. Channel lengths of 0.25 μm or less can be fabricated by present-day technology so that an electron can travel from one contact to another ballistically with essentially no scattering—elastic or inelastic.

We consider a generic structure [Fig. 1(a)] with two end regions ($x<0$ and $x>L$) and a middle region ($0<x<L$) consisting of two isolated channels.[10] In each of the three regions we can calculate a set of transverse modes in z (or subbands) by solving the eigenvalue equation

$$\left[-\frac{\hbar^2}{2m^*}\frac{d^2}{dz^2}+E_c(z)\right]|n\rangle=\varepsilon_n|n\rangle, \tag{1}$$

where the potential energy $E_c(z)$ includes conduction-band discontinuities as well as any band bending due to space charge. Since the function $E_c(z)$ is different in the end and middle regions, the transverse modes obtained from Eq. (1) are also different. Each subband $|n\rangle$ has a parabolic dispersion given by

$$E(k_x,k_y)=\varepsilon_n+\hbar^2k_x^2/2m^*+\hbar^2k_y^2/2m^*. \tag{2}$$

These are sketched in Fig. 1(b). It will be noted that at equilibrium the relative positioning of the subbands in the three regions is fixed by the requirement of a constant Fermi level for the specified doping densities. The current, I, through the structure for a small applied potential, V, is given by[11]

$$I=\frac{2c}{h}\int dE\int\frac{W_y dk_y}{2\pi}\left[f(E)-f(E+eV)\right]\sum_{n',n''}|T_{n'',n'}|^2. \tag{3}$$

Here W_y is the width of the structure in the y direction. $T_{n'',n'}$ is the transmission coefficient from subband n' in the left end to subband n'' in the right end. E and k_y are the energy and the transverse wave vector of the electrons as they enter from the left end.

We now make four assumptions. (1) Only the lowest subband is occupied everywhere in the device; this is ensured if the doping density N_D is less than $\pi/2W^3$. (2) The magnetic field B_y is low enough that the transverse-mode wave functions are not significantly perturbed, so that its effect is described by lowest-order perturbation theory. (3) The transport from $x=0$ to $x=L$ is "ballistic"; that is, there is no elastic or inelastic scattering and both E and k_y are conserved from one end to the other. (4) $E_c(z)$ is symmetric about $z=0$. With these assumptions we will show that a large conductance modulation (approaching 100%) can be obtained even if the width W_y is large and k_BT is greater than the correlation energy. The absolute conductance modulation ΔG can thus be made arbitrarily large by our increasing W_y. This is very different from diffusive transport and is particularly important if the Aharonov-Bohm effect is to find practical applications.[12] The large conductance

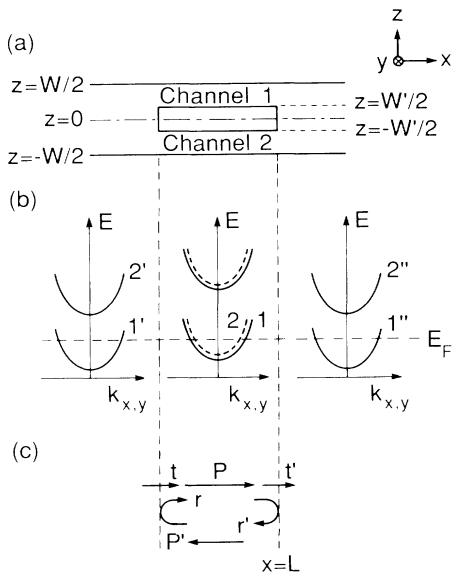

FIG. 1. (a) A two-channel structure suitable for observation of the Aharonov-Bohm effect. The structure is uniform in y with a width of W_y. (b) Dispersion curves $E(k_x,k_y)$ for the different subbands in the three regions. In the middle region $(0<x<L)$ the subbands corresponding to channels 1 and 2 are indicated by solid and dotted lines, respectively. (c) Definition of the transmission and reflection matrices t (2×1), t' (1×2), r (2×2), and r' (2×2) at the two junctions and the propagation matrices P (2×2) and P' (2×2) between $x=0$ and $x=L$.

modulation is obtained even if the aspect ratio W'/W approaches zero.

In view of assumption (1), there is only one subband to consider in each of the end regions. In the middle region there are two subbands corresponding to channels 1 and 2.[13] The transmission coefficient T from one end to the other [dropping the subscripts n' and n'' in Eq. (3)] can be written as[14]

$$T=t'[I-PrP'r']^{-1}Pt. \tag{4}$$

The matrices t,t',r,r',P, and P' are defined in Fig. 1(c). t is a 2×1 matrix describing the transmission from the left end into the two channels while t' is a 1×2 matrix describing the transmission from the channels into the right end:

$$t\equiv\begin{pmatrix}t_1\\t_2\end{pmatrix},\quad t'\equiv(t_1'\ \ t_2'). \tag{5}$$

Similarly r and r' are 2×2 matrices describing the reflections at the two junctions from the channels back into the channels. P and P' are 2×2 matrices describing forward and reverse propagation between $x=0$ and $x=L$. Since the channels are isolated and the transport is ballistic, these matrices are diagonal and the diagonal elements describe the phase shifts in the two channels:

$$P\equiv\begin{pmatrix}P_1&0\\0&P_2\end{pmatrix},\quad P'\equiv\begin{pmatrix}P_1'&0\\0&P_2'\end{pmatrix}, \tag{6}$$

where

$$P_{1,2}=\exp(ik_{x1,2}L), \tag{7}$$

$$P_{1,2}'=\exp(ik_{x1,2}'L). \tag{8}$$

k_{x1} and k_{x2} are the wave numbers in the x direction in channels 1 and 2, respectively, for a given E and k_y, for propagation from $x=0$ to $x=L$; k_{x1}' and k_{x2}' are the corresponding wave numbers for reverse propagation from $x=L$ to $x=0$.

Using Eqs. (5) through (8) in Eq. (4) and taking the magnitude squared we get

$$|T|^2=|a|^2+|b|^2+2|a||b|\cos(\theta+\phi), \tag{9}$$

where

$$a=(Dt_1't_1+Ct_2't_1)/(AD-BC), \tag{10a}$$

$$b=(Bt_1't_2+At_2't_2)/(AD-BC), \tag{10b}$$

$$\phi=\text{phase}\,(a^*b), \tag{10c}$$

$$\theta=(k_{x2}-k_{x1})L, \tag{10d}$$

$$A=1-P_1P_1'r_{11}r_{11}'-P_1P_2'r_{12}r_{21}', \tag{11a}$$

$$B=P_1P_1'r_{11}r_{12}'+P_1P_2'r_{12}r_{22}', \tag{11b}$$

$$C=P_2P_2'r_{22}r_{21}'+P_2P_1'r_{21}r_{11}', \tag{11c}$$

$$D=1-P_2P_2'r_{22}r_{22}'-P_2P_1'r_{21}r_{12}'. \tag{11d}$$

For a given E and k_y the parameters t,t',r, and r' can be calculated by our solving the interface matching problem.[15] $|T|^2$ can then be calculated from Eq. (9) and used in Eq. (3) to calculate the current for a given voltage. $|T|^2$ depends on a, b, ϕ, and θ. The quantities a, b, and ϕ will, in general, vary with E and k_y. However, if $E_c(z)$ is symmetric about $z=0$ then $t_1=t_2$ since channels 1 and 2 are symmetrically disposed while the lowest subband in the end region has a symmetric wave function.[16] Similarly $t_1'=t_2'$, $r_{11}=r_{22}$, etc. As a result, it can be seen from Eqs. (10) and (11) that $a=b$; hence,

$$|T|^2=2|a|^2(1+\cos\theta). \tag{12}$$

We will show shortly that the Aharonov-Bohm phase shift θ is linearly proportional to the magnetic field and is independent of E and k_y. We can then take the factor $(1+\cos\theta)$ outside the integral in Eq. (3):

$$I=(1+\cos\theta)\frac{4e}{h}\int dE\int\frac{W_y dk_y}{2\pi}[f(E)-f(E+eV)]|a|^2. \tag{13}$$

$|a|^2$ depends on θ, so that the current does not vary sinusoidally with the magnetic field, as expected because of multiple reflections; the current modulation is 100% and is not diluted by the integration over E and k_y (ensemble averaging). This is true even if k_BT is greater than the correlation energy. Moreover, the absolute modulation of the conductance ΔG can be made arbitrarily large by our increasing the width W_y.

We will now show that the phase shift θ is indeed independent of E and k_y for a given magnetic field B_y. In the presence of a magnetic field the Hamiltonian H is modified to

$$H=(\mathbf{p}+e\mathbf{A})^2/2m^*+E_C(z). \tag{14}$$

Assuming the vector potential $\mathbf{A}$ to be given by $(B_yz,0,0)$ we can find the modified dispersion relation $E(k_x,k_y)$ for channels 1 and 2 from lowest-order perturbation theory, assuming that the eigenmodes $|1\rangle$ and $|2\rangle$ are unperturbed by the magnetic field:

$$E_{1,2}=\varepsilon_{1,2}+\frac{\hbar^2k_{y1,2}^2}{2m^*}+\frac{(\hbar k_{x1,2}+eB_y\langle z_{1,2}\rangle)^2}{2m^*}+\frac{e^2B_y^2}{2m^*}\langle z_{1,2}^2\rangle, \tag{15}$$

where ε_1 and ε_2 are the energies at the bottom of the band in channels 1 and 2,

$$\varepsilon_{1,2}=\langle 1,2|(p_z^2/2m^*)+E_C(z)|1,2\rangle; \tag{16}$$

$\langle z_1\rangle$ and $\langle z_2\rangle$ are the mean locations of $|1\rangle$ and $|2\rangle$,

$$\langle z_{1,2}\rangle=\langle 1,2|z|1,2\rangle, \tag{17}$$

and $\langle z_1^2\rangle$ and $\langle z_2^2\rangle$ are the spread in the wave functions $|1\rangle$ and $|2\rangle$,

$$\langle z_{1,2}^2\rangle=\langle 1,2|z^2|1,2\rangle-\langle z_{1,2}\rangle^2. \tag{18}$$

Assuming equal E and k_y we can calculate $k_{x2}-k_{x1}$ from Eq. (15); using Eq. (10d) we get for the phase shift θ

$$\theta=eB_yL(\langle z_1\rangle-\langle z_2\rangle)/\hbar+(\varepsilon_1-\varepsilon_2)L/\hbar v_x, \tag{19}$$

where $v_x=\hbar(k_{x1}+k_{x2})/2m^*$. We have set $\langle z_1^2\rangle=\langle z_2^2\rangle$ and $\langle z_1\rangle+\langle z_2\rangle=0$. From Eq. (19) we note that if $\varepsilon_1=\varepsilon_2$, the phase shift θ, to lowest order in perturbation theory, is independent of k_x and hence of E and k_y; it is thus precisely the same for every electron in the ensemble. This is, however, not true if $\varepsilon_1\neq\varepsilon_2$, since the second term is nonzero. In fact, this is the term describing the electrostatic "Aharonov-Bohm" effect which may find applications in switching devices. An applied electric field $\mathcal{E}_z$ modifies $E_c(z)$ in Eq. (14) producing a difference ΔV between the average potentials in channels 1 and 2 so that $\varepsilon_1-\varepsilon_2=e\Delta V$. A similar effect has been observed with neutrons in a gravitational field.[17]

Let us now examine the four assumptions that we made in order to derive Eq. (13). First is the single-modedness of the wave functions in the z direction. If multiple subbands are occupied in the end regions then we need to sum over pairs of subbands n',n'' as indicated in Eq. (3). The wave functions for the different subbands are either symmetric or antisymmetric about $z=0$. It can be shown that $a=b$ if $|n'\rangle$ and $|n''\rangle$ have the same parity in z, and that $a=-b$ if $|n'\rangle$ and $|n''\rangle$ have opposite parity. Consequently, the current is proportional to $1+\cos\theta$ for certain pairs of modes and to $1-\cos\theta$ for others, thus reducing the conductance modulation. Our second assumption involves the rigidity of the wave functions in the z direction. At large magnetic fields the wave functions will resemble cyclotron orbits associated with Landau levels.[18] If the cyclotron radius $r_c\ll W$ then the electron wave functions are localized near one channel or the other and the interference between the two channels should be reduced significantly. It thus appears that in order to observe many periods of Aharonov-Bohm oscillations, long structures ($L\gg 2\pi W$) should be used.

Finally we have the two most stringent assumptions, namely, ballistic transport and symmetric $E_c(z)$. These assumptions ensure that the two arms of the interferometer are perfectly symmetric so that for every path from the left end to the right end, there is a complementary path obtained by substitution of channel 2 for channel 1 and vice versa. It can be shown that these complementary paths have equal transmission amplitudes but differ in phase by $n\theta$ where n is odd, so that they cancel exactly when θ is an odd multiple of π. This is the physical reason why the conductance goes to zero and the modulation is 100%. The question is whether sufficiently symmetric structures can be fabricated by present-day technology so that $a\simeq b$ and large interference effects can be observed. In metals this appears impossible because the phase shift across one arm of the interferometer is typically thousands of radians and even a small percentage deviation would cause a phase jitter greater than 2π. But in semiconductors the de Broglie wavelength is much longer, so that the phase shift across one arm can be only a few radians and it should be possible to control the phase more accurately. However, further theoretical and experimental work is needed to answer this question convincingly. It should be noted that a scattering center of a given dimension has a far smaller scattering cross section in semiconductors compared to metals; this implies a smaller phase shift.[19] Moreover, because of the small screening in semiconductors, the scattering potential in one channel could be strongly correlated with that in a nearby channel ~ 100 Å away. The interference is not

affected if the scattering potential is the same in both channels, even if it is inelastic; it is the differential potential that matters. There is already some experimental evidence[6] for conductance modulations far in excess of $2e^2/h$ even though the end regions were multimoded and transport was probably only partially ballistic.

To summarize, in this Letter we have presented a simple theory for the Aharonov-Bohm effect in semiconductor microstructures assuming ballistic transport and have shown that large interference effects with useful applications could be possible with well-designed symmetric structures.

This work was supported by the National Science Foundation under Grant No. ECS-83-51-036, and by the Semiconductor Research Corporation under Contract No. 83-01-001.

[1]Y. Aharonov and D. Bohm, Phys. Rev. **115**, 485 (1959).

[2]N. B. Sandesara and R. W. Stark, Phys. Rev. Lett. **53**, 1681 (1984).

[3]D. Yu. Sharvin and Yu. V. Sharvin, Pis'ma Zh. Eksp. Teor. Fiz. **34**, 285 (1981) [JETP Lett. **34**, 272 (1981)]; M. Gijs, C. Van Haesendonck, and Y. Bruynseraede, Phys. Rev. Lett. **52**, 2069 (1984).

[4]B. Pannetier, J. Chaussy, R. Rammal, and P. Gandit, Phys. Rev. Lett. **53**, 718 (1984); D. J. Bishop, J. C. Licini, and G. J. Dolan, Appl. Phys. Lett. **46**, 1000 (1985).

[5]R. A. Webb, S. Washburn, C. P. Umbach, and R. B. Laibowitz, Phys. Rev. Lett. **54**, 2696 (1985); V. Chandrasekhar, M. J. Rooks, S. Wind, and D. E. Prober, Phys. Rev. Lett. **55**, 1610 (1985).

[6]S. Datta, M. R. Melloch, S. Bandyopadhyay, R. Noren, M. Vaziri, M. Miller, and R. Reifenberger, Phys. Rev. Lett. **55**, 2344 (1985); S. Bandyopadhyay, B. Das, R. Reifenberger, S. Datta, W. P. Hong, and P. Bhattacharya, to be published.

[7]B. L. Alt'shuler, A. G. Aronov, and B. J. Spivak, Pis'ma Zh. Eksp. Teor. Fiz. **33**, 101 (1981) [JETP Lett. **33**, 94 (1981)].

[8]M. Büttiker, Y. Imry, and R. Landauer, Phys. Lett. **36A**, 365 (1983); M. Büttiker, Y. Imry, and M. Y. Azbel, Phys. Rev. A **30**, 1982 (1984); Y. Gefen, Y. Imry, and M. Y. Azbel, Phys. Rev. Lett. **52**, 129 (1984); M. Büttiker, Y. Imry, R. Landauer, and S. Pinhas, Phys. Rev. B **31**, 6207 (1985).

[9]A. D. Stone, Phys. Rev. Lett. **54**, 2692 (1985); P. A. Lee and A. D. Stone, Phys. Rev. Lett. **55**, 1622 (1985); A. D. Stone and Y. Imry, Phys. Rev. Lett. **56**, 189 (1986).

[10]Usually in metallic rings the end regions are tapered down to a width comparable to the channel widths. Since the de Broglie wavelength of electrons in semiconductors is ~ 500 Å (much larger than in metals), it appears that very few electrons would get from one contact to the other in such a ring structure unless its radius is very much larger than 500 Å. Consequently, the structure shown in Fig. 1(a) is possibly a better choice than the conventional ring structure. The discussion in this paper, however, is quite general and would apply to the latter as well.

[11]R. Tsu and L. Esaki, Appl. Phys. Lett. **22**, 562 (1973). For simplicity we are using this formula rather than the multichannel Landauer formula.

[12]S. Datta *et al.*, Appl. Phys. Lett. **48**, 487 (1986), and Surf. Sci. **174**, 439 (1986); S. Bandyopadhyay *et al.*, in Proceedings of the Second International Conference on Superlattices, Microstructures, and Microdevices, Göteborg, Sweden, 1986 (to be published), and in Technical Digest of the International Electron Devices Meeting, Los Angeles, 1986 (to be published).

[13]We assume that L is large enough that evanescent modes can be neglected.

[14]P. W. Anderson, Phys. Rev. B **23**, 4828 (1981).

[15]A. M. Kriman and P. Ruden, Phys. Rev. B **32**, 8013 (1985).

[16]This is strictly true only when $B_y = 0$; however, the change in t, t', r, r' can be neglected as long as B_y is small.

[17]R. Colella, A. W. Overhauser, and S. A. Werner, Phys. Rev. Lett. **34**, 1472 (1975).

[18]J. Yoshino, H. Sakaki, and T. Hotta, Surf. Sci. **142**, 326 (1984).

[19]See, for example, Akira Ishimaru, *Wave Propagation and Scattering in Random Media* (Academic, New York, 1978), Vol. 1, p. 13.