

Aharonov-Bohm effect and electron correlation in quantum dots

Hiroshi Akera

Faculty of Engineering, Hokkaido University, Sapporo 060, Japan

(Received 14 December 1992)

Transport through two quantum dots in parallel is studied theoretically in the presence of a magnetic flux between dots and in the limit of large one-electron level separations in each dot. Electron-electron scattering at a dot breaks the phase coherence of an electron in the tunneling process through the dot and suppresses the Aharonov-Bohm oscillations of the conductance. It is shown that the amplitude of the Aharonov-Bohm oscillations depends strongly on the Fermi level relative to the resonant levels because of a two-electron phase interference.

Interference phenomena of electron waves, such as the Aharonov-Bohm (AB) effect, have been observed in a mesoscopic solid-state device with a dimension of a few micrometers.¹ In such a small system and at low temperatures an electron keeps the phase coherence during its transport because of the absence of inelastic scatterings by phonons, etc. As well as the phase coherence, the Coulomb repulsion is known to be a key ingredient in the mesoscopic physics.² Electron-electron scatterings will suppress the phase coherence as expected from the Fermi liquid theory, but the way to suppress the phase coherence in such a confined system may be qualitatively different from that in a macroscopic system in which the Fermi liquid theory is valid. In the present paper, the suppression of the AB effect due to the electron correlation is studied in a system consisting of two quantum dots.

The transport through a mesoscopic grain, or through two tunnel junctions, has been investigated extensively in metallic systems with grain sizes much larger than the Fermi wavelength,² in which single-electron energy levels are continuous and possibly the Fermi liquid theory is applicable. Recently the study was extended to semiconductors in both theory^{3,4} and experiment.⁵ In semiconductors the discreteness of energy levels is more important because of small effective mass, and the level separations in the smallest dot are reported to be comparable to or larger than the charging energy.^{5,6}

In this paper the quantum limit is considered, for simplicity, in which energy separations ΔE between single-electron levels are much larger than the other energy scales, the charging energy U , the temperature $k_B T$, and the level broadening Γ due to tunneling. In this quantum limit, only one state contributes to the current. Two electrons can occupy this state and the second electron needs an excess energy U due to the Coulomb interaction. Such a quantum dot coupled to leads is essentially the Anderson model for a single magnetic impurity⁷ and has recently been explored again by many theoreticians.⁸⁻¹⁴

The transport depends on the number N of electrons in each dot and the Fermi level ϵ_F . When $N=0$, the effect of the electron correlations is small. The same is true for $N=2$ because of the electron-hole symmetry. The strongest electron correlation is expected at $N=1$, where

a localized spin is in each dot. When ϵ_F coincides with one of the energy levels separated by U , resonant tunneling through the level occurs and between these two levels the conductance is smaller but nonzero. At low temperatures quantum transport via intermediate states dominates the transport between the resonant conductance peaks and has been studied for a metallic dot theoretically¹⁵⁻¹⁸ and observed experimentally.¹⁹⁻²¹ In this paper such quantum transport is investigated in quantum dots.

We consider a system of dots connected to left and right leads. It is assumed that, when the voltage is applied between leads, the potential drop occurs exclusively between dots and leads, and that the local equilibrium is maintained in each lead. In this situation the current is evaluated from the tunneling probability between two leads.

First, we describe the single-dot case as an illustration. Our model Hamiltonian consists of three parts: the Hamiltonian of leads, that of a dot, and the tunneling term:

$$H = H_l + H_d + H_t, \quad (1)$$

$$H_l = \sum_{k\sigma} [(\epsilon_{ak\sigma} - eV)a_{k\sigma}^\dagger a_{k\sigma} + \epsilon_{bk\sigma} b_{k\sigma}^\dagger b_{k\sigma}], \quad (2)$$

$$H_d = \sum_{\sigma} \epsilon_{d\sigma} d_{\sigma}^\dagger d_{\sigma} + U d_{\uparrow}^\dagger d_{\uparrow} d_{\downarrow}^\dagger d_{\downarrow}, \quad (3)$$

$$H_t = \sum_{k\sigma} (V_{ak\sigma} a_{k\sigma}^\dagger d_{\sigma} + V_{bk\sigma} b_{k\sigma}^\dagger d_{\sigma}) + \text{H.c.}, \quad (4)$$

where $a_{k\sigma}^\dagger$ ($a_{k\sigma}$) and $b_{k\sigma}^\dagger$ ($b_{k\sigma}$) are creation (annihilation) operators of an electron with spin σ in a state k in the left and right leads, respectively, and d_{σ}^\dagger (d_{σ}) is that in the dot. V is the voltage applied between the left and right leads and $e > 0$.

Transition rates between the left and right leads are calculated by treating H_t as a perturbation. In the following the lowest order in H_t is considered and the Kondo effect due to the correlation between leads and dots is neglected assuming the temperature is higher than the Kondo temperature. The transition rate between $i = |ak_1\sigma_1s_1\rangle$, in which an electron with spin σ_1 is in state k_1 in the left lead and an electron in the dot is in spin state s_1 , and $f = |bk_2\sigma_2s_2\rangle$, in which an electron is

in the right lead, is expressed in terms of the transition matrix \hat{T} :

$$P_{fi} = (2\pi/\hbar) |\langle f | \hat{T} | i \rangle|^2 \delta(E_i - E_f), \quad (5)$$

and \hat{T} in the lowest order is

$$I = (-e) \sum_{k_1 \sigma_1 s_1} \sum_{k_2 \sigma_2 s_2} [P_{bk_2 \sigma_2 s_2, ak_1 \sigma_1 s_1} (1 - f_{bk_2 \sigma_2}) f_{ak_1 \sigma_1} g_{s_1} - P_{ak_1 \sigma_1 s_1, bk_2 \sigma_2 s_2} (1 - f_{ak_1 \sigma_1}) f_{bk_2 \sigma_2} g_{s_2}], \quad (7)$$

where g_s is the probability of state s and $f_{ak\sigma} = f(\epsilon_{ak\sigma})$ the Fermi distribution function. The same equation is valid for the many-dot case if we consider s_1 and s_2 as a quantum number specifying one degenerate state in a system of dots. If there are m_d degenerate states, $g_s = 1/m_d$, and for a single dot with one electron, $g_s = \frac{1}{2}$. Due to the detailed balance $P_{fi} = P_{if}$,

$$I = (-e/m_d) \sum P_{bk_2 \sigma_2 s_2, ak_1 \sigma_1 s_1} (f_{ak_1 \sigma_1} - f_{bk_2 \sigma_2}). \quad (8)$$

Assuming $V_{ak\sigma}$ and $V_{bk\sigma}$ depend only on $\epsilon_{ak\sigma}$ and $\epsilon_{bk\sigma}$, respectively,

$$I = (-e/m_d) \sum_{\sigma_1 s_1 \sigma_2 s_2} \int d\epsilon_1 D_{a\sigma_1}(\epsilon_1) \int d\epsilon_2 D_{b\sigma_2}(\epsilon_2) (2\pi/\hbar) |\langle bk_2 \sigma_2 s_2 | \hat{T} | ak_1 \sigma_1 s_1 \rangle|^2 \delta(\epsilon_1 - eV - \epsilon_2) [f(\epsilon_1) - f(\epsilon_2)]. \quad (9)$$

When the densities of states $D_{a\sigma}(\epsilon)$ and $D_{b\sigma}(\epsilon)$ do not depend on spin σ , the conductance becomes

$$G = (\partial I / \partial V)(V=0) = (2\pi e^2 / \hbar) \int d\epsilon D_a(\epsilon) D_b(\epsilon) (-\partial f / \partial \epsilon) S(\epsilon), \quad (10)$$

with

$$S(\epsilon) = (1/m_d) \sum_{\sigma_1 s_1 \sigma_2 s_2} |\langle bk_2 \sigma_2 s_2 | \hat{T} | ak_1 \sigma_1 s_1 \rangle|^2. \quad (11)$$

The conductance at temperatures much lower than U is proportional to $S(\epsilon_F)$ and calculated in the following. For a while, the Zeeman splitting is neglected and $\epsilon_{d\sigma} = \epsilon_d$. When $\epsilon_F < \epsilon_d$, there are no electrons in the dot and the transport is of one-electron character. The transition matrix is

$$\begin{aligned} \langle bk_2 \sigma | \hat{T} | ak_1 \sigma \rangle &= \frac{\langle bk_2 \sigma | H_t | d\sigma \rangle \langle d\sigma | H_l | ak_1 \sigma \rangle}{\epsilon_F - \epsilon_d} \\ &= T_\alpha, \end{aligned} \quad (12)$$

with

$$T_\alpha = -V_b V_a^* / (\epsilon_d - \epsilon_F), \quad (13)$$

where $V_a = V_a(\epsilon_F)$ and $V_b = V_b(\epsilon_F)$, and the conductance is proportional to $S(\epsilon_F) = 2|T_\alpha|^2$. When $\epsilon_d < \epsilon_F < \epsilon_d + U$, one electron in the dot and another electron from the left lead are involved in the transition and there are three types of transitions as illustrated in Fig. 1 for an up-spin incident electron. The transition matrices for these processes are T_α ,

$$T_\beta = V_b V_a^* / (\epsilon_F - \epsilon_d - U), \quad (14)$$

and

$$T_\gamma = T_\alpha - T_\beta, \quad (15)$$

respectively, and $S(\epsilon_F) = |T_\alpha|^2 + |T_\beta|^2 + |T_\gamma|^2$. When $\epsilon_F > \epsilon_d + U$, the transport becomes of one-electron char-

$$\langle f | \hat{T} | i \rangle = \sum_n \langle f | H_t | n \rangle \langle n | H_l | i \rangle / (E_i - E_n), \quad (6)$$

where intermediate states n have no electron or two electrons in the dot. The current is then expressed as

acter again and $S(\epsilon_F) = 2|T_\beta|^2$.

Now we move on to the transport through two dots and discuss the AB effect. The Hamiltonian becomes

$$H_d = \sum_{n=1}^2 \left[\sum_{\sigma} \epsilon_{n\sigma} d_{n\sigma}^\dagger d_{n\sigma} + U_n d_{n\uparrow}^\dagger d_{n\uparrow} d_{n\downarrow}^\dagger d_{n\downarrow} \right], \quad (16)$$

$$\begin{aligned} H_t &= \sum_{k\sigma} [e^{-i\varphi/4} (V_{a1k\sigma} a_{k\sigma}^\dagger d_{1\sigma} + V_{b2k\sigma} b_{k\sigma}^\dagger d_{2\sigma}) \\ &\quad + e^{i\varphi/4} (V_{a2k\sigma} a_{k\sigma}^\dagger d_{2\sigma} + V_{b1k\sigma} b_{k\sigma}^\dagger d_{1\sigma})] \\ &\quad + \text{H.c.}, \end{aligned} \quad (17)$$

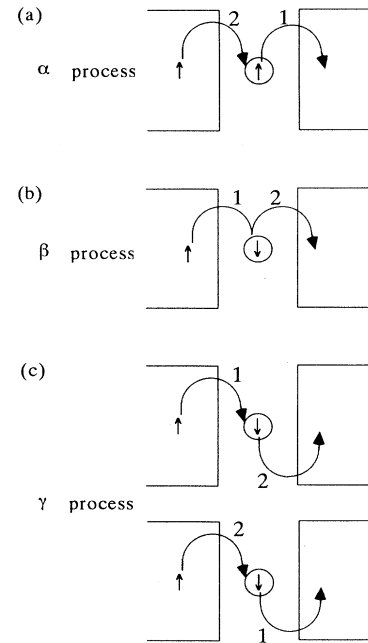


FIG. 1. Transport through an intermediate state in a single dot. Arrows represent the spin direction and numbers show the order of transitions.

where $\varphi=2\pi\Phi/\Phi_0$ with Φ the magnetic flux between dots and $\Phi_0=hc/e$ the flux quantum. For each dot, there are three types of transitions as in Fig. 1 and the transition matrices are written in terms of

$$T_{\alpha n} = -V_{bn}V_{an}^*/(\varepsilon_n - \varepsilon_F), \quad (18)$$

$$T_{\beta n} = V_{bn}V_{an}^*/(\varepsilon_F - \varepsilon_n - U_n), \quad (19)$$

$$T_{\gamma n} = T_{\alpha n} - T_{\beta n}, \quad (20)$$

with $n=1,2$. In the following we assume that two dots are identical and $\varepsilon_n = \varepsilon_d$, etc. Of course qualitative parts of the conclusions obtained below are also the case for nearly identical dots and a detailed adjustment may be possible in the experiment by applying the gate voltage on one of the dots. When $\varepsilon_F < \varepsilon_d$,

$$\begin{aligned} S(\varepsilon_F) &= 2|e^{i\varphi/2}T_{\alpha 1} + e^{-i\varphi/2}T_{\alpha 2}|^2 \\ &= 4|T_{\alpha}|^2(\cos\varphi + 1). \end{aligned} \quad (21)$$

The term with $\cos\varphi$ represents the AB oscillations. When $\varepsilon_d < \varepsilon_F < \varepsilon_d + U$, there are eight processes, four of which have an interference between paths through dots 1 and 2 and the transition matrices are

$$\begin{aligned} e^{i\varphi/2}T_{\alpha 1} + e^{-i\varphi/2}T_{\alpha 2}, \quad e^{i\varphi/2}T_{\beta 1} + e^{-i\varphi/2}T_{\beta 2}, \\ e^{i\varphi/2}T_{\alpha 1} + e^{-i\varphi/2}T_{\beta 2}, \quad e^{i\varphi/2}T_{\beta 1} + e^{-i\varphi/2}T_{\alpha 2}. \end{aligned} \quad (22)$$

The other four are spin-flip processes through dots 1 and 2 with transition matrices

$$e^{i\varphi/2}T_{\gamma 1}, \quad e^{-i\varphi/2}T_{\gamma 2}. \quad (23)$$

(There are two processes through dot 1 in which only spin orientations in dot 2 are different and the matrix elements are identical. The same is true for the processes through dot 2.) Therefore we obtain

$$\begin{aligned} S(\varepsilon_F) &= |V_a V_b|^2 [(1/\varepsilon_\alpha - 1/\varepsilon_\beta)^2 \cos\varphi \\ &\quad + 4(1/\varepsilon_\alpha^2 + 1/\varepsilon_\beta^2 + 1/\varepsilon_\alpha \varepsilon_\beta)], \end{aligned} \quad (24)$$

with $\varepsilon_\alpha = \varepsilon_F - \varepsilon_d > 0$ and $\varepsilon_\beta = \varepsilon_d + U - \varepsilon_F > 0$. If we put $S(\varepsilon_F) = A(R \cos\varphi + 1)$, the relative amplitude of the AB oscillations is

$$R = (\varepsilon_\alpha - \varepsilon_\beta)^2 / 4(\varepsilon_\alpha^2 + \varepsilon_\beta^2 + \varepsilon_\alpha \varepsilon_\beta). \quad (25)$$

$R(\varepsilon_F)$ ($\varepsilon_d < \varepsilon_F < \varepsilon_d + U$) is symmetric around $\varepsilon_F = \varepsilon_d + U/2$ and has a minimum $R=0$ at $\varepsilon_F = \varepsilon_d + U/2$ and maxima $R = \frac{1}{4}$ at $\varepsilon_F = \varepsilon_d$ and $\varepsilon_d + U$. Due to spin-flip processes R is reduced from unity, the value in the one-electron transport Eq. (21). The maximum value $R = \frac{1}{4}$ at $\varepsilon_F \rightarrow \varepsilon_d$ is obtained from the following simple argument.²² In this limit, processes through intermediate states with no electron in a dot (α processes) are dominant in comparison with those through states with two electrons (β processes). In order to avoid a spin flip in the α processes, both spins in the dots in the initial state should have the same orientation as the spin of the incident electron. The probability to have such an initial state is $(\frac{1}{2})^2$, giving $R = \frac{1}{4}$. The same argument is true in a $2m$ -dot ring, for which $R = (\frac{1}{2})^{2m}$.²³

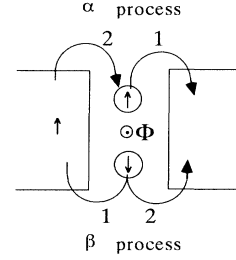


FIG. 2. Two-electron phase interference in processes through dots. Φ denotes the magnetic flux.

It is remarkable that the AB oscillations become smaller as ε_F moves away from $\varepsilon_F = \varepsilon_d$ and $\varepsilon_F = \varepsilon_d + U$ and vanish completely at $\varepsilon_F = \varepsilon_d + U/2$. The reason is as follows. Except for the vicinity of $\varepsilon_F = \varepsilon_d$ and $\varepsilon_F = \varepsilon_d + U$, both α and β processes contribute to the transport. Because of the sign difference between T_α and T_β , the processes involving both T_α and T_β shown in Eq. (22) and in Fig. 2 give oscillations with quite different phases from those involving two T_α 's and two T_β 's. At $\varepsilon_F = \varepsilon_d + U/2$, $T_\alpha = -T_\beta$ and these two kinds of oscillations have a phase difference of π and the same amplitude, resulting in the disappearance of the AB effect. Note that the additional minus sign in T_α comes from an exchange of electrons involved in the transport. Therefore, this disappearance of the AB effect is a manifestation of a two-electron interference.

The spin state in dots is drastically changed by the magnetic field and by the spin-spin interaction. When $\mu_B H \gg k_B T$, both spins in the dots align along the magnetic field in both the initial and the final state and the phase coherence is recovered, giving

$$S(\varepsilon_F) = 2|V_a V_b|^2 (1/\varepsilon_\alpha^2 + 1/\varepsilon_\beta^2) (\cos\varphi + 1). \quad (26)$$

The spin-spin interaction due to hopping between dots and leads is of the order of $(V_a^4 + V_b^4)/U^3$ and can be much smaller than $k_B T$, but the interaction due to direct hopping between dots can be large than $k_B T$. Here we assume a spin-spin interaction without asking its origin. The ground state is either spin-singlet or spin-triplet depending on the sign of interaction. For the singlet, it is shown that

$$S(\varepsilon_F) = |V_a V_b|^2 (1/\varepsilon_\alpha - 1/\varepsilon_\beta)^2 (\cos\varphi + 1), \quad (27)$$

and $R=1$ because of the absence of a spin flip. The conductance is zero at $\varepsilon_F = \varepsilon_d + U/2$ because of the two-electron interference mentioned above. For the triplet, in spite of the presence of spin-flip processes,

$$S(\varepsilon_F) = |V_a V_b|^2 (3/\varepsilon_\alpha^2 + 3/\varepsilon_\beta^2 + 2/\varepsilon_\alpha \varepsilon_\beta) (\cos\varphi + 1), \quad (28)$$

and $R=1$. This is because spin flips at different dots are correlated. For example, the matrix element between $(d_{1\uparrow}^\dagger d_{2\downarrow}^\dagger + d_{1\downarrow}^\dagger d_{2\uparrow}^\dagger)/\sqrt{2}$ and $d_{1\uparrow}^\dagger d_{2\uparrow}^\dagger$ is

$$\langle \langle b \downarrow 1 \uparrow 2 \uparrow | \hat{T} | a \uparrow 1 \uparrow 2 \downarrow \rangle + \langle b \downarrow 1 \uparrow 2 \uparrow | \hat{T} | a \uparrow 1 \downarrow 2 \uparrow \rangle \rangle / \sqrt{2}, \quad (29)$$

and two processes with a spin flip at dots 1 and 2 inter-

ferre with each other.

In conclusion, the Aharonov-Bohm oscillation of the conductance has been investigated in a system consisting of two quantum dots. Separations between one-electron levels in each dot are assumed to be much larger than the charging energy and the transport through a single-electron state near the Fermi level is considered. The amplitude of the Aharonov-Bohm oscillations divided by the average conductance is calculated as a measure of the phase coherence and is shown to decrease compared to that in the absence of the Coulomb repulsion when the state in each dot is half-filled. The phase coherence is strongly suppressed, but not completely destroyed in such a small system. There is no significant temperature dependence in the region $T_K \ll T \ll U/k_B$ with T_K the Kondo temperature in remarkable contrast with the Fermi liquid theory. A tunneling process with an exchange of electrons gives a different component of conductance

oscillations with an opposite phase compared to the usual AB effect. These two oscillations cancel each other and the conductance oscillations disappear when the Fermi level comes at the middle point of the two levels separated by U . The Zeeman splitting and the spin correlation between two dots recover the phase coherence.

By making two dots nearly identical using the recent microfabrication technique, it will be possible to observe a suppression of the AB oscillations even if they do not vanish because of small inevitable differences between them. In the measurement of the conductance as a function of the strength of a uniform magnetic field, this suppression will be observed when the effect of the Zeeman splitting is reduced by the temperature.

The author would like to thank T. Ando, H. Fukuyama, L. I. Glazman, A. H. MacDonald, H. Matsukawa, O. Sakai, and N. S. Wingreen for valuable discussions.

¹For example, S. Washburn, in *Mesoscopic Phenomena in Solids*, edited by B. L. Altshuler *et al.* (Elsevier, Amsterdam, 1991), p. 1.

²D. V. Averin and K. K. Likharev, in *Mesoscopic Phenomena in Solids* (Ref. 1), p. 173.

³C. W. J. Beenakker, *Phys. Rev. B* **44**, 1646 (1991).

⁴D. V. Averin *et al.*, *Phys. Rev. B* **44**, 6199 (1991).

⁵See, for example, B. Su *et al.*, *Phys. Rev. B* **46**, 7644 (1992), and references therein.

⁶The smallest dot is made by vertical etching of double-barrier resonant-tunneling structures. In this type of device, electrons occupy the dot only by applying the bias. Recently, however, a device has been made by S. Tarucha, T. Honda, Y. Hirayama, T. Saku, and Y. Tokura (unpublished), in which barriers are modulation doped and electrons are occupying the dot even at zero bias.

⁷P. W. Anderson, *Phys. Rev.* **124**, 41 (1961).

⁸L. I. Glazman and M. É. Raïkh, *Pis'ma Zh. Eksp. Teor. Fiz.* **47**, 378 (1988) [*JETP Lett.* **47**, 452 (1988)].

⁹T. K. Ng and P. A. Lee, *Phys. Rev. Lett.* **61**, 1768 (1988).

¹⁰A. Groshev *et al.*, *Phys. Rev. Lett.* **66**, 1082 (1991).

¹¹Y. Meir *et al.*, *Phys. Rev. Lett.* **66**, 3048 (1991).

¹²L. Y. Chen and C. S. Ting, *Phys. Rev. B* **44**, 5916 (1991).

¹³A. Kawabata, *J. Phys. Soc. Jpn.* **60**, 3222 (1991).

¹⁴S. Hershfield *et al.*, *Phys. Rev. Lett.* **67**, 3720 (1991).

¹⁵D. V. Averin and A. A. Odintsov, *Phys. Lett. A* **140**, 251 (1989).

¹⁶D. V. Averin and Y. V. Nazarov, *Phys. Rev. Lett.* **65**, 2446 (1990).

¹⁷L. I. Glazman and K. A. Matveev, *Pis'ma Zh. Eksp. Teor. Fiz.* **51**, 425 (1990) [*JETP Lett.* **51**, 484 (1990)].

¹⁸B. Laikhtman, *Phys. Rev. B* **43**, 2731 (1991).

¹⁹L. J. Geerligs *et al.*, *Phys. Rev. Lett.* **65**, 3037 (1990).

²⁰D. C. Glattli *et al.*, *Z. Phys. B* **85**, 375 (1991).

²¹T. M. Eiles *et al.*, *Phys. Rev. Lett.* **69**, 148 (1992).

²²Since $R(\epsilon_F) = 1$ for $\epsilon_F < \epsilon_d$ and $\epsilon_F > \epsilon_d + U$, there is a jump of $R(\epsilon_F)$ at $\epsilon_F = \epsilon_d, \epsilon_d + U$. This is an artifact of the lowest-order perturbation theory and $R(\epsilon_F)$ changes continuously if the Kondo effect is taken into account which becomes important when ϵ_F is within the level broadening Γ of ϵ_d and $\epsilon_d + U$ (Refs. 8 and 9). The following argument is applicable outside these regions and in the vicinity of $\epsilon_F = \epsilon_d, \epsilon_d + U$.

²³L. I. Glazman (unpublished).