Comments on the Aharonov-Bohm Effect

Kazuyasu Shigemoto and Kunihiko Uehara

Abstract

In the original setting of the Aharonov-Bohm, the gauge invariant physical longitudinal mode of the vector potential, which is expressed by the gauge invariant physical current, gives the desired contribution to the Aharonov-Bohm effect. While the scalar mode of the vector potential, which changes under the gauge transformation so that it is the unphysical mode, give no contribution to the Aharonov-Bohm effect. Then Aharonov-Bohm effect really occurs by the physical longitudinal mode in the original Aharonov-Bohm’s setting. In the setting of Tonomura et al., where the magnet is shielded with the superconducting material, not only the magnetic field but also the longitudinal mode of the vector potential become massive by the Meissner effect. Then not only the magnetic field but also the physical longitudinal mode does not come out to the region where the electron travels. In such setting, only the scalar mode of the vector potential exists in the region where the electron travels, but there is no contribution to the Aharonov-Bohm effect from that mode.

Index Terms

Aharonov-Bohm effect, non-simply connected region, physical longitudinal mode, Stokes’ theorem

I. INTRODUCTION

In 1959, Aharonov and Bohm publish the paper on the issue of the vector potential, which changes under the gauge transformation, so that the vector potential is unphysical, causes the physical observable Aharonov-Bohm effect [1], [2], [3]. There is another issue that the Aharonov-Bohm effect is the genuine quantum effect or not[1], [2], [3]. This is because, in the classical theory, there does not give any force through the Lorentz force with zero magnetic field in the region where the electron travels.

After their paper, various experimental test of the Aharonov-Bohm effect were reported [4], [5], [6], [7]. There also came various theoretical explanations different from that of Aharonov-Bohm, that is, the explanation by the non-local field instead of the local vector potential, the explanation by a new force in addition to the Lorentz force, the explanation by including the back-reaction of the solenoid etc..

Then there has come the theoretical paper that the Aharonov-Bohm effect itself does not occur [8], [9]. Furthermore there have appeared various theoretical and experimental papers to discuss whether the Aharonov-Bohm effect really exists or not and whether the Aharonov-Bohm effect is the genuine quantum effect or not.

In order to clarify such situations, Tonomura et al. proposed the experiment to shield a magnet with the superconducting material [10], [11], because the main objection of the non-existence of the Aharonov-Bohm effect is that the experimental evidence comes from the leakage of the magnetic field.

In order to clarify the argument of the Aharonov-Bohm effect, we consider the Aharonov-Bohm effect caused by the infinitely long solenoid with the steady current.

In order to make the argument clear, we physically decompose the vector potential \( A \) into various modes as the solution of the Maxwell equation. Time-dependent modes are two transverse modes \( A^T \), which become the photon when quantized. The rests are the longitudinal mode \( A^L \) which is expressed by the integral of the current, and the scalar mode \( A^S \) which is the pure gauge mode. As we focus on the Aharonov-Bohm effect, we consider only the longitudinal mode \( A^L(x) \) and the scalar mode \( A^S(x) \) of the vector potential \( A(x) \).

The longitudinal mode \( A^L(x) \) is the projected mode from \( A(x) \) in the momentum space in the form \( A^L(x) = (\delta_{ij} - \frac{\partial_i}{\partial x_j} + \frac{1}{4\pi} \int \nabla x \cdot j(x', t) \frac{d^3 x'}{|x - x'|} \), \( x' \neq x \)

For the electric current, we have the explicit form \( j(x, t) = (-e) \bar{\psi}(x, t) \gamma \psi(x, t) \) by using the electron field \( \psi(x, t) \). Though the phase of the electron changes, the current does not change under the gauge transformation, which means that the longitudinal mode does not change under the gauge transformation. The scalar mode \( A^S(x) \) is the pure gauge mode of the form

\[
A^S(x, t) = \nabla \theta(x, t),
\]

where \( \theta(x, t) \) is the scalar function for the gauge transformation. We call the physical mode which does not change, and the unphysical mode which changes, under the gauge transformation. Then the transverse and the longitudinal modes are physical mode and the scalar mode is the unphysical mode.
II. THE AHARONOV-BOHM EFFECT IN VARIOUS SETTINGS

A. The Original Aharonov-Bohm setting

We review the original Aharonov and Bohm setting, and consider the effect from the infinitely long solenoid.

For the infinite long solenoid with the radius \( a \) of the form \( \rho = a, \varphi = [0, 2\pi], z = [-\infty, +\infty] \) in the cylindrical coordinate \( x = \rho \cos \varphi, y = \rho \sin \varphi, z = z \) and the longitudinal mode of the vector potential outside the solenoid is given by

\[
A^L(x)_\rho = 0, \quad A^L(x)_z = 0, \quad A^L(x)_\varphi = \frac{\mu_0 n I a^2}{2\rho},
\]  

(3)

where \( I \) is the steady current, \( n \) is the number of turns of the solenoid in the unit length for the \( z \)-direction. The contribution to the magnetic field \( B^L(x) \) from the longitudinal mode of the vector potential \( A^L(x) \) is given to be zero in the following way

\[
B^L_\rho = \frac{1}{\rho} \frac{\partial A^L_\rho}{\partial \varphi} - \frac{\partial A^L_\varphi}{\partial z} = 0, \quad B^L_\varphi = \frac{\partial A^L_\rho}{\partial z} - \frac{\partial A^L_\varphi}{\partial \rho} = 0,
\]

(4)

\[
B^L_z = \frac{1}{\rho} \frac{\partial (\rho A^L_\rho)}{\partial \varphi} - \frac{\partial A^L_\varphi}{\partial \rho} = \frac{1}{\rho} \frac{\partial (\mu_0 n I a^2/2)}{\partial \rho} = 0.
\]

(5)

The contribution to the magnetic field \( B^S(x) \) from the scalar mode of the vector potential \( A^S(x) = \nabla \theta(x) \) is trivially given to be zero. In this way, we confirm that there is no magnetic field \( B^S(x) = 0, B^L(x) = 0 \), around the region where the electron travels. While the line integral of the longitudinal mode around the closed curve \( C \) gives

\[
\oint_C A^L(x) \cdot dx \equiv \int_0^{2\pi} A^L(x)_{\varphi} \rho \, d\varphi = \frac{\mu_0 n I a^2}{2} \int_0^{2\pi} d\varphi = \mu_0 n I a^2 \pi.
\]

(6)

We have the same result by using the Stokes’ theorem.

While the line integral of the scalar mode around the closed curve \( C \) gives

\[
\oint_C A^S(x) \cdot dx \equiv \oint_C \nabla \theta(x) \cdot dx = \oint_C d\theta(x) = \theta(P) - \theta(P) = 0,
\]

(7)

where the closed curve \( C \) starts from the point \( P \) and ends at the same point \( P \). Then, in the quantum theory, the contribution from only the physical mode, the longitudinal mode, gives the change of the phase of the electron \( \Phi = (-e) \bar{h} \oint A^L(x) \cdot dx = \frac{(-e)\mu_0 I a^2 \pi}{\bar{h}} \), which gives the Aharonov-Bohm effect.

In the classical theory, as the Lorentz force to an electron is given by

\[
F = (-e)(E + v \times B) = ma,
\]

(8)

and \( \phi(x) = 0, B(x) = B^L(x) + B^S(x) = 0 \) in the Aharonov-Bohm setting, we have

\[
F = (-e) \frac{dA^L(x(t))}{dt} = m \frac{dv(t)}{dt},
\]

(9)

where we use the longitudinal mode of Eq.(3) with the case of a very slow changing current, which becomes the steady current in the final limit. Then, classically, we have the conservation of the total momentum

\[
p(t) = mv(t) + (-e)A^L(x(t)) = (\text{const.}).
\]

(10)

We denote \( p^L(t) = (-e)A^L(x(t)) \) as the momentum of the longitudinal mode of the vector potential. Classically, as the electron is treated as the particle, we cannot include the scalar mode \( A^S(x, t) = \nabla \theta(x, t) \) in the momentum because there is no concept of the phase factor of the electron. (In the quantum theory, the scalar mode and the gradient of the phase factor of the electron, which change under the gauge transformation, compensate to give the theory gauge invariant.) Classically, there is no concept of the phase of the electron, there is not Aharonov-Bohm effect, but the physically meaningful longitudinal mode gives the change of the angular momentum \( \oint p^L(x) \cdot dx = (-e) \oint A^L(x) \cdot dx = \frac{(-e)\mu_0 I a^2 \pi}{\bar{h}} \). As is well-known, \( \oint p(x) \cdot dx = nh \) gives the Bohr’s semi-classical quantization condition. In this way, in the classical level, the physically meaningful longitudinal mode of the vector potential gives the change of the angular momentum.

On the issue that the vector potential, which changes under the gauge transformation, gives the physical Aharonov-Bohm effect, we have the following conclusion. In the quantum theory, the corresponding vector potential is composed of the longitudinal and the scalar modes. The longitudinal mode, which does not change under the gauge transformation, gives the observed Aharonov-Bohm effect. The scalar mode, which changes under the gauge transformation, gives no contribution to the Aharonov-Bohm effect.

In the classical theory, there does not appear the scalar mode in the momentum. Thus the physical longitudinal mode only contributes the physical change of the angular momentum.
B. The Tonomura et al.’s Setting

In this case, we consider the setting that the infinitely long solenoid is shielded by the superconducting material. The theoretical model is known as the Ginzburg-Landau-Abrikosov-Gor’kov theory [12], [13] in the form

\[
\frac{1}{2m} \left| (-ie \nabla - (-e^*) \right) \mathbf{A}_L \right|^2 \psi + \alpha \psi + \beta |\psi|^2 \psi = 0, \tag{11}\]

where \((-e^*)\) is the effective charge for the Cooper pair with \((-e^*) = (-2e)\), which explains the Meissner effect and the London equation. The Ginzburg-Landau paper is the origin of the Higgs mechanism. In the Tonomura et al.’s setting, the longitudinal mode of the vector potential becomes massive through the Higgs mechanism, which is known as the Meissner effect, that is, \(|\mathbf{A}_L(x)| \approx \exp(-|x|/\lambda_p)\) as \(|x| \to \infty\), with the penetration length \(\lambda_p = \sqrt{m/4\mu_0(e^*)^2 |\psi(\infty)|}\). Then, not only the magnetic field but also the longitudinal mode of the vector potential well decays at a long distance. So that not only the magnetic field but also the longitudinal mode does not exist, that is, \(B(x) = 0, \mathbf{A}_L(x) = 0\) in the region where the electron travels. Idealistically, we consider that there is the forbidden region, and inside the forbidden region, where the magnetic field and the longitudinal mode of the vector potential exist. The electron travels in the allowed region. In this allowed region, \(B(x) = 0, \mathbf{A}_L(x) = 0\), (neither leakage of the magnetic field nor the longitudinal mode of the vector potential). While the scalar mode is possible to exist \(\mathbf{A}_S(x) = \nabla \theta(x) \neq 0\), but as we have already shown, the scalar mode trivially does not contribute to the Aharonov-Bohm effect.

In this Tonomura et al.’s setting, when we apply the Stokes’ theorem, we must carefully use the theorem for the non-simply connected region in the form

\[
\int \int_D \mathbf{B}_n \cdot dS = \oint_{C_1} \mathbf{A} \cdot dx + \oint_{C_2} \mathbf{A} \cdot dx, \tag{12}\]

where \(\mathbf{B} = \nabla \times \mathbf{A}\) and \(C_1, C_2\) and \(D\) are shown in Figure 1 of the next section. We give the explanation of the generalized Stokes’ theorem in the next section.

In Tonomura et al.’s setting of the Aharonov-Bohm effect, we have \(\mathbf{A}(x) = \mathbf{A}_S(x) = \nabla \theta(x)\), and we have

\[
\oint_{C_1} \mathbf{A} \cdot dx = \int_{C_1} \nabla \theta(x) \cdot dx = \int_{C_1} d\theta(x) = 0, \tag{13}\]

\[
\oint_{C_2} \mathbf{A} \cdot dx = \int_{C_2} d\theta(x) = 0. \tag{14}\]

While, as there is no magnetic field in the region \(D\), we have

\[
\int \int_D \mathbf{B}_n \cdot dS = 0. \tag{15}\]

Then the Stokes’ theorem of Eq.(12) is trivially satisfied by using relations

\[
\int \int_D \mathbf{B}_n \cdot dS = 0, \quad \oint_{C_1} \mathbf{A} \cdot dx = 0, \quad \oint_{C_2} \mathbf{A} \cdot dx = 0. \tag{16}\]

In the Stokes’ theorem in the Tonomura et al.’s setting, we must notice the following fact

\[
\int \int_{D_0 + D} \mathbf{B}_n \cdot dS \neq \oint_{C_1} \mathbf{A} \cdot dx. \tag{17}\]

In the Tonomura et al.’s setting, we have \(\mathbf{A}(x) = \mathbf{A}_S(x) = \nabla \theta(x)\) in \(D\), then the right-hand side of Eq.(17) gives zero. While the magnetic field inside the forbidden region \(D_0\) gives the contribution to the left-hand side of the value \(\mu_0 n I a^2 \pi\). Therefore, the left-hand side is different from the right-hand side of Eq.(17).

In order to avoid the mistake, though it may be mathematically complicated, it will be better to calculate the phase shift only by using the vector potential, but not by using the magnetic field with the help of the Stokes’ theorem.

III. THE STOKES’ THEOREM IN THE NON-SIMPLY CONNECTED REGION

The Stokes’ theorem in the non-simply connected region of Figure 1 is given in the form

\[
\int \int_D \mathbf{B}_n \cdot dS = \oint_{C_1} \mathbf{A} \cdot dx + \oint_{C_2} \mathbf{A} \cdot dx, \tag{18}\]

where \(\mathbf{B} = \nabla \times \mathbf{A}\).

We explain the Stokes’ theorem in the non-simply connected region in the following. We start from

\[
\oint_{C_1} \mathbf{A} \cdot dx + \oint_{C_2} \mathbf{A} \cdot dx, \tag{19}\]

Then the Stokes’ theorem of Eq.(17) is trivially satisfied by using relations

\[
\int \int \mathbf{B}_n \cdot dS = 0, \quad \oint \mathbf{A} \cdot dx = 0. \tag{16}\]

In the Stokes’ theorem in the Tonomura et al.’s setting, we must notice the following fact

\[
\int \int_{D_0 + D} \mathbf{B}_n \cdot dS \neq \oint \mathbf{A} \cdot dx. \tag{17}\]

In the Tonomura et al.’s setting, we have \(\mathbf{A}(x) = \mathbf{A}_S(x) = \nabla \theta(x)\) in \(D\), then the right-hand side of Eq.(17) gives zero. While the magnetic field inside the forbidden region \(D_0\) gives the contribution to the left-hand side of the value \(\mu_0 n I a^2 \pi\). Therefore, the left-hand side is different from the right-hand side of Eq.(17).

In order to avoid the mistake, though it may be mathematically complicated, it will be better to calculate the phase shift only by using the vector potential, but not by using the magnetic field with the help of the Stokes’ theorem.
which is rewritten into the infinitely many line integrals around closed curves as in Figure 2, which is equivalent to the sum of the magnetic field perpendicular to the surface \( \int_{D} \mathbf{B} \cdot d\mathbf{S} \). While the two line integrals around closed curves \( C_1 \) and \( C_2 \) in Figure 1 can be transformed into the single line integral of the closed curve \( C = C_1 + C_4 + C_2 + C_3 \) as is given in Figure 3.

\[
\oint_{C_1} \mathbf{A} \cdot d\mathbf{x} + \oint_{C_2} \mathbf{A} \cdot d\mathbf{x} = \oint_{C_4} \mathbf{A} \cdot d\mathbf{x} + \oint_{C_2} \mathbf{A} \cdot d\mathbf{x} + \oint_{C_3} \mathbf{A} \cdot d\mathbf{x} = \oint_{C} \mathbf{A} \cdot d\mathbf{x},
\]

where \( C = C_1 + C_4 + C_2 + C_3 \) is the single closed curve, which does not enclose the forbidden region \( D_0 \). In this way, even for the non-simply connected region, the Stokes’ theorem is the theorem of the single line integral of the closed curve, which does not enclose the forbidden region. We can topologically shrink this closed line integral around the closed curve \( C \) into the infinitesimal closed line integral of the closed curve around any one point in the region \( D \).

For the Stokes’ theorem applied to the general non-simply connected region, we must take the single closed curve \( C \) in such a way as the closed curve can be topologically transformed to the infinitesimal closed curve around any one point in the region \( D \).
IV. SUMMARY AND DISCUSSIONS

We clarify issues of the Aharonov-Bohm effect. In the original setting of the Aharonov-Bohm, the longitudinal mode of the vector potential, which does not change under the gauge transformation so that it is the physical mode, gives the contribution of the observed Aharonov-Bohm effect. While the scalar mode of the vector potential, which changes under the gauge transformation so that it is the unphysical mode, gives no contribution to the Aharonov-Bohm effect. Then Aharonov-Bohm effect really occurs by the physical longitudinal mode in the original Aharonov-Bohm setting. In the setting of Tonomura et al., where the magnet is shielded by the superconducting material, not only the magnetic field but also the longitudinal mode of the vector potential become massive by the Meissner effect. Then not only the magnetic field but also the longitudinal mode does not come out to the region where the electron travels. In such setting, only the scalar mode of the vector potential exists, but the contribution to the Aharonov-Bohm effect from that mode is zero. Then, theoretically, the Aharonov-Bohm effect does not occur for Tonomura et al.’s setting. As people misuse the Stokes’ theorem for the non-simply connected region, we explain the correct Stokes’ theorem for the non-simply connected region.

In the quantum theory, the electron is treated as the wave, and the longitudinal mode gives the change of the phase, which causes the Aharonov-Bohm effect. In the classical theory, the electron is treated as the particle, and the longitudinal mode gives the change of the angular momentum. For the particle, there is no concept of the phase, so that there is no Aharonov-Bohm effect.

REFERENCES

[1] Aharonov Y, Bohm D. Significance of Electromagnetic Potentials in the Quantum Theory. Phys. Rev., 1959; 115: 485-491.
[2] Aharonov A, and Bohm D. Further Considerations on Electromagnetic Potentials in the Quantum Theory. Phys. Rev., 1961, 123: 1511-1524.
[3] Aharonov Y, Bohm D. Remarks on the Possibility of Quantum Electrodynamics without Potentials. Phys. Rev., 1962; 125: 2192-2193.
[4] Chambers RG, Shift of an Electron Interference Pattern by Enclosed Magnetic Flux. Phys. Rev. Lett., 1960; 5: 3-5.
[5] Fowler HA, Marton L, Simpson JA, Saddeth JA. Electron Interferometer Studies of Iron Whiskers. J. Appl. Phys., 1961; 32: 1153-1155.
[6] Möllenstedt G, Bayh W. Messung der kontinuierlichen Phasenschiebung von Elektronenwellen im kraftfeldfreien Raum durch das magnetische vektorpotential einer Luftspule. Naturwissenschaften, 1962; 49: 81-82.
[7] Boersch H, Hamisch H, Grohmann K, Wohleben D. Experimenteller Nachweis der Phasenschiebung von Elektronenwellen durch das magnetische Vektorpotential. Z. Phys., 1961; 165: 79-93.
[8] Bocchieri P, Loinger A. Nonexistence of the Aharonov-Bohm Effect. Nuovo Cimento, 1978; 47A: 475-482.
[9] Bocchieri P, Loinger A, Siragusa G. On the Aharonov-Bohm Effect. Nuovo Cimento. 1979; 51A: 1-17.
[10] Tonomura A, Osakabe N, Matsuda T, Kawasaki T, Endo J, Yano S, Yamada H. Evidence for Aharonov-Bohm Effect with Magnetic Field Completely Shielded from Electron Wave. Phys. Rev. Lett., 1986; 56: 792-795.
[11] Osakabe N, Matsuda T, Kawasaki T, Endo J, Tonomura A, Yano S, Yamada H. Experimental Confirmation of Aharonov-Bohm Effect using a Toroidal Magnetic Field Confined by a Superconductor. Phys. Rev., 1986; A34: 815-822.
[12] Ginzburg VL, Landau LD. On the Theory of Superconductivity. Zh. Eksp. Teor. Fiz., 1950; 20: 1064-1082.
[13] Abrikosov AA, Gor’kov LP. Contribution to the Theory of Superconducting Alloys with Paramagnetic Impurities. Sov. Phys. JETP, 1961; 12: 1243-1253.