



# The Aharonov–Bohm effect in a closed flux line

Ricardo Heras<sup>a</sup>

School of Physical Sciences, The Open University, Walton Hall, Milton Keynes MK7 6AA, UK

Received: 24 January 2022 / Accepted: 15 May 2022

© The Author(s), under exclusive licence to Società Italiana di Fisica and Springer-Verlag GmbH Germany, part of Springer Nature 2022

**Abstract** The Aharonov–Bohm (AB) effect was convincingly demonstrated using a micro-sized toroidal magnet but it is almost always explained using an infinitely-long solenoid or an infinitely-long flux line. The main reason for this is that the formal treatment of the AB effect considering a toroidal configuration turns out to be too cumbersome. But if the micro-sized toroidal magnet is modelled by a closed flux line of arbitrary shape and size then the formal treatment of the AB effect is exact, considerably simplified, and well-justified. Here we present such a treatment that covers in detail the electromagnetic, topological, and quantum-mechanical aspects of this effect. We demonstrate that the AB phase in a closed flux line is determined by a linking number and has the same form as the AB phase in an infinitely-long flux line which is determined by a winding number. We explicitly show that the two-slit interference shift associated with the AB effect in a closed flux line is the same as that associated with an infinitely-long flux line. We emphasise the topological nature of the AB phase in a closed flux line by demonstrating that this phase is invariant under deformations of the charge path, deformations of the closed flux line, simultaneous deformations of the charge path and the closed flux line, and the interchange between the charge path and the closed flux line. We also discuss the local and nonlocal interpretations of the AB effect in a closed flux line and introduce a non-singular gauge in which the vector potential vanishes in all space except on the surface surrounded by the closed flux line, implying that this vector potential is zero along the trajectory of the charged particle except on the crossing point where this trajectory intersects the surface bounded by the closed flux line, a result that questions the alleged physical significance of the vector potential and thereby the local interpretation of the AB effect.

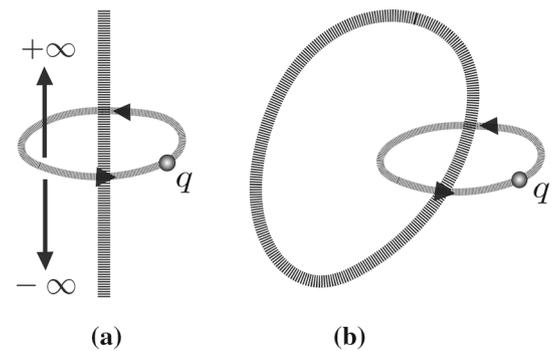
## 1 Introduction

Quantum mechanics predicts that the wave function of a charged particle encircling an infinitely-long solenoid enclosing a uniform magnetic flux accumulates the AB phase [1]. The charged particle moves in a non-simply connected region where there is no magnetic field and therefore there is no Lorentz force acting on the charge but there is a nonzero vector potential. The AB phase is topological because it depends on the winding number representing the number of times the charge carries out around the solenoid [2]. This topological feature is manifested in the fact that this phase is independent of the dynamics of the encircling charge. The AB phase admits a nonlocal interpretation according to which the magnetic field of the solenoid acts on the charged particle in regions for which this field is excluded (see, for example, the textbook of Rohrlich and Aharonov [2] for a representative view of this nonlocal interpretation). However, the most popular interpretation of this phase is that it is originated by the local action of the vector potential, in whose case this potential acquires a physical significance (see, for example, the textbook of Feynman [3] for a representative view of this local interpretation). The AB phase is physically manifested in a modified two-slit interference experiment, in which a shift in the interference pattern proportional to the AB phase is observed. This is the AB effect. Regardless of its physical interpretation, the AB effect has become an influential effect in many branches of physics from condensed matter physics to high-energy particle physics, fluid mechanics, gravitation and cosmology (see, for example, Cohen et al. [4], and references therein).

On the experimental side, the first reports on the detection of the AB effect were due to Chambers [5], Fowler et al. [6] and Boerch et al. [7], who used finite magnetic devices like magnetised whiskers and long solenoids. Möllenstedt and Bayh [8] used a tiny solenoid with a diameter as small as one micron. The use of a finite solenoid raised questions by several authors (see, for example, Peshkin and Tonomura [9], and Tonomura [10], and references therein) regarding the experimental verification of the AB effect by arguing that the electrons may not have been completely shielded from the magnetic field of the finite solenoid. But the concerns about the presence of a magnetic field in the electron paths were removed by a series of experiments reported by Tonomura et al. [11–13]. In a first instance, Tonomura et al. [11] employed a squared micro-sized toroidal magnet whose leakage field effects were confined to be sufficiently small in order to verify the AB effect. In a second instance, Tonomura et al. [12, 13] used a circular micro-sized toroidal magnet covered with a superconducting layer which, due to the Meissner effect, essentially confined the magnetic field of the toroidal magnet. This allowed a more definite experimental verification of the AB effect.

<sup>a</sup> e-mail: [ricardo.heras@ou.ac.uk](mailto:ricardo.heras@ou.ac.uk) (corresponding author)

**Fig. 1** **a** A charge moving around an infinitely-long flux line. **b** A charge moving around a closed flux line. Both configurations are defined in non-simply connected regions in which the AB effect exists



But why being experimentally demonstrated in a conclusive form using a toroidal configuration, is the AB effect generally explained using an idealised infinitely-long solenoid? We think that the answer deals with the fact that an exact treatment of the AB effect using a toroidal configuration and covering on equal footing its related electromagnetic, topological and quantum-mechanical features is a cumbersome task, which does not seem to have been reported so far. On the other hand, the posed question leads us to one of the more peculiar features of the AB effect: whenever the charged particle encircles a line of singularity, i.e. whenever it lies on a non-simply connected region, the existence of the AB effect does not depend on the particular geometry of the solenoid, which may be seen as a consequence of the topological character of this effect. Stated differently, in order for the wave function of the charged particle to accumulate the AB phase, the requirement of an idealised infinitely-long solenoid is sufficient but not necessary since this phase also arises by considering a less-idealised toroidal solenoid [14–16].

However, although in both an infinitely-long solenoid and a toroidal solenoid the AB effect arises, the mathematical treatment using the former solenoid is, as we have said before, considerably simpler than that using the latter solenoid because, among other reasons, the computation of the vector potential of the infinitely-long solenoid is much simpler compared to that of the toroidal solenoid [17–20]. For this reason we think that the AB effect is generally explained considering an infinitely-long solenoid instead of a toroidal solenoid.

An infinitely-long flux line and a closed flux line of arbitrary shape are also electromagnetic configurations defined in non-simply connected regions (see Fig. 1). As expected, in both configurations the AB phase arises. However, while the AB phase in an idealised infinitely-long flux line has been extensively discussed, the AB phase in the less-idealised closed flux line has received much less attention and only a few authors have laterally addressed it [21–23]. Since the more definite experimental verification of the AB effect relies on the use of a micro-sized superconducting toroidal magnet then the idea of modelling this magnet by a closed flux line of arbitrary shape and size is well justified.

In this paper, we present a detailed discussion of the AB effect in a closed flux line. The main purposes of this paper are the following: (i) to show that an exact and non-cumbersome treatment of the AB effect can be accomplished by considering a closed flux line of arbitrary shape, (ii) to emphasise the topological nature of the AB effect by showing that the AB phase arising in a closed flux line (determined by a linking number) has the same form than the AB phase arising in an infinitely-long flux line (determined by a winding number) and by demonstrating that the AB phase in a closed flux line is invariant under deformations of the charge path, deformations of the closed flux line, simultaneous deformations of the charge path and the closed flux line, and the interchange between the charge path and the closed flux line, (iii) to argue in favour of a nonlocal interpretation of the AB effect in a closed flux line by introducing a gauge that questions the local interpretation of this effect, and (iv) to discuss the difference in applying non-singular and singular gauge transformations in the AB effect and show that the latter transformations modify the magnetic field.

Our paper is organised as follows. In Sect. 2 we discuss the formal aspects of the electrodynamics of a closed magnetised flux line. We derive the vector potential of a closed flux line, show that it may be expressed as the gradient of a multi-valued function, and prove that its circulation is gauge invariant. In Sect. 3 we show that the circulation of the vector potential is topological because it depends on a linking number and is nonlocal because it is delocalised with respect to the magnetic flux, a result arising from the Stokes theorem applied to the examined non-simply connected region. In Sect. 4 we derive the AB phase that accumulates the wave function of a charged particle upon continuously encircling the closed flux line and show that this phase exhibits the same form as the AB phase in an infinitely-long flux line modulo a linking number that specifies the former phase and a winding number that specifies the latter phase. In Sect. 5 we present a novel treatment of the AB two-slit interference experiment using a closed flux line and show that the corresponding shift detected on the second screen of the interference device coincides with that using an infinitely-long flux line. In Sect. 6 we introduce four topological invariances of the AB phase in a closed flux line that enlighten the topological nature of this phase. In Sect. 7 we discuss the local and nonlocal interpretations of the AB effect in a closed flux line. We argue in favour of the latter interpretation and against the former interpretation by stressing that the vector potential is gauge-dependent and its circulation is spatially delocalised. In Sect. 8 we strengthen our objection against the physical reality of the vector potential and the local interpretation of the AB effect by introducing a non-singular gauge in which the vector potential vanishes in all space except on the surface surrounded by the closed flux line which implies that as the charge encircles the closed flux line, the vector potential is zero along the trajectory of the charge except on a point of this trajectory. In Sect. 9 we discuss the

subtle differences that exist in applying singular and non-singular gauge transformations in the AB effect and argue that only the latter transformations can consistently be applied in this effect because the former transformations modify the confined magnetic field. In Sect. 10 we summarise our main results. In Appendices A-E we demonstrate some equations relevant in our discussion.

## 2 Vector potential of a closed flux line

We can think of a closed flux line of arbitrary shape either as an infinitesimally thin closed magnetised solenoid or a closed line of magnetic dipoles. Any of these equivalent representations can be modelled using the steady electric current density

$$\mathbf{J} = \frac{c\Phi}{4\pi} \nabla \times \oint_{\mathcal{C}} \delta(\mathbf{x} - \mathbf{x}') d\mathbf{x}', \tag{1}$$

where  $\delta(\mathbf{x} - \mathbf{x}')$  is the Dirac delta function with  $\mathbf{x}$  being the field point and  $\mathbf{x}'$  the source point,  $\Phi = 4\pi\lambda$  is the flux through the closed magnetised line with  $\lambda$  being the magnetic dipole moment density per unit length, the line integral is evaluated along the closed flux line represented by the curve  $\mathcal{C}$ , the direction of the current is specified by the direction of the curve  $\mathcal{C}$ , and Gaussian units are adopted. The current satisfies  $\nabla \cdot \mathbf{J} = 0$  and is a magnetisation current  $\mathbf{J} = c\nabla \times \mathbf{M}$  where

$$\mathbf{M} = \frac{\Phi}{4\pi} \oint_{\mathcal{C}} \delta(\mathbf{x} - \mathbf{x}') d\mathbf{x}', \tag{2}$$

is the magnetisation vector confined along the curve  $\mathcal{C}$ . The associated magnetostatic equations are

$$\nabla \cdot \mathbf{B} = 0, \quad \nabla \times \mathbf{B} = \Phi \nabla \times \oint_{\mathcal{C}} \delta(\mathbf{x} - \mathbf{x}') d\mathbf{x}', \tag{3}$$

whose solution is given by the magnetic field

$$\mathbf{B} = \Phi \oint_{\mathcal{C}} \delta(\mathbf{x} - \mathbf{x}') d\mathbf{x}', \tag{4}$$

which is confined along the curve  $\mathcal{C}$ . From Eqs. (2) and (4) it follows the relation  $\mathbf{B} = 4\pi\mathbf{M}$  connecting the magnetic field with the magnetisation vector. To verify the homogeneous equation appearing in Eq. (3), we use  $\nabla \cdot \delta(\mathbf{x} - \mathbf{x}') d\mathbf{x}' = \nabla \delta(\mathbf{x} - \mathbf{x}') \cdot d\mathbf{x}'$  and  $\nabla \delta(\mathbf{x} - \mathbf{x}') = -\nabla' \delta(\mathbf{x} - \mathbf{x}')$  so that  $\nabla \cdot \mathbf{B} = -\Phi \oint_{\mathcal{C}} \nabla' \delta(\mathbf{x} - \mathbf{x}') \cdot d\mathbf{x}' = 0$  which holds because  $\oint_{\mathcal{C}} \nabla' \delta(\mathbf{x} - \mathbf{x}') \cdot d\mathbf{x}' = 0$  on account of the gradient theorem and the fact that  $\delta(\mathbf{x} - \mathbf{x}')$  is a single-valued function of  $\mathbf{x}'$ . The result  $\oint_{\mathcal{C}} \nabla \delta(\mathbf{x} - \mathbf{x}') \cdot d\mathbf{x}' = 0$  has also been shown by Kleinert [24, 25], deWit [26], and Kunin [27]. From the homogeneous equation in Eq. (3) it follows  $\mathbf{B} = \nabla \times \mathbf{A}$  where  $\mathbf{A}$  is the associated vector potential. This relation, the inhomogeneous equation in Eq. (3), the identity  $\nabla^2 \mathbf{F} = \nabla(\nabla \cdot \mathbf{F}) - \nabla \times (\nabla \times \mathbf{F})$ , and the adoption of the Coulomb gauge condition  $\nabla \cdot \mathbf{A} = 0$ , yield the Poisson equation

$$\nabla^2 \mathbf{A} = -\Phi \nabla \times \oint_{\mathcal{C}} \delta(\mathbf{x} - \mathbf{x}') d\mathbf{x}', \tag{5}$$

whose solution is given by the vector potential

$$\mathbf{A} = \frac{\Phi}{4\pi} \oint_{\mathcal{C}} \frac{(\mathbf{x}' - \mathbf{x}) \times d\mathbf{x}'}{|\mathbf{x} - \mathbf{x}'|^3}. \tag{6}$$

Using  $\nabla \times (d\mathbf{x}'/|\mathbf{x} - \mathbf{x}'|) = \nabla(1/|\mathbf{x} - \mathbf{x}'|) \times d\mathbf{x}'$  and  $\nabla(1/|\mathbf{x} - \mathbf{x}'|) = -\nabla'(1/|\mathbf{x} - \mathbf{x}'|)$  we can write Eq. (6) in the following form

$$\mathbf{A} = \frac{\Phi}{4\pi} \nabla \times \oint_{\mathcal{C}} \frac{d\mathbf{x}'}{|\mathbf{x} - \mathbf{x}'|}. \tag{7}$$

Considering  $\nabla^2(\nabla \times \mathbf{F}) = \nabla \times (\nabla^2 \mathbf{F})$  and  $\nabla^2(1/|\mathbf{x} - \mathbf{x}'|) = -4\pi\delta(\mathbf{x} - \mathbf{x}')$  we can verify that Eq. (7) satisfies Eq. (5):  $\nabla^2 \mathbf{A} = (\Phi/4\pi)\nabla \times \oint_{\mathcal{C}} \nabla^2(d\mathbf{x}'/|\mathbf{x} - \mathbf{x}'|) = -\Phi \nabla \times \oint_{\mathcal{C}} \delta(\mathbf{x} - \mathbf{x}') d\mathbf{x}'$ . Equation (7) satisfies the Coulomb gauge  $\nabla \cdot \mathbf{A} = 0$  on account of  $\nabla \cdot (\nabla \times \mathbf{F}) = 0$ . In Appendix A we show that the curl of Eq. (7) gives Eq. (4) while in Appendix B we show that Eq. (7) can be written as

$$\mathbf{A} = \frac{\Phi}{4\pi} \nabla \Omega_0 + \Phi \delta_S, \tag{8}$$

where  $\Omega_0$  is the single-valued solid angle subtended by the curve  $\mathcal{C}$  and defined by [24, 25]

$$\Omega_0 = \int_S \frac{(\mathbf{x}' - \mathbf{x}) \cdot d\mathbf{S}'}{|\mathbf{x} - \mathbf{x}'|^3}, \tag{9}$$

and  $\delta_S$  is the Dirac delta surface vector function defined by [24–27]

$$\delta_S = \int_S \delta(\mathbf{x} - \mathbf{x}') d\mathbf{S}', \tag{10}$$

where  $\mathcal{S}$  is the surface enclosed by the curve  $\mathcal{C}$  and  $d\mathcal{S}'$  the differential surface vector normal to  $\mathcal{S}$ . Accordingly, the vector potential of the closed flux line can be expressed as the sum of a term involving the gradient of the single-valued solid angle  $\nabla\Omega_0$  plus a term involving the delta function  $\delta_{\mathcal{S}}$ , which is localised on the surface  $\mathcal{S}$  enclosed by the curve  $\mathcal{C}$ . The function  $\Omega_0$  is said to be single-valued because it satisfies in all space the Schwarz integrability condition, according to which the crossed second partial derivatives applied to the function  $\Omega_0$  commute

$$(\partial^i \partial^j - \partial^j \partial^i)\Omega_0 = 0, \tag{11}$$

where index notation has been adopted in which summation on repeated indices is understood and  $\partial^i = (\nabla)^i$ —this condition for the single valuedness of functions based on the Schwarz integrability condition is discussed in Kleinert’s book on multi-valued fields [24]. In Appendix C we prove the relations  $\oint_{\mathcal{C}} \nabla\Omega_0 \cdot d\mathbf{x} = 0$  and  $\nabla \times \nabla\Omega_0 = 0$ , which allow us to demonstrate Eq. (11). On the other hand, we can verify that the curl of the potential defined in Eq. (8) gives the magnetic field specified in Eq. (4). With this purpose we first consider the relation

$$\nabla \times \delta_{\mathcal{S}} = \delta_{\mathcal{C}}, \tag{12}$$

where  $\delta_{\mathcal{C}}$  is a vector line Dirac delta defined along the curve  $\mathcal{C}$  and given by [24–27]

$$\delta_{\mathcal{C}} = \oint_{\mathcal{C}} \delta(\mathbf{x} - \mathbf{x}') d\mathbf{x}'. \tag{13}$$

Equation (12) has been mentioned by Kleinert [24,25] and explicitly demonstrated by deWit [26] and Kunin [27]. A proof of Eq. (12) goes as follows. Using  $\nabla \times [\delta(\mathbf{x} - \mathbf{x}')d\mathcal{S}'] = -d\mathcal{S}' \times \nabla\delta(\mathbf{x} - \mathbf{x}')$  and  $\nabla\delta(\mathbf{x} - \mathbf{x}') = -\nabla'\delta(\mathbf{x} - \mathbf{x}')$ , the curl of Eq. (10) gives  $\nabla \times \delta_{\mathcal{S}} = \int_{\mathcal{S}} d\mathcal{S}' \times \nabla'\delta(\mathbf{x} - \mathbf{x}')$  and using the Stokes theorem  $\int_{\mathcal{S}} d\mathcal{S}' \times \nabla'\delta(\mathbf{x} - \mathbf{x}') = \oint_{\mathcal{C}} \delta(\mathbf{x} - \mathbf{x}')d\mathbf{x}'$ , where  $\mathcal{C}$  is the boundary of  $\mathcal{S}$ , we obtain Eq. (12). The curl of Eq. (8) yields  $\nabla \times \mathbf{A} = [\Phi/(4\pi)]\nabla \times \nabla\Omega_0 + \Phi \nabla \times \delta_{\mathcal{S}}$  and since  $\Omega_0$  is single-valued then  $\nabla \times \nabla\Omega_0 = 0$  so that  $\nabla \times \mathbf{A} = \Phi \nabla \times \delta_{\mathcal{S}}$ . Using Eq. (12) we obtain  $\nabla \times \mathbf{A} = \Phi \delta_{\mathcal{C}}$ , whose right-hand side identifies with the magnetic field in Eq. (4).

Although the solid angle  $\Omega_0(\mathbf{x})$  is a single-valued function, it is a discontinuous function as it jumps by  $4\pi$  when the observation point  $\mathbf{x}$  crosses the surface  $\mathcal{S}$  [24,25]. This discontinuity has led several authors to the misconception that  $\Omega_0$  is a multi-valued function (see, for example, Zangwill [28], Schwinger et al. [29], and Eyges [30]). The fact that  $\Omega_0$  is single-valued and discontinuous has been emphasised by Kleinert [24,25] and has been explicitly demonstrated by Djurić [31]. We follow Kleinert [24,25] and make use of the Schwarz integrability condition to prove in Appendix C that  $\Omega_0$  is a single-valued function. We also note that the function  $\Omega_0$  depends on the choice of the shape of the surface  $\mathcal{S}$  [24,25]. We can express Eq. (8) in terms of the gradient of a multi-valued representation of the solid angle denoted as  $\Omega$ , which is continuous and independent of the choice of the surface  $\mathcal{S}$ . Kleinert [24,25] has shown the following result:

$$\nabla\Omega = \nabla\Omega_0 + 4\pi\delta_{\mathcal{S}}. \tag{14}$$

The function  $\Omega$  is a multi-valued function because it violates the Schwarz integrability condition [i.e.  $(\partial^i \partial^j - \partial^j \partial^i)\Omega \neq 0$ ]. Kleinert has pointed out the result [24,25]

$$(\partial^i \partial^j - \partial^j \partial^i)\Omega = 4\pi\varepsilon^{ijk}(\delta_{\mathcal{C}})_k, \tag{15}$$

where  $\varepsilon^{ijk}$  is the Levi-Civita symbol and  $(\delta_{\mathcal{C}})_k$  denotes the components of Eq. (13). In Appendix D we explicitly demonstrate Eq. (15). Using Eq. (15) we can express Eq. (8) in the useful form

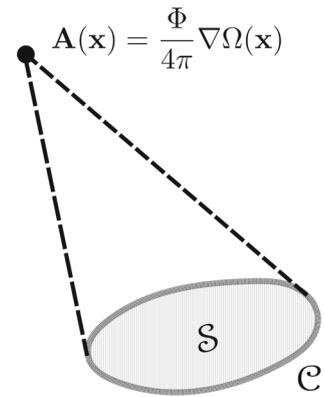
$$\mathbf{A} = \frac{\Phi}{4\pi} \nabla\Omega. \tag{16}$$

A pictorial description of the vector potential given by Eq. (16) is shown in Fig. 2. Since Eq. (16) can be written as the gradient of a function then we would have  $\nabla \times \mathbf{A} = 0$  and hence the vanishing of the magnetic field. Indeed, this is the case in all space except along  $\mathcal{C}$ . In fact, multiplying Eq. (15) by  $\varepsilon_{mij}$  we have  $\varepsilon_{mij}(\partial^i \partial^j - \partial^j \partial^i)\Omega = 4\pi\varepsilon_{mij}\varepsilon^{ijk}(\delta_{\mathcal{C}})_k$ , which implies  $2\varepsilon_{mij}\partial^i \partial^j \Omega = 8\pi(\delta_{\mathcal{C}})_m$  and therefore  $(\nabla \times \nabla\Omega)_m = 4\pi(\delta_{\mathcal{C}})_m$ , or equivalently

$$\nabla \times \nabla\Omega = 4\pi\delta_{\mathcal{C}}, \tag{17}$$

which gives  $\nabla \times \mathbf{A} = [\Phi/(4\pi)]\nabla \times \nabla\Omega = \Phi\delta_{\mathcal{C}}$  in agreement with Eq. (4). We should note that multi-valued functions, such as  $\Omega$  in Eq. (16), are characteristic in the electrodynamics of flux lines or magnetised strings which are defined in non-simply connected regions [24,25,32,33]. Two further examples of multi-valued functions appear in the following configurations: the vector potential outside an infinitely-long flux line in cylindrical coordinates [1]:  $\mathbf{A} = \Phi\hat{\phi}/(2\pi\rho)$  and the vector potential of the Dirac monopole in spherical coordinates [34]:  $\mathbf{A}_D = g(1 - \cos\theta)\hat{\phi}/(r \sin\theta)$ , where  $g$  is the magnetic charge. The former vector potential may be expressed as  $\mathbf{A} = \nabla\chi$  where  $\chi = \Phi\phi/(2\pi)$  is a multi-valued function that satisfies  $(\partial^i \partial^j - \partial^j \partial^i)\chi \neq 0$ . The latter vector potential satisfies  $\nabla \cdot (\nabla \times \mathbf{A}_D) \neq 0$  implying [24]:  $(\partial^i \partial^j - \partial^j \partial^i)(\mathbf{A}_D)_k \neq 0$ , which makes the Dirac monopole potential  $\mathbf{A}_D$  a multi-valued function. Discussions on electromagnetic aspects of the Dirac monopole can be found in Kleinert [24,25], Heras [35], and Shnir [36].

**Fig. 2** Vector potential of a closed flux line evaluated at the point  $x$ . The solid angle is subtended by the surface  $S$  enclosed by the curve  $C$  representing the shape of the closed flux line



Let us now discuss the gauge invariance of the circulation of the vector potential of the closed flux line. The circulation  $\oint_C A \cdot dx$  taken along an arbitrary closed path  $C$  is invariant under the gauge transformation  $A' = A + \nabla \Lambda$  where  $\Lambda$  is the corresponding gauge function. The gauge transformation must be a non-singular gauge transformation, i.e. one in which the gauge function  $\Lambda$  is a single-valued function satisfying the Schwarz integrability condition [24,25]  $(\partial^i \partial^j - \partial^j \partial^i) \Lambda = 0$ . Moreover, this condition should hold in all space and not only within a finite region. Therefore,

$$\oint_C A' \cdot dx = \oint_C A \cdot dx + \oint_C \nabla \Lambda \cdot dx = \oint_C A \cdot dx, \tag{18}$$

which follows because  $\oint_C \nabla \Lambda \cdot dx = 0$  on account of the single-valuedness of  $\Lambda$ . Let us insist that Eq. (18) holds whenever  $\Lambda$  is a single-valued function. If  $\Lambda$  were a multi-valued function then we would have  $\oint_C \nabla \Lambda \cdot dx \neq 0$  and this would imply  $\oint_C A' \cdot dx \neq \oint_C A \cdot dx$  and therefore the breaking of the gauge invariance of the circulation of the vector potential.

We also note that if the Coulomb gauge condition in the transformed potential  $\nabla \cdot A' = 0$  is preserved then the corresponding gauge function, in addition to be a single-valued function, it should be a restricted gauge function satisfying  $\nabla^2 \Lambda = 0$ . Nevertheless, this additional requirement is not necessary for the validity of Eq. (18) since  $A'$  need not be in the Coulomb gauge as we will see in Sec. 8 of this paper.

### 3 Topology and nonlocality of the circulation of the vector potential

The space containing a closed flux line is non-simply connected because there is a non-removable line of singularity along the curve  $C$  where the magnetic field is confined. This non-simply connected space implies some interesting topological and nonlocal features of the circulation of the vector potential of the closed flux line which will now be discussed. Let us first discuss the topological aspect. The circulation of the vector potential given by Eq. (6) along an arbitrary closed path  $C$  gives

$$\oint_C A \cdot dx = \frac{\Phi}{4\pi} \oint_C \oint_e \frac{[(x' - x) \times dx'] \cdot dx}{|x - x'|^3}, \tag{19}$$

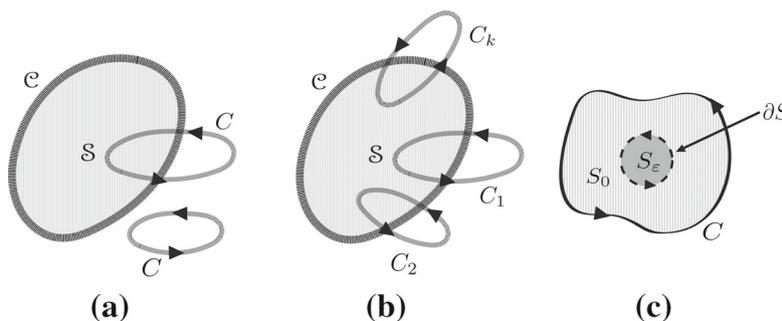
where it is assumed that the path  $C$  does not intersect the curve  $C$  where the closed flux line is defined. Using  $[(x' - x) \times dx'] \cdot dx = (x - x') \cdot (dx \times dx')$  in Eq. (19) we obtain

$$\oint_C A \cdot dx = \Phi \left[ \frac{1}{4\pi} \oint_C \oint_e \frac{(x - x') \cdot (dx \times dx')}{|x - x'|^3} \right]. \tag{20}$$

The quantity within the brackets is identified in the general case with the Gauss linking number  $l$  (or linking integral) which is defined as [17,24,37]

$$\frac{1}{4\pi} \oint_C \oint_e \frac{(x - x') \cdot (dx \times dx')}{|x - x'|^3} = \begin{cases} l(C, C) & \text{if } C \text{ encloses } C \\ 0 & \text{otherwise} \end{cases} \tag{21}$$

and represents the number of times the path  $C$  encloses the curve  $C$ . The linking number can be positive or negative depending on the direction of  $C$  and  $C$  and whether if  $C$  crosses above or below the surface  $S$  bounded by  $C$  as projected in a two-dimensional plane. From this result it follows that changing the direction of  $C$  and  $C$  will change the sign of the linking number, i.e.  $l(-C, C) = -l(C, C)$  and  $l(C, -C) = -l(C, C)$ . Moreover changing simultaneously the direction of  $C$  and  $C$  leaves the linking number invariant:  $l(-C, -C) = l(C, C)$ . Inserting Eq. (21) in Eq. (20) it follows



**Fig. 3** **a** If the circulation of the vector potential  $A$  along the path  $C$  encircles the closed flux line of shape  $\mathcal{C}$  then this configuration is defined in a non-simply connected region and  $\oint_C A \cdot dx = l\Phi$  where  $l$  is the linking number of the paths  $C$  and  $\mathcal{C}$ . If the circulation of the vector potential  $A$  along the path  $C$  does not encircle the closed flux line of shape  $\mathcal{C}$  then the configuration is simply-connected and  $\oint_C A \cdot dx = 0$ . **b** The circulation of the vector potential  $A$  along the different closed paths  $C_1, C_2, \dots, C_k$  encircling the closed flux line of shape  $\mathcal{C}$  and corresponding to the same linking number  $l$  are equivalent:  $\oint_{C_1} A \cdot dx = \oint_{C_2} A \cdot dx = \dots = \oint_{C_k} A \cdot dx$ . **c** The circulation  $\oint_C A \cdot dx$  is taken along a path  $C$  greater than the boundary  $\partial S_\epsilon$  of the infinitesimal surface  $S_\epsilon$  pierced by the closed flux line. Since  $\oint_{C > \partial S_\epsilon} A \cdot dx = \int_{S_\epsilon} B \cdot dS$  holds then the circulation of the vector potential is spatially delocalised from the surface where the magnetic flux is non-vanishing

$$\oint_C A \cdot dx = \begin{cases} l\Phi & \text{if } C \text{ encloses } \mathcal{C} \\ 0 & \text{otherwise} \end{cases} \tag{22}$$

which shows the topological nature of the circulation of the vector potential of the closed flux line: if the path  $C$  encloses the closed flux line of shape  $\mathcal{C}$  then this configuration is non-simply connected and the circulation of the vector potential  $A$  accumulates  $l$  times the magnetic flux  $\Phi$ . The product  $l\Phi$  is a constant quantity and therefore it is independent of the path  $C$ . If the path  $C$  does not encircle the closed flux line of shape  $\mathcal{C}$  then the configuration is locally simply connected and the circulation of  $A$  vanishes (see Fig. 3a). By locally simply connected we mean here any finite region of space that does not include the closed flux line, i.e. the closed line of singularity  $\mathcal{C}$ . In such regions  $\delta_S = \int_S \delta(x - x') dS' = 0$  because  $x$  is not on any point on the surface  $S$  and therefore Eq. (8) takes the form  $A(x \notin S) = [\Phi/(4\pi)]\nabla\Omega_0$  whose circulation vanishes because  $\Omega_0$  is a single-valued function.

A comment on the interpretation of the path  $C$  and the curve  $\mathcal{C}$  in Eq. (20) is pertinent. The path  $C$  identifies with the closed trajectory on which is defined the circulation of the vector potential  $A$  while the curve  $\mathcal{C}$  identifies with the closed magnetic flux line which is assumed to be stationary. The physical quantity that can be identified along the direction of the curve  $\mathcal{C}$  is the steady current  $J$  defined in Eq. (1) which flows continuously within the closed flux line. In short:  $C$  is the path of the circulation of  $A$  while  $\mathcal{C}$  is the closed flux line whose direction is followed by its steady current  $J$ .

Equation (14) can be used to obtain two convenient ways to define the linking number, which in turn can be applied to the circulation of the vector potential of the closed flux line. We first observe that Eq. (16) implies  $\nabla\Omega = 4\pi A/\Phi$  which combines with Eq. (6) to give the relation

$$\nabla\Omega = \oint_{\mathcal{C}} \frac{(x' - x) \times dx'}{|x - x'|^3}, \tag{23}$$

whose circulation reads

$$\oint_C \nabla\Omega \cdot dx = \oint_C \oint_{\mathcal{C}} \frac{(x - x') \cdot (dx \times dx')}{|x - x'|^3}, \tag{24}$$

where we have used  $[(x' - x) \times dx'] \cdot dx = (x - x') \cdot (dx \times dx')$ . The right-hand side of Eq. (24) is equal to  $4\pi$  times the linking number specified in Eq. (21) and therefore it follows that

$$\frac{1}{4\pi} \oint_C \nabla\Omega \cdot dx = \begin{cases} l & \text{if } C \text{ encloses } \mathcal{C} \\ 0 & \text{otherwise} \end{cases} \tag{25}$$

which shows that the circulation of the gradient of the multi-valued solid angle along a path  $C$  enclosing  $\mathcal{C}$  is non-vanishing and proportional to the linking number. On the other hand, Eq. (14) tells us that  $\nabla\Omega = \nabla\Omega_0 + 4\pi\delta_S$  which is used in the left-hand side of Eq. (25) to obtain

$$\frac{1}{4\pi} \oint_C \nabla\Omega_0 \cdot dx + \oint_C \delta_S \cdot dx = \begin{cases} l & \text{if } C \text{ encloses } \mathcal{C} \\ 0 & \text{otherwise} \end{cases} \tag{26}$$

But  $\oint_C \nabla\Omega_0 \cdot dx = 0$  because  $\Omega_0$  is a single-valued function and therefore

$$\oint_C \delta_S \cdot dx = \begin{cases} l & \text{if } C \text{ crosses } S \\ 0 & \text{otherwise} \end{cases} \tag{27}$$

which shows that the circulation of the Dirac surface vector  $\delta_S$  is non-vanishing when the closed path  $C$  crosses the surface  $S$ . Kleinert [24,25] has discussed Eqs. (25) and (27). The use of Eq. (25) in Eq. (22) gives

$$\oint_C \mathbf{A} \cdot d\mathbf{x} = \frac{\Phi}{4\pi} \oint_C \nabla\Omega \cdot d\mathbf{x} = \begin{cases} l\Phi & \text{if } C \text{ encloses } \mathcal{C} \\ 0 & \text{otherwise} \end{cases} \tag{28}$$

while the use of Eq. (27) in Eq. (22) gives

$$\oint_C \mathbf{A} \cdot d\mathbf{x} = \Phi \oint_C \delta_S \cdot d\mathbf{x} = \begin{cases} l\Phi & \text{if } C \text{ crosses } S \\ 0 & \text{otherwise} \end{cases} \tag{29}$$

Equation (28) tells us that the circulation of  $\mathbf{A}$  is proportional to the circulation of  $\nabla\Omega$  which is non-vanishing when the path  $C$  encircles the curve  $\mathcal{C}$ . Equation (29) tells us that the circulation of  $\mathbf{A}$  is non-vanishing and proportional to the circulation of  $\delta_S$  along any path  $C$  crossing the surface  $S$ —which enlightens the fact that the path  $C$  is irrelevant as long as it crosses the surface  $S$ . Since the circulation of  $\mathbf{A}$  is insensitive to the path  $C$  then we can consider  $C_1, C_2 \dots C_k$  different paths each one encircling the closed flux line and corresponding to the same linking number. Therefore,

$$\oint_{C_1} \mathbf{A} \cdot d\mathbf{x} = \oint_{C_2} \mathbf{A} \cdot d\mathbf{x} = \dots = \oint_{C_k} \mathbf{A} \cdot d\mathbf{x}. \tag{30}$$

The paths  $C_1, C_2 \dots C_k$  are homotopically equivalent and therefore we could not distinguish if the flux  $\Phi$  is connected with the circulation of  $\mathbf{A}$  along  $C_1$  or along  $C_2$  or along  $C_k$  (see Fig. 3b). This indistinguishability is a manifestation of the topology of the circulation of the vector potential of the closed flux line which lies in a non-simply connected region.

Let us now discuss the nonlocal aspect of the circulation of the vector potential. We can apply the Stokes theorem to the left-hand side of Eq. (22) but we need to be careful in doing so when the path  $C$  encloses the closed flux line of shape  $\mathcal{C}$  because this circulation is defined in a non-simply connected region. Applying the Stokes theorem we obtain

$$\oint_{C=\partial S} \mathbf{A} \cdot d\mathbf{x} = \int_S \nabla \times \mathbf{A} \cdot d\mathbf{S}, \tag{31}$$

where  $\partial S$  is the boundary of the total surface  $S$  enclosed by the path  $C$ . We can write this total surface as  $S = S_0 + S_\epsilon$ , where  $S_0$  is the surface that excludes the closed flux line and  $S_\epsilon$  is the infinitesimal surface pierced by the closed flux line (see Fig. 3c). Of course,  $C = \partial S > \partial S_\epsilon$  because the path  $C$  encircles the closed flux line. The notation  $C > \partial S_\epsilon$  states that the length of the curve  $C$  is greater than the length of the boundary  $\partial S_\epsilon$  of the infinitesimal surface  $S_\epsilon$  pierced by the closed flux line. It then follows that Eq. (31) takes the form

$$\oint_{C=\partial S} \mathbf{A} \cdot d\mathbf{x} = \int_{S_0} \nabla \times \mathbf{A} \cdot d\mathbf{S} + \int_{S_\epsilon} \nabla \times \mathbf{A} \cdot d\mathbf{S}. \tag{32}$$

Since the surface  $S_0$  excludes the closed flux line then the first term on the right-hand side vanishes  $\int_{S_0} \nabla \times \mathbf{A} \cdot d\mathbf{S} = 0$  because  $\nabla \times \mathbf{A} = 0$  along the surface  $S_0$ . Therefore, the Stokes theorem gives

$$\oint_{C>\partial S_\epsilon} \mathbf{A} \cdot d\mathbf{x} = \int_{S_\epsilon} \nabla \times \mathbf{A} \cdot d\mathbf{S}, \tag{33}$$

which expresses a nonlocal relation: while the circulation  $\oint_{C>\partial S_\epsilon} \mathbf{A} \cdot d\mathbf{x}$  is defined outside the closed flux line, the flux  $\int_{S_\epsilon} \nabla \times \mathbf{A} \cdot d\mathbf{S}$  is defined on the infinitesimal surface  $S_\epsilon$  pierced by the closed line of singularity  $\mathcal{C}$ , i.e. the sides of Eq. (33) are defined in different regions of space implying a nonlocal connection between them. This nonlocality is a consequence of having applied the Stokes theorem in a non-simply connected region. To have a better physical picture of Eq. (33) we can write in the right-hand side the magnetic field  $\nabla \times \mathbf{A} = \mathbf{B}$  which is non-vanishing along the infinitesimal surface  $S_\epsilon$ . This gives the relation

$$\oint_{C>\partial S_\epsilon} \mathbf{A} \cdot d\mathbf{x} = \int_{S_\epsilon} \mathbf{B} \cdot d\mathbf{S}, \tag{34}$$

and this shows that the circulation of the vector potential evaluated outside the closed flux line is delocalised with respect to the flux of the magnetic field confined along  $\mathcal{C}$ . In other words, there is a nonlocal relation between the circulation of  $\mathbf{A}$  and the flux of  $\mathbf{B}$ .

By applying the Stokes theorem to the vector  $\delta_S$ , it follows that

$$\oint_{C=\partial S} \delta_S \cdot d\mathbf{x} = \int_S \nabla \times \delta_S \cdot d\mathbf{S}, \tag{35}$$

where  $C$  is the boundary of the surface  $S$ . Using Eq. (35) and the relation  $\nabla \times \delta_S = \delta_{\mathcal{C}}$  given in Eq. (12), we obtain

$$\oint_{C=\partial S} \delta_S \cdot d\mathbf{x} = \int_S \delta_{\mathcal{C}} \cdot d\mathbf{S}, \tag{36}$$

which connects the circulation of the vector  $\delta_S$  with the surface integral of the vector  $\delta_C$ . Equations (36) and Eq. (29) give an alternative definition of the linking number

$$\int_S \delta_C \cdot dS = \begin{cases} l & \text{if } C \text{ crosses } S \\ 0 & \text{otherwise} \end{cases} \tag{37}$$

The surface in Eq. (37) can be written as  $S = S_0 + S_\varepsilon$  where  $S_0$  is the surface not pierced by the closed line of singularity  $C$  while  $S_\varepsilon$  is the infinitesimal surface pierced by the curve  $C$ . Thus

$$\int_{S_0} \delta_C \cdot dS + \int_{S_\varepsilon} \delta_C \cdot dS = \begin{cases} l & \text{if } C \text{ crosses } S \\ 0 & \text{otherwise} \end{cases} \tag{38}$$

But  $\int_{S_0} \delta_C \cdot dS = 0$  because  $\delta_C = 0$  along  $S_0$  ( $C$  never crosses  $S_0$ ) and therefore Eq. (38) reduces to

$$\int_{S_\varepsilon} \delta_C \cdot dS = \begin{cases} l & \text{if } C \text{ crosses } S_\varepsilon \\ 0 & \text{otherwise} \end{cases} \tag{39}$$

where  $S_\varepsilon \in S$ . Equations (37) and (39) are equivalent representations of the linking number with the former being discussed by Kleinert [25]. Using Eq. (39) and the relation  $\Phi \delta_C = \mathbf{B}$  it follows that

$$\int_{S_\varepsilon} \mathbf{B} \cdot dS = \begin{cases} l\Phi & \text{if } C \text{ crosses } S_\varepsilon \\ 0 & \text{otherwise} \end{cases} \tag{40}$$

and this shows that the quantity  $\int_{S_\varepsilon} \mathbf{B} \cdot dS$  is the accumulated magnetic flux along the closed line of singularity  $C$ . Both Eq. (22) and Eq. (40) are equivalent and both are in agreement with Eq. (34). Now, from Eqs. (30) and (34) we obtain the relation

$$\oint_{C_k > \partial S} \mathbf{A} \cdot d\mathbf{x} = \dots = \oint_{C_2 > \partial S_\varepsilon} \mathbf{A} \cdot d\mathbf{x} = \oint_{C_1 > \partial S_\varepsilon} \mathbf{A} \cdot d\mathbf{x} = \int_{S_\varepsilon} \mathbf{B} \cdot dS. \tag{41}$$

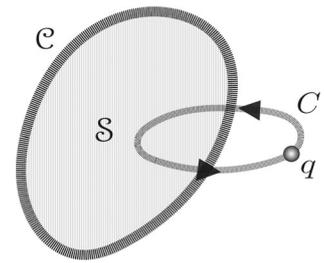
in which the equalities on the left-hand side express a manifest ambiguity because we cannot distinguish if the flux  $\int_{S_\varepsilon} \mathbf{B} \cdot dS$  is connected with the circulation of  $\mathbf{A}$  along  $C_1 > \partial S$  or along  $C_2 > \partial S$ , or along  $C_k > \partial S_\varepsilon$ . In other words, the circulations in Eq. (41) are delocalised with respect to the magnetic flux, an expected result since they are not functions of point. We should emphasise the difference between applying the Stokes theorem in a simply connected region and in the non-simply connected region considered here. In the former application of the theorem there is only a single curve  $C = \partial S$  representing the boundary  $\partial S$  of the surface  $S$ . In the latter application of the theorem there can be  $k$  curves  $C_k > \partial S$  all of them greater (i.e. delocalised) than the infinitesimal boundary  $\partial S_\varepsilon$  of the surface  $S_\varepsilon$  pierced by the closed line of singularity  $C$ . Interestingly, similar features can be found in other electromagnetic configurations defined in non-simply connected regions. In two recent papers [38, 39] we have shown that a similar relation to that in Eq. (41) holds in the circulations of vector potentials define in three non-simply connected configurations: (i) the circulation of the magnetic vector potential outside an infinitely-long magnetic solenoid, (ii) the circulation of the electric vector potential outside an infinitely-long electric solenoid, and (iii) the circulation of the sum of the magnetic and electric vector potentials outside an infinitely-long dual solenoid which confines its electric and magnetic fields. The fact that the circulations of vector potential of different geometrical configurations are delocalised enlightens their topological character.

We can now draw the lessons we have learned so far about the peculiarities of the electromagnetism of a closed flux line. The magnetisation current  $\mathbf{J}$  in Eq. (1) yields the vector potential  $\mathbf{A}$  in Eq. (6) and the magnetic field  $\mathbf{B}$  in Eq. (4). The closed flux line involves a closed line of singularity  $C$  and then one can apply the definition of the Gauss linking number [Eq. (21)] and the Stokes theorem [Eq. (31)] in this non-simply connected region. Both mathematical tools lead to Eq. (41) which unambiguously shows the nonlocality of the circulation of  $\mathbf{A}$  with respect to the flux of the confined magnetic field  $\mathbf{B}$ . One can then argue that the background of this nonlocality is of topological nature. To see this we use Eqs. (27) and (39) to obtain the beautiful topological relation

$$\oint_{C > \partial S_\varepsilon} \delta_S \cdot d\mathbf{x} = l = \int_{S_\varepsilon} \delta_C \cdot dS, \tag{42}$$

which contains all the geometrical information about the nonlocality of the electromagnetic relation in Eq. (34). Let us emphasise that Eq. (42) is a purely topological quantity devoid of any physical content. When multiplied by the flux  $\Phi$ , Eq. (42) yields the electromagnetic relation in Eq. (34)—see, Eqs. (22) and (40). After seeing these results, we cannot avoid saying that *topology dictates nonlocality!*

**Fig. 4** As the charged particle encircles the closed flux line along the path  $C$  its wave function accumulates the AB phase  $\delta = l\Phi/(\hbar c)$  where  $l$  is the linking number representing the number of times  $C$  encircles the closed flux line of shape  $\mathcal{C}$  and which encloses the surface  $\mathcal{S}$



### 4 AB phase in a closed flux line

Consider a non-relativistic particle of mass  $m$  and charge  $q$  that is continuously moving around the closed flux line (see Fig. 4). The corresponding time-dependent Schrödinger equation is given by

$$i\hbar \frac{\partial \Psi}{\partial t} = \frac{1}{2m} \left( -i\hbar \nabla - \frac{q}{c} \mathbf{A} \right)^2 \Psi, \tag{43}$$

where  $\mathbf{A}$  is the vector potential of the closed flux line. Since  $\mathbf{A}$  can be written as the gradient of a function as seen in Eq. (16) then a solution of Eq. (43) can be obtained by multiplying the free solution  $\Psi_0$ —which satisfies Eq. (43) when  $\mathbf{A} = 0$ —by a suitable local phase factor

$$\Psi(\mathbf{x}, t) = e^{[iq/(\hbar c)] \int_{\Gamma} \mathbf{A}(\mathbf{x}') \cdot d\mathbf{x}'} \Psi_0(\mathbf{x}, t), \tag{44}$$

where the line integral in the phase is taken along the charge path  $\Gamma$  from a fixed reference point  $\mathcal{O}$  to the variable point  $\mathbf{x}$ . We assume that the points  $\mathcal{O}$  and  $\mathbf{x}$  never lie on the closed flux line while the path  $\Gamma$  never crosses it. As the charge continuously encircles the closed flux line, it follows that any charge path  $\Gamma$  can be decomposed as  $\Gamma = C + \gamma$  where  $C$  is any closed path that accounts for the number of times the charge encircles the closed flux line and  $\gamma$  is any non-closed path that accounts for the open trajectory that the charge takes before completing another turn around the closed flux line. Thus, we can write

$$\int_{\Gamma} \mathbf{A} \cdot d\mathbf{x}' = \oint_C \mathbf{A} \cdot d\mathbf{x}' + \int_{\gamma} \mathbf{A} \cdot d\mathbf{x}'. \tag{45}$$

Using this relation the wave function in Eq. (44) takes the form  $\Psi = e^{[iq/(\hbar c)] [\oint_C \mathbf{A} \cdot d\mathbf{x}' + \int_{\gamma} \mathbf{A} \cdot d\mathbf{x}']} \Psi_0$ . As the path  $C$  encircles the closed flux line of shape  $\mathcal{C}$  it follows from Eq. (22) that  $\oint_C \mathbf{A} \cdot d\mathbf{x}' = l\Phi$  and thus,

$$\int_{\Gamma} \mathbf{A} \cdot d\mathbf{x}' = l\Phi + \int_{\gamma} \mathbf{A} \cdot d\mathbf{x}'. \tag{46}$$

Using Eqs. (44) and (46) we can write

$$\Psi(\mathbf{x}, t) = [e^{ilq\Phi/(\hbar c)}] e^{[iq/(\hbar c)] \int_{\gamma} \mathbf{A}(\mathbf{x}') \cdot d\mathbf{x}'} \Psi_0(\mathbf{x}, t), \tag{47}$$

which states that after the charged particle takes  $l$  turns around the closed flux line, its wave function picks up the phase factor  $e^{ilq\Phi/(\hbar c)}$  and thereby it accumulates the AB phase

$$\delta = l \frac{q\Phi}{\hbar c}, \tag{48}$$

where  $l$  is the linking number of the charge path  $C$  around the closed flux line of shape  $\mathcal{C}$ . A first observation is that the derived AB phase in a closed flux line has the same form than the AB phase in an infinitely-long flux line localised along the  $z$ -axis. In fact, the latter phase is given by  $\delta = nq\Phi/(\hbar c)$  where  $n$  is the winding number representing the number of times the charge encircles the infinitely-long flux line. Thus, the only difference between the AB phase in a closed flux line and the AB phase in an infinitely-long flux line is the linking number that specifies the former phase and the winding number that specifies the latter phase. However, both AB phases are equivalent in the sense that the topological numbers  $l$  and  $n$  have the same interpretation in these phases: the number of times a charged particle encircles a line of singularity (closed or infinitely-long). The fact that the AB phase arises in different geometrical configurations is a manifestation of its topological nature.

We can write the AB phase in Eq. (48) in other different forms each of which enlightens some of its properties. The form in Eq. (48) shows that the AB phase in a closed flux line is topological because it depends on the linking number  $l$  and is independent of the dynamics of the encircling charge, i.e. independent of the charge path  $C$ . A manifestly gauge-invariant form of this phase is obtained by using Eqs. (22) and (48),

$$\delta = \frac{q}{\hbar c} \oint_C \mathbf{A} \cdot d\mathbf{x}, \tag{49}$$

which is gauge invariant on account of the gauge invariance of the circulation of the vector potential [see Eq. (18)]. Another form of the AB phase is obtained by using Eqs. (40) and (48):

$$\delta = \frac{q}{\hbar c} \int_{S_\varepsilon} \mathbf{B} \cdot d\mathbf{S}. \tag{50}$$

This form of the AB phase is conceptually interesting because it clearly admits a nonlocal interpretation: while the charge  $q$  is moving along the path  $C$  defined outside the closed flux line, the flux  $\int_{S_\varepsilon} \mathbf{B} \cdot d\mathbf{S}$  is defined along the infinitesimal surface  $S_\varepsilon$  where the magnetic field  $\mathbf{B}$  is non-vanishing. This suggests a nonlocal interaction between the electric charge and the magnetic field confined along the closed flux line. Another form of the AB phase may be obtained by inserting the flux  $\Phi = 4\pi\lambda$  in Eq. (48):  $\delta = 4\pi l q \lambda / (\hbar c)$ , according to which the AB phase may be seen as the result of the nonlocal interaction between the electric charge  $q$  moving outside the closed flux line and the magnetic dipole moment linear density  $\lambda$  confined along the closed flux line. Another interesting form of the AB phase can be obtained using Eq. (25) and Eq. (48):

$$\delta = \frac{q\Phi}{4\pi\hbar c} \oint_C \nabla\Omega \cdot d\mathbf{x}, \tag{51}$$

where  $\Omega$  is the multi-valued solid angle defined by Eq. (14). This form shows that the AB phase in a closed flux line has a high range of validity since Eq. (51) is coordinate-independent. This fact shows that the AB phase is a quite general phenomenon that occurs whenever a charged particle encircles a magnetised closed flux line. From Eqs. (27) and (48) we can obtain another form of the AB phase,

$$\delta = \frac{q\Phi}{\hbar c} \oint_C \delta_S \cdot d\mathbf{x}, \tag{52}$$

which is interesting in the sense that it explicitly shows that the AB phase in a closed flux line is not continuously accumulated as the charged particle travels the path  $C$  but discretely picked up as the charged particle crosses the surface  $S$  along which the function  $\delta_S$  is non-vanishing. This result is independent of the shape of the surface  $S$ .

### 5 Quantum interference

The AB phase is physically manifested in a modified two-slit interference experiment for charged particles in which an infinitely-long solenoid that confines its magnetic flux is placed between the two screens in the interference device. The presence of the solenoid causes a shift in the corresponding interference pattern which is proportional to the AB phase. This experiment has been discussed in early treatments of the AB effect (see, for example, Feynman’s textbook [3]). However, a detailed theoretical treatment of this interference effect has not been done for the case of the closed flux line.

Consider a double-slit interference experiment in which identical charged particles propagate from a source, pass into the two slits of a first screen, and are finally detected on the second screen. The computation of the corresponding wave function is somewhat laborious but we can take advantage of a similar calculation given by Kobe [40]. Therefore, we will follow arguments similar to those given by Kobe [40]. The following assumptions are made: (i) Charged particles are localised in the  $x$ - $y$  plane and their motion perpendicular to the screens is treated classically while their motion parallel to the screens is treated quantum-mechanically. This assumption is valid if the velocity of the charged particles parallel to the screens is sufficiently high [40]. Under this condition, the charges will follow classical paths along the  $y$ -axis. (ii) Charged particles are emitted with constant velocity  $v$  in the  $y$ -direction which is perpendicular to the screens and have random velocity along the  $x$ -axis. This will allow us to use a one-dimensional wave function along the  $x$ -axis to describe the charged particles. (iii) The slits are infinitely-long in the  $z$ -direction and are Gaussian slits instead of rectangular slits. This will not change our main conclusions but will help in simplifying calculations [40].

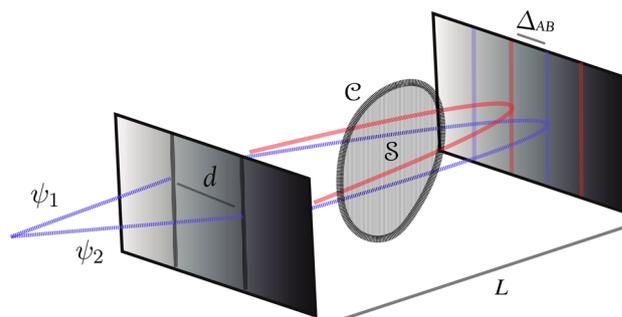
In order to obtain the wave function for points on the second screen let us first consider the case in which there is only one slit localised at  $+x_0$  and the closed flux line is absent. A charged particle propagates at the time 0 from the origin to a point  $x_a$  at the later time  $t_a$  on the first screen localised a distance  $y_a = vt_a$  where  $v$  is the constant velocity along the  $y$ -axis. The charge then passes through a slit centred at  $x_a = +x_0$  and continues to propagate until reaching a variable point  $x_b$  on the second screen at the later variable time  $t_b$  located a distance  $y_b = vt_b$ . The time taken for the charge to travel from  $x_a$  to  $x_b$  is then  $t_b - t_a$ . The corresponding wave function at  $x_b$  and  $t_b$  should then satisfy the one-dimensional time-dependent Schrödinger equation

$$i\hbar \frac{\partial \Psi_0}{\partial t_b} = -\frac{\hbar^2}{2m} \frac{\partial^2 \Psi_0}{\partial x_b^2}. \tag{53}$$

Kobe [40] has calculated the corresponding wave function

$$\Psi_0(x_b, t_b) = \sqrt{\frac{m}{2\pi i\hbar[t_b + i\alpha t_a(t_b - t_a)]}} \exp\left[-\frac{(1 - i\chi)(x_b - v_0 t_b)^2}{2(\Delta x)^2} + i\beta \frac{(x_b - x_0)^2}{(t_b - t_a)} + i\beta \frac{x_0^2}{t_a}\right], \tag{54}$$

**Fig. 5** Two-slit interference setup. Blue lines indicate partial wave packets that pass through the slits when the closed flux line is absent while red lines indicate partial wave packets that pass through the slits when the closed flux line is inserted between the two screens. The presence of the closed flux line gives rise to the additional shift  $\Delta_{AB}$  detected on the second screen



where  $\beta = m/2\hbar$ ,  $\alpha = \hbar/mv_0$  in which  $b$  corresponds to a rectangular slit of width  $\sqrt{\pi}b$ ,  $v_0 = x_0/t_a$  is the average velocity of the charge in the  $x$ -direction,  $(\Delta x)^2 = (bt_b/t_a)^2 + [\hbar(t_b - t_a)/mb]^2$  is the square of the total broadening where  $(bt_b/t_a)$  is the classical Gaussian broadening arising due to the propagation of the particle from the origin to the slit at time  $t_a$  and from this slit to the observation point on the second screen at the later time  $t_b - t_a$  and  $\hbar(t_b - t_a)/mb$  is the quantum-mechanical broadening due to the uncertainty of the momentum of the charged particle along the  $x$ -axis  $m\Delta v_x = \hbar/b$  in going through a slit of width  $b$ , and  $\chi = (bt_b/t_a)/[\hbar(t_b - t_a)/mb]$  is the ratio of the classical to the quantum-mechanical broadening. Let us now carry out the two-slit interference experiment in which the first slit is localised at  $+x_0$  and the second slit is localised at  $-x_0$ . From the principle of superposition the wave function can be written as  $\Psi_0 = \Psi_0^1 + \Psi_0^2$ , where  $\Psi_0^1$  is the partial wave passing through the first slit centred at  $+x_a$  and  $\Psi_0^2$  is the partial wave passing through the second slit centred at  $-x_a$ . The partial wave  $\Psi_0^1$  is of the same form as Eq. (54) while the partial wave  $\Psi_0^2$  can be obtained by making the replacements  $+x_0 \rightarrow -x_0$  and  $v_0 \rightarrow -v_0$  in Eq. (54). Therefore

$$\begin{aligned} \Psi_0(x_b, t_b) = & \sqrt{\frac{m}{2\pi i \hbar [t_b + i\alpha t_a(t_b - t_a)]}} \exp\left[i\beta \frac{x_0^2}{t_a}\right] \\ & \times \left\{ \exp\left[-\frac{(1 - i\chi)(x_b - v_0 t_b)^2}{2(\Delta x)^2} + i\beta \frac{(x_b - x_0)^2}{(t_b - t_a)}\right] \right. \\ & \left. + \exp\left[-\frac{(1 - i\chi)(x_b + v_0 t_b)^2}{2(\Delta x)^2} + i\beta \frac{(x_b + x_0)^2}{(t_b - t_a)}\right] \right\}. \end{aligned} \tag{55}$$

Using Eq. (55) and doing some simplifications, we obtain the probability density

$$\begin{aligned} |\Psi_0(x_b, t_b)|^2 = & \frac{m}{4\pi^2 \hbar^2 [t_b^2 + \alpha^2 t_a^2 (t_b - t_a)^2]} \left\{ \exp\left[-\frac{(x_b - v_0 t_b)^2}{(\Delta x)^2}\right] + \exp\left[-\frac{(x_b + v_0 t_b)^2}{(\Delta x)^2}\right] \right. \\ & \left. + 2 \exp\left[-\frac{[(x_b - v_0 t_b)^2 + (x_b + v_0 t_b)^2]}{2(\Delta x)^2}\right] \right. \\ & \left. \times \cos\left[\frac{\chi[(x_b - v_0 t_b)^2 - (x_b + v_0 t_b)^2]}{2(\Delta x)^2} + \frac{\beta[(x_b - x_0)^2 - (x_b + x_0)^2]}{(t_b - t_a)}\right] \right\}. \end{aligned} \tag{56}$$

In order for the interference effect to be significant it is necessary that [40]:  $v_0 t_b \gg \Delta x$ , i.e. that the total broadening  $\Delta x$  should be much larger than the average distance travelled by the charge in the  $x$ -direction through one of the slits. Under this condition we can write  $(x_b - v_0 t_b)^2/(\Delta x)^2 \approx x_b^2/(\Delta x)^2$  and  $(x_b + v_0 t_b)^2/(\Delta x)^2 \approx x_0^2/(\Delta x)^2$  and thus the probability density reduces to

$$|\Psi_0(x_b, t_b)|^2 = \frac{m e^{-x_0^2/(\Delta x)^2}}{\pi^2 \hbar^2 [t_b^2 + \alpha^2 t_a^2 (t_b - t_a)^2]} \cos(\delta_0/2), \tag{57}$$

where  $\delta_0 = [(2m x_b x_0)/\hbar(t_a - t_b)]$  is the phase angle. Using the de Broglie relation  $\lambda = \hbar/p$  where  $p = mv_y$  is the constant momentum along the  $y$ -axis and letting  $d = 2x_0$  be the distance between the two slits and  $L = v(t_a - t_b) = y_a - y_b$  the distance in the  $y$ -direction between the two screens, it follows that  $\delta_0 = dx_b/L\lambda$  and therefore a point  $x_b$  detected on the second screen reads

$$x_b = \frac{L\lambda}{d} \delta_0. \tag{58}$$

Let us now repeat the two-slit interference experiment but this time with the closed flux line

of shape  $\mathcal{C}$  inserted within the two slits as seen in Fig. 5. As in the case without the closed flux line, we first consider the solution when there is one slit only localised at the point  $x_a = +x_0$ . The magnetic field vanishes outside the closed flux line but the vector potential is non-vanishing. Therefore we are interested in solving the time-dependent Schrödinger equation

$$i\hbar \frac{\partial \Psi}{\partial t_b} = \frac{1}{2m} \left( -i\hbar \frac{\partial}{\partial x_b} - \frac{q}{c} A_x \right)^2 \Psi, \tag{59}$$

where  $A_x = A_x(x_b, y_b, 0)$  is the  $x$  component of the vector potential evaluated at  $x_b, y_b,$  and  $0$ . Since the vector potential is the gradient of a scalar function  $A = [\Phi/(4\pi)]\nabla\Omega$  as seen in Eq. (16) then we can follow the same approach as that in Eq. (43) but applied to our one-dimensional problem. We multiply  $\Psi_0$  in Eq. (54) times a suitable phase factor and obtain

$$\Psi(x_b, t_b) = e^{[iq/(\hbar c)]\int_0^{(x_b, y_b, 0)} A(x') \cdot dx'} \Psi_0(x_b, t_b), \tag{60}$$

where the path of the line integral goes from the origin, passes through the slit centred at  $x_a = +x_0$ , and ends at  $x_b, y_b, 0$ , with  $x_b$  being a variable point. Considering the two slits it follows that we can write  $\Psi = \Psi_1 + \Psi_2$  where  $\Psi_1$  is the partial wave going through the first slit and  $\Psi_2$  is the partial wave going through the second slit. We can then write  $\Psi_1 = e^{[iq/(\hbar c)]\int_1 A \cdot dx'} \Psi_0^1$  and  $\Psi_2 = e^{[iq/(\hbar c)]\int_2 A \cdot dx'} \Psi_0^2$  where  $\int_1$  denotes the path of the line integral that goes from the origin, passes through the first slit centred at  $+x_0$ , and ends at  $x_b, y_b, 0$ , while  $\int_2$  denotes the path of the line integral that goes from the origin, passes through the second slit centred at  $-x_0$ , and ends at  $x_b, y_b, 0$ . Therefore we can write

$$\Psi(x_b, t_b) = [\Psi_0^1(x_b, t_b) + e^{[iq/(\hbar c)]\int_2 A(x') \cdot dx' - \int_1 A(x') \cdot dx'} \Psi_0^2(x_b, t_b)] e^{[iq/(\hbar c)]\int_1 A(x') \cdot dx'}. \tag{61}$$

Assuming that the difference of the trajectories 1 and 2 forms a closed path  $C = 2 - 1$  encircling the closed flux line once then it follows from Eq. (22) that the phase difference in Eq. (61) becomes

$$\frac{q}{\hbar c} \left[ \int_2 A(x') \cdot dx' - \int_1 A(x') \cdot dx' \right] = \frac{q}{\hbar c} \oint_{C=2-1} A(x') \cdot dx' = \frac{q\Phi}{\hbar c}, \tag{62}$$

where the closed path  $C = 2 - 1$  corresponds to the linking number  $l = 1$  around the closed flux line of shape  $\mathcal{C}$  and we assume that the direction of  $\mathcal{C}$  is properly oriented so that the linking number is positive. Using Eq. (62) the wave function in Eq. (61) becomes

$$\Psi(x_b, t_b) = [\Psi_0^1(x_b, t_b) + e^{iq\Phi/(\hbar c)} \Psi_0^2(x_b, t_b)] e^{[iq/(\hbar c)]\int_1 A(x') \cdot dx'}. \tag{63}$$

The partial wave  $\Psi_0^1$  is of the same form as Eq. (54) while the partial wave  $\Psi_0^2$  can be calculated by making the replacements  $+x_0 \rightarrow -x_0$  and  $v_0 \rightarrow -v_0$  in Eq. (54). After some simplifications, the probability density becomes

$$\begin{aligned} |\Psi(x_b, t_b)|^2 &= \frac{m}{4\pi^2\hbar^2[t_b^2 + \alpha^2 t_a^2(t_b - t_a)^2]} \left\{ \exp\left[-\frac{(x_b - v_0 t_b)^2}{(\Delta x)^2}\right] + \exp\left[-\frac{(x_b + v_0 t_b)^2}{(\Delta x)^2}\right] \right. \\ &\quad \left. + 2 \exp\left[-\frac{[(x_b - v_0 t_b)^2 + (x_b + v_0 t_b)^2]}{2(\Delta x)^2}\right] \right. \\ &\quad \left. \times \cos\left[\frac{\chi[(x_b - v_0 t_b)^2 - (x_b + v_0 t_b)^2]}{2(\Delta x)^2} + \frac{\beta[(x_b - x_0)^2 - (x_b + x_0)^2]}{(t_b - t_a)} - \frac{q\Phi}{\hbar c}\right] \right\}. \tag{64} \end{aligned}$$

Since the AB phase  $q\Phi/(\hbar c) = \delta$  is generally not a multiple of  $2\pi$  (this case will be shortly addressed) it follows that there is a non-vanishing interference effect attributed to the this phase which should manifest in the two-slit interference effect. The total interference will be non-negligible provided  $v_0 t_b \gg \Delta x$  and under this condition we can approximate  $(x_b - v_0 t_b)^2/(\Delta x)^2 \approx x_b^2/(\Delta x)^2$  and  $(x_b + v_0 t_b)^2/(\Delta x)^2 \approx x_0^2/(\Delta x)^2$ , and therefore the probability density in Eq. (64) reduces to

$$|\Psi(x_b, t_b)|^2 = \frac{m e^{-x_0^2/(\Delta x)^2}}{\pi^2\hbar^2[t_b^2 + \alpha^2 t_a^2(t_b - t_a)^2]} \cos(\delta'/2), \tag{65}$$

where  $\delta' = [(2mx_b x_0)/\hbar(t_a - t_b)] + (q\Phi/\hbar c)$  is the phase angle. Using  $\tilde{\lambda} = \hbar/p$  and letting  $d = 2x_0$  be the distance between the two slits and  $L = v_y(t_a - t_b) = y_b - y_a$  the distance in the  $y$ -direction between the two screens, it follows that  $\delta' = d[x_b + L\tilde{\lambda}q\Phi/(d\hbar c)]/(L\tilde{\lambda})$ . Therefore the effect of the closed flux line physically manifests in a shift of the observation point  $x_b \rightarrow x_b + \Delta_{AB}$ , where

$$\Delta_{AB} = \frac{L\tilde{\lambda} q\Phi}{d \hbar c}, \tag{66}$$

denotes the AB shift. A first observation is that the derived AB shift in a closed flux line is the same as the AB shift corresponding to an infinitely-long flux line [40]. This result enlightens the topological nature of the AB effect since  $\Delta_{AB}$  arises in different geometrical configurations. On the other hand, Eq. (66) can be interpreted as a nonlocal effect: the shift  $\Delta_{AB}$  explicitly depends on the flux  $\Phi$  but it is evaluated in a point on the second screen where this flux is zero. This suggests that the shift  $\Delta_{AB}$  manifests a nonlocality in which the particles of charge  $q$  are affected by the magnetic field in a region for which this field vanishes.

We can write Eq. (66) in other different forms each one which enlightens some of its properties. Inserting the flux  $\Phi = 4\pi\lambda$  in Eq. (66) it follows that

$$\Delta_{AB} = 4\pi \frac{L\tilde{\lambda} q\lambda}{d \hbar c}, \tag{67}$$

which admits a nonlocal interpretation according to which the shift  $\Delta_{AB}$  originates from the nonlocal interaction between the electric charge  $q$  and the magnetic dipole moment linear density  $\lambda$ . Equation (66) can be written in terms of the AB phase  $\delta = q\Phi/(\hbar c)$ , i.e.

$$\Delta_{AB} = \frac{L\bar{\lambda}}{d}\delta. \tag{68}$$

Using Eqs. (51) and (68) together with the de Broglie relation  $\bar{\lambda} = \hbar/p$  we obtain

$$\Delta_{AB} = \frac{Lq\Phi}{4\pi pcd} \oint_{C=2-1} \nabla\Omega \cdot dx. \tag{69}$$

This equation shows that the AB shift  $\Delta_{AB}$  depends on the circulation of the gradient of the multi-valued representation of the solid angle  $\Omega$  defined in Eq. (14). This result is independent of the path  $C$ , i.e. independent of the dynamics of the charged particles having the momentum  $p$ —interestingly, Eq. (69) may be interpreted as a classical relation in the sense that it explicitly involves only classical pieces. Another form of  $\Delta_{AB}$  follows from using Eqs. (52) and (68) together with  $\bar{\lambda} = \hbar/p$ ,

$$\Delta_{AB} = \frac{Lq\Phi}{4\pi pcd} \oint_{C=2-1} \delta_S \cdot dx. \tag{70}$$

As may be seen, it only suffices for one of the partial wave packets to cross the surface  $S$  along the path 1 while the other to surround it along the path 2 to produce the interference shift in Eq. (70) (see Fig. 5). This result is independent of the paths of the partial waves or the surface  $S$ .

Let us now discuss the case in which the interference effect reflected in the shift in Eq. (66) is non-observable. This happens when we assume that the flux through the closed flux line is quantised, i.e. when  $\Phi = N\Phi_0$  where  $N$  is an integer and  $\Phi_0 = 2\pi\hbar c/e$  is the flux quantum with  $e$  being the electron’s charge. Moreover the electric charge is quantised  $q = n_e e$  with  $n_e$  being an integer and therefore the probability density reads  $|\Psi|^2 = |\Psi_0^1 + e^{iq\Phi/(\hbar c)}\Psi_0^2|^2 = |\Psi_0^1 + \Psi_0^2|^2$  where  $e^{iq\Phi/(\hbar c)} = e^{i2\pi n_e N} = 1$  on account of  $n_e N$  being an integer. It then follows that any interference effect due to the magnetic flux becomes unobservable. This is reflected in the corresponding shift  $x_b + \Delta_{AB} = x_b + n_e N 2\pi L\bar{\lambda}/d$  which is equivalent to Eq. (58) when the closed flux line is absent. In other words, shifting the phase angle  $\delta_0 \rightarrow \delta_0 + 2\pi n_e N$  does not produce any observable effect because the spectral lines detected on the second screen of the two-slit device merely get relabelled but are otherwise unchanged. In the experiments of Tonomura [12, 13] et al. a micro-sized ferromagnetic toroidal magnet covered with a superconducting layer was used. Characteristically, in superconducting rings the flux is quantised in the form  $\Phi = N\Phi_0/2$  where the factor 2 arises due to the corresponding Cooper pairs in the superconductor. If in addition we only consider electrons  $q = e$  then it follows that  $|\Psi|^2 = |\Psi_0^1 + e^{iq\Phi/(\hbar c)}\Psi_0^2|^2 = |\Psi_0^1 + e^{i\pi N}\Psi_0^2|^2$ . Thus, if  $N$  is even then  $e^{i\pi N} = 1$  and there is no AB effect but if  $N$  is odd then  $e^{i\pi N} = -1$  and there is the AB effect. In the practice, Tonomura et al. [12, 13] did not use the two-slit setup discussed here but a novel interference method based on electron holography [10–13]. However, the success of Tonomura et al. [12, 13] relied on verifying that when  $N$  was odd then the AB effect was detected.

### 6 Topological invariances of the AB phase in a closed flux line

In this section, we will introduce four topological invariances of the AB phase in a closed flux line. Let  $C$  be the charge path and  $\mathcal{C}$  the shape of the closed flux line. The following topological invariances of the AB phase in a closed flux line follow:

**(a) Deformations of the charge path.** The AB phase in Eq. (48) is invariant under deformations of the charge path from the initial path  $C$  to the final path  $C'$ . This invariance is expressed by

$$\delta(C, \mathcal{C}) = \delta(C', \mathcal{C}); \quad \{C \rightarrow C'\}, \tag{71}$$

where  $\delta(C, \mathcal{C})$  is the AB phase corresponding to the charge path  $C$  around the closed flux line of shape  $\mathcal{C}$  and  $\delta(C', \mathcal{C})$  is the AB phase corresponding to the deformed charge path  $C'$  around the closed flux line. Both phases can be written as

$$\delta(C, \mathcal{C}) = \frac{q\Phi}{4\pi\hbar c} \oint_C \oint_{\mathcal{C}} \frac{(\mathbf{x} - \mathbf{x}') \cdot (d\mathbf{x} \times d\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|^3}, \quad \delta(C', \mathcal{C}) = \frac{q\Phi}{4\pi\hbar c} \oint_{C'} \oint_{\mathcal{C}} \frac{(\mathbf{x} - \mathbf{x}') \cdot (d\mathbf{x} \times d\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|^3}. \tag{72}$$

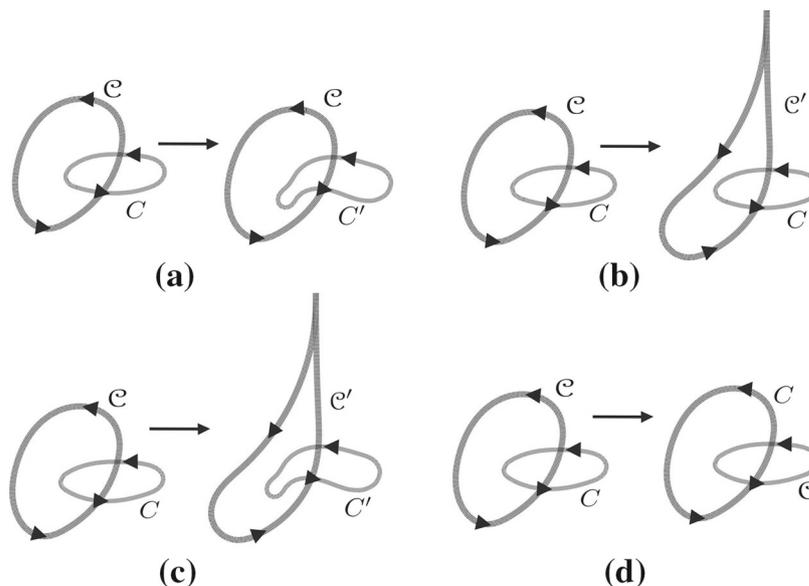
To demonstrate Eq. (71) we apply the transformation  $C \rightarrow C'$  to the quantity

$$\frac{1}{4\pi} \oint_C \oint_{\mathcal{C}} \frac{(\mathbf{x} - \mathbf{x}') \cdot (d\mathbf{x} \times d\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|^3}, \tag{73}$$

and obtain

$$\frac{1}{4\pi} \oint_C \oint_{\mathcal{C}} \frac{(\mathbf{x} - \mathbf{x}') \cdot (d\mathbf{x} \times d\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|^3} = \frac{1}{4\pi} \oint_{C'} \oint_{\mathcal{C}} \frac{(\mathbf{x} - \mathbf{x}') \cdot (d\mathbf{x} \times d\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|^3}, \tag{74}$$

**Fig. 6** **a** The AB phase is invariant under deformations of the charge path. **b** The AB phase is invariant under deformations of the closed flux line. **c** The AB phase is invariant under simultaneous deformations of the charge path and the closed flux line. **d** The AB phase is invariant under the interchange between the charge path and the closed flux line



which is explicitly proved in Appendix E. We also note that the relation in Eq. (74) has been demonstrated by Gelca [41]. The use of Eq. (74) implies Eq. (71) because of Eq. (72).

**(b) Deformations of the closed flux line.** The AB phase in Eq. (48) is invariant under deformations of the closed flux line from the initial shape  $\mathcal{C}$  to the final shape  $\mathcal{C}'$

$$\delta(C, \mathcal{C}) = \delta(C, \mathcal{C}'); \quad \{\mathcal{C} \longrightarrow \mathcal{C}'\}, \tag{75}$$

where  $\delta(C, \mathcal{C})$  is the AB phase corresponding to the charge path  $C$  around the closed flux line of shape  $\mathcal{C}$  and  $\delta(C, \mathcal{C}')$  is the AB phase corresponding to the charge path around the closed flux line of deformed shape  $\mathcal{C}'$ . These phases can be written as

$$\delta(C, \mathcal{C}) = \frac{q\Phi}{4\pi\hbar c} \oint_C \oint_{\mathcal{C}} \frac{(\mathbf{x} - \mathbf{x}') \cdot (d\mathbf{x} \times d\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|^3}, \quad \delta(C, \mathcal{C}') = \frac{q\Phi}{4\pi\hbar c} \oint_C \oint_{\mathcal{C}'} \frac{(\mathbf{x} - \mathbf{x}') \cdot (d\mathbf{x} \times d\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|^3}. \tag{76}$$

To demonstrate Eq. (75) we apply the transformation  $\mathcal{C} \rightarrow \mathcal{C}'$  to Eq. (73) and obtain

$$\frac{1}{4\pi} \oint_C \oint_{\mathcal{C}} \frac{(\mathbf{x} - \mathbf{x}') \cdot (d\mathbf{x} \times d\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|^3} = \frac{1}{4\pi} \oint_C \oint_{\mathcal{C}'} \frac{(\mathbf{x} - \mathbf{x}') \cdot (d\mathbf{x} \times d\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|^3}, \tag{77}$$

a result proved Appendix E. The use of Eq. (77) implies Eq. (75) because of Eq. (76).

**(c) Deformations of the charge path and the closed flux line.** The AB phase in Eq. (48) is invariant under simultaneous deformations of the charge path from  $C$  to  $C'$  and the shape of the closed flux line from  $\mathcal{C}$  to  $\mathcal{C}'$ . This invariance is expressed by the transformation

$$\delta(C, \mathcal{C}) = \delta(C', \mathcal{C}'); \quad \{C \longrightarrow C', \mathcal{C} \longrightarrow \mathcal{C}'\}, \tag{78}$$

where  $\delta(C, \mathcal{C})$  is the AB phase corresponding to the charge path  $C$  around the closed flux line of shape  $\mathcal{C}$  and  $\delta(C', \mathcal{C}')$  is the AB phase corresponding to the deformed charge path  $C'$  around the deformed closed flux line of shape  $\mathcal{C}'$ . These two AB phases can be written in the following form

$$\delta(C, \mathcal{C}) = \frac{q\Phi}{4\pi\hbar c} \oint_C \oint_{\mathcal{C}} \frac{(\mathbf{x} - \mathbf{x}') \cdot (d\mathbf{x} \times d\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|^3}, \quad \delta(C', \mathcal{C}') = \frac{q\Phi}{4\pi\hbar c} \oint_{C'} \oint_{\mathcal{C}'} \frac{(\mathbf{x} - \mathbf{x}') \cdot (d\mathbf{x} \times d\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|^3}. \tag{79}$$

To demonstrate Eq. (78) we apply the transformations  $C \rightarrow C'$  and  $\mathcal{C} \rightarrow \mathcal{C}'$  to Eq. (73) and obtain

$$\frac{1}{4\pi} \oint_C \oint_{\mathcal{C}} \frac{(\mathbf{x} - \mathbf{x}') \cdot (d\mathbf{x} \times d\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|^3} = \frac{1}{4\pi} \oint_{C'} \oint_{\mathcal{C}'} \frac{(\mathbf{x} - \mathbf{x}') \cdot (d\mathbf{x} \times d\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|^3}, \tag{80}$$

which is proved in Appendix E. The use of Eq. (80) implies Eq. (78) because of Eq. (79).

**(d) Interchange between the charge path and the closed flux line.** The AB phase in Eq. (48) is invariant under the simultaneous interchange between the charge path  $C$  and the closed flux line of shape  $\mathcal{C}$ . This interchange is represented by the transformation

$$\delta(C, \mathcal{C}) = \delta(\mathcal{C}, C); \quad \{C \longleftrightarrow \mathcal{C}\}, \tag{81}$$

where the AB phases  $\delta(C, \mathcal{C})$  and  $\delta(\mathcal{C}, C)$  can be written in the following form

$$\delta(C, \mathcal{C}) = \frac{q\Phi}{4\pi\hbar c} \oint_C \oint_{\mathcal{C}} \frac{(\mathbf{x} - \mathbf{x}') \cdot (d\mathbf{x} \times d\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|^3}, \quad \delta(\mathcal{C}, C) = \frac{q\Phi}{4\pi\hbar c} \oint_{\mathcal{C}} \oint_C \frac{(\mathbf{x} - \mathbf{x}') \cdot (d\mathbf{x} \times d\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|^3}. \tag{82}$$

The proof of Eq. (81) follows from the well-known relation [37,41]

$$\frac{1}{4\pi} \oint_C \oint_{\mathcal{C}} \frac{(\mathbf{x} - \mathbf{x}') \cdot (d\mathbf{x} \times d\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|^3} = \frac{1}{4\pi} \oint_{\mathcal{C}} \oint_C \frac{(\mathbf{x} - \mathbf{x}') \cdot (d\mathbf{x} \times d\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|^3}, \tag{83}$$

which implies Eq. (81) because of Eq. (82).

As may be seen, the AB phase in a closed flux line is a truly topological quantity which is manifested in the fact that this phase is invariant under the topological transformations given in Eqs. (71), (75), (78), and (81) (a pictorial description of these topological invariances is shown in Fig. 6). The reason for these topological invariances is simple: the linking number  $l(C, \mathcal{C})$  of the curves  $C$  and  $\mathcal{C}$  defined in Eq. (21) is invariant under deformations of  $C$ , deformations of  $\mathcal{C}$ , simultaneous deformations of  $C$  and  $\mathcal{C}$ , and the simultaneous interchange between  $C$  and  $\mathcal{C}$ . Accordingly,

$$l(C, \mathcal{C}) = l(C', \mathcal{C}), \quad l(C, \mathcal{C}) = l(C, \mathcal{C}'), \quad l(C, \mathcal{C}) = l(C', \mathcal{C}'), \quad l(C, \mathcal{C}) = l(\mathcal{C}, C), \tag{84}$$

which follow from Eqs. (74), (77), (80), and (83). Since the AB phase in a closed flux line is proportional to the linking number then the topological invariances of the linking number are translated into the AB phase. Moreover, any quantity proportional to the AB phase will also share these topological invariances. For example, consider the AB shift in Eq. (66) which can be written as  $\Delta_{AB} = L\bar{\chi}\delta(C, \mathcal{C})/d$  and which corresponds to  $l = 1$ . Considering Eq. (84) it follows that

$$\Delta_{AB} = \frac{L\bar{\chi}}{d}\delta(C, \mathcal{C}) = \frac{L\bar{\chi}}{d}\delta(C', \mathcal{C}) = \frac{L\bar{\chi}}{d}\delta(C, \mathcal{C}') = \frac{L\bar{\chi}}{d}\delta(C', \mathcal{C}') = \frac{L\bar{\chi}}{d}\delta(\mathcal{C}, C). \tag{85}$$

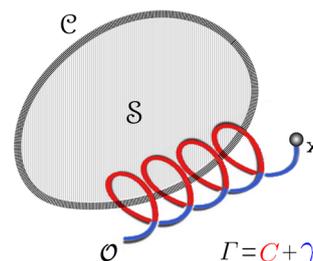
The second equality in Eq. (85) shows that the shift  $\Delta_{AB}$  is invariant under deformations of the path  $C = 2 - 1$  formed by the difference of the trajectories of the partial wave packets  $\Psi_1$  and  $\Psi_2$  in the two-slit interference effect. The third equality in Eq. (85) shows that the shift  $\Delta_{AB}$  is insensitive to deformations of the closed flux line of shape  $\mathcal{C}$ . The fourth equality in Eq. (85) shows that the shift  $\Delta_{AB}$  is invariant under simultaneous deformations of the path  $C$  and the closed flux line of shape  $\mathcal{C}$ . The fifth equality in Eq. (85) shows that the shift  $\Delta_{AB}$  is invariant under the simultaneous interchange of the path  $C$  and the closed flux line of shape  $\mathcal{C}$ . In particular, the third equality in Eq. (85) can be useful from an experimental viewpoint given that a small toroidal magnet (such as that employed by Tonomura et al. [11–13]) may be modelled by a closed flux line and the topological invariance  $L\bar{\chi}\delta(C, \mathcal{C})/d = \Delta_{AB} = L\bar{\chi}\delta(C, \mathcal{C}')/d$  implies that any possible deviation from a circular toroidal magnet in an experiment is insensitive to the AB shift.

The invariance  $\delta(C, \mathcal{C}) = \delta(\mathcal{C}, C)$  is physically interesting. The basis of this invariance is the relation  $l(C, \mathcal{C}) = l(\mathcal{C}, C)$  which is well-known in knot theory [41] and was discovered by Gauss in 1833 [37]! As already noted, this Gaussian relation admits a simple mathematical interpretation: the linking number  $l$  is invariant under the simultaneous interchange of  $C$  and  $\mathcal{C}$ , i.e.  $C \rightarrow \mathcal{C}$  and  $\mathcal{C} \rightarrow C$ . It is clear that  $\delta(C, \mathcal{C})$  is the AB phase that accumulates the wave function of a charge  $q$  when it travels the path  $C$  around the closed flux line of shape  $\mathcal{C}$ . Therefore the relation  $\delta(C, \mathcal{C}) = \delta(\mathcal{C}, C)$  tells us that  $\delta(\mathcal{C}, C)$  is the AB phase that accumulates the wave function of a charge  $q$  when it travels the path  $\mathcal{C}$  around the closed flux line of shape  $C$ . This means that the charge  $q$  and the steady current  $\mathbf{J}$  interchange their original curves. However, we must say that this interchange is rather a mathematical procedure than a physical one—it is assumed that the charge  $q$  first travels on the curve  $C$  and after on the curve  $\mathcal{C}$  and that the current  $\mathbf{J}$  flows first on the curve  $\mathcal{C}$  and after on the curve  $C$ . This requires that the charge  $q$  and the current  $\mathbf{J}$  are free to travel different curves, i.e. that the charge  $q$  and the current  $\mathbf{J}$  are not rigidly “endowed” to their initial curves.

### 7 A-explanation vs B-explanation: local and nonlocal interpretations of the AB effect

Although throughout this paper we have not hesitated to interpret the AB phase as a consequence of the nonlocal action of the magnetic field on the moving charge, we must recognise that this is not the most popular interpretation of the AB phase which considers either an infinitely-long solenoid or an infinitely-long flux line. In connection with this point we must say that there has been a longstanding debate between those who argue that the AB effect is caused by the local action of the vector potential on the moving charge (the **A**-explanation), in which case this potential must be considered as a real physical quantity despite its gauge dependence, and those who argue that the AB effect is caused by the nonlocal action of the magnetic field on the moving charge (the **B**-explanation), in whose case a form of action-at-a-distance is physically feasible (a short review of this debate can be found in Eynck et al. [42]). This debate is currently unsettled. Many papers have addressed this debate [43–52] and proposed other models that intend to explain the AB effect [53–76]. This debate may be naturally translated to the case of the AB effect involving a vector potential outside a closed flux line and a magnetic field confined along this closed flux line. In this section we will address the **A**-explanation and the **B**-explanation of the AB phase in the context of the closed flux line.

**Fig. 7** The charged path  $\Gamma$  can be decomposed as the sum of closed paths  $C$  (red line) enclosing the closed flux line plus the open path  $\gamma$  (blue line) that remains after the charge takes another turn around the solenoid. Both paths  $\Gamma$  and  $\gamma$  start from  $\mathcal{O}$  to the variable point  $x$ . The path  $\Gamma$  encloses the closed flux line while the path  $\gamma$  does not



We have seen that the AB phase can be expressed as  $\delta = [q/(\hbar c)] \oint_C \mathbf{A} \cdot d\mathbf{x}$ . If we apply the Stokes theorem in the considered non-simply connected region:  $\oint_{C > \partial S_e} \mathbf{A} \cdot d\mathbf{x} = \int_{S_e} \mathbf{B} \cdot d\mathbf{S}$  then the AB phase can be expressed as  $\delta = [q/(\hbar c)] \int_{S_e} \mathbf{B} \cdot d\mathbf{S}$ . Combining both expressions for the AB phase we have

$$\frac{q}{\hbar c} \oint_{C > \partial S_e} \mathbf{A} \cdot d\mathbf{x} = \delta = \frac{q}{\hbar c} \int_{S_e} \mathbf{B} \cdot d\mathbf{S}. \tag{86}$$

The **A**-explanation is supported by the first equality in Eq. (86). This states that the vector potential locally acts through its circulation on the charged particle originating the AB phase  $\delta$  and therefore it makes sense to claim that the vector potential locally influences the phase of the wave function of the charged particle. Argued differently, the AB phase must be physically originated by the local action of an electromagnetic quantity in the region outside the closed flux line and since the only electromagnetic quantity defined in each point of that region is the vector potential  $\mathbf{A}$  then this potential should produce the AB phase  $\delta$ . In short: the vector potential  $\mathbf{A}$  exists in each point of the trajectory of the charge  $q$  and therefore  $\mathbf{A}$  locally acts on  $q$  producing  $\delta$ . On the other hand, the **B**-explanation is supported by the second equality in Eq. (86), which states that the magnetic field nonlocally acts through its flux on the charged particle originating the AB phase  $\delta$ . In short:  $\mathbf{B}$  exists along the closed flux line and not in each point of the trajectory of the charge  $q$  outside this closed flux line and therefore  $\mathbf{B}$  nonlocally acts on  $q$  producing  $\delta$ .

In this paper we have supported the **B**-explanation and rejected the **A**-explanation. Let us point out three arguments that strengthen our support for the former explanation:

(I) The vector potential  $\mathbf{A}$  is gauge-dependent and therefore has no physical meaning.

(II) The Stokes theorem applied in the non-simply connected region discussed in this paper (see Eq. (41)) implies the relation

$$\frac{q}{\hbar c} \oint_{C_k > \partial S_e} \mathbf{A} \cdot d\mathbf{x} = \dots = \frac{q}{\hbar c} \oint_{C_2 > \partial S_e} \mathbf{A} \cdot d\mathbf{x} = \frac{q}{\hbar c} \oint_{C_1 > \partial S_e} \mathbf{A} \cdot d\mathbf{x} = \delta = \frac{q}{\hbar c} \int_{S_e} \mathbf{B} \cdot d\mathbf{S}, \tag{87}$$

where the different charge paths  $C_1 > \partial S_e, C_2 > \partial S_e, \dots, C_k > \partial S_e$  possessing all of them the same linking number are homotopically equivalent. If we consider the equalities on the left-hand side of  $\delta$  in Eq. (86) then there is a manifest ambiguity because we cannot distinguish if this phase is connected with the circulation of the vector potential  $\mathbf{A}$  along  $C_1 > \partial S_e$  or  $C_2 > \partial S_e$ , or along  $C_k > \partial S_e$ . These circulations are spatially delocalised with respect to the closed flux line (they are not functions of point) and therefore we cannot know which of the charge paths is locally connected with the circulation of the vector potential. In short: the vector potential  $\mathbf{A}$  is ambiguous due to its gauge-dependence and its circulation  $\oint_C \mathbf{A} \cdot d\mathbf{x}$  is gauge invariant but it is ambiguous due to its spatial delocalisation (indistinguishability of the curve  $C$ ). In consequence the **A**-explanation does not hold because the alleged local action of the vector potential is ambiguous. But if we consider the last equality in Eq. (86) then we conclude that the AB phase evaluated outside the closed line of singularity  $\mathcal{C}$  is unambiguously connected with the confined magnetic flux along this closed line of singularity. Since the charge and the closed flux line lie in different spatial regions this magnetic flux nonlocally acts on the charge producing the AB phase. Thus, the **B**-explanation holds. Let us remark that to understand the AB effect, we should be clear about the difference between applying the Stokes theorem in a simply connected region than in a non-simply connected region. This is the key to see why the AB phase is a nonlocal phase.

(III) Equation (45) implies the relation

$$\frac{q}{\hbar c} \int_{\Gamma} \mathbf{A} \cdot d\mathbf{x}' - \frac{q}{\hbar c} \int_{\gamma} \mathbf{A} \cdot d\mathbf{x}' = \frac{q}{\hbar c} \oint_C \mathbf{A} \cdot d\mathbf{x}', \tag{88}$$

which can be written as

$$\frac{q}{\hbar c} \int_{\mathcal{O}[\Gamma]}^x \mathbf{A} \cdot d\mathbf{x}' + \frac{q}{\hbar c} \int_x^{\mathcal{O}[\gamma]} \mathbf{A} \cdot d\mathbf{x}' = \frac{q}{\hbar c} \oint_C \mathbf{A} \cdot d\mathbf{x}', \tag{89}$$

where  $[\Gamma]$  and  $[\gamma]$  denote different paths going from the point  $\mathcal{O}$  to the point  $x$  (see Fig. 7).

Note that

$$\oint_C \mathbf{A} \cdot d\mathbf{x}' = \int_{\mathcal{O}[\Gamma]}^{\mathcal{O}[\gamma]} \mathbf{A} \cdot d\mathbf{x}' \neq 0, \tag{90}$$

because the function  $\Omega$  in the potential  $A = \Phi \nabla \Omega / (4\pi)$  is a multi-valued function. Moreover, since  $x$  is a variable point it follows that the quantities on the left-hand side of Eq. (89) are local functions which may be defined as

$$\frac{q}{\hbar c} \int_{\mathcal{O}[\Gamma]}^x A \cdot dx' = \delta_\Gamma(x), \quad \frac{q}{\hbar c} \int_x^{\mathcal{O}[\gamma]} A \cdot dx' = \delta_\gamma(x). \tag{91}$$

Considering these definitions of the functions  $\delta_\Gamma(x)$  and  $\delta_\gamma(x)$ , it follows that the AB phase  $\delta = [q/(\hbar c)] \oint_C A \cdot dx'$  in Eq. (89) can be expressed as

$$\delta_\Gamma(x) + \delta_\gamma(x) = \delta, \tag{92}$$

and thus one would be tempted to say that the AB phase is locally generated, a statement that supports the **A**-explanation of the AB effect. Nevertheless, the fundamental problem with this argument is that both functions  $\delta_\Gamma(x)$  and  $\delta_\gamma(x)$  are not gauge invariant. In fact, by applying the gauge transformation  $A' = A + \nabla \Lambda$  to the functions in Eq. (91) we have

$$\frac{q}{\hbar c} \int_{\mathcal{O}[\Gamma]}^x A' \cdot dx' = \frac{q}{\hbar c} \int_{\mathcal{O}[\Gamma]}^x A \cdot dx' + \frac{q}{\hbar c} [\Lambda(x) - \Lambda(\mathcal{O})], \tag{93}$$

$$\frac{q}{\hbar c} \int_x^{\mathcal{O}[\gamma]} A' \cdot dx' = \frac{q}{\hbar c} \int_x^{\mathcal{O}[\gamma]} A \cdot dx' + \frac{q}{\hbar c} [\Lambda(\mathcal{O}) - \Lambda(x)]. \tag{94}$$

These equations show that both  $\delta_\Gamma(x)$  and  $\delta_\gamma(x)$  are not gauge invariant and therefore they cannot be considered the physical causes of the AB phase, i.e. these functions are not measurable quantities and consequently they are devoid of any physical meaning. Only the sum of these local functions is physically meaningful because it is gauge invariant—this sum is no longer a function of point but a constant quantity implying the vanishing of the locality of the AB phase. Accordingly, on the basis of the arguments (I)-(III) we admit the **B**-explanation and reject the **A**-explanation.

As already pointed out, the nonlocal interpretation of the AB effect is much less popular than its local interpretation. One of the supporters of the former interpretation is Aharonov who in 1983 claimed [77]: “*In this talk we will review the AB effect, which ... provides a particularly clear example of nonlocal phenomena... One hopes that the arguments in the preceding section have convinced the reader that the A-B effect is indeed nonlocal*”. Here Aharonov clearly adopted the nonlocal explanation of the AB effect, and in doing so he disregarded the local explanation he had initially held together with Bohm [1]. More recently, Aharonov et al. have pointed out [78]: “*The [AB] phase is topological because it is determined by the number of windings the charge carries out around the solenoid, and is independent of the details of the trajectory. The phase is also nonlocal: while the magnetic flux in the solenoid clearly affects the resulting interference pattern, it has no local observable consequences along any point on the trajectory*”. Aharonov and Rohrlich have also pointed out [2]: “*Thus, instead of concluding that A and V are physical variables in quantum mechanics, we state a conclusion ...: Only E and B are physical quantities, but they act nonlocally—a magnetic field here has physical effects on electrons there, and so on. Such action at a distance by a field is completely nonclassical*”.

The idea that the AB effect is purely quantum has been emphasised by many authors and experiments confirming such an idea have been carried out [79–81]. Considering that the AB effect is a quantum-mechanical effect one could come to the conclusion that the nonlocality of the magnetic field in this effect is also inherently quantum-mechanical, as pointed out by Aharonov and Rohrlich [2]. However, we have emphasised here that the nonlocality of the AB effect is of topological nature and therefore one could reasonably expect that this nonlocality could arise in other branches of physics—after all topology is a branch of mathematics independent of quantum mechanics. With regard to this point, it is pertinent to say that in two recent papers [38,39] we have discussed three classical electromagnetic configurations defined in non-simply connected regions: (i) an electric charge encircling an infinitely-long solenoid, (ii) a magnetic charge encircling an infinitely-long electric solenoid, and (iii) a dyon encircling an infinitely-long dual solenoid enclosing magnetic and electric fluxes. We have shown that the electromagnetic angular momenta arising from these configurations describe nonlocal interactions between the encircling charges (electric, magnetic, or dyon) outside the respective solenoids and the corresponding fluxes (magnetic, electric, or dual) confined inside these solenoids. In particular, we have argued [38] that the electromagnetic angular momentum of the configuration (i) may be considered as the classical counterpart of the AB effect. It is pertinent to note here that a number of authors have proposed classical analogues of the AB effect [78,82–90]. We think that the AB effect is a purely quantum effect but the nonlocality of this effect also appears in classical effects. Anyway, the fact that this nonlocal feature arises in both classical and quantum physics illustrates the power of topology in physics. As already noted: *topology dictates nonlocality!*

### 8 A gauge that eliminates the vector potential in all space except in a finite region

In the previous section we have emphasised three formal arguments against the local **A**-explanation of the AB effect. In this section we will present a fourth formal argument, which may be even stronger and convincing than the first three.

Consider the vector potential of the closed flux line  $\mathbf{A} = \Phi \nabla \Omega / (4\pi) = \Phi \nabla \Omega_0 / (4\pi) + \Phi \delta_S$ . This potential satisfies  $\nabla \times \mathbf{A} = \Phi \delta_C = \mathbf{B}$ . Let us write the expression of this potential as

$$\mathbf{A} + \nabla \Lambda = \Phi \delta_S, \tag{95}$$

where  $\Lambda$  is a scalar function defined as

$$\Lambda = -\frac{\Phi}{4\pi} \Omega_0. \tag{96}$$

A crucial observation is that the function  $\Lambda$  is single-valued and this means that this function satisfies the Schwarz integrability condition in all space  $(\partial^i \partial^j - \partial^j \partial^i) \Lambda = 0$ , a condition that follows from the single-valuedness of the solid angle function  $\Omega_0$  which satisfies  $(\partial^i \partial^j - \partial^j \partial^i) \Omega_0 = 0$ . This is explicitly demonstrated Appendix C. Considering Eq. (95) we may interpret the function  $\Lambda$  as the single-valued gauge function of the non-singular gauge transformation

$$\mathbf{A}' = \mathbf{A} + \nabla \Lambda, \tag{97}$$

that transforms the Coulomb-gauge potential  $\mathbf{A} = \Phi \nabla \Omega / (4\pi)$  into the potential

$$\mathbf{A}' = \Phi \delta_S. \tag{98}$$

Remarkably, the gauge function  $\Lambda$  in Eq. (96) has the unusual property of changing the domain of definition of the potential  $\mathbf{A}$ , i.e. while the potential  $\mathbf{A}$  is defined in all space outside the closed flux line, the potential  $\mathbf{A}'$  vanishes in all space except on the surface  $S$  surrounded by the closed flux line. To see this peculiar behaviour more clearly, we can use Eq. (10) in Eq. (98) to obtain

$$\mathbf{A}' = \Phi \int_S \delta(\mathbf{x} - \mathbf{x}') dS', \tag{99}$$

which shows that for any point  $\mathbf{x}$  not on the surface  $S$  the potential  $\mathbf{A}'$  vanishes, i.e.  $\mathbf{A}'(\mathbf{x} \notin S) = 0$ . Evidently,  $\nabla \times \mathbf{A}' = \nabla \times \mathbf{A} = \mathbf{B}$  because  $\nabla \times \nabla \Lambda = 0$  ( $\Lambda$  is single-valued). While the vector potential  $\mathbf{A}$  satisfies the Coulomb gauge condition  $\nabla \cdot \mathbf{A} = 0$ , the vector potential  $\mathbf{A}'$  satisfies the gauge condition

$$\nabla \cdot \mathbf{A}' = \Phi \nabla \cdot \delta_S, \tag{100}$$

or equivalently

$$\nabla \cdot \mathbf{A}' = -\Phi \int_S \nabla' \delta(\mathbf{x} - \mathbf{x}') \cdot dS', \tag{101}$$

where we have taken the divergence to Eq. (99) and used  $\nabla \cdot [\delta(\mathbf{x} - \mathbf{x}') dS'] = \nabla \delta(\mathbf{x} - \mathbf{x}') \cdot dS'$  and  $\nabla \delta(\mathbf{x} - \mathbf{x}') = -\nabla' \delta(\mathbf{x} - \mathbf{x}')$ . Since  $\mathbf{A}$  and  $\mathbf{A}'$  are equivalent potentials it follows that the AB phase calculated with the transformed potential  $\mathbf{A}'$  is the same as that calculated with the original potential  $\mathbf{A}$ . This may be seen by using Eq. (98):

$$\frac{q}{\hbar c} \oint_C \mathbf{A}' \cdot d\mathbf{x} = \frac{q\Phi}{\hbar c} \oint_C \delta_S \cdot d\mathbf{x} = l \frac{q\Phi}{\hbar c}, \tag{102}$$

where we have used the definition of the linking number given in Eq. (27). Accordingly, the equivalence  $[q/(\hbar c)] \oint_C \mathbf{A}' \cdot d\mathbf{x} = \delta = [q/(\hbar c)] \oint_C \mathbf{A} \cdot d\mathbf{x}$  holds and thereby those gauge-invariant quantities discussed in this paper involving the Coulomb-gauge potential  $\mathbf{A}$ , for example the two-slit interference shift  $\Delta_{AB} = Lxq\Phi/d$ , are identical if we use the potential  $\mathbf{A}'$  in the new gauge specified by Eq. (101). We should also note that Eq. (102) verifies our earlier assertion made in Eq. (51) that the AB phase in a closed flux line is not continuously accumulated along the charge path  $C$  but discretely picked up as the charge crosses the closed flux line.

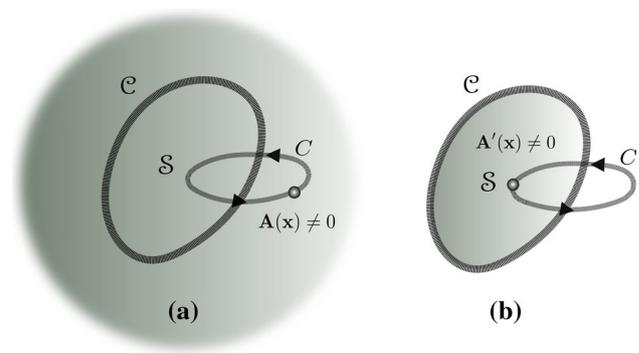
Interestingly, Eq. (98) can be expressed as  $[\Phi/(4\pi)] \nabla(\Omega - \Omega_0) = \mathbf{A}'$  which shows that the vector potential in the introduced gauge may be written as

$$\mathbf{A}' = \frac{\Phi}{4\pi} \nabla \Theta, \tag{103}$$

where  $\Theta = \Omega - \Omega_0$  is the function formed by the difference between the multi-valued representation of the solid angle  $\Omega$  and the single-valued representation of the solid angle  $\Omega_0$ . Since  $\Omega$  is a multi-valued function it follows that  $\Theta$  is a multi-valued function and therefore  $(\partial^i \partial^j - \partial^j \partial^i) \Theta \neq 0$ .

According to the **A**-explanation of the AB effect, the vector potential exists in every point of the trajectory of the charged particle and then quantum mechanics may be invoked to argue that this potential, through its gauge invariant circulation, influences the wave function of this particle producing the AB phase (see Fig. 8a). However, we have introduced here a non-singular gauge in which the vector potential [Eq. (99)] vanishes in all space except on the surface surrounded by the closed flux line implying that every time the charge takes a turn around the closed flux line the vector potential is zero along the trajectory of the charge except on a single point of this trajectory—the crossing point where the trajectory of the charge intersects the surface surrounded by the closed

**Fig. 8** **a** The potential  $A(x)$  is non-vanishing in every point  $x$  along the path  $C$  of the charged particle. **b** The potential  $A'(x)$  is zero in every point  $x$  along the path  $C$  of the charged particle except at a single point localised on the surface  $S$



flux line (see Fig. 8b). In short: the vector potential exists only at a point of the charge path. The widely spread idea that the vector potential “acts” through its circulation on the charge at every point of its path then collapses.

At the risk of being too reiterative, let us insist in our fourth argument against the **A**-explanation. As is well-known, the **A**-explanation of the AB effect tells us that the original vector potential **A** exists at every point of the trajectory of the charge. But we have showed here that the transformed vector potential  $A'$  in Eq. (99) is zero along this trajectory except on one of its points. Since we can always transform **A** into  $A'$  via a non-singular gauge transformation [Eq. (97)] then, after this transformation, the domain of definition of the vector potential has been modified. This behaviour of the vector potential does not correspond to any physical local quantity since the domain of definition of a physical quantity (in a reference frame) should remain fixed in order for the local interaction of this quantity to make sense. In this regard, it is interesting to note that Paiva et al. [91] have recently pointed out a similar argument in favour of a nonlocal interpretation of the AB effect (they considered an infinitely-long flux line): “...because there always exists a gauge in which vector potential vanishes in an arbitrary region that does not completely enclose the solenoid, a priori, the AB effect cannot be seen as the result of the local interaction between the charge and the vector potential”. However, these authors did not provide any explicit gauge that proves their statement. Here we have explicitly shown that such a gauge exists in the AB effect in a closed flux line (the corresponding gauge condition is given in Eq. (101)). In a few words: the original potential **A** exists at every point on the charge path but the transformed potential  $A'$  exists at a single point of the charge path! The alleged local interaction of this vector potential with the particle could be anything but a physical interaction.

### 9 Singular and non-singular gauge transformations in the AB effect

The term “singular gauge transformations” deserves a clarification. In proper physical jargon, gauge transformations are transformations of potentials that leave invariant their associated fields. End of story. But physicists are sloppy in the use of language and some of them have wanted to extend the story by talking about “singular” gauge transformations. In fact, the term singular gauge transformations is an oxymoron because these transformations are not true gauge transformations—they modify their corresponding fields! If we admit such singular gauge transformations then we could appear or disappear magnetic fields depending on the chosen gauge. In Kleinert’s words [33]: “...Obviously, this terminology [singular gauge transformation] is misleading and must be rejected. After all, if we were to allow for such “singular” (i.e. nonintegrable) transformations...we could reach an arbitrary field  $F_{\mu\nu}$  starting from  $F_{\mu\nu} \equiv 0$ , and the physics would certainly not be invariant under this”. The failure to specify the difference that exists between applying either a singular or non-singular gauge transformations in the context of the AB effect may lead to the misleading conclusion that the vector potential in the AB effect can be eliminated via a gauge transformation.

Consider the following statement: unlike a non-singular gauge, a singular gauge modifies the magnetic field and therefore it cannot be considered a symmetry of this field. This assertion should be included in textbooks and then one could understand that the possible disappearance of the vector potential in a given configuration via a singular gauge entails a modification of the magnetic field—which totally breaks the spirit of gauge invariance. The puzzling question arises: how to understand a supposed gauge transformation that changes the magnetic field?

In order to see how a singular gauge was introduced to interpret the AB effect, let us do some history. In 1979 Bocchieri and Loinger [92] surprised theoretical physicists by claiming that they had found a gauge in which the vector potential is zero outside the infinitely-long solenoid and as a consequence the AB effect disappeared in this gauge! These authors then concluded that the AB effect [92]: “...has a purely mathematical origin”, i.e. this effect was a sort of mathematical artefact devoid of physical meaning. The argument of Bocchieri and Loinger can be reconstructed as follows.

Consider first the vector potential of an infinitely-long solenoid of radius  $R$ . In the Coulomb gauge this vector potential can be written in cylindrical coordinates as [38,93]

$$A = \frac{\Phi}{2\pi} \left[ \frac{\Theta(\rho - R)}{\rho} + \frac{\rho \Theta(R - \rho)}{R^2} \right] \hat{\phi}, \tag{104}$$

where  $\Phi = \pi R^2 B$  is the flux through the solenoid with  $B$  being the magnitude of the magnetic field inside the solenoid and  $\Theta$  is the Heaviside step function. The curl of Eq. (104) gives [38,93]

$$\nabla \times \mathbf{A} = \frac{\Phi}{\pi R^2} \Theta(R - \rho) \hat{z}, \tag{105}$$

whose right-hand side identifies with the magnetic field confined in the solenoid [38]:  $\mathbf{B} = \Phi \Theta(R - \rho) \hat{z} / (\pi R^2)$ . Following Bocchieri and Loinger [92], let us apply a “presumable” gauge transformation to transform the potential  $\mathbf{A}$  into the potential  $\mathbf{A}'$  by adding to it the gradient of the function  $\chi$

$$\mathbf{A}' = \frac{\Phi}{2\pi} \left[ \frac{\Theta(\rho - R)}{\rho} + \frac{\rho \Theta(R - \rho)}{R^2} \right] \hat{\phi} + \nabla \chi, \tag{106}$$

where  $\chi = -\Phi \phi / (2\pi)$  with  $\phi$  being the azimuthal angle.

By invoking the usual result that the curl of a gradient identically vanishes, it follows that

$$\nabla \times \mathbf{A}' = \frac{\Phi}{\pi R^2} \Theta(R - \rho) \hat{z}, \tag{107}$$

and then  $\mathbf{A}$  and  $\mathbf{A}'$  are equivalent potentials in the sense that the curl of both yields the same magnetic field.

Now  $\nabla \chi = -\Phi \hat{\phi} / (2\pi \rho)$  and therefore Eq. (106) can be written as

$$\mathbf{A}' = \frac{\Phi}{2\pi \rho} \left[ \Theta(\rho - R) - 1 \right] \hat{\phi} + \frac{\Phi}{2\pi} \frac{\rho \Theta(R - \rho)}{R^2} \hat{\phi}. \tag{108}$$

Outside the solenoid [ $\rho > R$ ,  $\Theta(\rho - R) = 1$  and  $\Theta(R - \rho) = 0$ ] the potential vanishes:  $A'_{\text{out}} = 0$  while inside the solenoid [ $R > \rho$ ,  $\Theta(R - \rho) = 1$  and  $\Theta(\rho - R) = 0$ ] the potential takes the form  $A'_{\text{in}} = [\rho \Phi / (2\pi R^2) - \Phi / (2\pi \rho)] \hat{\phi}$ . Since the charge path  $C$  is outside the solenoid then it follows that  $\delta = [q / \hbar c] \oint_C A'_{\text{out}} \cdot d\mathbf{x} = 0$  and therefore the AB effect vanishes. The conclusion arises: the AB phase exists for the potential  $\mathbf{A}$  but not for the equivalent potential  $\mathbf{A}'$ . This means that the AB phase is gauge dependent and therefore the AB effect seems to be a mathematical construct devoid of physical meaning. As Bocchieri and Loinger claimed [92]: “... Obviously, in this gauge [ $A'_{\text{out}} = 0$ ] there is no Aharonov–Bohm effect. This way of reasoning shows that the effect is gauge-dependent; this seems astonishing, at first, as in [ $A_{\text{out}} = \Phi \hat{\phi} / (2\pi \rho)$ ] only  $\Phi$  appears, which is a gauge-invariant concept”.

The argument of Bocchieri and Loinger seems to be correct except for one thing: the function  $\chi = -\Phi \phi / (2\pi)$  is a multi-valued function and its gradient  $\nabla \chi = -\Phi \hat{\phi} / (2\pi \rho)$  is a singular function. As already noted, if  $\Lambda$  is a single-valued function then it satisfies the Schwarz integrability condition:  $(\partial^i \partial^j - \partial^j \partial^i) \Lambda = 0$ . The single-valuedness of  $\Lambda$  implies  $\nabla \times \nabla \Lambda = 0$ . But if  $\Lambda$  is a multi-valued function then it violates the Schwarz condition:  $(\partial^i \partial^j - \partial^j \partial^i) \Lambda \neq 0$  and therefore  $\nabla \times \nabla \Lambda \neq 0$ . The function considered by Bocchieri and Loinger:  $\chi = -\Phi \phi / (2\pi)$  is a multi-valued function that satisfies

$$\nabla \times \nabla \chi = -\frac{\Phi}{2\pi} \frac{\delta(\rho)}{\rho} \hat{z}, \tag{109}$$

which can be verified as follows. From the Stokes theorem we can write  $\oint_C \nabla \chi \cdot d\mathbf{x} = \int_S \nabla \times \nabla \chi \cdot d\mathbf{S}$ , where  $C$  is the boundary of  $S$  and it is assumed  $C$  encircles the solenoid. Using  $\chi = -\Phi \phi / (2\pi)$  we obtain  $\oint_C \nabla \chi \cdot d\mathbf{x} = [-\Phi / (2\pi)] \oint_C \nabla \phi \cdot d\mathbf{x}$ . Inserting  $\nabla \phi = \hat{\phi} / \rho$  and  $d\mathbf{x} = d\rho \hat{\rho} + \rho d\phi \hat{\phi} + dz \hat{z}$  it follows  $\oint_C \nabla \phi \cdot d\mathbf{x} = \oint_C d\phi$  and since  $\oint_C d\phi = 2\pi n$  where  $n$  is the winding number of the path  $C$ , it follows that  $\oint_C \nabla \chi \cdot d\mathbf{x} = -n\Phi$  which together with the Stokes theorem imply the relation  $\int_S \nabla \times \nabla \chi \cdot d\mathbf{S} = -n\Phi$ . This relation is satisfied by Eq. (109) which verifies the validity of this equation. Therefore the correct expression for the curl of the potential  $\mathbf{A}'$  in Eq. (108) reads

$$\nabla \times \mathbf{A}' = \frac{\Phi}{\pi R^2} \Theta(R - \rho) \hat{z} - \frac{\Phi}{2\pi} \frac{\delta(\rho)}{\rho} \hat{z}. \tag{110}$$

The first term on the right-hand side identifies with the magnetic field confined in the solenoid  $\mathbf{B} = \Phi \Theta(R - \rho) \hat{z} / (\pi R^2)$  while the second term identifies with the negative of the singular magnetic field  $\mathbf{B}_{\text{string}} = \Phi \delta(\rho) \hat{z} / (2\pi \rho)$  due to an infinitely-long flux line (or a magnetised string) localised along the  $z$ -axis. Expressed more compactly Eq. (110) reads

$$\mathbf{B}' = \mathbf{B} - \mathbf{B}_{\text{string}}. \tag{111}$$

It becomes now evident that the argument of Bocchieri and Loinger for the elimination of the AB effect is inconsistent because it breaks the gauge invariance of the magnetic field of the solenoid:  $\mathbf{B} = \nabla \times \mathbf{A} \neq \mathbf{B}' = \nabla \times \mathbf{A}'$ —several authors [93–96] have noted this inconsistency of the supposed gauge transformation proposed by Bocchieri and Loinger.

The gauge transformation applied by Bocchieri and Loinger is an example of the so-called singular gauge transformations. The wrong idea that singular gauge transformations do not modify the magnetic field has been something recurrent in the literature. For example, Wilczek [97] followed an argument similar to that of Bocchieri and Loinger to eliminate the vector potential outside an infinitely-long solenoid by means of a non-singular gauge transformation in order to obtain the eigenvalues of the  $z$ -component

of the angular momentum of the system formed by a charge encircling this solenoid. The inconsistency of Wilczek’s approach was pointed out by Kobe [98]. Nambu [99] also used a singular gauge transformation to study the AB effect. Other authors have also used singular gauge transformations to study the quantum mechanics of a charge encircling an infinitely-long solenoid (see, for example, Ref. [100]).

Similar arguments to those of Bocchieri and Loinger can be applied to the vector potential of the closed flux line and its associated AB effect. In the previous section we have applied the non-singular gauge transformation  $A' = A + \nabla \Lambda$  with  $\Lambda = -\Phi \Omega_0 / (4\pi)$  to transform the Coulomb-gauge potential  $A = \Phi \nabla \Omega / (4\pi)$  into the potential  $A' = \Phi \delta_S$ . Suppose now that instead of considering the gauge function  $\Lambda = -\Phi \Omega_0 / (4\pi)$  we consider the “gauge” function  $\chi = -\Phi \Omega / (4\pi)$ . It follows that

$$A' = \frac{\Phi}{4\pi} \nabla \Omega + \nabla \chi = \frac{\Phi}{4\pi} \nabla \Omega - \frac{\Phi}{4\pi} \nabla \Omega = 0. \tag{112}$$

Apparently, we have found a gauge in which the vector potential is zero and therefore the AB effect in the closed flux line is seen to disappear! Nevertheless, we had to pay a very high price: while the curl of the original potential  $A$  in Eq. (16) yields the non-vanishing magnetic field  $B = \Phi \oint_C \delta(x - x') dx'$ , the curl of the transformed potential  $A' = 0$  yields a vanishing magnetic field  $B' = 0$ . The transformation we have applied here is a singular gauge transformation which modifies the magnetic field—the function  $\chi = -\Phi \Omega / (4\pi)$  is a multi-valued function.

An alternative way to see the difference between non-singular and singular gauge transformations may be given using index notation in which the magnetic field is represented by the axial component  $F^{ij} = -\varepsilon^{ijk} B_k$  (where  $B_k$  represents the components of the field  $B$ ) of the electromagnetic field tensor  $F^{\mu\nu}$  (for notation and conventions in the Minkowski spacetime see Ref. [101]). In our case this axial component [the magnetic field in Eq. (4)] reads

$$F^{ij} = -\Phi \varepsilon^{ijk} (\delta_C)_k, \tag{113}$$

From  $F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu$  and  $\partial^\mu = (\partial^0, -\partial^i)$  it follows that the tensor  $F^{ij}$  can be expressed in terms of the components  $A^j$  of the vector potential  $A$  as

$$F^{ij} = \partial^j A^i - \partial^i A^j, \tag{114}$$

and in terms of the primed components  $A'^j$  of the primed vector potential  $A'$  by  $F'^{ij} = \partial^j A'^i - \partial^i A'^j$ . Let us apply the transformation  $A^j \rightarrow A'^j = A^j - \partial^j \Lambda$ , where  $\Lambda$  is an arbitrary function of space (at this stage this transformation is not necessarily a gauge transformation), to Eq. (114). We obtain

$$F'^{ij} = F^{ij} + (\partial^i \partial^j - \partial^j \partial^i) \Lambda. \tag{115}$$

If  $\Lambda$  is a single-valued function then it satisfies the Schwarz condition  $(\partial^i \partial^j - \partial^j \partial^i) \Lambda = 0$ . In this case the corresponding potentials are gauge potentials and their associated transformation is a gauge transformation which leaves invariant the axial component of the electromagnetic field:  $F'^{ij} = F^{ij}$ . But if  $\Lambda$  is a multi-valued function then it does not satisfy the Schwarz condition  $(\partial^i \partial^j - \partial^j \partial^i) \Lambda \neq 0$ . In this case the corresponding potentials are not gauge potentials and its associated transformation is not a gauge transformation because it does not leave invariant the axial component of the electromagnetic field:  $F'^{ij} \neq F^{ij}$ . In our particular case, we have the specific multi-valued gauge function  $\Lambda = -\Phi \Omega / (4\pi)$  which satisfies  $(\partial^i \partial^j - \partial^j \partial^i) \Lambda = -\Phi \varepsilon^{ijk} (\delta_C)_k$ . When this result and Eq. (113) are used in Eq. (115) we conclude that  $F'^{ij} = 0$ . In this case we have a singular gauge transformation that transforms the non-vanishing field  $F^{ij} = -\Phi \varepsilon^{ijk} (\delta_C)_k$  into the vanishing field  $F'^{ij} = 0$ ! As Sidney Coleman said [102]: “...singular gauge transformations make people uneasy”.

### 10 Concluding remarks

Millikan once said [103]: “Science walks forward on two feet, namely theory and experiment, ... Sometimes it is one foot that is put forward first, sometimes the other, but continuous progress is only made by the use of both -by theorising and then testing, or by finding new relations in the process of experimenting and then bringing the theoretical foot up and pushing it on beyond, and so on in unending alterations”. Millikan’s quote applies very well to the case of the AB effect:

#### The first step was with the theoretical foot:

The story began in 1959 when Aharonov and Bohm [1] theoretically predicted the now known as AB effect in which a relative phase shift proportional to the magnetic flux of an “idealised” infinitely-long solenoid can be observed as a displacement of interference fringes, even when the interfering charged particles pass through a spatial region where there is no magnetic field but there is a vector potential. Two problems did not take long to appear:

- (i) The effect was predicted considering an idealised infinitely-long solenoid which was evidently unavailable.
- (ii) The effect admitted two different and excluding interpretations: it was either attributed to the local action of the vector potential (**A**-explanation) or to the nonlocal action of the magnetic field (**B**-explanation). The solution to the problem (i) would be to replace the non-available infinitely-long solenoid with a long but finite solenoid which was attainable. However, the finite long

solenoid does not confine completely the magnetic field and therefore the use of this solenoid does not allow to theoretically predict the AB effect. No conclusive formal argument was proposed to find the correct explanation of the AB effect and thus to solve the problem (ii).

### The second step was with the experimental foot:

In the 1960's several experiments that tried to test the AB effect were reported [5–8] in which the corresponding displacement of interference fringes were observed using finite magnetic devices like magnetised whiskers and long solenoids. But these displacements were questioned [9, 10] due to the unavoidable magnetic field leakage rather than to the AB effect. However, in the 1980's Tonomura and collaborators [11–13] decisively confirmed the AB effect in a series of experiments using a toroidal configuration. In a first instance [11] they used a squared micro-sized toroidal magnet whose field leakage was sufficiently small to verify the AB effect with good accuracy. In a second instance [12, 13], they used a circular micro-sized toroidal magnet covered with a superconducting layer which avoided the field leakage due to the Meissner effect. Thus, the experimental foot suggested a way to solve the problem (i) by detecting the AB effect using a tiny magnet whose shape was roughly that of a toroid. However, two questions arise:

- (iii) Is there a well-established theoretical treatment of the AB effect using a toroidal configuration?
- (iv) How to explain that the AB effect arises in two configurations of different geometry? Expectably, the theoretical foot could solve the problems (i)-(iv).

### The third step is with the theoretical foot:

We have addressed here the problem (i) by showing that the AB effect is theoretically predicted using a less-idealised closed flux line instead of using a highly-idealised infinitely long solenoid. With regard to the problem (ii) we have argued that the **B**-explanation should be seen as the correct explanation of the AB effect. We have questioned the **A**-explanation by introducing a non-singular gauge in which the vector potential vanishes in all space except on the surface surrounded by the closed flux line and therefore, as the charge encircles the closed flux line, the vector potential is zero along the trajectory of the charge except on a point of this trajectory, which questions its alleged physical significance. We have emphasised the difference in applying singular and non-singular gauge transformations in the AB effect and argue that only the latter are consistent with the gauge invariance of the magnetic field. With regard to the question (iii) we should first say that there have been some theoretical treatments on the AB effect in a toroidal solenoid [14, 15, 17, 104–111]. However these treatments are too cumbersome to the extent that most authors (if not all) of standard textbooks on quantum mechanics (see, for example, Weinberg [112] and Sakurai [113]) discussing theoretical aspects of the AB effect still prefer the use of an infinitely-long solenoid instead of a toroidal solenoid even though these authors justify the experimental validity of the AB effect with the experiments of Tonomura et al. [11–13] who clearly used a toroidal configuration. Here we have addressed the question (iii) by showing that if a thin toroidal solenoid is modelled by a closed flux line of arbitrary shape and size then the theoretical treatment of the AB effect is exact and considerably accessible to the extent that it may be included in textbooks on quantum mechanics. We have demonstrate that the AB phase in a closed flux line is determined by a linking number and exhibits the same form as the AB phase in an infinitely-long flux line which is determined by a winding number. We have showed that the shift in the interference pattern associated to the AB effect in a closed flux line is the same as that associated to an infinitely-long flux line. We have addressed the question (iv) by emphasising that the AB effect arises in different geometries because it is of topological nature. We have stressed the topological character of the AB phase in a closed flux line by introducing four topological invariances of this phase: invariance under deformations of the charge path, invariance under deformations of the closed flux line, invariance under simultaneous deformations of the charge path and the closed flux line, and invariance under the interchange between the charge path and the closed flux line.

Finally, we would like to comment on four further and feasible applications of the closed-flux-line model discussed here. The first application deals with the scattering amplitude of charges outside the closed flux line. Just like the AB scattering in an infinitely-long flux line, in an infinitely-long solenoid, or in a toroidal solenoid, we expect that the scattering amplitude of charged particles outside a closed flux line should have a non-vanishing effect due to the confined magnetic flux. The second application deals with the energy levels of a charged particle encircling the closed flux line. Similarly to the energy levels of the system formed by a charge encircling an infinitely-long flux line or an infinitely-long solenoid, we expect the energy levels of the system formed by a charged particle encircling the closed flux line to have an explicit dependence on the magnetic flux. The third application deals with demonstrating that the AB phase in a closed flux line is an example of the Berry phase [114]. The fourth application deals with anyons or composite systems formed by a charge encircling a flux line which obey fractional statistics. Wilczek's [115] original anyon model is based on a charge encircling an infinitely-long solenoid. The anyon model was later extended to charges encircling a toroidal solenoid [116–119], i.e. toroidal anyons. The quantum mechanics of closed-flux line anyons would be an interesting theoretical possibility.

**Acknowledgements** I thank my father José A. Heras for the insightful and often enjoyable discussions we had about the AB effect. Perhaps the most entertaining part of our conversations was trying to understand why nonlocality seems to be such an unthinkable concept for many physicists. As an undergraduate student, I wrote this paper in my free time and with the interest of understanding the AB effect without any kind of prejudice.

**Data Availability Statement** No data sets were generated or analysed in this paper.

**Appendix A. Derivation of Eq. (4)**

Consider the potential of the closed flux line defined in Eq. (7)

$$\mathbf{A} = \frac{\Phi}{4\pi} \nabla \times \oint_{\mathcal{C}} \frac{d\mathbf{x}'}{|\mathbf{x} - \mathbf{x}'|}. \tag{A1}$$

The curl of this potential and the use of the identity  $\nabla \times (\nabla \times \mathbf{F}) = \nabla(\nabla \cdot \mathbf{F}) - \nabla^2 \mathbf{F}$  yield

$$\begin{aligned} \nabla \times \mathbf{A} &= \frac{\Phi}{4\pi} \nabla \times \left[ \nabla \times \left( \oint_{\mathcal{C}} \frac{d\mathbf{x}'}{|\mathbf{x} - \mathbf{x}'|} \right) \right] \\ &= \frac{\Phi}{4\pi} \left[ \nabla \oint_{\mathcal{C}} \nabla \cdot \left( \frac{d\mathbf{x}'}{|\mathbf{x} - \mathbf{x}'|} \right) - \oint_{\mathcal{C}} \nabla^2 \left( \frac{d\mathbf{x}'}{|\mathbf{x} - \mathbf{x}'|} \right) \right]. \end{aligned} \tag{A2}$$

Inserting  $\nabla \cdot (d\mathbf{x}'/|\mathbf{x} - \mathbf{x}'|) = -\nabla'(1/|\mathbf{x} - \mathbf{x}'|) \cdot d\mathbf{x}'$  and  $\nabla^2(1/|\mathbf{x} - \mathbf{x}'|) = -4\pi\delta(\mathbf{x} - \mathbf{x}')$  in Eq. (A2), we obtain

$$\nabla \times \mathbf{A} = \frac{\Phi}{4\pi} \left[ -\nabla \oint_{\mathcal{C}} \nabla' \left( \frac{1}{|\mathbf{x} - \mathbf{x}'|} \right) \cdot d\mathbf{x}' + 4\pi \oint_{\mathcal{C}} \delta(\mathbf{x} - \mathbf{x}') d\mathbf{x}' \right]. \tag{A3}$$

The first term on the right-hand side vanishes because  $\oint_{\mathcal{C}} \nabla'(1/|\mathbf{x} - \mathbf{x}'|) \cdot d\mathbf{x}' = 0$  on account of the gradient theorem and the fact that  $1/|\mathbf{x} - \mathbf{x}'|$  is a single-valued function of  $\mathbf{x}'$ . Thus, Eq. (A3) becomes

$$\nabla \times \mathbf{A} = \Phi \oint_{\mathcal{C}} \delta(\mathbf{x} - \mathbf{x}') d\mathbf{x}' = \mathbf{B}, \tag{A4}$$

which shows that the curl of Eq. (7) yields the magnetic field given in Eq. (4).

**Appendix B. Derivation of Eq. (8)**

Consider the potential of the closed flux line defined in Eq. (7)

$$\mathbf{A} = \frac{\Phi}{4\pi} \nabla \times \oint_{\mathcal{C}} \frac{d\mathbf{x}'}{|\mathbf{x} - \mathbf{x}'|}. \tag{B1}$$

Using the Stokes theorem in the closed line integral of this potential we obtain

$$\oint_{\mathcal{C}} \frac{d\mathbf{x}'}{|\mathbf{x} - \mathbf{x}'|} = \int_{\mathcal{S}} d\mathbf{S}' \times \nabla' \left( \frac{1}{|\mathbf{x} - \mathbf{x}'|} \right), \tag{B2}$$

where  $\mathcal{S}$  is the surface enclosed by the curve  $\mathcal{C}$ . Making use of the relations  $\nabla \times (d\mathbf{S}'/|\mathbf{x} - \mathbf{x}'|) = -d\mathbf{S}' \times \nabla(1/|\mathbf{x} - \mathbf{x}'|)$  and  $\nabla(1/|\mathbf{x} - \mathbf{x}'|) = -\nabla'(1/|\mathbf{x} - \mathbf{x}'|)$  in Eq. (B2), we obtain

$$\oint_{\mathcal{C}} \frac{d\mathbf{x}'}{|\mathbf{x} - \mathbf{x}'|} = \nabla \times \int_{\mathcal{S}} \frac{d\mathbf{S}'}{|\mathbf{x} - \mathbf{x}'|}, \tag{B3}$$

which allows us to write Eq. (B1) as

$$\mathbf{A} = \frac{\Phi}{4\pi} \left[ \nabla \times \left( \nabla \times \int_{\mathcal{S}} \frac{d\mathbf{S}'}{|\mathbf{x} - \mathbf{x}'|} \right) \right]. \tag{B4}$$

The use of the identity  $\nabla^2 \mathbf{F} = \nabla(\nabla \cdot \mathbf{F}) - \nabla \times (\nabla \times \mathbf{F})$  in Eq. (B4) gives

$$\mathbf{A} = \frac{\Phi}{4\pi} \left[ \nabla \int_{\mathcal{S}} \nabla \cdot \left( \frac{d\mathbf{S}'}{|\mathbf{x} - \mathbf{x}'|} \right) - \int_{\mathcal{S}} \nabla^2 \left( \frac{d\mathbf{S}'}{|\mathbf{x} - \mathbf{x}'|} \right) \right]. \tag{B5}$$

Considering the results  $\nabla \cdot (d\mathbf{S}'/|\mathbf{x} - \mathbf{x}'|) = \nabla(1/|\mathbf{x} - \mathbf{x}'|) \cdot d\mathbf{S}'$  and  $\nabla(1/|\mathbf{x} - \mathbf{x}'|) = -\nabla'(1/|\mathbf{x} - \mathbf{x}'|)$  in Eq. (B5), it becomes

$$\mathbf{A} = \frac{\Phi}{4\pi} \left[ -\nabla \int_{\mathcal{S}} \nabla' \left( \frac{1}{|\mathbf{x} - \mathbf{x}'|} \right) \cdot d\mathbf{S}' - \int_{\mathcal{S}} \nabla^2 \left( \frac{d\mathbf{S}'}{|\mathbf{x} - \mathbf{x}'|} \right) \right]. \tag{B6}$$

Using  $\nabla'(1/|\mathbf{x} - \mathbf{x}'|) = (\mathbf{x} - \mathbf{x}')/|\mathbf{x} - \mathbf{x}'|^3$  and  $\nabla^2(1/|\mathbf{x} - \mathbf{x}'|) = -4\pi\delta(\mathbf{x} - \mathbf{x}')$  in Eq. (B6), we obtain

$$\mathbf{A} = \frac{\Phi}{4\pi} \left[ \nabla \int_{\mathcal{S}} \frac{(\mathbf{x}' - \mathbf{x}) \cdot d\mathbf{S}'}{|\mathbf{x} - \mathbf{x}'|^3} + 4\pi \int_{\mathcal{S}} \delta(\mathbf{x} - \mathbf{x}') d\mathbf{S}' \right]. \tag{B7}$$

The first integral is identified with the single-valued solid angle  $\Omega_0$  defined in Eq. (9) while the second integral is identified with the surface vector Dirac delta  $\delta_{\mathcal{S}}$  specified in Eq. (10). Thus, we get Eq. (8):  $\mathbf{A} = \Phi \nabla \Omega_0 / (4\pi) + \Phi \delta_{\mathcal{S}}$ .

### Appendix C. Proof of Eq. (11)

The proof of Eq. (11) will be developed in two parts. In the first part we will explicitly demonstrate that the circulation of the gradient of  $\Omega_0$  along an arbitrary closed path  $C$  vanishes  $\oint_C \nabla \Omega_0 \cdot d\mathbf{x} = 0$ . In the second part we will transform this circulation using the Stokes theorem  $\oint_C \nabla \Omega_0 \cdot d\mathbf{x} = 0 = \int_S \nabla \times \nabla \Omega_0 \cdot d\mathbf{S}$  to show  $\nabla \times \nabla \Omega_0 = 0$ . This last result will be used to demonstrate Eq. (11).

Let us obtain a suitable form of the gradient of  $\Omega_0$ . The gradient of Eq. (9) gives

$$\nabla \Omega_0 = \nabla \int_S \frac{(\mathbf{x}' - \mathbf{x}) \cdot d\mathbf{S}'}{|\mathbf{x} - \mathbf{x}'|^3}. \tag{C1}$$

Using  $\nabla(1/|\mathbf{x} - \mathbf{x}'|) = -(\mathbf{x} - \mathbf{x}')/|\mathbf{x} - \mathbf{x}'|^3$ , Eq. (C1) becomes

$$\nabla \Omega_0 = -\nabla \int_S \nabla' \left( \frac{1}{|\mathbf{x} - \mathbf{x}'|} \right) \cdot d\mathbf{S}'. \tag{C2}$$

Considering the relations  $\nabla'(1/|\mathbf{x} - \mathbf{x}'|) = -\nabla(1/|\mathbf{x} - \mathbf{x}'|)$  and  $\nabla(1/|\mathbf{x} - \mathbf{x}'|) \cdot d\mathbf{S}' = \nabla \cdot (d\mathbf{S}'/|\mathbf{x} - \mathbf{x}'|)$  in Eq. (C2), we obtain

$$\nabla \Omega_0 = \nabla \int_S \nabla \cdot \left( \frac{d\mathbf{S}'}{|\mathbf{x} - \mathbf{x}'|} \right). \tag{C3}$$

When the identity  $\nabla(\nabla \cdot \mathbf{F}) = \nabla \times (\nabla \times \mathbf{F}) + \nabla^2 \mathbf{F}$  is used in Eq. (C3), it becomes

$$\nabla \Omega_0 = \nabla \times \left[ \nabla \times \int_S \frac{d\mathbf{S}'}{|\mathbf{x} - \mathbf{x}'|} \right] + \nabla^2 \int_S \frac{d\mathbf{S}'}{|\mathbf{x} - \mathbf{x}'|}. \tag{C4}$$

Inserting  $\nabla \times (d\mathbf{S}'/|\mathbf{x} - \mathbf{x}'|) = -d\mathbf{S}' \times \nabla(1/|\mathbf{x} - \mathbf{x}'|)$  together with  $\nabla(1/|\mathbf{x} - \mathbf{x}'|) = -\nabla'(1/|\mathbf{x} - \mathbf{x}'|)$  in the quantity within the brackets and using  $\nabla^2(1/|\mathbf{x} - \mathbf{x}'|) = -4\pi\delta(\mathbf{x} - \mathbf{x}')$  on the second term of Eq. (C4), we obtain

$$\nabla \Omega_0 = \nabla \times \left[ \int_S d\mathbf{S}' \times \nabla' \left( \frac{1}{|\mathbf{x} - \mathbf{x}'|} \right) \right] - 4\pi \int_S \delta(\mathbf{x} - \mathbf{x}') d\mathbf{S}'. \tag{C5}$$

The quantity within the brackets in Eq. (C5) can be transformed into a closed line integral via the Stokes theorem

$$\int_S d\mathbf{S}' \times \nabla' \left( \frac{1}{|\mathbf{x} - \mathbf{x}'|} \right) = \oint_{\mathcal{C}} \frac{d\mathbf{x}'}{|\mathbf{x} - \mathbf{x}'|}, \tag{C6}$$

where  $\mathcal{C}$  is the boundary of  $S$ . When Eq. (C6) and the surface vector Dirac delta  $\int_S \delta(\mathbf{x} - \mathbf{x}') d\mathbf{S}' = \delta_S$  given in Eq. (10) are used in Eq. (C5), it takes the form

$$\nabla \Omega_0 = \nabla \times \oint_{\mathcal{C}} \frac{d\mathbf{x}'}{|\mathbf{x} - \mathbf{x}'|} - 4\pi \delta_S. \tag{C7}$$

Considering  $\nabla \times (d\mathbf{x}'/|\mathbf{x} - \mathbf{x}'|) = \nabla(1/|\mathbf{x} - \mathbf{x}'|) \times d\mathbf{x}'$  and  $\nabla(1/|\mathbf{x} - \mathbf{x}'|) = (\mathbf{x}' - \mathbf{x})/|\mathbf{x} - \mathbf{x}'|^3$  we can write Eq. (C7) as

$$\nabla \Omega_0 = \oint_{\mathcal{C}} \frac{(\mathbf{x}' - \mathbf{x}) \times d\mathbf{x}'}{|\mathbf{x} - \mathbf{x}'|^3} - 4\pi \delta_S, \tag{C8}$$

which is a suitable form of the gradient of  $\Omega_0$ . Let us now take the circulation to Eq. (C8) along an arbitrary closed path  $C$

$$\oint_C \nabla \Omega_0 \cdot d\mathbf{x} = \oint_C \oint_{\mathcal{C}} \frac{[(\mathbf{x}' - \mathbf{x}) \times d\mathbf{x}'] \cdot d\mathbf{x}}{|\mathbf{x} - \mathbf{x}'|^3} - 4\pi \oint_C \delta_S \cdot d\mathbf{x}. \tag{C9}$$

Making use of the relation  $[(\mathbf{x}' - \mathbf{x}) \times d\mathbf{x}'] \cdot d\mathbf{x} = (\mathbf{x} - \mathbf{x}') \cdot (d\mathbf{x} \times d\mathbf{x}')$  in the first term of the right-hand side of Eq. (C9), we obtain

$$\oint_C \nabla \Omega_0 \cdot d\mathbf{x} = 4\pi \left[ \frac{1}{4\pi} \oint_C \oint_{\mathcal{C}} \frac{(\mathbf{x} - \mathbf{x}') \cdot (d\mathbf{x} \times d\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|^3} \right] - 4\pi \left[ \oint_C \delta_S \cdot d\mathbf{x} \right]. \tag{C10}$$

The first quantity within the brackets is the linking number  $l$  defined by Eq. (21). The second quantity within the brackets is another equivalent form of the linking number defined by Eq. (27). Since  $C$  corresponds to the same path in the two closed line integrals on the right-hand side of Eq. (C10) then it follows that

$$\oint_C \nabla \Omega_0 \cdot d\mathbf{x} = \begin{cases} 4\pi l - 4\pi l = 0 & \text{if } C \text{ encloses } \mathcal{C} \\ 0 & \text{otherwise} \end{cases} \tag{C11}$$

We observe that regardless of the path  $C$  we have the vanishing of the circulation

$$\oint_C \nabla \Omega_0 \cdot d\mathbf{x} = 0, \tag{C12}$$

which is the first step in the proof of Eq. (11). In the second step we transform the left-hand side of Eq. (C12) into a surface integral via the Stokes theorem

$$\oint_C \nabla \Omega_0 \cdot d\mathbf{x} = \int_S \nabla \times \nabla \Omega_0 \cdot d\mathbf{S}, \tag{C13}$$

where  $S$  is the surface enclosed by  $C$ . Equations (C12) and (C13) imply

$$\oint_C \nabla \Omega_0 \cdot d\mathbf{x} = 0 = \int_S \nabla \times \nabla \Omega_0 \cdot d\mathbf{S}. \tag{C14}$$

Since this result holds for any path  $C$  then it follows that the second equality in Eq. (C14) is valid for any surface  $S$  implying the vanishing of the curl of the gradient of  $\Omega_0$  in all space

$$\nabla \times \nabla \Omega_0 = 0. \tag{C15}$$

To show Eq. (11) let us write Eq. (C14) in index notation

$$\oint_C \partial_k \Omega_0 dx^k = 0 = \int_S \varepsilon_{kmn} \partial^m \partial^n \Omega_0 dS^k. \tag{C16}$$

Consider now the antisymmetric tensor  $dS_{ij} = \varepsilon_{ijk} dS^k$  representing an infinitesimal element of the surface  $S$ . In terms of  $dS_{ij}$  we may write the differential surface vector in the following form  $dS^k = (1/2)\varepsilon^{kij} dS_{ij}$ . Using this result together with the identity  $\varepsilon_{kmn}\varepsilon^{kij} = \delta_m^i \delta_n^j - \delta_m^j \delta_n^i$  we obtain  $\varepsilon_{kmn} \partial^m \partial^n \Omega_0 dS^k = (1/2)(\partial^i \partial^j - \partial^j \partial^i) \Omega_0 dS_{ij}$ , which is used in the second equality in Eq. (C16) to obtain the relation

$$2 \oint_C \partial_k \Omega_0 dx^k = 0 = \int_S (\partial^i \partial^j - \partial^j \partial^i) \Omega_0 dS_{ij}. \tag{C17}$$

Since the first equality is valid for any path  $C$  then the second equality is valid for any surface  $S$  and this implies Eq. (11):  $(\partial^i \partial^j - \partial^j \partial^i) \Omega_0 = 0$  in all space.

**Appendix D. Proof of Eq. (15)**

Our approach to show Eq. (15) is as follows. We will show that the circulation of the gradient of the solid angle  $\Omega$  along an arbitrary closed path  $C$  is non-vanishing:  $\oint_C \nabla \Omega \cdot d\mathbf{x} \neq 0$ . Then we will transform this circulation via the Stokes theorem  $\oint_C \nabla \Omega \cdot d\mathbf{x} = \int_S \nabla \times \nabla \Omega \cdot d\mathbf{S}$  to show  $\nabla \times \nabla \Omega = \delta_{\mathcal{C}}$ . We will use this result to demonstrate Eq. (15).

Using Eq. (14), the circulation of the gradient of  $\Omega$  takes the form

$$\oint_C \nabla \Omega \cdot d\mathbf{x} = \oint_C \nabla \Omega_0 \cdot d\mathbf{x} + 4\pi \oint_C \delta_{\mathcal{S}} \cdot d\mathbf{x}. \tag{D1}$$

We can transform the first circulation on the right-hand side of Eq. (D1) using Eq. (C10). This gives

$$\begin{aligned} \oint_C \nabla \Omega \cdot d\mathbf{x} &= \oint_C \oint_{\mathcal{C}} \frac{(\mathbf{x} - \mathbf{x}') \cdot (d\mathbf{x} \times d\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|^3} - 4\pi \oint_C \delta_{\mathcal{S}} \cdot d\mathbf{x} + 4\pi \oint_C \delta_{\mathcal{S}} \cdot d\mathbf{x} \\ &= 4\pi \left[ \frac{1}{4\pi} \oint_C \oint_{\mathcal{C}} \frac{(\mathbf{x} - \mathbf{x}') \cdot (d\mathbf{x} \times d\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|^3} \right]. \end{aligned} \tag{D2}$$

The quantity within the brackets is the Gauss linking number defined in Eq. (21). Therefore

$$\oint_C \nabla \Omega \cdot d\mathbf{x} = \begin{cases} 4\pi l & \text{if } C \text{ encloses } \mathcal{C} \\ 0 & \text{otherwise} \end{cases} \tag{D3}$$

Using the Stokes theorem we can transform the left-hand side of Eq. (D3),

$$\oint_C \nabla \Omega \cdot d\mathbf{x} = \int_S \nabla \times \nabla \Omega \cdot d\mathbf{S}, \tag{D4}$$

where  $C$  is the boundary of the surface  $S$ . When  $C$  does not enclose  $\mathcal{C}$  then from Eq. (D3) we have

$$\oint_C \nabla \Omega \cdot d\mathbf{x} = 0 = \int_S \nabla \times \nabla \Omega \cdot d\mathbf{S}, \tag{D5}$$

and thus  $\nabla \times \nabla \Omega = 0$  locally holds for any surface  $S$  not pierced by  $\mathcal{C}$ . However, this result does not hold in all space (i.e. globally) because if  $C$  encloses  $\mathcal{C}$  then the left-hand side of Eq. (D5) is non-vanishing and from Eq. (D3) we obtain

$$\oint_C \nabla \Omega \cdot d\mathbf{x} = 4\pi l = \int_S \nabla \times \nabla \Omega \cdot d\mathbf{S}, \tag{D6}$$

and thus the relation  $\nabla \times \nabla \Omega \neq 0$  holds. To find the explicit form of this relation, we use Eq. (14):  $\nabla \Omega = \nabla \Omega_0 + 4\pi \delta_{\mathcal{S}}$  and therefore  $\nabla \times \nabla \Omega = 4\pi \delta_{\mathcal{C}}$  where we have used Eq. (C15):  $\nabla \times \nabla \Omega_0 = 0$  and Eq. (12):  $\nabla \times \delta_{\mathcal{S}} = \delta_{\mathcal{C}}$ , where  $\delta_{\mathcal{C}} = \oint_{\mathcal{C}} \delta(\mathbf{x} - \mathbf{x}') d\mathbf{x}'$  is a line Dirac delta along the closed path  $\mathcal{C}$  which forms the boundary of  $\mathcal{S}$ . Thus

$$\int_S \nabla \times \nabla \Omega \cdot d\mathbf{S} = 4\pi \int_S \delta_{\mathcal{C}} \cdot d\mathbf{S}, \tag{D7}$$

which implies

$$\nabla \times \nabla \Omega = 4\pi \delta_{\mathcal{C}}. \tag{D8}$$

To prove Eq. (15), we write Eq. (D4) in index notation

$$\oint_C \partial_k \Omega dx^k = \int_S \varepsilon_{kmn} \partial^m \partial^n \Omega dS^k. \tag{D9}$$

Consider now the antisymmetric tensor  $dS_{ij} = \varepsilon_{ijk} dS^k$  representing an infinitesimal element of the surface  $S$ . Using this result and the identity  $\varepsilon_{kmn} \varepsilon^{kij} = \delta_m^i \delta_n^j - \delta_m^j \delta_n^i$ , we obtain  $\varepsilon_{kmn} \partial^m \partial^n \Omega dS^k = (1/2)(\partial^i \partial^j - \partial^j \partial^i) \Omega dS_{ij}$ , which is used in the second equality in Eq. (D9), obtaining

$$2 \oint_C \partial_k \Omega dx^k = \int_S (\partial^i \partial^j - \partial^j \partial^i) \Omega dS_{ij}. \tag{D10}$$

When the path  $C$  does not enclose the curve  $\mathcal{C}$  then from Eq. (D3) we have

$$2 \oint_C \partial_k \Omega dx^k = 0 = \int_S (\partial^i \partial^j - \partial^j \partial^i) \Omega dS_{ij}, \tag{D11}$$

which implies  $(\partial^i \partial^j - \partial^j \partial^i) \Omega = 0$  for any surface  $S$  not pierced by  $\mathcal{C}$ . In this case  $\Omega$  is locally single-valued. However, this is not the global case for if  $C$  encloses  $\mathcal{C}$  then the left-hand side of Eq. (D10) is non-vanishing and from Eq. (D3) we obtain

$$2 \oint_C \partial_k \Omega dx^k = 8\pi l = \int_S (\partial^i \partial^j - \partial^j \partial^i) \Omega dS_{ij}. \tag{D12}$$

which implies  $(\partial^i \partial^j - \partial^j \partial^i) \Omega \neq 0$  when  $C$  encircles  $\mathcal{C}$ , or equivalently stated, when  $C$  crosses  $S$ . To find the explicit form of  $(\partial^i \partial^j - \partial^j \partial^i) \Omega$  we use Eqs. (D12) and (D9) to obtain

$$\int_S (\partial^i \partial^j - \partial^j \partial^i) \Omega dS_{ij} = 2 \int_S \varepsilon_{kmn} \partial^m \partial^n \Omega dS^k. \tag{D13}$$

Equation (D8) in index notation reads  $4\pi(\delta_{\mathcal{C}})_k = \varepsilon_{kmn} \partial^m \partial^n \Omega$ . This result and  $dS^k = (1/2)\varepsilon^{kij} dS_{ij}$  give the relation  $\varepsilon_{kmn} \partial^m \partial^n \Omega dS^k = 2\pi \varepsilon^{ijk} (\delta_{\mathcal{C}})_k dS_{ij}$  so that Eq. (D13) reduces to

$$\int_S (\partial^i \partial^j - \partial^j \partial^i) \Omega dS_{ij} = 4\pi \int_S \varepsilon^{ijk} (\delta_{\mathcal{C}})_k dS_{ij}, \tag{D14}$$

and this implies Eq. (15):  $(\partial^i \partial^j - \partial^j \partial^i) \Omega = 4\pi \varepsilon^{ijk} (\delta_{\mathcal{C}})_k$ .

### Appendix E. Proofs of Eqs. (74), (77), and (80)

The proof of Eq. (74) is based on a proof given by Gelca [41]. Similar proofs for Eqs. (77) and (80) will be given. Our general strategy is as follows: we will apply topological transformations to the linking number  $l$  (i.e. deformations of the associated curves in  $l$ ) and show that these transformations leave the linking number invariant.

**Proof of Eq. (74).** Consider the linking number of the curves  $\mathcal{C}$  and  $\mathcal{C}$

$$\frac{1}{4\pi} \oint_C \oint_{\mathcal{C}} \frac{(\mathbf{x} - \mathbf{x}') \cdot (d\mathbf{x} \times d\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|^3} = l(\mathcal{C}, \mathcal{C}). \tag{E1}$$

Let  $C$  be a closed path encircling the curve  $\mathcal{C}$  and let us deform the path  $C$  into the path  $C'$  via the transformation  $C \rightarrow C'$  and let  $\mathbb{C} = C \cup (-C')$  be the union of  $C$  and  $(-C')$  which bounds the surface  $\mathbb{S}$  traced by  $C$  while being deformed into  $C'$ . Accordingly,

$\mathbb{C} = C \cup (-C') = \partial\mathbb{S}$  where  $\partial\mathbb{S}$  is the boundary of  $\mathbb{S}$ . We also assume  $C$  and  $C'$  encircle the same number of times  $\mathbb{C}$ . Using the properties  $\oint_{\mathbb{C}} = \oint_C \cup \oint_{-C'} = \oint_C + \oint_{-C'}$  and  $\oint_{-C'} = -\oint_{C'}$  it follows that Eq. (E1) can be decomposed as

$$\frac{1}{4\pi} \oint_C \oint_{\mathbb{C}} \frac{(\mathbf{x} - \mathbf{x}') \cdot (d\mathbf{x} \times d\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|^3} = \frac{1}{4\pi} \oint_C \oint_{\mathbb{C}} \frac{(\mathbf{x} - \mathbf{x}') \cdot (d\mathbf{x} \times d\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|^3} - \frac{1}{4\pi} \oint_{C'} \oint_{\mathbb{C}} \frac{(\mathbf{x} - \mathbf{x}') \cdot (d\mathbf{x} \times d\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|^3}, \tag{E2}$$

or equivalently,

$$l(\mathbb{C}, \mathbb{C}) = l(C, \mathbb{C}) - l(C', \mathbb{C}), \tag{E3}$$

where

$$\frac{1}{4\pi} \oint_C \oint_{\mathbb{C}} \frac{(\mathbf{x} - \mathbf{x}') \cdot (d\mathbf{x} \times d\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|^3} = l(C, \mathbb{C}), \quad \frac{1}{4\pi} \oint_{C'} \oint_{\mathbb{C}} \frac{(\mathbf{x} - \mathbf{x}') \cdot (d\mathbf{x} \times d\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|^3} = l(C', \mathbb{C}), \tag{E4}$$

are the linking numbers of  $C$  and  $\mathbb{C}$ , and  $C'$  and  $\mathbb{C}$ , respectively. Therefore if  $l(\mathbb{C}, \mathbb{C}) = 0$  then  $l(C, \mathbb{C}) = l(C', \mathbb{C})$  and this would prove Eq. (74). In Appendix C we demonstrated Eq. (C8) which can be re-arranged to obtain the relation

$$\oint_{\mathbb{C}} \frac{(\mathbf{x}' - \mathbf{x}) \times d\mathbf{x}'}{|\mathbf{x} - \mathbf{x}'|^3} = \nabla\Omega_0(\mathbb{S}) + 4\pi\delta_{\mathbb{S}}, \tag{E5}$$

where  $\Omega_0(\mathbb{S}) = \int_{\mathbb{S}} \{(\mathbf{x}' - \mathbf{x}) \cdot d\mathbf{S}' / |\mathbf{x} - \mathbf{x}'|^3\}$  is the single-valued solid angle subtended by  $\mathbb{C}$  and  $\delta_{\mathbb{S}} = \int_{\mathbb{S}} \delta(\mathbf{x} - \mathbf{x}') d\mathbf{S}'$  is the surface vector Dirac delta defined along the surface  $\mathbb{S}$  bounded by  $\mathbb{C}$ . Using Eq. (E5) in the left-hand side of Eq. (E1) and  $\{(\mathbf{x}' - \mathbf{x}) \times d\mathbf{x}'\} \cdot d\mathbf{x} = (\mathbf{x} - \mathbf{x}') \cdot (d\mathbf{x} \times d\mathbf{x}')$  it follows

$$l(\mathbb{C}, \mathbb{C}) = \frac{1}{4\pi} \oint_{\mathbb{C}} \nabla\Omega_0(\mathbb{S}) \cdot d\mathbf{x} + \oint_{\mathbb{C}} \delta_{\mathbb{S}} \cdot d\mathbf{x}. \tag{E6}$$

The first line integral in the right-hand side vanishes because  $\Omega_0(\mathbb{C})$  is a single-valued function. Applying the Stokes theorem to the second line integral in the right-hand side, we obtain

$$l(\mathbb{C}, \mathbb{C}) = \int_{\mathbb{S}} \delta_{\mathbb{C}} \cdot d\mathbf{S}, \tag{E7}$$

where  $\mathbb{S}$  is the surface bounded by  $\mathbb{C}$  and we have used Eq. (12):  $\nabla \times \delta_{\mathbb{S}} = \delta_{\mathbb{C}}$  where  $\delta_{\mathbb{C}} = \oint_{\mathbb{C}} \delta(\mathbf{x} - \mathbf{x}') d\mathbf{x}'$  is a line vector Dirac delta defined along  $\mathbb{C}$ . The surface  $\mathbb{S}$  corresponds to the surface traced by the path  $C$  while being deformed into the path  $C'$  and therefore the curve  $\mathbb{C}$  never crosses the surface  $\mathbb{S}$ . Accordingly, the function  $\delta_{\mathbb{C}}$  vanishes along the surface  $\mathbb{S}$  and therefore  $\int_{\mathbb{S}} \delta_{\mathbb{C}} \cdot d\mathbf{S} = 0$  which gives  $l(\mathbb{C}, \mathbb{C}) = 0$ . This result and Eq. (E3) imply  $l(C, \mathbb{C}) = l(C', \mathbb{C})$  and this proves Eq. (74).

**Proof of Eq. (77).** Consider the linking number of the curves  $C$  and  $\mathbb{C}$

$$\frac{1}{4\pi} \oint_C \oint_{\mathbb{C}} \frac{(\mathbf{x} - \mathbf{x}') \cdot (d\mathbf{x} \times d\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|^3} = l(C, \mathbb{C}). \tag{E8}$$

Let  $C$  be a closed path encircling the curve  $\mathbb{C}$ . Let us deform the curve  $\mathbb{C}$  into the curve  $\mathbb{C}'$  via the transformation  $\mathbb{C} \rightarrow \mathbb{C}'$  and let  $\mathcal{C} = \mathbb{C} \cup (-\mathbb{C}')$  be the union of  $\mathbb{C}$  and  $(-\mathbb{C}')$  which bounds the surface  $\mathcal{S}$  traced by  $\mathbb{C}$  while being deformed into  $\mathbb{C}'$ . It follows that  $\mathcal{C} = \mathbb{C} \cup (-\mathbb{C}') = \partial\mathcal{S}$  where  $\partial\mathcal{S}$  is the boundary of  $\mathcal{S}$ . We assume  $C$  encircles the same number of times  $\mathbb{C}$  and  $\mathbb{C}'$ . Using the properties  $\oint_{\mathcal{C}} = \oint_{\mathbb{C}} \cup \oint_{-\mathbb{C}'} = \oint_{\mathbb{C}} + \oint_{-\mathbb{C}'}$  it follows that Eq. (E8) can be decomposed as

$$\frac{1}{4\pi} \oint_C \oint_{\mathcal{C}} \frac{(\mathbf{x} - \mathbf{x}') \cdot (d\mathbf{x} \times d\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|^3} = \frac{1}{4\pi} \oint_C \oint_{\mathbb{C}} \frac{(\mathbf{x} - \mathbf{x}') \cdot (d\mathbf{x} \times d\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|^3} - \frac{1}{4\pi} \oint_C \oint_{\mathbb{C}'} \frac{(\mathbf{x} - \mathbf{x}') \cdot (d\mathbf{x} \times d\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|^3}, \tag{E9}$$

or equivalently,

$$l(C, \mathcal{C}) = l(C, \mathbb{C}) - l(C, \mathbb{C}'), \tag{E10}$$

where

$$\frac{1}{4\pi} \oint_C \oint_{\mathcal{C}} \frac{(\mathbf{x} - \mathbf{x}') \cdot (d\mathbf{x} \times d\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|^3} = l(C, \mathcal{C}), \quad \frac{1}{4\pi} \oint_C \oint_{\mathbb{C}'} \frac{(\mathbf{x} - \mathbf{x}') \cdot (d\mathbf{x} \times d\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|^3} = l(C, \mathbb{C}'), \tag{E11}$$

are the linking numbers of  $C$  and  $\mathcal{C}$ , and  $C$  and  $\mathbb{C}'$ , respectively. Therefore if  $l(C, \mathcal{C}) = 0$  then  $l(C, \mathbb{C}) = l(C, \mathbb{C}')$  and this would prove Eq. (77). Following the same line of arguments that led to Eq. (C8), it follows that we can make the replacement  $\mathbb{S} \rightarrow \mathcal{S}$  in Eq. (C8) and obtain

$$\oint_C \frac{(\mathbf{x}' - \mathbf{x}) \times d\mathbf{x}'}{|\mathbf{x} - \mathbf{x}'|^3} = \nabla\Omega_0(\mathcal{S}) + 4\pi\delta_{\mathcal{S}}, \tag{E12}$$

where  $\Omega_0(S) = \int_S \{(\mathbf{x}' - \mathbf{x}) \cdot d\mathbf{S}' / |\mathbf{x} - \mathbf{x}'|^3\}$  is the single-valued solid angle function subtended by the curve  $C$  and  $\delta_S = \int_S \delta(\mathbf{x} - \mathbf{x}') d\mathbf{S}'$  is the surface vector Dirac delta defined along the surface  $S$  bounded by  $C$ . Using Eq. (E12) and the relation  $[(\mathbf{x}' - \mathbf{x}) \times d\mathbf{x}'] \cdot d\mathbf{x} = (\mathbf{x} - \mathbf{x}') \cdot (d\mathbf{x} \times d\mathbf{x}')$  we obtain

$$l(C, C) = \frac{1}{4\pi} \oint_C \nabla \Omega_0(S) \cdot d\mathbf{x} + \oint_C \delta_S \cdot d\mathbf{x}. \tag{E13}$$

The first line integral on the right-hand side vanishes because  $\Omega_0(S)$  is single-valued. On the other hand, the surface  $S$  corresponds to the surface traced by the curve  $C$  while being deformed into the curve  $C'$  and therefore the path  $C$  never crosses the surface  $S$ . Consequently, the function  $\delta_S$  vanishes along the path  $C$  so that  $\oint_C \delta_S \cdot d\mathbf{x} = 0$  and this gives  $l(C, C) = 0$ . This result and the right-hand side of Eq. (E10) imply  $l(C, C) = l(C, C')$ , result that proves Eq. (77).

**Proof of Eq. (80).** Consider the linking number of the curves  $C$  and  $C'$

$$\frac{1}{4\pi} \oint_C \oint_{C'} \frac{(\mathbf{x} - \mathbf{x}') \cdot (d\mathbf{x} \times d\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|^3} = l(C, C'). \tag{E14}$$

Let  $C$  be a closed path encircling the curve  $C'$ . Let us simultaneously deform the path  $C$  into the path  $C'$  via the transformation  $C \rightarrow C'$  and deform the curve  $C'$  into the curve  $C'$  via the transformation  $C' \rightarrow C'$ . Let  $C = C \cup (-C')$  be the union of  $C$  and  $(-C')$  which bounds the surface  $\mathbb{S}$  traced by  $C$  while being deformed into  $C'$  and let  $C = C' \cup (-C)$  be the union of  $C'$  and  $(-C)$  which bounds the surface  $S$  traced by  $C'$  while being deformed into  $C'$ . Accordingly,  $C = C \cup (-C') = \partial\mathbb{S}$  where  $\partial\mathbb{S}$  is the boundary of  $\mathbb{S}$  and  $C = C' \cup (-C) = \partial S$  where  $\partial S$  is the boundary of  $S$ . We assume the path  $C$  encircles  $C'$  and  $C'$  the same number of times the path  $C'$  encircles  $C$  and  $C'$ . Using  $\oint_{C=C \cup (-C')} = \oint_C - \oint_{C'}$ ,  $\oint_{C=C' \cup (-C)} = \oint_{C'} - \oint_C$ ,  $\oint_{-C'} = -\oint_{C'}$ ,  $\oint_{-C} = -\oint_C$ , and  $\oint_{-C'} = -\oint_{C'}$ , we can decompose Eq. (E14) as

$$\begin{aligned} \frac{1}{4\pi} \oint_C \oint_C \frac{(\mathbf{x} - \mathbf{x}') \cdot (d\mathbf{x} \times d\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|^3} &= \frac{1}{4\pi} \oint_C \oint_{C'} \frac{(\mathbf{x} - \mathbf{x}') \cdot (d\mathbf{x} \times d\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|^3} - \frac{1}{4\pi} \oint_C \oint_{C'} \frac{(\mathbf{x} - \mathbf{x}') \cdot (d\mathbf{x} \times d\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|^3} \\ &\quad - \frac{1}{4\pi} \oint_{C'} \oint_{C'} \frac{(\mathbf{x} - \mathbf{x}') \cdot (d\mathbf{x} \times d\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|^3} + \frac{1}{4\pi} \oint_{C'} \oint_{C'} \frac{(\mathbf{x} - \mathbf{x}') \cdot (d\mathbf{x} \times d\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|^3}, \end{aligned} \tag{E15}$$

or equivalently,

$$l(C, C) = l(C, C') - l(C, C') - l(C', C) + l(C', C'), \tag{E16}$$

where the corresponding linking numbers are defined by

$$\frac{1}{4\pi} \oint_C \oint_{C'} \frac{(\mathbf{x} - \mathbf{x}') \cdot (d\mathbf{x} \times d\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|^3} = l(C, C'), \quad \frac{1}{4\pi} \oint_{C'} \oint_{C'} \frac{(\mathbf{x} - \mathbf{x}') \cdot (d\mathbf{x} \times d\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|^3} = l(C', C'), \tag{E17}$$

$$\frac{1}{4\pi} \oint_{C'} \oint_C \frac{(\mathbf{x} - \mathbf{x}') \cdot (d\mathbf{x} \times d\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|^3} = l(C', C), \quad \frac{1}{4\pi} \oint_C \oint_{C'} \frac{(\mathbf{x} - \mathbf{x}') \cdot (d\mathbf{x} \times d\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|^3} = l(C, C'). \tag{E18}$$

Now, we have the result  $l(C', C) = l(C, C')$  because of Eqs. (E3) and (E7) (which follows from the transformation  $C \rightarrow C'$ ). Also, we have the result  $l(C, C) = l(C, C')$  because of Eq. (E10) and (E13) (which follows from the transformation  $C \rightarrow C'$ ). Using these results Eq. (E16) reduces to

$$l(C, C) = l(C', C') - l(C, C'). \tag{E19}$$

Accordingly, if the left-hand side of Eq. (E19) vanishes then  $l(C, C) = l(C', C')$  and this would prove Eq. (80). Using Eq. (E12) together with  $[(\mathbf{x}' - \mathbf{x}) \times d\mathbf{x}'] \cdot d\mathbf{x} = (\mathbf{x} - \mathbf{x}') \cdot (d\mathbf{x} \times d\mathbf{x}')$  we can write

$$l(C, C) = \frac{1}{4\pi} \oint_C \nabla \Omega_0(S) \cdot d\mathbf{x} + \oint_C \delta_S \cdot d\mathbf{x}. \tag{E20}$$

The first line integral on the right-hand side vanishes because  $\Omega_0(S)$  is single-valued. This result and the relations  $\oint_{C=C \cup (-C')} = \oint_C - \oint_{C'}$  and  $\oint_{-C'} = -\oint_{C'}$  yield  $l(C, C) = \oint_C \delta_S \cdot d\mathbf{x} - \oint_{C'} \delta_S \cdot d\mathbf{x}$ . The surface  $S$  corresponds to the surface traced by the curve  $C'$  while being deformed into the curve  $C'$ . Accordingly, neither the path  $C$  nor the path  $C'$  cross the surface  $S$  along which  $\delta_S$  is non-vanishing. Therefore  $\oint_C \delta_S \cdot d\mathbf{x} = 0$  and  $\oint_{C'} \delta_S \cdot d\mathbf{x} = 0$  which implies  $l(C, C) = 0$ . This result and Eq. (E20) give  $l(C, C) = l(C', C')$  and this proves Eq. (80).

**References**

1. Y. Aharonov, D. Bohm, Significance of electromagnetic potentials in the quantum theory. *Phys. Rev.* **115**, 485–491 (1959)
2. D. Rohrlich, Y. Aharonov, *Quantum Paradoxes: Quantum Theory for the Perplexed* (Wiley-VCH, Weinheim, 2005)
3. R.P. Feynman, R.B. Leighton, M. Sands, *The Feynman Lectures on Physics*, vol. 2 (Addison-Wesley, Reading, 1964)

4. E. Cohen et al., Geometric phase from Aharonov-Bohm to Pancharatnam-Berry and beyond. *Nat. Rev. Phys.* **1**, 437–449 (2019)
5. R.G. Chambers, Shift of an electron interference pattern by enclosed magnetic flux. *Phys. Rev. Lett.* **5**, 3–5 (1960)
6. H.A. Fowler et al., Electron interferometer studies of iron whiskers. *J. Appl. Phys.* **32**, 1153–1155 (1961)
7. H. Boersch, H. Hamisch, K. Grohmann, Experimenteller Nachweis der Phasenverschiebung von Elektronenwellen durch das magnetische Vektorpotential. II. *Z. Phys.* **169**, 263–272 (1962)
8. G. Möllenstedt, W. Bayh, Messung der kontinuierlichen Phasenschiebung von Elektronenwellen im kraftfeldfreien Raum durch das magnetische vektorpotential einer Luftspule. *Naturwissenschaften* **49**, 81–82 (1962)
9. M. Peshkin, A. Tonomura, *The Aharonov-Bohm Effect* (Springer, Berlin, 1989)
10. A. Tonomura, The AB effect and its expanding applications. *J. Phys. A Math. Theor.* **43**, 354021 (2010)
11. A. Tonomura et al., Observation of Aharonov-Bohm effect by electron holography. *Phys. Rev. Lett.* **48**, 1443–1446 (1982)
12. A. Tonomura et al., Evidence for Aharonov-Bohm effect with magnetic field completely shielded from electron wave. *Phys. Rev. Lett.* **56**, 792–795 (1986)
13. N. Osakabe et al., Experimental confirmation of Aharonov-Bohm effect using a toroidal magnetic field confined by a superconductor. *Phys. Rev. A* **34**, 815–822 (1986)
14. V.L. Lyuboshitz, Ya. A. Smorodinskii, Aharonov-Bohm effect in a toroidal solenoid. *Zh. Eksp. Teor. Fiz.* **75**, 40–45 (1978) [*Sov. Phys. JETP* **48**, 19–24 (1978)]
15. V.V. Lyuboshitz, V.L. Lyuboshitz, Momentum-transfer scattering cross section and the Aharonov-Bohm effect on a toroidal solenoid. *Zh. Eksp. Teor. Fiz.* **118**, 777–786 (2000) [*Sov. Phys.* **91**, 673–681 (2000)]
16. N.A. Nemkov, A.A. Basharin, V.A. Fedotov, Nonradiating sources, dynamic anapole, and Aharonov-Bohm effect. *Phys. Rev. B* **95**, 165134 (2017)
17. G.N. Afanasiev, *Topological Effects in Quantum Mechanics* (Springer, Netherlands, 1999)
18. G.N. Afanasiev, Closed analytical expressions for some useful sums and integrals involving Legendre functions. *J. Comput. Phys.* **69**, 196–208 (1987)
19. G.N. Afanasiev, D.M. Dubovik, Electromagnetic properties of a toroidal solenoid. *J. Phys. A Math. Gen.* **25**, 4869–4886 (1992)
20. N.J. Carron, On the fields of a torus and the role of the vector potential. *Am. J. Phys.* **63**, 717–729 (1995)
21. F.W. Wiegel, Path integrals with topological constraints: Aharonov-Bohm effect and polymer entanglements. *Physica A* **109**, 609–617 (1981)
22. J. Maeda, K. Shizuya, The Aharonov-Bohm and Aharonov-Casher effects and electromagnetic angular momentum. *Zeitschrift für Physik C* **60**, 265–270 (1993)
23. H.J. Rothe, Comments on the theory of the Aharonov-Bohm effect. *Nuov. Cim. A* **62**, 54–67 (1981)
24. H. Kleinert, *Multivalued Fields in Condensed Matter, Electromagnetism, and Gravitation* (World Scientific, Singapore, 2008)
25. H. Kleinert, *Path Integrals in Quantum Mechanics, Statistics, Polymer Physics, and Financial Markets* (World Scientific, Singapore, 2009)
26. R. DeWit, Theory of disclinations: II. Continuous and discrete disclinations in anisotropic elasticity. *J. Res. Nat. Bur. Stand. (USA)* **77A**, 49–100 (1973)
27. I.A. Kunin, Fields of randomly distributed dislocations and force dipoles in an infinite elastic anisotropic medium. *J. Appl. Mech. Tech. Phys.* **6**, 49–53 (1965)
28. A. Zangwill, *Modern Electrodynamics* (Cambridge University Press, Cambridge, 2012)
29. J. Schwinger, L.L. De Raad, Jr., K. A. Milton, and W. Y. Tsai, *Classical Electrodynamics* (Perseus, Reading, 1998)
30. L. Eyges, *The Classical Electromagnetic Field* (Addison-Wesley, Reading, 1972)
31. J. Djurić, Double layers and solid angle in potential theory. *Am. J. Phys.* **35**, 411–415 (1967)
32. H. Kleinert, Double-gauge invariance and local quantum field theory of charges and Dirac magnetic monopoles. *Phys. Lett. B* **246**, 127–130 (1990)
33. H. Kleinert, The extra Gauge symmetry of string deformations in electromagnetism with charges and Dirac monopoles. *Int. J. Mod. Phys. A* **7**, 4693–4705 (1992)
34. P.A.M. Dirac, Quantised singularities in the electromagnetic field. *Proc. R. Soc. Lond. A* **133**, 60–72 (1931)
35. R. Heras, Dirac quantisation condition: a comprehensive review. *Contemp. Phys.* **59**, 331–355 (2018)
36. Y.M. Shnir, *Magnetic Monopoles* (Springer, Berlin, 2005)
37. R.L. Ricca, B. Nipoti, Gauss' linking number revisited. *J. Knot Theory Ramifications* **20**, 1325–1343 (2011)
38. J.A. Heras, R. Heras, Can classical electrodynamics predict nonlocal effects? *Eur. Phys. J. Plus* **136**, 847 (2021)
39. J.A. Heras, R. Heras, Topology, nonlocality and duality in classical electrodynamics. *Eur. Phys. J. Plus* **137**, 157 (2022)
40. D.H. Kobe, Aharonov-Bohm effect revisited. *Ann. Phys.* **123**, 381–410 (1979)
41. R. Gelca, *Theta Functions and Knots* (World Scientific, Singapore, 2014)
42. T. O. Eynck, H. Lyre, N. V. Rummell, *A versus B! Topological nonseparability and the Aharonov-Bohm effect*, E-print PITT-PHIL-SCI00000404 (2001)
43. Y. Aharonov, E. Cohen, D. Rohrlich, Nonlocality of the Aharonov-Bohm effect. *Phys. Rev. A* **93**, 042110 (2016)
44. E. Shech, Idealizations, essential self-adjointness, and minimal model explanation in the Aharonov-Bohm effect. *Synthese* **195**, 4839–4863 (2018)
45. P. Pearle, Feynman's lecture utilizing the Aharonov-Bohm effect. *Quantum Stud. Math. Found.* **5**, 391–398 (2018)
46. T. Maudlin, Ontological clarity via canonical presentation: electromagnetism and the Aharonov-Bohm effect. *Entropy* **20**, 465 (2018)
47. S.C. Tiwari, Physical reality of electromagnetic potentials and the classical limit of the Aharonov-Bohm effect. *Quant. Stud. Math. Found.* **5**, 279–295 (2018)
48. M. Wakamatsu et al., The role of electron orbital angular momentum in the Aharonov-Bohm effect revisited. *Ann. Phys.* **38**, 259–277 (2018)
49. J. Earman, The role of idealizations in the Aharonov-Bohm effect. *Synthese* **196**, 1991–2019 (2019)
50. J. Dougherty, The non-ideal theory of the Aharonov-Bohm effect. *Synthese* **198**, 12195–12221 (2021)
51. V. Ardourel, A. Guay, Why is the transference theory of causation insufficient? The challenge of the Aharonov-Bohm effect. *Stud. Hist. Philos. Sci. B* **63**, 12–23 (2018)
52. R.A. Mulder, Gauge-underdetermination and shades of locality in the Aharonov-Bohm effect. *Found. Phys.* **51**, 48 (2021)
53. L. Vaidman, Role of potentials in the Aharonov-Bohm effect. *Phys. Rev. A* **86**, 040101 (2012)
54. Y. Aharonov, E. Cohen, D. Rohrlich, Comment on “role of potentials in the Aharonov-Bohm effect”. *Phys. Rev. A* **92**, 026101 (2015)
55. L. Vaidman, Reply to “Comment on ‘role of potentials in the Aharonov-Bohm effect’”. *Phys. Rev. A* **92**, 026102 (2015)
56. M. Bunge, Does the Aharonov-Bohm effect occur? *Found. Sci.* **20**, 129–133 (2015)
57. R.F. Wang, A possible interplay between electron beams and magnetic fluxes in the Aharonov-Bohm effect. *Front. Phys.* **10**, 358–363 (2015)
58. K. Kicheon, Locality of the Aharonov-Bohm-Casher effect. *Phys. Rev. A* **91**, 052116 (2016)
59. P. Pearle, A. Rizzi, Quantum-mechanical inclusion of the source in the Aharonov-Bohm effects. *Phys. Rev. A* **95**, 052123 (2017)
60. P. Pearle, A. Rizzi, Quantized vector potential and alternative views of the magnetic Aharonov-Bohm phase shift. *Phys. Rev. A* **95**, 052124 (2017)
61. B. Li, D.W. Hewak, Q.J. Wang, The transition from quantum field theory to one-particle quantum mechanics and a proposed interpretation of Aharonov-Bohm effect. *Found. Phys.* **48**, 837–852 (2018)

62. M. El Atiki, M. Bendahane, A. Kassou-Ou-Ali, Aharonov-Bohm effect in the ghost interference. *Pramana J. Phys.* **91**, 76 (2018)
63. M. Becker et al., Asymmetry and non-dispersivity in the Aharonov-Bohm effect. *Nat. Commun.* **10**, 1700 (2019)
64. K.J. Kasunic, Magnetic Aharonov-Bohm effects and the quantum phase shift: a heuristic interpretation. *Am. J. Phys.* **87**, 745–751 (2019)
65. D.A. Slavnov, The Aharonov-Bohm Effect: an algebraic approach. *Phys. Part. Nucl.* **50**, 77–86 (2019)
66. J. Bernabeu, J. Navarro-Salas, A non-local action for electrodynamics: duality symmetry and the Aharonov-Bohm effect. Revisited. *Symmetry* **11**, 1191 (2019)
67. C. Marletto, V. Vedral, Aharonov-Bohm phase is locally generated like all other quantum phases. *Phys. Rev. Lett.* **125**, 040401 (2020)
68. G. Hetzroni, Relativity and equivalence in Hilbert space: a principle-theory approach to the Aharonov-Bohm effect. *Found. Phys.* **50**, 120–135 (2020)
69. C.R. de Oliveira, R.G. Romano, A new version of the Aharonov-Bohm effect. *Found. Phys.* **50**, 137–146 (2020)
70. K. Bhattacharya, Demystifying the nonlocality problem in Aharonov-Bohm effect. *Phys. Scr.* **96**, 084011 (2021)
71. P.L. Saldanha, Local description of the Aharonov-Bohm effect with a quantum electromagnetic field. *Found. Phys.* **51**, 6 (2021)
72. M.I. Wanas, M.M. Kamal, Z.A. Ismail, A pure geometric approach to the Aharonov-Bohm effect. *Indian J. Phys.* **95**, 2865–2871 (2021)
73. G. Spavieri et al., Effective interaction force between an electric charge and a magnetic dipole and locality (or nonlocality) in quantum effects of the Aharonov-Bohm type. *Chin. Phys. Lett.* **38**, 034101 (2021)
74. P.L. Saldanha, Aharonov-Casher and shielded Aharonov-Bohm effects with a quantum electromagnetic field. *Phys. Rev. A* **104**, 032219 (2021)
75. V. Vedral, *A Classical (Local) Account of The Aharonov-Bohm Effect* (2021). [arXiv:2111.00476](https://arxiv.org/abs/2111.00476)
76. K. J. Kasunic, *Shear of the vector potential in the Aharonov-Bohm effect* (2021). [arXiv:2112.10611](https://arxiv.org/abs/2112.10611)
77. Y. Aharonov, *Non-local phenomena and the Aharonov-Bohm effect*, Proc. Int. Symp. Foundations of Quantum Mechanics, Tokyo, 10–19 (1983). Reproduced in *Foundations of Quantum Mechanics in the Light of New Technology*, 8–17, World Scientific (1997)
78. Y. Aharonov et al., Classical analog to topological nonlocal quantum interference effects. *Phys. Rev. Lett.* **92**, 020401 (2004)
79. H. Batelaan, A. Tonomura, The Aharonov-Bohm effects: variations on a subtle theme. *Phys. Today* **62**, 38–43 (2009)
80. A. Caprez, B. Barwick, H. Batelaan, Macroscopic test of the Aharonov-Bohm effect. *Phys. Rev. Lett.* **99**, 210401 (2007)
81. M. Becker, H. Batelaan, Experimental test for approximately dispersionless forces in the Aharonov-Bohm effect. *Europhys. Lett.* **115**, 10011 (2016)
82. D. Dragoman, M. Dragoman, *Quantum Classical Analogies* (Springer, Berlin, 2004)
83. G. Rizzi, M.L. Ruggiero, The Sagnac phase shift suggested by the Aharonov-Bohm effect for relativistic matter beams. *Gen. Rel. Grav.* **35**, 1745–1760 (2003)
84. C.-H. Tsai, D. Neilson, New quantum interference effect in rotating systems. *Phys. Rev. A* **37**, 619–621 (1988)
85. M.V. Berry et al., Wavefront dislocations in the Aharonov-Bohm effect and its water wave analogue. *Eur. J. Phys.* **1**, 154–162 (1980)
86. M.V. Berry, The adiabatic phase and Pancharatnam's phase for polarized light. *J. Mod. Opt.* **34**, 1401–1407 (1987)
87. J.H. Hannay, Angle variable holonomy in adiabatic excursion of an integrable Hamiltonian. *J. Phys. A* **18**, 221–230 (1985)
88. N. Satapathy et al., Classical light analogue of the non-local Aharonov-Bohm effect. *Europhys. Lett.* **97**, 50011 (2012)
89. H. Davidowitz, V. Steinberg, On an analog of the Aharonov-Bohm effect in superfluid helium. *Europhys. Lett.* **38**, 297–300 (1997)
90. G. Rousseaux, R. Kofman, O. Minazzoli, The Maxwell-Lodge effect: significance of electromagnetic potentials in the classical theory. *Eur. Phys. J. D* **49**, 249–256 (2008)
91. I.L. Paiva, R. Lenny, E. Cohen, Geometric phases and the Sagnac effect: Foundational aspects and sensing applications. *Adv. Quantum Technol.* **2100121** (2022)
92. P. Bocchieri, A. Loinger, Nonexistence of the Aharonov-Bohm effect. *Nuov. Cim. A* **47**, 475–482 (1978)
93. E.G.P. Rowe, Consequences of a singular gauge transformation. *Nuov. Cim. A* **56**, 16–20 (1980)
94. A. Zeilinger, On the Aharonov-Bohm effect. *Lett. Nuov. Cim.* **25**, 333–336 (1979)
95. M. Bawin, A. Burnel, Further comments on the Aharonov-Bohm effect. *Lett. Nuov. Cim.* **27**, 4–6 (1980)
96. J.A. Mignaco, C.A. Novaes, Remarks on the possibility of nonexistence of the Aharonov-Bohm effect (ESAB effect). *Lett. Nuov. Cim.* **26**, 453–456 (1979)
97. F. Wilczek, Magnetic flux, angular momentum, and statistics. *Phys. Rev. Lett.* **48**, 1144–1146 (1982)
98. D.H. Kobe, Comment on “Magnetic flux, angular momentum, and statistics.”. *Phys. Rev. Lett.* **49**, 1592 (1982)
99. Y. Nambu, The Aharonov-Bohm problem revisited. *Nucl. Phys. B* **579**, 590–616 (2000)
100. T. Chakraborty, *Nanoscopic Quantum Rings: A New Perspective*. In: Kramer B. (eds) *Advances in Solid State Physics. Advances in Solid State Physics*, vol 43 (Springer, Berlin, 2003)
101. R. Heras, The Helmholtz theorem and retarded fields. *Eur. J. Phys.* **37**, 065204 (2016)
102. S. Coleman, *The Magnetic Monopole Fifty Years Later, in The Unity of the Fundamental Interactions*, 21–117 (Springer, Boston, 1983)
103. R. A. Millikan, Nobel prize lecture 1923. May 23 (1924)
104. G.N. Afanasiev, The scattering of charged particles on the toroidal solenoid. *J. Phys. A* **21**, 2095–2110 (1988)
105. G.N. Afanasiev, Theoretical description of Tonomura-like experiments (electron scattering on a toroidal solenoid). *Phys. Lett. A* **142**, 222–226 (1989)
106. M. Qian, Z. Gu, C. Miao, Aharonov-Bohm scattering on thin toroidal magnetic flux without toroidal shielding. *Commun. Theor. Phys.* **34**, 135–142 (2000)
107. J. Hamilton, *Aharonov-Bohm and other Cyclic Phenomena, Springer Tracts in Modern Physics*, vol. 139 (Springer, Berlin, 1997)
108. M. Ballesteros, R. Weder, High-velocity estimates for the scattering operator and Aharonov-Bohm effect in three dimensions. *Commun. Math. Phys.* **285**, 345–398 (2009)
109. M. Ballesteros, R. Weder, The Aharonov-Bohm effect and Tonomura et al. experiments: rigorous results. *J. Math. Phys.* **50**, 122108 (2009)
110. P. Roux, Scattering by a toroidal coil. *J. Phys. A* **36**, 5293–5304 (2003)
111. G.N. Afanasiev, V.M. Shilov, Numerical investigation of Tonomura experiments. *J. Phys. A* **26**, 743–750 (1993)
112. S. Weinberg, *Lectures on Quantum Mechanics* (Cambridge University Press, Cambridge, 2013)
113. J.J. Sakurai, *Modern Quantum Mechanics* (Addison-Wesley, Reading, 1994)
114. M.V. Berry, Quantal phase factors accompanying adiabatic changes. *Proc. R. Soc. Lond. A* **392**, 45–57 (1984)
115. F. Wilczek, Quantum mechanics of fractional-spin particles. *Phys. Rev. Lett.* **49**, 957–959 (1982)
116. X.G. Wen, E. Dagotto, E. Fradkin, Anyons on a torus. *Phys. Rev. B* **42**, 6110–6123 (1990)
117. R. Iengo, K. Lechner, Quantum mechanics of anyons on a torus. *Nucl. Phys. B* **346**, 551–575 (1991)
118. G.N. Afanasiev, Quantum mechanics of toroidal anyons. *J. Phys. A Math. Gen.* **24**, 2517–2528 (1991)
119. Y. Hatsugai, M. Kohmoto, Y.-S. Wu, Anyons on a torus: Braid group, Aharonov-Bohm period, and numerical study. *Phys. Rev. B* **43**, 10761–10768 (1991)