# Does the Dressing Field Method Explain the Aharonov-Bohm Effect? A Critical Analysis

Shan Gao
Research Center for Philosophy of Science and Technology,
Shanxi University, Taiyuan 030006, P. R. China
E-mail: gaoshan2017@sxu.edu.cn

April 4, 2026

**Abstract**

The dressing field method (DFM) has been proposed as a manifestly gauge-invariant reformulation of quantum electrodynamics, with the claim that it provides a physical explanation of the Aharonov-Bohm (AB) effect. In this paper, we rigorously examine this claim. We first provide a comprehensive review of the dressing field formalism, from Dirac's foundational work to contemporary developments. We then present a decisive argument demonstrating that DFM fails to explain the AB effect: before interference, the dressed field is physically identical to the bare field; during interference, the dressing factors cancel completely; and the AB phase arises solely from the bare wave function's path integrals. We conclude that DFM is a mathematically equivalent reformulation that provides no new explanatory power regarding the AB effect.

## 1 Introduction

The Aharonov-Bohm (AB) effect remains one of the most conceptually challenging phenomena in quantum physics [1]. It demonstrates that electrons are influenced by electromagnetic potentials in regions where the magnetic field vanishes, challenging classical notions of locality and the physical significance of gauge potentials. Recently, the dressing field method (DFM) has been advanced as a manifestly gauge-invariant reformulation of quantum electrodynamics. Proponents argue that by "dressing" the electron field with a nonlocal phase factor, one obtains a gauge-invariant physical electron field that can explain the AB effect. According to this view, the dressed electron is an extended object carrying a "Dirac string" that encodes the magnetic flux, thus providing a local interaction mechanism for the AB effect [6].

However, doubts have been raised about whether DFM offers genuine explanatory power or merely a mathematical reformulation [11, 12]. In particular, a no-go theorem has recently been proven to exclude all gauge-invariant dynamical explanations of the generalized AB effect [11]. In this paper, we provide a focused analysis specifically targeting DFM's claimed explanation of the AB effect.

Our argument proceeds in three steps. First, we provide a detailed review of the dressing field formalism, tracing its development from Dirac's foundational work to contemporary formulations. Second, we present a rigorous analysis showing that before interference, the dressed field is physically identical to the bare field, and during interference the dressing factors cancel completely. Third, we demonstrate that the AB phase arises solely from the bare wave function's path integrals. We conclude that DFM adds no explanatory power and is merely a mathematical reshuffling.

# 2 DFM: A Historical and Conceptual Overview

DFM has a rich intellectual history that spans seven decades, yet its recent application to the AB effect marks a notable departure from its original concerns. To evaluate whether the method can genuinely explain the AB effect, we must first understand what it was designed to do and by whom. This section provides a comprehensive review of the dressing field formalism's development, beginning with Dirac's foundational insight that gauge-invariant electron fields could be constructed by attaching a "Coulomb field" to the bare electron. We then examine how Lavelle, McMullan, and their collaborators systematized and extended Dirac's idea, applying it to problems of infrared divergences, renormalization, and confinement in both QED and QCD. Finally, we survey contemporary formulations and trace how the method came to be linked, by some authors, to the AB effect. This historical groundwork will prove essential for distinguishing between the method's legitimate technical achievements and the more questionable interpretive claims that have recently been attached to it.

## 2.1 Dirac's 1955 Proposal: Gauge-Invariant Electron Fields

The concept of gauge-invariant electron fields originated with Dirac's seminal 1955 paper "Gauge-Invariant Formulation of Quantum Electrodynamics" [7]. Dirac recognized a fundamental tension: the electron field in standard QED is gauge-dependent, yet physical electrons must be described by gauge-invariant quantities. He proposed a solution by constructing what he called a "gauge-invariant operation of creation of an electron." In his own words: "Electrodynamics is formulated so as to be manifestly invariant under general gauge transformations, through being built up entirely in terms of gauge-invariant dynamical variables" [7]. The key insight was that a gauge-invariant electron field could be constructed by attaching a "Coulomb field" to the bare electron:[1]

$$\psi^P(x) = \exp\left(-ie\int_{\Gamma_x} A_\mu(y)dy^\mu\right)\psi(x) \tag{1}$$

Dirac emphasized a crucial physical implication: "The gauge-invariant operation of creation of an electron involves the simultaneous creation of an electron and of the Coulomb field around it. The requirement of manifest gauge invariance prevents one from using the concept of an electron separated from its Coulomb field" [7]. This passage is often cited by proponents of DFM as evidence that the dressed electron is an extended, physically meaningful object.

However, Dirac himself was cautious about the interpretation. He noted that the quantization of this formulation "meets with the usual difficulties" and did not claim that it resolved conceptual puzzles like the AB effect—which had not yet been discovered. Dirac's work was primarily concerned with establishing a manifestly gauge-invariant formalism, not with providing explanations for specific quantum phenomena.

[1] Throughout this paper we set $\hbar = 1$.

## 2.2 The Lavelle-McMullan School: Dressed Charges in QED and QCD

DFM was significantly developed and extended by Martin Lavelle, David McMullan, and their collaborators in a series of papers throughout the 1990s and early 2000s [2, 3, 15]. They introduced the terminology of "dressed charges" and "dressing fields," and extended the construction from QED to QCD.

In their 1996 paper "How do constituent quarks arise in QCD?" [2], they argued: "We motivate the use of dressed charges by arguing that such objects are needed to describe, e.g., constituent quarks and, in general, physical charged states in gauge theories." The authors emphasized the infrared properties of dressed charges, demonstrating that "the one loop propagator of a relativistic dressed charge can be renormalized in the mass shell scheme with no infra-red divergence showing up."

Their comprehensive 1997 review "Constituent quarks from QCD" [15] systematized the dressing field approach. They explicitly connected the construction to Dirac's original idea: "In QED electrons must also be dressed with photons... This directly implies that the fundamental matter fields in the original Lagrangian of QED should not be identified with the asymptotic physical fields. In particular all charged fields, whether quarks, gluons, or electrons, must always carry with them a chromo-or electromagnetic cloud and only these systems - the matter and its associated dressing cloud taken together - can have any physical meaning."

The Lavelle-McMullan school made several technical advances. First, they established the relationship between dressings and gauge fixings, showing that "to every gauge fixing there corresponds a dressing" [15]. Second, they analyzed the Gribov ambiguity's implications for dressings in non-Abelian theories, concluding that "it is not possible to construct a non-perturbative asymptotic quark field" in QCD, which they proposed "is a direct proof of quark confinement." Third, they extended the construction to include moving charges and studied their infrared properties.

Crucially for our purposes, Lavelle and McMullan did not claim that DFM explains the AB effect. Their focus was on infrared structure, renormalization, and confinement. The application to the AB effect appears to be a later development.

## 2.3 Contemporary Formulations and Extensions

In recent years, DFM has been further developed and applied to various contexts. Gomes and Riello (2018) provided a "Unified geometric framework for boundary charges and particle dressings" [13], connecting the dressing construction to asymptotic symmetries and boundary charges in gauge theories and gravity. They showed that dressings can be understood geometrically in terms of field-space connections, providing a deeper mathematical foundation.

The Fröhlich-Morchio-Strocchi (FMS) mechanism, developed in the early 1980s [8, 9] and extensively studied by Maas and collaborators [16, 17], represents a related but distinct gauge-invariant approach. While the FMS mechanism addresses the BEH effect and the construction of physical states in the electroweak sector, it shares with DFM the emphasis on composite, gauge-invariant operators. However, as Maas notes, the FMS mechanism demonstrates that "the success of PT [perturbation theory] is not accidental, but rather a consequence of the structure of the standard model" [17].

More recently, Berghofer and François [5] have linked DFM to interpretive issues in gauge theories, arguing that it provides a way to "extract and interpret gauge-invariant content." (see also [4]) Berghofer's 2024 paper [6] explicitly states that DFM explains the AB effect:

> "In QED, it is actually possible to transform the path integral into the variables [of the physical photon (2.7) and the physical electron (2.8)], i.e. physical degrees of freedom. Of course, because of the Dirac string, this will no longer lead to a local Lagrangian. On the other hand, the Dirac string explains how a non-local effect like the Aharonov-Bohm effect can emerge."

This claim—that DFM explains the AB effect—is the target of our critique.

### 2.4 The Dressing Field Account of the AB Effect

The proposed explanation of the AB effect via DFM can be reconstructed as follows:

1. Gauge-Invariant Physical Fields: The bare electron field $\psi(x)$ is gauge-dependent and therefore cannot represent a physical electron. The physical electron is instead represented by the dressed field $\psi^P(x) = D(x)\psi(x)$, which is gauge-invariant.

2. Extended Nature of the Electron: The dressing factor $D(x) = \exp(-ie\int_{\Gamma_x} A)$ is nonlocal, involving an integral along a path to infinity. This means the dressed electron is an extended object, not a point particle. In Dirac's words, it carries its "Coulomb field" with it.

3. The Dirac String: The path $\Gamma_x$ functions as a "Dirac string" attached to the electron. When the electron moves, this string moves with it. The string can interact with magnetic fields even when the electron's location is field-free.

4. Topological Phase Encoding: For points on opposite sides of the solenoid, the paths to infinity necessarily have different topologies (one must go around the solenoid). This topological difference encodes the magnetic flux $\Phi$ into the relative phase of the dressed fields on opposite sides.

5. Explanation of the AB Effect: When two dressed electron wave packets from opposite sides meet at the detector, their relative phase contains the factor $e^{-ie\Phi}$ due to the topological difference in their defining paths. Thus, the AB phase shift is explained as arising from the extended, topologically nontrivial nature of the dressed electron.

This reasoning appears plausible at first glance. However, as we shall demonstrate through rigorous analysis, it fails at every crucial step.

## 3 Why DFM Fails to Explain the AB Effect

We now present our rigorous argument demonstrating that DFM fails to explain the AB effect.

### 3.1 Before Interference: The Dressed Field Reduces to the Bare Field

Consider the standard magnetic AB setup: an electron beam is split into two wave packets that travel on opposite sides of a solenoid containing magnetic flux $\Phi$. Before the packets overlap at the detector, they propagate in a region that is simply connected (the two paths can be contracted to a point without crossing the solenoid).

**Lemma 1.** *In a simply connected region where* $\mathbf{B} = 0$*, there exists a gauge transformation such that* $\mathbf{A} = 0$ *along the entire path.*

*Proof.* In a simply connected region with $\mathbf{B} = 0$, $\mathbf{A}$ is pure gauge: $\mathbf{A} = \nabla\chi$ for some function $\chi$. The gauge transformation $\psi \to e^{-ie\chi}\psi$ sets $\mathbf{A} \to \mathbf{A} - \nabla\chi = 0$. Since the region is simply connected, $\chi$ is single-valued and the transformation is globally defined. □

In this gauge, the bare wave function $\psi$ satisfies the free Schrödinger equation. Moreover, the dressing factor $D(x)$ becomes 1, because $\mathbf{A} = 0$ along the chosen path to infinity (if we also adjust the path to lie within the region where $\mathbf{A} = 0$). Therefore:

$$\psi^P(x) = D(x)\psi(x) = \psi(x) \quad \text{(in this gauge)}. \tag{2}$$

**Corollary 1.** *Before interference, the dressed field* $\psi^P$ *is physically identical to the bare field* $\psi$*. It contains no additional information about the magnetic flux* $\Phi$*.*

This is the first crucial point: dressing adds nothing before interference. Any claim that the dressed electron "carries" information about the flux is gauge-dependent and can be eliminated by a suitable gauge choice. The "Dirac string" attached to the electron is not a physical structure but a gauge artifact—it can be gauged away in simply connected regions.

## 3.2 At Interference: The Dressing Factor Cancels

At the detector point $x_f$, the two wave packets overlap. The total wave function in the dressed formalism is:

$$\Psi^P(x_f) = \psi_1^P(x_f) + \psi_2^P(x_f) = D(x_f)\psi_1(x_f) + D(x_f)\psi_2(x_f) = D(x_f)[\psi_1(x_f) + \psi_2(x_f)]. \tag{3}$$

The interference pattern depends on:

$$|\Psi^P(x_f)|^2 = |D(x_f)|^2 |\psi_1(x_f) + \psi_2(x_f)|^2 = |\psi_1(x_f) + \psi_2(x_f)|^2, \tag{4}$$

since $|D(x_f)| = 1$. The dressing factor $D(x_f)$ is a common factor that cancels completely in the probability density.

**Theorem 1.** *The dressing factor* $D(x_f)$ *does not contribute to the relative phase between the two paths.*

*Proof.* The relative phase is determined by the ratio:

$$\frac{\psi_1^P(x_f)}{\psi_2^P(x_f)} = \frac{D(x_f)\psi_1(x_f)}{D(x_f)\psi_2(x_f)} = \frac{\psi_1(x_f)}{\psi_2(x_f)}. \tag{5}$$

The dressing factors cancel exactly, leaving the ratio of bare wave functions. □

This cancellation is mathematically inevitable because $D(x_f)$ is the same for both paths—it depends only on the detector point $x_f$, not on which path the electron took to get there. The Dirac string attached to the electron at the detector point is identical for both wave packets and therefore cannot produce a relative phase.

### 3.3 Source of the AB Phase

The bare wave functions $\psi_i(x_f)$ acquire phases from propagation along their respective paths. In the region outside the solenoid ($\mathbf{B} = 0$), the propagator for a charged particle takes the form:

$$\psi_i(x_f) = \psi_0(x_f) \exp\left(-ie \int_{C_i} \mathbf{A} \cdot d\mathbf{l}\right), \tag{6}$$

where $\psi_0(x_f)$ is the free propagator (independent of the path) and the integral is along path $C_i$ from source to detector.

The relative phase between the two paths is therefore:

$$\frac{\psi_1(x_f)}{\psi_2(x_f)} = \exp\left(-ie \int_{C_1} \mathbf{A} \cdot d\mathbf{l} + ie \int_{C_2} \mathbf{A} \cdot d\mathbf{l}\right) = \exp\left(-ie \oint_{C} \mathbf{A} \cdot d\mathbf{l}\right) = e^{-ie\Phi}, \tag{7}$$

where $C = C_1 - C_2$ is the closed loop encircling the solenoid, and $\Phi$ is the enclosed magnetic flux.

**Theorem 2.** *The AB phase $e^{-ie\Phi}$ arises solely from the bare wave function's path integrals and is independent of the dressing construction.*

*Proof.* From equations (5)-(7), the interference term in the dressed formalism is:

$$\begin{aligned} \psi_1^{P*}(x_f)\psi_2^P(x_f) &= |D(x_f)|^2 \psi_1^*(x_f)\psi_2(x_f) \\ &= |\psi_0|^2 \exp\left(ie \int_{C_1} \mathbf{A} \cdot d\mathbf{l} - ie \int_{C_2} \mathbf{A} \cdot d\mathbf{l}\right) \\ &= |\psi_0|^2 e^{ie\Phi}. \end{aligned} \tag{8}$$

The dressing factor $D(x_f)$ has completely canceled, and the phase factor $e^{ie\Phi}$ comes entirely from the bare wave function's path integrals. □

### 3.4 Summary of the Argument

We have established three key facts:

1. Before interference: In a simply connected region, a gauge exists where $\mathbf{A} = 0$ along each path. In this gauge, $D(x) = 1$ and $\psi^P(x) = \psi(x)$. The dressed field contains no flux information.

2. At interference: The dressing factor $D(x_f)$ is a common factor that cancels in both the probability density and the relative phase.

3. Source of AB phase: The AB phase $e^{-ie\Phi}$ arises from the bare wave function's path integrals $\exp(-ie \int_{C_i} \mathbf{A} \cdot d\mathbf{l})$, which are present regardless of dressing.

Therefore, DFM adds nothing to the explanation of the AB effect. It is a mathematically equivalent reformulation that, despite its gauge invariance, provides no new physical insight.

## 4 Further Discussion

The central claim of this paper is that DFM, despite its mathematical elegance and gauge invariance, does not explain the AB effect. The previous section provided the core argument: before interference the dressed field reduces to the bare field, at interference the dressing cancels completely, and the AB phase arises solely from bare path integrals. In this discussion, we unpack the conceptual significance of these findings. We first examine why mathematical equivalence is often mistaken for explanatory gain, then diagnose the specific reasoning that makes the dressing field account seem plausible, address historical nuances regarding Dirac's original intentions, and finally compare our critique with analyses of other gauge-invariant approaches.

## 4.1 Mathematical Equivalence vs. Physical Explanation

DFM is mathematically equivalent to the standard formulation of QED. This equivalence can be seen by noting that:

$$\psi^P(x) = \exp\left(-ie\int_{\Gamma_x} \mathbf{A}\cdot d\mathbf{l}\right)\psi(x) \tag{9}$$

is simply a field redefinition. All physical quantities computed with $\psi^P$ are identical to those computed with $\psi$. As we have shown, the AB phase in particular is unchanged.

The recent no-go theorem strengthens this conclusion: no gauge-invariant quantity can dynamically explain the AB effect, because during the phase-generating period all gauge-invariant observables of the electron remain identical to those in a zero-flux experiment [11]. DFM, despite its gauge invariance, falls under this theorem.

## 4.2 The Source of the Misinterpretation

Why might one think DFM explains the AB effect? The reasoning appears to be that the dressed field $\psi^P$ is gauge-invariant and thus "physical", and the dressing factor $D(x)$ contains a line integral that can pick up the flux $\Phi$, and therefore, the dressed electron "carries" the flux information and can interact locally with the magnetic field.

However, this reasoning is fallacious for several reasons. First, the line integral in $D(x)$ is along a path to infinity, not along the electron's path. The flux information in $D(x)$ is not dynamically acquired during propagation; it is artificially inserted by the choice of path. Second, the "Dirac string" attached to the electron is a gauge artifact, not a physical structure. In simply connected regions, it can be completely gauged away. Third, at the detector, the dressing factor $D(x_f)$ is common to both paths and cancels completely in the interference pattern. The flux information it supposedly carries does not contribute to the relative phase. Finally, the actual AB phase comes from the path integrals along $C_1$ and $C_2$, which are present in the bare wave functions and are completely independent of the dressing construction.

DFM thus commits what philosophers call the "fallacy of misplaced concreteness": it reifies a mathematical construction (the path-dependent phase factor) and mistakes it for a physical mechanism.

## 4.3 Dirac's Caution and Subsequent Developments

There is a certain historical nuance worth noting regarding claims that DFM explains the AB effect. Dirac himself, in his 1955 paper, was careful not to over-interpret his construction. He noted that his gauge-invariant formulation "meets with the usual difficulties" of quantization and did not claim that it resolved any conceptual puzzles. Moreover, the

AB effect was not discovered until 1959, four years after Dirac's paper, so he naturally did not address it.

Dirac's emphasis was on the fact that "the gauge-invariant operation of creation of an electron involves the simultaneous creation of an electron and of the Coulomb field around it." This is a statement about the construction of physical states in a manifestly gauge-invariant formalism. It highlights that in such a description, one cannot separate the electron from its accompanying field—but this is a feature of the mathematical representation, not necessarily a claim about the mechanism of nonlocal effects.

The subsequent development of DFM by Lavelle, McMullan, and others focused primarily on technical issues such as infrared divergences, renormalization, and the structure of physical charges in gauge theories. Their work did not emphasize the AB effect as a primary motivation or application. The connection between DFM and the AB effect appears to be a more recent interpretive suggestion, which—as we have argued—faces significant conceptual difficulties upon closer examination.

### 4.4 Comparison with Other Gauge-Invariant Approaches

Our critique extends beyond DFM to other gauge-invariant approaches. For example, the holonomy interpretation [14] faces similar problems: holonomies are nonlocal objects that encode the flux, but they do not explain how the phase difference arises dynamically. It has been shown that the holonomy interpretation fails to account for the time-dependent AB effect, where the flux varies during the electron's flight [11, 12].

The unitary gauge approach, recently analyzed in [10], suffers from analogous issues. While it transforms the Schrödinger equation into the Madelung equations expressed in local gauge-invariant variables, it requires an additional quantization condition that is inherently nonlocal. The unitary gauge does not eliminate nonlocality; it merely relocates it to a quantization condition.

The fundamental issue is that the AB effect is inherently topological and nonlocal. Any attempt to explain it through local, gauge-invariant quantities must fail because the effect itself is not locally generated in gauge-invariant formulations. As we have demonstrated, DFM merely relocates the nonlocality from the interaction term to the definition of the field, without providing any dynamical mechanism.

## 5 Conclusion

We have presented a rigorous analysis demonstrating that DFM does not explain the AB effect. Our argument rests on three simple but decisive points:

1. Before interference, the dressed field is physically identical to the bare field and contains no flux information.

2. During interference, the dressing factors cancel completely and do not contribute to the relative phase.

3. The AB phase arises solely from the bare wave function's path integrals $\exp(-ie\int \mathbf{A}\cdot d\mathbf{l})$.

DFM is mathematically equivalent to standard QED but provides no new explanatory power. Our historical review shows that the originators of the method—Dirac, Lavelle, McMullan, and their collaborators—were primarily concerned with technical issues like gauge invariance, infrared divergences, and renormalization. They did not claim that the method explains the AB effect. That claim is a recent innovation that does not withstand rigorous scrutiny.

Our analysis supports the broader conclusion that no gauge-invariant explanation of the AB effect is possible [11]. The AB effect demonstrates the physical reality of gauge potentials, which cannot be reduced to gauge-invariant quantities without losing explanatory content. Future work should focus on understanding the ontological implications of this result: if gauge potentials are physically real despite their gauge dependence, what does this tell us about the nature of gauge theories? DFM, while mathematically interesting, offers no answers to these fundamental questions.

# References


[1] Y. Aharonov and D. Bohm, "Significance of Electromagnetic Potentials in the Quantum Theory," Physical Review 115, 485-491 (1959).

[2] E. Bagan, M. Lavelle, D. McMullan, B. Fiol, and N. Roy, "How do constituent quarks arise in QCD? Perturbation theory and the infra-red," Nuclear Physics B - Proceedings Supplements 54, 208-212 (1997) [arXiv:hep-ph/9609333].

[3] E. Bagan, M. Lavelle, and D. McMullan, "Charges from dressed matter: Construction," Annals of Physics 282, 471-501 (2000).

[4] P. Berghofer, J. François, S. Friederich, H. Gomes, G. Hetzroni, A. Maas, and R. Sondenheimer, *Gauge Symmetries, Symmetry Breaking, and Gauge-Invariant Approaches*, Cambridge Elements: Elements in the Foundations of Contemporary Physics, Cambridge University Press (2023). Chapter 5.

[5] P. Berghofer and J. François, "Dressing vs. Fixing: On How to Extract and Interpret Gauge-Invariant Content," Foundations of Physics 54, 72 (2024) [arXiv:2404.18582].

[6] P. Berghofer, "Motivating Gauge-Invariant Approaches to Particle Physics," arXiv:2408.11705 [physics.hist-ph] (2024).

[7] P. A. M. Dirac, "Gauge-Invariant Formulation of Quantum Electrodynamics," Canadian Journal of Physics 33, 650-660 (1955).

[8] J. Fröhlich, G. Morchio, and F. Strocchi, "Higgs phenomenon without symmetry breaking order parameter," Physics Letters B 97, 249-252 (1980).

[9] J. Fröhlich, G. Morchio, and F. Strocchi, "Higgs phenomenon without symmetry breaking order parameter," Nuclear Physics B 190, 553-582 (1981).

[10] S. Gao, "Can Unitary Gauge Provide a Local and Gauge-Invariant Explanation of the Aharonov-Bohm Effect?" https://philsci-archive.pitt.edu/24678/ (2025).

[11] S. Gao, "Impossibility of Gauge-Invariant Explanations of the Aharonov–Bohm Effect," https://philsci-archive.pitt.edu/28224/ (2026).

[12] S. Gao, "Revisiting Modular Momentum in the Aharonov-Bohm Effect: A Resolution of Aharonov's Paradox," https://philsci-archive.pitt.edu/26613/ (2025).

[13] H. Gomes and A. Riello, "Unified geometric framework for boundary charges and particle dressings," Physical Review D 98, 025013 (2018).

[14] R. Healey, *Gauging What's Real: The Conceptual Foundations of Contemporary Gauge Theories*, Oxford University Press (2007).

[15] M. Lavelle and D. McMullan, "Constituent quarks from QCD," Physics Reports 279, 1-65 (1997).

[16] A. Maas, "Brout-Englert-Higgs physics: From foundations to phenomenology," Progress in Particle and Nuclear Physics 106, 132-209 (2019).

[17] A. Maas and P. Schreiner, "The manifestly gauge-invariant spectrum of the minimal supersymmetric standard model," Annals of Physics 479, 170050 (2025) [arXiv:2307.10282].