American Journal of Physics

NOTES AND DISCUSSIONS | JUNE 01 2024

Mutual inductance between an infinite solenoid and a surrounding loop—A paradox resolved FREE

Ian C Malcolm

Am. J. Phys. 92, 473–475 (2024)
https://doi.org/10.1119/5.0172133

Mutual inductance between an infinite solenoid and a surrounding loop—A paradox resolved

Ian C Malcolm[a)]
School of Physical Sciences, The Open University, Walton Hall, Milton Keynes MK7 6AA, United Kingdom

(Received 13 August 2023; accepted 17 March 2024)

Students may calculate the mutual inductance between an infinite solenoid and a surrounding circular loop using an incorrect method that yields the correct answer, causing frustration and confusion. A physical explanation for the agreement of the two methods is provided that is based on the translational symmetry of the solenoid. As a consequence, the conclusion generalizes to infinite uniform solenoids of arbitrary cross section.
https://doi.org/10.1119/5.0172133

I. INTRODUCTION

Misconceptions or errors in physics understanding often lead to paradoxes. This is particularly true in electromagnetism where errors appear to be of two types, which I will call "incomplete system" and "unjustified simplification." Examples of the former can be found in the paradox of a force on a stationary charge due to a current-carrying wire,[1] where the paradox arises from neglect of the part of the system composed of the source and sink of current, and in the paradox of non-conservation of angular momentum,[2] which arises from neglect of the angular momentum of electromagnetic fields. An example of the second type is the paradox of differing readings on different voltmeters connected between the same points of a circuit.[3] The simplification of equal values for the relevant line integrals, which leads to equal readings, is unjustified in a multiply connected region. In this note, I discuss another example of the latter type of error, the paradox here being that, rather than giving the wrong answer, the unjustified simplification gives the correct one.

II. THE PROBLEM

In the undergraduate electromagnetism curriculum, as an application of Faraday's law, students are often asked to find the mutual inductance M_{12} of the axially symmetric system shown in Fig. 1, where a varying current, I_2, in the outer circular loop (the primary) induces an emf in the inner, co-axial and closely wound circular solenoid (the secondary) composed of n circular loops per unit length. The center of the outer loop defines the origin, and we assume that the radius R_2 of the loop is greater than the radius R_1 of the solenoid.

In the author's experience, students attempting a solution fall into two categories: those who use Method A—solving the problem as it stands but by incorrectly assuming that the field due to the loop takes its on-axis value everywhere inside the solenoid, and those who use Method B—solving the reciprocal problem with the solenoid as the primary and the loop as the secondary.

Solvers using Method A start with the on-axis field of the loop,

$$b_{2z}(z) = \frac{\mu_0 {R_2}^2 I_2}{2\left(z^2 + {R_2}^2\right)^{3/2}}. \tag{1}$$

(I use the lowercase b to represent the field due to a source that has only a single turn; this notation will be useful later in the paper.) They then calculate the flux through the solenoid due to the loop with the assumption that the field maintains its on-axis value everywhere inside the solenoid,

$$\Phi_{12} = \pi {R_1}^2 \int_{-\infty}^{\infty} b_{2z}(z) n \mathrm{d}z = \mu_0 n I_2 \pi {R_1}^2, \tag{2}$$

leading to the Method A result

$$M_{12} \equiv \frac{\Phi_{12}}{I_2} = \mu_0 n \pi {R_1}^2. \tag{3}$$

This is the correct result. Nevertheless, Method A is not justified because of its assumption that the field of the loop (or at least its axial component) is independent of distance from the z-axis.

Method B considers the reciprocal system where a current I_1 in the solenoid produces a (constant) field $B_1 = \mu_0 n I_1$ inside the solenoid (here the capital B denotes the field due to an infinite stack of turns), producing a flux $\Phi_{21} = \mu_0 n I_1 \pi {R_1}^2$ through the loop, and leading to the mutual inductance $M_{21} = \mu_0 n \pi {R_1}^2$. Using the principle of reciprocity[4] one obtains $M_{12} = M_{21} = \mu_0 n \pi {R_1}^2$.

There is no doubt that Method B is correct, based as it is on sound fundamental principles. However, Method A's result is the same as Method B's result. Imagine a Method A solver's surprise, dismay, and confusion when told that their answer is correct but that their method is not justified. They might reasonably ask, "Why am I wrong?" (Awarding marks here presents an interesting challenge for the instructor.)

As far as the author is aware, this apparent paradox has never been noted in the treatment of mutual inductance in the literature.

III. IS THERE A PARADOX?

On being reminded of the limited applicability of Eq. (2), the student's surprise could be even greater. At least in the

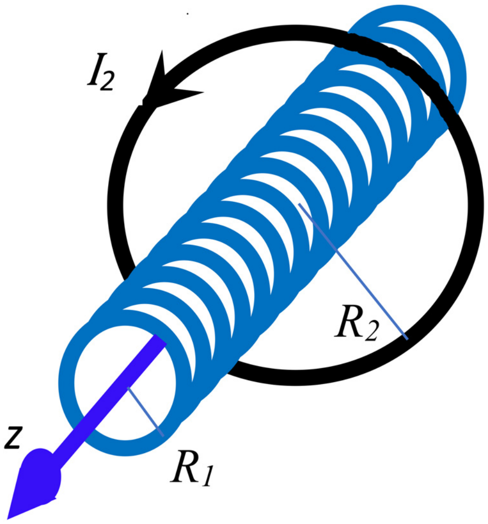

Fig. 1. A circular solenoid of radius R_1 with its axis along z passes through a loop of radius R_2 carrying a current I_2.

plane of the loop, the on-axis field is less than the off-axis field, so the student might expect that use of the on-axis field should underestimate M_{12}. This question can be addressed by a first-principles calculation using the Biot–Savart law in cylindrical coordinates with origin at the center of the loop. The only relevant field component within the loop is the axial one $b_{2z}(\rho, z)$, and it is given as a function of ρ and z by Dasgupta,[5]

$$b_{2z}(\rho, z) = \frac{\mu_0 R_2 I_2}{2\pi} \int_0^{\pi} \frac{(R_2 - \rho\cos\alpha)}{\left[\left(R_2{}^2 + \rho^2 + z^2\right) - 2R_2\rho\cos\alpha\right]^{3/2}} \mathrm{d}\alpha, \tag{4}$$

where α is the azimuthal angle between the radial vector ρ to the field point in the solenoid and a radial vector to an arbitrary source point on the loop. This integral can be expressed only in terms of elliptic integrals,[5] but a numerical evaluation assuming $R_2 = 2R_1$ is displayed in Fig. 2.

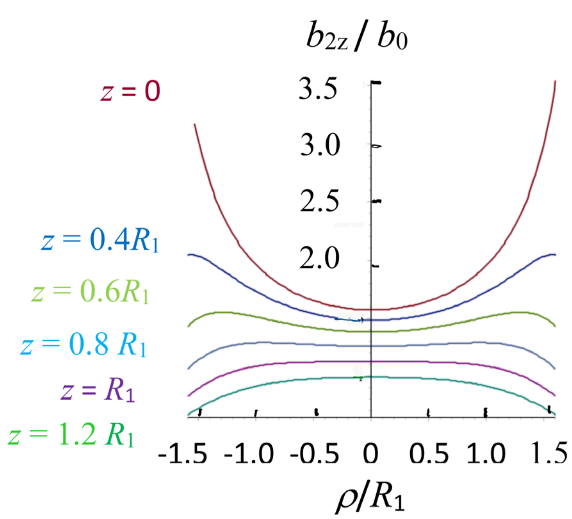

Fig. 2. Axial field component b_{2z} (normalized to $b_0 = \mu_0 I_2 / 2\pi R_1$) due to a current loop of radius $R_2 = 2R_1$, plotted as a function of radial coordinate ρ, for various axial distances z from the origin.

For $z = 0$, the radial minimum in b_{2z} occurs on the axis, with the value at the minimum decreasing with z as expected. However, this position of the radial minimum only persists out to an axial distance $|z| \approx 0.8R_1$. Beyond this distance, the on-axis value of b_{2z} becomes a maximum. This observation should reduce the student's surprise and eliminate the paradox, since using the axial field underestimates the flux for $|z| < 0.8a$ but overestimates it for larger values of $|z|$, making it possible for the axial field to correctly predict the flux. The careful calculation of the flux requires a quadruple integral[5] and yields $M_{12} = \mu_0 n\pi R_1{}^2$ in agreement with Eqs. (2) and (3) (see the supplementary material).[6]

IV. FUNDAMENTAL REASON FOR THE AGREEMENT

While the agreement between Methods A and B is explained by this integration, one cannot avoid the feeling that there must be a deeper reason. An explanation is provided here, which applies even if the cross sections are non-circular, as long as the solenoid has translational symmetry and the loop lies in a plane perpendicular to the axis of translational symmetry (taken to lie within the solenoid). Under these conditions is is straightforward to show using Ampere's law that the magnetic field inside the solenoid is uniform and axial.[7,8]

Figure 3 shows a generalized version of the problem. Taking the z-axis to be parallel to the direction of translational symmetry of the solenoid and to lie within the cross section of the solenoid with origin in the plane of the loop, the flux through the solenoid due to the loop field is

$$\Phi_{12} = \int_{\rho,\varphi} \mathrm{d}S_1 \int_{-\infty}^{\infty} n b_{2z}(\rho,\ \phi,\ z)\mathrm{d}z, \tag{5}$$

where, as in Eq. (4), b_{2z} represents the z-component of the magnetic field at each point (ρ,ϕ,z) inside the solenoid, due to a current I_2 in the loop, and where S_1 is the solenoid's cross-sectional area normal to the z-axis. Normally, one might first integrate over the area of each coil ($\mathrm{d}S_1$) and then over all the loops in the solenoid ($\mathrm{d}z$), but the order of integration can be changed because of the translational symmetry: the limits of ρ and ϕ are independent of z.

Now, consider a new situation, shown in Fig. 4. In this case, the loop has been extended into a (fictitious) solenoid

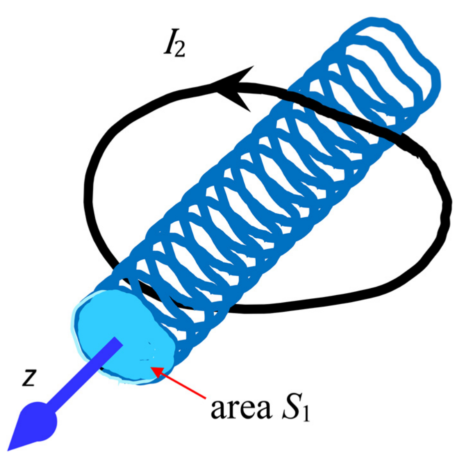

Fig. 3. System related to Eq. (5), consisting of an arbitrary loop surrounding an arbitrary but uniform solenoid, carrying current I_2.

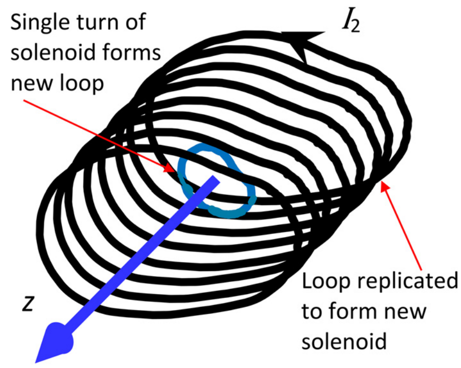

Fig. 4. System relating to Eq. (6), dual to that of Fig. 3: a single turn of the original solenoid inside a uniform, infinite solenoid carrying a current I_2, formed by replicating the original loop along the z-axis.

with n turns per unit length, and the solenoid is reduced to a (fictitious) single turn loop whose plane contains the origin. We know from Ampère's law and the symmetry of the solenoid that the axial field component (the only component) inside this fictitious solenoid is independent of the position (ρ,ϕ,z) and is equal to $B_2 = \mu_0 n I_2$, where I_2 is the current through each of the loops of the fictitious solenoid. However, we can also write this field as an integral,

$$B_2(\rho,\,\phi) = \int_{-\infty}^{\infty} n b_{2z}(\rho,\,\phi,\,z)\mathrm{d}z, \tag{6}$$

where b_{2z} still refers to the field at a point $(\rho,\,\phi)$ due to the current I_2 in one loop (one turn of the fictitious solenoid). Here, the integration variable z represents the distance from each turn of the fictitious solenoid to the plane of the loop, and the integral over z takes into account the fact that the solenoid consists of an infinite stack of loops at different positions z.

Here is the key insight. Even though they represent two different situations, the integral in Eq. (6) is mathematically identical to the interior integral in Eq. (5). Since we know from Ampère's law that the integral in Eq. (6) yields $B_2 = \mu_0 n I_2$, independent of the position $(\rho,\,\phi)$ inside the loop, it follows that the interior integral in Eq. (5) is also equal to $\mu_0 n I_2$, independent of $(\rho,\,\phi)$. That makes the full integral in Eq. (5) trivial to compute: the flux $\Phi_{12} = \mu_0 n I_2 S_1$.

Because the integrated effect of the single-loop field in Eq. (6) has a form independent of ρ and ϕ, any combination of the coordinates lying within the cross section of the solenoid could be used to evaluate the integral in Eq. (5). The simplest choice is $\rho = 0$ (i.e., using the on-axis field of the original loop, replacing $b_{2z}(\rho,\,\phi,\,z)$ in Eq. (5) by $b_{2z}(z)$ as given by Eq. (1)), leading to a trivial integration over ρ and ϕ. This was the implicit choice of students using Method A, and it works because of the translational symmetry of the original solenoid. Generally, given the field distribution as a function of z along any axis parallel to the axis of translational symmetry, Eq. (5) can be evaluated to give the correct flux.

This result is applied to the example of a square loop in supplement 2 of the supplementary material. The proviso that the solenoid lies within the loop for the success of Method A translates into $R_1 < R_2$. If it were otherwise, then as pointed out in supplement 3 of the supplementary material for the example of the analogous scenario of Fig. 1, it would not be clear whether the integrated on-axis loop field should be multiplied by $\pi {R_1}^2$ or $\pi {R_2}^2$ to obtain Φ_{12}.

V. CONCLUSION

What do we conclude from all this? Including the first-principles calculation, we now have three methods at our disposal, all giving the same correct answer. Which of these should be adopted? Although the first-principles calculation is sound and yields the correct answer, we immediately rule it out because it involves a quadruple integration. In deciding between Method A and the reciprocity method (B), the latter wins "hands down" because, notwithstanding its correct result, Method A contains an "unjustified simplification" type of misconception. The result of this method can only be justified and trusted with the hindsight of the reciprocity result.

However, there is a more important principle at stake if Method A is accepted. In the example of the configuration dual to Fig. 1, where the loop lies within the solenoid, supplement 3 of the supplementary material shows that this method is ambiguous with regard to which cross-sectional area should be used to obtain Φ_{12}. Allowing the use of Method A to go unchallenged might lead its adherents to the erroneous belief that their method is always applicable, but, as we have seen, this cannot always be guaranteed.

ACKNOWLEDGMENTS

The author acknowledges several anonymous students whose assignment solutions inspired this article. He also acknowledges the valuable comments and suggestions of the anonymous reviewers.

AUTHOR DECLARATIONS

Conflict of Interest

The author has no conflicts of interest to disclose.

[a)]ORCID: 0000-0002-0997-0818.

[1]Denise C. Gabuzda, "Magnetic force due to a current-carrying wire," Am. J. Phys. **55**(5), 420–422 (1987).

[2]Emerson M. Pugh and George E. Pugh, "Physical significance of the poynting vector in static fields," Am. J. Phys. **35**(2), 153–156 (1967).

[3]Robert H. Romer, "What do voltmeters measure?: Faraday's law in a multiply-connected region," Am. J. Phys. **50**(12), 1089–1093 (1982).

[4]E. M. Purcell and D. J. Morin, *Electricity and Magnetism*, 3rd ed. (Cambridge U. P., New York, 2013), pp. 362–364.

[5]B. B. Dasgupta, "Magnetic field due to a solenoid," Am. J. Phys. **52**(3), 258–258 (1984).

[6]See the supplementary material online for Supplement 1: the original problem of Sec. II (a solenoid radius R_1, within a loop radius R_2) can be solved exactly from first principles by performing a quadruple integration giving ϕ_{12} leading to M_{12}. Details of this are given in this supplement. Supplement 2: works through unjustified Method A showing that it also gives the correct result of M when applied to a square loop. Supplement 3: performs a similar calculation to Supplement 1, but now with the loop within the solenoid.

[7]D. J. Griffiths, *Introduction to Electrodynamics*, 4th ed. (Cambridge U. P., Cambridge, 2017), pp. 235–237.

[8]W. Hauser, "Magnetic field due to a solenoid," Am. J. Phys. **53**, 774 (1985).