The transient magnetic field outside an infinite solenoid

Raymond J. Protheroe and Don Koks
Department of Physics and Mathematical Physics, The University of Adelaide, Adelaide, Australia 5005

(Received 14 August 1995; accepted 27 February 1996)

The electromotive force (emf) in a loop outside an infinite solenoid with changing current is usually calculated using the vector potential because the magnetic field outside an infinite solenoid is supposed to be zero. However, the magnetic field will only be zero for steady currents. A change in the applied voltage will give rise to a change in the current, which will propagate along the solenoid in the same way as a wave on a transmission line. This gives rise to a transient magnetic field outside the solenoid. It is quite possible to calculate this transient magnetic field and use it in Faraday's law to calculate the emf directly without using the vector potential. In practice, it is usually simpler to use the vector potential. However, care should be taken to ensure that students are not given the impression that there is no magnetic field and that it is the vector potential that acts on charges in the loop. We give examples of the magnetic field configuration outside an infinite solenoid for a steplike change in driving voltage and for an ac driving voltage.

I. INDUCED ELECTROMOTIVE FORCE IN A LOOP AROUND A SOLENOID

This paper concerns an alternative way of looking at and deriving a result which is often explained in physics texts by invoking what is to us an unnecessary hypothesis. The situation may be described as follows. Suppose we have an infinitely long solenoid along the z axis, carrying some current. Around the outside of this solenoid, coaxial with it, is a single loop of wire. Now it is known from Faraday's law that while the current in the solenoid is changing an emf will be induced in the loop.

The commonly asked question is "We know that charges in the loop move in response to an emf produced by a changing magnetic field; but since there is no magnetic field outside the solenoid, how do the charges in the loop know they must move?" Conventional wisdom has it that there can be no **B** field outside the loop while there certainly will be a vector potential **A**, and so it must be this vector potential which is giving rise to the pulse of current. This is sometimes stated as proof that in this sense the **A** field is more real than the **B** field.[1] Recently, the question of whether the emf can be related to the integrated effect of the acceleration fields of the moving electrons in the solenoid, if proper account is taken of retardation effects, has also been asked.[2]

In this paper we show that the usual results can be obtained by a careful consideration of what happens when the current is changing, without reference to the vector potential **A** and without use of retarded potentials. Thus we conclude that, classically at least, there is no reason to assume **B** is any less real than **A**.[3]

It should be noted that our calculation is model dependent and confined to a solenoid; we have not treated any more complicated arrangements. However our calculation relies on the fact that in electromagnetism, a change to the current or charge distribution will inevitably lead to the electric and magnetic fields adjusting themselves to the new configuration, and in so doing there will always be a transient effect. Here we are describing just such an effect for the solenoid.

II. THE MODEL USED IN THIS PAPER

Later in this paper we will consider the situation discussed in Sec. I, but for now we would like to be more realistic and ask how we could carry out such an experiment in practice. A solenoid is too simple a system in itself as we also need a return wire to close the circuit so that a current can flow. When the voltage applied at one end of the solenoid is changed, the resulting change in current propagates along the solenoid-return wire combination in the same way as a wave on a transmission line. To calculate the speed of the wave, we need a simple configuration of solenoid and return wire. We can produce such a configuration by making the return wire into another solenoid, coaxial with the first and of slightly smaller diameter. That is, we have two solenoids, one fitting inside the other, carrying current in opposite directions. At one end of this double solenoid pair is a variable voltage supply while the other end is at infinity or, in practice, a long way away with a terminating impedance equal to the characteristic impedance, Z_0, to ensure no waves are reflected.

This model is useful because we can easily solve for the speed of the waves produced as we alter the driving voltage, and we will show that this speed is much less than the speed of light; this means that we need not use the full treatment of retarded times in our analysis, which is a big simplification. Once this has been shown, we would argue that this applies also to the case of a single solenoid, i.e., the change in current propagates along the solenoid at some speed $v \ll c$, and that we can use this change in current to calculate the transient magnetic field.

For a lossless transmission line the wave speed and characteristic impedance are

$$v=(LC)^{-1/2}, \qquad (2.1)$$

$$Z_0=(L/C)^{1/2}, \qquad (2.2)$$

where L, C are the inductance per unit length and capacitance per unit length.[4] Suppose the two solenoids each carry current i through n turns per unit length, the inner solenoid having diameter a and the outer solenoid only slightly larger at $a+\Delta a$; then we can calculate these quantities as follows.

First, within the inner solenoid there will be no magnetic field, since the contributions from each solenoid cancel there. So the field is constrained to lie between the inner and outer solenoids, and just equals the field due to the outer one there. In that case the flux present is

$$\Phi \approx \mu_0 i n 2\pi a \Delta a. \tag{2.3}$$

The inductance per unit length is

$$L = n\Phi / i \approx \mu_0 n^2 2\pi a \Delta a. \tag{2.4}$$

To obtain the capacitance we consider each solenoid to be a cylindrical conductor. For $\Delta a \ll a$ we can approximate this by a parallel plate capacitor of area $2\pi a$ per unit length and separation Δa giving a capacitance per unit length of

$$C = \varepsilon_0 2\pi a / \Delta a. \tag{2.5}$$

Hence the wave speed and characteristic impedance become

$$v \approx c/(2\pi a n), \tag{2.6}$$

$$Z_0 \approx n \Delta a \mu_0 c. \tag{2.7}$$

Such a wave speed is well below the speed of light, which means that for all practical purposes we do not need a treatment of retarded times. Bearing this major simplification in mind, we now proceed to analyze the double solenoid in terms of its two components, obtaining our results by superposing the results for each solenoid.

We start with a single solenoid. Let us picture the situation with a circular loop of radius b in the x–y plane, with an infinitely long solenoid of radius a that extends along the z axis, and consider an increase in the current from I to $I+\Delta I$. Our argument hinges on the fact that *the current in the solenoid cannot change simultaneously along its length.* We suppose there is a wave of current ΔI which is moving from $z=-\infty$ toward increasing z such that behind the wave front the current is $I+\Delta I$ and ahead of the wave the solenoid carries its old current I. For simplicity we suppose that this wave moves with a sharp edge like a step function. Now because the **B** field outside a solenoid with constant current is zero, we can disregard the ever-present current I in the infinite solenoid when calculating emfs, and further simplify the situation by examining only a semi-infinite solenoid, extending from $z=-\infty$ (where the variable voltage supply is located) to $z=z_0$ (the edge of the wave), which carries current ΔI. Our analysis is concerned only with this semi-infinite solenoid; the original one carrying constant current I plays no role and so does not enter into our calculations.

III. MAGNETIC FIELD OF THE SEMI-INFINITE SOLENOID

For the purpose of illustration, we would like to plot some magnetic field lines, and for this we must calculate the magnetic field of a semi-infinite solenoid. We approximate this by the field of a series of single coils of radius a spaced equally apart between $z=-30a$ and $z=0$ each carrying the same current, plus a series of coils spaced logarithmically apart between $z=-30a$ and $z=-\infty$ each carrying a current proportional to the spacing between coils. We then use a Runge–Kutta scheme with adaptive step size[5] to trace the **B** field lines. We have done this in Fig. 1 for a solenoid made up of current loops of radius $a=1$ and separation $0.01a$. We plot eight field lines emerging from the solenoid with $z_0=0$. The field lines radiate outwards, becoming isotropic at large distances, so that at distances large compared to its radius the end of the solenoid looks just like a magnetic monopole. This is also noted by Jackson[6] who discusses the field emerging from a semi-infinite solenoid in connection with magnetic monopoles.

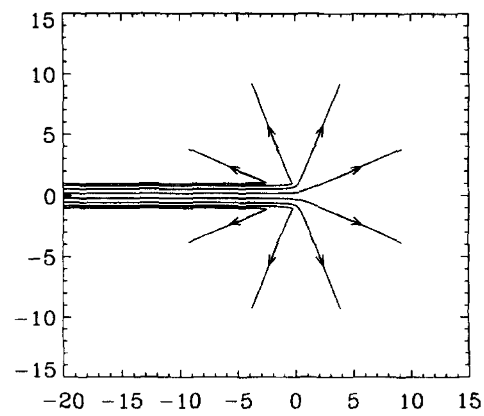

Fig. 1. Representative magnetic field lines of a semi-infinite solenoid with $a=1$ and the end located at the origin.

For a semi-infinite double solenoid, we could superpose the fields of the two solenoids which are of slightly differing strengths and opposite signs. Alternatively we can consider this as a boundary value problem in magnetostatics. The semi-infinite double solenoid can be approximated by a semi-infinite hollow tube of magnetic material with constant magnetization

$$M = \Phi_{\rm tot} / (\mu_0 2\pi a \Delta a) = nI \tag{3.1}$$

directed along its axis, where

$$\Phi_{\rm tot} = \mu_0 n I 2\pi a \Delta a \tag{3.2}$$

is the net flux through the double solenoid. Thus the magnetic field outside of the magnetic material, $\mathbf{B}_{\rm out}$, is identical to that of a ring of magnetic charge $\Phi_{\rm tot}/\mu_0$ of radius a located at the end of the tube. The electrostatic version of this problem is discussed by Jackson.[7] Inside the magnetic material the magnetic field is $\mathbf{B}_{\rm out} + \mu_0 \mathbf{M}$. The magnetic field lines are shown in Fig. 2 and again, at large distances, the field looks like that of a monopole.

Here is the important point: *as the semi-infinite solenoid advances along the z-axis, the field lines due to this open end*

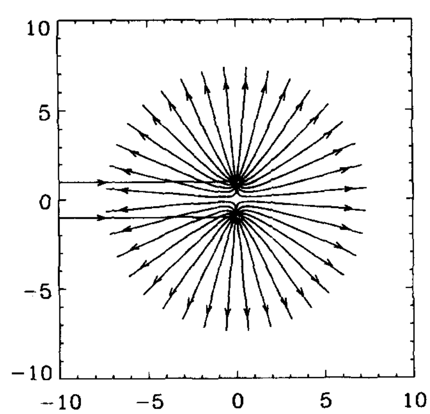

Fig. 2. Magnetic field of a semi-infinite double solenoid with $a=1$ and $\Delta a \to 0$, with the end located at the origin.

cut the loop, and in so doing they produce an emf which drives the electrons around the loop. As the current in the original infinite solenoid increases from I to $I+\Delta I$, our model has the semi-infinite solenoid's monopole-like end passing from $z_0=-\infty$ to $z_0=\infty$, and so the total increase in flux $\Delta\Phi$ through the loop is just the net flux emanating from this end, which for a single solenoid carrying current ΔI with n turns per unit length and cross sectional area πa^2 is well known to be $\mu_0 n\Delta I\pi a^2$, or $\mu_0 n\Delta I 2\pi a\Delta a$ for the semi-infinite double solenoid. Of course this is the same result as the usual one, which is normally derived for the original solenoid by noting that the total flux through the loop, with **B** wholly confined inside the solenoid, has nevertheless changed by

$$\Delta\Phi=\Phi_{\text{final}}-\Phi_{\text{initial}}=\mu_0 n(I+\Delta I)\pi a^2-\mu_0 nI\pi a^2 = \mu_0 n\Delta I\pi a^2. \quad (3.3)$$

We now see that there really is a **B** field which cuts the loop, and in so doing produces the emf. This is to be expected: after all, a changing current should radiate, so we should not be surprised to see the static picture of $\mathbf{B}=0$ outside the solenoid break down. There is *expected* to be a transient pulse of magnetic field which is cut by the loop, producing an emf in the process.

IV. AN EXPRESSION FOR THE ELECTROMOTIVE FORCE

As further illustration, we shall now calculate the expression for the emf induced in the surrounding loop. Again we do this calculation only for the single solenoid case and superpose to obtain the answer for the double solenoid. We have used two methods. It turns out that both of them assume the loop radius b is much larger than the solenoid radius a; we have assumed this purely to make the mathematics easier.

To calculate the emf we consider the semi-infinite solenoid as it moves toward the direction of increasing z. Suppose the wave of increasing current moves along the solenoid with some speed v calculated in (2.6), so that the end of the solenoid is located at

$$z_0(t)=vt. \quad (4.1)$$

The emf equals minus the rate of change of flux through the loop; so we need first to calculate this flux due to the semi-infinite solenoid as a function of time.

A. The solid angle approach

Suppose the loop subtends some solid angle $\Omega(\theta)$ as seen by the end of the solenoid, as shown in Fig. 3. If $b\gg a$, the field lines will be isotropic so that we can make use of the symmetry as follows. If the total flux through the solenoid is $\Phi_{\text{tot}}\equiv\mu_0 n\Delta I\pi a^2$, then the flux enclosed by the loop is given by

$$\Phi(t)=\Phi_{\text{tot}}\Omega(\theta)/(4\pi)=(\Phi_{\text{tot}}/2)[1-\cos\theta(t)] = (\Phi_{\text{tot}}/2)[1+vt(b^2+v^2t^2)^{-1/2}]. \quad (4.2)$$

It follows that the total increase in flux is

$$\Phi_{\text{final}}-\Phi_{\text{initial}}=\Phi(\infty)-\Phi(-\infty)=\Phi_{\text{tot}} \quad (4.3)$$

as expected.

The emf is given by

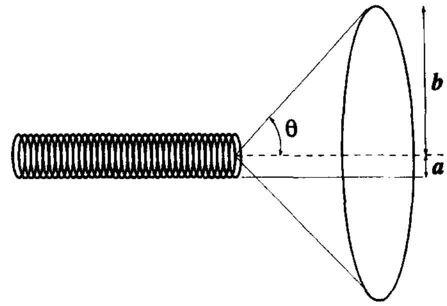

Fig. 3. Semi-infinite solenoid and loop.

$$\mathcal{E}=-\Phi'(t)=-(\Phi_{\text{tot}}/2)b^2v(b^2+v^2t^2)^{-3/2}. \quad (4.4)$$

As expected, the emf reaches a peak at $t=0$ and is negative, showing Lenz's law in action. The emf for the double solenoid case is of course obtained by modifying Φ_{tot}:

$$\Phi_{\text{tot}}\rightarrow\mu_0 n\Delta I\pi[(a+\Delta a)^2-a^2]\simeq\mu_0 n\Delta I 2\pi a\Delta a. \quad (4.5)$$

For a double solenoid, plots of the magnitude of the emf from a numerical calculation are shown in Fig. 4 for various radii, and are compared with the approximate results of Eq. (4.4).

B. Magnetic dipole approach

We can calculate the emf from first principles, and the fact that this calculation will be seen to agree with the above method supports the approach of that method. We need to obtain the flux $\Phi(t)$ through the loop, as the solenoid's end moves from $z_0=-\infty$ to $z_0=\infty$. What we shall do is consider the semi-infinite solenoid at some time t, and slice it up infinitesimally. Each slice can be thought of as an infinitesimal magnetic dipole with dipole moment

$$d\mathbf{m}=n\Delta I\pi a^2 d\mathbf{z}. \quad (4.6)$$

The magnetic field at some point in space connected to the dipole by an outward vector $\mathbf{r}$ is given by[8]

$$d\mathbf{B}=\nabla\times d\mathbf{A};\quad d\mathbf{A}\simeq\mu_0/(4\pi)d\mathbf{m}\times\mathbf{r}/r^3. \quad (4.7)$$

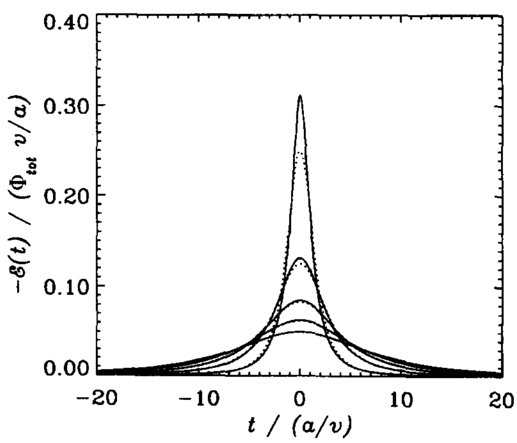

Fig. 4. emf induced in loop for $b/a=2$ (top curves), 4, 6, 8, and 10 (bottom curves): solid line—numerical result for double solenoid; dotted line—Eq. (4.4).

Note that this expression for $d\mathbf{A}$ is only valid for the solenoid if $r \gg a$; however, as shown below, it will suffice for our calculation and we choose to use it since it is relatively simple. The exact expressions for **A**, **B** are immensely more complicated—being infinite sums of Legendre polynomials[9]—and using them would only obscure the main point of this calculation.

The flux $\Phi(t)$ is first calculated for this slice of the solenoid, and then the total flux cutting the loop at time t is given by summing the contributions from each slice:

$$\Phi(t) = \int_{z=-\infty}^{z_0(t)} \int d\mathbf{B} \cdot d\mathbf{S}, \tag{4.8}$$

where $d\mathbf{S}$ is the usual element of any surface bounded by the loop and dz is contained in $d\mathbf{B}$ [see Eqs. (4.7) and (4.6)].

The surface S through which we are calculating the flux must always be drawn such that the condition $r \gg a$ is met. When the solenoid is far from the loop, i.e., $z_0 \ll 0$, this surface can be chosen to be the simplest: a plane. However, as z_0 approaches zero we can no longer integrate over this plane because a calculation of the flux through it requires a knowledge of $d\mathbf{B}$ at every point on it, and in this case there will be points for which the condition $r \gg a$ breaks down. As z_0 approaches zero, if we are still to use the $d\mathbf{B}$ calculated from (4.7) we must deform the surface in order to keep all points on it far from the origin, i.e., so that $r \gg a$ for all these points.

In principle the calculation would become very difficult here. However we are saved by Stokes' theorem. Invoking this enables us to write

$$\int d\mathbf{B} \cdot d\mathbf{S} = \int \nabla \times d\mathbf{A} \cdot d\mathbf{S} = \oint d\mathbf{A} \cdot d\mathbf{l}, \tag{4.9}$$

where $d\mathbf{l}$ is a line element around the loop's perimeter. We can now use (4.7) since we will only ever require the value of $d\mathbf{A}$ at the loop, and the expression for $d\mathbf{A}$ in (4.7) becomes more exact when b is much larger than a. This is the crucial point—we are *not* saying **A** is any more real than **B**; rather, we are using a very commonly known fact in electromagnetism, which is that in general **A** is easier to work with than **B**.[10]

So consider the infinitesimal slice of the solenoid. It carries current $n\ dz\ \Delta I$, and then we have

$$\Phi(t) = \mu_0/(4\pi) \int_{z=-\infty}^{z_0(t)} \oint d\mathbf{m} \times \mathbf{r} \cdot d\mathbf{l}/r^3. \tag{4.10}$$

For

$$d\mathbf{m} = (0,\ 0,\ n\Delta I \pi a^2\, dz),$$

$$\mathbf{r} = (b \cos\theta,\ b \sin\theta, -z)$$

$$d\mathbf{l} = (-b\ d\theta \sin\theta,\ b\ d\theta \cos\theta,\ 0) \tag{4.11}$$

we can write (with $\Phi_{\text{tot}} \equiv \mu_0 n \Delta I \pi a^2$ as previously)

$$\Phi(t) = \Phi_{\text{tot}}/(4\pi) b^2 \int_0^{2\pi} d\theta \int_{-\infty}^{z_0(t)} dz (b^2 + z^2)^{-3/2}$$

$$= (\Phi_{\text{tot}}/2)[1 + vt(b^2 + v^2 t^2)^{-1/2}] \tag{4.12}$$

as was found in (4.2) using the solid angle approach. The double solenoid result also follows immediately as before; alternatively we can substitute

$$d\mathbf{m} \to (0,\ 0,\ n\Delta I 2\pi a \Delta a\ dz) \tag{4.13}$$

and the calculation will carry through to give the same result.

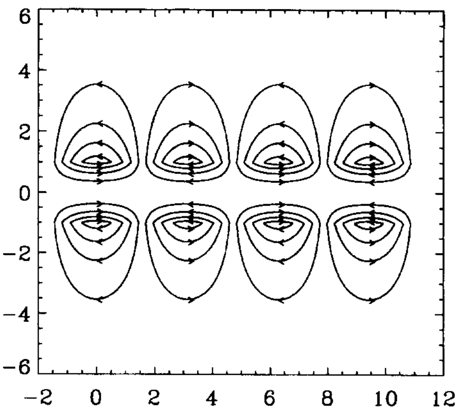

Fig. 5. Magnetic field of double solenoid of radius $a = 1$ carrying ac current such that the wave number is $k = 2\pi/\lambda = 2\pi f/v = 1$.

V. DOUBLE SOLENOID WITH AC CURRENT

As a final note we consider the most easily realizable setup for what we have been calculating. This concerns the emf induced in the loop due to an ac current with frequency f in the double solenoid. In this case we can again use a Runge–Kutta method to trace the magnetic field lines which are shown in Fig. 5 for wave number $k = 2\pi/\lambda = 1$. The magnetic field pattern repeats with wavelength v/f, and as the current oscillates in direction, the magnetic field pattern keeps its form but moves along the solenoid with speed v. Thus any particular piece of the loop is subjected to a field which alternates in direction with the same frequency as the current. Again we see here that the induced emf in the loop is produced by a very real magnetic field outside the solenoids.

VI. CONCLUDING REMARKS

We have shown that the usually accepted idea of the magnetic field being zero outside an infinite solenoid is only true for a static current. We have demonstrated that we can avoid what is often seen as a necessity, that is we have demonstrated that we can avoid invoking the vector potential to explain induction outside of the solenoid.

The main text of this paper has dealt with an infinite solenoid whose current is abruptly changed in a step-like way. For this case we have given both numerical and analytical evidence that allows us to say the following.

(1) The magnetic field outside the solenoid looks like that of a moving magnetic monopole.
(2) As the current changes, the radially outward directed magnetic field lines which make up this "monopole" will cut a loop outside the solenoid and induce an emf in the usual way as predicted by Faraday's law.
(3) A demonstration of this need not lean on any reference to the vector potential. The effect is directly linked to the **B** field lines cutting the loop. However, the vector potential can be extremely useful, but its use in this paper arises from the application of Stokes' theorem and as such is a mathematical device only.[11]
(4) This being said, this scenario presents no evidence for the **B** field being any less real than the **A** field.

After submitting our paper, the paper "Exact solution of the field equations in the case of an ideal infinite solenoid" by Templin[12] appeared in this journal. Templin considers the case of a steady alternating current in an infinite solenoid of radius a in which the current at any instant is *independent of position,* i.e., $I(z,t)=I_0 \cos \omega t$, or equivalently, the surface current in the surface of the cylinder defined by $\rho=a$ is

$$\mathbf{K}(\phi,z)=I_0 n \cos \omega t \hat{\phi},$$

where (ρ,ϕ,z) are cylindrical coordinates. He states that because of the simplification that $I(z,t)$ is independent of z the model lacks physical reality as it would require "an infinite set of synchronized radio frequency generators driving each segment of the solenoid." In addition, for $\mathbf{K}(\phi,z)$ to be independent of ϕ one requires "the propagation time around the solenoid to be much smaller than the period of the oscillations" giving $2\pi a/c \ll 2\pi/\omega$, or equivalently $\omega a \ll c$. For this idealized model, Templin obtains an exact solution of the field equations and shows that an alternating magnetic field exists outside the solenoid. Because of the assumed simultaneous change of current along the solenoid, the resulting magnetic field in Templin's model is everywhere parallel to the axis of the solenoid, independent of z, and oscillates with angular frequency ω, i.e., it is of the form

$$\mathbf{B}(\rho,\phi,z,t)=(B_0(\rho)\cos \omega t + B_1(\rho)\sin \omega t)\hat{z}.$$

Our approach is fundamentally different to that of Templin inasmuch as our work is motivated by the fact that the current in the solenoid at any instant is *not identical at all points along the solenoid,* but rather any change in current must *propagate along the solenoid* in the same way that a wave propagates along a transmission line. This results in the magnetic field configuration moving in the z direction with the speed of electromagnetic waves in the circuit, $v \approx c/(2\pi a n)$. Our basic assumption is that we can ignore retardation effects, and to justify this we require the wave speed along the solenoid, v, to be much less than c, i.e., we require $2\pi a n \gg 1$. For any tightly wound solenoid $2\pi a \gg 1/n$, and so this condition is usually satisfied. We also assume that the current depends only on z and t, and is independent of ϕ. In the case of a steady ac driving voltage we would then require that the wavelength $\lambda = 2\pi v/\omega = c/(\omega a n)$ be much greater than the spacing between adjacent coils, $1/n$, giving $\omega a \ll c$, interestingly the same range of validity as Templin's result. For the case of a steady ac driving voltage the magnetic field configuration in our model is as given in Fig. 5, and the whole pattern moves in the z direction with speed v.

[1]George E. Owen, *Introduction to Electromagnetic Theory* (Allyn & Bacon, Boston, 1963), p. 244; Albert Shadowitz, *The Electromagnetic Field* (McGraw-Hill, New York, 1975), p. 389; William T. Scott, *The Physics of Electricity and Magnetism* (Wiley, New York, 1966), 2nd ed., p. 345; Paul Lorrain, Dale P. Corson, and François Lorrain, *Electromagnetic Fields and Waves* (W. H. Freeman and Company, New York, 1988), 3rd ed., p. 440.

[2]A. P. French, "Faraday's law, question #6," Am. J. Phys. **62**, 972 (1994).

[3]For an account of the significance of **B** vs that of **A** in quantum mechanics, see R. P. Feynman, R. B. Leighton, and M. Sands, *The Feynman Lectures on Physics* (Addison-Wesley, Reading MA, 1964), Vol. 2. pp. 15-7–15-14.

[4]See for example, Ref. 3, pp. 24-1–24-3.

[5]William H. Press, Brian P. Flannery, Saul A Teukolsky, and William A. Vetterling, *Numerical Recipes* (Cambridge U.P., Cambridge, 1986), pp. 554–560.

[6]John D. Jackson, *Classical Electrodynamics* (Wiley, New York, 1975), 2nd ed., pp. 257–258.

[7]Reference 6, pp. 90–93.

[8]This is a standard result which can be found in most texts on electromagnetism, for example, Ref. 3, pp. 14-7–14-8.

[9]George Arfken, *Mathematical Methods for Physicists* (Academic, London, 1966), 1st ed., pp. 439–443.

[10]An excellent discussion can be found in Ref. 3, p. 15-7.

[11]A similar point of view is expressed in Sergei Schelkunoff, *Electromagnetic Fields* (Blaisdell, New York, 1963), 1st ed., p. 87.

[12]Jacques D. Templin, "Exact solution of the field equations in the case of an ideal infinite solenoid," Am. J. Phys. **63**, 916–920 (1995).

Why are the energy levels of the quantum harmonic oscillator equally spaced?

Antoine Royer

Département de Génie Physique, Ecole Polytechnique, Montréal, Québec H3C 3A7, Canada

(Received 9 November 1995; accepted 21 February 1996)

The quantum harmonic oscillator has the unique property that its energy levels are equally spaced. This property is connected, via the Wigner function, to the circular symmetry of the Hamiltonian in *phase space*, hence to the close relation between harmonic oscillations and rotations, very familiar at the classical level. We here have an example of a spectral regularity which is due to a symmetry in phase space, rather than in configuration space, as is more usual. © *1996 American Association of Physics Teachers.*

...we consider—and this is the most important point of the entire calculation—E as being composed of a completely definite number of finite, equal parts, and make use for that purpose of the natural constant $h=6.55\times10^{-27}$ [erg. sec]. This constant, when multiplied by the common frequency ν of the resonators, yields the energy element ϵ...[1]

...I can characterize the whole procedure as an act of