NOTES AND DISCUSSIONS

Ideal capacitor circuits and energy conservation

K. Mita and M. Boufaida
Department of Physics, St. Mary's College of Maryland, St. Mary's City, Maryland 20686

(Received 20 July 1998; accepted 8 December 1998)

In some introductory physics textbooks, authors point out that there is missing energy in the charging process of the capacitor circuit in Fig. 1(a).[1] In some other textbooks, they present a problem in which a portion of the electric charge is transferred from a fully charged capacitor to an empty one, as shown in Fig. 1(b).[2] In either case, half the energy of the circuit is missing in the final state after the switch S is thrown. The usual way to circumvent this difficulty is to introduce a small amount of resistance in the connecting wires to explain that the missing energy is dissipated as Joule heating. While this argument appears correct, it leaves us feeling unsettled. We wonder why is it that the energy is missing in the circuits in Fig. 1(a) or (b). We also ask why the introduction of resistance accounts for the missing energy. Is this the only way to resolve the difficulty? We will explore these questions in this note and present the general condition under which energy transfer to a capacitor is complete, without missing energy.

To be sure, implicit in Fig. 1(a) and (b) are idealizations and approximations of actual physical circuits so that we may describe the situations with simple mathematics. For example, in Fig. 1(a), the connecting wires are without resistance, and we ignore a small amount of radiation due to accelerated charges.[3] As for the capacitor itself, a real capacitor has resonances and losses with corresponding effective inductance and resistance.[4] If one includes these factors, one would not encounter the difficulty of missing energy. But in physics courses, we often deal with idealized circuits, and thus our focus here is to investigate the peculiarity of the idealized capacitor circuits with regard to energy conservation.

First we examine two other idealized, elementary circuits which do not exhibit the difficulty of missing energy. In Fig. 2(a) and (b) are shown simple resistance and inductance circuits, respectively. By comparing these two circuits with the capacitor circuit, we may find a clue as to why the capacitor circuit has the aforementioned peculiarity.

For the resistance circuit, there exists a simple differential equation

$$R\,dq/dt = V_0, \tag{1}$$

where $q(t)$ is the electric charge passing through the resistor R from time $t=0$ to some arbitrary time t, and V_0 is the voltage of the battery. When the switch S is thrown at $t=0$, we have $q=V_0 t/R$, and since

$$\int_0^t V_0 \dot{q}\,dt = \int_0^t R\dot{q}^2\,dt = qV_0, \tag{2}$$

the energy supplied by the battery is equal to the Joule dissipation in R. For the inductor circuit, the differential equation becomes

$$L\,d^2q/dt^2 = V_0, \tag{3}$$

with the particular solution $q=V_0 t^2/2L$, and since

$$\int_0^t V_0 \dot{q}\,dt = \frac{1}{2} L\dot{q}^2 = qV_0, \tag{4}$$

the energy supplied by the battery is equal to the magnetic energy stored in L at time t.

In each of these two cases the equation is differential and the charge q is a monotonically increasing, continuous function of time t. That is, the charge delivery from the battery is gradual. For the capacitor circuit, on the other hand, the equation is not differential and the charge, $q_0 = CV_0$, stored in the capacitor, rises abruptly, as a step function at time $t=0$. While the energy supplied by the battery is $q_0 V_0$, the energy stored in the capacitor is $q_0 V_0/2$. These facts suggest that the missing energy may stem from the instantaneous charging of the capacitor. Mechanically, Fig. 1(a) is equivalent to an idealized spring of spring constant k without mass or friction. When a constant force F_0 is suddenly applied to this spring and it is either compressed or stretched by a distance d, the same situation results. That is, the work done by the force is $F_0 d$, while the energy stored in the spring is $kd^2/2 = F_0 d/2$. In either case, the missing energy is due to too much idealization of the real system, resulting in an instantaneous process.

Mathematically, the instantaneous charging process in Fig. 1(a) may be expressed as

$$q = CV_0\,\theta(t), \tag{5}$$

where $\theta(t)$ is the Heaviside step-function defined by

$$\theta(t) = \begin{cases} 1, & t>0 \\ 0, & t<0. \end{cases} \tag{6}$$

Then the energy supplied by the battery is

$$E_B = \int_{-\infty}^{\infty} V_0\,\theta(t)\dot{q}\,dt = CV_0^2 \int_{-\infty}^{\infty} \theta(t)\,\delta(t)\,dt, \tag{7}$$

where $\delta(t)$ is the Dirac δ function, and because of the nature of the functions involved, we extend the integration range to $(-\infty, \infty)$. Here we encounter a difficulty. In order for the integral to exist, $\theta(t)$ must be continuous at $t=0$.[5] But since this is not the case, the integral does not exist. This indicates that the introduction of the δ function to analyze this particular singular phenomenon is not useful because of the discontinuity of the step function. A pathology caused by an idealization is not cured by mathematics in this case.

As is well known, the missing energy can be accounted for by introducing a small amount of resistance R_W associated with the connecting wires in the elementary capacitor circuit shown in Fig. 1(a).[6] Then the equation for the charge q becomes

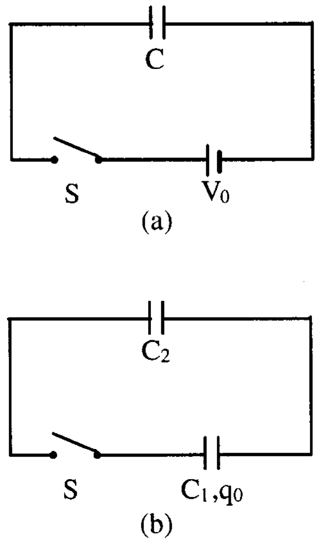

Fig. 1. Capacitor circuits with missing energy.

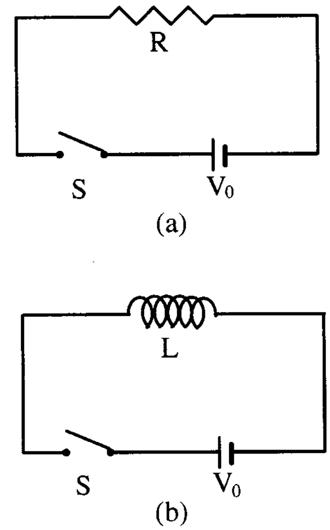

Fig. 2. Energy conserving elementary circuits.

$$R_W\dot{q}+q/C=V_0, \tag{8}$$

which is now a differential equation. Consequently, the charge $q(t)$ stored in the capacitor rises over a finite amount of time, though it may be a small time interval depending on the parameter R_W.

However, this is not the only way to account for the missing energy. We now introduce a small amount of inductance L_C of the closed circuit and ignore R_W. Replacing $R_W\dot{q}$ in Eq. (8) with $L_C\ddot{q}$, we have

$$L_C\ddot{q}+q/C=V_0. \tag{9}$$

The particular solution of Eq. (9) satisfying $q=\dot{q}=0$ at $t=0$ is given by

$$q=CV_0(1-\cos\omega t), \quad \omega=1/\sqrt{L_C C}. \tag{10}$$

The time t_0 for which the charge stored in the capacitor assumes the value q_0 for the first time is[7]

$$t_0=\cos^{-1}(1-q_0/CV_0)/\omega. \tag{11}$$

At this time the electric current is given by ωq_0, and the magnetic energy stored in the inductor is

$$E_L=L_C i_0^2/2=q_0V_0/2, \tag{12}$$

which is equal to the missing energy.

In each of the two examples cited above, the voltage across the capacitor changes from the step function to a continuous function of time when R_W or L_C is introduced, and the charging of the capacitor does not take place instantly. The same is true when both R_W and L_C are used. As the circuit is viewed in a more realistic manner, the difficulty of missing energy disappears.

Though the following is another idealization, one can think of still more variations of the capacitor circuit in Fig. 1(a), in which energy delivery to the capacitor takes place without the missing energy. This time we replace the battery with a special kind of power supply. Heinrich[8] showed that in the *RC* circuit, when V_0 is divided into N equal substeps, the energy dissipation in the resistor becomes smaller compared to the one-step charging process, and, in the limit as $N\to\infty$, the energy dissipation in the resistor tends to vanish. In order for this to take place, the capacitive time constant *RC* should be small compared to the time interval of each substep t_0/N, where t_0 is the total time taken for the charging process. This means that in the limit as $N\to\infty$, the voltage across the capacitor is written as

$$V=V_0t/t_0, \quad 0\leqslant t\leqslant t_0, \tag{13}$$

and $R\to 0$. Then the energy provided by the power supply is

$$E_{\mathrm{PS}}=\int_0^{t_0}V\dot{q}\,dt=\frac{1}{2}CV_0^2, \tag{14}$$

which is equal to the energy stored in the capacitor. Note that the discontinuities of the step functions are eliminated in the limit as $N\to\infty$.

We now generalize this point. In the capacitor circuit in Fig. 1(a), let us replace the battery with a power supply which provides the voltage of

$$V=V_0f(t), \tag{15}$$

where $f(t)$ is any monotonically increasing, continuous function of time with the properties

$$f(0)=0, \quad f(t_0)=1, \tag{16}$$

for the time interval $0\leqslant t\leqslant t_0$. The parameter t_0 may be finite or infinite. Then at time t_0, the energy stored in the capacitor is $CV_0^2/2$ and the energy provided by the power supply is

$$E_{\mathrm{PS}}=\int_0^{t_0}V\dot{q}\,dt=CV_0^2\int_0^{t_0}f(t)\dot{f}(t)dt=CV_0^2\int_0^1 f\,df=\frac{1}{2}CV_0^2, \tag{17}$$

where it is assumed that $f(t)$ is differentiable on $[0,t_0]$. When the function $f(t)$ contains a discontinuity on $[0,t_0]$, Eq. (17) no longer holds valid and the energy delivery from the power supply to the capacitor is not complete.

Coming back to Eq. (7), if we now employ a mathematical ‘‘fudging’’ and use the expression of the step function[9]

$$\theta(t)=\lim_{n\to\infty}\tfrac{1}{2}(1+\tanh nt), \tag{18}$$

without taking the limit, then, for large, finite n, the step function becomes a continuous function with the value 1/2 at time $t=0$, and $E_B=CV_0^2/2$. This observation, however, is a pure mathematical curiosity and it may not resolve the issue at hand.

To state our conclusion in a general manner, so long as the voltage across a capacitor is a continuous and differentiable function of time, the energy delivery to the capacitor is complete without the missing energy. The idealized circuits in Fig. 1(a) and (b) happen to be exceptions.

ACKNOWLEDGMENTS

We thank C. Adler and T. Darkhosh for enlightening conversations.

[1]For example, see A. Hudson and R. Nelson, *University Physics* (Saunders, Philadelphia, 1990), p. 675.

[2]For example, see D. Halliday, R. Resnick, and J. Walker, *Fundamentals of Physics* (Wiley, New York, 1997), 5th ed., p. 649; C. Zucker, ''Condenser problem,'' Am. J. Phys. **23**, 469 (1955).

[3]For a detailed account of radiation from an accelerated charge, see, for example, R. Ehrlich, J. Tuszynski, L. Roelofs, and R. Stoner, *Electricity and Magnetism Simulations* (Wiley, New York, 1995), pp. 137–149.

[4]For an excellent discussion of real circuit elements, see R. P. Feynman, R. B. Leighton, and M. Sands, *The Feynman Lectures on Physics* (Addison–Wesley, Reading, MA, 1964), Vol. II, Chap. 23.

[5]G. Arfken, *Mathematical Methods for Physicists* (Academic, Orlando, 1985), p. 481.

[6]We will examine the circuit in Fig. 1(a) only. The extension of our analysis to the circuit in Fig. 1(b) is straightforward.

[7]To avoid the complications due to the oscillatory nature of this circuit, we may restrict time to the interval $0\leqslant t\leqslant \pi/\omega$.

[8]F. Heinrich, ''Entropy change when charging a capacitor: A demonstration experiment,'' Am. J. Phys. **54**, 742–744 (1986); I. Fundaun, C. Reese, and H. H. Soonpaa, ''Charging a capacitor,'' *ibid.* **60**, 1047–1048 (1992).

[9]See Ref. 5, p. 490.

An unusual feature of charge densities for two-particle bound states

Horace W. Crater

The University of Tennessee Space Institute, Tullahoma, Tennessee 37388

(Received 2 October 1998; accepted 17 November 1998)

Two oppositely charged particles moving nonrelativistically in a two-particle Coulomb bound state produce a charge distribution that is equivalent to that of two single-particle bound states of opposite charge moving in separate external potentials. With examples, we show that this is unlikely to be true if the binding force is not pure Coulombic.

The main focus of the two-body bound state problem in nonrelativistic quantum mechanics is the calculation of the energy levels and their associated eigenstates. In this paper we develop an expression for the effective charge distribution of two charged particles bound in such a state by an arbitrary two-body potential. That we obtain a sum of two distributions each related to the absolute value squared of the bound state wave function is not surprising. What is rather unusual is that the mass factors that accompany this total distribution have the effect, for Coulomb bound states, of replacing the reduced mass in the bound state wave function with the constituent mass for each distribution in the sum. Thus one can picture the charge distribution as equivalent to that produced by a particle of mass m_1 and charge e_1, bound to a *fixed* center with charge e_2 plus that produced by a particle of mass m_2 and charge e_2 bound to a *fixed* center with charge e_1. The fixed center in both cases is the center of mass. The picture is thus that of two noninteracting charged particles moving independently in separate external potentials. This picture does not hold in general for potentials that are not pure Coulombic.

Let P be a field point located at a position $\mathbf{R}$ relative to the center of mass chosen as the origin of our coordinate system. The coordinates $\mathbf{r}_1$ and $\mathbf{r}_2$ are the locations of the particles relative to the c.m. and $\mathbf{r}$ is their relative coordinate. Thus

$$\mathbf{r}_1=\frac{m_2}{M}\mathbf{r}, \tag{1a}$$

$$\mathbf{r}_2=\frac{-m_1}{M}\mathbf{r}, \tag{1b}$$

where $M=m_1+m_2$. The charge density operator is defined as

$$\rho(\mathbf{r}_1,\mathbf{r}_2)=e_1\delta(\mathbf{R}-\mathbf{r}_1)+e_2\delta(\mathbf{R}-\mathbf{r}_2). \tag{2}$$

The charge density $\rho(\mathbf{R})$ at P, is the expectation value[1]

$$\begin{aligned}\rho(\mathbf{R})&=\langle\psi|\rho(\mathbf{r}_1,\mathbf{r}_2)|\psi\rangle\\&=\int(d\mathbf{r})|\psi(\mathbf{r})|^2(e_1\delta(\mathbf{R}-\mathbf{r}_1)+e_2\delta(\mathbf{R}-\mathbf{r}_2))\\&=\int(d\mathbf{r})|\psi(\mathbf{r})|^2\left(e_1\delta\left(\mathbf{R}-\frac{m_2}{M}\mathbf{r}\right)+e_2\delta\left(\mathbf{R}+\frac{m_1}{M}\mathbf{r}\right)\right)\\&=\int(d\mathbf{r})|\psi(\mathbf{r})|^2\left(e_1\delta\left(\mathbf{r}-\frac{M}{m_2}\mathbf{R}\right)\left(\frac{M}{m_2}\right)^3\right.\\&\quad\left.+e_2\delta\left(\mathbf{r}+\frac{M}{m_1}\mathbf{R}\right)\left(\frac{M}{m_1}\right)^3\right)\\&=e_1\left|\psi\left(\frac{M}{m_2}\mathbf{R}\right)\right|^2\left(\frac{M}{m_2}\right)^3+e_2\left|\psi\left(-\frac{M}{m_1}\mathbf{R}\right)\right|^2\left(\frac{M}{m_1}\right)^3.\end{aligned} \tag{3}$$

For $e_1=Ze$ and $e_2=-e$ and $V(r)=-Ze^2/r$ the stationary states are

$$\psi_{nlm}(\mathbf{r}) = R_{nl}(r) Y_{lm}(\theta,\phi) \equiv \frac{1}{a_0^{3/2}} \phi_{nlm}(\mathbf{r}/a_0), \tag{4}$$

where

$$R_{nl}(r) = -\left\{ \left(\frac{2Z}{na_0}\right)^3 \frac{(n-l-1)!}{2n[(n+l)!]^3} \right\}^{1/2} e^{-1/2\rho} \rho^l L_{n+l}^{2l+1}(\rho) \tag{5}$$

and

$$a_0 = \frac{\hbar^2}{\mu e^2}, \quad \rho = \frac{2Z}{na_0} r,$$

with $\mu = m_1 m_2 / M$. The total charge density will, of course, always be the sum of two single-particle bound state charge densities. However, what makes the situation presented by the two-body Coulomb bound state unique is the dependence of the wave function on the mass factors in (3). As a consequence, the total charge density becomes equivalent to that of the direct sum of two charged single-particle states, each bound to an *infinitely heavy* center of force at the c.m. That is

$$\rho(R) = e\left(\frac{Z}{a_1^3} \left| \phi_{nlm}\left(\frac{\mathbf{R}}{a_1}\right) \right|^2 - \frac{1}{a_2^3} \left| \phi_{nlm}\left(\frac{R}{a_2}\right) \right|^2 \right) \tag{6}$$

in which

$$a_1 = \frac{m_2}{M} a_0 = \frac{\hbar^2}{m_1 e^2} \tag{7a}$$

and

$$a_2 = \frac{m_1}{M} a_0 = \frac{\hbar^2}{m_2 e^2} \tag{7b}$$

are the only mass-dependent factors in the wave functions. In the ground state, for example,

$$\rho(\mathbf{R}) = e\left(Z \frac{e^{-2R/a_1}}{\pi a_1^3} - \frac{e^{-2R/a_2}}{\pi a_2^3} \right).$$

Note that in the static limit $m_1 \to \infty$ the first distribution reduces to a delta function (Ref. 2 gives the charge density for an electron bound to a fixed nucleus with charge *Ze*). For finite m_1, say m_p, we have $a_1 \sim 27 fm$. This is significantly larger than the proton's ''free'' size of about 1 F. However, this size would be hard to detect by e^- scattering because charged probes energetic enough to have such a small wavelength would ionize the atom, thus effectively reducing the ''size'' of the proton to its free size. We point out that the effective size of this charge distribution does not, of course, affect the electron's binding energy, unlike a real size. One could say that there is a correlation to the motional effects responsible for the reduced mass dependence of the binding energy, in that as the size decreases the reduced mass dependence collapses toward the electron's mass. But this is only a consequence of the fact that both of these aspects result from the inner workings of the Schrödinger equation for an electron in the field of a point charge nucleus.

If the potential deviates from Coulombic then there are other mass-dependent factors than the Bohr radius. For example, if

$$V(r) = \frac{-Ze^2}{r} + \frac{\beta}{r^2}, \tag{8}$$

then the spectrum and wave function are related to the Coulombic one by analytic continuation of the angular momentum,

$$l(l+1) \to l(l+1) + 2\mu\beta \equiv \lambda(\lambda+1), \tag{9}$$

$$\lambda = \lambda(l,\mu) = -1/2 + \sqrt{(l+1/2)^2 + 2\mu\beta}. \tag{10}$$

That is

$$\psi_{nlm}(\mathbf{r}) = R_{n\lambda(l,\mu)}(r) Y_{lm}(\theta,\phi). \tag{11}$$

The mass factor in (3) cannot produce changes in the reduced mass factor in (9) to the individual constituent masses as occurs in (6) unless it just happens to be inversely proportional to the reduced mass. Assuming that is not the case, for the potential (8) the total charge density is not equivalent to the direct sum of two single-particle bound states of opposite charge.

Note that it might be argued that in reality, such corrections to the Coulomb potential generally arise from relativistic effects. In that case β may be, to lowest approximation, inversely proportional to the reduced mass so that λ becomes independent of μ. However, this would take us beyond the scope of this work into considerations of the relativistic two-body problem.[3] The final example we give has no such ambiguities.

Consider the case in which the potential energy includes a portion which can be approximated by the simple harmonic oscillator potential. Suppose further that its effects dominate the Coulomb potential in computing the bound state wave function so that to a good approximation the bound state wave function is that for the harmonic oscillator potential alone. For $V = \frac{1}{2}kr^2$ that wave function is

$$\phi_{nlm} = R_{nl}(r) Y_{lm}(r), \tag{12}$$

where

$$R_n(r) = \left(\frac{\mu k}{\hbar^2}\right)^{3/4} \sqrt{2} \rho^l e^{-\rho^2/2} \Gamma\left(l + \frac{1}{2}\right) \times \begin{pmatrix} \frac{1}{2}(N+l+1) \\ \frac{1}{2}(N-l) \end{pmatrix} L_{1/2(N-l)}^{l+1/2}(\rho^2) \tag{13}$$

and

$$\rho = \left(\frac{\sqrt{2\mu k}}{\hbar}\right)^{1/2} r. \tag{14}$$

The mass dependence in (13) and (14) is not of the correct type that lets the mass factors in (3) replace the reduced mass with the constituent particle masses m_1 and m_2, as occurs in Eqs. (3)–(6). Although it is evident that if $k = \mu\omega^2$ with ω a fixed constant then this property displayed by the Coulomb potential alone persists, this will not occur in general in nature because k is determined by the behavior of a realistic potential near equilibrium, which has no such special mass dependence. In general, the appearance of parameters in the potentials that are not dimensionless (in natural units) and do not depend on the reduced mass would also not be of the correct type.

This feature of the two-body bound state charge density for the Coulomb potential adds to other unique quantum features of this potential such as producing cross sections that agree with the classical result and displaying dynamical symmetries that result in ‘‘accidental’’ degeneracies.

[1] For stationary states the two-particle wave function $\Psi(\mathbf{r}_1,\mathbf{r}_2) = \exp^{i\mathbf{P}\cdot\mathbf{R}_0}\psi(\mathbf{r})$ in which $\mathbf{R}_0$ is the c.m. position and **P** is the total momentum of the system. The c.m. portion of the wave function will not contribute to the expectation value.

[2] A. Messiah, *Quantum Mechanics* (North-Holland, Amsterdam, 1963), Vol. II, p. 815.

[3] H. W. Crater and P. Van Alstine, ‘‘Quantum constraint dynamics for two spinless particles under vector interaction,’’ Phys. Rev. D **30**, 2585–2594 (1984).

LIMITED BANDWIDTH

Young scientists who have difficulty in finding acceptance for their work are likely to blame such troubles on their exclusion from ‘‘the establishment.’’ But in fact, the effective establishments for such purposes are very narrow. Even a famous scientist who presents a case before the wrong audience can be ignored. In 1917, when he was already a noted scientist and retiring as president of the German Physical Society, Albert Einstein presented a paper pointing to the difficulty caused by chaos in Sommerfeld—Wilson quantization. That paper remained obscure for decades. My own PhD thesis was related to this subject, yet I was unaware of Einstein’s work. Even when I saw a reference to it some years later, I did not bother to examine it. All of us have limited bandwidth for information intake; there are no simple villains when new concepts are ignored.

Rolf Landauer, ‘‘Fashions in Science and Technology,’’ Phys. Today **50** (12), 61–62 (1997).