Did you know?

Some puzzles in radio engineering fundamentals

by Epsilon

THE ART of radio engineering is now well into its second half-century; many of the fundamentals, probably once well understood, are perhaps too easily accepted today and seldom explained adequately in basic engineering courses. Take, for example, a capacitor. Readers will know that there are quite fundamental laws which describe its behaviour. These are the laws of charge and energy, and they are often used to solve certain problems in much the same manner as momentum and kinetic energy are used in mechanics.

A typical problem is shown in Fig. 1.

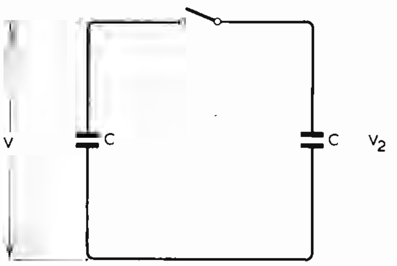


Fig. 1. *What is the voltage on the right-hand capacitor after the switch is closed?*

Here a capacitor of capacitance C is charged to a voltage V. At a particular time, the first capacitor is connected to a second capacitor, also of value C, but containing no charge. By the law of charge conservation, the charge before and after the connection is the same, and is given by

$$Q = CV = 2CV_2$$

and therefore the voltage V_2 is equal to $V/2$. But by the law of energy conservation,

$$E = CV^2/2 = (2C)V_2^2$$

and the voltage V_2 is equal to $V/\sqrt{2}$.

Now since capacitors are essentially lossless, energy cannot vanish without trace, and so the second answer should be the correct one. But this would imply that charge had increased by a factor of $\sqrt{2}$, thus violating the law of conservation of charge. The problem becomes really ridiculous if one capacitor is charged to $+V$, the other to $-V$. The net charge is then zero, and the use of one method would predict a final voltage of zero, the other a finite voltage of indeterminate sign.

Unfortunately, *both* conservation laws happen to be cornerstones of electrical engineering theory and are not to be discarded lightly. Some means must be found to reconcile the two laws, but how? No doubt readers will reassure themselves at this point by claiming that all real capacitors have resistance, and that the losses associated with this account for the discrepancies between application of the two laws.* A natural reply is to make the capacitors operate at superconducting temperatures and to reconsider the problem. Another ingenious way out might be to note that it is very difficult to discharge a capacitor without forming an arc (and hence getting rid of excess energy). Unfortunately for this suggestion, semiconductor technology does enable an arcless contact to be made, and so this explanation is at best a weak one. Accepting that simple measurements with a voltmeter show that the charge is conserved, what is the explanation of the apparent disappearance of the energy?

A second example, representing such an everyday feature of electronic equipment that its correct operation is taken for granted, is shown in principle in Fig. 2. A coaxial cable takes an r.f. signal from one part of a system to another. The system has a metallic ground plane which can be considered as being infinite in extent. Standard practice dictates that the outer braid of the coaxial cable is connected to the ground plane at both ends. An interesting question now emerges: what path does the return current from the load take? One answer (which is certainly true at d.c.), is that it takes the path of least resistance, or rather, it shares itself between the outer braid of the coaxial cable and the ground plane in the ratio of conductances. At a.c., the impedance between two points on a ground plane is effectively zero, whereas between the two ends of the outer conductor of a coaxial cable it is roughly

$$X = 0.21f\log_e(2l/D)$$

where l is the length in metres, D is the braid radius, and f is the frequency in MHz. Clearly the impedance increases with length and frequency, and therefore most, if not all, of the return current does not flow in the outer conductor at all, but in the ground plane. Of course, this situation does not happen; if it did, the outer conductor would be redundant, the cable would be unscreened, and the concept of characteristic impedance would be quite meaningless. Obviously, no matter what the cable length or the frequency, *none* (well, almost none) of the current flows in the ground plane. Why?

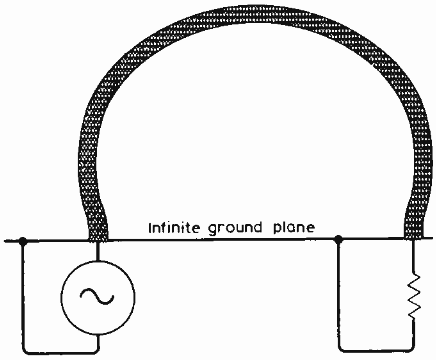


Fig. 2. *Does the return current from the load flow in the ground plane or the outer braid of the cable?*

A third example is not so much one of fundamental principle as one of observed fact. It concerns switch-on surges in transformers. If a transformer (the larger the better) is connected directly to a mains supply, a distinctive hum is often heard which decays away over a period of tens of cycles. If the transformer is large enough it may blow a quite substantial fuse. Why? Those who have experienced the effect will mutter "switching-on surge," but that is a description of the problem and not a quantifiable explanation of its cause. A related problem, which will help to obtain the answer, concerns the circuit shown in Fig. 3. Assuming that the switch is closed when the applied sine wave is at zero and, for the present, that the resistance is zero, is the current in

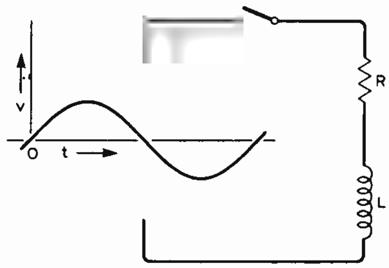
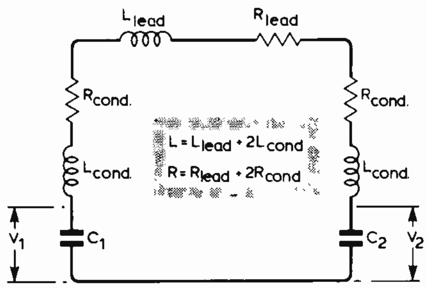


Fig. 3. *The switch is closed at t = 0. What is the current in the inductance?*

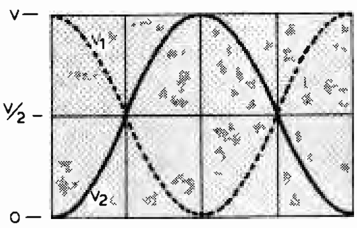
*It so happens that this statement is exactly true for any finite value of resistance.

the inductance sinusoidal and does it have an average d.c. value of zero?

Before continuing to two more problems less related to real life, let us examine the answers to the questions already presented. The first example is an interesting one, if only because it is so fundamental. As a first step it can be noted that all capacitors must have physical size; *it is simply not possible to make a finite capacitance of infinitely small dimensions;* secondly, whenever a capacitor is discharged the current must flow through a finite distance, and thirdly, current flowing through a distance generates a magnetic field, which in practical terms means that every capacitor possesses a small inductance.



(a)



(b)

Fig. 4. *The solution to Fig. 1 with $R = 0$*

Fig. 4(a) shows the equivalent circuit of Fig. 1, and Fig. 4(b) shows the actual voltages of the two capacitors as a function of time. The whole circuit is resonant at a frequency given by

$$\omega^2 LC/2 = 1$$

V_2 starts at zero, oscillating between a value of zero and V, the mean value being $V/2$. V_1 has the same sinusoidal form as V_2 but starts at V and decreases down to zero. Both charge and energy can now be accounted for. The total charge in the two capacitors remains constant at $Q = CV$, thus satisfying the requirement of conservation of charge, but the charge in each separate capacitor oscillates from one to the other about the mean.

The energy flow is more complicated, there being a continual transference between capacitive and inductive storage. Thus, beginning at time $t = 0$ in Fig. 4(b) capacitor C_2 starts with zero voltage and zero energy. C_1 starts at V. One quarter-cycle later both capacitors have the same voltage $V/2$, and there is zero voltage across the inductance. Since the current in an inductance lags the voltage by a phase angle of 90°, the former is now at a maximum and it will be found that exactly one half the energy resides in the inductance. This accounts for the "missing" energy. One half a cycle later the voltage across C_2 is a maximum and equal to V, and that across C_1 is zero; all the energy has now been removed from the inductance and resides in C_2. The reader can follow the remainder of the cycle.

No matter how small the lead inductance, the oscillation just described is always present, and, taken with the steady voltage, it fully accounts for both the original charge and the energy. The reason that an erroneous result can be obtained, in this case by neglecting inductance, is because the laws of conservation do not tell us *how* charge or energy may be stored, only that they cannot disappear.

The explanation may now be developed a little further, to begin with by allowing a small series resistance to be present as shown in Fig. 4(a). The oscillation, instead of persisting indefinitely as before, now decays exponentially (the multiplying factor is actually $\exp(-Rt/2L)$) and leaves a steady state voltage of $V/2$ on both capacitors. This is, of course, the voltage measured by a d.c. meter. Taking a further step, just as a capacitor possesses inductance, so it also possesses radiating properties, and there will in general be an apparent resistive loss because of this. Taking a third step, capacitance, inductance and radiation resistance are not lumped circuit elements but are distributed, and so even the reasoning given above is at best an approximation.

The explanation of the screening properties of the coaxial cable concerns the self and mutual inductance of the two conductors of which it is comprised. Fig. 5 shows a longitudinal cross section of the cable, large letters being used for the outer conductor and small letters for the inner.

A current i flowing upwards in the inner conductor sets up a magnetic field whose lines of force go around it. As a simplification we shall assume that the conductor is straight and long, and then these force lines have a magnetic field strength of

$$h = i/2\pi x \text{ ampere turns per metre}$$

where x is the distance from the centre of the conductor (but note that this expression is not applicable when x is less than d). A current I flowing downwards in the outer conductor also causes a magnetic field strength, this time given by

$$H = -I/2\pi x,$$

where x is measured from the centre of the hollow tube which forms the outer conductor and I is distributed uniformly around the periphery. Inside the tube the field due to I is everywhere zero.

The magnetic fields cause back voltages to be generated in each conductor whenever they change with time. The current I causes a large back voltage to appear in the outer conductor (the self inductance term) but this is cancelled by the back voltage caused by the current in the inner conductor (the mutual inductance term). Now it so happens that the magnetic field *inside* the outer conductor does not have any effect, and the total back voltage is thus proportional to $(h-H)$ or $(i-I)/2\pi x$. If $i = I$, all external magnetic fields are exactly equal to zero (in other words, the cable is properly screened) and the back voltage is also zero. Each end of the cable is at exactly the same potential and no current flows in the earth plane; if it did, a potential in the correct sense to cancel it would appear along the cable. We can all breathe a sigh of relief at the result, because otherwise r.f. engineering would be impossible. However, at low frequencies, particularly audio, cable resistance starts to become important and screening against magnetic pick-up is not at all so easy.

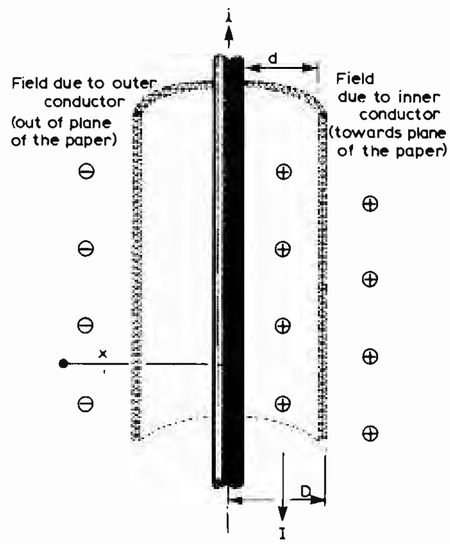


Fig. 5. *Fields due to the inner and outer conductors of a coaxial cable go around the cable. The left and right sides show the field separately.*

The full explanation for the third example can only be found satisfactorily by recourse to differential equations. Before giving the solution, it is as well to recall that the properties associated with inductance are expressed solely in terms of the back voltage developed when a current experiences a rate of change. Specifically:

$$V = L \times \text{no. of amperes changed per second}$$

It is important to realize that the voltage V is not a function of any constant current, which could be infinite without altering V in any way. With this in mind, the full solution for the current in Fig. 3 becomes understandable. It is

$$I = (E/Z) \sin(\omega t - \phi) + (E/Z/ \exp(Rt/L) \sin \phi$$

where $\tan\phi = \omega L/R$ and $Z = \sqrt{R^2 + (\omega L)^2}$.

The solution will be seen to consist of two terms, the first being the one generally used in a.c. impedance calcu-

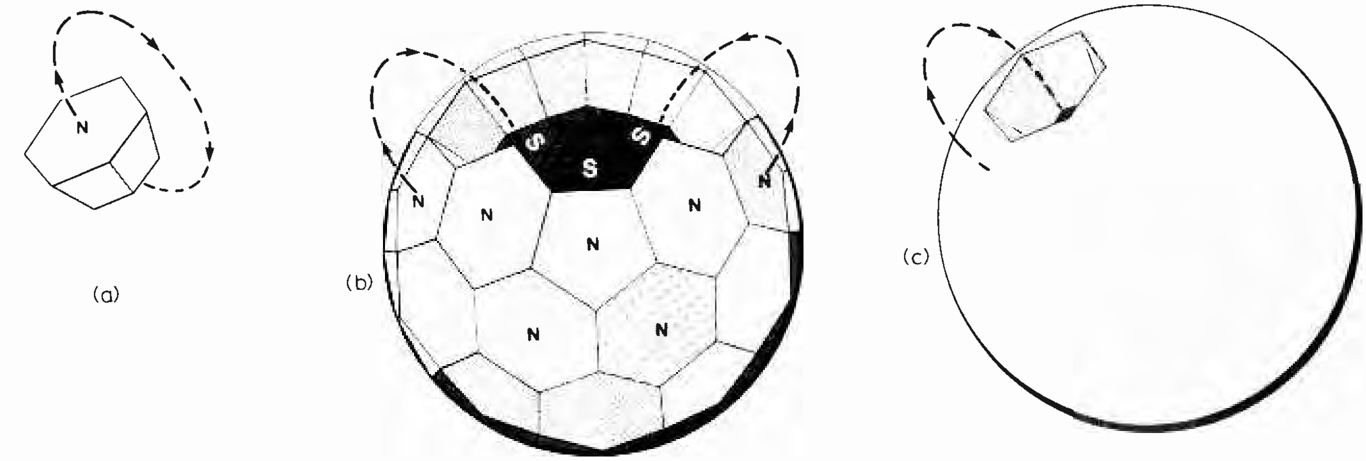


Fig. 6. *What is the external magnetic field when the last plug (a) is placed into the hole in the hollow sphere (c)?*

lations, the second being a transient term which is d.c. with an exponentional decay.

The solution for the case when R is zero is given by:

$$I = -(E/Z)\cos\omega t + (E/Z)$$

and there is thus a standing d.c. term equal in magnitude to the peak alternating current. The total current starts at zero and builds up to twice the value normally expected, but it never reverses in sign. There is a constant direct current circulating on a nominally a.c. supply and this persists indefinitely. In real life, resistance is always present and the d.c. term decays to zero; the lower the resistance, the lower the rate of decay. Knowing this, an explanation of why a switch-on occurs can now be given.

When a transformer is connected to the mains supply, a circulating direct current is set up as just described. If the switch-on occurs at or near the zero voltage point of the a.c. cycle, the d.c. is at a maximum, and the total current runs up to nearly twice the normal value given by $I = E/\omega L$. Twice the normal magnetization current is often more than sufficient to run the iron core of the transformer into saturation, and the laminations start to protest loudly. With the transformer iron saturated, the instantaneous value of L is grossly reduced and so the magnetizing current must increase to generate a back voltage which is equal to the mains supply. The effect persists until the direct current dies away or until the fuse blows.

After these questions and answers on rather everyday topics, here are two problems of a more thought-provoking nature.

A small bar magnet is launched into outer space, where it can be assumed to be free of any external influences. The magnet is set spinning about an axis which passes through its middle and is perpendicular to the line joining the two poles. What happens to the rotational speed of the magnet with the passage of time? No mechanical forces on the magnet (such as air resistance) need be considered.

Our second problem is also concerned with permanent magnets. Fig. 6 (a) shows a magnet made in the form of a six sided tapered plug. The lines of force of this magnet run from north to south (by convention). A number of these plugs† can be assembled as in Fig. 6 (b), and the result will be part of a hollow sphere. Lines of force will emanate from the outside of the sphere and will enter the inside as shown in Fig. 6 (b). The assembly of the sphere can continue until the situation in Fig. 6 (c) is reached, at which point lines emanating from the outer surface still return via the single hole to the south pole of the inner surface. A compass needle passed anywhere near the outer surface would record that it behaved as a magnetic north pole except near to the hole. The final plug is now inserted into the hole. What is the external magnetic field of the sphere at large, intermediate, and zero distances from the surface?

On a superficial level the two answers happen to be rather obvious: the spinning bar magnet slows down and the magnetic field outside the sphere is everywhere zero. The more quantitative explanations are as follows.

The spinning magnet generates an alternating magnetic field that gives rise to an electromagnetic effect and hence to radio waves. The power associated with these comes from the only available source, the kinetic energy of the spinning magnet, which therefore slows down. The explanation is rather an interesting one, because it shows that there is no reason why mechanical energy should not be turned directly into radiation without the use of electronic devices. However, I should point out that the idea is intriguing rather than practical!

To carry the explanation a little further, the magnet can be assumed to be replaced by a solenoid carrying a current I whose field matches that of a magnet. Next, the rotating magnetic field can equally well be created by two such solenoids fixed in an inertial reference frame at right angles to one another and carrying currents.

$$I_1 = I\sin(2\pi v_0 t),$$

$$I_2 = I\cos(2\pi v_0 t)$$

v_0 being the speed of the real magnet in revolutions per second. By this substitution the problem has been reduced to one of radio engineering. Each solenoid acts as a small loop antenna, the radiation from which results in a circularly polarized radio wave. Now the radiation resistance of an electrically small loop is given by

$$R = 31,200\, A^2N^2V^4/c^4$$

where A is the area of the solenoid, N is the number of turns, V is the frequency in cycles per second, and c is the velocity of light. By equating the kinetic energy stored in the rotating magnet to the I^2R losses in radiation it can be shown that the speed after a time t is

$$v = v_0(1 + kt)^{-1/2}$$

$$k = 31,200\, (IAN)^2(2\pi v_0)^2/(c^4W)$$

where W is the moment of inertia of the magnet.

The explanation for the magnetic field outside the sphere being zero can be given by reducing the problem to absurdity. Since the sphere is perfectly symmetrical in a three-dimensional sense there can be no preferred axis of magnetization; if lines of force do exist, they can only be perpendicular to the surface and they must all have the same direction of flow. But then this is tantamount to saying that the sphere acts as if it were a unit magnetic pole. Now man has been searching for unit magnetic poles for a long time, and, like the philosopher's stone, they have never been found (except possibly at the subatomic level). Unless you believe otherwise, the only possible solution is for the field outside the sphere to be everywhere zero. A more formal proof exists.

†In practice, and in theory, a sphere cannot be assembled from six-sided plugs alone. To get a perfect fit it must be done with a mixture of six-sided and five-sided plugs (see Fig. 6 (b)), as in the truncated icosahedron. However, this awkward fact should not affect the author's discussion. — Ed.