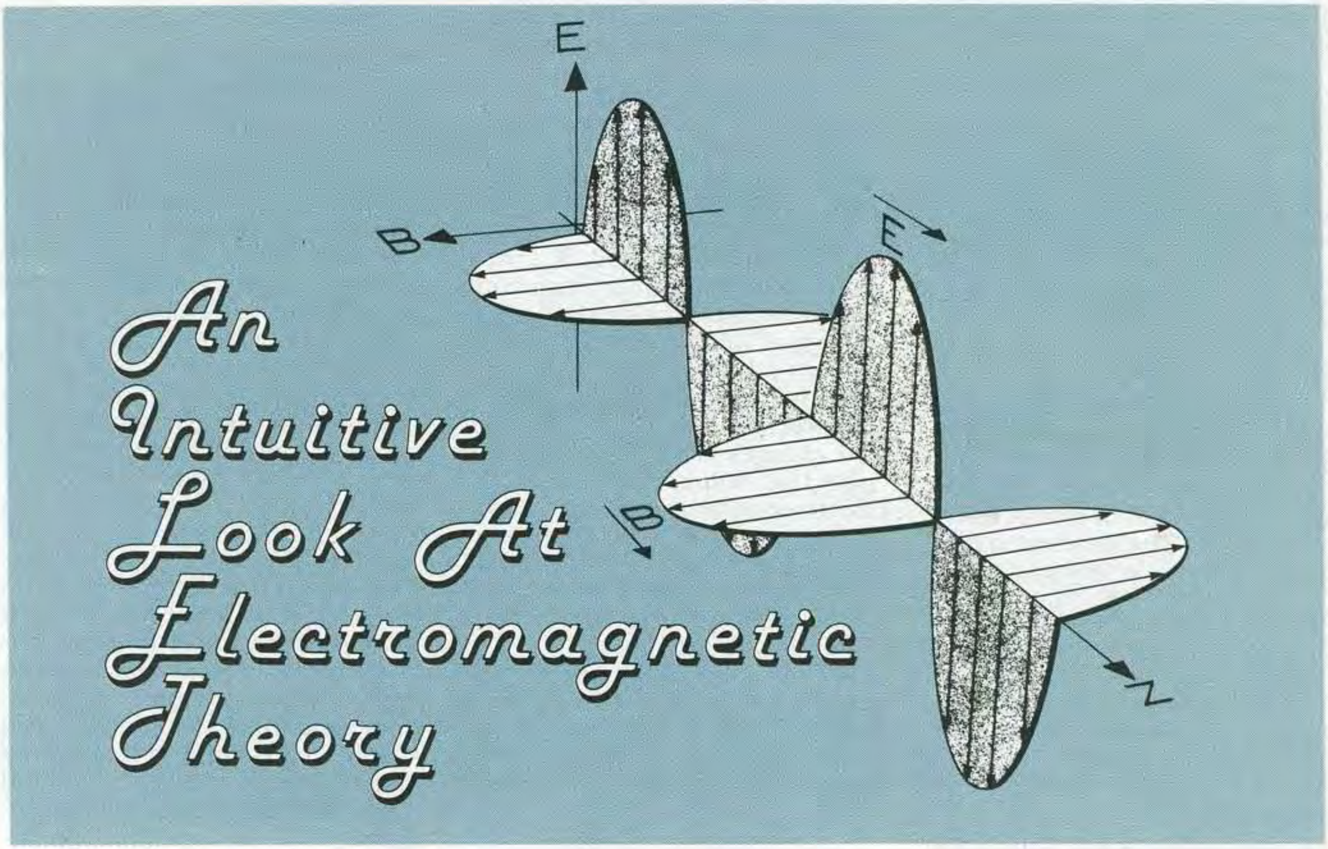
An Intuitive Look At Electromagnetic Theory



WILLIAM P. RICE

SOME OF OUR READERS MAY NOT BE aware that 1991 marks the 100th anniversary of the third edition of James Clerk Maxwell's *A Treatise on Electricity and Magnetism*, the ultimate reference on electromagnetic theory. What better way to recognize the impact that Maxwell had on the study of electromagnetics than to present the first of a series of articles in **Radio-Electronics** on the subject? In this edition, we will take a physical, intuitive look at the basics of electromagnetism, how they relate to some common electronic components, and how to interpret some of the complex mathematical symbolism. Only a familiarity with vector algebra is needed.

Maxwell's equations were first formulated in 1873. In his first publication, a mathematical foundation for relating electric and magnetic effects were given. In the Preface to the 1891 edition, J.J. Thomson noted that most of his students had difficulty with some aspects of electromagnetic theory. One hundred years later, not much has changed in that regard. One reason is that electromagnetic theory requires knowledge of some involved mathematics such as vector and tensor calculus and integral-differential equations.

Maxwell's idea that a changing electric field gives rise to an associated magnetic field developed from an intuitive sense for the natural order in the world. By presenting physical concepts in such an "intuitive" way, the reader will find it easier to understand Maxwell's equations, and his mathematical approach. Let's begin by examining the concept of an electric field.

The electric field

A scaler can be thought of as a quantity that can be completely characterized by its magnitude. Some examples of scaler quantities are mass, time, and volume. A scaler field is simply an extension of the scaler concept. It is a function of position that is specified by its magnitude at all points in a region of space. Land elevation is a two-dimensional scaler field because at each point of latitude and longitude there is an associated height above sea level. Air temperature is an example of a three-dimensional scaler field. With the appropriate instrument one could measure the height, or temperature, at each point. A scaler quantity is symbolized by a letter, such as h, for height.

A vector is a quantity that is characterized by its magnitude and direction. Some examples of vectors are velocity, acceleration, and force. A vector field is a function of position that is specified by its magnitude and direction at all points in a region of space. An example of a vector field is air velocity, where at each point in space, the magnitude and direction of air flow can be measured with the proper instrument. Vectors are often symbolized by letters with arrows above them, however, we will use boldface letters to indicate vectors.

When using vector notation, **A** is a vector with a specific magnitude and direction, and −**A** is a vector of the same magnitude but pointing in the opposite direction. Vectors are illustrated graphically by arrows, which have a direction and a corresponding length, which is proportional to the magnitude.

The field concept allows us to associate something that happens at one point with what happens at another point even though there may be no material objects connecting those points; examples are air temperature and velocity fields. Although we will not be directly concerned with them here, there are other types of fields, such as tensor fields, that assign a set of three vectors to each point in space, or quantum fields that assign mathematical operations to each point in space-time.

Electric charges

Experiments have shown that electric charges are either positive or negative. Like charges repel each other, unlike charges attract. The unit of charge is the Coulomb, C. The smallest magnitude of charge, *e*, is equal to

$$e = 1.60 \times 10^{-19} \text{C}.$$

Charge follows the principle of conservation, which states that the net sum of all charges in an isolated system remains constant. A charge can be moved, but it cannot be created without the creation of an equal and opposite charge.

Experiments by Coulomb showed that if a charge, q_1, was placed at a point in empty-space, nothing appears to happen. But if another charge, q, is placed at some other point, as shown in Fig. 1-*a*, it will experience a force in newtons

$$\mathbf{F}_c = \left[k \frac{qq_1}{r_1^2} \right] \mathbf{r}_1$$

where $\mathbf{r}_1$ represents a vector of magnitude 1 (a unit vector which defines the direction) directed from q_1 to q. $\mathbf{r}_1$ is the separation distance in meters. The constant of proportionality, k, is a number that is chosen to make the units work out. Coulomb's constant, k, has the value in a vacuum of

$$k = 9.0 \times 10^9 \ \text{N·m}^2/\text{C}^2$$

The value of k in air is slightly greater. Using the mks system, the constant k can also be written as

$$k = 1/(4\pi\epsilon_0) \ \text{N·m}^2/\text{C}^2$$

which will give familiar units such as volts, ohms, and amperes. ϵ_0 is the permittivity of free space, and is equal to

$$\epsilon_0 = 8.85 \times 10^{-12}$$

The Coulomb force, $\mathbf{F}_c$, on a charge, q, will have a magnitude proportional to the product of the charges, and inversely proportional to the square of the separation distance. That force will also be directed away from q_1. If one of the charges is negative, then the direction will be opposite. That force tends to provide an acceleration, **a**, to q in the same direction. There is, of course, an equal and opposite force on q_1, and Coulomb's law for that is written by simply redefining **r**.

If there is a number, n, of point charges instead of just q_1 present, as shown in Fig. 1-*b*, the force vector of each would all add vectorally to give the total force

$$\mathbf{F}_c = \left[k \frac{qq_1}{r_1^2} \right] \mathbf{r}_1 + \ldots \left[k \frac{qq_n}{r_n^2} \right] \mathbf{r}_n = kq \sum_{k=1}^{n} \left[\frac{1}{r_k^2} \right] \mathbf{r}_k q_k$$

The fact that the vector forces add in this manner is called linear superposition.

If a charge q is spread out over a region of space instead of being located at one point, we consider the charge by dividing it up into an infinite number of infinitesimal charges, dq, and sum the contributions from each. The force that is exerted on a charge q_0 at another point is given by the calculus notation

$$\mathbf{F}_c = kq_0 \int (1/r^2) \mathbf{r} \, dq$$

where the integration symbol $\int$ can be "read" as the sum of an infinite number of infinitesimal contributions.

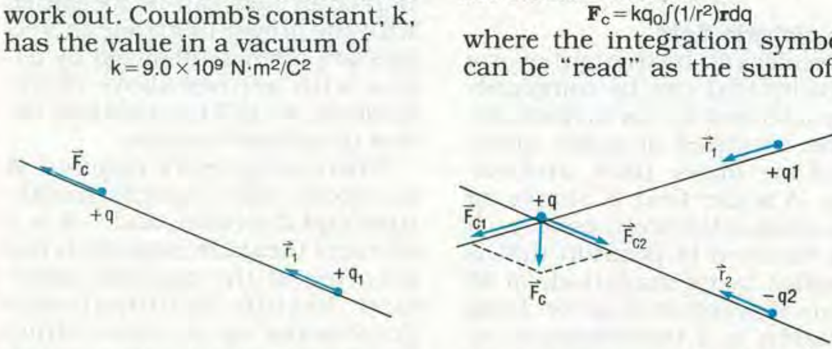


FIG. 1—COULOMB'S EXPERIMENTS showed that a static electric charge produces a force F_c on another charge. A positive-point charge $+q_1$ produces a force on another positive charge $+q$ in the direction of the unit vector r_1 (*a*). A positive charge $+q_1$ produces a force F_{c1} on $+q$ in the direction of the unit vector r_1. A negative charge $-q_2$ produces a force F_{c2} directed opposite to r_2. The total force on $+q$ is the vector sum F_c (*b*).

The electric field E

Coulomb's law defines the force only at one point where q is located. It does not define a field in the sense used here, but it provides a starting point to develop the idea of an electric field. Suppose we make q a very small positive charge and use it as an instrument to explore all points other than where q_1 is located. Since q experiences a force $\mathbf{F}_c$ at every point it is placed, we get the impression that the condition of space is affected by the presence of q_1. We can amend the statement that "if q_1 were alone in space, nothing appears to happen" to "if q_1 were alone in space, then space has the propensity to exert a force on another charge, if it is present, according to Coulomb's law." Since that inclination appears to apply to space, independent of any q, we divide q out of Coulomb's law to obtain a definition of the electric field (also called electric field intensity)

$$\mathbf{E} = \frac{\mathbf{F}_c}{q} = \left[k \frac{q_1}{r_1^2} \right] \mathbf{r}_1$$

which can be thought of as a measure of the propensity. $\mathbf{r}_1$ is a unit vector pointing from q_1 to whatever point in space is being considered and r_1 is the distance. That assigns an **E** vector to every point in space (except at q_1 where $r_1 = 0$).

In the case of a number of point charges, n, the **E** field is obtained by linear superposition

$$\mathbf{E} = k \sum_{k=1}^{n} \left[\frac{1}{r_k^2} \right] \mathbf{r}_k q_k$$

For a spread out charge distribution, summing by integration gives

$$\mathbf{E} = k \int \left[\frac{1}{r^2} \right] \mathbf{r} \, dq$$

Figure 2 illustrates the **E** fields for a number of charge configurations. The Coulomb force on any charge q at a point is just $\mathbf{F}_c = q\mathbf{E}$ where **E** is evaluated at that point.

A small charge q is used to explore the field so that it has a minimal effect upon the object it is measuring. Suppose we let q ap-

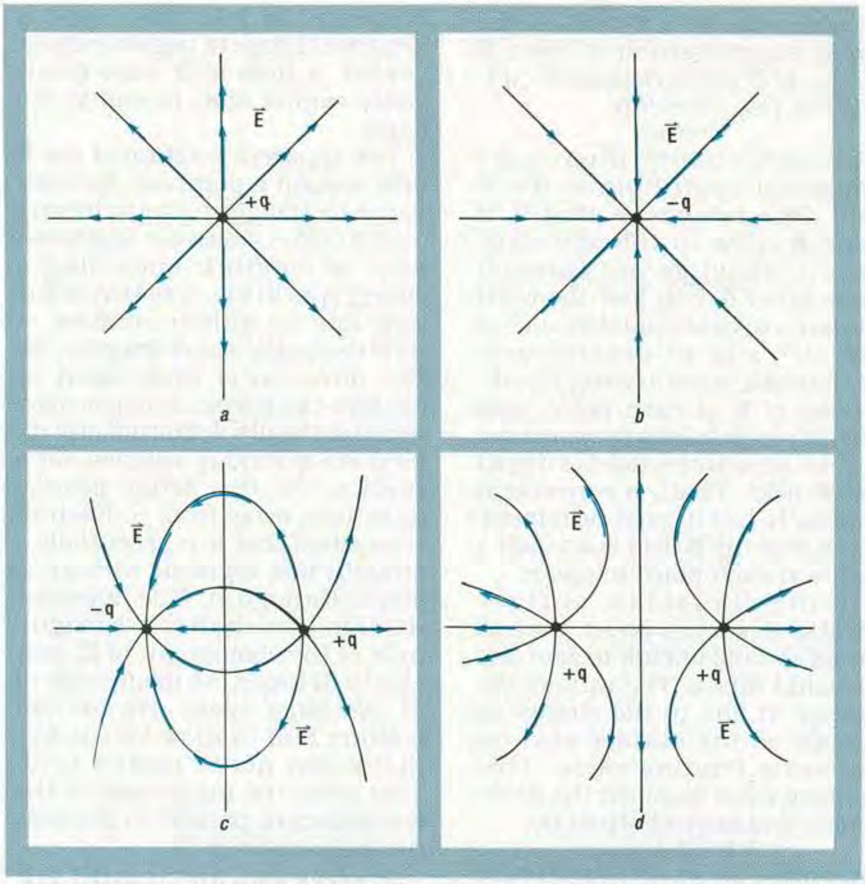


FIG. 2—THE E FIELD IS A RESULT OF the forces between static electric charges. Field vectors are shown in a cross section of a 3-dimensional space for a static positive-point charge (*a*) and for a static negative-point charge (*b*). In (*c*) and (*d*) the E fields for two static charges are shown; the vectors are located at their tail points.

proach 0. In reality, we can't vary the charge continuously since charge appears to come in multiples of *e*, but we can idealize the process. The force felt by that charge will decrease as the charge decreases, but the ratio of the change in force to the change in charge will reach some limiting value. That relationship is written in the calculus notation

$$E = \lim_{\Delta q \to 0} \frac{\Delta F}{\Delta q} = \frac{dF}{dq}$$

Very small positive point charges (so small that their **E** fields can be neglected) can be thought of as ideal devices to explore the **E** field.

Field characteristics

A scaler field, as shown in Fig. 3, can be characterized by the fact that a scaler value can change by a certain amount in a particular direction. In any real field, the values differ little from one point to neighboring points. The gradient of a scaler field is a mathematical operation. It gives a vector that points in the direction for which the value undergoes the largest change, and whose magnitude is that rate of change. The gradient of the scaler field h is symbolized by ∇h. If ∇h equals zero, then the neighboring points must all equal h values. If ∇h is non-zero at a point, then the neighboring points at right angles to ∇h have the same value h.

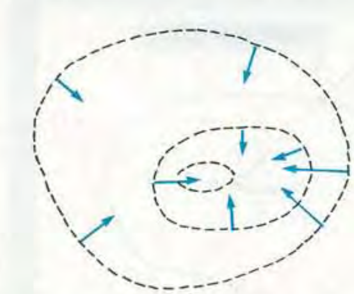


FIG. 3—THE GRADIENT OF A SCALER FIELD is a vector field. The scaler value is the same along each dashed line called an equi-line. Each of the vectors have a magnitude proportional to the greatest rate of change in scaler value per unit distance, and point in the direction of the greatest change. The vectors are perpendicular to the equi-line at their respective points.

For example, imagine standing at a point on a hillside with the height, h, at every point known. ∇h would point in the direction of maximum increase in h, and the maximum decrease would be in the opposite direction, -∇h. If you walked at right angles to ∇h at each point, you would walk along a level or equi-height line. If ∇h equals zero, you would be at a flat spot. ∇h is a vector field since it gives a vector for each point.

Vector fields can be characterized by the fact that they give the impression of flow, as shown in Fig. 4-*a*–*e*. In general, near any point the apparent flow diverges away from (or toward) the point, rotates or curls around a point, or is a combination of both. If the field describes a material, such as air velocity, then there is an actual flow of material.

To measure the apparent flow, or spreading out of the **E** field from a point, imagine an arbitrary closed surface, called a

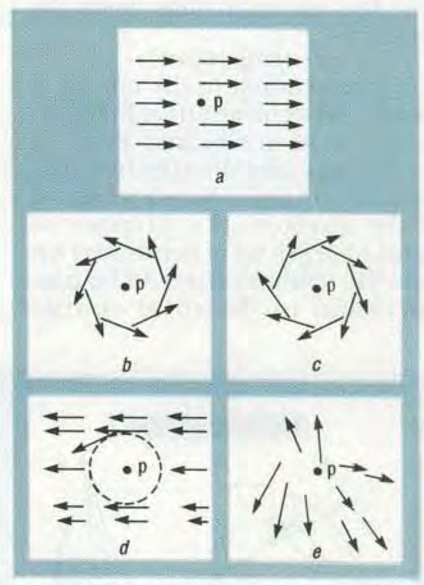


FIG. 4—VECTOR FIELDS GIVE the impression of flow that diverges from, or curls around, an arbitrary point p. Both the divergence and curl of the field are zero in (*a*). There is zero divergence and non-zero curl in (*b*); the curl is a vector out of the page at the point. In (*c*), the direction is reversed, and the vector points into the page. In (*d*) there is zero divergence but non-zero curl since there are non-symmetrical contributions around the closed line. Both the divergence and curl are non-zero in (*e*); these fields could not be static E fields.

Gaussian surface enclosing a charge q, as in Fig. 5-*a*. Divide the surface into an infinite number of infinitesimal surface areas **ds**. Area is a vector because it has a magnitude and also a direction, or orientation in space which is taken as normal (perpendicular) to the surface, and pointing outward away from the enclosed volume. Each infinitesimal area is essentially a small plane with an **E** vector through it. Because the surface is arbitrary, each **ds** and its **E** vector does not have to be parallel. In other words, **E** may not be normal to the plane.

To find the apparent outflow, we need to consider only the component of **E** normal to the plane; the rest is just flowing over the surface. The scaler, or dot product, **E**·**ds**, does that by giving the product of the magnitude of **E** parallel with **ds** times the magnitude of **ds**. That is the same as the product of the magnitude of the effective area (the projected area with **ds** parallel to **E**) times the magnitude of **E**. The apparent flow is electric flux. Summing the contributions from each **ds** over the entire surface gives the total flux

$$\psi = \int \mathbf{E}\cdot d\mathbf{s} \text{ (N/C m}^2\text{)}$$

ψ is proportional to the charge q within the volume since **E** is proportional to q. Because **E** obeys the $1/r^2$ law, and the effective area obeys the r^2 law, ψ is independent of the surface. If a number of point charges were contained inside the volume, ψ would be proportional to the total charge because the total **E** field is the linear superposition of their **E** fields. The proportionality constant is $1/\epsilon_0$, therefore

$$\psi = q/\epsilon_0$$

Charges outside the volume would not contribute to the **E** field. The reason for that is if some **E** came in through some **ds**'s, it would go out through some other **ds**'s in just the right amounts to cancel out because of the $1/r^2$ and r^2 dependence. Graphically, lines having the direction of **E** at each point, and with their closeness proportional to ψ are sometimes used to depict the **E** field. That's a convenient approach, but it must be remembered that the **E** field is actually a vector at each point in space.

If the Gaussian surface shrinks down to a point, then all the ds's would shrink to zero and so would flux ψ. The ratio of the change in flux to the change in volume as the surface shrinks reaches a limiting value. That limiting value is called the divergence, and is symbolized by

$$\nabla\cdot\mathbf{E} = d\psi/d_{volume}$$

That must be proportional to the charge per unit volume

$$dq/d_{volume} = \rho,$$

which is called charge density within the surface, therefore

$$\nabla\cdot\mathbf{E} = \rho/\epsilon_0 \text{ (N/C m)}$$

Since ϵ_0 is a constant and is independent of the volume, the above equation could be written as

$$\nabla\cdot\epsilon_0\mathbf{E} = \rho$$

A number of **E** field instruments (small +q's) scattered around a region, would diverge away from a positive charge (a positive divergence) or converge upon a negative charge (a negative divergence). A field with zero divergence cannot start or end at the point.

The apparent rotation of the **E** field around a point can be measured by imagining an arbitrary closed curve, called an amperean loop, of length **l**, encircling a charge q as in Fig. 5-*b*. Divide the loop into an infinite number of infinitesimally small lengths, **dl**. The direction of **dl** is taken as counter-clock wise. A loop is used because the **dl**'s define general directions around q, whereas for a surface, the **ds**'s define general directions away from q. Each **dl** is so small that it is essentially a straight line segment with an **E** vector through it. The apparent rotation at each **dl** is the magnitude of the component of **E**, parallel to **dl** times the magnitude of **dl**. We must again use the dot product **E**·**dl** to allow for the fact that **E** may not be parallel to **dl**. That gives the magnitude of the **E** component parallel to **dl** times the magnitude of **dl**.

Imagine moving around the loop, summing up **E**·**dl** to obtain the total apparent rotation, or electric circulation. Since **E** points radially along **r**, the only place **E**·**dl** is non-zero is where **dl** has a component parallel to **r**. But the entire loop is closed, so for any amount it moves out radially, it must at some place move that same amount inward. The field is symmetrical, therefore whenever **E**·**dl** is positive along some **dl**'s, it is negative by the same amount along other **dl**'s, with a net result of zero. In calculus notation

$$\oint \mathbf{E}\cdot d\mathbf{l} = 0$$

The circle on the integration symbol reminds us that the loop is closed. Again, by linear superposition, that is true for any static charge configuration.

If the amperean loop shrinks down to a point, all the **dl**'s would shrink to zero, and so would $\int \mathbf{E}\cdot d\mathbf{l}$ (even if it weren't already zero). But the ratio of the change in $\oint \mathbf{E}\cdot d\mathbf{l}$ to the change in the enclosed area as the loop shrinks reaches a limiting value. That limiting value is called the curl, and is symbolized by

$$\nabla\times\mathbf{E} = d(\oint \mathbf{E}\cdot d\mathbf{l})/d_{area}.$$

The curl is a vector, since area is a vector. It's direction is taken as

continued on page 79

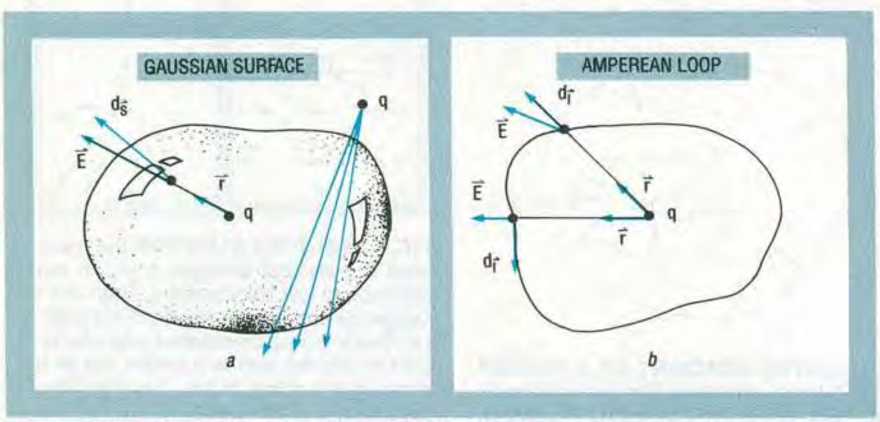


FIG. 5—CHARACTERISTIC OF AN E FIELD. In (*a*) a Gaussian surface composed of an infinite number of infinitesimal areas d_s surrounds a positive charge q. The total apparent flow of the electric field and the electric flux is the sum of $E\cdot d_s$ over the entire surface, which is proportional to q. Flux from charges outside the surface does not contribute because whatever flux "flows" through the surface must also flow back out. In (*b*), an amperean loop composed of an infinite number of infinitesimal lengths, dl, encircles the charge. The electric circulation around the loop $\int E\cdot d_l$ is zero.

continued from page 68

the direction of the extended thumb of the right hand with the fingers wrapped in the general direction taken around the loop. In the case of the static **E** field

$$\oint \mathbf{E} \cdot d\mathbf{l} = 0,$$

therefore

$$\nabla \times \mathbf{E} = 0.$$

The curl is a vector measure of the apparent rotation of the field about a point. If a number of **E** field instruments were scattered around a region, the group would not rotate.

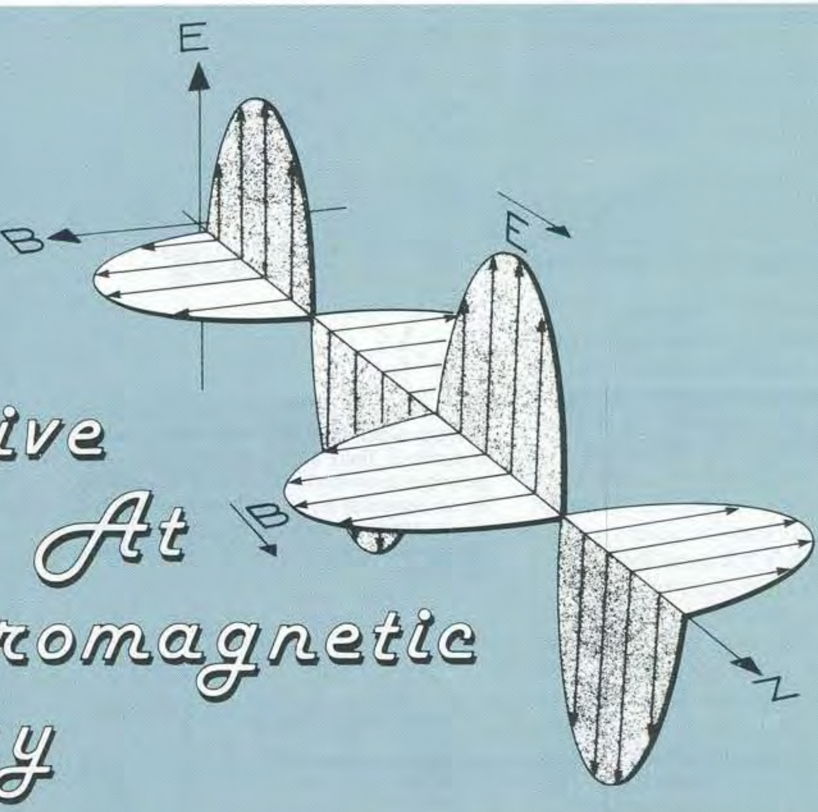
The divergence and curl of the types of fields we're discussing completely characterize the field; **E** can be found if $\nabla \cdot \mathbf{E}$ and $\nabla \times \mathbf{E}$ are known. This is known as Helmholtz's theorem.

The curl of a vector field is always zero, if, and only if the field is the gradient of some scaler field. Consider our h field example. If $\nabla \times \nabla h$ were non-zero, then in following a closed path from some point and back to the beginning, one encounters different rates of change of height times distance when taking different paths. $\oint \nabla h \cdot d\mathbf{l}$ would be path dependent. That would amount to leaving from a point at, for instance, 50 meters in elevation and returning only to find the elevation is 300 meters, or 2 meters, depending upon what path was taken!

The divergence of a field is always zero only if the field is the curl of another field. Imagine the fields of Fig. 4 in 3-dimensional space. Curl the right-hand fingers in the direction of the apparent rotation around the point. The extended thumb is the direction of the curl vector at that point. Conversely, consider the vectors shown as curl vectors. Direct the thumb along them and the fingers will curl in the direction of the field vectors. The field vectors seem to cancel, and not spread out. Those fields are the curl of another vector field. Try that with Figs. 2 or 4 and you'll get conflicting results.

Next time, we'll develop Ohm's law and look at the **E** field in materials, which will provide further insight into Maxwell's equations. R-E

An Intuitive Look At Electromagnetic Theory



WILLIAM P. RICE

LAST TIME WE PRESENTED GENERAL concepts of electric fields and how they are related to static electric charges. We saw that the **E** field in empty space accounts for the forces between such charges. In this article, we'll see how the familiar units of volts and amperes are related to each other. Ohm's law and the concept of an **E** field in materials will be discussed with the help of a simple quantum theory viewpoint.

Potential

To quasi-statically move a charge q from point a to point b in an **E** field, a force that is infinitely close to being equal and opposite to the Coulomb force must be applied to q. That force is $-q\mathbf{E} = -\mathbf{F}_c$, as shown in Fig. 1. As we discussed in our previous article, when moving around a closed path

$$\oint \mathbf{E}\cdot d\mathbf{l} = 0,$$

or

$$\nabla \times \mathbf{E} = 0$$

at all points. So in moving the charge around a closed path

$$-\oint q\mathbf{E}\cdot d\mathbf{l} = 0.$$

The dot product gives the magnitude of force times distance in the direction moved, which is the work done or change in the potential energy ΔU. The energy expended in moving along the path from a to b is just the sum of the contributions along that path, as defined in the calculus notation

$$\Delta U_{ab} = -\int_a^b q\mathbf{E}\cdot d\mathbf{l} \text{ (newton} \times \text{meters = joules).}$$

The energy change is independent of the path taken from point a to b, and the **E** field follows the laws of conservation; whatever energy is expended in moving the charge from point a to b is recovered when the charge moves from b to a. The energy is said to be stored in the **E** field since the field is responsible for the force.

Dividing by the charge gives us the change in energy per unit charge, the potential or voltage at point b with respect to a is

$$V_{ab} = \frac{\Delta U_{ab}}{q} =$$

$$-\int_a^b \mathbf{E}\cdot d\mathbf{l} \text{ (joules / coulomb = volts).}$$

The use of the name potential is perhaps unfortunate because it's easy to confuse the term with potential energy.

Recall also that since $\nabla \times \mathbf{E} = 0$, **E** must be the gradient of a scaler field, which we now see is the potential V, therefore

$$\mathbf{E} = -\nabla V \text{ (volts/meter = newtons/coulomb).}$$

Along a surface of equal potential, there would be no change in V per length d**l**. Perpendicular to that surface the change in V per length would be a maximum, which is what the gradient tells us.

Since the field is obtainable by linear superposition, the potential difference is simply the sum of the potentials. For example, $V_{ac} = V_{ab} + V_{bc}$. That analysis is the basis of Kirchoff's voltage law, which states that the algebraic sum of the voltage rises and drops around a closed path must equal zero.

Electric current

Imagine a Gaussian surface in space through which a number of q charges are moving, as shown in Fig. 2. (We are not concerned with the type of field influencing the motion, only that there is motion.) The current across that surface is defined as the charge per unit time (in seconds) crossing the surface. In order to calculate that, divide the surface into an infinite number of infinitesimal surfaces, d**s**. The charges move with velocity **v** through each surface. If there are n charges per unit volume, then the current density, or charge per unit area is

$$\mathbf{J} = nq\mathbf{v} = \rho\mathbf{v} \text{ (C/m}^2\text{s)}.$$

Multiplying that by the effective area and summing the contributions by integration gives the total current

$$I = \int \mathbf{J}\cdot d\mathbf{s} \text{ (C/s = amperes)}.$$

Positive charges flowing in one direction can be considered equivalent to negative charges flowing in the opposite direction (the Hall effect is a common exception) since both **J** and d**s** would then be negative. That is why a circuit can be analyzed in terms of either conventional currents or electron currents.

The way current is defined is similar to the way we explained electric flux ω except that flux is an apparent flow while current is due to an actual flow of charge. Charge conservation tells us that whatever charge flows into the surface must also flow out unless the current density inside is changing in time. That is the basis of Kirchhoff's current law, which tells us that the sum of the currents flowing into a junction is equal to the sum of the currents flowing out of that junction. Shrinking the Gaussian surface down to a single point and taking the ratio of the rate of change in current to the rate of change in volume gives the divergence

$$\nabla\cdot\mathbf{J} = -\frac{\partial\rho}{\partial t} \text{ (C/m}^3\text{s)}.$$

The partial differential symbol ∂, as in d, means an infinitesimal change in something. It also reminds us that we're only interested in ρ's change with respect to time, t. The negative sign indicates that a decrease in ρ, a negative $\partial\rho/\partial t$, gives a positive divergence. The net charge must therefore flow out through the surface.

Conductivity

Up until this point we have been concerned only with charges in empty space. The space of solid materials, however, is far from empty. Atoms are located at positions called lattice points. An external **E** field applied to a solid material causes the electrons with a $-e$ charge to

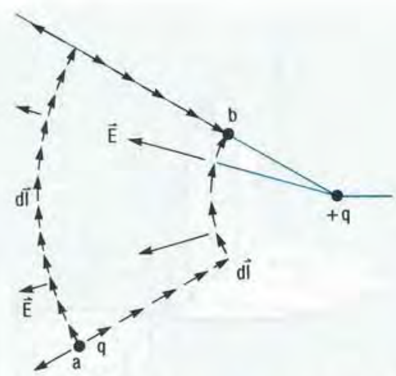


FIG. 1—AN ELECTRIC CHARGE q is moved quasi-statically from point a to b in a static E field along either path, composed of an infinite number of lengths dl, by an external force qE (not shown). The work done or change in energy is the negative of the sum of all the qE·dl's along the path.

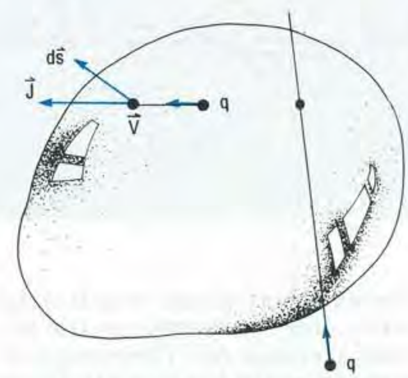


FIG. 2—CURRENT DENSITY J is the number of charges q per unit volume moving with velocity v through an infinitesimal section ds of the Gaussian surface. The total current is found by summing J·ds over the entire surface. Any charge that comes in through one ds must leave through another. Any net outflow must be at the expense of the charge density enclosed by the surface.

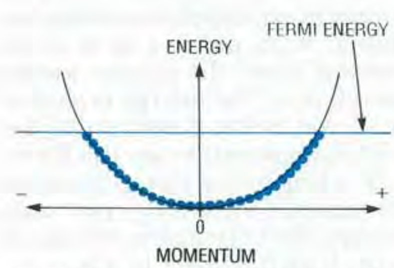


FIG. 3—ENERGY VERSUS MOMENTUM for electrons in a material. Temperature and lattice effects are neglected. Each electron, represented by a dot on the curve, has a unique energy state. Those are the lowest states available. The highest occupied energy is called the Fermi energy.

move. Quantum theory must be used to describe the effects of temperature and the lattice upon the motion of charges.

The electrons are in a state described by their energy, momentum, and spin. No two electrons can be in the same state. Electrons can change energy only by moving to a neighboring unoccupied energy state. Figure 3 shows the energy versus momentum states, neglecting the effects of temperature and the lattice. The two possible spin states for each electron are not shown for clarity.

The more electrons there are in the material, the higher the highest occupied energy state, or Fermi level. Only electrons near the Fermi level can respond to external effects such as thermal energy and electric fields. Supplying thermal energy excites some electrons to energies just above the Fermi level, leaving unoccupied states just below. The Fermi level is then taken as the energy with 50% occupancy. Electrons that can change energy, and hence momentum, are called conduction electrons. Thermally excited electrons have random momentum and velocity, and do not produce a net current.

Electrons act as waves and, therefore, experience interference effects due to interaction with the lattice. At certain wavelengths, standing waves result which produce energy gaps, as shown in Fig. 4. If only some of the energy states up to the gap are occupied or the gap is very small, the material will have many conduction electrons since little external energy is required to excite an electron to a higher state. Such materials are good electrical conductors. A good insulator (or dielectric) has occupied states up to a relatively large gap. A large amount of external energy is required to excite electrons to higher energies in a dielectric material. A material with a large gap and many occupied lower states exhibits noticeable electrical resistance.

If a potential difference is maintained across a material, an electric field is established. Conduction electrons will be subjected to a force **F**, which is equal to $-e\mathbf{E}$. Electrons tend to accelerate, and then "collide" and lose energy to the lattice. If τ is the average time between collisions, which is temperature dependent due to thermal motion of the lat-

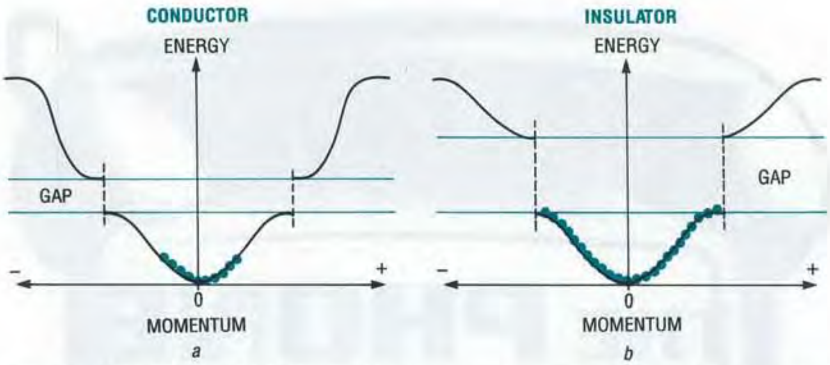


FIG. 4—ENERGY VERSUS MOMENTUM for electrons is a lattice of atoms. The gaps in the curves result from interference effects with the electron waves. In a conductor (*a*) the levels below the gap are partially occupied. External energy excites electrons to the unoccupied energy states. That allows them to participate in an electric current. In an insulator (*b*) the levels below the gap are filled and the energy gaps are large. Electrons cannot participate in a current unless a large amount of external energy is supplied.

tice atoms, then the average electron momentum is

$$\mathbf{F}\tau = -e\mathbf{E}\tau = m + \mathbf{v} \text{ (N·s = kg·m/s)}$$

where m is the electron mass, and **v** is the average velocity. Solving for the velocity and substituting into the equation for current density gives us

$$\mathbf{J} = \frac{ne^2\tau}{m}\mathbf{E}$$

which is the vector form of Ohm's law. Since the number of electrons n and τ are properties of the material, the conductivity

$$\sigma = ne^2\tau/m \text{ (C}^2\text{s/kg} = 1/(\Omega\cdot\text{m}))$$

is a property of the material. The resistivity is defined as $r = 1/\sigma$. If the material is of uniform cross-sectional area S and of length L, **J** is uniform and normal to d**s**, therefore the current is

$$I = JS = \sigma\frac{V}{L}S$$

or V = IR where R = rL/S is resistance in more familiar units of ohms.

In metals, increasing the thermal energy excites electrons mainly into the unoccupied states of the lower band, but the time between lattice collisions decreases. Increasing the temperature increases the resistance. In some other materials resistance decreases with increasing temperature because the number of conduction electrons exceeds the effect of increased collision time.

Due to the low velocity of electrons in most solids, the magnetic effects can be neglected. Conduction becomes more complicated in gases and liquids since the atoms can also move, and velocities can become greater than in solids.

The electric field in materials

When a material is placed in an external electric field $\mathbf{E}_o$, the wave functions of the atoms are changed. The net effect is that the regions with probability of finding electrons are shifted in the $-\mathbf{E}_o$ direction while the regions with probability of finding the positively charged nuclei are shifted in the direction of $+\mathbf{E}_o$ (Fig. 5). The shifts may not exactly align parallel to $\mathbf{E}_o$, and may not all be uniform except in what we call simple materials. A negative surface charge develops on the material near the source of $\mathbf{E}_o$, and a positive surface charge develops on the opposite side. We say the material has an induced charge, or that it is electrically polarized.

The induced charges produce a field $\mathbf{E}_d$ in the opposite direction to $\mathbf{E}_o$ in the material. In a very good conductor, there are enough free charges so that $\mathbf{E}_d$ equals $\mathbf{E}_o$, and the average field inside is zero. That is why metal is an effective shielding material, at least for static fields. Outside the conductor the $\mathbf{E}_o$ field vectors are changed so that they are normal to the surface.

In dielectrics, the large energy gap means the electrons are elastically attached to the lattice and only slight shifts are experienced. $\mathbf{E}_o$ and $\mathbf{E}_d$ don't cancel each other completely. In a simple dielectric, pairs of internal charges, $-q$ and $+q$, are separated by a distance **R** taken in the direction of $\mathbf{E}_o$, from $-q$ to $+q$. Those pairs of negative $-q$ and positive $+q$ charges are called electric dipoles. The vector quantity, q**R**, is called the electric dipole moment. If there are n dipoles per unit volume, then a measure of the polarization can be expressed as

$$\mathbf{P} = n(q\mathbf{R})\zeta \text{ (C·m/m}^3 = \text{C/m}^2\text{)},$$

which is called the dipole moment per unit volume. ζ is a function of the alignment and ranges from 0 to 1. For simple materials $\zeta = 1$. Since n, q, **R**, and ζ depend on the material,

$$\mathbf{P} = \epsilon_o \varkappa \mathbf{E}$$

where $\varkappa$, the electric susceptibility, is a measure of the ease of polarization of the material. $\mathbf{E}_o$ is present to maintain correct units. The so called depolarization field $\mathbf{E}_d$ is equal to $-\gamma\mathbf{P}/\epsilon_o$, where γ is a number between 0 and 1, and is related to the geometry of the material. $\mathbf{E}_d$ is not, in general, very useful.

The surface charge σ_b is an actual accumulation of charges

continued on page 82

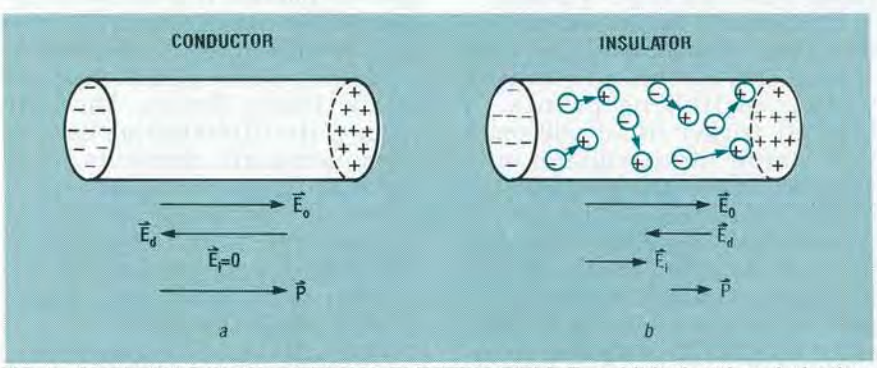


FIG. 5—MATERIALS IN AN EXTERNAL ELECTRIC FIELD E_o exhibit electric polarization. The resulting separation of positive and negative charge regions produce electric dipole moments qR, where q is taken as positive. In a conductor (*a*), enough electrons are free to move to create a depolarization field E_d equal and opposite to E_o. The internal electric field $E_i = E_o - E_d$ is zero. In an insulator or dielectric (*b*), electrons are restricted in movement and E_i is non zero. In both cases, the polarization or dipole moment per unit volume P is related to $-E_d$. The vectors are shown outside the material for clarity.

ELECTROMAGNETIC THEORY

continued from page 59

that are bound directly to the atom and cannot flow. If **N** is of magnitude 1 and is normal to the surface then

$$\sigma_b = \mathbf{P}\cdot\mathbf{N} \text{ (C/m}^2\text{)}.$$

Imagine a Gaussian surface inside the dielectric. With a nonuniform charge distribution some of the bound charges will be displaced across the surface by **P**, leaving a net charge within the surface. In the same manner that we found $\nabla\cdot\mathbf{E} = \rho\epsilon_0$, where ρ is the volume charge density of all the charges contributing to **E**, we can see that the volume charge density in the dielectric is

$$\nabla\cdot\mathbf{P} = -\rho_b \text{ (C/m}^3\text{)}.$$

The negative sign means that the dipole moment per unit volume, **P**, points from negative to positive in the dipoles.

It is customary and convenient to consider a field associated with just the free charge density ρ_f since ρ_b is due to the response of the material. That field must be due to the total charge density less the bound charge density, therefore

$$\rho_f = \rho - \rho_b = \nabla\cdot\epsilon_0\mathbf{E} + \nabla\cdot\mathbf{P} = \nabla\cdot[\epsilon_0\mathbf{E} + \mathbf{P}].$$

The term in brackets is called the displacement field vector

$$\mathbf{D} = \epsilon_0\mathbf{E} + \mathbf{P} \text{ (C/m}^2\text{)}.$$

In simple dielectrics, **P** and **E** are parallel, and the following relation holds true

$$\mathbf{D} = \epsilon_0(1 + \varkappa)\mathbf{E} = \epsilon\mathbf{E}.$$

ϵ_0 can now be interpreted as the ability of empty space to support an electric field, and is called the permittivity of free space. ε is the permittivity of the material. A commonly used quantity is the dielectric constant

$$K = 1 + \varkappa = \epsilon/\epsilon_0.$$

K is greater than 1 for any material, and goes to infinity for a conductor because **E** = O in a conductor. K can be thought of as a measure of the modification of free space by the presence of a material.

From our previous analysis, we have obtained one of Maxwell equations, Gauss' law which reads

$$\nabla\cdot\mathbf{D} = \rho_f.$$

Gauss' law says that the apparent spreading out of the displacement field vector **D** through a Gaussian surface is due to the density of free charges inside. Gauss' law doesn't say, however, that **D** is not producing a swirl. The static **E** contribution can't produce swirling, but the **P** contribution can.

Capacitance

We know that two conductors, separated by a dielectric with dielectric constant *k*, form a capacitor. If one conductor has charge +q and the other −q, the measure of the amount of charge that must be placed on a conductor to change its potential by one volt is called the capacitance, which is in units of coulombs per volt

$$C = q/V \text{ (farads)}.$$

If the free charge q increases, the displacement field vector **D**, which equals the $\epsilon_0 k$ field also increases. That causes a proportionate increase in voltage as **E** rises. Given a particular charge q, the only way to change the capacitance is to change the voltage. That can be done by changing the charge separation distances or by changing the properties of space to give different **E**'s. Simply filling the separation space with a material of greater dielectric constant reduces the **E** field in that space, which reduces the voltage and increases the capacitance.

We can use Gauss' law, without involved calculations, to determine the change in the electric field when any capacitor is filled with a dielectric. In empty space, **P** = O and all the charges are free charges, therefore

$$\nabla\cdot\mathbf{D}/\epsilon_0 = \nabla\cdot\mathbf{E} = \rho_f/\epsilon_0,$$

and

$$\nabla\cdot\mathbf{D}/\epsilon_0 = \nabla\times\mathbf{E} = O.$$

If the space is filled with a simple dielectric, $\mathbf{D} = \epsilon_0 k\mathbf{E}$, therefore

$$\nabla\cdot\mathbf{D}/\epsilon_0 = \nabla\cdot k\mathbf{E} = \rho_f/\epsilon_0.$$

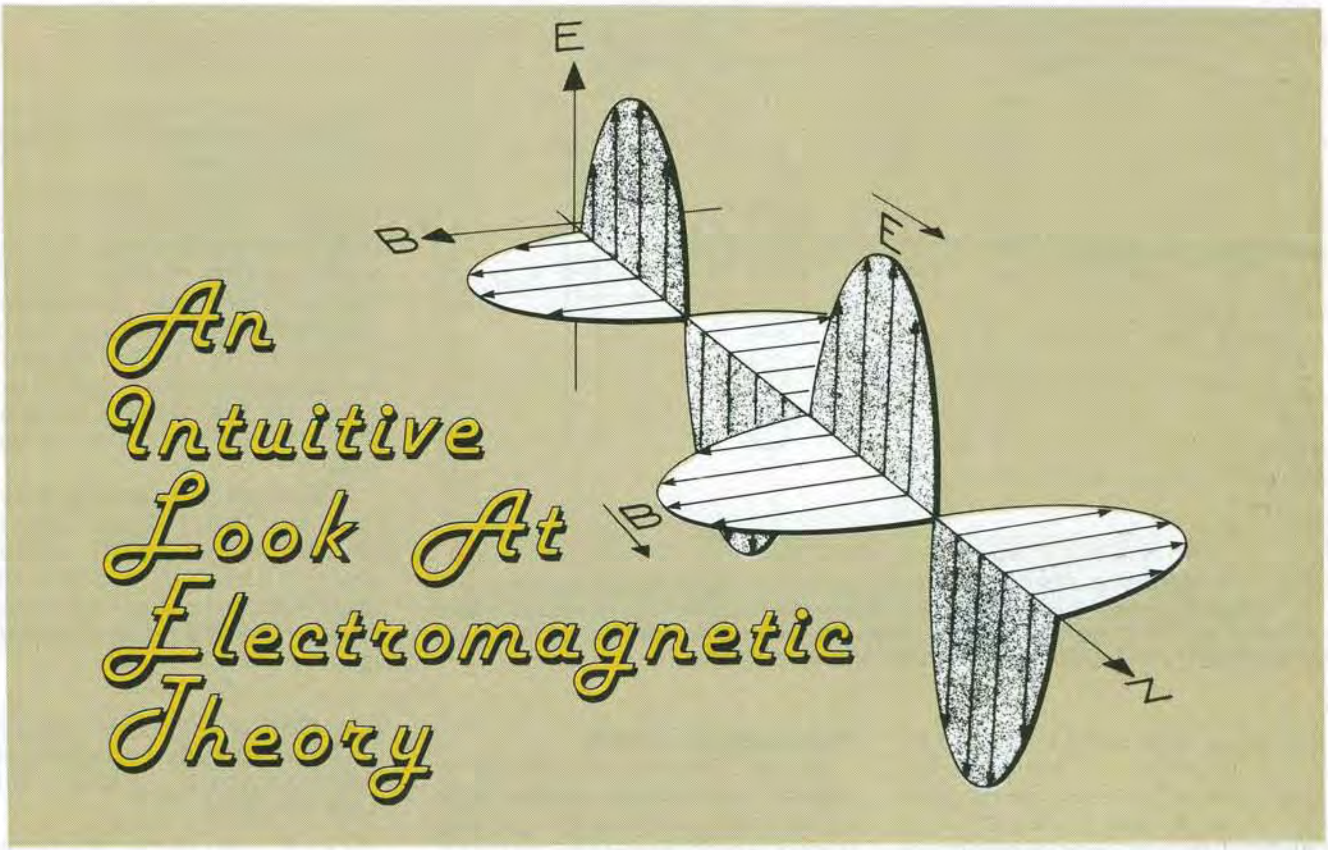
P is aligned with **E** so there is no apparent rotation and

$$\nabla\times\mathbf{D}/\epsilon_0 = \nabla\times k\mathbf{E} = O.$$

The divergence and curl of **E** completely characterize the field. By comparison, the **E** for a charged capacitor with empty space as a dielectric is the same as *k***E** for the same charged capacitor with a dielectric constant *k*. In a capacitor filled with a dielectric, **E** is reduced by 1/*k*. The capacitance C = q/V is increased by *k* since the voltage potential V is reduced by 1/*k*.

In our next edition, we'll look at the effects of electric charges in motion. We'll see that another type of field, the **B** field, is required to describe the magnetic forces associated with them. **R-E**

An Intuitive Look At Electromagnetic Theory



Learn the meaning behind magnetic charges and "B" fields.

WILLIAM P. RICE

IN OUR LAST ARTICLE, WE DISCUSSED the general concepts of an electric field and how they applied to forces between static electric charges. We'll now develop an intuitive picture of how charges moving with a constant velocity produce an additional force, and how that force leads to the concept of a magnetic field.

Magnetic "charges"

Early experiments showed that if two permanent magnets were near each other, each experienced a force. In each magnet there appears to be two regions, called the north and south poles, that contain the source of the force. A pole of one magnet attracts the opposite pole of the other magnet but repels the other pole, thereby creating a torque. Apparently, a magnet produces something similar, but not identical, to that produced by an electric-charge distribution. A basic difference is that electric charge distribution can be separated into two distinct regions of positive and negative charge, while experiments show that cutting a magnet into smaller and smaller pieces simply result in more magnets, each having two poles. No matter how small a piece is taken, there's always an equal amount of north and south magnetic "charge." Experiments have shown that there's no such thing as an isolated single magnetic charge, or a magnetic monopole, only magnetic dipoles.

Hans Christian Oersted conducted experiments which showed that a permanent magnet near a conductor carrying a constant electric current I_1 experienced a similar force as shown in Fig. 1-*a*. Experiments by French physicist Andre Marie Ampere showed that when a conductor carries a constant current I along an infinitesimal length **dl** and another conductor caries a constant current I_1 along an infinitesimal length $\mathbf{dl}_1$, the length **dl** experiences an infinitesimal force in newtons

$$dF_m = k_m \frac{Idl \times (I_1\, dl_1 \times r_1)}{r_1^2}$$

as shown in Fig. 1-*b*. $\mathbf{r}_1$ is a unit vector directed from $\mathbf{dl}_1$ to **dl** and r_1 is the separation distance. In the mks units, k_m is equal to

$$k_m = \mu_0/4\pi \text{ (webers/(ampere} \times \text{meter).}$$

Idl and I_1 $\mathbf{dl}_1$, in units of m·C/s = A·m, are infinitesimal lengths of positive current in the direction of **dl** and $\mathbf{dl}_1$. In conductors, there are equal distributions of positive and negative electric charges even though the negative charges are moving. The **E** fields from the charges must sum to zero, so the $\mathbf{dF}_m$ must be distinct from the Coulomb force $\mathbf{F}_c$.

The infinitesimal force equation mentioned above is more complicated than for the static electric force since the direction of charge motion must be taken into account by vector multiplication, also known as the cross product ×. The direction of $I_1\ \mathbf{dl}_1 \times \mathbf{r}_1$ is defined by the right hand rule: curl the fingers of the right hand through the smallest angle from the vector $I_1\ \mathbf{dl}_1$ to the vector $\mathbf{r}_1$; the extended thumb points in the direction of $I_1\ \mathbf{dl}_1 \times \mathbf{r}_1$. The magnitude of $I_1\ \mathbf{dl}_1 \times \mathbf{r}_1$ is the area of a paral-

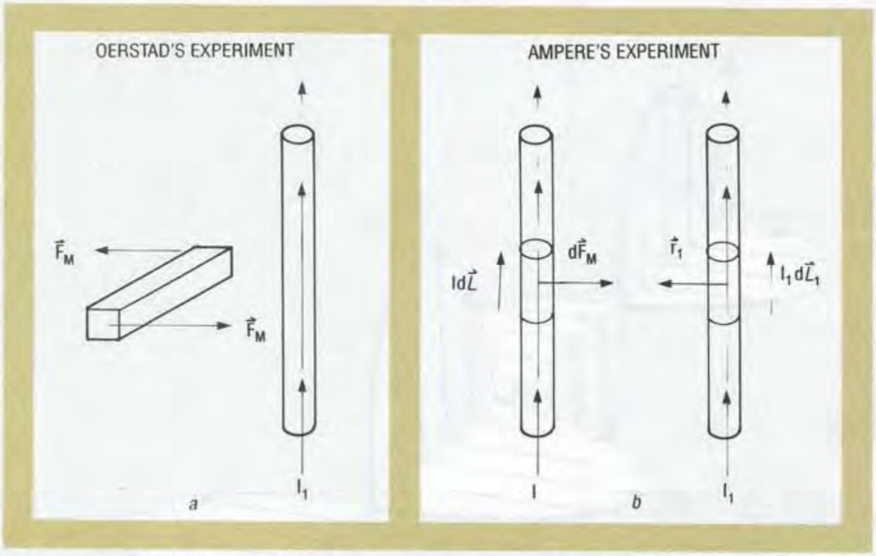


FIG. 1—A MAGNETIC FORCE F_m is produced by a constant electric current I_1 in a conductor. A permanent magnet experiences a torque due to the force on each pole (*a*). In (*b*) a small current segment I dl of another conductor experiences a force due to the segment I_1 dl_1. I_1 dl_1 experiences an equal and opposite force.

lelogram with sides I_1 $\mathbf{dl}_1$ and $\mathbf{r}_1$. That is similar to scalar multiplication where A times B gives the area of the rectangle with sides A and B. The direction of $\mathbf{dF}_m$ is that of the extended right hand thumb with the fingers wrapped through the smallest angle from I **dl** to I_1 $\mathbf{dl}_1 \times \dot{\mathbf{r}}_1$.

In Fig. 1-*b*, the current segment I **dl** experiences a force towards I_1 $\mathbf{dl}_1$. I_1 $\mathbf{dl}_1$ experiences an equal and opposite force towards I **dl**. For other current segments, the force on an I **dl** is not equal and opposite to that on an I_1 $\mathbf{dl}_1$. That may appear to be a violation of Newton's third law, however the actual constant currents exist only in closed loops or circuits as dictated by charge conservation. The I **dl** and I_1 $\mathbf{dl}_1$ are only a part of each loop. The total force is found by summing up all the infinitesimal contributions around each closed loop. We must sum twice by integration, first to find the forces of all the I_1 $\mathbf{dl}_1$'s on an I **dl**, and then to sum the forces on each I **dl**

$$F_m = \frac{\mu_0}{4p}\oint\oint \frac{I Dl \times (I_1\, dl_1 \times r_1)}{r_1^2}.$$

The force on the entire loop composed of I **dl** is always equal and opposite to that on the entire loop composed of I_1 $\mathbf{dl}_1$.

Ampere went on to suggest that in a permanent magnet, the force $\mathbf{F}_m$ is produced by some sort of closed current loops that exist in the material.

The magnetic field B

Figure 2 shows that the space around a constant current segment I_1 $\mathbf{dl}_1$ can be explored using a very small constant current loop obtained by adding all its I **dl** contributions and symbolized by $\oint$ I **dl**. Since each I **dl** will experience a force due to the presence of I_1 $\mathbf{dl}_1$ even though nothing material connects them, one has the impression that the condition of space itself is affected by the presence of the I_1 $\mathbf{dl}_1$. We can say that a constant current gives space the propensity to exert a force on another constant current, if it were present, according to Ampere's force law.

To find that propensity, we remove I **dl** from the force law to obtain the definition of the magnetic field (also called magnetic flux density) in units of webers/meter², which equals the tesla

$$dB = \frac{\mu_0}{4p}\ \frac{I_1\ dl_1 \times r_1}{r_1^2}.$$

This is called the Biot-Savart law. The force on each I **dl** is $\mathbf{dF}_m$, which equals I $\mathbf{dl} \times \mathbf{B}$. Opposite sides of the loop will experience forces in the opposite direction since the **dl**'s are in opposite directions. The loop will, therefore, experience a torque. Since I_1 $\mathbf{dl}_1$ exists only as a part of a closed loop, the total **B** at any point in space is

$$B = \frac{\mu_0}{4p}\oint \frac{I_1\ dl_1 \times r_1}{r_1^2}.$$

Any current loop is called a magnetic dipole because it results in a **B** field.

The magnetic-field test instrument must be a very small magnetic dipole, just as a very small positive charge +q is the electric-field test instrument. The distinction is that $\oint$ I **dl** is a sum of all the vectors for which magnitudes and directions must be taken into account, whereas +q has only a magnitude.

If a current I is considered as just an individual electric charge q, moving with constant velocity through a point, the magnetic force it would experience in the **B** field at that point is

$$\mathbf{F}_m = q\mathbf{v} \times \mathbf{B}.$$

If an electric field is also present, q would experience an additional electric force $\mathbf{F}_c$, and the total force would be

$$\mathbf{F} = \mathbf{F}_c + \mathbf{F}_m = q\mathbf{E} + q\mathbf{v} \times \mathbf{B} = q[\mathbf{E} + (\mathbf{v} \times \mathbf{B})].$$

This equation is known as the Lorentz force law.

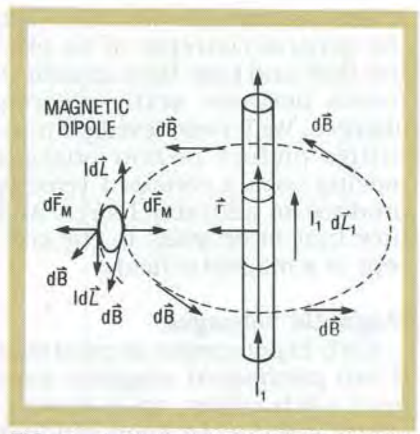


FIG. 2—A SMALL MAGNETIC DIPOLE, composed of current segments I dl, near a current-carrying conductor experiences a torque due to opposite forces dF_m on opposite sides. Since nothing material is pushing the dipole, the concept of the B field is used to account for the forces.

B field characteristics

The apparent flow of the **B** field from an infinitesimal volume about a point can be found by the same method used to find the electric flux. Imagine a Gaussian surface around a current loop composed of an infinite number of current segments I **dl** each producing a **dB** field as shown in Fig. 3-*a*. Divide the surface into an infinite number of infinitesimal **ds** areas. Through each **ds** there are an infinite number of **dB**'s. The total **B** field at each **ds** is $\mathbf{B} = \int d\mathbf{B}$ by linear superposition. Taking $\mathbf{B} \cdot d\mathbf{s}$ gives the magnitude of **B** times the magnitude of the effective area parallel to **B**. That is the apparent flow of **B** through **ds**. Summing those factors by integration over the entire surface gives the total apparent flow, or magnetic flux

$$\phi = \int \mathbf{B} \cdot d\mathbf{s} \text{ (webers).}$$

Imagine moving over the surface, adding up the $d\mathbf{B} \cdot d\mathbf{s}$ contributions from each I **dl**. At each **ds**, **r** points from I **dl** towards **ds**. Since **dB** is perpendicular to **r**, the only place $d\mathbf{B} \cdot d\mathbf{s}$ is non-zero is where **ds** is not directed along **r**. That is where **ds** moves away from or toward I **dl**. Since the surface is closed, for each place we move away from I **dl** by a certain amount and direction, there must be another place that we move back in towards I **dl** by the same amount and in opposite direction. Whatever $\mathbf{B} \cdot d\mathbf{s}$ contribution is found over some of the surface is canceled by a $-\mathbf{B} \cdot d\mathbf{s}$ contribution over another part of the surface, therefore we can say that

$$\phi = 0.$$

As in electric flux, any **B** produced by currents outside the surface will not contribute to the total.

If the original Gaussian surface is shrunk so the volume enclosed approaches zero, the ratio of the change in flux to the change in volume would reach a limiting value even if the flux were not zero. That is the divergence of **B**, and since $\phi = 0$ for any Gaussian surface

$$\nabla \cdot \mathbf{B} = 0 \text{ (T/m}^3\text{).}$$

That is the unnamed Maxwell equation. It simply says that the total spreading out, or divergence, of the **B** field through an infinitesimal closed surface about any point is zero. Whatever **B** field appears to leave from a particular point must return to that same point. Magnetic monopoles, therefore, cannot exist. That relationship allows magnetic dipoles, which produce equal amounts of outward and inward magnetic flux from a point. If a number of our **B**-field instruments were scattered about the point, they would not spread out.

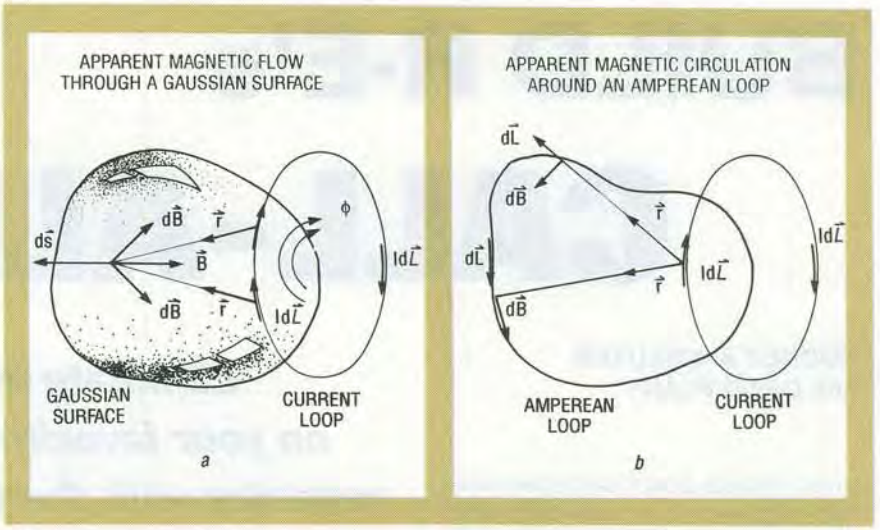


FIG. 3—CHARACTERISTICS OF A B FIELD. In (*a*), a Gaussian surface composed of an infinite number of infinitesimal ds areas surrounding part of a current loop. The total apparent flow of B, the magnetic flux $\int \mathbf{B} \cdot d\mathbf{s}$, through the surface is zero. Flux from currents outside the surface does not contribute since whatever flux "flows" in through the surface also flows back out. In (*b*), an amperean loop composed of an infinite number of infinitesimal dL lengths encircles a current segment. The magnetic circulation around the loop $\oint \mathbf{B} \cdot d\mathbf{L}$ is proportional to the current encircled. For currents not encircled, $\oint \mathbf{B} \cdot d\mathbf{L}$ is zero.

The apparent rotation of the **B** field around an infinitesimal area containing a point can be found by imagining an amperean loop about some current loop as shown in Fig. 3-*b*. Divide the amperean loop into an infinite number of infinitesimal lengths **dL**. The **B** field at each **dL** is again just the sum of each of the infinitesimal **dB** contributions from each I **dl**, where $B = \int d\mathbf{B}$. The magnetic circulation around the loop is proportional to the current encircled.

If you take $\mathbf{B} \cdot d\mathbf{L}$ you get the magnitude of **B** times the magnitude of the effective length parallel to **B**, which is the apparent flow along **dL**. The direction of **dL** is taken as the direction of the curled fingers of the right hand with the extended thumb pointing in the direction of I **dl**. The total apparent rotation, also called the magnetic circulation, around the amperean loop is found by adding those parts by integration over the entire closed loop $\oint \mathbf{B} \cdot d\mathrm{L}$.

Imagine moving along the loop, in the direction of **dL**, adding up the $d\mathbf{B} \cdot d\mathrm{L}$'s. **r** points from I **dl** to **dL**. When we move at right angles to **dB**, that is along **r** or I **dl**, where $d\mathbf{B} \cdot d\mathbf{L}$ is zero. At all other places there will be a non-negative contribution since we are always traveling in one direction around the loop. The contributions are proportional to the current I through the loop since $\mathbf{B} = \int d\mathbf{B}$ is proportional to that current. The proportionality constant is μ_0, so

$$\oint \mathbf{B} \cdot d\mathbf{L} = \mu_0 I \text{ (T·m).}$$

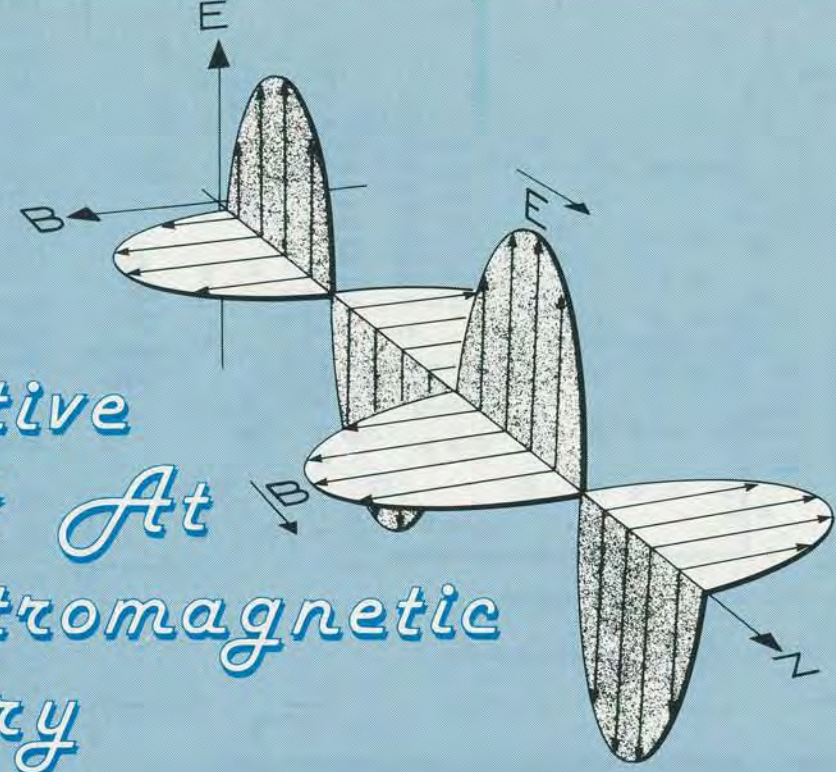
If the amperean loop is shrunk so the area enclosed approaches zero, the ratio of the change in circulation to the change in area reaches a limiting value. That is the curl of **B**, which must be proportional to the current per unit area **J** through the loop, therefore

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{J} \text{ (T/m}^3\text{).}$$

That relationship is called Ampere's law for constant currents. It simply says that the total apparent rotation, or curl of **B**, around any point is proportional to the constant current density at that point. The right hand rule gives the direction of apparent rotation. If a number of the **B**-field instruments were scattered about a point, they would rotate.

Next time, we'll discuss some magnetic phenomena and how inductance is related to the magnetic field. The concept of a magnetic circuit will be developed based on an analogy to the electric circuit. We'll see that in matter, the magnetic field can be considered as the linear superposition of two fields, similar to what was shown with the electric field. **R-E**

An Intuitive Look At Electromagnetic Theory



Find out more about magnetic phenomena and how inductance is related to the magnetic field.

WILLIAM P. RICE

IN OUR LAST EDITION, WE DISCUSSED the characteristics of a static magnetic field in empty space. In this article we'll look further into the **B** field and its effects on matter. Of particular importance, we will show that the magnetic field in matter can be found by using the linear superposition of free and bound current densities.

Potential

If you recall, the expression $\nabla \times \mathbf{B} = \mu_0 \mathbf{J}$ says that the apparent rotation of the **B** field around a small region about a point is proportional to the current density in that region. Unless the current density or charge per unit area **J** is zero, **B** cannot be the gradient of a scalar potential and therefore is not a conservative field. However, in regions that have no current flow, $\nabla \times \mathbf{B} = 0$. In that case, the field is conservative and a scalar potential can be defined. Suppose a small current loop, the **B**-field instrument $\oint I d\mathbf{l}$, is moved quasistatically from point A to B in such a region as shown in Fig. 1. The force in the direction of motion d**L** gives the work done or change in magnetic potential energy

$$\Delta U_{ab} = -\int_a^b \left[(\oint I d\mathbf{l}) \times \mathbf{B} \right] \cdot d\mathbf{l} \ .$$

The work depends not only on the path taken but on the orientation of $\oint I d\mathbf{l}$ along the path. No work is done if $(\oint I d\mathbf{l}) \times \mathbf{B}$ is always perpendicular to d**L**. Work is done if, at any place along the path, $\oint I d\mathbf{l}$ is rotated so that $(\oint I d\mathbf{l}) \times \mathbf{B}$ has some component parallel to d**L**. That is the mechanical energy due to the work done against the torque.

Additional energy is required to maintain the current I in the loop. If the loop has resistance R, then I^2R is the rate of thermal energy loss. That energy must come from someplace, and if the magnetic field enclosed by the loop changes, more energy is required. We'll discuss the reason why additional energy is required in our next article.

Previously, we saw that any field with zero divergence is the curl of some other field. Since $\nabla \cdot \mathbf{B} = 0$, it must be that $\mathbf{B} = \nabla \times \mathbf{A}$. The **A** field is called the magnetic vector potential. It is not an energy field (energy is a scalar quantity), but it can be used in energy calculations. The main advantage in using the **A** field is that calculations required to solve many real-world problems are simplified. Since we won't be doing any calculations here, we will just say that the **A** field is real in the same sense as the **B** field.

We can use the analogy that the **A** field describes action at a distance from the **B** field just as the **B** field describes action at a distance from a current loop. The **E** field is also used to describe action at a distance from an electric charge. An appropriate instrument can be placed in a region of an **A** field, even through the **E** and **B** fields are

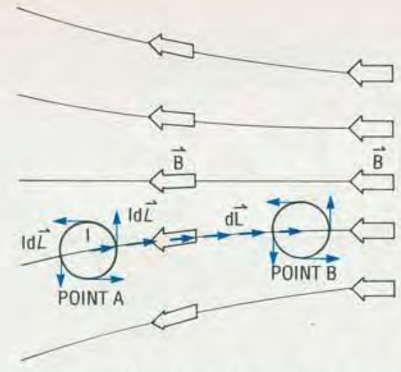


FIG. 1—A MAGNETIC DIPOLE IN A B FIELD is moved from point A to B along the path composed of d_L. The force vector on any small segment of the current loop is $dF_m = Idl \times B$. dF_m is directed out of the page as is the total force $F = \oint Idl \times B$. The force vector is perpendicular to dL, so the work done on $F \cdot d_L$ is zero. If the dipole is rotated so that F was not normal to the paper, then work would be done.

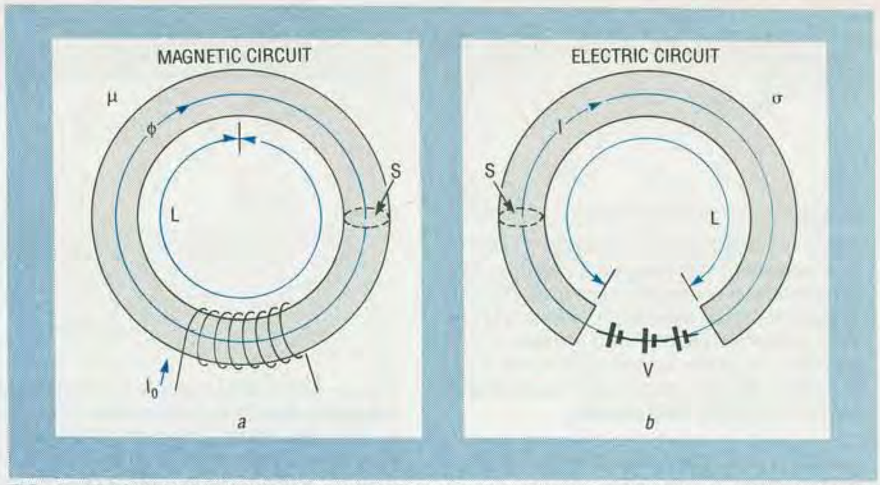


FIG. 2—MAGNETIC FLUX IS ANALOGOUS TO ELECTRIC CURRENT. In (*a*) the magnetic path of length L and cross-sectional area S is in a material of permeability μ. The source of magnetomotive force nI_o is the current I_o encircling the material n times. In (*b*) the electrical path is in a material of conductivity σ. The source of electromotive force nV is a battery of n cells each with a voltage V.

zero there, and an influence can be measured. The Bohm-Aharanov effect is an example.

Magnetic "current"

Recall that $\nabla \cdot \mathbf{B} = 0$ says that the lines of magnetic flux are closed lines. Nothing material flows along these lines but we can make an analogy with the closed path of a constant electric current. The magnitude of **B** in the magnetic circuit of Fig. 2–*a* can be found from $\oint \mathbf{B} \cdot dL = \mu I$, where L is the total length of the magnetic path, μ describes a property of the path material to be discussed later, and I is the total electric current enclosing the path. There are n turns of wire each carrying current I_0 so $I = nI_0$. Since the material is uniform, the magnitude of **B** must be independent when d**L** is being summed. So, denoting the magnitude of **B** as B and summing by integration gives

$$BL = \mu n I_0.$$

The magnetic flux is

$$\phi = \int \mathbf{B} \cdot d\mathbf{s}$$

where **s** is the cross-sectional area of the path. Since the area is uniform

$$\phi = BS = \frac{nI_o}{L/\mu S}.$$

In the circuit shown in Fig. 2–*b*, a current I exists in a material of length L, conductivity σ, and cross-sectional area S. The voltage is supplied by n cells, consisting of V volts each. From Ohm's law

$$I = \frac{nV}{L/\sigma S} = \frac{nV}{R}.$$

The so-called magnetomotive force nI_0 can be compared to the voltage nV. The magnetomotive force is summed in the same way voltages are summed. μ is similar to σ, which suggests that L/μS is a magnetic resistance R_M, called reluctance. Those facts, along with the motivating fact that electric current and magnetic flux form closed paths (implying a conservation of something), allow analogous magnetic circuit equations to be developed.

Magnetic field in materials

In any material there are small current loops or magnetic dipoles formed by the atomic-scale rotational and orbital motions of the electrons and charges in the nuclei, as shown in Fig. 3. The vector quantity I**s** (where **s** is the area of each atomic-current loop), is the magnetic dipole moment. Normally the magnetic dipole moments have random orientations, so no average or macroscopic magnetic field is present.

When a material is placed in an external magnetic field $\mathbf{B}_o$, the quantum-wave functions are changed in such a way that there is a higher probability of the magnetic dipole moments being aligned antiparallel to the $\mathbf{B}_o$, as shown in Fig. 4–*a*. The directions may not all exactly align and may not be uniform except in what we call simple magnetic materials. The net effect is that magnetic poles appear at the ends of the material. We say the material has an induced magnetic field, a magnetic polarization, or simply that it is magnetized. This induced magnetic field is called the demagnetization field $\mathbf{B}_d$. The total magnetic field in the material is $\mathbf{B}_i = \mathbf{B}_o + \mathbf{B}_d$. $\mathbf{B}_d$ is antiparallel to $\mathbf{B}_o$ so $\mathbf{B}_i$ has a smaller magnitude than $\mathbf{B}_o$. Such a material exhibiting those characteristics is called diamagnetic.

In some materials there are additional magnetic dipoles resulting from electrons with unpaired spins. Their magnetic dipole moments are normally oriented randomly. When placed in an external magnetic field, the wave functions are changed in such a way that there is a higher probability of the magnetic dipole moments being aligned parallel to the $\mathbf{B}_o$ as shown in Fig.–4*b*. $\mathbf{B}_d$ is aligned parallel to $\mathbf{B}_o$, so $\mathbf{B}_i$ has greater magnitude than $\mathbf{B}_o$. A material exhibiting those characteristics is called paramagnetic.

In many materials, when the external $\mathbf{B}_o$ field is removed, the wave functions return to their original form within a short time and $\mathbf{B}_d$ becomes zero. However, in ferromagnetic materials the wave functions don't return completely and in some regions, called magnetic domains, residual alignment remains. It is as if each domain supplies a $\mathbf{B}_o$ to all other domains, thus maintaining some $\mathbf{B}_i$ in each.

$\mathbf{B}_d$ is not a particularly useful quantity. If there are n magnetic dipoles per unit volume, then a

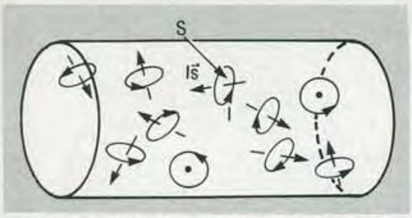


FIG. 3—ATOMIC-SCALE CURRENT LOOPS in a material form magnetic dipoles. The magnitude and direction are given by the magnetic dipole moment Is, where s is the area enclosed by the loop current I. The direction is given by the right-hand rule. Normally, the directions are random and no net magnetic field results.

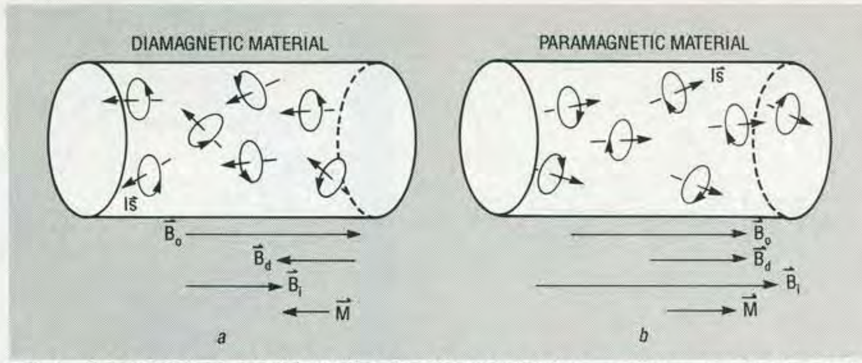


FIG. 4—MATERIALS IN AN EXTERNAL MAGNETIC FIELD B_o exhibit magnetization. In (*a*), magnetic dipole moments tend to align antiparallel to B_o. Demagnetization B_d opposes B_o and the internal magnetic field B_i is smaller in magnitude than B_o. In (*b*), the dipole moments tend to align parallel to B_o due to unpaired electrons. B_i is greater in magnitude than B_o. In both cases the magnetization per unit volume M is related to B_d. The vectors are shown outside of the material for clarity.

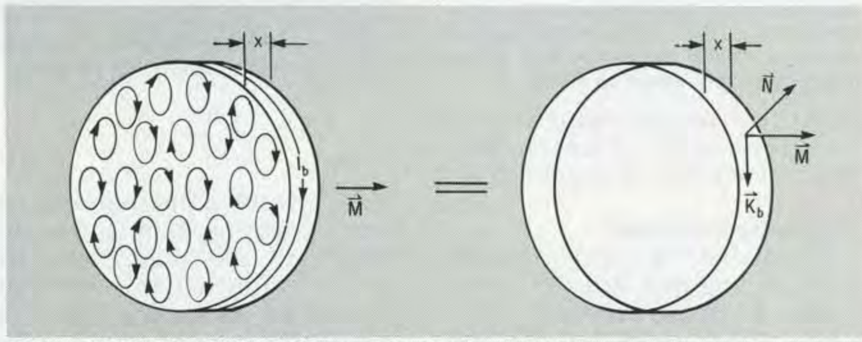


FIG. 5—ELECTRIC CURRENTS associated with individual magnetic dipoles cancel inside the material. At the surface, however, the currents are in the same direction resulting in a net surface current I_b. I_b is bound to the surface since it consists of pieces of the dipole currents bound in the material.

measure of the total magnetic polarization is

$$\mathbf{M} = n(I\mathbf{s})\zeta \text{ (A/m)}$$

called the magnetic dipole moment per unit volume (or just magnetization). ζ is a function of the average alignment of the dipoles with the external field and takes on values from -1 for total antiparallel alignment to $+1$ for total parallel alignment. $\mathbf{B}_d$ and **M** are related by a factor that takes into account properties of the material.

We can use the idea of Ampere's law, which says the apparent rotation of a magnetic field around a small region is proportional to the current per unit area in that region, to account for the **M** field. On an average, the atomic-scale magnetic-dipole currents cancel everywhere in a material except at the surface, as shown in Fig. 5. **M** can therefore be attributed to a bound surface current I_b around an area of magnitude S in a material of length x. The magnitude of **M** is simply the magnetic dipole moment per unit volume as illustrated by

$$I_bS/(xS) = I_b/x.$$

It's sometimes convenient to define a lineal-surface current density as

$$\mathbf{K}_b = \mathbf{M} \times \mathbf{N} \text{ (A/m)}$$

where **N** is a unit vector normal to the surface. The curl of **M** is found the same way Ampere's law for static currents was derived, except the current density of concern is the average atomic-scale volume current density bound in the material $\mathbf{J}_b$. That gives us the formula:

$$\nabla \times \mathbf{M} = \mathbf{J}_b \text{ (A/m}^3\text{)}.$$

A convenient way to separate the external and internal contributions is to consider the total current density **J** as a linear superposition of $\mathbf{J}_b$ due to the material and all other currents called the free current density $\mathbf{J}_f$. From Ampere's law, it can then be concluded that

$$\mathbf{J}_f = \mathbf{J} - \mathbf{J}_b = \frac{1}{\mu_o}(\nabla \times \mathbf{B}) - \nabla \times \mathbf{M} = \nabla \times \left[\frac{1}{\mu_o}\mathbf{B} - \mathbf{M}\right].$$

The term in brackets is called the magnetic-field intensity or just the magnetic field (not to be confused with the **B** field)

$$\mathbf{H} = \frac{1}{\mu_o}\mathbf{B} - \mathbf{M}.$$

In simple materials, **B** and **M** are along the same line so $\mathbf{B} = \mu_0(1 + \varkappa_m)$ and $\mathbf{H} = \mu\mathbf{H}$. $\varkappa_m$ is called the magnetic susceptibility and μ is the magnetic permeability of the material. A commonly used quantity is the relative permeability which can be written as

$$\mu_r = 1 + \varkappa_m = \mu/\mu_0.$$

μ_r is less than 1 for diamagnetic materials and greater than 1 for paramagnetic materials. In ferromagnetic materials, μ_r is very large but the **H** and **M** relationship is generally more complicated and μ_r is not a simple constant.

Ampere's law now says

$$\nabla \times \vec{H} = \vec{J}_f$$

This says that the apparent rotation of the H field around a small region is due to the density of free current through that region. One of Maxwell's great contribuiton was the modification of Ampere's law.

Inductance

We know that a conductive loop, enclosing empty space or some material, forms an inductor. If the loop is carrying a constant current I, then a proportional magnetic flux exists through the area **s** enclosed by the loop. The constant of proportionality is the inductance, in units of webers per ampere, or henrys

continued on page 87

ELECTROMAGNETICS

continued from page 69

$$L = \Phi/I = (\int \mathbf{B} \cdot d\mathbf{s})/I \text{ (H)}.$$

Since **B** may not be constant across the area, we sum each infinitesimal contribution by integration. Note that we're concerned with the flux through the enclosed area, not the total flux through a Gaussian surface enclosing the loop, which is zero. For simple materials, L is independent of I since the equation $\mathbf{B} = \mu\mathbf{H}$ is proportional to I. However, L is dependent upon the area since the equation $\int \mathbf{B} \cdot d\mathbf{s}$ depends on the total area being summed. The inductance (L) is also dependent on μ.

We can use Ampere's law to see that effect. In empty space, $\mathbf{M} = 0$ and there are no bound currents, so we can say

$$\nabla \times \mathbf{H} = \nabla \times \mathbf{B}/\mu_0 = \mathbf{J}_f$$

and

$$\nabla \cdot \mathbf{H} = \nabla \cdot \mathbf{B} = 0.$$

With a simple material filling space, $\mathbf{H} = \mathbf{B}$/mu, so

$$\nabla \times \mathbf{H} = \nabla \times \mathbf{B}/\mu = \mathbf{J}_f$$

and

$$\nabla \cdot \mathbf{H} = \nabla \cdot \mathbf{B}/\mu = 0.$$

Since the divergence and curl of the field completely characterize the fields, **B** is larger by $\mu/\mu_0 = \mu_r$ in a filled inductor.

In our next article, we'll look at the effects of electric and magnetic fields as they change with time. We'll see that these fields are so closely related to each other that they lead to a single electromagnetic field. **R-E**