Why Antennas Radiate

Antenna theory is a popular subject with hams. We love to read and talk about it. Now put on your thinking cap and fasten your safety belt for a review of the math and science behind it.

By Stuart G. Downs, WY6EE

Mastering an understanding of electric and magnetic fields is not easy. The electric field E, the magnetic field B and the vector magnetic potential **A**, are abstract mathematical concepts that make practical presentations difficult. Those fields, however, have everything to do with why antennas radiate. Explanations for the qualitative and quantitative relationships between electric, magnetic and potential fields are presented.

All electromagnetic field equations are interrelated. Each represents a different aspect of the same thing. Indeed, we may derive one from another! Here we shall examine the dc case, and then move on to RF. First, we owe it to ourselves to cover the fundamentals.

11581 Aspendell Dr
San Diego, CA 92131

We shall begin with E and B. Assume that both fields are constant in magnitude (uniform) and observable. In our first example, they are produced by a single charged particle moving at a constant velocity **v** in a vacuum. Next, changing magnetic and electric fields produced by time varying antenna RF currents are discussed. In all cases, fields are produced both by stationary charge and charge in motion. All charge we assume to be connected through fields. We observe that the fields change when charge is in motion relative to an observer.

It is very important to realize that both constant and time-varying fields cause action at a distance. That is to say, an electron's field affects other electrons some distance away.

What is a field? No one really knows. Field lines were visualized by Michael Faraday. The idea came from the orientation of iron filings that he observed on top a piece of paper with a magnet (lodestone) placed beneath it. According to the Richard Feynman,[1] a field is a mathematical function we use to avoid the idea of action at a distance. We can state that a field connecting charged particles causes them to interact because the field exerts a force on charged particles.

Does this mean that all matter in the universe is connected through fields? As one so elegantly put it:

All things by immortal power,
Near or far,
Hiddenly
To each other linked are,
That thou canst not stir a flower
Without troubling of a star...."

[1]Notes appear on page 42.

From "Mistress of Vision" by English poet Francis Thompson (1857-1907)

Charge in Motion Gives Rise to a Magnetic Field

All charged particles produce an electric field E. The E field can point inward or outward, depending on the sign of the charge, and is infinite in extent. The field causes action at a distance. See Fig 1. Because the field has both a magnitude and direction, we shall represent it as a vector. The same is true of a particle's velocity, so it is also a vector. Vectors are henceforth indicated by boldface letters.

The only magnetic field associated with a stationary charged particle is its spin magnetic dipole moment, but we shall ignore that for now. No other magnetic field is produced by stationary charged particles because the particle and its **E** field are not in motion relative to the observer.

When a particle with charge q moves with velocity v, its **E** field changes: It becomes dynamic. The dynamic **E** field gives rise to a **B** field as seen by "Joe Ham," a stationary observer, in Fig 2. Joe observes that constant-magnitude electric and magnetic fields are present simultaneously. The magnitude of **B** depends upon the velocity **v** of the particle. The magnetic field that Joe observes is:[2]

$$\mathbf{B} = \frac{1}{c^2}\mathbf{v} \times \mathbf{E} \qquad \text{(Eq 1)}$$

where c is the speed of light in m/s, **v** is charge velocity in m/s and **E** is the E field in V/m. Note that if **v** = 0 then **B** = 0 (no magnetic field). If the velocity of the E field were c, then we would have **E**=c**B**, which is what we have with a freely propagating electromagnetic wave. The cross product, designated by x, between the velocity vector **v** and the **E** vector indicates that particle velocity is perpendicular to both the E and B fields. The B field in Eq 1 is bound to the moving charged particle, as observed by Joe. The B field comes from a relativistic transformation of the E field involving the ratio ***v***/c. Einstein introduced the world to relativity 100 years ago, in 1905!

The **E** field multiplied by **v**/c² that transforms to the become the B field in Eq 1 is:[2]

$$\mathbf{E} = \frac{q}{4\pi\varepsilon_0 r^2}\frac{(1-\beta^2)}{(1-\beta^2 \sin\theta)^{3/2}}\hat{\mathbf{r}} \qquad \beta = \frac{v}{c} \qquad \text{(Eq 2)}$$

where ε_0 is the permitivity of a vacuum, r is the distance from the particle's line of travel, θ is the angle between the **E** field and the particle's direction of travel, and $\hat{\mathbf{r}}$ is a vector of unit length pointing in the direction from the particle to the place where **E** is evaluated. The magnitude of **B** (see Note 2) is found by substituting Eq 2 into Eq 1:

$$B = \frac{q}{4\pi\varepsilon_0 r^2}\frac{v}{c^2}\frac{(1-\beta^2)\sin\theta}{(1-\beta^2 \sin\theta)^{3/2}} \qquad \text{(Eq 3)}$$

Therefore, for a moving charged particle in free space with no external influences, a portion of its E field gets transformed, becomes dynamic, and appears as the B field multiplied by the coefficient v/c². This means that if particle velocity increases, the magnitude of the B field increases proportionally. An observer sees both the B field and the E field at the same time with charge in motion. Note that the **B** vector is perpendicular to the **E** vector, to the velocity vector v. Einstein recognized this very fact by thinking about one of Maxwell's equations and the result was relativity. The picture assumes that θ in Eq 3 is 90°.

To summarize, the magnitude of the resulting magnetic field depends upon the velocity of the charge and the amount of charge. This means that the B field really is the relativistically transformed E field! The two fields must always change together and they do. If the B field source is the E field, can there be such a thing as a B field by itself without an E field? The answer may surprise you.

Charge Moving at Constant Velocity Produces a Constant Magnetic Field

Charge flow in a wire is obviously very much different from that of an isolated charge moving in free space. However, exactly what is different and why?

Assume that there are no unmatched charges in our wire. That means the number of electrons (nega-

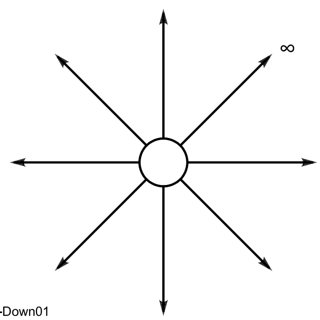

Fig 1—An isolated charged particle with its E-field lines. E fields cause action at a distance. The effect is very small for small charges a great distance apart.

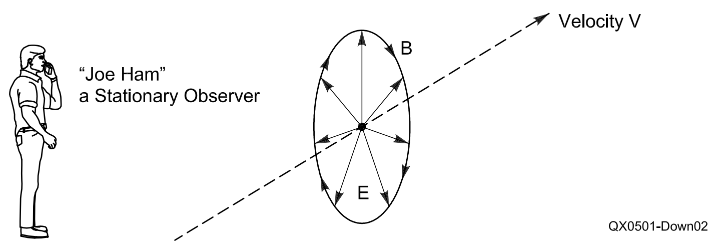

Fig 2—The B and E fields of an electron moving with a constant velocity *v*.

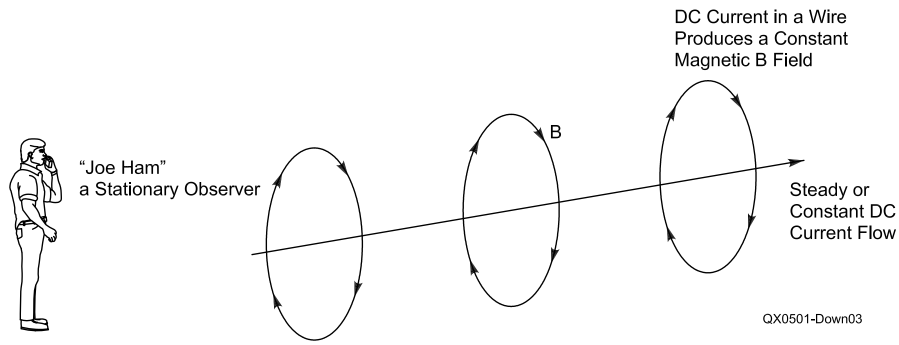

Fig 3—Magnetic field lines around a wire with dc current.

tive charges) equals the number of protons (positive charges) in the wire. Experiments have confirmed that current flow in such a wire produces only an observable magnetic B field at a distance *r* from the wire. The electrons' radial E fields are not observed at all. If Joe Ham were the observer, he would say there were no E fields present. The drift velocity of the moving charge is constant, so the B fields is just a function of the number of charges in motion and their velocity. See Fig 3.

Constant DC current in an infinitely long wire produces a uniform magnetic field along the wire's entire axial length. The magnitude of **B** everywhere along the axis at a distance *r* from the wire is the same. The magnetic field lines extend to great distances and they always close on themselves, unlike electric field lines, which always terminate at a charge.

Joe might ask the question, "If the B field came from the E field, then how is there now only a B field observable?" Perhaps we should state that there is no *net* E field present. Now I would like to propose another question: If there is no net E field present, is there an E-field energy density present?

The Wire's Unobserved Net E Field

Let's say that there is a moving radial E field that accompanies charge flow in a wire and it cannot be observed for some reason. We know that protons are fixed in the wire's metallic lattice and that the electrons are the charges that actually move. The proton static radial E field is uniform and produces no B field because there is no relative motion with respect to the observer. The electrostatic E field from all the wire's protons and electrons cancel and we are just left with the B field. The E fields of the protons and electrons, of the same magnitude but of opposite directions, must be zero or very close to it for Joe not to observe them.

Is there a basis for this in physics? The answer is yes. The big question is: How do we know this? The energy density of each field must be present. If this were not the case, then the law of superposition would be violated.

Superposition of Like Field Quantities

We know from the principle of field superposition that the net vector field produced by two separate vector fields of the same kind (either E or B), at the same time and place, is the vector sum of those individual fields. That may help additionally to explain why no net E field is observed when current flows in a wire. The E field from a fixed proton must exactly cancel the nonrelativistically transformed E field from the moving electrons.

The superposition principle also works for magnetic fields according to Feynman (See Note 2). That is to say, individual small amounts of magnetic field strength *d*B combine to produce the macroscopic strength that we observe. Is it possible to test for energy density where the net E field is zero and is not observable?

Gauss's Law

Let's use superposition and the concept of the closed surface to prove what we discussed earlier and what we saw in Fig 4. To do it, we extend the discussion to another term called flux. Flux is the electric field strength associated with each unit area through a surface. Feynman[3] relates flux and charge in the following way: "The E-field flux Φ through any closed surface is equal to the net charge inside that surface divided by the permittivity of free space." That is Gauss's law. The closed surface could be a sphere containing some charge, or any other shape so long as it fully encloses the charge. In mathematical terms:

$$\Phi_{total} = \frac{q_{total}}{\varepsilon_0} \qquad \text{(Eq 4)}$$

Thus the net E-field flux passing through the surface enclosing our wire, which has a net charge of zero, is zero.

It is interesting to note that if we used superposition to determine energy density (energy/volume) by summing the E-field energies of protons and electrons, the energy densities do not cancel, they add! The reason is that in the calculation of energy density, the E-field is squared. This yields two positive numbers so there can be no net density cancellation. The proof that field energy density is present is beyond the scope of this paper. I suspect that it has something to do with gravity. Didn't Einstein show us that matter and fields were both essentially forms of energy?

The Biot-Savart Law

It turns out that there is a law that may be derived from Eq 3 that supports the idea that all B fields arise from changing E fields. Everything about it is self-consistent. It is called the Biot-Savart law.

The Biot-Savart law allows determination of B a distance r away from a wire for a given dc current. To perform an actual calculation, one would break the wire into infinitesimally small segments *dl* and integrate along the entire length of the wire all the infinitesimally small *dB* contributions they produced at some distance *r* from the wire. The sum would be the magnetic field strength B at that distance.

To derive this law, we begin with Eq 3 and assume that the magnitude of B is changing with time. In the notation of calculus, the rate of change would be dB/dt, where *t* is time. Taking the time derivative of the magnetic field, we get:

$$\frac{dB}{dt} = \frac{d}{dt}\left[\frac{q}{4\pi\varepsilon_0 r^2}\frac{v}{c^2}\frac{(1-\beta^2)\sin\theta}{(1-\beta^2\sin\theta)^{3/2}}\right] \qquad \text{(Eq 5)}$$

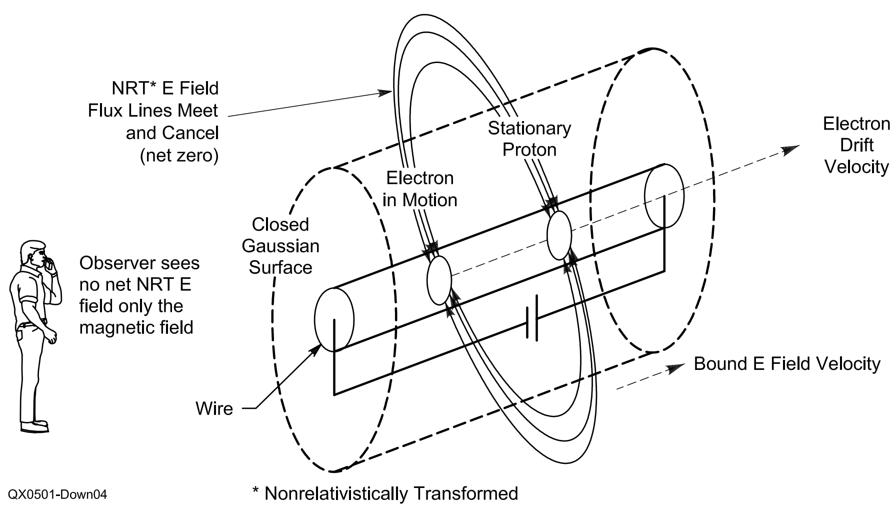

Fig 4—Electron motion in a wire and particle E fields.

Assuming $\theta = 90°$, it is easily shown that:

$$\frac{dB}{dt} = \frac{d}{dt}\left[\frac{q}{4\pi\varepsilon_0 r^2}\frac{v}{c^2}\frac{1}{(1-\beta^2)^{1/2}}\right] \quad \text{(Eq 6)}$$

Note that $\gamma = 1/(1-\beta^2)^{1/2}$ is a commonly used relativistic term. Using the binomial expansion theorem for the relativistic term and $\beta^2 = v^2/c^2$, we see that:

$$\left(1-\beta^2\right)^{1/2} \approx \frac{1}{\left(1-\frac{v^2}{c^2}\right)} \quad \text{(Eq 7)}$$

when v is small with repect to c. Substituting in the relativistic gamma term γ, we see:

$$\frac{dB}{dt} = \frac{d}{dt}\left(\frac{v}{c^2}\frac{\gamma q}{4\pi\varepsilon_0 r^2}\right) = \frac{d}{dt}\left(\frac{v}{c^2}\gamma E\right) \quad \text{(Eq 8)}$$

The gamma function γ stretches the electrostatic field $q/(4\pi\varepsilon r^2)$. Now substituting in Eq 7:

$$\frac{dB}{dt} = \frac{d}{dt}\left[\frac{v}{c^2}\frac{q}{4\pi\varepsilon_0 r^2}\left(1-\frac{v^2}{c^2}\right)\right] \quad \text{(Eq 9)}$$

Knowing that $v \ll c$, we use a simplification so that the inside term in parentheses of Eq 9, goes to zero. Additionally, we multiply both sides by *dt* and we end up with:

$$dB = \frac{d}{dt}\left(\frac{q}{4\pi r^2}\frac{vdt}{\varepsilon_0 c^2}\right) \quad \text{(Eq 10)}$$

Lastly, we know that $vdt = dl$ (velocity times time equals distance); current $I = dq/dt$ (charge per unit time) and $\mu_0 = 1/\varepsilon_0 c^2$ (the permeability of a vacuum), we arrive at the magnitude form of the Biot-Savart law with $\theta = 90°$:[4]

$$dB = \frac{\mu_0 I dl}{4\pi r^2} \quad \text{(Eq 11)}$$

We derived the Biot-Savart law from charge in motion in free space. Biot and Savart experimentally deduced[4] this relationship that links a current segment in a wire to an infinitesimally small magnetic field *dB* a distance *r* away from the wire. It is taught in physics textbooks today[4, 5, 6] and forms the basis for magnetic induction. For example, Ampere's and Faraday's law can be derived from the Biot-Savart law. So magnetic fields are related directly to currents. See Fig 5.

Electromagnetic Radiation

Up to now, we have been looking at fields bound to or coupled to charges moving at constant velocities. We have also seen that a stationary observer sees only the B field for charge moving at constant velocity in a wire, and that velocity has been slow. Now what would happen if the charge velocity rapidly changed, as in the case of an RF signal? What if charge velocity were changing at a high rate?

When its velocity is changing, we say a particle is accelerating. Deceleration is just negative acceleration—it's just a matter of signs. Acceleration of charges is what launches radio signals. However, the E field here is one of a different color. We have seen that as field velocity goes to c, both E and B propagate together at the speed of light so that E=cB. However, how exactly does an antenna launch such an electromagnetic wave?

What really happens is that electrons in an antenna accelerate and decelerate because of the application of some time-varying electromotive force (EMF or voltage) to the antenna. A time-varying EMF implies the presence of a time-varying electric field E. Each electronic charge q in the antenna experiences force $F = qE$ and therefore accelerates and decelerates according to $F = mA$, where m is the mass of the electron and A is acceleration.

Thus, we have an alternating current in the antenna. The simplest alternating current is sinusoidal since it consists of a single frequency. We have shown that a changing E field gives rise to a B field. The fields propagate with velocity c, but the drift velocity of the electrons in the antenna travel at some much lower velocity v. The magnitude of the fields must vary in a sinusoidal fashion from point to point along the wire's length. It is interesting to note that the electron speed and the wave speed appear to detach from each other.

The radiated B field is perpendicular to the antenna while the E field is

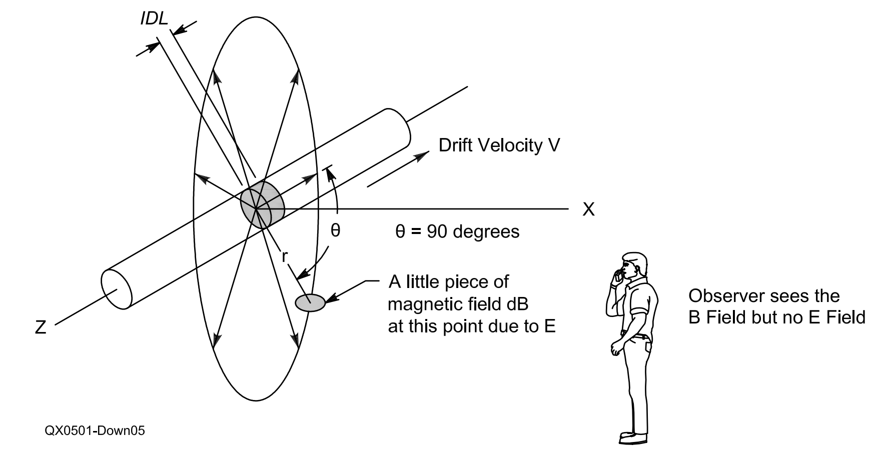

Fig 5—A pictorial representation of the Biot-Savart law.

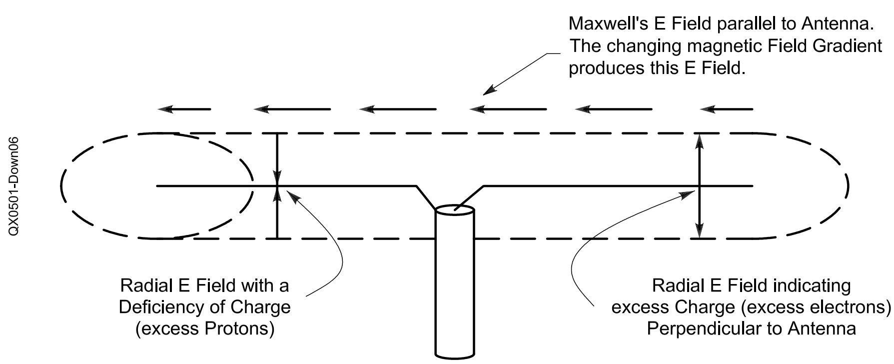

Fig 6—A dipole antenna and surrounding fields.

parallel to it. Both move outward from the antenna with speed c with respect to the antenna. See Fig 6. Nevertheless, how do we get a parallel E field when we know that the E-fields of the electrons themselves are perpendicular?

We saw above that in a wire having zero net charge, the E field caused by the electrons is not measurable macroscopically. Recall that is because the E field from the protons cancel it. All that is measurable outside the wire is the B field. But it has been shown that just as changing E fields produce B fields, the reverse is also true: Changing B fields produce E fields. The orientation of whichever field is so produced actually *counteracts* the change in the field producing it.

It might seem at first that this reciprocity would prevent anything from happening, since one effect tends to cancel the other, and that tendency is true. It's called Lenz's law and it is an essential physical fact, not just some arbitrary convention about signs or directions. It is an interesting manifestation of physical systems' resistance to change, akin to Newton's first law of motion. It implies that some energy must be added to the system to build fields. It is convenient to think of that energy as being stored in the field.

In our antenna, we start with only a B field outside the wire, but it is changing and propagating away from the wire at speed c. When the current in the wire alternates rapidly enough, the changing B field propagates away before the Lenz effect can cancel it. Since the B field is also alternating, it is accompanied by an alternating E field whose peak magnitude grows to its final value as the wave is launched. That first stage of wave formation takes place in what we call the near field.

The near field is generally considered within about 10 wavelengths. Outside the near field is the far field. In the far field, where the electromagnetic wave freely propagates outward, E=cB at all times everywhere. Thus the near field is chiefly magnetic and E<B.

For more detailed discussion and mathematics surrounding the above topics, navigate to **www.arrl.org/qex/.** Look for 0501Downs.zip.

Summary and Conclusion

The radial electrostatic field in motion around a current carrying wire is not observed. This is so because there is no net charge in the wire, and the net E field from all of the wire's protons and electrons cancel. To observe any net E field would violate Gauss's law.

In an electromagnetic wave, E drops off as $1/r$ in the far field whereas an electrostatic field drops off as $1/r^2$. Therefore, E in the travelling wave cannot be an electrostatic field. It is the changing field generated by the changing B field around the antenna. Likewise, E in the travelling wave is generated by the changing B field. Mutual recreation occurs perpetually and the wave travels at velocity c. James Maxwell proved it.

E and B fields are mathematical constructs we use to describe action at a distance. They are really manifestations of the same thing. There are particle theories of electromagnetic radiation, too, but we chose not to discuss them here.

Acknowledgments

I am grateful to Glenn Borland, KF6ZVK, for his friendship and insight, which has helped with this paper's formulation. Indeed, over the years we have had many great conversations. A hardy thanks also to Julius Tolan and Santos Cota for their suggestions.

Notes

[1]Feynman, Leighton, Sands, *The Feynman Lectures on Physics*, Volume II, Addison-Wesley, p 15-7.

[2]R.D. Field, University of Florida. Physics 2061 class notes, **www.phys.ufl.edu/~rfield/classes/fall02/images/relativity_21.pdf**

[3]Feynman, Leighton, Sands, *The Feynman Lectures on Physics*, Volume II, Addison-Wesley, 13-12.

[4]Halliday, Resnick, and Walker, *Fundamentals of Physics* Sixth Edition, Extended Version, Wiley 2001, pp 687-688.

[5]Fishbane, Gasiorowicz and Thornton, *Physics for Scientists and Engineers*, Prentice-Hall 1996, pp 819-896

[6]Hayt, *Engineering Electromagnetics*, Fifth Edition, McGraw Hill 1989, pp 216-224.

[7]Halliday, Resnick, and Walker, *Fundamentals of Physics* Sixth Edition, Extended Version, Wiley 2001, p 690.

[8]Feynman, Leighton, Sands, *The Feynman Lectures on Physics*, Volume II, Addison-Wesley, p 13-2.

[9]Hayt, *Engineering Electromagnetics*, Fifth Edition, McGraw Hill, 1989, p 251.