THE QUEEN'S AWARDS FOR ENTERPRISE 2001

BBC

R&D White Paper

WHP 103

December 2004

The theory of electrically-short antennas

R.H.M. Poole

Research & Development
BRITISH BROADCASTING CORPORATION

BBC Research and Development
White Paper WHP 103

The theory of electrically-short antennas

R H M Poole

Abstract

Electromagnetic theory is difficult: it involves an off-putting combination of calculus and vector algebra. Even people with an engineering background often feel intimidated by the first few pages of a textbook on the subject. An unfortunate consequence is that it is possible to offer 'magic' antennas for sale to a gullible world.

This report attempts to explain how short antennas work in straightforward terms, and concludes with a discussion of the practical problems associated with using such devices. The hope is that the reader will be able to judge the claims made for 'magic' antennas more intelligently.

White Papers are distributed freely on request.

Authorisation of the Chief Scientist is required for publication.

1. Introduction

Electromagnetic theory has a reputation for being difficult: it involves an off-putting combination of calculus and vector algebra. The first few pages of a textbook on the subject are sufficient to intimidate anyone without a mathematical background. An unfortunate consequence is that 'magic' antennas get offered for sale to a gullible world. The claim usually made is that the antenna in question offers a desirable combination of highly directional radiation pattern, high efficiency and small physical size.

This report attempts to explain how short antennas work in straightforward terms. It is hoped that the reader will gain enough theoretical knowledge to discuss the merits of such devices intelligently. The report does not enter into the controversy associated with any particular 'magic' antenna. However, it does show that claims made for 'magic' antennas in general should be regarded with scepticism.

2. Radiation from a Doublet

The key principle when dealing with problems of electromagnetism is 'superposition'. In essence, superposition means that the whole is the sum of the parts. For instance, when calculating the performance of an antenna, it is generally convenient to divide the antenna up into a large number of short lengths, or elements. If it is possible to calculate the field generated by each of the elements, the job is done: the overall field is simply the vector sum of the individual fields. Incidentally, the principle of superposition has been in existence for hundreds of years, and has been applied successfully to wave disturbances in any linear medium.

An appropriate antenna element is the doublet, which is simply a short length of wire carrying a current. Suppose, for the moment, that the doublet includes a battery, so that a direct current flows if the two ends of the doublet are placed in a resistive medium. The current may be thought of as being forced into the surrounding medium at one end and drawn out at the other end:

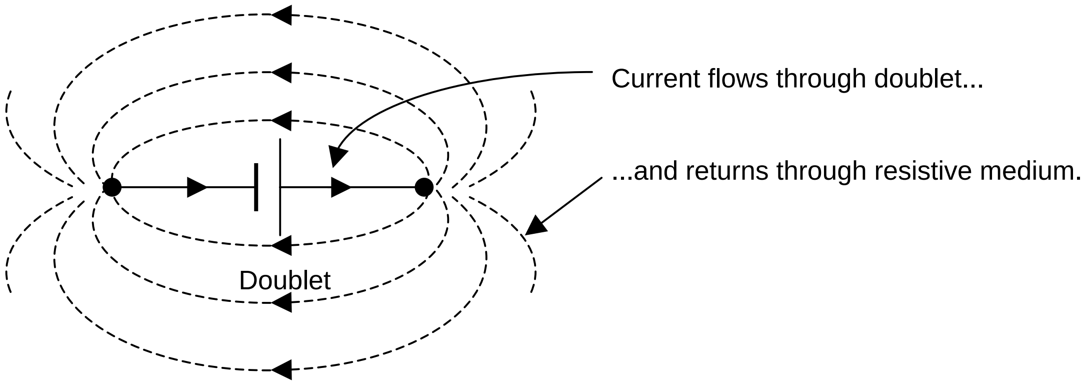


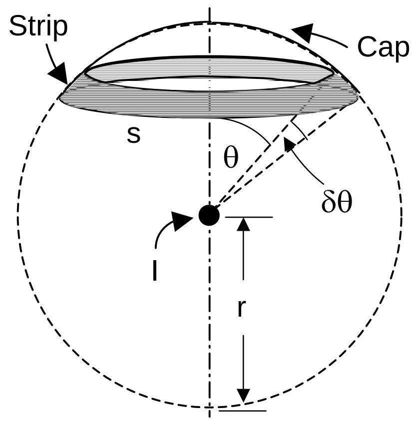
Figure 2.1: Current Flow through Doublet and Surrounding Resistive Medium

A consequence of the current flow is the generation of a magnetic field. It is this field that is of interest when calculating radiation.[1] Because a battery is used to provide the current, an electric field must also be present in the surrounding medium. It would be just as valid to calculate the radiation in terms of voltages and electric fields, but the usual convention is to think in terms of currents and magnetic fields.

[1] To be strictly correct, we should say that only *alternating* currents are capable of generating radiation.

The first step in calculating radiation from the doublet is to work out the current distribution in the surrounding medium. This appears difficult to do until one realises that the calculations for the individual poles are straightforward. Once obtained, the two resulting contributions can be summed to give the overall current distribution — an instance of superposition.

Consider the point source of current, or monopole, shown below. The current I spreads evenly in all directions through the surrounding resistive medium.



The current flowing through the cap bounded by circle of circumference s is calculated by simple integration. A strip of length s and width $r\delta\theta$ possesses area $s\, r\delta\theta$. The current density at distance r equals $I/4\pi r^2$, and so the current through the strip is

$$\delta i_{cap} = I\, s\, r\delta\theta / 4\pi r^2.$$

Substituting $2\pi r \sin\theta$ for s gives

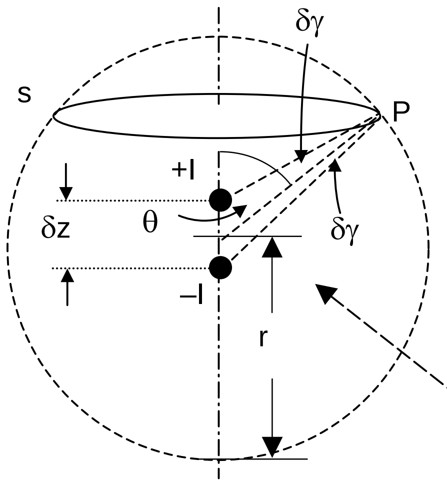
$$\delta i_{cap} = (I/2) \sin\theta\, \delta\theta\ .$$

Integrating between the limits 0 and θ gives

$$i_{cap} = (I/2)\ \{\ 1 - \cos\theta\ \}\ .$$

Figure 2.2: Current Spreading out from a Point Source

Now consider the doublet, which is simply two such point sources of current, +I and –I, separated by distance δz. The current through the cap is now the resultant of the two contributions:



$$i_{cap} = \underbrace{(I/2)\{1 - \cos(\theta+\delta\gamma)\}}_{\textit{positive pole}} - \underbrace{(I/2)\{1 - \cos(\theta-\delta\gamma)\}}_{\textit{negative pole}},$$

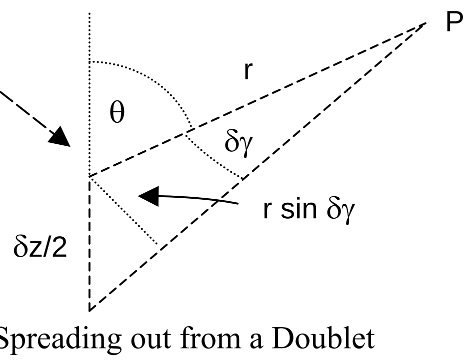
where 2δγ is the angle subtended by the doublet at P.

The two I/2 terms cancel, so

$$i_{cap} = (I/2)\ \{\cos(\theta-\delta\gamma) - \cos(\theta+\delta\gamma)\}.$$

Since $\cos(A \pm B) = \cos A \cos B \mp \sin A \sin B$,

$$i_{cap} = I \sin\theta \sin\delta\gamma.$$



Because $r \sin\delta\gamma = (\delta z/2) \sin\theta$,

$$i_{cap} = (I/2r)\ (\sin^2\theta)\ \delta z.$$

Figure 2.3: Current Spreading out from a Doublet

A magnetic field H results from the current i_{cap}. The relationship between current and magnetic field is given by Ampère’s Law, which states that the current through the cap equals the line integral of H around the periphery of the cap; in vector notation, $i_{cap} = \int \mathbf{H.ds}$. The scalar product **H.ds** indicates that only the component of **H** in the direction of **ds** is relevant. It is easier to think in terms of the individual polar coordinates:

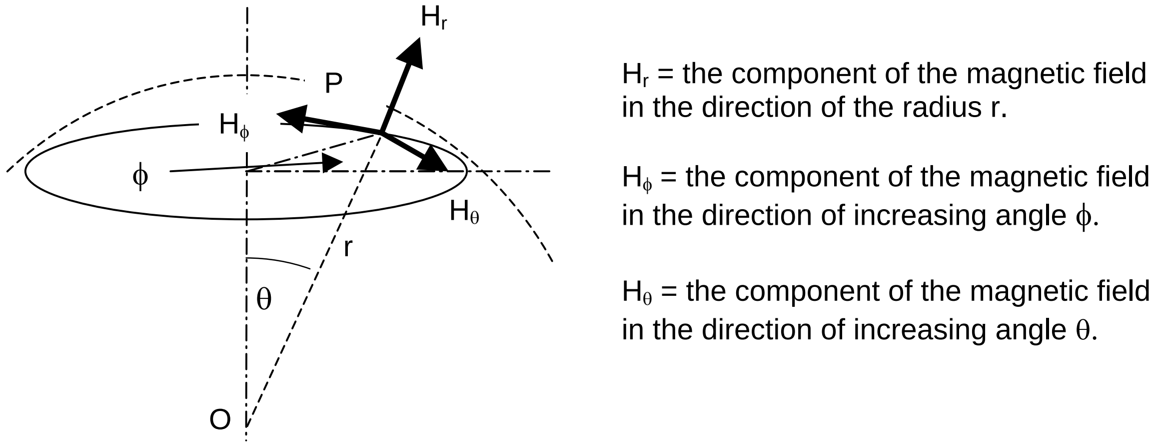


Figure 2.4: Polar Coordinates and Associated Magnetic Field Components

In differential form, Ampère's Law becomes

$$H_\phi = di_{cap}/ds = (d/ds)\{(I/2r)(\sin^2\theta)\,\delta z\}\ .$$

Since $s = 2\pi r \sin\theta$, $ds = (2\pi\sin\theta)\,dr$. Substituting for ds in the equation for H_ϕ gives

$$H_\phi = \{(I\sin\theta)/2\pi\}\ \delta z\,(d/dr)(1/2r) = (I\sin\theta)\,\delta z/4\pi r^2\ .$$

Thus the magnetic field is proportional to the current and the length of the doublet, and falls away as the inverse square of the distance. The field is zero along the axis of the doublet.

Now suppose that the doublet is fed with an alternating current $I_0\cos\omega t$. H_ϕ is calculated in the same way as for the static case, except that I must be replaced with $I_0\cos(\omega t - 2\pi r/\lambda)$: the phase lag $2\pi r/\lambda$ results from the finite speed at which the current propagates. Hence the expression for H_ϕ becomes

$$H_\phi = \{\delta z\sin\theta)/4\pi\}\ (d/dr)\{(1/r)(I_0\cos(\omega t - 2\pi r/\lambda))\}\ .$$

The quantity $2\pi/\lambda$ is conventionally referred to as β. The differentiation with respect to r is carried out by treating (1/r) and $(I_0\cos(\omega t - \beta r))$ as the two components of a product. Then

$$H_\phi = \{I_0\,\delta z\sin\theta)/4\pi\}\ \{((-1/r^2)\cos(\omega t + \beta r)) + (1/r)(+\beta\sin(\omega t - \beta r))\}\ .$$

This expression is often written, more conveniently, in complex notation:

$$H_\phi = \{I_0\,\delta z\sin\theta)/4\pi\}\ \{(j\beta/r) + (1/r^2)\}\ \exp j(\omega t - \beta r)\ .$$

The expression for H_ϕ is the same as for the static case, except for the new term $(j\beta/r)$.

It is possible to dispense with the conducting medium by imagining the doublet to comprise two halves of a dumbbell. Because the ends of the dumbbell possess capacitance, a current may be sustained by applying an alternating voltage between them. The $(j\beta/r)$ term now corresponds to radiation from the doublet.[2] There is a 90° phase difference between the radiation field and the $1/r^2$ 'induction' field, and the induction field is said to be reactive.

A similar analysis of the associated electric fields yields the results **[2]**

$$E_r = \{60\,I_0\,\delta z\cos\theta\}\ \{(1/r^2) - (j/\beta r^3)\}\ \exp j(\omega t - \beta r) \qquad \text{and}$$

$$E_\theta = \{30\,I_0\,\delta z\sin\theta\}\ \{(j\beta/r) + (1/r^2) - (j/\beta r^3)\}\ \exp j(\omega t - \beta r)\ .$$

The radial component of the field, E_r , does not contain a term in 1/r.

2 No radiation is associated with the term containing $1/r^2$. It is obvious that at a great distance from the doublet, only the 1/r term is significant. The total power radiated must then be associated with this term.

Apart from the induction fields, there are 'electrostatic' fields that fall away according to a $1/r^3$ law. Provided that r exceeds the wavelength of the radiation, λ, the induction and electrostatic fields are normally negligible, and the field strength magnitudes can be written as

$$E_\theta = \{30\ I_0\ \delta z \sin\theta\ \}\ (\beta\ /r) \text{ and } H_\phi = \{I_0\ \delta z \sin\theta\ )\ /4\pi\}\ (\beta\ /r).$$

The θ and ϕ subscripts indicate that the E and H fields are orthogonal. There is a constant ratio between the two fields of 120π ohms. This quantity is known as the impedance of free space, Z_0, and is analogous to the resistance of a 'real' conductor.

The results of the above analysis are well known, and have been used extensively to calculate the performance of antenna systems. There is no reason to doubt that they are correct. Note that the induction and electrostatic terms are essential companions to the radiation process: they are not simply unwanted artefacts that can be eliminated by 'good' antenna design.[3]

3. The Concept of Radiation Resistance

The previous section has shown how to calculate the magnetic field strength associated with a doublet radiator. It has also quoted the electric field strength. Another important parameter is the radiation resistance, since this determines the ease with which the doublet may be matched to a source of power. It also has an important influence on the amount of power wasted in the associated system, particularly the ground. The following calculation of the radiation resistance is little more than a sketch; there are many textbooks providing a thorough treatment. **[1]**

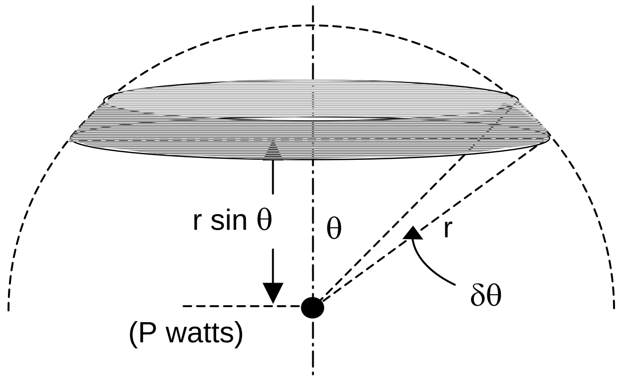


Figure 3.1 shows the doublet radiating a power P at the centre of a sphere of radius r. Assuming r is large, the magnitude of the magnetic field is given by

$$H_{\phi\ rms} = \{I_{rms}\ \delta z \sin\theta\ )\ /4\pi\}\ (\beta\ /r).$$

(It is convenient here to consider RMS values of I and H_ϕ .)

Figure 3.1: Power Flow from the Doublet Radiator

The energy associated with $H_{\phi\ rms}$ is (1/2) $\mu_0\ H_{\phi\ rms}^2$ joules per cubic metre, where μ_0 is the permeability of free space. There is also an exactly equal amount of energy (1/2) $\varepsilon_0\ E_{\theta\ rms}^2$ associated with the electric field E_θ,[4] giving a total energy of (1/2) $\mu_0\ H_{\phi\ rms}^2$ + (1/2) $\varepsilon_0\ E_{\theta\ rms}^2$. However, because it is simpler to think in terms of one of the fields only, the total energy will be taken as $\mu_0\ H_{\phi\ rms}^2$. If the energy is radiated with velocity c, the power flow will be c $\mu_0\ H_{\phi\ rms}^2$ watts per square metre. Since c equals 3×10^8 (very closely) and μ_0 equals $4\pi \times 10^{-7}$ (by definition), the product c μ_0 can be written as 120π . This quantity is simply the impedance of free space, Z_0, the concept of which was introduced in the previous section.

---

[3] This point is made because the proponents of some 'magic' antennas claim that, by generating E and H fields in the correct proportions, the near fields can be greatly reduced.

[4] The remarkable symmetry that exists between electric and magnetic fields is not surprising to believers in the Theory of Relativity. Einstein claimed that his main motivation for developing the theory was to show that electric and magnetic fields are really different manifestations of the same thing. **[4]** The appendix in this document provides an illustration of the equivalence of electricity and magnetism.

The length of the circular strip shown in Figure 3.1 is $2\pi r \sin\theta$, and its width $r\,\delta\theta$; hence its surface area is $2\pi r^2 \sin\theta\,\delta\theta$. Since the power flow per unit area is $Z_0 H_{\phi\,rms}^2$, the power δW flowing through the strip is given by[5]

$$\delta W = (2\pi r^2 \sin\theta\,\delta\theta) \times Z_0 H_{\phi\,rms}^2 \text{, where } H_{\phi\,rms} = \{ I_{rms}\,\delta z \sin\theta\,)/4\pi\}\,(\beta/r).$$

Thus $\delta W = Z_0 (I_{rms}\,\beta\,\delta z)^2 (1/8\pi) \sin^3\theta\,\delta\theta$,

and $W = Z_0 (I_{rms}\,\beta\,\delta z)^2 (1/8\pi) \int \sin^3\theta\,\delta\theta$

The integration needs to be carried out over the range $0 \le \theta \le \pi$.

$\int \sin^3\theta = -(3/4)\cos\theta + (1/12)\cos 3\theta = +(4/3)$, and so the expression for W simplifies to

$$W = 20\,(I_{rms}\,\beta\,\delta z)^2 .$$

(120π has been substituted for Z_0.)

The radiation resistance R_r is defined such that the power supplied to the doublet equals $I_{rms}^2 R_r$, by analogy with a 'real' resistor. This power must equal that calculated above:

$$W = 20\,(I_{rms}\,\beta\,\delta z)^2 = I_{rms}^2 R_r .$$

Hence $R_r = 20\,(\beta\,\delta z)^2$.

If β is written as $2\pi/\lambda$, the radiation resistance is given by

$$R_r = 790\,(\delta z/\lambda)^2 .$$

The radiation resistance is very small for electrically short antennas. If $\delta z/\lambda$ is taken as 1/20 — which still corresponds to a length of 15 m at 1 MHz — R_r is only about 2 Ω.

4. Virtual Poles and Wire Antennas

At MF, the ground is more-or-less conductive. This means that the transmitted signal must be vertically polarised, since any electric field component along the ground would be heavily attenuated. Hence the antenna must also be vertical.

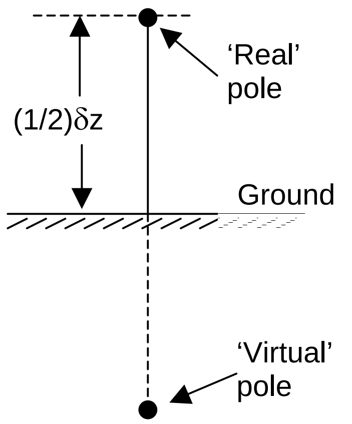


Figure 4.1: A 'Virtual' Doublet

Figure 4.1 illustrates the effect of the ground. At actual transmitting stations, the ground conductivity is enhanced in the neighbourhood of the antenna, usually by burying radial copper strips to form an 'earth-mat'. The transmitter power is then applied between the earth-mat and the antenna.

For the doublet radiator, the equivalent situation is a single 'real' pole $\delta z/2$ above the ground and a 'virtual' pole an equal distance beneath: the ground behaves as a mirror. Note that the distance δz now refers to the distance between the real and virtual poles.

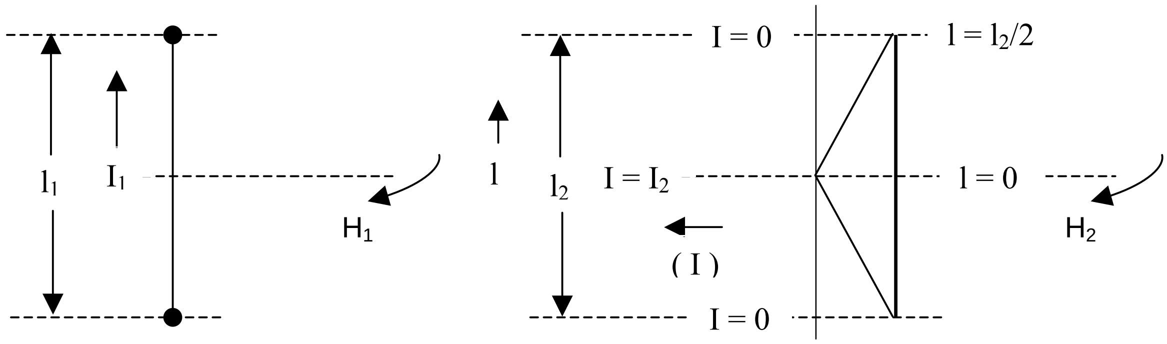
---

5 Conventional treatments of electromagnetism generally introduce the Poynting vector concept. The close relationship between the electric and magnetic fields allows the expression for power flow to be written as $E_\theta \times H_\phi$, with the direction of power flow being mutually perpendicular to the two fields. In vector terms, $\mathbf{S} = \mathbf{E} \times \mathbf{H}$, with **S** being the Poynting vector.

Note that $\mathbf{E} \times \mathbf{H}$ represents the power flow associated with an established wave. It should not be assumed that, if **E** and **H** can be made to exist independently at a point in space, a power flow will result.

The conductivity of the ground is useful in doubling the effective length of the antenna, hence increasing the radiation resistance and easing the match to the transmitter. Also, the effective power output of the transmitter is doubled, as power is not radiated into the ground.

However, even a doublet with a virtual pole is not representative of a real antenna, because in practice the current distribution is unlikely to be constant. If the antenna takes the form of a length of wire, the current must be a maximum at its feed-point (close to the ground) and zero at the end. The difference this makes is illustrated below:



Doublet *Wire Antenna*

Figure 4.2: Comparison of Doublet and Wire Antennas

The calculations for the true doublet are straightforward. The field δH_1 resulting from current I_1 flowing through an infinitesimal doublet of length δl is $k\ I_1\ \delta l$, where k is a constant. Integrating this shows that H_1 equals $k\ I_1\ l_1$. The calculations for the wire antenna are just the same, except that, if the current is taken to vary linearly between its ends,[6]

$$\delta H_2 = k\, I\, \delta l \text{ , where } I = I_2\,(l_2 - 2l)/\,l_2 \;\}\text{ between } l = \pm\,(1/2)\,l_2 \,.$$

Integrating gives $H_2 = k\, I_2\, l_2\, (1/2)$.

If the feed-point currents and resultant fields are taken to be the same in the two cases, the calculations show that $l_2 = 2\ l_1$ — the wire antenna must be twice as long as the doublet. The previously derived expression for radiation resistance becomes

$$R_r = 790\,(l_2 /2\,\lambda\,)^2 \quad = 197.5\,(l_2 /\lambda\,)^2 \,.$$

Note that, as already explained, the use of an earth-mat doubles the effective length of the antenna: l_2 refers to the total length of the antenna and its image.

A common technique is to add a 'capacity hat' to the top of the antenna. The capacitance of this has the effect of making the current distribution more nearly uniform. In the limit, when the current is constant along the wire, the radiation resistance equals that of a doublet of the same length. The extra capacitance has the added advantage of easing the match to the transmitter.

[6] The current distribution would be perfectly sinusoidal in a lossless resonator of length $\lambda/2$, as illustrated below. An antenna may be thought of as a resonator from which a small proportion of the stored energy is leaked each cycle . Where the antenna is electrically short, only the tips of the sinusoidal distribution are active; the remainder is replaced, in effect, by the impedance matching circuitry.

In the diagram, the missing part of the distribution is shown dashed. The tips of the sinusoid can be approximated by straight lines, since $\sin x \approx x$.

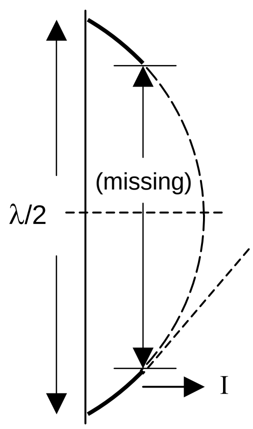


Figure 4.3: Current Distribution in an Antenna

5. Some Practical Considerations

Now that the basic theory is understood, there are three practical problems to consider: the vertical radiation pattern, efficiency, and the match of the antenna to the transmitter.

The vertical radiation pattern (VRP) is important because any power radiated skywards may be reflected by the ionosphere. Beyond a critical distance from the transmitter, the reflected sky wave and the direct ground wave will be of comparable strengths, and the resultant signal will be subject to deep fades. Traditional mast radiators are often deliberately designed to reduce the sky wave as far as possible. The technique is to make the antenna physically large — usually between 0.5λ and 0.6λ tall. Calculation then shows that the field contributions from the infintesimal elements of the antenna tend to cancel, except at low angles.**[2]** Unfortunately, as shown earlier, the VRP of an electrically short antenna always possesses a sin θ characteristic.[7]

A question that might be asked is whether a large antenna could be simulated by a combination of two short antennas and an electrical delay. In the example below, the delay is equivalent to λ/2, and the current has the same magnitude in each element:

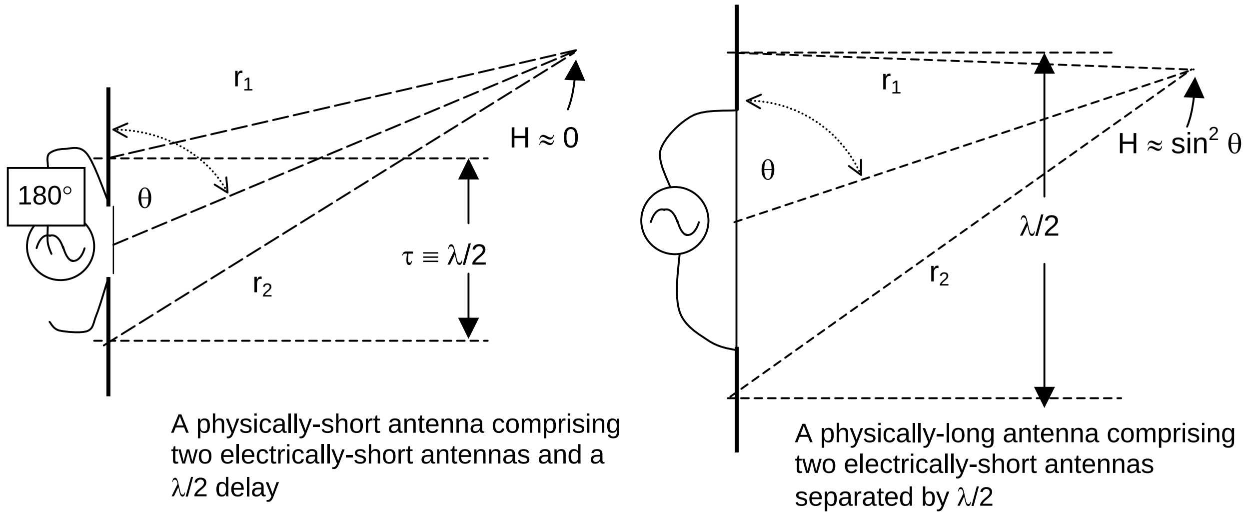


Figure 5.1: Comparison of Antennas with Lumped and Distributed Delays

The field associated with each element of the physically-short antenna possesses a sin θ dependence. However, because the two path-lengths r_1 and r_2 are almost equal, the phase difference between the contributions is always close to 180° and the contributions cancel. Clearly, this antenna configuration would not be very effective.

On the other hand, where the two elements of the antenna are separated by λ/2, the two contributions are in phase when θ is close to 90°, and consequently they add. As θ decreases, the phase difference increases with path difference $|r_1 - r_2|$. When θ = 90°, the phase difference is 180° and the contributions cancel completely. For small values of (90 – θ)°, the overall VRP approximates to $\sin^2\theta$, since the VRP of an individual element and the phase difference between the elements both possess sin θ dependences.

[7] This suggests a good test for 'magic' antennas. Some of these are claimed to possess much more directivity than this, despite their small size. If this is true for a particular antenna, magic really is involved. On the other hand, if the radiation pattern shows the expected sin θ dependence, the antenna is not magic but electrically short.

Power loss can occur both in the ground and in the network matching the antenna to the transmitter. These two components can be considered together.

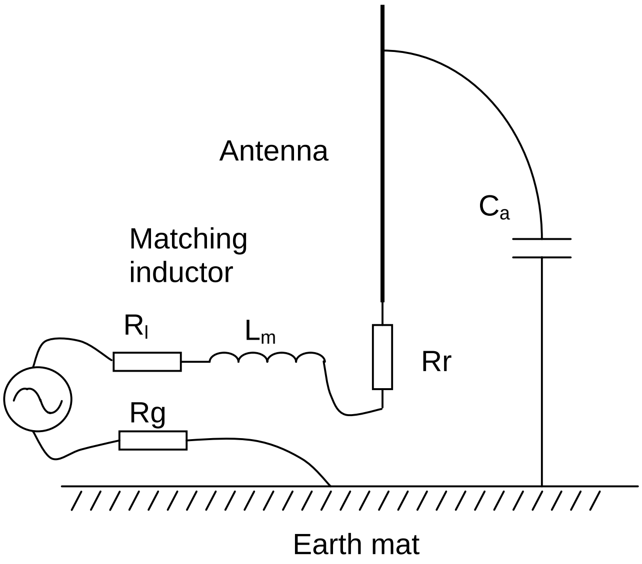


Figure 5.2 shows the antenna, which possesses a radiation resistance of R_r ohms and capacitance C_a. The effective ground resistance is R_g. Since the transmitter will probably wish to 'see' a resistive load, C_a is shown neutralised by inductor L_m. The equivalent series resistance of the inductor is R_l.

The signal must be applied with respect to ground. 'Ground' usually takes the form of a 'earth mat' of radial copper tapes extending as far as possible from the base of the antenna.

Figure 5.2: An Electrically-short Antenna with Matching Components

It is helpful to consider some representative values for an antenna designed to work at the top end of the MF band. Suppose the angular frequency ω_0 is 10^7, and $\delta z/\lambda$ is 1/40, so that the antenna is about 4.5 m tall. A good earth-mat might possess an R_g of 5 Ω. Since the R_r corresponding to $\delta z/\lambda = 1/40$ is about 0.5 Ω, the antenna must be less than 10% efficient: for every unit of power that is radiated, 10 units will be lost in the ground resistance.

The matching inductor will lose further power. L_m is simply given by the formula

$$L_m = 1/(\omega_0^2 C_a)$$ where C_a is the capacitance of the antenna.

If C_a is 100 pF, L_m should be 100 μH. Such an inductor can be made with a Q of about 500. Since R_l equals $\omega_0 L_m/Q$, R_l is then 2 Ω. This value is small compared with the ground resistance, but is still four times the radiation resistance. Note that there is a double advantage in adding a capacity hat: not only is the radiation resistance greater, but also L_m and its associated losses are smaller.

Another important issue is the bandwidth of the matching network. Assuming that the matching network is driven from a voltage source, the effective Q of the network is given by $\omega_0 L_m/R$, where R is the total series resistance. Taking $\omega_0 L_m$ as 1000 Ω and R as 5 Ω gives a Q of 200. This corresponds to a bandwidth of about 8 kHz, which is barely acceptable for AM broadcasting. Any attempt to improve the efficiency by reducing losses would serve to make the problem worse.[8]

Since the total of R_r, R_l and R_g is likely to be low, a practical installation would include a transformer to provide a match to (say) 50 Ω. Further losses could occur in the transformer.

---

[8] An idea worth exploring is to make the transmitter appear as a current source. The Q of the matching network would not then reduce the bandwidth of the transmission. Of course, the *voltage* appearing at the output of the transmitter would still be frequency dependent.

6. Field Strength at the Receiver

The efficiency of an antenna is conventionally defined as follows:

$$\eta = \frac{\text{effective monopole radiated power (EMRP)}}{\text{antenna input power}} \times 100\%$$

In practice, the EMRP is deduced from a measurement of the field strength at a distant point. The mathematics relating the theoretical EMRP to the field strength are straightforward.

Suppose for the moment that the transmitter power P is radiated from a point source or isotropic radiator. Energy will then pass uniformly through any sphere centred on the point source. At a distance r, the power flow S_i per unit area is given by

$$S_i = P/4\pi r^2 .$$

As discussed in Section 3, S is related to the electric field strength E_i as follows:

$$S_i = E_i^2/120\pi \text{, where } 120\pi \text{ is the impedance of free space.}$$

Combining these two expressions for S gives

$$E_i = \sqrt{(30\,P)}/r \text{, or } P = (1/30)\,E_i^2\,r^2 .$$

However, the maximum field strength generated by a doublet (or electrically-short antenna) must be greater than this, because the doublet does not radiate energy along its axis and the total energy must be the same in the two cases.

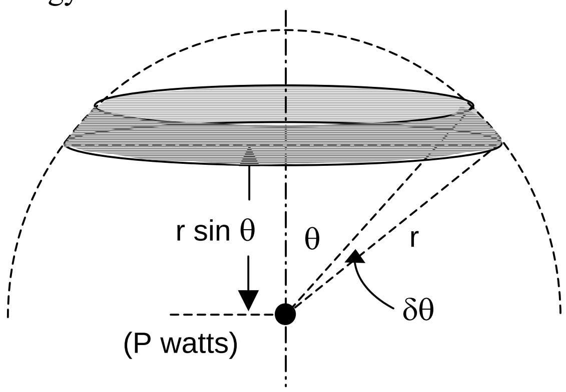


Figure 6.1: Calculation of Doublet Effective Gain

Suppose the electric field generated by the doublet is E_d . Since this possesses a sin θ dependence, $E_d = E_0 \sin\theta$, and the power density S_d is given by

$$S_d = (E_0 \sin\theta)^2/120\pi .$$

The increment of area shown in Figure 6.1 equals $(2\pi r \sin\theta)(r\,\delta\theta)$, or $2\pi r^2 \sin\theta\,\delta\theta$. Hence the power flowing through the increment of area is given by

$$\delta P = ((E_0 \sin\theta)^2/120\pi)\,2\pi r^2 \sin\theta\,\delta\theta .$$

Integrating the expression for δP between the limits θ = 0 and θ = π must give the total power P:

$$P = (2\pi r^2 E_0^2/120\pi) \int \sin^3\theta\,d\theta .$$

As quoted in Section 3, $\int \sin^3\theta = 4/3$, and so the expression for P becomes

$$P = (2\pi r^2 E_0^2/120\pi)(4/3) .$$

This expression for P must equal that previously derived for the isotropic case; therefore,

$$E_i^2 r^2/30 = (2\pi r^2 E_0^2/120\pi)(4/3) \text{, or}$$

$$E_0^2 = 1.5\,E_i^2 .$$

Hence the field E_0 generated by the doublet is √1.5 that of the isotropic radiator for θ equals 90°.

A monopole generates an even greater field strength, since all the power is radiated into the upper hemisphere alone. Consequently, the field strength should be multiplied by √2. Thus the final expression for field strength becomes

$$E_0 = \sqrt{(90\,P)}/r .$$

7. Conclusion

This document has attempted to provide some theoretical understanding of the properties of electrically-short antennas, and hence a means of interpreting future practical measurements. The conclusions are as follows:

- Electrically-short antennas all possess a sin θ vertical radiation pattern, where θ is the angle from the vertical. The large proportion of power heading skywards means that such antennas are not suitable for anti-fading applications.
- Although electrically-short antennas are not inefficient in themselves, the small value of the radiation resistance implies that much of the transmitter power will be lost in the ground resistance. Some power will also be lost in the impedance matching network. If the height of the antenna is λ/40, the overall efficiency is unlikely to exceed 10%.
- The small size of the antenna implies large near fields and consequent dielectric losses in the ground. A grounded conducting sheet underneath the antenna can be helpful.
- Unfortunately, minimising system losses is likely to result in an excessive Q of the matching network, and consequent narrow bandwidth. Not only does a poor bandwidth cause loss of modulation sidebands, it also results in 'touchy' transmitter matching.
- Although the antenna itself might be electrically small, the need for an efficient grounding system implies that the complete installation must still be large.
- Where a 'magic' antenna is claimed to provide a narrow vertical beamwidth, the vertical radiation pattern should be checked with the help of a helicopter or balloon.

8. Acknowledgements

The author would like to thank those of his colleagues who read through the draft White Paper and offered useful suggestions and corrections.

9. References and Further Reading

Note that not all of the following are referred to in the previous text.

1. KRAUSS, J D. Antennas (second edition). McGraw-Hill International Editions, ISBN 0-07-100482-3.
2. PAGE, H, 1963. 'An Introduction to the Theory of Aerials.' BBC Research Department Report No. E-085.
3. MORTON, A H, 1966. 'Advanced Electrical Engineering' (Chapter 12). Pitman Press, FS–(T.1310).
4. EINSTEIN, A, 1905. 'On the Electrodynamics of Moving Bodies.' Annalen der Physik, 17:891, 1905.

   The above appears in English translation in the following:

   'Principle of Relativity.' Methuen, 1923. It can also be found on the web:

   http://www.fourmilab.ch/etexts/einstein/specrel/www/

   A good, readable account of magnetism is also available on the web:

   http://www.ph.unimelb.edu.au/~dnj/teaching/160mag/160mag.htm
5. ARRL, 1995. 'The ARRL Antenna Book.' Published by the American Radio Relay League, ISBN 0-87259-206-5.

   The above is a good general introduction to antenna theory. Chapter 3 deals with the effects of the ground. Unfortunately, the ground resistance measurements quoted all apply to the United States.
6. TRAINOTTI, V, 2001. 'Short Medium Frequency AM Antennas.' IEEE Transactions on Broadcasting, Volume 47, September 2001, pages 216–218.

Note Other engineers have been investigating the performance of electrically-short antennas. Many references can be found on the web. The one below contains a further useful bibliography:

http://www.arising.com.au/people/holland/ralph/ShortVert.htm

This page has been left blank deliberately.

Appendix 1: The Relationship between Electricity and Magnetism

One of Einstein's main motivations in developing the Theory of Relativity was to unify electricity and magnetism. Einstein's original paper is available **[4]**, and anyone wishing for mathematical rigour should examine this. However, it is hardly an easy read, and so the following is offered as a simple — perhaps oversimplified — interpretation.

Suppose that two wires s_1 and s_2 are separated by distance r, and that they carry currents I_1 and I_2 respectively. Wire s_1 then experiences a force F_1 defined by Ampère's law:

$$\mathbf{dF_1} = \frac{\mu_0 I_1 I_2}{4\pi r^3} \{\mathbf{ds_1} \times (\mathbf{ds_2} \times \mathbf{r})\}$$

where μ_0 is the permeability of free space.[9] The vector quantities are represented conventionally in bold. In the case of two parallel wires, distance a apart and carrying equal currents in the same direction, the law reduces to

$$\frac{dF}{ds} = \frac{\mu_0 I^2}{2\pi a} \; ;$$

that is, the force per unit length equals $\mu_0 I^2 / 2\pi a$. In the SI system, μ_0 is defined as $4\pi \times 10^{-7}$ exactly, and the formula is then used to define the unit of current.

A convenient if simplistic model of a metal is that of a lattice of fixed positive ions surrounded by a 'sea' of mobile electrons. The electrons are free to move under the influence of an electric field, which might be created by applying a battery across a piece of the metal. If the parallel wires each contain n free electrons per unit length, the average velocity of the electrons is given by

$$v = I / n e,$$ where e is the charge on a single electron.

Because there are so many free electrons in a good conductor, v is generally small for any reasonable value of I.

Now imagine yourself to be an electron moving along one of the parallel wires. If you looked across to the other wire, you would see a large body of fellow electrons moving at the same speed as you; that is, relative to you, the electrons would be stationary. You would also see the lattice of positive ions apparently moving away from you with velocity v. Einstein tells us that this column of ions, of stationary length l, will have an apparent length l′ given by the Lorentz contraction formula:

$$l' = l \{ 1 - (v/c)^2 \}^{1/2}$$, where c is the velocity of light.

Because the column of ions has shrunk, whereas the corresponding column of electrons has not, you see a greater density of ions, and therefore experience an electrostatic attraction. An illustration of this is shown overleaf, in Figure A1.1. The diagram is a picture of the world from the point of view of an electron in Wire 1. However, an ion in Wire 1 will also experience an attraction, since there is apparently a greater density of electrons in Wire 2.

---

9 Ampère formulated his law after making numerous practical measurements.

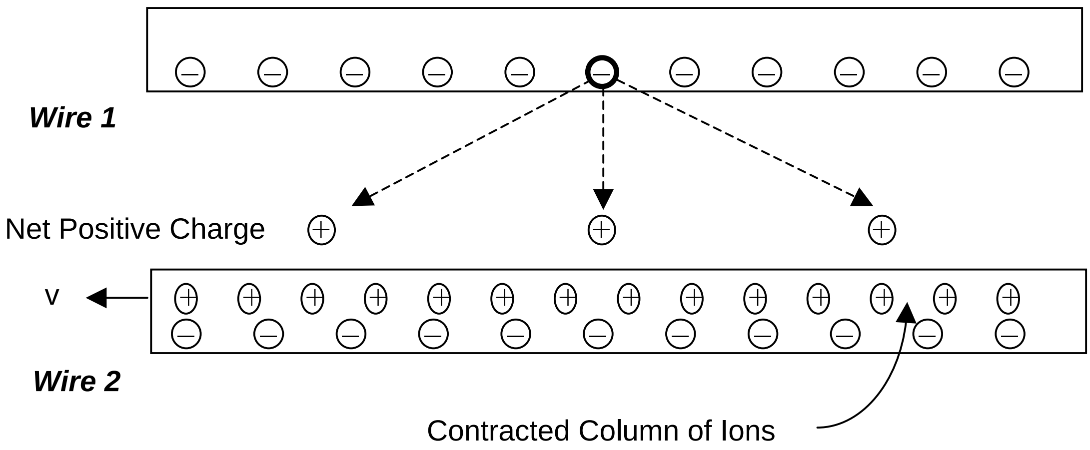


Figure A1.1: Illustration of Attraction between Wires Carrying Electric Current

It is straightforward to calculate the attraction. Firstly, the Lorentz contraction formula can be simplified with the help of the binomial expansion, because v/c is so small:

$$l' \approx l \{ 1 - (1/2) (v/c)^2 \} ;$$

so one unit of length is now shorter by $(1/2) (v/c)^2$. In other words, if there were originally n ions and n electrons per unit length, there now appears to be a surplus of $n (1/2) (v/c)^2$ ions per unit length. Secondly, we need to invoke Coulomb's law.[10] This states that two charges q_1 and q_2, separated by distance r, will attract each other with a force F given by

$$\mathbf{F} = \frac{q_1 q_2 \mathbf{r}}{4\pi\varepsilon_0 r^3}$$

where ε_0 is the permittivity of free space. To calculate the attraction between the electron and the surplus ions in the neighbouring wire, we can substitute e for q_1. q_2 represents the surplus ions. However, because these are distributed, it is necessary to carry out a simple integration:

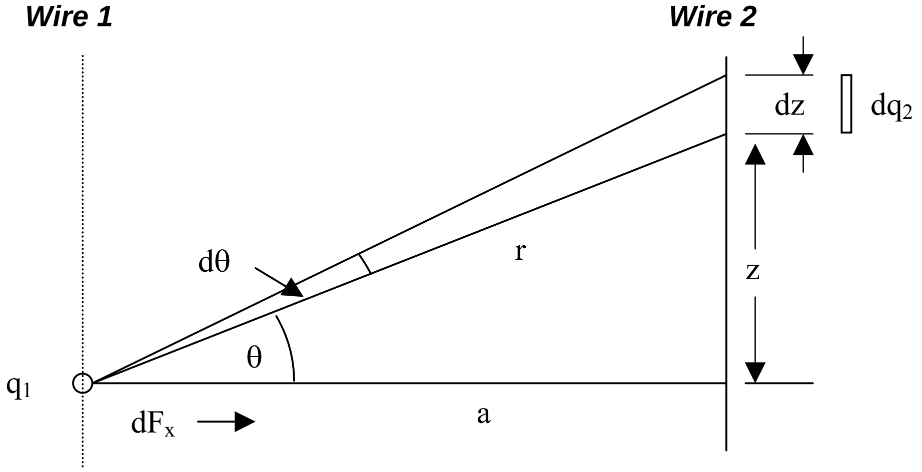


Figure A1.2: Calculation of Attraction between Point and Line Charges

In the diagram, q_1 is taken to be a single point charge in Wire 1, whilst q_2 represents the charge per unit length in Wire 2.

[10] Like Ampère's law, Coulomb's law was the result of experiment.

The horizontal component of the force dF between q_1 and the element of charge dq_2 is given by

$$dF_x = \frac{q_1 (q_2\, dz)}{4\pi\varepsilon_0\, r^2} .\cos\theta .$$

$$= \frac{q_1 q_2}{4\pi\varepsilon_0} . \frac{(\cos^2\theta)}{a^2} . \frac{a\, d\theta}{\cos\theta} ,$$

since $r = a / \cos\theta$, and $dz = r\, d\theta/ \cos\theta = a\, d\theta / \cos^2\theta$.

Hence $$dF_x = \frac{q_1 q_2 \cos\theta\, d\theta}{4\pi\varepsilon_0\, a} .$$

Integrating dF_x between the limits $\theta = -\pi/2$ and $\theta = +\pi/2$ gives the result

$$F_x = q_1 q_2 / 2\pi\varepsilon_0\, a .$$

Substituting e for q_1 and $n\, e\, (1/2)\, (v/c)^2$ for q_2 gives

$$F_x = e\, n\, e\, (1/2)\, (v/c)^2 / 2\pi\varepsilon_0\, a;$$

and substituting $I / n\, e$ for v,

$$F_x = (1/2)\, (1/n)\, (I/c)^2 / 2\pi\varepsilon_0\, a.$$

Since there are n electrons per unit length, the total force per unit length becomes

$$nF_x = (1/2)\, (I/c)^2 / 2\pi\varepsilon_0\, a .$$

So far, we have only considered the attraction between the electrons in the first wire and the apparent excess of positive ions in the second wire. Of course, the situation is similar for the ions in the first wire, which will 'see' an apparent excess of electrons in the second wire; hence the actual force measured will be twice that calculated above, or $(I/c)^2 / 2\pi\varepsilon_0\, a$. Since c equals $1 / \sqrt{(\varepsilon_0 \mu_0)}$, the force can be written as $\mu_0 I^2 / 2\pi a$, which is just as calculated from Ampère's law.

People are often surprised that the electric and magnetic components of an electromagnetic wave are so interdependent. Perhaps the interdependence is less mysterious if one remembers that electricity and magnetism are really different aspects of the same thing.