Chapter 1

Introduction

A communications link consists of a source of information, transmitter electronics including modulators, mixers, power amplifiers, or other components, a transmitting antenna or array, the propagation environment, a receiving antenna or array and associated receiver electronics, and signal processing to detect and decode information from the received signal. Examples range from broadcast radio and wireless handsets to sending commands to a deep space probe millions of miles from earth or controlling a smart drill bit at the bottom of a thousand foot deep oil well. The purpose of this book is to develop the analytical tools required for an end-to-end model of such a communications link, including antenna and propagation effects as well as signal processing.

In communication theory, a link between a source of information and a destination is a channel. The key measures of the goodness of a channel are the signal to noise ratio (SNR) and channel capacity. Channel capacity is the maximum bit rate that can be reliably sent from transmitter to receiver. The capacity is determined by the bandwidth of the channel and the signal to noise ratio (SNR) at the receiver output. The SNR is influenced by the following factors:

Transmitter: Total radiated power, antenna radiation pattern, gain, and polarization.

Propagation environment: Distance between transmitter and receiver, multipath, blockage, loss, noise, and interference.

Receiver: Antenna characteristics, amplifiers and receiver electronics, and signal processing.

The actual data rate achieved when a channel is used in practice is bounded above by the capacity. The achieved data rate is determined by the signal modulation used to send information over the channel, so coding and modulation theory are also important. Our goal is to understand each of these aspects of the electromagnetic propagation channel and to model the overall performance of a communications channel in terms of the transmitter characteristics, propagation environment, and receiver system.

In order to develop a complete channel model, we must consider antenna theory for both transmitting and receiving antennas, specific antenna types and array antennas, noise theory, propagation channels, and communication theory for both single antennas and multi-antenna systems. This will provide the tools necessary to determine the SNR at the output of a communications link, the channel capacity, and the bit error rate realized with the channel for a specific modulation scheme. These tools will also allow synthesis of a communications system which meets a desired performance criterion.

1.1 Link Budget Analysis

A simple tool for propagation analysis is a link budget, or a formula for the SNR at the receiving end of a communication channel in terms of transmitted power, antenna gain, propagation loss, noise, and other factors. The link budget combines all the major factors that determine the SNR for a communications system. In this book, we will develop models for the various contributions to a system link budget. For complex propagation environments and multiple input multiple output (MIMO) communications systems, a simple link budget analysis is inadequate, and for those situations, we will develop more sophisticated channel capacity models.

1.2 Applications

Applications of antennas and propagation modeling include everything from a basic point to point microwave communications link or radio broadcast system to modern technologies such as wireless local area networks, satellite uplinks and downlinks, deep space communications, and MIMO systems. Many of the same principles are applicable to other fields beyond voice and data transmission, such as receivers for radio astronomy observations, magnetic resonance imaging, radar, global position systems, and remote sensing.

Chapter 2

Antennas

As a device that transforms a wave on a transmission line to a wave in the space around the antenna, an antenna has two key properties: the input impedance it presents to the transmission line, and the pattern of the radiated fields. The goal of antenna analysis is to determine the impedance and radiation pattern from the geometry and composition of the antenna structure. Designing an antenna based on desired impedance and radiation properties is antenna synthesis.

Configured as a receiver, the antenna can be modeled as an equivalent voltage or current source connected to a transmission line, with an open circuit voltage or short circuit current induced by an incident field and a given source impedance. Typically, an antenna is modeled as a transmitter, and its receiving properties are inferred using the electromagnetic reciprocity principle.

An antenna radiation problem is a boundary value problem, where the fields radiated by the antenna are determined by Maxwell's equations with material properties given by the shape and composition of the antenna structure and a source excitation connected to the antenna terminals. Maxwell's equations then determine the electromagnetic fields around the antenna. From these fields, the voltage and current at the antenna terminals can be computed to determine the antenna impedance, and the far fields determine the antenna radiation pattern.

2.1 Antenna Analysis

One way to find the fields around the antenna is to use a numerical method to solve the boundary value problem directly. All antenna parameters, including the antenna impedance, can be found in this way. For simple antenna types, it is more convenient to develop approximate formulas for the antenna parameters using analytical techniques.

One of the basic analytical techniques of antenna theory is to model the antenna as an equivalent current distribution, which when impressed in free space radiates the same fields as the antenna structure with a given excitation at the terminals. The radiation integral can be used to find the far fields, from which the radiation pattern and radiation resistance can be computed. For a lossless antenna, the radiation resistance is equal to the real part of the antenna impedance. The current is also sometimes used to estimate the additional part of the antenna resistance due to ohmic losses in the antenna structure. A more sophisticated analysis or a numerical method is usually required to model the antenna reactance and obtain an accurate value for the antenna impedance.

Analytical current models are typically approximate and can be found only for simple antenna geometries. For complex antennas, analytical current models are not available, and numerical methods are used to solve Maxwell's equations and find the field radiated by the antenna.

Common analytical current models and numerical methods used for antenna analysis include:

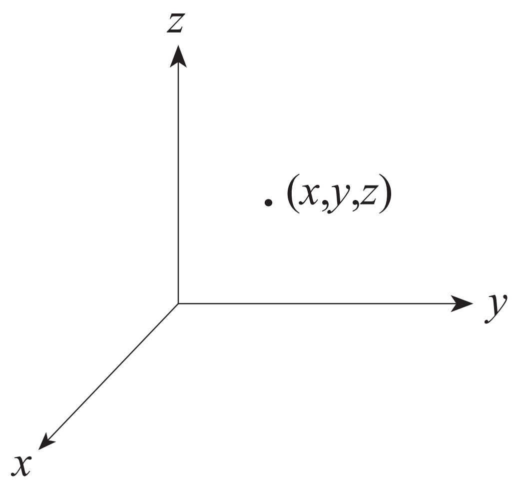

Figure 2.1: The rectangular coordinate system.

Analytical approximations:

Hertzian dipole model (delta function current).

Linear antennas: triangular or sinusoidal current distributions.

Aperture antennas: aperture field is approximated by the incidence field that illuminates the aperture.

Patch antennas: cavity model for fields under the patch.

Numerical methods

1D method of moments (MOM) for thin wires.

Surface method of moments (MOM) for perfect electric conductor (PEC) objects.

Volume method of moments (MOM) for composite dielectric and conducting structures.

Finite difference time domain (FDTD)

Finite element method (FEM)

The analytical approximations provide an equivalent current representation for the antenna, from which the fields radiated by the current source can be found using a Green's function and the radiation integral. A Green's function is the field radiated by a point or delta function source for a given set of boundary conditions. It can be thought of as the impulse response of space. Boundary conditions may include dielectric interfaces, conductors, and the radiation boundary condition at infinity. The most common case is the free space Green's function, which is available in analytic form. The field is then given by a radiation integral, which is the convolution of the Green's function with a current source.

2.1.1 Coordinate Systems

To give explicit numerical values for the sources and fields and to compute the derivatives in Maxwell's equations, a coordinate system is required. The most basic coordinate system is the rectangular coordinate system. More generally, curvilinear coordinate systems can be defined, which simplifies the analysis of electromagnetics problems involving curved geometries such as spheres and cylinders.

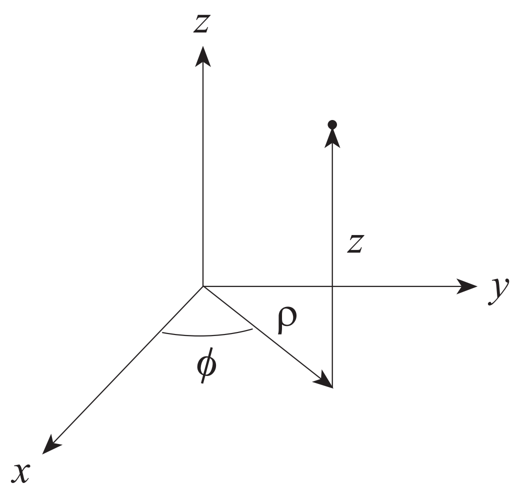

Figure 2.2: The cylindrical coordinate system.

Rectangular Coordinates

The rectangular coordinate system is shown in Figure 2.1. Using this coordinate system, a point in a three-dimensional (3-D) space can be represented by its distances x, y, and z from the origin of the coordinate system along each of three orthogonal axes.

We define the three orthogonal unit vectors $\hat{x}$, $\hat{y}$, and $\hat{z}$, which point in the directions of each of the three coordinate axes, respectively, and which have unit length, so that $\|\hat{x}\| = \|\hat{y}\| = \|\hat{z}\| = 1$. Using these unit vectors, an arbitrary vector $\overline{A}$ can be expressed in terms of its length along each of the axes as

$$\overline{A} = A_x\hat{x} + A_y\hat{y} + A_z\hat{z} \tag{2.1}$$

where A_x, A_y, and A_z are the components of the vector with respect to the rectangular coordinate system.

The dot product of two vectors can be represented as

$$\overline{A} \cdot \overline{B} = A_xB_x + A_yB_y + A_zB_z \tag{2.2}$$

and the cross-product is

$$\overline{A} \times \overline{B} = \hat{x}(A_yB_z - A_zB_y) + \hat{y}(A_zB_x - A_xB_z) + \hat{z}(A_xB_y - A_yB_x) \tag{2.3}$$

The magnitude or length of the vector $\overline{A}$ is

$$\|\overline{A}\| = (\overline{A} \cdot \overline{A})^{1/2} = (A_x^2 + A_y^2 + A_z^2)^{1/2} \tag{2.4}$$

If the components of the vector are complex, then the magnitude is

$$\|\overline{A}\| = (\overline{A} \cdot \overline{A}^*)^{1/2} = (|A_x|^2 + |A_y|^2 + |A_z|^2)^{1/2} \tag{2.5}$$

Cylindrical Coordinates

The circular cylindrical coordinate system is shown in Figure 2.2. The coordinates ρ, ϕ, z are related to the rectangulr coordinates x, y, z by the formulas

$$x = \rho\cos\phi \tag{2.6a}$$

$$y = \rho\sin\phi \tag{2.6b}$$

$$z = z \tag{2.6c}$$

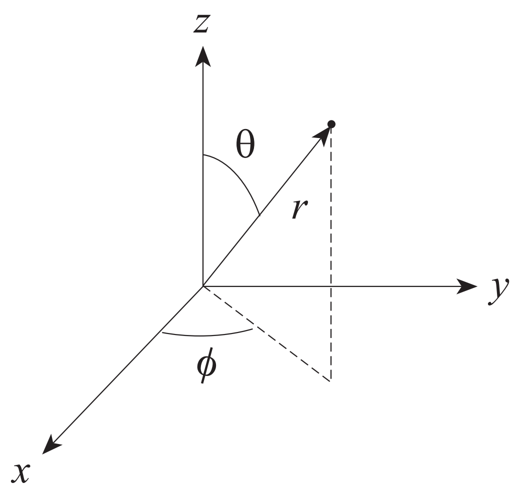

Figure 2.3: The spherical coordinate system.

and

$$\rho = \sqrt{x^2 + y^2} \tag{2.7a}$$

$$\phi = \tan^{-1}\frac{y}{x} \tag{2.7b}$$

$$z = z \tag{2.7c}$$

The unit vectors for this coordinate system are $\hat{\rho}$, $\hat{\phi}$, and $\hat{z}$. Unlike the unit vectors associated with the rectangular coordinate system, which are fixed vectors and independent of position, the vectors $\hat{\rho}$ and $\hat{\phi}$ change direction depending on the angle ϕ. An arbitrary vector is expressed in terms of its cylindrical components as

$$\overline{A} = A_\rho\hat{\rho} + A_\phi\hat{\phi} + A_z\hat{z} \tag{2.8}$$

A vector is a coordinate-independent object, so a given vector can be represented in any given coordinate system by transforming its components in one coordinate system to a new set of components with respect to the basis vectors of another coordinate system. Tables of transformations among the representations of vector components in rectangular, cylindrical, and spherical coordinates can be found in many references, including [1].

Spherical Coordinates

The spherical cylindrical coordinate system is shown in Figure 2.3. The coordinates r, θ, ϕ are related to the rectangular coordinates x, y, z via

$$x = r\sin\theta\cos\phi \tag{2.9a}$$

$$y = r\sin\theta\sin\phi \tag{2.9b}$$

$$z = r\cos\theta \tag{2.9c}$$

and

$$r = \sqrt{x^2 + y^2 + z^2} \tag{2.10a}$$

$$\theta = \tan^{-1}\frac{\sqrt{x^2 + y^2}}{z} \tag{2.10b}$$

$$\phi = \tan^{-1}\frac{y}{x} \tag{2.10c}$$

The unit vectors for this coordinate system are $\hat{r}$, $\hat{\theta}$, and $\hat{\phi}$. An arbitrary vector can be expressed as

$$\overline{A} = A_r \hat{r} + A_\theta \hat{\theta} + A_\phi \hat{\phi} \tag{2.11}$$

2.1.2 Vector Analysis

Electromagnetic fields are represented mathematically as vector fields. A vector field is a mapping from a space such as R^3 to a vector space. At each point in the space, the mapping evaluates to a vector. A vector field can be represented as a vector with components that are functions of position, so that in the rectangular coordinate system,

$$\overline{E}(x, y, z) = E_x(x, y, z)\hat{x} + E_y(x, y, z)\hat{y} + E_z(x, y, z)\hat{z} \tag{2.12}$$

where $\hat{x}$ is a vector of length one, or unit vector, pointing in the direction of increase of the x coordinate. The other unit vectors $\hat{y}$ and $\hat{z}$ are defined similarly.

The variation of a vector field with respect to position can be analyzed using the vector derivative operator

$$\nabla = \frac{\partial}{\partial x}\hat{x} + \frac{\partial}{\partial y}\hat{y} + \frac{\partial}{\partial z}\hat{z} \tag{2.13}$$

This operator can be applied to a scalar $f(x, y, z)$ to compute the gradient

$$\nabla f(x, y, z) = \frac{\partial f}{\partial x}\hat{x} + \frac{\partial f}{\partial y}\hat{y} + \frac{\partial f}{\partial z}\hat{z} \tag{2.14}$$

The operator can be applied to a vector field using the cross product to produce the curl operation

$$\nabla \times \overline{A} = \hat{x}\left(\frac{\partial A_z}{\partial y} - \frac{\partial A_y}{\partial z}\right) + \hat{y}\left(\frac{\partial A_x}{\partial z} - \frac{\partial A_z}{\partial x}\right) + \hat{z}\left(\frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y}\right) \tag{2.15}$$

Or, it can be applied to a vector field using the dot product, to produce the divergence

$$\nabla \cdot \overline{A} = \frac{\partial A_x}{\partial x} + \frac{\partial A_y}{\partial y} + \frac{\partial A_z}{\partial z} \tag{2.16}$$

of the vector field.

For a vector field expressed in terms of its cylindrical components,

$$\nabla \times \overline{A} = \hat{\rho}\left(\frac{1}{\rho}\frac{\partial A_z}{\partial \phi} - \frac{\partial A_\phi}{\partial z}\right) + \hat{\phi}\left(\frac{\partial A_\rho}{\partial z} - \frac{\partial A_z}{\partial \rho}\right) + \hat{z}\frac{1}{\rho}\left(\frac{\partial}{\partial \rho}\rho A_\phi - \frac{\partial A_\rho}{\partial \phi}\right) \tag{2.17a}$$

$$\nabla \cdot \overline{A} = \frac{1}{\rho}\frac{\partial \rho A_\rho}{\partial \rho} + \frac{1}{\rho}\frac{\partial A_\phi}{\partial \phi} + \frac{\partial A_z}{\partial z} \tag{2.17b}$$

In the spherical coordinate system,

$$\nabla \times \overline{A} = \frac{\hat{r}}{r\sin\theta}\left[\frac{\partial}{\partial \theta}(A_\phi \sin\theta) - \frac{\partial A_\theta}{\partial \phi}\right] + \frac{\hat{\theta}}{r}\left[\frac{1}{\sin\theta}\frac{\partial A_r}{\partial \phi} - \frac{\partial}{\partial r}(rA_\phi)\right] + \frac{\hat{\phi}}{r}\left[\frac{\partial}{\partial r}(rA_\theta) - \frac{\partial A_r}{\partial \theta}\right] \tag{2.18a}$$

$$\nabla \cdot \overline{A} = \frac{1}{r^2}\frac{\partial}{\partial r}r^2 A_r + \frac{1}{r\sin\theta}\frac{\partial}{\partial \theta}\sin\theta A_\theta + \frac{1}{r\sin\theta}\frac{\partial}{\partial \phi}A_\phi \tag{2.18b}$$

The gradient, curl, and divergence satisfy the identities

$$\nabla \times \nabla f = 0 \tag{2.19}$$

$$\nabla \cdot (\nabla \times \overline{A}) = 0 \tag{2.20}$$

as long as the function f and the components of the vector field $\overline{A}$ are smooth enough that partial derivative operators can be interchanged without changing the result.

The Laplacian operator in the rectangular coordinate system is

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial z^2} + \frac{\partial^2}{\partial z^2} \tag{2.21}$$

When applied to a vector field $\overline{A}$, the Laplacian operator is related to the gradient, curl, and divergence operators by

$$\nabla^2 \overline{A} = -\nabla \times \nabla \times \overline{A} + \nabla\nabla \cdot \overline{A} \tag{2.22}$$

For a scalar quantity, this simplifies to

$$\nabla^2 f = \nabla \cdot \nabla f \tag{2.23}$$

Another convenient notation is the position vector

$$\overline{r} = x\hat{x} + y\hat{y} + z\hat{z} \tag{2.24}$$

This is not a true vector in the usual sense, but instead is a compact way to represent the point with coordinates (x, y, z) in three-dimensional space. Often, it is useful to represent the components of this vector in the spherical coordinate system, so that

$$\overline{r} = r\sin\theta\cos\phi\hat{x} + r\sin\theta\sin\phi\hat{y} + r\cos\theta\hat{z} \tag{2.25}$$

It is also useful to define the unit vector

$$\hat{r} = \frac{\overline{r}}{\|\overline{r}\|} = \sin\theta\cos\phi\hat{x} + \sin\theta\sin\phi\hat{y} + \cos\theta\hat{z} \tag{2.26}$$

The length $\|\overline{r}\| = \sqrt{x^2 + y^2 + z^2}$ is the distance from the point (x, y, z) to the origin, which is also given by the symbol r, or the radial coordinate in the spherical coordinate system. The unit vector $\hat{r}$ points from the origin to the point (x, y, z), but has unit length. With this notation, we can represent position-dependent functions such as the phase of a plane wave field solution in a compact, intuitive way using the dot product of $\overline{r}$ and another vector.

2.1.3 Maxwell's Equations

Maxwell's equations for the electromagnetic field are expressed in terms of electric and magnetic vector fields $\overline{E}$, $\overline{D}$, $\overline{H}$, and $\overline{B}$, and the source quantities $\overline{J}$ and ρ. These vector fields and one scalar quantity are defined in Table 2.1 along with the units of each. Other systems of units are available, but the International System of Units (SI) convention is used throughout this book.

The vector fields used to represent electromagnetic fields are governed by Maxwell's equations. These equations can be written in differential or integral form. For wave propagation analysis, it is convenient to begin with the differential or point forms of the equations. Maxwell's equations in point form are

$$\nabla \times \overline{E} = -\frac{\partial \overline{B}}{\partial t} \tag{2.27a}$$

$$\nabla \times \overline{H} = \frac{\partial \overline{D}}{\partial t} + \overline{J} \tag{2.27b}$$

$$\nabla \cdot \overline{D} = \rho \tag{2.27c}$$

$$\nabla \cdot \overline{B} = 0 \tag{2.27d}$$

The first pair of equations are Faraday's and Ampére's laws, respectively, and the second pair are Gauss's laws for the electric and magnetic flux density. This set of coupled differential equations governs the behavior of static fields due to fixed charge sources as well as the waves radiated by dynamic sources.

Table 2.1: The Field and Source Quantities of Electromagnetics

Name	Symbol	Dimension
Electric Field Intensity	$\overline{E}$	V/m
Magnetic Field Intensity	$\overline{H}$	A/m
Electric Flux Density	$\overline{D}$	C/m^2
Magnetic Flux Density	$\overline{B}$	Wb/m^2
Electric Current Density	$\overline{J}$	A/m^2
Electric Charge Density	ρ	C/m^3

2.1.4 Constitutive Relations

In free space or a vacuum, the flux density and field intensity vectors are related by the constitutive relations

$$\overline{D} = \epsilon_0 \overline{E} \tag{2.28a}$$
$$\overline{B} = \mu_0 \overline{H} \tag{2.28b}$$

where the permittivity of free space is $\epsilon_0 \simeq 8.854 \times 10^{-12}$ F/m and the permeability is $\mu_0 = 4\pi \times 10^{-7}$ A/m. The constitutive relations can be modified to introduce models for the effect on the electromagnetic field of materials such as dielectrics. In an inhomogeneous dielectric medium, the permeability and permittivity are functions of position, so that the constitutive relations become

$$\overline{D} = \epsilon(x, y, z)\overline{E} \tag{2.29a}$$
$$\overline{B} = \mu(x, y, z)\overline{H} \tag{2.29b}$$

The relative permittivity ϵ_r and relative permeability μ_r are defined according to

$$\epsilon = \epsilon_r \epsilon_0 \tag{2.30a}$$
$$\mu = \mu_r \mu_0 \tag{2.30b}$$

Conductive materials can be modeled by adding a conduction current or induced current term to the right-hand side of Ampére's law. The induced current density is

$$\overline{J} = \sigma(x, y, z)\overline{E} \tag{2.31}$$

where σ is the conductivity of the material. When analyzing electromagnetic systems, it is often convenient to introduce an impressed current to represent sources driven by excitations outside of the modeled region of interest. Strictly speaking, all electric currents are induced by electric fields according to (2.31). For engineering work, it is impractical to model all currents using (2.31). The electric fields inside the signal generator that drive a current in a transmission line, for example, are nearly completely shielded from the fields radiated by an antenna, so it makes little sense to model the antenna, transmission line, and signal generator as a complete system. It is more efficient to simply consider the current source that excites the antenna to be fixed. A fixed current that excites a system is referred to as an impressed current. Mathematically, an impressed current is a given source term in Ampère's law (2.27b), or a forcing function in differential equation terminology. The total current on the right-hand side of Ampère's law is the sum of the induced and impressed currents

$$\overline{J} = \overline{J}_{\text{ind}} + \overline{J}_{\text{imp}}$$

where $\overline{J}_{\text{ind}} = \sigma\overline{E}$.

2.1.5 Time- and Frequency-Domain Representations

For many applications, the signals transmitted in communications systems are narrowband. For narrowband signals, it is often a good approximation to treat the signal as a pure tone or sinusoidal waveform. This approximation is based on the assumption that the antenna characteristics in the system and the propagation environment are close to constant over the bandwidth of the signal. If this assumption holds, then Maxwell's equations can be transformed to what is variously known as the frequency domain, sinusoidal steady state form or phasor domain. This representation simplifies the mathematical treatment of electromagnetic fields for narrowband systems.

The time-domain and phasor forms of the electric field intensity are related by

$$\overline{E}(x, y, z, t) = \mathrm{Re}[\underline{\overline{E}}(x, y, z)e^{j\omega t}] \tag{2.32}$$

where $\overline{E}(x, y, z, t)$ is the time-varying electric field and $\underline{\overline{E}}(x, y, z)$ is the phasor representation. In some treatments, particulary in the physics literature, the time variation is of the form $e^{-i\omega t}$, in which case phasors are complex conjugated relative to the convention of (2.32). This same definition holds for the electric flux density, magnetic field intensity and flux density, and the source quantities $\overline{J}$ and ρ.

Because we are assuming that the signal is a pure sinusoid, the phasor in (2.32) is constant. If the signal were modulated, so that it is not a pure sinusoid, then the complex quantity $\underline{\overline{E}}(x, y, z)$ would be a function of time, $\underline{\overline{E}}(x, y, z, t)$. This is referred to as the complex analytic or complex baseband signal representation, and is covered in Section 8.1.8. For much of this book, the pure sinusoidal approximation is valid, and the phasor is independent of time. In (2.32), an underline is used to distinguish the phasor quantity from the time varying field intensity. Some texts use a different typeface to represent time-domain fields and phasors, such as $\overline{\mathcal{E}}(x, y, z, t)$ in the time domain and $\overline{E}(x, y, z)$ for phasors. Because we will use the phasor form of the field quantities exclusively in this text, the underline will be omitted and the phasor electric field represented simply as $\overline{E}$.

In the phasor or frequency domain, Maxwell's equations are

$$\nabla \times \overline{E} = -j\omega \overline{B} \tag{2.33a}$$

$$\nabla \times \overline{H} = j\omega \overline{D} + \overline{J} \tag{2.33b}$$

$$\nabla \cdot \overline{D} = \rho \tag{2.33c}$$

$$\nabla \cdot \overline{B} = 0 \tag{2.33d}$$

where all field and source quantities are complex-valued vectors or scalars. The time derivatives in (2.27) are no longer present and have been replaced by factors of $j\omega$. For the remainder of this book, we will use Maxwell's equations in frequency domain form.

2.1.6 Free Space Green's Function

A fundamental tool used to analyze electromagnetic systems is the field radiated by a point source in a system with a given set of boundary conditions. This is referred to as a Green's function. The fields radiated by an arbitrary, distributed source can be found by convolving the source distribution with the Green's function.

One approach to finding the fields radiated by a point source is to transform the first order system of Maxwell's equations into a single second order partial differential equation (PDE). Taking the curl of both sides of Faraday's law (2.33a) and using the constitutive relation (2.30b) yields

$$\nabla \times \nabla \times \overline{E} = -j\omega\mu\nabla \times \overline{H} \tag{2.34}$$

We have assumed that the permeability μ is invariant of position (the medium or material in which the wave propagates is spatially homogeneous), which allowed the curl operator to be moved past the factor of μ on the right-hand side of this expression. Substituting Ampère's law into this expression with $\overline{J} = 0$ leads to

$$\nabla \times \nabla \times \overline{E} = \omega^2 \mu\epsilon \overline{E} - j\omega\mu\overline{J} \tag{2.35}$$

Here, we have assumed that the material parameters are constant, which allowed a time derivative to be moved past the factor of ϵ.

Since the Laplacian operator as given by (2.21) is much simpler than the double curl operator in (2.35), the next step is to apply the identity (2.22) to eliminate the double curl. Using Gauss's law for the electric field, and assuming that there are no sources ($\rho = 0$) and the medium is homogeneous, we have

$$\nabla \cdot \overline{E} = \frac{1}{\epsilon}\nabla \cdot \overline{D} = 0 \tag{2.36}$$

so that the identity (2.22) applied to $\overline{E}$ becomes $\nabla \times \nabla \times \overline{E} = -\nabla^2 \overline{E}$. Using this to simplify (2.35) produces the Helmholtz equation

$$\nabla^2 \overline{E} + \omega^2 \epsilon\mu \overline{E} = j\omega\mu\overline{J} \tag{2.37}$$

We define the wavenumber to be $k = \omega\sqrt{\epsilon\mu}$, or $k = \omega/c$, where the constant c is the speed of wave propagation,

$$c = \frac{1}{\sqrt{\mu\epsilon}} \tag{2.38}$$

In a vacuum, $c \simeq 2.998 \times 10^8$m/s. This simplifies the Helmholtz equation to

$$[\nabla^2 + k^2]\overline{E}(\overline{r}) = j\omega\mu\overline{J}(\overline{r}) \tag{2.39}$$

which is the governing PDE for the electric field intensity radiated by a time-harmonic source.

To find a Green's function for this PDE, we need to solve for the electric field radiated by a point source of the form

$$\overline{J}(\overline{r}) = \hat{p}\delta(\overline{r} - \overline{r}') \tag{2.40}$$

where $\overline{r}'$ is the location of the source and $\hat{p}$ is the orientation, and the electric field also satisfies a given boundary condition, either at a conductive surface, a radiation boundary condition at infinity, or some other condition that together with Maxwell's equations uniquely defines the solution to the boundary value problem. The Green's function is the electric field intensity vector as a function of position $\overline{r}$, the location of the point source $\overline{r}'$, and the orientation $\hat{p}$ of the source. With the Green's function, the field due to an arbitrary source can be found by convolving the source with the Green's function.

Unfortunately, there are mathematical difficulties associated with finding the Green's function for either (2.35) or (2.39). The derivative operator in (2.35) is complicated, so it is difficult to solve this equation directly for the Green's function. Equation (2.39) has a simpler derivative operator, but the Helmholtz equation has more solutions than Maxwell's equations (longitudinal waves), so those nonphysical solutions must be eliminated from the convolution of the Green's function with the source to obtain a valid electric field.

To overcome these difficulties, we can define an auxiliary potential that also satisfies a Helmholtz type PDE from which valid electric and magnetic fields can be derived. Gauss's law for the magnetic flux density is

$$\nabla \cdot \overline{B} = 0 \tag{2.41}$$

Using a theorem from differential geometry, it follows that $\overline{B}$ is the curl of some vector field, which means that the magnetic flux density can be represented in the form

$$\overline{B} = \nabla \times \overline{A} \tag{2.42}$$

where $\overline{A}$ is called the magnetic vector potential. Using this in Faraday's law, we find that

$$\nabla \times (\overline{E} + j\omega\overline{A}) = 0 \tag{2.43}$$

Using another theorem from differential geometry, the quantity in parenthesis in this expression must be the gradient of some scalar function, so that

$$\overline{E} + j\omega\overline{A} = -\nabla\phi \tag{2.44}$$

In the static case ($\omega = 0$), ϕ is the electric potential. The negative sign in (2.44) is included so that ϕ agrees with the usual sign convention for electric potential.

Using (2.44) in Ampere's law leads to

$$[-\nabla \times \nabla \times +k^2]\overline{A} = j\omega\mu\epsilon\nabla\phi - \mu\overline{J} \tag{2.45}$$

Using the identity (2.22) for the Laplacian operator,

$$[\nabla^2 + k^2]\overline{A} = -\mu\overline{J} + j\omega\mu\epsilon\nabla\phi + \nabla\nabla\cdot\overline{A} \tag{2.46}$$

The last two terms on the right are inconvenient, but we can eliminate them. Since there are vector fields with zero curl, it follows that there are many vector fields $\overline{A}$ that satisfy (2.42) for a given magnetic flux density $\overline{B}$. If we choose the particular vector potential for which

$$\nabla\cdot\overline{A} = -j\omega\epsilon\mu\phi \tag{2.47}$$

then the undesirable terms in (2.46) are zero. This choice for the extra degree of freedom in the magnetic vector potential is known as the Lorenz gauge. Using this in (2.46) leads to the Helmholtz equation

$$[\nabla^2 + k^2]\overline{A} = -\mu\overline{J} \tag{2.48}$$

This is similar in form to (2.39), but the solution to the PDE is the magnetic vector potential, rather than the electric field intensity. If we solve (2.48) for the magnetic vector potential and then find the electric field using (2.47) and (2.44), we are guaranteed to obtain a valid electromagnetic field solution. $\overline{A}$ and $\overline{E}$ are similar as vector fields, except that the transformation from $\overline{A}$ to $\overline{E}$ removes the longitudinal wave part of the magnetic vector potential, so that $\overline{E}$ represents a true physical solution to Maxwell's equations.

Scalar Green's Function

To find a Green's function for Maxwell's equations, we first need to find a second Green's function for the PDE (2.39). This second Green's function allows us to find the magnetic vector potential due to an arbitrary source, and is referred to as the scalar Green's function.

Finding a Green's function for a PDE means finding solution to the PDE with the right-hand side a point source or delta function. The delta function is a three-dimensional source, which can be written as

$$\overline{J}(\overline{r}) = \delta(x - x')\delta(y - y')\delta(z - z') \tag{2.49}$$

where the coordinates (x', y', z') give the location of the point source. As a shorthand notation, we can express the 3D source as

$$\delta(x - x')\delta(y - y')\delta(z - z') = \delta(\overline{r} - \overline{r}') \tag{2.50}$$

where $\overline{r}' = x'\hat{x} + y'\hat{y} + z'\hat{z}$. A boundary value problem requires both a PDE and a boundary condition. We will take the boundary condition to be the radiation boundary condition, which means that fields propagate outwards as the distance from the source approaches infinity.

We will label the scalar Green's function or solution $\overline{A}(\overline{r})$ for a point source located at $\overline{r}'$ as $g(\overline{r}, \overline{r}')$. The scalar Green's function is defined by the PDE

$$[\nabla^2 + k^2]g(\overline{r}, \overline{r}') = -\delta(\overline{r} - \overline{r}') \tag{2.51}$$

Since free space is homogeneous, the solution to the PDE is shift invariant. Without loss of generality, we can shift $\overline{r}'$ temporarily to zero, so that $g(\overline{r}, 0) = g(\overline{r}) = g(r)$, and we have

$$[\nabla^2 + k^2]g(r) = -\delta(\overline{r}) \tag{2.52}$$

With this adjustment, the Laplacian becomes considerably simpler, since g is now only a function of r. For $r > 0$, the PDE reduces to the ordinary differential equation (ODE)

$$\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial g}{\partial r}\right) + k^2 g(r) = 0 \tag{2.53}$$

Using a change of variables, we can transform this to a more familiar and easily solved ODE. If we let $u(r) = rg(r)$, then the differential equation becomes

$$\frac{d^2}{dr^2}u(r) + k^2 u(r) = 0 \tag{2.54}$$

which has the general solution

$$u(r) = Ae^{-jkr} + Be^{jkr} \tag{2.55}$$

Transforming back to the original variable, we arrive at the general solution

$$g(r) = A\frac{e^{-jkr}}{r} + B\frac{e^{jkr}}{r} \tag{2.56}$$

Since the original ODE is second order, the general solution has two independent terms with unknown coefficients A and B. The values of these coefficients are fixed by the boundary condition and the strength of the delta function source.

The radiation boundary condition at infinity implies that waves must be outgoing as the distance r from the source becomes large, so we must have $B = 0$, and the Green's function solution reduces to

$$g(r) = A\frac{e^{-jkr}}{r} \tag{2.57}$$

It now remains to find the constant A.

We will do this by ensuring that the left-hand side of (2.52) integrates to -1 over a volume containing the origin. Integrating both sides of (2.52) over a ball V of radius r centered at the origin leads to

$$\int_V [\nabla^2 + k^2]A\frac{e^{-jkr}}{r}\,d\overline{r} = -\int_V \delta(\overline{r} - \overline{r}')\,d\overline{r} \tag{2.58}$$

By integrating each of the delta functions in (2.50) over x', y', and z', respectively, the right-hand side evaluates to -1. On the left-hand side of (2.58), the integral of the k^2 term of the Helmholtz operator applied to the Green's function vanishes as the radius r becomes small, whereas the ∇^2 term is more strongly singular, and when integrated over V evaluates to a constant, no matter how small the radius of the ball. In fact, $\nabla^2 r^{-1}$ evaluates to a delta function at the origin, as implied by (2.52). By this reasoning, if the radius of V is small, the left-hand side of (2.58) evaluates to

$$\begin{aligned}\int_V [\nabla^2 + k^2] A \frac{e^{-jkr}}{r}\, d\overline{r} &\simeq \int_V \nabla^2 A \frac{1}{r}\, d\overline{r} \\ &= A \int_V \nabla \cdot \nabla \frac{1}{r}\, d\overline{r} \\ &= A \oint_S \nabla \frac{1}{r} \cdot d\overline{S} \\ &= -A \oint_S \frac{1}{r^2} r^2 \sin\theta \, d\theta\, d\phi \\ &= -4\pi A\end{aligned}$$

where we have made use of the divergence theorem to transform the volume integral into a surface integral. Using this result together with (2.58) shows that the value of the constant is $A = 1/(4\pi)$. Shifting the source point from the origin back to $\overline{r}'$ leads to the final result

$$g(\overline{r}, \overline{r}') = \frac{e^{-jk|\overline{r}-\overline{r}'|}}{4\pi|\overline{r}-\overline{r}'|} \tag{2.59}$$

where $|\overline{r} - \overline{r}'|$ is the distance between the points (x, y, z) and (x', y', z'). This is the scalar free space Green's function, or the Green's function for the Helmholtz equation with constant k and the radiation boundary condition at infinity.

2.1.7 Radiation Integral

A linear system can be characterized by its output for a delta function input, or the impulse response. The impulse response can be used to find the output of the system for an arbitrary input, by convolving the input with the impulse response. The boundary value problem consisting of Maxwell's equations as the governing PDE together with the radiation boundary condition can be viewed as a linear system, where the input is an impressed current source and the output is the field radiated by the source. A Green's function can be viewed as the spatial impulse response of this system. The field radiated by an arbitrary source can be found by convolving the source distribution with the Green's function. Physically, the convolution process adds up the fields radiated by many small point sources that combine to make up the source distribution.

By this reasoning, the magnetic vector potential associated with an arbitrary source distribution $\overline{J}$ is given by the integral of the scalar Green's function $g(\overline{r}, \overline{r}')$ weighted by the source distribution $\overline{J}(\overline{r})$. This leads to the radiation integral

$$\overline{A}(\overline{r}) = \mu \int g(\overline{r}, \overline{r}') \overline{J}(\overline{r}')\, d\overline{r}' \tag{2.60}$$

Since free space is a shift-invariant medium, the Green's function can be written in the form $g(\overline{r} - \overline{r}')$, which places the radiation integral into a convolution form. This can be seen in (2.59), which is only a function of the distance between $\overline{r}$ and $\overline{r}'$, and not the absolute locations of either point. The shift invariance of empty space is analogous to a time-invariant linear system, for which the impulse response $h(t, t')$ is only a function of the difference between t and t', so that the impulse response can be given in the form $h(t - t')$. In free space, the radiation integral for $\overline{A}$ can be simplified slightly to

$$\overline{A}(\overline{r}) = \mu \int g(\overline{r} - \overline{r}') \overline{J}(\overline{r}')\, d\overline{r}' \tag{2.61}$$

If inhomogeneous materials are added to the problem, the Green's function is no longer given by (2.59) nor is it shift-invariant.

The electric field can be found in terms of the magnetic vector potential using (2.44) and the Lorenz gauge, so that

$$\begin{aligned}\overline{E} &= -j\omega\overline{A} - \nabla\phi \\ &= -j\omega\overline{A} + \frac{1}{j\omega\epsilon\mu}\nabla\nabla\cdot\overline{A} \\ &= -j\omega\left[1 + \frac{1}{k^2}\nabla\nabla\cdot\right]\overline{A}\end{aligned} \tag{2.62}$$

Inserting (2.60) for the vector potential,

$$\overline{E}(\overline{r}) = -j\omega\mu\left[1 + \frac{1}{k^2}\nabla\nabla\cdot\right]\int g(\overline{r},\overline{r}')\overline{J}(\overline{r}')\,d\overline{r}' \tag{2.63}$$

This is the free space radiation integral for the electric field in terms of the scalar Green's function and the electric current density. This expression is similar to (2.60) for the magnetic vector potential, which shows the close connection between $\overline{E}$ and $\overline{A}$. The component of $\overline{E}$ that arises from the unity term in the square brackets is identical to (2.60) with the exception of an additional scale factor of $-j\omega$. The key difference is that the "$\nabla\nabla$" term in (2.63) removes the longitudinal wave component of the magnetic vector potential, making $\overline{E}$ a valid solution to Maxwell's equations.

2.1.8 Far Field Approximation

Often in antenna analysis we are only interested in the fields far from the antenna. In this case, we can simplify the radiation integral considerably. The approximation is based on the first two terms of the expansion of a binomial,

$$(1+x)^p \simeq 1 + px \tag{2.64}$$

If the source is near the origin and the field observation point $\overline{r}$ is far from the origin, then we can use this approximation for the binomial twice to obtain the result

$$\begin{aligned}|\overline{r} - \overline{r}'| &= \sqrt{(x-x')^2 + (y-y')^2 + (z-z')^2} \\ &\simeq \sqrt{x^2 - 2xx' + y^2 - 2yy' + z^2 - 2zz'} \\ &= \sqrt{r^2 - 2\overline{r}\cdot\overline{r}'} \\ &= r\sqrt{1 - 2\hat{r}\cdot\overline{r}'/r} \\ &\simeq r(1 - \hat{r}\cdot\overline{r}'/r) \\ &= r - \hat{r}\cdot\overline{r}'\end{aligned} \tag{2.65}$$

In making this approximation, we have dropped terms in a larger Taylor expansion of order $1/r^2$ or smaller and retained the constant terms and terms that behave as $1/r$ as $r \to \infty$. In other words, the far field approximation is accurate to order $1/r$ as r becomes large.

The next step is to use the far field expansion (2.65) in the scalar Green's function (2.59). The distance $|\overline{r} = \overline{r}'|$ appears twice in the scalar Green's function, once in the phase and again in the denominator, but we do not need to use approximations of the same accuracy in both places. In the phase term, a small offset matters even if the wave has propagated a long distance, so we must use both terms of (2.65). The denominator does not vary as much if the distance $|\overline{r} = \overline{r}'|$ changes slightly, so there we only need the leading term of the far field expansion. Making these substitutions into the scalar Green's function leads to

$$g(\overline{r}, \overline{r}') \simeq \frac{e^{-jkr}}{4\pi r} e^{jk\hat{r}\cdot\overline{r}'} \tag{2.66}$$

Another way to look at this is that the far field approximation is accurate to order $1/r$, and when inserted into the denominator of the scalar Green's function the second term of the far field expansion leads to a correction that is only of order $1/r^2$, so it can be dropped.

The derivative operators in (2.63) can also be simplified when the observation point is far from the source. The terms in the gradient of the scalar Green's function in the spherical coordinate system can be divided into terms of order $1/r$ and higher order terms according to

$$\begin{aligned}
\nabla \frac{e^{-jkr}}{r} &= \hat{r} \frac{\partial}{\partial r} \frac{e^{-jkr}}{r} + O(1/r^2) \\
&= \hat{r} \left(-jk \frac{e^{-jkr}}{r} - \frac{e^{-jkr}}{r^2} \right) + O(1/r^2) \\
&\simeq -jk\hat{r} \frac{e^{-jkr}}{r}
\end{aligned}$$

This result suggests that the ∇ operator can be replaced with $-jk\hat{r}$ when r is large.

Using this approximation in the radiation integral (2.63) leads to the far field radiation integral

$$\overline{E}(\overline{r}) = -j\omega\mu(1 - \hat{r}\hat{r}\cdot) \frac{e^{-jkr}}{4\pi r} \int e^{jk\hat{r}\cdot\overline{r}'} \overline{J}(\overline{r}')\, d\overline{r}' \tag{2.67}$$

Each term in this expression has a physical meaning. The leading constant adjusts the units of the electric field intensity. The $\hat{r}\hat{r}$ term subtracts out waves with electric field in the r direction, since these are longitudinal waves and are not valid solutions of Maxwell's equation. The identity term in $1 - \hat{r}\hat{r}\cdot$ is what we would have obtained if we had solved (2.39) directly without the use of the magnetic vector potential. The term e^{-jkr} represents the phase of a spherical wave as it moves away from the origin, and the factor of $1/r$ accounts for spreading of energy radiated by the source over a sphere of radius r.

Vector Current Moment

The remaining integration at the far right of (2.67) is essentially a Fourier transform of the source distribution. This integral is referred to as the vector current moment $\overline{N}$, or

$$\overline{N}(\hat{r}) = \int e^{jk\hat{r}\cdot\overline{r}'} \overline{J}(\overline{r}')\, d\overline{r}' \tag{2.68}$$

After the source point $\overline{r}'$ is integrated out, the result of the integration depends on the unit vector $\hat{r}$, which is

$$\hat{r} = \hat{x}\sin\theta\cos\phi + \hat{y}\sin\theta\sin\phi + \hat{z}\cos\theta \tag{2.69}$$

It can be seen that $\hat{r}$ depends only on the spherical angles θ and ϕ. The vector current moment therefore captures the angular dependance of the fields radiated by the source $\overline{J}$, and it can be written alternately as $\overline{N}(\theta, \phi)$. The integral can be viewed as a transformation from the spatial dependence of the source distribution to the angular dependence of the field. Comparing this integral with the Fourier transform of a time-domain signal, the spatial dependence of the source is analogous to time, and the angles θ and ϕ are analogous to frequency. This connection between the integral in (2.68) and the temporal Fourier transform means that all of the intuition, theorems, results, and concepts associated with the Fourier transform of a time signal can be used to understand the angular dependence of the fields radiated by a given source. As an example, a source with a small spatial extent, such as a delta function or point source, corresponds to a radiated field with very broad angular distribution. This is analogous to the Fourier transform of a delta function in time leading to a constant in frequency. Conversely, a signal that is nearly constant in time corresponds to a narrowband signal with limited frequency content, which implies that a very large antenna aperture is required to radiate a narrow beam pattern in angle. We will return to this concept later when discussing specific types of antennas.

It often is convenient in practice to first compute the vector current moment, and then find the electric field in terms of $\overline{N}$. In terms of the vector current moment, the far electric field intensity is

$$\begin{aligned}\overline{E} &= -j\omega\mu(1 - \hat{r}\hat{r}\cdot)\frac{e^{-jkr}}{4\pi r}(\hat{\theta}N_\theta + \hat{\phi}N_\phi + \hat{r}N_{\mathrm{r}}) \\ &= -j\omega\mu\frac{e^{-jkr}}{4\pi r}(\hat{\theta}N_\theta + \hat{\phi}N_\phi) \qquad (2.70)\end{aligned}$$

This expression shows that the electric field is similar to the vector current moment, but with the radial component removed.

When working with far field quantities, the spherical coordinate system is used almost exclusively. This means that the observation point $\overline{r}$ is represented by the coordinates (r, θ, ϕ). For the source point $\overline{r}'$, on the other hand, the coordinate system should match the geometry of the source. If we use spherical coordinates for $\overline{r}$ and rectangular coordinates for $\overline{r}'$, then the dot product in the phase term of the vector current moment integral is

$$\begin{aligned}\hat{r}\cdot\overline{r}' &= (\hat{x}\sin\theta\cos\phi + \hat{y}\sin\theta\sin\phi + \hat{z}\cos\theta)\cdot(x'\hat{x} + y'\hat{x} + z'\hat{z}) \\ &= x'\sin\theta\cos\phi + y'\sin\theta\sin\phi + z'\cos\theta \qquad (2.71)\end{aligned}$$

This form of of the dot product is useful for many antenna analysis problems. Using this in (2.68) leads to the expression

$$\overline{N}(\hat{r}) = \int e^{jk_x x'}e^{jk_x y'}e^{jk_z z'}\overline{J}(x', y', z')\,dx'dy'dz' \qquad (2.72)$$

This shows explicitly that the vector current moment is a 3D Fourier transform of the current source distribution. The variables x', y', and z' are integrated, leaving a result that is a function of $k_x = k\sin\theta\cos\phi$, $k_y = k\sin\theta\sin\phi$, and $k_z = k\cos\theta$. The only difference between this expression and a fully general 3D Fourier transform is that the transformed variables k_x, k_y, and k_z are arbitrary in the general case, whereas in the vector current moment, they lie on a sphere of radius k. This is closely related to the dispersion relation for plane waves propagating in free space, which is considered in Section 4.1.

2.1.9 Magnetic Current

A basic tool for solving electromagnetic problems is to transform a complex structure and source distribution to a simpler source, often in free space, through the use of equivalent currents. This approach is based on the equivalence theorem of electromagnetic theory. Often in using the equivalence theorem it is convenient to use equivalent magnetic currents in addition to electric currents. There are apparently no isolated magnetic charges in nature, but fictitious magnetic currents can still be introduced mathematically by adding a source term $\overline{M}$ into Faraday's law:

$$\nabla\times\overline{E} = -j\omega\overline{B} - \overline{M} \qquad (2.73)$$

The magnetic current density carries units Wb/m^2/s, where Wb represents the weber, or the standard unit of electric charge. Wb/m^2/s is equivalent to V/m^2, as can be seen by balancing the units with the left-hand side of Faraday's law, recalling that a spatial derivative introduces a unit factor of inverse length.

By following an approach similar to that of Section 2.1.6, we can develop a radiation integral that finds the electric field produced by the magnetic volume current distribution $\overline{M}$. If there are no unbalanced electric charges in a given region of interest, then $\nabla \cdot \overline{D} = 0$, and therefore we can represent $\overline{D}$ as a curl according to

$$\overline{D} = -\nabla \times \overline{F} \tag{2.74}$$

By analogy with (2.48), a Helmholtz type PDE for $\overline{F}$ can be obtained. Since we have already found the Green's function for the Helmhotlz equation, we know that the solution can be found by convolving the Green's function with the magnetic current distribution according to

$$\overline{E}(\overline{r}) = -\nabla \times \int g(\overline{r}, \overline{r}')\overline{M}(\overline{r}')\, d\overline{r}' \tag{2.75}$$

In the far field limit, this expression simplifies to

$$\overline{E}(\overline{r}) \simeq jk\frac{e^{-jkr}}{4\pi r}\hat{r} \times \overline{L}(\overline{r}) \tag{2.76}$$

where the magnetic vector current moment is

$$\overline{L}(\overline{r}) = \int e^{jk\hat{r}\cdot\overline{r}'}\overline{M}(\overline{r}')\, d\overline{r}' \tag{2.77}$$

These results allow the far field contribution to be computed for magnetic current source distributions.

2.1.10 Electric and Magnetic Currents

For most types of materials, the constitutive relations together with Maxwell's equations are linear in the source distributions, which means that the fields radiated by electric and magnetic sources can be found separately and added to find the total radiated field. Combining (2.70) and (2.76) leads to a combined far field radiation integral for the electric field due to electric and magnetic currents,

$$\overline{E}(\overline{r}) \simeq jk\frac{e^{-jkr}}{4\pi r}\left[\hat{\theta}(-\eta N_\theta - L_\phi) + \hat{\phi}(-\eta N_\phi + L_\theta)\right] \tag{2.78}$$

The far magnetic field is

$$\overline{H}(\overline{r}) \simeq jk\frac{e^{-jkr}}{4\pi r}\left[\hat{\theta}(N_\phi - L_\theta/\eta) + \hat{\phi}(-N_\theta - L_\phi/\eta)\right] \tag{2.79}$$

These expressions represent the most general possible form for a spherical wave, which decays in amplitude as $1/r$ and has spherical phase fronts (i.e., the phase depends on the radial coordinate r, from which it follows that the surfaces of constant phase are spheres). It can be seen by inspecting these two expressions that the electric and magnetic far fields are orthogonal in the far field and have no longitudinal component in the $\hat{r}$ direction. The angular dependence of the radiated fields on θ and ϕ as well as the polarization of the fields are contained in the magnetic and electric vector current moments $\overline{N}$ and $\overline{L}$. These results provide important analytical tools for working with the fields radiated by antenna systems.