CHAPTER 26

Antenna Design Considerations for MIMO and Diversity Systems

MICHAEL A. JENSEN and JON W. WALLACE

26.1 INTRODUCTION

Because of the signal fading induced by multipath propagation, it can be difficult to offer reliable wireless communication in many practical environments. However, because multipath propagation is characterized by electromagnetic waves that depart at a variety of different transmit angles and arrive at different angles at the receiver, it is theoretically possible to achieve high capacity in these scenarios by exploiting these spatial propagation characteristics. For example, if the transmit and receive antennas were to provide infinite spatial selectivity (antenna beams that could excite a single plane wave), unique data streams could be transmitted on each multipath component. While practical considerations clearly make this infeasible, this simple conceptual illustration reveals the high potential capacity available in a rich scattering environment.

While infinite antenna resolution is infeasible, some degree of spatial selectivity can be obtained using multiple antenna elements at the transmit and receive ends of a link. Such multiple-input multiple-output (MIMO) wireless communications exploit the multipath channel characteristics to provide a new resource, namely, spatial processing, that allows improvement of the system performance. This new resource can be used to increase data throughput, improve signal reliability, or reduce transmitted power (leading to extended battery life in mobile devices), all without requiring an increase in the spectrum used for communication. Given these benefits, MIMO has received considerable attention in the research community and is now integrated into emerging communications standards.

This chapter focuses on antenna issues related to MIMO systems [1]. Specifically, we explore the interaction of the antenna elements with the electromagnetic propagation, a study that reveals principles that aid in the design of antennas suitable for MIMO systems. Antenna radiation characteristics such as pattern shape and polarization as well as array configuration naturally represent an important part of the discussion. We also focus on mutual coupling and array supergain, which represent significant issues when implementing MIMO technology on mobile devices where the array must be compact. Finally, the discussion turns to the synthesis of optimal antennas for MIMO communication and provides a way to compare practical designs to this optimal benchmark.

Modern Antenna Handbook, Edited by Constantine A. Balanis
Copyright © 2008 John Wiley & Sons, Inc.
26.2 MIMO SYSTEMS

It is difficult to fully appreciate the impact of the antenna properties on MIMO communication performance without at least a basic understanding of a MIMO radio architecture as well as the underlying principles that enable improved communication performance using appropriate MIMO signal processing. While this discussion focuses on a system with multiple antennas at both the transmitter and receiver, single-input multiple-output (SIMO) and multiple-input single-output (MISO) systems, also known respectively as receive and transmit diversity systems, are simply special cases of the full MIMO representation. The only fundamental difference between MIMO, SIMO, and MISO systems is in the types of signaling and detection algorithms employed.

For this discussion and throughout this chapter, boldface uppercase and lowercase letters represent matrices (matrix \( H \) with \( m \)th element \( H_{mn} \)) and column vectors (vector \( h \) with \( m \)th element \( h_m \)), respectively. When dealing with electromagnetic polarizations, we use the notation \( \mathbf{x} \) and \( \mathbf{G} \) for vectors and dyads, respectively, to clearly distinguish between polarization vectors and array signal vectors. Finally, we use the explicit functional notation \( x(t) \) and \( x(\omega) \) to indicate that a waveform is to be interpreted in the time domain with time variable \( t \) or in the frequency domain with radian frequency \( \omega \). If a waveform is given with no functional notation, it can be assumed to be represented in the frequency domain.

26.2.1 MIMO System Model

Figure 26.1 depicts a portion of a radio appropriate for MIMO communication, where the remainder of the radio is typically identical to what one would find in a traditional single-input single-output (SISO) link. To simplify the discussion, we assume that the transmitter has performed the bit-to-symbol mapping so that the input consists of a discrete sequence of complex numbers or communication symbols. Each block of \( Q \) symbols is formed into a \( Q \times 1 \) vector \( b(k) \), where \( k \) is an integer time index, which is fed into the space–time encoder to create the \( NT \times 1 \) complex output vector \( x(k) \), where \( NT \) is the number of transmit antennas. Each element of this vector is fed into a pulse shaping circuit that transforms the discrete sequence into a time-domain signal, resulting in the \( NT \times 1 \) vector function \( x(t) \). This function is upconverted to the appropriate carrier frequency to form the vector function \( x_A(t) \), which drives the transmit array to generate the vector radiated field \( \mathbf{x}_P(t, \Omega_T) \), where \( \Omega_T = (\theta_T, \phi_T) \) is a solid angle.

![Figure 26.1](image.png)
coordinate expressed in the transmit array coordinate frame with $\theta_T$ and $\phi_T$ representing the elevation and azimuth angles, respectively.

The dyadic function $\mathbf{G}_P(t, \tau, \Omega_R, \Omega_T)$, where $\Omega_R = (\theta_R, \phi_R)$ is the solid angle referenced to the receive array coordinate frame, represents the impulse response relating field radiated by the transmit array to the field incident on the receive array. The dependence on time $t$ explicitly shows that this impulse response can be time variant due to motion of the transmitter and receiver or changing scatterers in the environment. The variable $\tau$ represents the time delay relative to the excitation time $t$. We make two assumptions relative to this impulse response: (1) $\mathbf{G}_P(t, \tau, \Omega_R, \Omega_T) = 0$ for $\tau > \tau_0$ (finite impulse response), and (2) $\mathbf{G}_P(t, \tau, \Omega_R, \Omega_T)$ remains constant over a time interval $(\tau)$ of duration at least as long as the larger of $\tau_0$ and the symbol duration $T_{\text{symbol}}$. The second assumption allows the physical channel to be considered a linear, time-invariant system over each symbol transmission. As we consider only a single symbol time in the following, we drop the explicit dependence of $\mathbf{G}_P$ on $t$, understanding, however, that the channel can change from symbol to symbol.

At the receive array, the field distribution $\mathbf{y}_P(t, \Omega_R)$ is constructed by a convolution integration in delay and superposition integration in transmit angle, or

$$
\mathbf{y}_P(t, \Omega_R) = \int_{\Omega_T} \int_{-\infty}^{\infty} \mathbf{G}_P(\tau, \Omega_R, \Omega_T) \cdot \mathbf{x}_P(t - \tau, \Omega_T) \, d\tau \, d\Omega_T \quad (26.1)
$$

where $\cdot$ represents a vector dot product. The $N_R$-element receive array spatially samples this field to generate the $N_R \times 1$ signal vector $\mathbf{y}_A(t)$ at the array terminals. Noise in the system is typically generated in the physical propagation channel (interference) and the receiver front-end electronics (thermal noise). To simplify the discussion, we lump all additive noise into a single contribution represented by the $N_R \times 1$ vector $\eta(t)$, which is injected at the receive antenna terminals. The resulting signal plus noise vector $\mathbf{y}_A(t)$ is downconverted to produce the $N_R \times 1$ baseband output vector $\mathbf{y}(t)$, which is passed through a matched filter and then sampled once per symbol to create $\mathbf{y}(k)$.

The space–time decoder produces the $Q \times 1$ vector $\hat{\mathbf{b}}(k)$ representing estimates of the originally transmitted symbols, which can then be converted to a serial symbol stream.

### 26.2.2 Antennas

The characteristics of the propagation channel fundamentally control the potential MIMO performance for a given environment. However, the transmit and receive antenna arrays must ultimately enable the system to properly interact with the propagation channel in order to achieve this performance. Since it is traditional to represent antenna radiation and impedance properties in the frequency domain, we take the Fourier transform of the relevant signals (note that the transform of $\mathbf{G}_P$ is performed with respect to the delay variable $\tau$). If $\mathbf{F}_{T,n}(\omega, \Omega_T)$ represents the vector radiation pattern of the $n$th transmit antenna, then the field radiated by the transmit array can be expressed as

$$
\mathbf{X}_P(\omega, \Omega_T) = \sum_{n=1}^{N_T} \mathbf{F}_{T,n}(\omega, \Omega_T) x_{A,n}(\omega) \quad (26.2)
$$

where $x_{A,n}(\omega)$ represents the antenna driving point excitation.
The choice of which type of driving point excitation (voltage, current, incident voltage wave) should be used in this analysis can vary depending on the situation. However, it must be recognized that the choice of excitation will impact the radiation pattern $\mathbf{\Gamma}_{T,n}(\omega, \Omega_T)$ used. For example, if $x_{A,n}(\omega)$ represents the antenna driving point voltage, current, or incident voltage wave, then Eq. (26.2) implies that $\mathbf{\Gamma}_{T,n}(\omega, \Omega_T)$ must be the radiation pattern for the $n$th element with all other elements terminated, respectively, in a short-circuit, open circuit, or the system characteristic impedance. For many antenna types, the radiation pattern of the driven element with all others terminated in an open circuit can be closely approximated by the radiation pattern of the element in isolation, since the open-circuit termination does not allow terminal currents on adjacent elements to flow. More discussion on this particular point appears in Section 26.4.3. For now we assume that these open-circuit patterns are used in the analysis, so that $x_{A,n}(\omega)$ is the antenna terminal current. It should also be mentioned that while here we are using generic input and output variable designations $x$ and $y$, we will alter this notation to explicitly represent voltages, currents, voltage waves, or electric fields as we transition into more detailed analysis.

Each element in the receive array samples the incident field $\mathbf{\Upsilon}_P(\omega, \Omega_R)$, leading to the received signal vector $\mathbf{y}_A(\omega)$. If $\mathbf{\Upsilon}_{R,m}(\omega, \Omega_R)$ is the radiation pattern of the $m$th receive element with all other elements terminated in an open circuit, then the open-circuit voltage on the $m$th receive antenna is

$$y_{A,m}(\omega) = \int_{\Omega_R} \mathbf{\Upsilon}_{R,m}(\omega, \Omega_R) \cdot \mathbf{\Upsilon}_P(\omega, \Omega_R) d\Omega_R$$  \hspace{1cm} (26.3)

Combining Eqs. (26.1), (26.2), and (26.3) leads to

$$\mathbf{y}_A(\omega) = \mathbf{H}_A(\omega)\mathbf{x}_A(\omega) + \mathbf{\eta}(\omega)$$  \hspace{1cm} (26.4)

where the channel transfer matrix (or simply channel matrix) has elements

$$H_{A,mn}(\omega) = \int_{\Omega_R} \int_{\Omega_T} \mathbf{\Upsilon}_{R,m}(\omega, \Omega_R) \cdot \mathbf{G}_P(\omega, \Omega_R, \Omega_T) \cdot \mathbf{\Upsilon}_{T,n}(\omega, \Omega_T) d\Omega_R d\Omega_T$$  \hspace{1cm} (26.5)

Equation (26.4) provides a simple transfer relationship for the MIMO system with Eq. (26.5) explicitly revealing that the physical channel includes the impact of both the propagation environment and the antennas.

It is important to recognize that since the radiation patterns are computed with other elements present but open-circuited, this analysis includes mutual coupling in the sense that it incorporates all appropriate electromagnetic boundary conditions for the antenna pattern determination. However, at this point Eq. (26.4) simply relates transmit currents to received open-circuit voltages. In Section 26.4, we formulate the channel to include at least a portion of the radiofrequency (RF) electronics, providing the opportunity to introduce the impact of the full antenna impedance matrix into the analysis.

While the waveform-based transfer relationship in Eq. (26.4) is useful for some analyses, signal-processing analyses often go one step further by considering the channel to relate the input $\mathbf{x}^{(k)}$ to the pulse shaping block and the sampled output $\mathbf{y}^{(k)}$ of the matched filter. Assuming that the frequency-domain channel transfer matrix remains approximately constant over the signal bandwidth, a scenario often described as frequency nonselective or flat fading, the frequency-domain transfer functions are treated as complex constants
that scale the complex input symbols. The discrete-time transfer relationship assumes the form

$$y^{(k)} = H^{(k)}x^{(k)} + \eta^{(k)}$$  \hspace{1cm} (26.6)$$

where $\eta^{(k)}$ contains the noise that has passed through the receiver and has been sampled at the matched-filter output. Despite the fact that for notational simplicity we sometimes drop the superscript $k$ representing the symbol index, it is important to emphasize that the channel matrix $H^{(k)}$ can change over time. It is also important to remember that $H^{(k)}$ is based on the value of $H_A(\omega)$ evaluated at the carrier frequency, but includes the additional effects of the transmit and receive electronics. For much of this chapter, we assume a narrowband system with ideal electronics, and therefore Eq. (26.6) is a relevant description.

### 26.2.3 Representative Propagation Channel

Because the MIMO physical channel depends critically on both the antennas and the propagation environment, it is important to discuss briefly how the propagation environment can be described mathematically. It is very common in signal-processing analyses to treat the physical channel transfer matrix $H_A$ (or discrete-time channel matrix $H$) as a random matrix whose elements are random variables, typically drawn from a Gaussian distribution. While this is convenient for analysis, incorporation of the antenna radiation properties requires a more descriptive physical channel representation to allow explicit formulation of the interaction between the antennas and the electromagnetic fields responsible for the propagation.

In such cases, it is common to assume that all scattering in the propagation channel is in the far field of the arrays and that a finite number of discrete propagation “paths” (plane waves) connects the transmit and receive arrays. (It is important to recognize that the far-field scattering and plane wave assumptions are not explicitly true for many environments. However, the operation of a MIMO system does not depend critically on this assumption being true. In fact, the rapid variation of near fields can enhance MIMO performance.) If $L$ denotes the number of propagating plane waves that link the transmitter to the receiver, the physical channel response can be expressed as

$$G_P(\tau, \Omega_R, \Omega_T) = \sum_{\ell=0}^{L-1} \beta_\ell \delta(\tau - \tau_\ell) \delta(\Omega_T - \Omega_{T,\ell}) \delta(\Omega_R - \Omega_{R,\ell})$$  \hspace{1cm} (26.7)$$

where $\delta(\cdot)$ represents the Dirac delta function and $\beta_\ell$ is the dyadic complex gain of the $\ell$th path with angle of departure (AoD) $\Omega_{T,\ell}$, angle of arrival (AoA) $\Omega_{R,\ell}$, and time delay of arrival (TDoA) $\tau_\ell$. Note that the dyadic gain $\beta_\ell$ incorporates polarization changes of the field due to scattering as the wave propagates from the transmitter to the receiver. Any channel time variation is included by making the multipath parameters $(L, \beta_\ell, \tau_\ell, \Omega_{T,\ell}, \Omega_{R,\ell})$ time dependent. With this channel model, the channel matrix has elements

$$H_{A, mn}(\omega) = \sum_{\ell=0}^{L-1} e^{-j\omega \tau_\ell} \bar{e}_{R,m}(\omega, \Omega_{R,\ell}) \cdot \beta_\ell \cdot \bar{e}_{T,n}(\omega, \Omega_{T,\ell})$$  \hspace{1cm} (26.8)$$

In many situations, it is useful to describe the propagation environment stochastically so that it is possible to assess the statistical behavior of a given MIMO system over an
ensemble of representative propagation environments. The mechanism for accomplishing this is to treat each multipath parameter as a random variable drawn from a specified distribution. Most current models are based on initial work in outdoor environments by Turin et al. [2], which demonstrated that multipath components are generally grouped into clusters that decay exponentially with increasing delay. Later work extended the model to indoor scenarios [3] and added directional information [4–7]. Provided that the underlying statistical distributions are properly specified, these models can offer highly accurate channel representations (in a statistical sense). In this section we detail one implementation of such a model which extends the well-known Saleh–Valenzuela model [3] to include AoA/AoD in addition to ToA and multipath amplitude. This model is referred to as the Saleh–Valenzuela model with angle or simply the SVA model. Naturally, there are numerous other mechanisms for describing the channel characteristics, and therefore this should be considered an introduction to the considerations relevant for incorporating the effect of the antenna in the MIMO system analysis rather than a comprehensive discussion on the topic. More details on this subject can be found in the scientific literature [1, 8].

The SVA model is based on the experimentally observed phenomenon that multipath arrivals appear at the receiver in clusters in both space and time. We refer to arrivals within a cluster as rays and restrict our discussion to the horizontal plane for simplicity ($\theta_T = \theta_R = \pi/2$). The model also assumes a single electromagnetic polarization, although it has been extended to include vector fields [4].

Assuming we have $N_c$ clusters with $N_r$ rays per cluster, then the scalar directional channel impulse response of Eq. (26.7) can be written

$$G_P(\tau, \phi_R, \phi_T) = \frac{1}{\sqrt{N_cN_r}} \sum_{\ell=0}^{N_c-1} \sum_{k=0}^{N_r-1} \beta_{\ell k} \delta(\tau - T_{\ell} - \tau_{k\ell}) \times \delta(\phi_T - \Phi_{T,\ell} - \phi_{T,k\ell}) \delta(\phi_R - \Phi_{R,\ell} - \phi_{R,k\ell})$$

where the summation explicitly reveals the concept of clusters (index $\ell$) and rays within the cluster (index $k$). The parameters $T_{\ell}$, $\Phi_{T,\ell}$, and $\Phi_{R,\ell}$ represent the initial arrival delay, mean departure angle, and mean arrival angle, respectively, of the $\ell$th cluster. The $k$th ray arrival delay $\tau_{k\ell}$, departure angle $\phi_{T,k\ell}$, and arrival angle $\phi_{R,k\ell}$ are taken relative to the corresponding cluster values.

The SVA model specifies the statistical distribution of the various multipath parameters. The delay of each cluster conditioned on the delay of the previous cluster satisfies the probability density function (PDF)

$$p(T_\ell | T_{\ell-1}) = \Lambda_T e^{-\Lambda_T(T_\ell - T_{\ell-1})}$$

where $\Lambda_T$ controls the cluster arrival rate for the environment and typically $T_0 = 0$. Similarly, the arrival time for the $k$th ray in the $\ell$th cluster obeys the conditional PDF

$$p(\tau_{k\ell} | \tau_{k-1,\ell}) = \lambda_\tau e^{-\lambda_\tau(\tau_{k\ell} - \tau_{k-1,\ell})}$$

where $\lambda_\tau$ controls the ray arrival rate.
The complex gain $\beta_{k\ell}$ has a magnitude that obeys the Rayleigh distribution with the expected power (or variance) satisfying
\[ E\{|\beta_{k\ell}|^2\} = E\{|\beta_{00}|^2\} e^{-\tau_T/\Gamma_T} e^{-\tau_r/\gamma_r} \] (26.12)
which makes the amplitudes of the clusters as well as the amplitudes of the rays within the clusters decay exponentially with the time constants $\Gamma_T$ and $\gamma_r$, respectively. The phase of the complex gain is assumed uniformly distributed on $[0, 2\pi]$.

Finally, the angles of departure and arrival must be specified. For indoor and dense urban areas where scattering tends to come from all directions, the cluster departure and arrival angles can be modeled as uniformly distributed random variables on $[0, 2\pi]$. Based on measured data taken in Ref. 5, a two-sided Laplacian distribution is assumed for the ray AoA/AoD distribution with PDF given by
\[ p(\phi_P) = \frac{1}{\sqrt{2} \sigma_{P,\phi}} \exp\left(-\sqrt{2}\frac{\phi_P}{\sigma_{P,\phi}}\right) \] (26.13)
where $P \in \{T, R\}$ and $\sigma_{P,\phi}$ is the standard deviation of angle in radians.

When this SVA model is used in simulations for this chapter, statistical parameters appropriate for an indoor environment, as measured in Ref. 5, are used. These parameters are given as follows:

<table>
<thead>
<tr>
<th>Parameter Value</th>
<th>Parameter Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_{T,\phi}$</td>
<td>$26^\circ$</td>
</tr>
<tr>
<td>$\sigma_{R,\phi}$</td>
<td>$26^\circ$</td>
</tr>
<tr>
<td>$1/\Gamma_T$</td>
<td>$34$ ns</td>
</tr>
<tr>
<td>$1/\gamma_r$</td>
<td>$29$ ns</td>
</tr>
<tr>
<td>$1/\lambda_T$</td>
<td>$17$ ns</td>
</tr>
<tr>
<td>$1/\lambda_r$</td>
<td>$5$ ns</td>
</tr>
</tbody>
</table>

### 26.2.4 Channel Capacity

In SISO systems, antenna designers typically focus only on the system gain (or received power) enabled by the antenna. However, in a MIMO system, the multivariate nature of the problem complicates the description, and therefore the community has resorted to the capacity as a metric for defining channel quality [9, 10]. This quantity represents the highest error-free MIMO transmission rate for a given transfer matrix and signal-to-noise ratio (SNR) under optimal space–time coding and modulation. Because capacity has become the standard evaluative metric for MIMO systems, it is constructive to discuss this parameter here.

We apply the following analysis to the discrete-time channel in Eq. (26.6) (we drop the time index $k$). In this case, the capacity represents the maximum number of bits conveyed per time slot (bits/use). If the capacity for the physical channel in Eq. (26.4) is computed, it gives the transmission rate per hertz of bandwidth (bits/s/Hz). Under ideal processing, the two capacities are the same provided that proper signal and noise power definitions are used.

Optimal communication performance is enabled when the transmitter and receiver both know the channel transfer matrix. Let the singular value distribution (SVD) of the channel matrix be denoted by $H = USV^\dagger$, where $[.]^\dagger$ denotes a conjugate transpose, the matrices $U$ and $V$ of singular vectors are unitary, and the diagonal matrix $S$ contains the real, nonnegative singular values ordered from largest to smallest. If $V_Q$ consists of
the first $Q$ columns of $V$, we can encode the input data (space–time encoding) using $x = V_Q Pb$, where $P$ is a diagonal $Q \times Q$ matrix of real, positive values, which specifies the power allocated to each symbol. The receiver performs the operation (space–time decoding) $\hat{b} = P^{-1}S_Q^{-1}U_Q^\dagger y$, where $S_Q$ is the upper left $Q \times Q$ block of $S$. Since $U$ and $V$ are unitary, the resulting output is

$$\hat{b} = P^{-1}S_Q^{-1}U_Q^\dagger [Hx + \eta]$$

$$= P^{-1}S_Q^{-1}U_Q^\dagger [(USV^\dagger) V_Q Pb + \eta]$$

$$= b + P^{-1}S_Q^{-1}U_Q^\dagger \eta$$

(26.14)

Our signaling strategy has resulted in a received signal vector that is a noisy replica of the transmitted signal vector. Figure 26.2 shows a block diagram of this signaling approach.

This signaling strategy is important as it provides physical intuition regarding MIMO communication. Since each (power-weighted) element of $b$ multiplies the corresponding column of $V_Q$, this operation suggests that each column of $V_Q$ represents the array weights for the input symbol. Since the operation $U_Q^\dagger$ creates a vector whose elements are scaled copies of the original transmitted vector, each row of $U_Q^\dagger$ (column of $U_Q^*$) represents the receive array weights for the corresponding symbol. We can think of the array patterns associated with each symbol as eigenpatterns that create independent (spatially orthogonal) parallel communication channels in the multipath environment (see Ref. 11 for detailed discussion). Note that each of the created eigenchannels (or eigenmodes or modes) in general excites all multipath components.

For a specific channel, $Q$ and $P$ can be chosen to achieve the channel capacity, which mathematically is the maximum mutual information between the transmit and receive signal vectors. If we assume that $\eta$ has independent zero-mean Gaussian-distributed elements with equal variance $\sigma_\eta^2$, then the capacity achieving signal vector $x$ has elements that are zero-mean complex Gaussian-distributed random variables with covariance $R_x = E\{xx^\dagger\}$, where $E\{\cdot\}$ represents the expectation. The expression for channel capacity is

$$C = \max_{\|R_t\|_F \leq P_T} \log_2 |I + HR_xH^\dagger/\sigma_\eta^2|$$

(26.15)

where $I$ is the identity matrix, $\|\cdot\|$ is the determinant, $\operatorname{Tr}[\cdot]$ is the trace, and $P_T$ is a measure of the total power. Note that $\operatorname{Tr}[R_t]$ constrains the sum of the squares of the antenna excitation signals, which in our case represents currents. We later discuss how this does not constrain the power radiated by the transmit array. However, this constraint is commonly used in practice, and therefore we use it for this brief capacity discussion. The off-diagonal elements of $R_t$ represent the correlation between the transmitted signal streams, with increased correlation generally resulting in decreased capacity.

![Figure 26.2](image-url)  
Figure 26.2  Signaling strategy for optimally communicating over the channel eigenmodes.
Our problem therefore becomes that of determining the covariance $R_x$ that maximizes the right-hand side of Eq. (26.15) subject to the constraint. First, it can be shown that $HR_xH^\dagger$ must be diagonal to maximize this expression \cite{12}. Using the SVD of $H$ and defining $A_x = V^\dagger R_x V$, we can write

$$|I + USA_xS^T U^\dagger| = |U(I + SA_xS^T)U^\dagger| = |I + SA_xS^T|$$  \hspace{1cm} (26.16)

where $\{\cdot\}^T$ is a transpose and we have used that $U$ is unitary (recall that $S$ is diagonal and real). Note that, given our discussion above, $A_x$ represents the covariance of the power-scaled vector $b$, and since $A_x$ is diagonal (to maximize the capacity expression) the input streams should be uncorrelated. Our capacity expression becomes

$$C = \max_{\{\Lambda_x, \sum_{mm} \Lambda_x,mm \leq PT\}} \sum_{m=1}^{N_R} \log_2 \left( 1 + \frac{S^2_{mm} \Lambda_x,mm}{\sigma_q^2} \right)$$  \hspace{1cm} (26.17)

where we have used that $\text{Tr} \{R_x\} = \text{Tr} \{A_x\}$ since $V$ is unitary. The terms $\Lambda_x,mm$ and $S^2_{mm}$ represent, respectively, the power to be allocated to the $m$th symbol (and therefore $m$th channel eigenmode) and the power gain of the $m$th eigenmode. Our problem is now reduced to finding the real, nonnegative values of the diagonal matrix $A_x$. This can be accomplished using Lagrange multipliers to obtain the water-filling solution \cite{9, 12–14}, which allocates power to the high gain modes and may not use weaker channels, so that $Q$ becomes the number of nonzero values of $\Lambda_x,mm$.

While this strategy is optimal, it requires that the transmitter be aware of the channel matrix $H$, something that may not be practical in systems where the channel state changes rapidly due to node motion or changes in scatterers. If the transmitter has no knowledge of the channel state information (CSI), then each equally weighted data stream can be transmitted from one of the antennas, yielding a transmit covariance of $R_x = (PT/NT)I$.

Substitution into Eq. (26.15) results in the uninformed transmit capacity \cite{15}

$$C_{UT} = \log_2 \left| I + \frac{PT}{NT \sigma_q^2} HH^\dagger \right|$$  \hspace{1cm} (26.18)

For full-rank channel matrices at high SNR, the penalty paid for an uninformed transmitter is relatively small \cite{14}. Note that if $N_R \geq N_T$, then the maximum likelihood estimate of the vector $x$ can be obtained (under the assumption of spatially white Gaussian noise) using

$$\hat{x} = H^+ y$$  \hspace{1cm} (26.19)

where $\{\cdot\}^+$ represents a pseudo-inverse operation. This is the basic principle behind the well publicized VBLAST algorithm \cite{16, 17}.

Since capacity depends on receive SNR, it is important to properly normalize channel matrices for correct interpretation of results. For channel matrices $H^{(k)}$, $1 \leq k \leq K$, normalized channel matrices are computed as

$$H^{(k)}_{\text{norm}} = H^{(k)} \left[ \frac{1}{KN_T N_R} \sum_{k=1}^{K} \|H^{(k)}\|_F^2 \right]^{-1/2}$$  \hspace{1cm} (26.20)
where $\| \cdot \|_F$ is the Frobenius norm. When normalized matrices are used in the capacity expressions, $\rho = P_T/\sigma_n^2$ represents the average SNR of a single antenna system and is referred to as the SISO SNR. The differences in path loss among a number of channel matrices may be removed by normalizing each matrix independently, or $K = 1$. Using $K > 1$ preserves the relative power levels among the $K$ channels.

26.2.5 Diversity and Spatial Multiplexing

One of the most important conclusions drawn from the analysis of this section is that MIMO systems exploit the channel spatial degrees of freedom to increase communication performance. The two signaling strategies discussed in Section 26.2.4 are focused on using the degrees of freedom to simultaneously communicate multiple independent data streams over the channel. This technique, referred to as spatial multiplexing, is only possible in a true MIMO system with multiple antennas at both ends of the link. Note that while this seems the most efficient use of the channel, accurate detection of the received streams is sensitive to noise and the accuracy of the CSI at the receiver (and possibly the transmitter). Therefore it is typically necessary to use a lower rate error control code to obtain reliable transmission.

An alternate approach is to transmit a single data symbol each symbol period, but encode it so that the spatial degrees of freedom provide redundancy. At the transmitter, this is equivalent to spatial error control coding, although we typically refer to this as transmit diversity. With multiple antennas at the receiver, further redundancy is assured using receive diversity. When the system relies on diversity rather than multiplexing, a higher-rate temporal error control code can typically be employed, and it is technically possible to achieve the same throughput using the two strategies. More generally, the temporal and spatial codes can be combined into a single matrix so that a single symbol is sent over multiple antennas and multiple symbol periods, leading to the concept of space–time codes. In this case, the discussion of the operations depicted in Figure 26.1 must be modified to allow transmission of multiple vectors in time, all based on a single input symbol.

The advantage of a diversity system is that its performance is often more robust to situations where one or more of the eigenmodes becomes unusable due to changes in the scattering environment. This is particularly important if the transmitter is unaware of the channel or unable to adapt the number of transmitted streams based on the channel quality. Furthermore, MISO and SIMO signaling are transmit and receive diversity, respectively, indicating that diversity signaling is relevant for systems with only a single antenna at one end of the link. It is also possible to combine multiplexing and diversity [18], which is particularly useful when $N_T \neq N_R$ so that the extra antennas at one end of the link can be used to advantage in a diversity fashion.

Despite the fundamental differences between these two signaling approaches, the goal in each is to use the antennas to exploit the spatial degrees of freedom inherent in the multipath propagation to increase communication performance. Therefore the physical antenna radiation characteristics impact the performance of SIMO, MISO, and MIMO systems in essentially the same manner. In fact, despite the meaning of the word diversity as a communication concept, in the context of antennas the word diversity simply means that each element of the array possesses unique radiation characteristics that allow it to excite (or sample) the multipath field in a unique way. Therefore, as this discussion transitions to focus additional attention on the role of antennas in determining system
performance, we often remove the complexity associated with considering both transmit and receive antenna arrays and focus only on the receive diversity performance of the array. Naturally, at other times, we show how the analysis applies to MIMO systems as well.

26.3 ANTENNA DIVERSITY

Armed now with a basic understanding of the principles governing MIMO operation, we can focus additional detail on the role of the antennas in determining system performance. Consider a simple example of two receive antennas with vector field patterns \( \mathbf{e}_1(\theta, \phi) \) and \( \mathbf{e}_2(\theta, \phi) \) and placed at the coordinates \((-d/2, 0, 0)\) and \((d/2, 0, 0)\) in Cartesian space. A set of \( L \) plane waves, with the \( \ell \)th plane wave characterized by complex strength \( E_\ell \), arrival angles \((\theta_\ell, \phi_\ell)\), and electric field polarization \( \hat{e}_\ell \), impinges on the antenna array. If each pattern is obtained with the other antenna terminated in an open circuit, then the open-circuit signal voltages on each antenna are given by

\[
\hat{v}_1 = \sum_{\ell=0}^{L-1} E_\ell \left[ \mathbf{e}_1(\theta_\ell, \phi_\ell) \cdot \hat{e}_\ell \right] e^{-j(\pi d/\lambda) \sin \theta_\ell \cos \phi_\ell}
\]

\[
\hat{v}_2 = \sum_{\ell=0}^{L-1} E_\ell \left[ \mathbf{e}_2(\theta_\ell, \phi_\ell) \cdot \hat{e}_\ell \right] e^{j(\pi d/\lambda) \sin \theta_\ell \cos \phi_\ell}
\]

(26.21)

where \( \lambda \) is the free-space wavelength. For the MIMO system to work effectively, the signals \( \hat{v}_1 \) and \( \hat{v}_2 \) must be unique, despite the fact that both antennas observe the same set of plane waves, which can be accomplished when each antenna provides a unique weighting to each of the plane waves. Equation (26.21) reveals that this can occur in three different ways:

1. Based on the different antenna element positions, each antenna places a unique phase on each multipath component based on its arrival angles. This is traditional spatial diversity.
2. If the radiation patterns \( \mathbf{e}_1(\theta, \phi) \) and \( \mathbf{e}_2(\theta, \phi) \) are different, then each multipath is weighted differently by the two antennas. When the antennas share the same polarization but have different magnitude and phase responses in different directions, this is traditional angle diversity.
3. If the two antennas have different polarizations, the dot product leads to a unique weighting of each multipath component. This is traditional polarization diversity.

It is noteworthy that both angle and polarization diversity are subsets of the more inclusive pattern diversity, which simply implies that the two antenna radiation patterns (magnitude, phase, and polarization) differ to create the unique multipath weighting.

26.3.1 Antenna Diversity Performance

While saying that an array design provides good diversity is qualitatively helpful, it does nothing to quantify the array performance. It should be clear based on this discussion that
the performance depends on the propagation environment as well as the antenna radiation properties. Because it is convenient to be able to discuss the performance of an array at one end of the link independent of the design at the other, we can consider a framework where a complex vector electromagnetic wave $\mathbf{e}(\Omega R)$ impinges on a receiving array. We assume that this field is represented as a zero-mean Gaussian stochastic process, which means that the complex field envelope varies slowly in time (slow relative to the sinusoidal variation of the carrier frequency). We also assume that the field arriving at one angle is uncorrelated with that arriving at another angle, or

$$E\{\mathbf{e}(\Omega R)\mathbf{e}^*(\Omega' R)\} = E\{\mathbf{e}(\Omega R)\mathbf{e}^*(\Omega' R)\} \delta(\Omega R - \Omega' R)$$

(26.22)

where the vector–vector product is a dyadic multiplication (essentially a vector outer product) and $\mathbf{P}(\Omega R)$ is the power angular spectrum, or the average power per unit angle, for the environment.

Again letting $\mathbf{e}_{R,m}(\Omega R)$ represent the field radiation pattern of the $m$th receive antenna with all other elements terminated in an open circuit, the random variable representing the open-circuit signal voltage on the $m$th receive antenna in this environment is

$$\hat{v}_{s,m} = \varphi \int_{\Omega R} \mathbf{e}_{R,m}(\Omega R) \cdot \mathbf{e}(\Omega R) \, d\Omega R$$

(26.23)

where the integration is over a full $4\pi$ steradians and $\varphi$ is a complex constant. The covariance of the random signal vector $\hat{v}_s$ has elements

$$R_{s,mp} = E\{\hat{v}_{s,m} \hat{v}_{s,p}^*\}$$

(26.24)

$$= |\varphi|^2 \int_{\Omega R} \int_{\Omega R} \mathbf{e}_{R,m}(\Omega R) \cdot E\{\mathbf{e}(\Omega R)\mathbf{e}^*(\Omega' R)\} \cdot \mathbf{e}_{R,p}(\Omega' R) \, d\Omega R \, d\Omega' R$$

(26.25)

$$= |\varphi|^2 \int_{\Omega R} \mathbf{e}_{R,m}(\Omega R) \cdot \mathbf{P}(\Omega R) \cdot \mathbf{e}_{R,p}(\Omega R) \, d\Omega R$$

(26.26)

where we have used the result in Eq. (26.22).

Any analysis of multiantenna system performance must incorporate the receiver noise. If the noise is generated by an interfering incident field $\mathbf{e}(\Omega R)$, then the open-circuit noise voltage on the $m$th antenna is given by

$$\hat{v}_{i,m} = \varphi \int_{\Omega R} \mathbf{e}_{R,m}(\Omega R) \cdot \mathbf{e}(\Omega R) \, d\Omega R$$

(26.27)

We assume that the interference is a zero-mean Gaussian random process and that the interference field arriving at one angle is uncorrelated with that arriving at another angle. The covariance of the open-circuit noise voltage is

$$R_{i,mp} = |\varphi|^2 \int_{\Omega R} \mathbf{e}_{R,m}(\Omega R) \cdot \mathbf{P}(\Omega R) \cdot \mathbf{e}_{R,p}(\Omega R) \, d\Omega R$$

(26.28)

where $\mathbf{P}(\Omega R)$ is the interference power angular spectrum and the matrix $R_i$ is in general full.
An alternate source of noise is thermal noise generated by the receiver front-end amplifiers. In this case, accurate injection of the noise into the model is somewhat complicated and depends on the receiver architecture and nature of the noise source. For simplicity, it is often assumed that the noise voltage referred to the open-circuit antenna terminals is a zero-mean Gaussian process with covariance

$$ R_t = \sigma_t^2 I $$

where \( \sigma_t^2 \) is the thermal noise variance. A more detailed method of incorporating realistic thermal noise models is discussed in Section 26.4.2.

We generally assume that the noise sources are independent of each other and are both independent of the signal, such that the overall open-circuit voltage and its covariance are

$$ \hat{v} = \hat{v}_s + \hat{v}_i + \hat{v}_t $$

$$ E\{ \hat{v} \hat{v}^\dagger \} = R_s + R_i + R_t $$

where the subscript \( \eta \) indicates that the quantity represents the total of all noise sources.

Since \( R_s \) is in general a full matrix, the open-circuit signal voltages on the receive antenna terminals are correlated, a fact that complicates the analysis of the system performance. To facilitate the analysis, we run the received open-circuit voltage through two beam formers. First, we route the signal through a spatial noise prewhitening filter expressed as

$$ y' = R_{\eta}^{-1/2} \hat{v} $$

which has covariance

$$ R'_{y} = R_{\eta}^{-1/2} R_s R_{\eta}^{-1/2} + I = R'_s + I $$

Next, because the covariance \( R'_s \) is positive semidefinite and Hermitian, it can be expressed in terms of its eigenvalue decomposition (EVD) \( R'_s = \xi_s \Lambda_s \xi_s^\dagger \), where \( \xi_s \) is a unitary matrix of eigenvectors and \( \Lambda_s \) is a diagonal matrix with real, nonnegative entries. If we pass the signal \( y' \) through a second beam former to obtain \( y = \xi_s^\dagger y' \), the signal \( y \) has covariance

$$ R_y = \Lambda_s + I $$

Figure 26.3 illustrates the beam-forming process described.

The convenience of this beam-forming process is that we have transformed our correlated signals to a set of uncorrelated ones, where the average SNR of the \( m \)th uncorrelated signal is given by the real value \( \Lambda_{s,m} \). The number and magnitudes of the nonzero eigenvalues are an indicator of the system performance. More precisely, we can assume that

![Figure 26.3 Beam-forming approach for uncorrelating the noisy output of the receive antenna array.](image)
we will combine the signals on the independent branches in an appropriate manner. There are several methods for performing this combining, which in order of increasing performance and implementation complexity include the following:

1. **Switched Combining:** Only a single branch is selected for connection to the receiver electronics at any given time. The currently selected branch is used until its SNR drops below a predefined threshold, at which time the next branch is connected to the electronics.

2. **Selection Combining:** All branches are constantly monitored, and the one with the highest SNR is selected for connection to the receiver electronics.

3. **Equal Gain Combining:** The phase of the carrier on each branch is shifted such that the signals on all branches share the same phase. The signals on all branches are then added together.

4. **Maximal Ratio Combining:** As in equal gain combining, the phase of the carrier on each branch is shifted such that the signals on all branches share the same phase. When the branch signals are added together, however, those with the highest SNR receive more weight, with the exact weight chosen according to an optimal algorithm. This approach is optimal in terms of maximizing the SNR at the output of the combining circuitry.

A more comprehensive examination of these diversity combining strategies can be found in Ref. 19. For the purposes of this chapter, we assume that the receiver uses the optimal maximal ratio combining strategy. If all of the branches have equal average SNR or $\Lambda_s = \Lambda_0 I$ and if the SNR on each branch satisfies a Rayleigh distribution, then the instantaneous SNR $\gamma$ at the output of the maximal ratio combiner satisfies the cumulative distribution function (CDF) [20]

$$
P_{MR}(\gamma \leq x) = 1 - e^{-x/\Lambda_0} \sum_{m=1}^{N_T} \frac{(x/\Lambda_0)^{m-1}}{(m-1)!} \tag{26.34}
$$

In the case of distinct values of average SNR (all elements of $\Lambda_s$ are unique), the instantaneous SNR at the combiner output satisfies the CDF [20]

$$
P_{MR}(\gamma \leq x) = \sum_{m=1}^{N_T} \epsilon_m \left( 1 - e^{-x/\Lambda_{s,mm}} \right) \tag{26.35}
$$

$$
\epsilon_m = \prod_{k=1}^{N_T} \frac{1}{1 - \frac{\Lambda_{s,\lambda k}}{\Lambda_{s,mm}}} \tag{26.36}
$$

Figure 26.4 illustrates the resulting CDF of the instantaneous SNR normalized by the average SNR ($\gamma/\Lambda_0$) at the combiner output for maximal ratio combining with up to four different branches with equal average SNR. The curves are plotted on a Rayleigh scale such that when the CDF represents a Rayleigh distribution (one branch), the curve is a straight line. The curve for a single diversity branch indicates that the normalized SNR will be below $-20$ dB 1% of the time, while for two branches the normalized SNR will
26.3 ANTENNA DIVERSITY

Figure 26.4  Cumulative distribution function of the instantaneous SNR (normalized by the average SNR) for maximal ratio diversity combining with different numbers of branches.

be below −9 dB 1% of the time. Effectively, the use of two antennas and maximal ratio combining represents a gain of 11 dB at the 1% probability level. This gain is referred to as the diversity gain, designated as “DG” in the figure. It is noteworthy that the value of the diversity gain depends on the probability level assumed.

While the diversity gain represents a useful metric for quantifying the performance of a diversity system, it is also possible to characterize performance using what is arguably a more intuitive quantity. Consider Figure 26.5, which shows the CDF curves for maximal ratio combining (two and three branches) when all branches have unity average SNR. This figure also shows the CDF for a system with three branches with average SNR values of 1, 0.7, and 0.5. Again considering the 1% probability level, the number of effective diversity branches represents the fractional number of uncorrelated antennas with equal average SNR required to achieve the same output SNR as that obtained with the three-branch system with unequal average SNR values. For example, this number of effective branches would be roughly 2.7 for the example case in Figure 26.5. We use this metric as a means for comparing the performance of different antenna topologies.

26.3.2 Pattern (Angle and Polarization) Diversity

As a first study of the diversity potential for different array topologies, it is interesting to explore the fundamental limits of achievable performance with a multiport antenna located at a single point in space. Clearly, this means that all diversity must be pattern diversity, which as previously mentioned includes both angle and polarization diversity. Polarization diversity can be particularly intriguing, since it can potentially lead to low correlation on at least two branches even when the channel is characterized by little or no multipath scattering. However, proper implementation of polarization diversity for MIMO systems requires understanding of the physics involved.

This analysis begins with infinitesimal electric and magnetic current elements (dipoles) radiating into free space. For a three-dimensional coordinate frame, we use three of each
current type (six total dipoles) oriented in the $\hat{x}$, $\hat{y}$, and $\hat{z}$ directions. Each current creates a unique vector far-field radiation pattern given as follows:

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{x}$</td>
<td>$e_{R,1} = -\hat{\theta}\cos\theta_R \cos\phi_R + \hat{\phi}\sin\phi_R$</td>
<td>$e_{R,4} = \hat{\phi}\sin\phi_R + \hat{\theta}\cos\theta_R \cos\phi_R$</td>
</tr>
<tr>
<td>$\hat{y}$</td>
<td>$e_{R,2} = -\hat{\theta}\cos\theta_R \sin\phi_R - \hat{\phi}\cos\phi_R$</td>
<td>$e_{R,5} = -\hat{\theta}\cos\phi_R + \hat{\phi}\cos\theta_R \sin\phi_R$</td>
</tr>
<tr>
<td>$\hat{z}$</td>
<td>$e_{R,3} = \hat{\phi}\sin\theta_R$</td>
<td>$e_{R,6} = -\hat{\phi}\sin\theta_R$</td>
</tr>
</tbody>
</table>

These results reveal that polarization diversity cannot be completely independent of angle diversity since the radiation pattern shape depends on the orientation (polarization) of the radiating current.

With the dipole patterns defined by this radiation analysis, we now consider all six antennas located at the same point in space and receiving an incident multipath field. We assume the power angular spectrum is

$$\overline{P}_s(\Omega_R) = \begin{cases} (1/D\Omega)(D\Omega), & |\theta_R - \pi/2| \leq D\theta, \ |\phi_R| \leq D\phi \\ 0, & \text{otherwise} \end{cases}$$

(26.37)

where the notation $\overline{\theta}$ simply means a dyad of all zeros and

$$D\Omega = \int_{-\Delta\phi}^{\Delta\phi} \int_{\pi/2 - \Delta\phi}^{\pi/2 + \Delta\phi} \sin^2\theta \, d\theta \, d\phi$$

(26.38)

We further assume that there is no external interference and that the thermal noise has covariance $R_t = \sigma^2_t I$. With this incident field and the radiation patterns specified, the covariance matrix elements can be computed in closed form [21]. For computation of
the number of effective diversity branches, all eigenvalues of $\Lambda_s$ are normalized by the largest eigenvalue, and the average SNR used for construction of the CDF for systems with equal branch SNR is $\Lambda_0 = 1$. This normalization makes the absolute value of $\sigma_t^2$ irrelevant.

Figure 26.6 plots the effective number of branches as a function of the angular extent of the power angular spectrum. As can be seen, for zero angle spread (which corresponds to a single plane wave), the system offers two effective branches corresponding to the two orthogonal polarizations available in the plane wave. As the angle spread increases, so does the number of diversity branches, with a maximum value of six resulting when the multipath arrives from all angles on the sphere [22]. One interpretation of this full angle spread result is that, under these conditions, the three electric and three magnetic field vector components are all mutually uncorrelated. Another equivalent interpretation is that the three orthogonal radiation patterns times two polarizations ($\hat{\theta}$ and $\hat{\phi}$) leads to the six degrees of freedom. Either way, it is important to recognize that practical environments are rarely characterized by full angle spread, with most of the energy confined to the horizontal plane. Therefore the achievement of six uncorrelated channels using a point sensor is not realized in most practical systems.

There are other practical challenges associated with this multipolarization concept. First, in most environments, transmission of the vertical polarization results in a difference of 3 to 10 dB in the received power levels for the vertical and horizontal polarizations. Mathematically, this means that, if we use multiple polarizations at both ends of the link in a MIMO system, the elements of the channel matrix corresponding to reception that is cross-polarized to the transmission will be weak relative to the copolarized reception [23]. This produces a decrease in the overall receiver SNR, which tends to compensate for the reduced signal correlation enabled by the multipolarized signaling.

Furthermore, while this analysis using infinitesimal dipoles is instructive, realizing the performance suggested in Figure 26.6 with practical antennas is difficult. For example, using half-wavelength dipoles and full-wavelength loops leads to strong mutual coupling and nonideal pattern characteristics that can reduce the number of independent channels.
For this reason, alternate topologies such as the MIMO cube antenna have been proposed, which achieve a combination of pattern and spatial diversity (see Section 26.3.3) [24]. This antenna consists of 12 dipoles placed along the edges of a cube, as depicted in Figure 26.7. Computational studies show that the pattern diversity allows this antenna to provide good performance even when the side length is as small as $\lambda/20$.

There are of course other ways of achieving pattern diversity. For example, it is possible to obtain different magnitude and polarization responses by exciting multiple modes of an antenna. Figure 26.8 illustrates the concept of using different coaxial feeds in a circular microstrip patch antenna positioned so that different relative excitations on the two ports excite different patch modes. The resulting radiation patterns have been shown to exhibit low correlation and reasonable capacity for many practical propagation scenarios [25]. However, when using this multimode principle, it is important to use modes that not only provide high diversity but also properly direct their energy in angular regions where multipath power is the highest. Failure to do so can reduce the effective received signal power so severely that the benefit gained by the low correlation created by the patterns can be outweighed by the loss in SNR due to poor excitation/reception within the environment of interest [26]. Naturally, another option is simply to use different antenna element types in the design or placing the same element with different orientations on the communications device. Regardless of the approach, the framework considered here provides a method for determining the performance of the array based on knowledge of the element radiation patterns.

![Figure 26.7](image1.png) MIMO cube antenna consisting of 12 dipole antennas placed along the edges of a cube.

![Figure 26.8](image2.png) Circular microstrip patch antenna with two different coaxial feed points positioned to excite two different radiation modes.
26.3.3 Spatial Diversity

The pattern diversity discussion in Section 26.3.2 is concerned with the types of elements that should be used in a multiantenna system. The concept of spatial diversity, on the other hand, emphasizes the array topology that should be used. The most straightforward illustration of spatial diversity assumes the propagation environment given in Eq. (26.37) assuming $\Delta \theta = 0$ and $\Delta \phi = \pi$ and with two Hertzian dipoles separated by a distance $d$ and placed along the $x$ axis. The covariance matrix for this scenario is given by

$$R_{s,mm} = \frac{1}{2\pi} \int_0^{2\pi} d\phi e^{i\frac{kd}{2}} \cos\phi = 1$$  \hspace{1cm} (26.39)

$$R_{s,mp} = \frac{1}{2\pi} \int_0^{2\pi} \left( e^{i\frac{kd}{2}} \cos\phi \right) e^{i\frac{kd}{2}} \cos\phi = J_0(kd)$$  \hspace{1cm} (26.40)

where $k$ is the free-space wavenumber and $J_0(\cdot)$ is the Bessel function of the first kind of order 0.

Figure 26.9 plots the number of effective diversity branches as a function of the dipole spacing for this configuration using a reference SNR of $\Lambda_0 = 1$ and thermal noise power of $\sigma_t^2 = 1$. It is rather intuitive that the number of branches is small at close spacing and increases to the value of two at larger spacing where the correlation is small. What may be surprising, however, is that the analysis predicts a number of diversity branches that is larger than unity as the spacing approaches zero. This phenomenon is a result of the fact that the analysis ignores electromagnetic effects (electromagnetic coupling), and therefore the two antennas both receive a full share of power even when the spacing is small. Naturally, a more comprehensive analysis, such as the one provided in Section 26.4.2, would limit the effective aperture of the coupled antennas so that this nonphysical behavior would be removed from the computation. Despite this limitation, this analysis reveals that as long as the antenna spacing is larger than roughly $\lambda/4$, the two dipoles essentially behave independently in terms of providing diversity performance.

![Figure 26.9](image-url)  

**Figure 26.9** Number of effective diversity branches versus element spacing for two Hertzian dipoles. The incident power is uniformly distributed in angle in the horizontal plane.
While this simple analysis provides insight into some considerations relevant to array design, the actual problem of array synthesis is generally complicated. There has been one notable study where several different array configurations were explored for both the base station and the mobile unit in an outdoor environment [27]. The base station antennas included single and dual polarization array and multibeam structures. The arrays on the mobile were constructed from monopoles to achieve spatial, angle, and/or polarization diversity. All of the array configurations provided very similar performance, with the exception of the multibeam base station antennas, which resulted in a 40–50% reduction in measured capacity since generally only one of the beams pointed in the direction of the mobile. These results suggest that average capacity is relatively insensitive to array configuration provided the signal correlation is adequately low and that all elements in the array have patterns that are effective for the environment of interest. More on the issue of antenna synthesis is found in Section 26.6.

26.4 MUTUAL COUPLING

Up to this point in the discussion, the concept of antenna electromagnetic coupling, generally referred to as mutual coupling, has been neglected in the analysis. In systems that can allow relatively large electrical spacing between the antenna elements, ignoring the impact of coupling is justified. However, in many compact devices, the antenna elements must be closely spaced, and the resulting antenna mutual coupling can impact communication performance. It is therefore important to have a framework for analyzing the impact of coupling on the system operation.

Before embarking on a detailed analysis of mutual coupling in MIMO and diversity systems, it is important to first understand the impact of coupling on the antenna radiation characteristics. We can divide the overall effect into two different physical phenomena:

1. Consider an antenna in isolation and driven by a terminal current $i_1(\omega)$. Electromagnetic analysis can be used to construct the radiation pattern for this system. However, if a second antenna is terminated in an open circuit and brought into proximity of the driven element, the boundary conditions used in the electromagnetic analysis change, thereby changing the effective radiation pattern of the driven element. We note, however, that the open-circuit termination does not allow terminal current to flow in the parasitic element, and for many types of elements (e.g., half-wave dipoles) the impact of the altered boundary conditions on the pattern is often quite minor. As a result, in some analyses, the radiation pattern for the element in isolation (isolated element pattern) is used in place of the pattern for the driven element with other elements present but terminated in an open circuit (open-circuit pattern). In the following, we assume that this open-circuit pattern with unit driving current for the $n$th element is denoted as $\mathbf{\bar{r}}_{T,n}(\Omega_T)$.

2. The open-circuit voltage $\hat{v}_2$ induced at the coupled antenna terminals is related to the current $i_1$ in the driven element according to $\hat{v}_2 = Z_{21}i_1$, where $Z_{21}$ is the mutual impedance. If the coupled antenna is now terminated with a load impedance $Z_L$, the induced voltage will create a current in the coupled antenna that depends on the termination impedance. Therefore the effective radiation pattern...
is the superposition
\[
\mathbf{e}_{T}(\Omega_T) = i_1 \left[ \mathbf{e}_{T,1}(\Omega_T) + \frac{Z_{21}}{Z_L} \mathbf{e}_{T,2}(\Omega_T) \right]
\] (26.41)

The composite pattern therefore depends on the load attached to the coupled element. Naturally, by reciprocity, the analysis at the receiver parallels that at the transmitter.

Item 2 makes it clear that ambiguity exists in defining the pattern to use in the correlation analysis expressed in Eq. (26.26). Furthermore, an impedance matching network can have impact on the ultimate superposition of the signals at the antenna terminals. Finally, the way in which this superposition is performed has impact on both the signal and the noise, and therefore any thorough analysis of diversity and MIMO systems with coupled antennas must consider the noise sources involved. In this section, we present approaches for incorporating antenna mutual coupling and realistic noise sources into the analysis. We begin our analysis by focusing on the diversity performance of a receiver subsystem. This is then extended in Section 26.4.3 to incorporate the transmitter and a MIMO propagation channel.

### 26.4.1 Receiver Modeling: Terminated Antennas

The traditional approach for incorporating mutual impedance into the analysis of receive diversity is to first compute the voltage \( \hat{v}_{s,m} + \hat{v}_{i,m} \) at the open-circuited terminals of the \( m \)th receive antenna with all other elements terminated in an open circuit using the integrations in Eqs. (26.23) and (26.27). Since the open circuit patterns are specified as the patterns with all other elements open-circuited, this step incorporates the impact of adjacent elements on the electromagnetic boundary conditions of the problem. The challenge now is simply to incorporate the impact of the antenna array impedance matrix into the analysis.

Figure 26.10 shows a block diagram of the receiver model, which includes the incident field, coupled receive array with full impedance matrix \( \mathbf{Z}_R \), and a termination characterized by the impedance matrix \( \mathbf{Z}_L \). Note that we are not necessarily constraining the load impedance matrix \( \mathbf{Z}_L \) to be diagonal, implying the possibility of a coupled load network in addition to the coupled antenna array. In this model, any thermal noise generated within the receiver network should be represented as an open-circuit noise voltage at

![Figure 26.10](image)

**Figure 26.10** Block diagram of a receiver diversity system including the incident field, mutually coupled array, and loads.
the antenna terminals represented as $\hat{v}_t$. Using simple network analysis, the vector $v_R$ of voltages at the terminals of the load network can be expressed in terms of the total open-circuit voltage vector $\hat{v}$ according to [28–33]

$$v_R = Z_L (Z_L + Z_R)^{-1} \hat{v} = C_R \hat{v} \quad (26.42)$$

where $C_R$ is a coupling matrix.

With this representation, the covariance matrix of the voltage across the receiver terminations is given as

$$E\{v_R v_R^\dagger\} = C_R R_s C_R^\dagger + C_R R_\eta C_R^\dagger = \tilde{R}_s + \tilde{R}_\eta \quad (26.43)$$

where the signal and noise covariance matrices are defined in Eqs. (26.26) and (26.28), respectively. The transformed covariance matrices $\tilde{R}_s$ and $\tilde{R}_\eta$ can then be used in the framework of Section 26.3.1 to analyze the system performance. This approach, which has been used to analyze the performance of adaptive array [28] and diversity [29] systems, effectively captures the physics of electromagnetic coupling into the formulation [30–33].

It should be emphasized that some analysis is required to reflect receiver front-end noise to an effective open-circuit antenna voltage. Furthermore, while it is tempting to simply use this coupling matrix formulation to analyze the impact of coupling termination on the signal and then assume that the noise voltages on the output branches are uncorrelated with each other and equal power ($\tilde{R}_\eta = \sigma_\eta^2 I$), this approach does not yield the correct noise properties. Finally, this approach does not conveniently allow incorporation of more sophisticated noise models such as are encountered in practical front-end low noise amplifiers.

### 26.4.2 Receiver Modeling: Front-End Network

Because the diversity or MIMO capacity performance depends critically on the SNR observed on each diversity branch or MIMO subchannel, properly incorporating appropriate thermal noise models into the analysis framework is important if the results are to properly reflect the achievable system performance. When considering such an analysis framework, it is important to remember that the noise performance of a transistor-based amplifier depends on the impedance of the source driving the amplifier. Therefore the model must properly include the impedance interface between the antenna terminals and the low noise amplifiers and accurately represent the nature of the amplifier thermal noise. Unfortunately, it is not obvious how the coupling matrix analysis approach discussed previously can easily be modified to include these effects.

Naturally, when increasing the level of modeling detail, one could include additional levels of complexity such as component nonlinearity, conversion loss, matched-filter imperfections, and sampling. While all of these issues can be important, we limit our focus to determining the branch SNR levels achieved by the antenna interfaced to the RF front-end. Figure 26.11 shows a block diagram of the resulting system model, which includes the incident field, coupled receive array, an impedance matching network, noisy amplifiers (used to model the source of thermal front-end noise), and a termination. Since impedance parameters do not lend themselves well to the analysis of cascaded system blocks, we instead use scattering parameters (S-parameters) referenced to a real impedance $Z_0$ [34] to describe the network signal flow wherein the incident and reflected
26.4 MUTUAL COUPLING

traveling waves are denoted as \( \mathbf{a} \) and \( \mathbf{b} \), respectively. The flow diagram for this network, with the various system blocks delineated by dashed lines, is shown in Figure 26.12. The specific traveling-wave vectors, S-parameter matrices (symbol \( \mathbf{S} \)), and reflection coefficient matrices (symbol \( \mathbf{\Gamma} \)) appearing in Figure 26.11 are used in the analysis.

26.4.2.1 Network Analysis The first step in our analysis is to properly transform the open-circuit antenna terminal voltage representing the signal plus external interference to a voltage across the receiver terminations and add to this the contribution due to amplifier thermal noise. We begin with the vector \( \mathbf{b}_R \), which represents the traveling wave resulting from the incident signal plus interference that will be delivered by the receive antenna terminals to a set of independent loads of resistance \( Z_0 \) (so that \( \mathbf{b}_1 = \mathbf{0} \)). Given this as a source wave, the general expression for the wave \( \mathbf{a}_1 \) incident on the matching network input is given by

\[
\mathbf{a}_1 = \mathbf{b}_R + \mathbf{S}_R \mathbf{b}_1
\]  

(26.44)

where \( \mathbf{S}_R \) represents the S-parameter matrix for the coupled array. If the antenna array is terminated in an open circuit, then \( \mathbf{b}_1 = \mathbf{a}_1 \). Substituting this observation into Eq. (26.44) produces

\[
\mathbf{b}_R = (\mathbf{I} - \mathbf{S}_R) \mathbf{a}_1
\]  

(26.45)
However, since we already know that the voltage $\hat{v}_s + \hat{v}_i$ represents the open-circuit voltage at the antenna terminals and since $\hat{v}_s + \hat{v}_i = Z_0^{1/2}(a_1 + b_1) = 2Z_0^{1/2}a_1$ (again using that $b_1 = a_1$ for an open-circuit termination), we can use this expression in Eq. (26.45) to obtain

$$b_R = \frac{1}{2Z_0^{1/2}}(I - S_R)(\hat{v}_s + \hat{v}_i)$$  \hspace{1cm} (26.46)

This gives us a mechanism for computing $b_R$ based on our knowledge of the open-circuit antenna voltage vector and the S-parameter matrix of the coupled array.

The multiport matching network is described by the block S-parameter matrix

$$S_M = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix}$$  \hspace{1cm} (26.47)

where 1 and 2 refer to input and output ports, respectively. With this representation, we use Eq. (26.44) with $b_1 = S_{11}a_1 + S_{12}a_2$ to obtain

$$a_1 = (I - S_RS_{11})^{-1}(b_R + S_RS_{12}a_2)$$  \hspace{1cm} (26.48)

Since $b_2 = S_{21}a_1 + S_{22}a_2$ we have

$$b_2 = S_{21}(I - S_RS_{11})^{-1}b_R + \left(S_{22} + S_{21}(I - S_RS_{11})^{-1}S_RS_{12}\right)a_2$$  \hspace{1cm} (26.49)

where we have used $\Gamma_0$ to represent the reflection coefficient at the matching network output (see Figure 26.11).

The amplifier block introduces noise into the system, with the $m$th amplifier contributing forward and reverse noise waves $a_{t,m}$ and $b_{t,m}$, respectively, at the amplifier input [35]. The negative branch gain used to connect $a_r$ into the flow diagram of Figure 26.12 is simply a convention used commonly in noise analysis [35]. Furthermore, the amplifier S-parameters are represented using the block matrix structure in Eq. (26.47), but with the blocks denoted by $\tilde{S}_{ij}$. Using this notation, the amplifier output waves are expressed as

$$a_2 = \tilde{S}_{11}b_2 + \tilde{S}_{12}b_L - \tilde{S}_{11}a_1 + b_t$$  \hspace{1cm} (26.50)

$$a_L = \tilde{S}_{21}b_2 + \tilde{S}_{22}b_L - \tilde{S}_{21}a_t$$  \hspace{1cm} (26.51)

Inserting Eq. (26.50) with $b_L = \Gamma_La_L$ into Eq. (26.49) and rearranging leads to

$$b_2 = (I - \Gamma_0\tilde{S}_{11})^{-1}\left[S_{21}(I - S_RS_{11})^{-1}b_R + \Gamma_0(b_t - \tilde{S}_{11}a_t + \tilde{S}_{12}\Gamma_la_L)\right]$$  \hspace{1cm} (26.52)

Similarly, using $b_L = \Gamma_La_L$ in Eq. (26.51) leads to

$$a_L = (I - S_{21}\Gamma_L)\left(\tilde{S}_{21}b_2 - \tilde{S}_{21}a_t\right)$$  \hspace{1cm} (26.53)

Substituting Eq. (26.52) into this result gives us the voltage $v_L$ across the loads as

$$v_L = Z_0^{1/2}(I + \Gamma_L)a_L$$

$$= Q\left[G(\hat{v}_s + \hat{v}_t) + \Gamma_0(b_t - a_t)\right]$$  \hspace{1cm} (26.54)
26.4 MUTUAL COUPLING

where

\[ G = \frac{1}{2Z_0^{1/2}} S_{21} (I - S_R S_{11})^{-1} (I - S_R) \]  \hspace{1cm} (26.55)

and

\[ Q = Z_0^{1/2} (I + \Gamma_L) [ (I - \Gamma_0 \tilde{S}_{11}) \tilde{S}_{21}^{-1} (I - \tilde{S}_{22} \Gamma_L) - \Gamma_0 \tilde{S}_{12} \Gamma_L]^{-1} \]  \hspace{1cm} (26.56)

We are now equipped with an expression representing the total signal plus noise across the loads given the signal and interference incident on the array and the thermal noise generated in the low noise amplifiers.

26.4.2.2 Matching Network Specification

Armed with this expression for the output signal and noise voltages, we are prepared to discuss specification of the matching network, which is in reality coupled with a discussion on amplifier design. Practical amplifier design involves specifying a performance goal and synthesizing the source and load terminations that achieve this goal. Signal amplifiers are typically designed to provide minimum noise figure, optimal power gain, or some compromise between the two 36. Our task is to define a desired value of \( \Gamma_0 \), which is the source termination seen by the amplifier, and use this value to determine \( S_M \) for the matching network.

To begin this analysis, we take the SVD of each subblock using the representation \( S_{ij} = U_{ij} \Lambda_{ij}^{1/2} V_{ij}^\dagger \), where the square root on the diagonal matrix of real singular values is used for notational convenience. Assuming that the matching network is lossless (\( S_{Mi} S_M = I \)) and reciprocal (\( S_M = S_M^T \)), it can be shown that the subblocks can be written

\[ S_{11} = U_{11} \Lambda_{11}^{1/2} U_{11}^T \]
\[ S_{12} = j U_{11} (I - A_{11})^{1/2} U_{22}^T \]
\[ S_{21} = j U_{22} (I - A_{11})^{1/2} U_{11}^T \]
\[ S_{22} = U_{22} \Lambda_{22}^{1/2} U_{22}^T \]  \hspace{1cm} (26.57)

With \( \Gamma_0 = U_0 \Lambda_0^{1/2} U_0^T \) representing the SVD of the reciprocal matrix \( \Gamma_0 \), we use the expression for \( \Gamma_0 \) in Eq. (26.49) and the relations in Eq. (26.57) to obtain

\[ \Gamma_0 = U_{22} [ \Lambda_{11}^{1/2} - (I - A_{11})^{1/2} T (I - A_{11})^{1/2} ] U_{22}^T \]  \hspace{1cm} (26.58)
\[ T = U_{11}^T (I - S_R U_{11} A_{11}^{1/2} U_{11}^T)^{-1} S_R U_{11} \]  \hspace{1cm} (26.59)

We have flexibility in specifying the singular vectors \( U_{11} \) and \( U_{22} \) and therefore choose representations that lead to mathematical simplicity. First, we see that if \( S_R = U_R \Lambda_R^{1/2} U_R^T \) represents the SVD of the reciprocal matrix \( S_R \), then by choosing \( U_{11} = U_R^T \), we obtain

\[ T = (I - \Lambda_R^{1/2} A_{11}^{1/2})^{-1} \Lambda_R^{1/2} \]  \hspace{1cm} (26.60)

which is diagonal. If we further choose \( U_{22} = U_0 \), we can solve Eq. (26.58) to obtain

\[ A_{11}^{1/2} = (A_0^{1/2} + \Lambda_R^{1/2}) (I + A_0^{1/2} \Lambda_R^{1/2})^{-1} \]  \hspace{1cm} (26.61)
We now have identified the values of all matrices in Eq. (26.57), which means that the matching network has been specified to achieve the design goals.

We generally assume that the amplifiers and loads are uncoupled ($\hat{S}_{ij}$ and $\Gamma_L$ are diagonal), so that typical design goals are achieved for diagonal $\Gamma_0$. If $\Gamma_{\text{opt}}$ and $\Gamma_{\text{MS}}$ represent the (scalar) source reflection coefficient for achieving amplifier minimum noise figure and maximum power gain [36], respectively, then these goals are achieved by setting $\Gamma_0 = \Gamma_{\text{opt}}I$ and $\Gamma_0 = \Gamma_{\text{MS}}I$. Since the performance of a multiantenna system depends on SNR, we expect a design for minimum noise figure to outperform one for maximum power gain. We also point out that if $\Gamma_0 = 0$, then any reverse traveling noise from the transistor will not be reflected back into the amplifier, leading to perhaps reduced SNR compared to a design for optimal power gain. This condition can be obtained by setting $S_{11} = S_R^\dagger$.

To achieve diagonal $\Gamma_0$, the matching network must be coupled to “undo” the coupling created by the antenna, and it therefore acts as an array combining network as well as an impedance transforming network. This is an important observation, since the linear combination of the signals in the matching network is performed before injection of the thermal noise by the amplifiers, implying that the matching network can have significant impact on the final system SNR. It is also interesting that if the network produces diagonal $\Gamma_0$, the effective radiation patterns observed at the output of the matching network are mutually orthogonal, indicating that they can achieve perfect radiation efficiency as demonstrated in Ref. 37 as well as zero correlation in an environment where the multipath field is equally likely to arrive at any angle on the sphere [29]. These features of a coupled antenna along with a coupled matching network can therefore result in a diversity performance that is higher than what could be obtained with uncoupled antennas.

Finally, while optimal matching networks must be coupled, practically speaking it is easier to design an uncoupled network. We assume that the coupled antenna impedance can be represented using the diagonal matrix $\tilde{Z}_R$ containing only the diagonal elements of the full impedance matrix $Z_R$. The resulting diagonal S-parameter matrix $\tilde{S}_R$ has elements $\tilde{S}_{R,ii} = (\tilde{Z}_{R,ii} - Z_0)/(\tilde{Z}_{R,ii} + Z_0)$. This value of $\tilde{S}_R$ is then used in place of $S_R$ when formulating the matching network S-parameter matrix. However, when analyzing the performance of this self-impedance match, the nondiagonal form of $S_R$ must be used in Eq. (26.54). We should also mention that while the self-impedance matching approach provides a mechanism for creating an uncoupled matching network for a coupled antenna array, this does not generally represent the optimal uncoupled matching network in terms of diversity performance. Analytic and numerical optimizations can yield matching networks that outperform this self-impedance match for highly coupled antenna arrays [38].

### Covariance Formation

Computing the impact of coupling and the matching network on the diversity performance requires that we formulate the covariance matrices for the signal and noise voltage vectors based on our transfer function of Eq. (26.54). For the thermal noise, we assume that the noise in each amplifier is uncorrelated with that of all other amplifiers, leading to the expressions

\[
\begin{align*}
\mathbb{E}\{a_i a_i^\dagger\} &= k_B BT_0^R I \\
\mathbb{E}\{b_i b_i^\dagger\} &= k_B BT_0^R I \\
\mathbb{E}\{a_i b_i^\dagger\} &= k_B BT_0^R I
\end{align*}
\]  

(26.62)
where $k_B$ is Boltzmann’s constant, $B$ is the system noise power bandwidth, and $T_{\alpha}$, $T_\beta$, and $T_\gamma = T_\gamma e^{i\phi_\gamma}$ are amplifier effective noise temperatures that can readily be computed from other transistor noise parameters [35]. The covariance of the thermal noise voltage at the terminations is therefore

$$\tilde{R}_t = Q \mathbf{E} \left\{ (\Gamma_0 b_t - a_t)(\Gamma_0 b_t - a_t)^\dagger \right\} Q$$

(26.63)

$$\tilde{R}_t = k_B B (T_\alpha + T_\beta \Gamma_0^\dagger - T_\gamma \Gamma_0 - T_\gamma^\dagger \Gamma_0^\dagger)$$

(26.64)

For the signal and external interference, the output voltage covariances become

$$\tilde{R}_s = Q G R_s G^\dagger Q^\dagger$$

(26.65)

$$\tilde{R}_i = Q G R_i G^\dagger Q^\dagger$$

(26.66)

where the signal and noise covariance matrices are defined in Eqs. (26.26) and (26.28), respectively. These transformed covariance matrices can then be used to compute the diversity performance using the framework of Section 26.3.1.

### 26.4.2.4 Computational Example

To demonstrate application of this framework and quantify the impact of coupling on the diversity performance, we conduct a simple computational example. The antennas used in this representative computation are two $z$-oriented half-wave (total length) dipoles, separated by the distance $d$, and each with a wire radius of $0.01\lambda$. To accurately characterize these coupled antennas, we use the finite-difference time-domain (FDTD) [39, 40] computational approach with single-frequency antenna excitation. The FDTD grid uses 80 cells per wavelength in the $z$ direction and 200 cells per wavelength in the $x$ and $y$ directions to adequately model the azimuthal current variations for close antenna spacings. A buffer region of a quarter-wavelength is placed between the antennas and the terminating eight-cell perfectly matched layer (PML) absorbing boundary condition (ABC). Pattern computations are performed when one antenna is excited while the second is terminated in an open circuit. The antenna S-parameter matrix $S_R$ is computed with the antennas terminated in $Z_0$.

The incident field is vertically polarized with a power angular spectrum given by Eq. (26.37) with $\Delta\theta = 0$ and $\Delta\phi = \pi$, which means that the field is equally likely to arrive from any angle in the horizontal plane. We also assume that there is no external interference present in the environment. To maintain a consistent reference as antenna parameters (such as spacing) are swept, we characterize an isolated dipole using the FDTD approach and then compute the scalar variance $\tilde{R}_s$ for this single antenna in the propagation environment. For simplicity, we use a matching network corresponding to $\Gamma_0 = 0$ in this computation, leading to an average SNR for the single isolated dipole of $\tilde{R}_s / k_B B T_\alpha$. The reference CDF curves for the independent branch signals used in computation of the effective number of branches for the coupled system are then constructed based on this value of SNR.

The transistor used for the amplifier is a BJT taken from an application note [41]. At a collector–emitter bias voltage of 10 V, collector current of 4 mA, frequency of 4 GHz, and reference impedance of $Z_0 = 50 \Omega$, the S-parameters and noise parameters are
given as

\[
\begin{align*}
\hat{S}_{11} &= 0.552 \angle 169^\circ \\
\hat{S}_{12} &= 0.049 \angle 23^\circ \\
\hat{S}_{21} &= 1.681 \angle 26^\circ \\
\hat{S}_{22} &= 0.839 \angle -67^\circ \\
F_{\text{min}} &= 2.5 \text{ dB} \\
\Gamma_{\text{opt}} &= 0.475 \angle 166^\circ \\
R_o &= 3.5 \, \Omega
\end{align*}
\]

(26.67)

where \( F_{\text{min}} \), \( \Gamma_{\text{opt}} \), and \( R_o \) are the device minimum noise figure, optimal source termination for noise figure, and effective noise resistance, respectively. These parameters are converted to the effective noise temperatures \( T_o, T_j, \) and \( T_R \) using algebraic relations [35].

We now explore the impact of matching on the diversity performance of the coupled antennas as a function of the dipole spacing. In the examples, we use matching networks designed to achieve optimal amplifier noise figure and optimal power gain. Matching network synthesis is based on the full antenna coupling matrix \( S_R \) as well as the diagonal coupling matrix \( \tilde{S}_R \) as discussed in Section 26.4.2.2 (self-impedance match).

Figure 26.13 plots the number of effective diversity branches as a function of dipole spacing for matching networks achieving optimal noise figure or power gain. Several observations regarding these results deserve attention. First, for very close antenna spacings, the two antennas behave largely as a single element, resulting in approximately one effective diversity branch. This is in contrast to the plot in Figure 26.9, where the fact that coupling was neglected resulted in the number of effective branches being larger than unity for zero spacing. The low diversity performance for close spacing increases rapidly with separation and for certain moderate spacings can actually exceed the diversity gain achieved for large element separation. This peak in the diversity performance stems from the pattern orthogonality created by the coupled antennas and matching network as discussed in Section 26.4.2.2, which leads to higher diversity than can be achieved with the uncoupled dipoles with the same spacing used as a reference in the diversity.
gain computation. The height of this peak is also influenced by the fact that the reference antenna used for this computation has a suboptimal match of \( \Gamma_0 = 0 \), which means that the optimal match of \( \Gamma_0 = \Gamma_{opt} \) for the coupled antennas should outperform the reference dipole for larger element spacings.

The results of Figure 26.13 also show that matching to the self-impedance creates relatively little degradation in performance, particularly for element spacings larger than about \( \lambda/4 \). It is also particularly revealing that while optimal power transfer is a typical design goal, it is dramatically suboptimal in terms of diversity performance. This is an intuitive result, since matching for maximum power transfer neglects the impact of the match on amplifier noise figure, which directly controls the received SNR, the key parameter in determining the overall communication performance. This superiority of matching for minimum noise figure is therefore general for any receiving system equipped with practical noisy amplifiers.

### 26.4.3 MIMO System Modeling

To extend our discussion of antenna mutual coupling to the case of MIMO communication, we must augment our modeling framework by adding the transmitter and properly characterizing the MIMO communication channel. We use the same framework to perform this augmentation for the models in Sections 26.4.1 and 26.4.2, and then use this augmentation to explore some implications of antenna coupling on MIMO system performance.

#### 26.4.3.1 Network Analysis

Figure 26.14 illustrates an augmentation of the diversity block diagram from Figure 26.10 appropriate for characterizing a MIMO system. In this representation, the transmit array is characterized by the impedance matrix \( Z_T \) and is fed through a source impedance network by a set of generators creating the voltage \( v_S \). The antenna terminal voltage is related to the generator voltage vector as

\[
v_T = Z_T (Z_T + Z_S)^{-1} v_S = \tilde{C}_T v_S \tag{26.68}
\]

where \( \tilde{C}_T \) is a coupling matrix.

We can now combine this treatment with the analysis of Section 26.4.1 to arrive at a MIMO system transfer function. For example, if we assume that \( H_{A,mn} \) represents the ratio of the open-circuit voltage \( \hat{v}_m \) on the \( m \)th receive antenna to the driving point voltage \( v_S \), we have

\[
H_{A,mn} = \frac{v_m}{v_S}.
\]

Figure 26.14

Augmentation of the block diagram of Figure 26.10 to represent a MIMO system with coupled transmit and receive arrays.
voltage $v_{T,n}$ on the $n$th transmit antenna, then the transfer function becomes simply (see Eq. (26.42))

$$v_R = C_R H_A C_T v_S$$ \hspace{1cm} (26.69)

At this point, we must revisit the more general analysis of Sections 26.2.2 and 26.2.3, where we chose to formulate the channel transfer matrix $H_A$ using transmit radiation patterns constructed with all other elements terminated in an open circuit. This allowed the superposition in Eq. (26.2) to properly represent the radiated field provided that the excitation vector $x_A$ represents a driving point current. If, instead, we let the excitation vector represent a driving point voltage, then the radiation patterns used in the construction of $H_A$ must be constructed with all other elements terminated in a short circuit. This approach is possible, but it has some practical disadvantages:

1. It is not particularly standard to characterize antenna patterns in this fashion, and in fact it can be challenging to obtain accurate numerical computations of such a scenario.

2. Since this short-circuit pattern will likely differ significantly from the pattern of the antenna in isolation, this approach does not lend itself to the approximation that the coupled element pattern is approximately the same as the isolated pattern.

3. It is common to assume that $H_A$ consists of independent and identically distributed (i.i.d.) Gaussian random variables, which has been shown to be a reasonable approximation for dipole antennas characterized by their isolated element patterns. The radiation patterns of the elements in the presence of other short-circuited elements will generally be significantly perturbed from their isolated values, which indicates that the i.i.d. Gaussian assumption may not be valid under these circumstances.

4. The pattern characterization at the receiver (which requires open-circuit patterns) differs from that at the transmitter, which may be inconvenient in practical scenarios.

We can, instead, assume that $H_{A,mn}$ represents the ratio of the open-circuit voltage $\hat{v}_{mn}$ on the $m$th receive antenna to the driving point current $i_{T,n}$ on the $n$th transmit antenna. We must modify Eq. (26.68) as

$$i_T = (Z_T + Z_S)^{-1} v_S = C_T v_S$$ \hspace{1cm} (26.70)

The transfer matrix $H_A$ now can be constructed using patterns with all other elements terminated in an open circuit, and the MIMO transfer relationship becomes

$$v_R = C_R H_A C_T v_S$$ \hspace{1cm} (26.71)

There have been a variety of studies based on this approach [30, 31, 33]. Perhaps most notable is the study of Janaswamy [31], which uses a detailed analysis based on the MIMO transfer relationship of Eq. (26.69). This analysis reveals that increasing the number of antennas in a uniform linear array of dipoles (of fixed aperture size) cannot increase the capacity beyond a certain limit due to the impact of coupling. This is intuitive,
since mutual coupling limits the ability of closely spaced elements to independently excite or sample the propagation environment.

The same approach can be used to augment the model of Section 26.4.2, which includes a detailed description of the receiver front-end network. Figure 26.15 shows the block diagram representation for the MIMO system, where the receiver architecture is identical to that detailed in Figure 26.11. Again assuming that the transmit element radiation patterns are obtained with all other elements terminated in an open circuit and that the elements of the transfer matrix $H_A$ represent the ratio of the open-circuit receive antenna voltages to the driving point currents, the model of Eq. (26.54) is simply augmented by recognizing that the open-circuit signal voltage is $\hat{v}_s = H_A C_T v_S$. The voltage transfer function therefore becomes

$$v_L = Q [G H_A C_T v_S + G \hat{v}_i + \Gamma_0 b_i - a_i]$$  \hspace{1cm} (26.72)

We use this expression for the remainder of this discussion on mutual coupling for MIMO systems since it provides a detailed representation of the interference and thermal noise.

### 26.4.3.2 MIMO System Capacity

Computing the capacity for the MIMO signal relationship in Eq. (26.72) requires that we once again formulate the covariance of $v_L$, assuming the input signal $v_S$, interference, and noise are drawn from zero-mean complex Gaussian distributions. The interference and noise covariances are given by Eq. (26.63) and (26.66), respectively. If we let $R_S = E\{v_S v_S^\dagger\}$ represent the covariance of the input signal, then the covariance of the signal at the receiver loads is given by

$$\tilde{R}_s = Q G H_A C_T R_S C_T^\dagger H_A^\dagger G^\dagger = Q \tilde{H}_A R_S \tilde{H}_A^\dagger Q^\dagger$$ \hspace{1cm} (26.73)

The capacity for this MIMO system is obtained by determining the matrix $R_S$ that maximizes the mutual information expression

$$I(v_L, v_S) = \log_2 \left| \frac{Q \left[ \tilde{H}_A R_S \tilde{H}_A^\dagger + \tilde{R}_\eta \right] Q^\dagger}{QR_S Q^\dagger} \right|$$ \hspace{1cm} (26.74)

$$= \log_2 \left| \tilde{R}_\eta^{-1/2} \tilde{H}_A R_S \tilde{H}_A^\dagger \left( \tilde{R}_\eta^{-1/2} \right)^\dagger + I \right|$$ \hspace{1cm} (26.75)

where $\tilde{R}_\eta = GR \tilde{G}^\dagger + \tilde{R}_i$ and we have assumed that the matrix $Q$ is nonsingular.

![Figure 26.15](image-url)  

**Figure 26.15** Block diagram of a MIMO system including the coupled transmit and receive arrays, MIMO channel, and the receiver network (detailed in Figure 26.11).
Equation (26.75) is now in a form that can be maximized using the water-filling procedure referred to in Section 26.2.4. We note, however, that this optimization is coupled with a power constraint that limits the diagonal elements of the transmit signal covariance matrix $R_S$. While this is a reasonable constraint for uncoupled antennas, it should be noted that for coupled antennas this will not limit the amount of power radiated by the array. Since this radiated power is what is specified by regulating authorities and is the power that must be provided by the generators, it is logical to formulate the capacity based on the power radiated.

Assuming lossless antennas, the power radiated by the array averaged over one sinusoidal cycle can be computed from

$$p_{\text{rad}} = \frac{1}{2} \text{Re}\{i_T^\dagger Z_T i_T\} = \frac{1}{2} \text{Re}\{i_T^\dagger Z_T i_T + i_T^\dagger Z_T i_T\}$$  \hspace{1cm} (26.76)

$$= \frac{1}{4} i_T^\dagger (Z_T + Z_T^*) i_T$$  \hspace{1cm} (26.77)

$$= \frac{1}{2} i_T^\dagger \text{Re}\{Z_T\} i_T = v_S^\dagger A v_S$$  \hspace{1cm} (26.78)

where we have used Eq. (26.70) and $A = \frac{1}{2} C_T^\dagger \text{Re}\{Z_T\} C_T$. In this formulation, we have taken the transpose of the second (scalar) term on the right-hand side of Eq. (26.76) and have used that $Z_T$ is symmetric for reciprocal antennas. For zero-mean signals, the average radiated power is given by

$$P_{\text{rad}} = E\{p_{\text{rad}}\} = \text{Tr}\{R_S A\} = \text{Tr}\left[A^{1/2} R_S A^{1/2}\right]$$  \hspace{1cm} (26.79)

where we have used the facts that $A$ is real and positive semidefinite and that matrices commute under the trace. Defining $\hat{R}_S = A^{1/2} R_S A^{1/2}$ and substitution into Eq. (26.75) leads to

$$I(v_L, v_S) = \log_2 \det\{H \hat{R}_S H^\dagger + I\}$$  \hspace{1cm} (26.80)

where $H = \hat{R}_S^{-1/2} H_A \hat{R}_S^{-1/2}$. The radiated power constraint is simply $\text{Tr}\{\hat{R}_S\} < P_T$. The problem is now in a form for which the capacity can be computed using the framework of Section 26.2.4.

**26.4.3.3 Computational Example** We can demonstrate the impact of coupling on the capacity using the receive antenna and amplifier configuration used in Section 26.4.2.4 with the addition of a transmit array, which is also a two-element array of half-wave dipoles with a half-wave element spacing. Once again, the interference is assumed to be zero. The channel matrix $H_A$ is constructed as a random realization of the SVA channel model discussed in Section 26.2.3, with the capacity results representing an average over 5000 such channel realizations.

For each physical channel realization, a single transmit and receive dipole are used in conjunction with a lossless receive matching network with $S_{11} = S_R^*$ so that $\Gamma_0 = 0$ (all terms are scalars). The single-input single-output (SISO) SNR is then taken as the square of the received signal voltage to the variance of the thermal noise, or

$$\text{SNR} = \frac{|GH_A C_T v_S|^2}{k_B B T_a} = \frac{|(1 - S_R) H_A|^2}{2Z_0 (1 - |S_R|^2) \text{Re}\{Z_T\}} \frac{P_T}{k_B B T_a}$$  \hspace{1cm} (26.81)
26.5 SUPERDIRECTIVITY IN MIMO SYSTEMS

where we have used that $|S_{21}|^2 = 1 - |S_{11}|^2$. This SNR value is then averaged in space by moving each dipole in $0.1 \lambda$ steps over a linear range of $1.5 \lambda$. For a given transmit power, the value of $k_B B$ can be computed to achieve an average SISO SNR (20 dB in this work) for the channel realization.

For each receive dipole spacing, we construct the matching network to achieve the specified design goal as outlined in Section 26.4.2.2. Figure 26.16 plots the capacity as a function of receive dipole spacing for matching networks that achieve minimum noise figure and maximum amplifier gain. Results for a coupled match and a simpler self-impedance match are included. We first observe that the match achieving minimum amplifier noise figure (noise figure of $F = F_{\text{min}} = 2.5$ dB) produces notably higher capacity than the match providing maximum power transfer, which generates a much higher noise figure of $F = 7.2$ dB. This result is intuitive, since ultimately capacity depends on SNR as opposed to absolute signal strength, as discussed in Section 26.4.2.4. We also observe that for close antenna spacings with high coupling, the shortcomings of the self-impedance match are evident. However, once the spacing reaches approximately $d = \lambda/4$, this match provides near optimal performance.

26.5 SUPERDIRECTIVITY IN MIMO SYSTEMS

The discussion in Section 26.4.3.2 revealed that constraining the square of the transmit antenna excitations, as is traditionally done in MIMO system capacity formulations, does not limit the power radiated by the antenna. Since it is this radiated power that is limited by regulatory agencies and since the transmit generators (amplifiers) must supply this power, it is logical that the capacity formulation should directly constrain the radiated power. This section discusses this issue in additional detail and demonstrates that when the radiated power is constrained the capacity solution specifies superdirective excitations [42–46] characterized by very high array directivity in preferred directions.
It also illustrates that external interference can lead to superdirective array weights at the receiver.

While the phenomenon of superdirectivity theoretically allows the system to advantageously exploit the propagation channel spatial characteristics, its implementation is typically considered impractical for a variety of reasons [43]. Therefore this section provides two mechanisms for computing the capacity when the level of superdirectivity is limited.

### 26.5.1 Capacity Expression

The goal of this formulation is to provide a simpler capacity expression than that obtained for a MIMO system with coupled antennas while exposing the mechanisms that lead to array superdirectivity. For this development, we assume that the excitation signal is the antenna driving point current $i_T$. As outlined in Section 26.4.3.2, one mechanism for computing the radiated power (averaged over one sinusoidal cycle) is to use the mutual impedance matrix $Z_T$ for the transmit array as

$$p_{rad} = \frac{1}{2} \text{Re} \{ Z_T \} i_T = i_T^\dagger A i_T \tag{26.82}$$

where $A = \frac{1}{2} \text{Re} \{ Z_T \}$.

In situations where the impedance matrix is unknown, it is also possible to compute the radiated power based on the transmitted field. If $\mathbf{e}_{T,n}(\Omega)$ represents the radiation pattern of the $n$th transmit element with all other elements terminated in an open circuit, the total transmitted field is

$$\mathbf{e}_T(\Omega) = \sum_{n=1}^{N_T} \mathbf{e}_{T,n}(\Omega) i_{T,n} \tag{26.83}$$

The radiated power is

$$p_{rad} = \frac{1}{2\eta_0} \int_{\Omega_T} \mathbf{e}_T^*(\Omega) \cdot \mathbf{e}_T(\Omega) \, d\Omega_T$$

$$= \sum_{n=1}^{N_T} \sum_{q=1}^{N_T} i_{T,n}^* \left[ \frac{1}{2\eta_0} \int_{\Omega_T} \mathbf{e}_{T,n}(\Omega) \cdot \mathbf{e}_{T,q}(\Omega) \, d\Omega_T \right] i_{T,q}^* \tag{26.84}$$

where $\eta_0$ is the intrinsic impedance of free space. In either case, the average radiated power is

$$P_{rad} = \text{Tr} [ \mathbf{R}_T A ] \tag{26.85}$$

where $\mathbf{R}_T = \mathbb{E} [ i_T i_T^\dagger ]$ is the covariance of the transmit currents. If we represent our transmit currents instead as $\mathbf{i}_T = A^{-1/2} \hat{\mathbf{i}}_T$, where $\hat{\mathbf{i}}_T$ has covariance $\hat{\mathbf{R}}_T = \mathbb{E} [ \hat{\mathbf{i}}_T \hat{\mathbf{i}}_T^\dagger ]$, then we can recast this power constraint to

$$P_{rad} = \text{Tr} \left[ A^{-1/2} \hat{\mathbf{R}}_T A^{-1/2} A \right] = \text{Tr} \left[ \hat{\mathbf{R}}_T \right] \tag{26.86}$$
where we have used the fact that \( A \) is real and positive semidefinite. It is this quantity that will be constrained in the capacity formulation.

We assume that the noise consists of external interference and that the element \( H_{mn} \) of the channel matrix represents the ratio of the open-circuit voltage on the \( m \)th receive antenna to the driving point current on the \( n \)th transmit antenna. In this case, the transfer relationship between the open-circuit receive voltage and the transmit current is simply

\[
\hat{v} = H_i \hat{v}_i = \hat{v}_s + \hat{v}_i \tag{26.87}
\]

Following the development of Eq. (26.28), the covariance of the open-circuit interference voltage \( \hat{v}_i \) at the receive antenna terminals is given by

\[
R_{i,m} = |\phi|^2 \int \overline{F}_{i}(\Omega) \cdot \overline{F}_{i}^{*}(\Omega_R) \cdot d\Omega_R \tag{26.88}
\]

where \( \overline{F}_{i}(\Omega) \) represents the interference power angular spectrum. Using a spatial pre-whitening filter for the noise and introducing our modified current definition transforms the transfer relationship to

\[
\hat{v}_0 = R^{-1/2}_i H A^{-1/2}_i + R^{-1/2}_i \hat{v}_i \tag{26.89}
\]

The mutual information for this transfer relationship can be maximized subject to a constraint on the radiated power defined in Eq. (26.86), as discussed in Section 26.2.4.

The effect of the radiated power constraint and interference can be observed from this mutual information expression. Small eigenvalues in \( A \) or \( R_i \) will lead to spatial subchannels with high gain in the effective channel matrix \( R^{-1/2}_i H A^{-1/2}_i \). When the transmit covariance \( R_T \) is constructed in the capacity solution, it will tend toward a solution that exploits these high gain channels. It is therefore instructive to understand the physical mechanisms leading to this high channel gain.

### 26.5.2 Superdirectivity

The matrix quantities \( A \) and \( R_i \) can be linked to the phenomenon of array superdirectivity. We assume that all of the elements are identical, allowing us to construct the pattern overlap matrix, which for the transmitter and receiver are

\[
A_T = A / A_{11} \tag{26.90}
\]

\[
A_R = R_i / R_{i,11} \quad \text{with} \quad R_i \text{ computed using } \overline{F}_{i}(\Omega) = 1 \tag{26.91}
\]

The superdirectivity \( Q \) factors [44–46] for the transmit array with a vector of transmit currents \( i_T \) and for the receive array with applied beam-forming weights \( w_R \) are then given as

\[
Q_T = \frac{i_T^* i_T}{i_T^* A_T i_T} \tag{26.92}
\]

\[
Q_R = \frac{w_R^* w_R}{w_R^* A_R w_R} \tag{26.93}
\]
The product of the array $Q$ factor and the quality factor $Q_e$ of the individual array elements approximates the quality factor of the antenna array for the excitation or weight vector [42]. Therefore a high $Q$ factor corresponds to a small usable bandwidth. For example, suppose we use an element that has $Q_e = 10$ when operating in isolation. This corresponds to a 10% frequency bandwidth, something easily obtainable by practical elements such as a half-wave dipole. If the array configuration leads to a modest $Q$ factor of 10, the overall array quality factor will be 100, leading to a frequency bandwidth of only 1%. Therefore the attempt to use superdirectivity to enhance system performance will in most cases fail due to this bandwidth reduction. Since the goal of using MIMO technology is to obtain high spectral efficiency, this severe bandwidth reduction can be considered counterproductive to this fundamental goal.

When the array is used for information communication, the excitation and weighting and therefore the array $Q$ factors are time variant. We can, however, gain insight into the value of the $Q$ factor from $A_T$ or $A_R$ (which depend only on the array properties). Considering the transmit array, we let the EVD of $A_T$ be represented by $A_T = \xi_T A_T \xi_T^\dagger$ [13], where $\xi_T$ is a unitary matrix of eigenvectors and $A_T$ is a diagonal matrix of real eigenvalues (since $A_T$ is Hermitian). Suppose that at one instant in time, the excitation is the $p$th eigenvector (column of $\xi_T$), which can be expressed mathematically as $i_T = \xi_T, p$. The $Q$ factor for this excitation will is $Q_T = 1/\Lambda_T, pp$. The array $Q$ factor is therefore large (indicative of superdirectivity) when the current is aligned with eigenvectors associated with small eigenvalues and reaches a maximum value of $Q_T, \text{max} = 1/\Lambda_T, \text{min}$, where $\Lambda_T, \text{min}$ represents the smallest eigenvalue in $A_T$. Referring to our discussion in Section 26.5.1, where we suggested that the small eigenvalues of $A$ create spatial subchannels with high gain, it becomes clear that this high gain is created by transmit array superdirectivity.

Naturally, an analogous discussion is possible for the receive array. We must recognize, however, that strictly speaking the matrix $R_i$, which accounts for the high gain discussed in Section 26.5.1, is not the same as the matrix $A_R$ unless the average multipath power arriving at the receiver is uniformly distributed in angle. However, the similarity between the definitions of $A_R$ and $R_i$ suggests that the small eigenvalues in $R_i$ leading to high gain are also representative of superdirective effects.

It is important to recognize that this array superdirectivity, while intriguing from a theoretical standpoint, is associated with several practical problems. For example, superdirective excitations are characterized by large antenna current magnitudes (which leads to high ohmic loss), extreme sensitivity to the excitation weights, and narrow operating bandwidth [43]. To gain a feel for the level of superdirectivity that is achievable for compact arrays, consider a uniform circular transmit array of 16 Hertzian dipoles, where the array diameter is $D = \lambda/2$. Figure 26.17 plots the inverse of the eigenvalues of $A_T$ for this array. These results clearly show that the array $Q$ factors can become very large. This motivates identifying a mechanism for limiting array superdirectivity in the capacity solution to enable identification of the maximum achievable capacity given practical constraints.

### 26.5.3 Limiting Superdirectivity: Beam Formers

One method for limiting the influence of superdirectivity on the capacity solution is to use transmit and receive beam formers that disallow superdirective weights [47, 48]. For the following, we define $Q_t$ and $Q_r$ as the highest allowable $Q$ factor for the transmit
and receive arrays, respectively. The matrix $\hat{\xi}_T$ represents the eigenvectors in $\xi_T$ associated with eigenvalues in $A_T$ that are greater than $1/Q_t$. Similarly, $\hat{\xi}_R$ represents the eigenvectors in $\xi_R$ associated with eigenvalues in $A_R$ that are greater than $1/Q_r$.

We can limit the transmit superdirectivity to have a $Q$ factor below $Q_t$ by requiring the excitation $i_T$ to remain within the subspace spanned by $\hat{\xi}_T$. We therefore can modify our representation of the transmit current to (see the discussion leading up to Eq. (26.86))

$$i_T = A^{-1/2} \hat{\xi}_T \hat{i}_T$$

(26.94)

The resulting modification to the power constraint becomes

$$\text{Tr} \left[ R_T A \right] = \text{Tr} \left[ A^{-1/2} \hat{\xi}_T \hat{R}_T \hat{\xi}_T A^{-1/2} \hat{A} \right] = \text{Tr} \left[ \hat{R}_T \right]$$

(26.95)

where we have used $\hat{\xi}_T^\dagger \hat{\xi}_T = I$.

At the receiver, we apply the beam former represented by $\hat{\xi}_R^\dagger$ to create the signal

$$\hat{v}' = \hat{\xi}_R^\dagger \hat{v} = \hat{\xi}_R^\dagger A^{-1/2} \hat{\xi}_T \hat{i}_T + \hat{\xi}_R^\dagger \hat{v}_i$$

(26.96)

where the transformed noise has covariance $\hat{R}_i = \hat{\xi}_R^\dagger \hat{R}_i \hat{\xi}_R$. With this projection, any subsequently applied receive beam-forming weights characterized by a $Q$ factor above $Q_t$ will lie in the null space of $\hat{v}'$ and therefore will not contribute to the capacity. Application of the prewhitening filter gives

$$\hat{v}_0 = \hat{R}_i^{-1/2} \hat{v}' = \hat{R}_i^{-1/2} \hat{\xi}_R^\dagger A^{-1/2} \hat{\xi}_T \hat{i}_T + \hat{R}_i^{-1/2} \hat{\xi}_R^\dagger \hat{v}_i$$

(26.97)
where the noise $\hat{v}_i$ has covariance $I$. The mutual information between the transmit current and the open-circuit received voltage is given by

$$ I(\hat{v}_0, \hat{i}_T) = \log_2 \left| \hat{H} \hat{R}_T \hat{H}^\dagger + I \right| $$

(26.98)

This expression can be maximized subject to the constraint $\text{Tr}[\hat{R}_T] \leq P_T$ to obtain the capacity using the water-filling solution discussed in Section 26.2.4.

We assume that the interference power is uniformly distributed in the horizontal plane or $P_I(\Omega_R) = \sigma_i^2 \delta(\theta_R - \pi/2)$. To specify the channel average SISO SNR, we construct the scalars $A$ and $R_i$ for a single dipole at transmit and receive. The average radiated power is $P_T = |i_T|^2 A$ and the received noise power is given by $R_i$. If the channel transfer function for this SISO case is the scalar $H$, then the received SISO SNR is

$$ \text{SNR}_{\text{isotropic}} = \frac{|H|}{R_i} = \frac{|H|^2 P_T}{A R_i} $$

(26.99)

The value of $\sigma_i^2$ can then be adjusted to produce the desired SISO SNR for each channel realization.

The transmit and receive arrays are the same 16-element circular array of Hertzian dipoles with an array diameter of $D = \lambda/2$, as was considered for Figure 26.17. We compute the capacity averaged over 500 channel realizations of the SVA channel model as a function of the threshold $Q$ factors $Q_t = Q_r$. Figure 26.18 plots this capacity for the isotropic external noise field using the water-filling and uninformed transmitter capacity solutions. The jumps in the capacity occur when the threshold is increased enough to increase the dimensionality of $\hat{\xi}_T = \hat{\xi}_R$. As expected, the water-filling solution, which exploits channel state information at the transmitter, is larger than the capacity for the uninformed transmitter, although the difference at this large SNR of 20 dB is relatively small [14]. This result illustrates the dramatic impact of superdirectivity on the capacity performance.

![Figure 26.18](image)

**Figure 26.18** Capacity (averaged over 500 channel realizations) for a 16-element circular array with diameter $D = \lambda/2$ as a function of $Q_t = Q_r$ for different capacity solutions.
26.5.4 Limiting Superdirectivity: Antenna Loss

While the beam-forming approach discussed in Section 26.5.3 clearly is effective for limiting the level of superdirectivity used by the MIMO system, this solution is not optimal. Specifically, it is possible to form currents from a linear combination of the vectors from the superdirective and nonsuperdirective spaces that achieve an overall $Q$ factor that is below the given threshold. Stated another way, the beam former limits the excitation currents or receive weights to lie in a subspace, while the actual constraint should limit these vectors to an ellipsoid in the multidimensional space. Unfortunately, there does not appear to be an obvious way to achieve the optimal solution using the beam-forming approach combined with the capacity solution.

Because transmit superdirective solutions are characterized by high current magnitudes (for a given radiated power), the loss introduced by even a small antenna resistance makes superdirective excitations inefficient and unfavorable relative to nonsuperdirective ones. At the receiver, ohmic loss leads to spatially white thermal noise that will remove receive superdirective solutions. Because of the difference in the effects at the transmitter and receiver, we develop the impact of antenna loss at each end of the link separately.

26.5.4.1 Transmitter

Incorporating transmit antenna loss as part of the channel implies that the capacity formulation power constraint must limit the power delivered to the transmit array rather than the power radiated, since some of the power is consumed by antenna loss. We formulate the capacity under this delivered constraint in this section and then discuss an approach for compensating for the reduction in radiated power in Section 26.5.4.3. Recognizing that the diagonal elements of $\mathbf{A}$ represent the radiation resistance of each antenna, we can write that the average power delivered to the antenna array:

$$P_{in} = \text{Tr} \left( \mathbf{R} \mathbf{\hat{A}} \right)$$

(26.100)

where $\mathbf{\hat{A}} = \mathbf{A} + \mathbf{L}_T$. The $n$th element of the diagonal matrix $\mathbf{L}_T$ represents one-half the antenna ohmic loss resistance for the $n$th transmit antenna. We emphasize that this is the physical antenna resistance, which can be obtained from radiation efficiency measurements for practical scenarios [49]. For arrays constructed of identical elements, this matrix is $\mathbf{L}_T = \mu_T \mathbf{I}$, where $\mu_T$ is half the loss resistance of each element.

If the delivered power is constrained in the capacity formulation, then $\mathbf{\hat{A}}$ replaces $\mathbf{A}$ in Eq. (26.89). The addition of the diagonal matrix $\mathbf{L}_T$ in $\mathbf{\hat{A}}$ eliminates the very small eigenvalues associated with superdirectivity and therefore regularizes the matrix inverse $\mathbf{\hat{A}}^{-1/2}$ even when the antenna loss is modest. This is a mathematical indication of the fact that superdirective solutions exhibit high loss and become unfavorable relative to more traditional excitations.

For the example computations that follow, all array elements are assumed identical so that $A_{nn}$ is the same for all $n$. We can therefore rearrange $\mathbf{\hat{A}}$ as

$$\mathbf{\hat{A}} = \mathbf{A} + A_{11} \left( 1/\mu_T - 1 \right) \mathbf{I}$$

(26.101)

where $\mu_T = A_{11}/(A_{11} + L_T)$ is the single element efficiency [49]. This allows demonstration of the impact of loss on superdirectivity as a function of this practical efficiency parameter.
26.5.4.2 Receiver  At the receiver, ohmic loss does not explicitly change the possibility of observing receive superdirectivity in the capacity solution since the loss operates identically on the signal and the external interference. However, in this case the resistance adds a thermal noise component to the received signal that must be modeled correctly. Specifically, if the receive array is characterized by a diagonal ohmic loss resistance matrix \( L_R \) (where each matrix element represents half the ohmic loss of the corresponding antenna), then an open-circuit noise voltage vector \( \hat{v}_t \) is introduced so that the received signal becomes

\[
\hat{v} = \mathbf{H} \hat{f} + \hat{v}_i + \hat{v}_t
\]  

(26.102)

Since the noise on each antenna is assumed independent of the noise on all other antennas, the covariance of this noise is [35]

\[
\mathbf{R}_t = 4k_B B T (2 \mathbf{L}_R)
\]  

(26.103)

Given this spatially white thermal noise contribution, which is assumed independent of the external interference, the total interference plus noise has covariance

\[
\mathbf{R}_\eta = \mathbf{R}_i + \mathbf{R}_t
\]  

(26.104)

The addition of the diagonal matrix \( \mathbf{R}_t \) provides the regularization required to avoid receive superdirectivity. Once again, for the computations shown in this chapter, all array elements are assumed identical so that \( \mathbf{R}_{i,mm} \) is the same for all \( m \) and \( \mathbf{L}_R = \mathbf{L}_R\). We can therefore write

\[
\mathbf{R}_\eta = \mathbf{R}_i + \frac{\mathbf{R}_{i,11}}{\text{INR}} \mathbf{I}
\]  

(26.105)

where \( \text{INR} = R_{i,11}/R_{i,11} \) is the interference-to-noise ratio.

26.5.4.3 Capacity  With the antenna ohmic loss now properly included, we can follow the procedure used to obtain Eq. (26.89) to generate our modified transfer relationship

\[
\hat{v}_0 = \mathbf{R}_\eta^{-1/2} \hat{\mathbf{H}} \mathbf{A}^{-1/2} \hat{f} + \mathbf{R}_\eta^{-1/2} (\hat{v}_i + \hat{v}_t)
\]  

(26.106)

The mutual information for this model is given by

\[
I(\hat{v}_0; \hat{f}) = \log_2 \left| \hat{\mathbf{H}} \mathbf{R}_T \hat{\mathbf{H}}^T + \mathbf{I} \right|
\]  

(26.107)

where \( \hat{\mathbf{H}} = \mathbf{R}_\eta^{-1/2} \hat{\mathbf{H}} \mathbf{A}^{-1/2} \). The capacity can be determined subject to the power constraint \( \text{Tr} [\mathbf{R}_{T}] \leq P_T \). Because the antenna loss results in reduced radiated power, we can construct \( \mathbf{R}_{T} \) from the formulation and then scale it by \( \alpha \) so that \( P_{\text{rad}} = \alpha \text{Tr} [\mathbf{R}_T \mathbf{A}] = P_T \). Using this scaled version when evaluating Eq. (26.107) then provides the capacity bound under the delivered power constraint (to suppress transmit superdirectivity) but with the impact of the reduced radiated power removed.

Figure 26.19 shows the water-filling and uninformed transmit capacity as a function of the radiation efficiency of the transmit elements in isolation for an interference-to-noise ratio (INR) of 10 dB and circular array diameter \( D = \lambda/2 \). Most apparent is the fact that the optimal water-filling capacity is somewhat larger than the corresponding value
26.5 SUPERDIRECTIVITY IN MIMO SYSTEMS

Figure 26.19 Capacity (averaged over 500 channel realizations) versus transmit isolated element efficiency for a 16-element circular array with diameter $D = \lambda / 2$ for INR = 10 dB and different capacity solutions.

obtained using the suboptimal uninformed transmitter solution. Otherwise, the curves show similar trends, implying that the same physical phenomena apply to both capacity solutions. These curves reveal that as the antenna efficiency is increased, the capacity increases, with the most dramatic impact occurring around 99% efficiency, where superdirective excitations dominate the solution. It is noteworthy that the 99% threshold efficiency is very difficult to achieve in practice, suggesting that true transmit superdirective behavior would not be observed in a practical system.

Figure 26.20 shows the water-filling capacity as a function of the INR for an isolated transmit element efficiency of 95% and three different circular array diameters (antenna

Figure 26.20 Water-filling capacity (averaged over 500 channel realizations) versus INR for a 16-element circular array for $\mu_T = 95\%$ and different array radii.
spacing reduces with diameter). The horizontal axis is actually expressed as 1/INR to emphasize the dramatic change in capacity as the thermal noise goes to zero (antenna becomes lossless). For small arrays, the close element spacing enables superdirectivity, which accounts for the sharp capacity increase for small antenna loss. The performance of the largest array, on the other hand, is less impacted by the reduced loss since the increased element spacing results in reduced superdirective effects.

26.6 MIMO ANTENNA SYNTHESIS

All of the tools we have discussed represent a framework for analyzing the impact of antennas on the performance of MIMO systems. However, this discussion has not touched on the issue of synthesizing antennas that are appropriate for MIMO communications. Naturally, any concept of optimality will be tied to specific characteristics of the propagation environment, although these characteristics can be specified stochastically to ensure that the final design is appropriate over an ensemble of channels. To explore this concept, we consider the covariance matrix computation specified by Eq. (26.26). An optimal set of antennas for maximizing diversity performance should generate a covariance matrix that satisfies two criteria:

1. $R_{s,m,m}$ large, indicating that each antenna element receives a large amount of signal power.
2. $R_{s,m,p} = 0$ for $m \neq p$, indicating that each antenna element samples the incident signal field in a unique way. In other words, if two antennas sample the field in largely the same way, then there is little diversity offered by the second antenna. With reference to Eq. (26.26), we see that this means that the radiation patterns are orthogonal with respect to the power angular spectrum of the incident field.

These observations indicate that the radiation patterns of the antennas effectively form a basis that should be able to reasonably represent the power angular spectrum of the field. The goal is to design the optimal set of antennas that accomplish these goals.

Simply defining the radiation patterns that optimally accomplish our goals is overly simplistic since the achievable patterns depend on practical issues such as the volumetric aperture in which the antennas reside. We therefore must formulate the problem by incorporating practical constraints to ensure that the antennas provide a reasonable design benchmark against which the performance of practical implementations can be compared. This section offers a framework for accomplishing this optimal antenna synthesis.

26.6.1 Pattern Synthesis

The first step in this formulation is to relate the radiation patterns used in Eq. (26.26) to the physical aperture to which the antennas are restricted. Patterns can be defined by considering either radiating currents (transmit perspective) or the way in which the fields incident on the aperture are sampled and weighted before they are added together (receive perspective), with reciprocity being a mechanism to tie these two perspectives into a single framework. While our discussion on diversity has been focused on receiving incident fields, we define our radiation patterns in terms of radiation currents as this is arguably a more intuitive perspective when considering the antenna synthesis problem.
Therefore consider a volume \( V \) containing an electric current distribution. While we could also consider magnetic current distribution, we ignore such currents for the sake of simplicity. We represent the current distribution as a sum of vector functions \( \mathbf{j}_m(\mathbf{r}) \), with the radiation pattern for the \( m \)th current function being given by

\[
\mathbf{e}_m(\Omega) = \int_V \overline{\mathbf{G}}(\Omega, \mathbf{r}) \cdot \mathbf{j}_m(\mathbf{r}) \, d\mathbf{r}
\] (26.108)

where \( \overline{\mathbf{G}}(\Omega, \mathbf{r}) \) is the dyadic Green’s function relating the currents to the far-field radiation. The goal is therefore to determine the optimal functions \( \mathbf{j}_m(\mathbf{r}) \) that create the desired radiation patterns.

To simplify this synthesis problem, we first represent each current function as a weighted sum of vector basis functions \( \mathbf{f}_n(\mathbf{r}) \), or

\[
\mathbf{j}_m(\mathbf{r}) = \sum_n B_{nm} \mathbf{f}_n(\mathbf{r})
\] (26.109)

where \( B_{nm} \) represents the unknown weighting coefficient for the \( n \)th basis function and the \( m \)th current function. Substitution of this expansion into Eq. (26.108) yields

\[
\mathbf{e}_m(\Omega) = \sum_n B_{nm} \int_V \overline{\mathbf{G}}(\Omega, \mathbf{r}) \mathbf{f}_n(\mathbf{r}) \, d\mathbf{r} = \sum_n B_{nm} \mathbf{z}_n(\Omega)
\] (26.110)

where the function \( \mathbf{z}_n(\Omega) \) physically represents the radiation pattern due to the \( n \)th basis function. Finally, substitution of this result into Eq. (26.26) gives

\[
R_{s,mp} = \sum_n \sum_q B_{nm} \int_V \mathbf{z}_n(\Omega) \cdot \overline{\mathbf{P}}_s(\Omega) \cdot \mathbf{z}_q(\Omega) \, d\Omega \left( \frac{1}{C_{nq}} B_{qp}^* \right)
\] (26.111)

or

\[
\mathbf{R}_s = \mathbf{B}^T \mathbf{C} \mathbf{B}^*
\] (26.112)

Our problem has now been reduced to identifying the coefficients contained in \( \mathbf{B} \) which accomplish our goal. However, we must recognize that even if the current basis functions \( \mathbf{f}_n(\mathbf{r}) \) are normalized, each resulting radiation pattern \( \mathbf{e}_m(\Omega) \) can have a unique total radiated power unless the pattern itself is properly normalized. This normalization is accomplished by constraining each resulting radiation pattern to satisfy

\[
\frac{1}{2\eta_0} \int \mathbf{e}_m(\Omega) \cdot \mathbf{e}_m(\Omega) \, d\Omega = P_{rad}
\] (26.113)

where \( P_{rad} \) is the desired total radiated power for each pattern. If we let the vector \( \mathbf{b}_m \) represent the \( m \)th column of the matrix \( \mathbf{B} \), then using Eq. (26.110) in this constraint leads to \( \mathbf{b}_m^T \mathbf{A} \mathbf{b}_m = P_{rad} \), where

\[
\mathbf{A}_{nq} = \frac{1}{2\eta_0} \int \mathbf{z}_n(\Omega) \cdot \mathbf{z}_q(\Omega) \, d\Omega
\] (26.114)

We must also consider that it is possible for the solution to use superdirective excitations to optimize the radiation patterns to the environment. To avoid this, we use
the approach detailed in Section 26.5.4, wherein antenna loss is introduced to avoid superdirective excitations. To this end, we introduce loss by assuming that the material in which the transmit currents flow is characterized by a conductivity $\sigma_T$. The loss for the $n$th basis function is then

$$P_{\text{loss},n} = \frac{1}{\sigma_T} \int |\mathbf{f}_n(r)|^2 \, dr$$

(26.115)

where the integration is over the antenna aperture. Since $A_{nn}$ represents the radiated power for the $n$th basis function, we can choose $\sigma_T$ such that the radiation efficiency for the $n$th mode $\mu_{T,n} = A_{nn}/(P_{\text{loss},n} + A_{nn})$ achieves a specified value. For many basis functions, a single value of $\sigma_T$ may result in different values of radiation efficiency for each basis function, and therefore the value of $\sigma_T$ can be chosen to set the radiation efficiency for a single basis function, likely the lowest order one, to a specified value.

We then use $\hat{\mathbf{A}} = \mathbf{A} + \mathbf{L}_T$, where $\mathbf{L}_T$ is a diagonal matrix with $L_{T,nn} = P_{\text{loss},n}$.

To satisfy both the supergain and the radiated power constraints, we can parameterize $b_m$ using

$$b_m = P_{\text{rad}}^{-1/2} \mathbf{A}^{-1/2} \hat{b}_m$$

(26.116)

which leads to $\hat{b}_m^* \hat{A} b_m = P_{\text{rad}} \hat{b}_m^* \hat{b}_m$ so that $\hat{b}_m^* \hat{b}_m = 1$. Since this also implies that

$$\mathbf{B} = P_{\text{rad}}^{-1/2} \mathbf{A}^{-1/2} \mathbf{C}$$

(26.117)

our covariance matrix in Eq. (26.112) can be written

$$\mathbf{R}_s = \mathbf{B}^T \frac{P_{\text{rad}} \hat{\mathbf{A}}^{-1/2} \hat{\mathbf{C}} \hat{\mathbf{B}}^*}{\mathbf{C}}$$

(26.118)

We are now prepared to determine the coefficients that generate the optimal radiation patterns according to our criteria. Because $\mathbf{C}$ and therefore $\hat{\mathbf{C}}$ are positive semidefinite and Hermitian, the EVD of $\hat{\mathbf{C}}$ has the structure $\hat{\mathbf{C}} = \mathbf{\xi}_C \hat{\mathbf{A}}_C \mathbf{\xi}_C^*$. It is important to recognize that the number of basis functions may (and generally should) be larger than the number of desired antennas. Therefore if $M$ represents the desired number of antennas, we let $\mathbf{\xi}_M$ contain the columns of $\mathbf{\xi}_C$ corresponding to the $M$ largest eigenvalues in $\hat{\mathbf{A}}_C$. If we therefore choose $\hat{\mathbf{B}} = \mathbf{\xi}_M^*$ then the covariance matrix becomes

$$\mathbf{R}_s = \mathbf{\xi}_M^* \mathbf{\xi}_C \hat{\mathbf{A}}_C \mathbf{\xi}_C^* \mathbf{\xi}_M = \mathbf{A}_M$$

(26.119)

where $\mathbf{A}_M$ represents the $M \times M$ diagonal matrix containing the $M$ dominant eigenvalues in $\hat{\mathbf{A}}_C$. We have therefore satisfied our criteria. Furthermore, since the eigenvectors are orthonormal, we also satisfy the constraint that $\hat{b}_m^* \hat{b}_m = 1$. The final coefficients are then constructed from Eq. (26.117).

### 26.6.2 Computational Example

We restrict ourselves to a two-dimensional scenario where the incident field and the antenna aperture lie in the horizontal plane and a single electromagnetic polarization
is present. The signal power angular spectrum is defined by the truncated Laplacian distribution shown in Figure 26.21. For a rectangular aperture of side lengths $L_x$ and $L_y$ in the x and y dimensions, respectively, the scalar basis functions used for the computation are “Fourier functions” defined by

$$f_n(x, y) = e^{j2\pi(n_xx/L_x + n_yy/L_y)}$$  \hspace{1cm} (26.120)

where $-N_x \leq n_x \leq N_x$, $-N_y \leq n_y \leq N_y$, and a unique value of $n$ is assigned to each unique pair $(n_x, n_y)$. For simplicity in the following computations, $L_x = L_y = 1\lambda$ and $N_x = N_y = 10$. Furthermore, we use the basis function corresponding to $n_x = n_y = 0$ in Eq. (26.115) for computing the value of loss to achieve the specified radiation efficiency.

Figure 26.22 shows the four best current distributions and the resulting radiation patterns for this scenario when the basis function radiation efficiency is set to $\mu_T = 99\%$.
For comparison, a computation is also performed using an array of four filamentary currents placed at the corners of the aperture. In this case, the diversity gain of the optimal antennas is 4.8 dB higher than that of the array of filamentary currents (at the 1% probability level). Figure 26.23 shows the same results when the basis function radiation efficiency is set to $\mu_T = 99.99\%$. In this case, it is interesting that the currents are more concentrated near the edges of the aperture and have much larger magnitudes. Furthermore, the optimal radiation patterns have narrower beams enabled by the array superdirectivity for this low loss scenario. In this case, the diversity gain of the optimal antennas is 6.7 dB higher than that of the array of filamentary currents.

26.7 SUMMARY

This chapter discusses the role of antennas in determining the communication performance of MIMO and diversity systems. After a brief introduction to MIMO system architectures and the signal processing used to enhance communication performance using antenna arrays in multipath propagation environments, the discussion turns to the physical mechanisms through which the antenna radiation characteristics impact the system performance. Issues such as achievable performance with polarization and spatial diversity, antenna mutual coupling, and array superdirectivity are incorporated in the analysis. The chapter concludes with a discussion on the synthesis of optimal antenna patterns for maximizing diversity performance.

REFERENCES


Queries in Chapter 26

Q1. OK to cite Refs. 30-33 here
Q2. AU: capital italic OK?
Q3. AU: a single transmit dipole and a single receive dipole are used
Q4. AU: matrices are
Q5. AU: inverse? see Fig 17 legend
Q6. AU: Inverse of eigenvalues?