Chapter 12

Multiple Input, Multiple Output (MIMO) Communications

Diversity uses multiple antennas to increase the performance of a channel, but the end result of the basic diversity techniques surveyed in the previous chapter is a single input, single output (SISO) channel with greater SNR. More generally, we can transmit different signals from each antenna and process the received signals using multiple detectors to achieve several effective parallel communication channels. Whereas higher SNR improves capacity only logarithmically, multiple parallel channels increases the achievable information rate multiplicatively. This is the motivation for a multiple input, multiple output (MIMO) communication channel.

We will first derive a bound on the capacity of a MIMO channel in terms of the properties of the propagation environment, and then explore specific channel types and methods for exploiting a MIMO channel to achieve capacity that approaches the theoretical limit. We will find that the performance of MIMO is related to the multipath richness of the environment. Being a theoretical framework for antenna diversity systems, MIMO will also be shown to be related to the combining methods of the previous chapter. MIMO also can be considered within a coding theory framework, and the spatial symbols that are transmitted at each array element can be linked to the temporal codes used for channel coding, leading to the concept of space-time codes.

12.1 MIMO Channel Model

For a multiple input, multiple output communication system, both the transmitter and receiver consist of antenna arrays. We will denote the number of transmit antennas as N_t and the number of receive antennas as N_r. The transmitted signals will be characterized in terms of the phasor or complex baseband representation of generator open circuit voltages attached to the input terminals of the transmit antennas. We will arrange these values into the column vector

$$\mathbf{s} = \begin{bmatrix} s_1(n) \\ s_2(n) \\ \vdots \\ s_{N_\mathrm{t}}(n) \end{bmatrix} \tag{12.1}$$

Each element is a complex voltage representing the symbol transmitted at the nth symbol period. At the receiver, we define

$$\mathbf{x} = \begin{bmatrix} x_1(n) \\ x_2(n) \\ \vdots \\ x_{N_\mathrm{r}}(n) \end{bmatrix} \tag{12.2}$$

to be a vector of voltages at the outputs of complex baseband receivers connected to each antenna.

In terms of these quantities, we can represent the propagation channel in matrix form,

$$\mathbf{x}(n) = \mathbf{H}(n)\mathbf{s}(n) + \boldsymbol{\nu}(n) \tag{12.3}$$

where $\mathbf{H}$ relates input currents or voltages at the transmit antenna ports to received voltages and $\boldsymbol{\nu}$, represented by the Greek letter corresponding to n to avoid confusion with the integer time index, is a random process representing external noise, interference, noise due to ohmic losses in the receiving antenna elements, and receiver noise. The noise vector $\mathbf{w}$ in this chapter is not to be confused with the beamformer weight vector in earlier chapters. The index n represents a given symbol period over time.

For a fixed channel, we can view $\mathbf{H}$ as a deterministic matrix with specified values, or for a variable channel we can model the elements of $\mathbf{H}$ as random processes. The transmitted symbols $\mathbf{s}$ can also be modeled as random processes, since the details of a specific data stream sent across the channel is not important in designing the communications systems. In the following, we will not consider dispersion and time-delay effects. Dispersion and time delay nearly always occur in multipath propagation environments, but we will assume that these effects can be dealt with and removed using signal processing.

12.1.1 Network Model for a MIMO Channel

To further understand the channel matrix, we can model a combined transceiver system as a multiport network with N_1 ports at the transmit side and N_2 ports at the receiver. The mutual impedance matrix can be written in block form as

$$\begin{bmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{Z}_{11} & \mathbf{Z}_{12} \\ \mathbf{Z}_{21} & \mathbf{Z}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{i}_1 \\ \mathbf{i}_2 \end{bmatrix} \tag{12.4}$$

$\mathbf{Z}_{11}$ is the mutual impedance matrix of array 1 and $\mathbf{Z}_{22}$ is the mutual impedance matrix of array 2. If array 1 is the transmitter, then the propagation channel is represented by $\mathbf{Z}_{21}$. This matrix relates the vector of input currents $\mathbf{i}_1$ into the ports of array 1 to the open circuit voltages $\mathbf{v}_2$ induced at the ports of array 2. In terms of the mutual impedances, the channel matrix is

$$\mathbf{H} = \mathbf{Q}_2\mathbf{Z}_{21}\mathbf{Q}_1 \tag{12.5}$$

where $\mathbf{Q}_1$ is the system transfer function matrix on the right side of the transmit network relationship (7.6) and $\mathbf{Q}_2$ is defined similarly by (8.5) for array 2 as the receiver. $\mathbf{Q}_1$ relates generator voltages to the transmit element input currents and $\mathbf{Q}_2$ relates the receive antenna open circuit voltages to receiver output voltages. The properties of the propagation environment are characterized by $\mathbf{Z}_{21}$. If the details of the transmitter and receiver electronics are not important, we could redefine $\mathbf{s}$ and $\mathbf{x}$ in (12.3) to be the transmit element input currents and the receiver element open circuit voltages, respectively, and represent the channel directly as $\mathbf{H} = \mathbf{Z}_{21}$.

12.1.2 Free Space Channel

Using the array formulation of Chapter 6, the open circuit voltage at the nth element of array 2 due to radiated fields from array 1 is

$$v_{\mathrm{oc},n} = \frac{4\pi j r_2 e^{jkr_2}}{\omega\mu I_1 I_2} \sum_{m=1}^{N_\mathrm{t}} i_m \overline{E}_m^1(\overline{r}_1) \cdot \overline{E}_n^2(\overline{r}_2) \tag{12.6}$$

where $\overline{E}_n^1$ is the embedded element radiation field pattern with current I_1 into element n of array 1 and the other elements open circuited. $\overline{E}_n^2$ is defined similarly for array 2. From this result, we can see that

$$Z_{21,mn} = \frac{4\pi j r_2 e^{jkr_2}}{\omega\mu I_1 I_2} \overline{E}_m^1(\overline{r}_1) \cdot \overline{E}_n^2(\overline{r}_2) \tag{12.7}$$

The point $\overline{r}_1$ is the location of array 2 in the coordinate system used to find the embedded element patterns of array 1. The point $\overline{r}_2$ has an arbitrary length r_2 and points in the direction of array 1 in the coordinate system of array 2. The decrease in signal strength with the separation between transmitter and receiver is represented by the r_1 dependence of the fields $\overline{E}_m^1(\overline{r}_1)$ radiated by array 1.

In earlier chapters, the embedded element patterns E_n^1 and E_n^2 were defined with the arrays in free space. Equation 12.7 could be modified to accommodate the actual environment by redefining E_n^1 to be the radiation pattern of the nth element of the transmit array in the presence of the multipath propagation environment. Another way to include multipath is to include an angular spectrum transfer function that relates a plane wave propagation direction away from the transmitter to an arrival direction at the receiver, along the lines of (9.40).

12.2 Capacity of the Gaussian MIMO Channel

In order to analyze the capacity of a MIMO system, we need to generalize the Shannon capacity in (10.52) to a multiple input, multiple output channel. The simplest approach would be to consider each transmitter and receiver pair as an independent channel, and multiply the SISO capacity bound by the smaller of N_t and N_r. The problem with this approach is that the received signals are a combination of all the transmitted signals, and the capacity can be increased by using the channels in a cooperative manner. We must take a more sophisticated approach to understand the actual capacity of a MIMO channel.

In the analysis of the capacity of a MIMO channel, we must maximize the mutual information (10.50) between the transmit symbol vector and the receiver outputs over all possible distributions for the signal symbols $\mathbf{x}$. The goal when designing a MIMO system is to select transmit symbol vectors $\mathbf{s}$ somehow to achieve a bit rate as close to the capacity as possible. If we chose a given modulation scheme such as QPSK, the resulting mutual information and bit rate would not be maximum, since a richer symbol constellation with more symbols could better exploit the propagation channel and increase the achievable bit rate. If we model $\mathbf{s}$ as a random process, we have essentially two degrees of freedom: the PDF of each element of $\mathbf{s}$ and the correlation matrix $\mathbf{R_s}$. It can be shown that for an AWGN channel, which has Gaussian distributed noise, the Gaussian distribution for the signal maximizes capacity. In practice, the actual transmitted symbol distribution is not Gaussian, so the realized bit rate does not reach the theoretical capacity that we are deriving here, but the capacity bound will allow us to assess the relative goodness of various types of channels and quantify how close a given modulation scheme comes to achieving the optimum.

Based on this argument, the transmitted symbols at each time step n in this analysis will be modeled as a random vector with zero mean complex Gaussian distribution. The covariance matrix of the transmit symbol vector,

$$\mathbf{R}_s = \mathrm{E}[\mathbf{s}\mathbf{s}^H] \tag{12.8}$$

must also be specified to fully define the Gaussian random vector. The covariance matrix will not be given at this stage of the derivation, but will be treated in Section 12.4. We will also assume that the symbol vector, channel, and noise are all independent.

The capacity is defined in terms of the mutual information as

$$C = \max_{p(\mathbf{s})} \mathrm{E}_H[I(\mathbf{s};\mathbf{x}|\mathbf{H})] \tag{12.9}$$

This expression assumes a fixed realization of the channel matrix $\mathbf{H}$. If the propagation channel is modeled stochastically, we can find the average capacity by taking an expectation over the channel matrix, but for now, we will treat the channel matrix as deterministic. From the definition of mutual information,

$$\begin{aligned} I(\mathbf{s};\mathbf{x}|\mathbf{H}) &= \int\int f_{\mathbf{s},\mathbf{x}|\mathbf{H}}(\mathbf{s},\mathbf{x}|\mathbf{H}) \log_2 \left[\frac{f_{\mathbf{s},\mathbf{x}|\mathbf{H}}(\mathbf{s},\mathbf{x}|\mathbf{H})}{f_{\mathbf{x}|\mathbf{H}}(\mathbf{x}|\mathbf{H}) f_{\mathbf{s}}(\mathbf{s})}\right] d\mathbf{s}\, d\mathbf{x} \\ &= H(\mathbf{x}|\mathbf{H}) - H(\mathbf{x}|\mathbf{s},\mathbf{H}) \\ &= H(\mathbf{x}|\mathbf{H}) - H(\mathbf{H}\mathbf{s}+\boldsymbol{\nu}|\mathbf{s},\mathbf{H}) \\ &= H(\mathbf{x}|\mathbf{H}) - H(\boldsymbol{\nu}) \end{aligned} \tag{12.10}$$

where the last equality holds because only the noise makes the received signal vector $\mathbf{x}$ uncertain given a transmitted signal and channel matrix.

This derivation reduces the problem to the determination of the entropy of a zero mean Gaussian random vector. For a real, zero mean Gaussian random vector, the entropy is

$$\begin{aligned} H(\mathbf{x}) &= -\mathrm{E}[\log_2 f_{\mathbf{x}}(\mathbf{x})] \\ &= -\int \log_2 \left(\frac{e^{-\mathbf{x}^T \mathbf{R}_{\mathbf{x}}^{-1}\mathbf{x}/2}}{\sqrt{(2\pi)^N \det \mathbf{R}_{\mathbf{x}}}}\right) f_{\mathbf{x}}(\mathbf{x})\, d\mathbf{x} \\ &= -\int \tfrac{1}{2}[-\log_2(2\pi)^N - \log_2 \det \mathbf{R}_{\mathbf{x}} - \mathbf{x}^H \mathbf{R}_{\mathbf{x}}^{-1}\mathbf{x} \log_2(e)] f_{\mathbf{x}}(\mathbf{x})\, d\mathbf{x} \\ &= \tfrac{1}{2}[N \log_2 2\pi + \log_2 \det \mathbf{R}_{\mathbf{x}} + N \log_2(e)] \end{aligned} \tag{12.11}$$

The last step follows for the first two terms because they are constants with respect to $\mathbf{x}$ and the PDF integrates to unity. The third term is $\mathrm{E}[\mathbf{x}^H \mathbf{R}_{\mathbf{x}}^{-1}\mathbf{x}] = \sum_{ij} R_{ij}^{-1} R_{ij} = N$. For the circular complex case, the entropy doubles and the factor of one half is removed.

Since the transmitted symbols are complex Gaussian, we can also assume that the received signal vector is also complex Gaussian, and (12.11) can be used to obtain the entropy of $\mathbf{s}$ as well as the noise vector $\boldsymbol{\nu}$, so that

$$\begin{aligned} H(\mathbf{x}|\mathbf{H}) &= N \log_2 2\pi + \log_2 \det \mathbf{R}_{\mathbf{x}} + N \log_2(e) \\ H(\boldsymbol{\nu}) &= N \log_2 2\pi + \log_2 \det \mathbf{R}_{\boldsymbol{\nu}} + N \log_2(e) \end{aligned}$$

The covariance matrix of the received signal in terms of the transmitted symbols and noise is

$$\begin{aligned} \mathbf{R}_{\mathbf{x}} &= \mathrm{E}[\mathbf{x}^H \mathbf{x}] \\ &= \mathrm{E}[(\mathbf{H}\mathbf{s}+\boldsymbol{\nu})^H (\mathbf{H}\mathbf{s}+\boldsymbol{\nu})] \\ &= \mathrm{E}[\mathbf{H}\mathbf{s}\mathbf{s}^H \mathbf{H}^H] + \mathrm{E}[\boldsymbol{\nu}\boldsymbol{\nu}^H] \\ &= \mathbf{H}\,\mathrm{E}[\mathbf{s}\mathbf{s}^H]\mathbf{H}^H + \mathrm{E}[\boldsymbol{\nu}\boldsymbol{\nu}^H] \\ &= \mathbf{H}\mathbf{R}_{\mathbf{s}}\mathbf{H}^H + \mathbf{R}_{\boldsymbol{\nu}} \end{aligned} \tag{12.12}$$

For a given channel matrix, the capacity is

$$\begin{aligned}
C &= \max_{\mathbf{R_s}}[N\log_2(e) + N\log_2 2\pi + \log_2\det\mathbf{R_x} - (N\log_2(e) + N\log_2 2\pi + \log_2\det\mathbf{R}_{\boldsymbol{\nu}})] \\
&= \max_{\mathbf{R_s}}\log_2\frac{\det\mathbf{R_x}}{\det\mathbf{R}_{\boldsymbol{\nu}}} \\
&= \max_{\mathbf{R_s}}\log_2\frac{\det(\mathbf{R}_{\boldsymbol{\nu}} + \mathbf{H}\mathbf{R_s}\mathbf{H}^H)}{\det\mathbf{R}_{\boldsymbol{\nu}}}
\end{aligned} \tag{12.13}$$

The noise covariance matrix must be invertible, or else there is effectively a zero noise channel with infinite capacity. Using the properties of the determinant,

$$C = \max_{\mathbf{R_s}}\log_2\det(\mathbf{I}_{N_\mathrm{r}} + \mathbf{R}_{\boldsymbol{\nu}}^{-1}\mathbf{H}\mathbf{R_s}\mathbf{H}^H) \tag{12.14}$$

This is the "log-det" time average or ergodic capacity bound for a MIMO channel. If the channel is stochastic, the capacity is the expectation of this expression over the channel distribution. Intuitively, the second term inside the determinant is similar to an SNR, because it is the product of the received signal covariance matrix and the inverse of the noise covariance matrix.

12.2.1 Power Constraint

The channel capacity increases with the magnitude of $\mathbf{s}$, which is determined by the strength of the transmitted signals. In order to obtain a finite capacity when maximizing (12.14) over $\mathbf{R_s}$, we must limit the time average transmitted power, which is

$$P_\mathrm{t} = \tfrac{1}{2}\mathrm{E}\left[\mathrm{Re}\{\mathbf{s}^H\mathbf{Q}_1^H\mathbf{Z}_{11}\mathbf{Q}_1\mathbf{s}\}\right] \tag{12.15}$$

Since the real part of the impedance matrix $\mathbf{Z}_{11}$ is determined by the element pattern overlap matrix, this can also be expressed using the overlap matrix according to (7.12).

If the transmit array is uncoupled and the elements are identical and resonant (so that the input reactance is zero), then $\mathrm{Re}\{\mathbf{Z}_{11}\} = R_\mathrm{rad}\mathbf{I}$. If the elements are conjugate matched to the generators and loads, then $\mathbf{Q}_1 = \mathbf{Q}_2 = \frac{1}{2}\mathbf{I}$. In this case, the transmitted power becomes

$$P_\mathrm{t} = \tfrac{1}{8}R_\mathrm{rad}\mathrm{E}[\mathbf{s}^H\mathbf{s}] = \tfrac{1}{8}R_\mathrm{rad}\,\mathrm{tr}\,\mathbf{R_s} \tag{12.16}$$

where tr denotes the matrix trace. So, in the uncoupled case we can express the power constraint in the form

$$\max_{\mathbf{R_s}}\mathrm{tr}\,\mathbf{R_s} \le P_\mathrm{t} \tag{12.17}$$

where we have lumped the scale factors into P_t. This is a trace constraint on the signal correlation matrix. If the array elements are mutually coupled, the trace constraint is only an approximation to the actual transmitted power.

12.3 Special Cases

In order to gain insight into the capacity expression for a Gaussian MIMO channel, we will consider a few special cases for which the capacity bound can be put into a simpler closed form.

12.3.1 Single Input Single Output (SISO) Channel

If $N_\mathrm{r} = N_\mathrm{t} = 1$, then the channel matrix has only one element H_{11}, and the capacity reduces to

$$\begin{aligned} C &= \log_2 \left[\frac{\det(\sigma_\nu^2 + H_{11}\sigma_\mathrm{s}^2 H_{11}^*)}{\det(\sigma_\nu^2)} \right] \\ &= \log_2 \left(\frac{\sigma_\nu^2 + |H_{11}|^2\sigma_\mathrm{s}^2}{\sigma_\nu^2} \right) \\ &= \log_2 \left(1 + \frac{|H_{11}|^2\sigma_\mathrm{s}^2}{\sigma_\nu^2} \right) \end{aligned}$$

where σ_s^2 is the transmitted power and σ_ν^2 is the noise power at the receiver load. The second term inside the logarithm is the SNR at the receiver, so this expression is the Shannon capacity (10.52) expressed in bits/sec/Hz for a single complex channel.

12.3.2 Receive Diversity

In this case, $N_\mathrm{t} = 1$, and the capacity is

$$\begin{aligned} C &= \log_2 \left\{ \frac{\det(\sigma_\nu^2 \mathbf{I} + \mathbf{H}\sigma_\mathrm{s}^2 \mathbf{H}^H)}{\det(\sigma_\nu^2 \mathbf{I})} \right\} \\ &= \log_2 \det \left(\mathbf{I} + \frac{\sigma_\mathrm{s}^2}{\sigma_\nu^2} \mathbf{H}\mathbf{H}^H \right) \end{aligned}$$

Because the channel matrix is a column vector, the rank-one matrix $\mathbf{H}\mathbf{H}^H$ has one eigenvalue equal to $\mathbf{H}^H\mathbf{H}$ and the other eigenvalues are zero. Consequently, the eigenvalues of the matrix under the determinant operation are all unity except for one which is equal to $1 + \mathbf{H}^H\mathbf{H}\sigma_\mathrm{s}^2/\sigma_\nu^2$. Since the determinant is equal to the product of the eigenvalues, the capacity is

$$\begin{aligned} C &= \log_2 \left(1 + \frac{\sigma_\mathrm{s}^2}{\sigma_\nu^2} \mathbf{H}^H\mathbf{H} \right) \\ &= \log_2 \left(1 + \frac{\sigma_\mathrm{s}^2}{\sigma_\nu^2} \sum_{n=1}^{N_\mathrm{r}} |H_{n1}|^2 \right) \end{aligned} \qquad (12.18)$$

The quantity $\sum_n |H_{n1}|^2\sigma_\mathrm{s}^2/\sigma_\nu^2$ is the SNR that is obtained with maximum ratio combining, as should be expected since maximum ratio combining maximizes the SNR and hence also the capacity of a channel with receive diversity.

12.3.3 Transmit Diversity

If $N_\mathrm{r} = 1$, the capacity reduces to

$$C = \log_2 \left(1 + \frac{\sigma_\mathrm{s}^2}{\sigma_\nu^2} \sum_{n=1}^{N_\mathrm{t}} |H_{1n}|^2 \right) \qquad (12.19)$$

which is nearly identical to (12.18), except that the transmit power in the uncoupled approximation is $P_t = \mathrm{tr}\,\mathbf{R}_\mathbf{s} = \sigma_\mathrm{s}^2 N_\mathrm{t}$. If we constrain the total transmit power for systems with transmit and receive diversity to be the same, the SNR for transmit diversity is lower by a factor of N_t.

12.3.4 Unknown Channel

If the channel matrix $\mathbf{H}$ is unknown, then we have no way to choose the transmit symbol vector s to maximize capacity. In this case, one can do no better than to choose transmit symbols such that the elements of s are uncorrelated, so that

$$\mathbf{R}_\mathbf{s} = \sigma_\mathrm{s}^2 \mathbf{I} \tag{12.20}$$

The capacity is

$$C = \log_2 \det \left(\mathbf{I} + \frac{\sigma_\mathrm{s}^2}{\sigma_\nu^2} \mathbf{H}\mathbf{H}^H \right) \tag{12.21}$$

If we set $P_\mathrm{t} = \mathrm{tr}\,\mathbf{R}_s = N_\mathrm{t}\sigma_\mathrm{s}^2$, then this becomes

$$C = \log_2 \det \left(\mathbf{I} + \frac{P_\mathrm{t}}{N_\mathrm{t}\sigma_\nu^2} \mathbf{H}\mathbf{H}^H \right) \tag{12.22}$$

12.4 Known Channel (Water-Filling Solution)

The channel scenarios we have looked at above represent special cases for which the capacity reduces to a simple closed form solution. Although the capacity for these cases is typically lower than that of a full MIMO implementation, these cases do have practical value, since the hardware required for a MIMO system can be expensive, physically large, or otherwise unrealizable for a given communication scenario. Both the transmitter and receiver require multiple antennas and receiver chains, and the system must include a periodic training function in order to estimate the channel matrix $\mathbf{H}$. Because channel bandwidth is a limited resource, and capacity grows only logarithmically as the transmitted power is increases, however, there are powerful motivations to exploit the capacity gain offered by MIMO in a multipath environment. In order to weigh these tradeoffs and engineer an optimal solution, we must understand the ultimate limits on average channel capacity for a full MIMO implementation.

From the perspective of the average capacity bound (12.14), maximizing the capacity requires choosing the transmit symbols $\mathbf{s}$ in an optimal way given a particular channel matrix $\mathbf{H}$. We will still assume that the distribution of symbol voltages for each transmit element is Gaussian, so the goal is to choose the symbols in a cooperative way to achieve the correlation matrix $\mathbf{R_s}$ that maximizes (12.14). To foreshadow the physical interpretation we will arrive at later, we observe that correlations between elements of $\mathbf{s}$ imply a given relative phase and magnitude between the input currents to the transmit antennas, meaning that $\mathbf{s}$ acts like a transmit array beamformer weight vector. If $\mathbf{R_s}$ were only rank one, the channel would be effectively a SISO channel. Because $\mathbf{R_s}$ can have higher rank, in a multipath environment several simultaneous, independent beams from the transmitter to the receiver can be achieved. The goal here is to determine the transmit beamformer weights so that the radiation patterns of the transmitted beams exploit the multipath structure of the propagation channel as represented by the channel matrix in a way that maximizes capacity.

Mathematically, the problem is to determine the capacity

$$C = \max_{\mathbf{R_s}} \log_2 \det \left(\mathbf{I}_{N_\mathrm{r}} + \frac{\mathbf{H}\mathbf{R_s}\mathbf{H}^H}{\sigma_\nu^2} \right) \tag{12.23}$$

subject to a transmitted power constraint $\operatorname{tr} \mathbf{R_s} = P_\mathrm{t}$ and the requirement that $\mathbf{R_s}$ is Hermitian and positive semidefinite. This expression does not have an expectation over $\mathbf{H}$ because we assume that the channel is known and fixed for at least one symbol period. To simplify the treatment, we have assumed that the noise is uncorrelated (IID), although the derivation can be extended to the case of correlated noise. The result can also be generalized to take into account mutual coupling in the transmitted power constraint.

The singular value decomposition (SVD) of the channel matrix is

$$\mathbf{H} = \mathbf{U}\mathbf{S}\mathbf{V}^H \tag{12.24}$$

where $\mathbf{S}$ is an $N_\mathrm{r} \times N_\mathrm{t}$ matrix with singular values on the diagonal and zeros off the diagonal, $\mathbf{U}$ is an $N_\mathrm{r} \times N_\mathrm{r}$ unitary matrix, and $\mathbf{V}$ is an $N_\mathrm{t} \times N_\mathrm{t}$ unitary matrix. Using the SVD in the capacity expression leads to

$$C = \log_2 \det \left(\mathbf{I}_{N_\mathrm{r}} + \sigma_\nu^{-2} \mathbf{U}\mathbf{S}\mathbf{V}^H \mathbf{R_s} \mathbf{V}\mathbf{S}^T\mathbf{U}^H \right) \tag{12.25}$$

Using the identity $\det(AB) = \det A \det B$ and $\det \mathbf{U}\mathbf{U}^H = \det \mathbf{I} = 1$, this becomes

$$C = \log_2 \det \left(\mathbf{I}_{N_\mathrm{r}} + \sigma_\nu^{-2} \mathbf{S}\mathbf{V}^H \mathbf{R_s} \mathbf{V}\mathbf{S}^T \right) \tag{12.26}$$

Using the identity $\det(\mathbf{I}_M + \mathbf{A}\mathbf{B}) = \det(\mathbf{I}_N + \mathbf{B}\mathbf{A})$ for any $M \times N$ matrix A and $N \times M$ matrix B,

$$C = \log_2 \det \left(\mathbf{I}_{N_\mathrm{t}} + \sigma_\nu^{-2} \mathbf{S}^T\mathbf{S} \underbrace{\mathbf{V}^H \mathbf{R_s} \mathbf{V}}_{\tilde{\mathbf{R}}_\mathbf{s}} \right) \tag{12.27}$$

From this point forward, we will work with the transformed signal correlation matrix $\tilde{\mathbf{R}}_\mathbf{s}$. Since $\mathbf{V}$ is unitary, $\operatorname{tr}\mathbf{R}_\mathbf{s} = \operatorname{tr}\tilde{\mathbf{R}}_\mathbf{s}$, so we can apply the trace constraint for transmitted power directly to $\tilde{\mathbf{R}}_\mathbf{s}$.

The Hadamard inequality for a positive semidefinite matrix is

$$\det \mathbf{A} \le \Pi_k A_{kk} \tag{12.28}$$

so that the determinant is bounded above by the product of the diagonal matrix elements. The right and left-hand sides are equal if the matrix is diagonal. Applying this bound to the capacity leads to

$$C = \log_2 \det(\mathbf{I}_{N_\mathrm{r}} + \sigma_\nu^{-2}\mathbf{\Lambda}\tilde{\mathbf{R}}_\mathbf{s}) \le \log_2 \prod_{k=1}^{N}(1 + \sigma_\nu^{-2}\lambda_k \tilde{R}_{\mathrm{s},kk}) \tag{12.29}$$

where λ_k are the diagonal elements of $\mathbf{S}^T\mathbf{S}$, or the squares of the singular values of the channel matrix. From the Hadamard inequality, the maximum value of the capacity is achieved when $\mathbf{R}_\mathbf{s}$ is chosen such that $\tilde{\mathbf{R}}_\mathbf{s}$ is a diagonal matrix.

This step is important, because it indicates that maximum capacity is achieved when the transmitted symbol vector has a special relationship to the unitary matrix $\mathbf{V}$, which is determined by the propagation channel. Since

$$\mathbf{R}_\mathbf{s} = \mathbf{V}\tilde{\mathbf{R}}_\mathbf{s}\mathbf{V}^H \tag{12.30}$$

where $\mathbf{V}$ is a unitary matrix and $\tilde{\mathbf{R}}_\mathbf{s}$ is diagonal, it follows that the columns of $\mathbf{V}$ from the channel matrix singular value decomposition become the eigenvectors of the optimal signal correlation matrix. Later, we will see that this implies a beamformer interpretation for the eigenvectors of the optimal signal correlation matrix.

The above argument fixes the eigenvectors of the signal correlation matrix $\mathbf{R}_\mathbf{s}$. The remaining degrees of freedom in the signal correlation matrix that we need to choose in order to maximize capacity are the eigenvalues, which determine the transmit power assigned to each singular vector of the channel matrix. From $\tilde{\mathbf{R}}_\mathbf{s} = \mathbf{V}^H\mathbf{R}_\mathbf{s}\mathbf{V}$ it follows that the eigenvalues of the optimal signal correlation matrix are the diagonal elements of $\tilde{\mathbf{R}}_\mathbf{s}$. Since $\mathbf{R}_\mathbf{s}$ is a Hermitian, positive semidefinite matrix, the eigenvalues must be real and nonnegative. To simplify the notation, we will denote the diagonal elements of $\tilde{\mathbf{R}}_\mathbf{s}$ as $R_{s,k}$. This leads to

$$\begin{aligned} C &= \log_2 \prod_k (1 + \sigma_\nu^{-2}\lambda_k \tilde{R}_{\mathrm{s},k}) \\ &= \sum_k \log_2(1 + \sigma_\nu^{-2}\lambda_k \tilde{R}_{\mathrm{s},k}) \\ &= \sum_k \log_2 \lambda_k + \sum_k \log_2(1/\lambda_k + \sigma_\nu^{-2}\tilde{R}_{\mathrm{s},k}) \end{aligned} \tag{12.31}$$

The capacity maximization problem is now reduced to the choice of the N real, nonnegative values $\tilde{R}_{\mathrm{s},k}$ subject to the power constraint

$$\sum_k \tilde{R}_{\mathrm{s},k} = P_\mathrm{t} \tag{12.32}$$

This is a classical constrained optimization problem, the solution of which give the power levels associated with each eigenvector of the optimal signal correlation matrix. Once the transformation $\mathbf{V}$ is applied to the signal correlation matrix, the MIMO channel becomes a collection of parallel, uncorrelated channels with different SNR values, and the problem is reduced to finding the best way to allocate the available transmit power to the parallel channels.

The standard approach to solving constrained optimization problems is the method of Lagrange multipliers. At the constrained maximum, the gradient of the function we want to maximize must be parallel to the gradient of the constraint function. If this were not the case, then we could shift the values of the independent variable to move the optimal point slightly along the constraint curve and increase the value of the function to be maximized. This implies that we can choose a scale factor so that the gradient of the function to be maximized, which is the capacity (12.31), and the gradient of the constraint (12.32) add to zero. The scale factor is referred to as the Lagrange multiplier. Combining the capacity and power constraint with the Lagrange multiplier γ leads to the function

$$f = \sum_k \log_2(1/\lambda_k + \sigma_\nu^{-2}\tilde{R}_{\mathrm{s},k}) + \gamma\left(\sum_k \tilde{R}_{\mathrm{s},k} - P_\mathrm{t}\right) \tag{12.33}$$

Based on the argument above, since the gradients of both functions are parallel at the optimal set of values for $\tilde{R}_{\mathrm{s},k}$, there must exist a value for the scalar γ that makes the gradient of f equal to zero.

Computing the gradient of f by taking the derivative with respect to $\tilde{R}_{\mathrm{s},k}$ leads to

$$\frac{\partial f}{\partial \tilde{R}_{\mathrm{s},k}} = \frac{\sigma_\nu^{-2}\lambda_k}{1+\sigma_\nu^{-2}\lambda_k\tilde{R}_{\mathrm{s},k}} + \gamma \tag{12.34}$$

If we set the derivative to zero and solve for $\tilde{R}_{\mathrm{s},k}$, we obtain

$$\tilde{R}_{\mathrm{s},k} = \underbrace{-\frac{1}{\gamma}}_{\alpha} - \frac{\sigma_\nu^2}{\lambda_k} \tag{12.35}$$

We will relabel the Lagrange multiplier as $\alpha = -1/\gamma$ to simplify the notation in later formulas.

Normally, the constrained optimization problem could be easily solved at this point, but there is actually another constraint that we have not accounted for yet. The sum of the values $\tilde{R}_{\mathrm{s},k}$ must equal the power constraint P_t, but the powers associated with each eigenvector of the signal correlation matrix must also be nonnegative, since a negative transmitted power has no useful physical meaning. To enforce the nonnegativity constraint, we choose the unknown eigenvalues using (12.35) for all k such that the expression is positive, and set $\tilde{R}_{\mathrm{s},k} = 0$ otherwise. This leads to the solution

$$\begin{aligned}\tilde{R}_{\mathrm{s},k} &= \max\left\{0, \alpha - \sigma_\nu^2/\lambda_k\right\} \\ &\equiv \left[\alpha - \sigma_\nu^2/\lambda_k\right]^+\end{aligned} \tag{12.36}$$

The Lagrange multiplier α is then determined by the power constraint

$$\begin{aligned}P_\mathrm{t} &= \sum_{k=1}^{N_\mathrm{t}} \left[\alpha - \sigma_\nu^2/\lambda_k\right]^+ \\ &= \sum_{k=1}^{q}(\alpha - \sigma_\nu^2/\lambda_k) \\ &= q\alpha - \sigma_\nu^2 \sum_{k=1}^{q}\frac{1}{\lambda_k}\end{aligned} \tag{12.37}$$

where we have assumed that the eigenvalues λ_k are ordered from largest to smallest, so that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{N_\mathrm{t}}$. The value of the integer q is important. This represents the number of nonzero eigenvalues, and q must be determined along with the eigenvalues themselves. As the transmit power increases, there is more power available to distribute among the eigenvectors of the signal correlation matrix, which means that q increases as well.

We can now solve for the Lagrange multipler,

$$\begin{aligned} \alpha &= \frac{1}{q}\left(\sigma_\nu^2 \sum_{k=1}^{q} \frac{1}{\lambda_k} + P_{\mathrm{t}}\right) \\ &= \frac{\sigma_\nu^2}{q}\left(\mathrm{SNR}_{\mathrm{t}} + \sum_{k=1}^{q} \frac{1}{\lambda_k}\right) \end{aligned} \tag{12.38}$$

where $\mathrm{SNR}_{\mathrm{t}} = P_{\mathrm{t}}/\sigma_\nu^2$ is the ratio of the total transmitted power to the noise power at one receive element. $\mathrm{SNR}_{\mathrm{t}}$ is not a true signal to noise ratio, because the signal and noise powers in the ratio are not computed at the same reference plane, but it is still a useful parameter that quantifies the power at the transmitter and the noisiness of the channel.

A computational procedure for determining the optimal value of q is to try possible values for the number of nonzero eigenvalues, beginning from the largest possible value, N_{t}, and decrementing q by one according to $N_{\mathrm{t}}, N_{\mathrm{t}} - 1, \ldots$, until the largest value of q is found for which the eigenvalues

$$\tilde{R}_{\mathrm{s},k} = \frac{\sigma_\nu^2}{q}\left(\mathrm{SNR}_t + \sum_{r=1}^{q} \frac{1}{\lambda_r}\right) - \frac{\sigma_\nu^2}{\lambda_k} \tag{12.39}$$

are greater than zero for all k from 1 to q. This provides both the optimal value of q and the nonzero signal covariance matrix eigenvalues $\tilde{R}_{\mathrm{s},k}$. The signal covariance can then be determined using (12.30) for $\mathbf{R}$ in terms of the singular vectors $\mathbf{V}$ of the channel matrix.

The channel capacity obtained with this solution is

$$\begin{aligned} C &= \sum_{k=1}^{N_{\mathrm{t}}} \log_2(1 + \lambda_k \tilde{R}_{\mathrm{s},k} \sigma_\nu^{-2}) \\ &= \sum_{k=1}^{q} \log_2[1 + \lambda_k(\alpha/\sigma_\nu^2 - 1/\lambda_k)] \\ &= \sum_{k=1}^{q} \log_2\left\{1 + \lambda_k\left[\frac{1}{q}\left(\sum_{r=1}^{q} \frac{1}{\lambda_r} + \mathrm{SNR}_t\right) - \frac{1}{\lambda_k}\right]\right\} \\ &= \sum_{k=1}^{q} \log_2\left[\frac{\lambda_k}{q}\left(\sum_{r=1}^{q} \frac{1}{\lambda_r} + \mathrm{SNR}_t\right)\right] \end{aligned} \tag{12.40}$$

where $\mathrm{SNR}_t = P_{\mathrm{t}}/\sigma_\nu^2$. Since we have now maximized (12.23) over the transmitter signal covariance, this is the capacity according to the information theoretic definition of the MIMO channel subject to a trace type power constraint with IID white Gaussian noise and known channel matrix.

This optimization procedure is known as the water filling solution for the capacity of a MIMO channel. We can view the optimal solution as dividing the transmitted power between the singular vectors, beginning with the one with the largest singular value, and adding additional power to singular vectors with smaller singular values, until the capacity is maximized and no transmit power remains to apply to singular vectors with the smallest singular values. Visually, the value of α represents the water level, and σ_ν^2/λ_k is the height of the bottom of a water vessel. The power levels $\tilde{R}_{\mathrm{s},k} = \left[\alpha - \sigma_\nu^2/\lambda_k\right]^+$ represent the water depth, or the difference between σ_ν^2/λ_k and the water level α. If σ_ν^2/λ_k is greater than the water level, no power is assigned to that channel. By the constraint, the sum of all the height differences between the bottom of the vessel and the water level equals P_{t}, or the total amount of available water. The value σ_ν^2/λ_k is the inverse SNR of the kth channel. The larger the SNR for the kth channel, the lower the bottom of the vessel for that channel, and the larger the water level or the power assigned to that channel. Low SNR channels have a value for the inverse SNR that is above the water level, and no power is assigned to those channels. The fundamental principle is that parallel channels or multipaths used cooperatively have greater capacity than a single channel, but if some of the multipaths have very low SNR at the receiver for a given transmitted power, higher capacity is achieved if those multipaths are not used and the power is applied to the stronger multipaths with higher SNR.

12.4.1 Eigenmodes and Spatial Coding

The singular vectors of the channel matrix can be viewed as transmit and receive beamformer weight vectors. From the orthogonality of the columns of $\mathbf{U}$ and $\mathbf{V}$, it follows that pairs of transmit and receive singular vectors represent orthogonal communication channels, or eigenchannels. To see this, we can insert the SVD of the channel matrix in (12.3) into the channel model to obtain

$$\mathbf{x} = \mathbf{U}\mathbf{S}\mathbf{V}^H\mathbf{s} + \boldsymbol{\nu} \tag{12.41}$$

Multiplying by $\mathbf{U}^H$,

$$\underbrace{\mathbf{U}^H\mathbf{x}}_{\tilde{\mathbf{x}}} = \mathbf{S}\underbrace{\mathbf{V}^H\mathbf{s}}_{\tilde{\mathbf{s}}} + \mathbf{U}^H\boldsymbol{\nu} \tag{12.42}$$

Since $\mathbf{S}$ is diagonal, in terms of $\tilde{\mathbf{x}}$ and $\tilde{\mathbf{s}}$, the channel consists of $N = \min\{N_\mathrm{r}, N_\mathrm{t}\}$ independent, parallel scalar channels. The signal strength in each of these eigenchannels is determined by $\mathbf{S}_{kk} = \sqrt{\lambda_k}$. If there are fewer than N multipaths, then some of the singular values are zero, and those channels cannot be used to carry information.

The eigenchannels can be used to transmit independent streams of information by judiciously choosing the vector $\mathbf{s}$. From the water filling solution, the optimal signal correlation matrix is $\mathbf{R}_\mathbf{s} = \mathbf{V}^H\tilde{\mathbf{R}}_\mathbf{s}\mathbf{V}$, where $\tilde{\mathbf{R}}_\mathbf{s}$ is a diagonal matrix with diagonal elements given by $\tilde{R}_{\mathrm{s},k}$. The transmit symbol vector which has this correlation matrix is

$$\mathbf{s} = \mathbf{V}\tilde{\mathbf{s}} \tag{12.43}$$

where $\mathrm{E}[\tilde{\mathbf{s}}\tilde{\mathbf{s}}^H] = \tilde{\mathbf{R}}_\mathbf{s}$. From this expression, we can see that for each symbol period, the excitation at the transmitter consists of a linear combination of the columns of $\mathbf{V}$. The kth column can be thought of as a beamformer weight vector. The beamformer weight vector is multiplied by the symbol $\tilde{s}_k$ and scaled so that the power radiated in the beam is equal to $\tilde{R}_{\mathrm{s},k}$. From the water filling solution, the number of beams or columns of $\mathbf{V}^H$ which are actually used and radiate nonzero power is the value of q as determined from (12.39), so the matrix $\mathbf{V}$ could be truncated to q columns and $\tilde{\mathbf{s}}$ shortened to a vector with only q elements. Each of the q beams is used to send a different symbol simultaneously with a given power. On the receive side, each element of $\tilde{\mathbf{x}}$ is formed by combining the receive element output voltages with a beamformer weight vector given by a column of $\mathbf{U}$. The receiver forms q beams, and the outputs of the q beams represent the received symbols $\tilde{\mathbf{x}}$.

Putting all of this together, if we change the channel model so that the channel matrix relates voltages at the inputs of N_t transmit beamforming networks with coefficients from the columns of $\mathbf{V}$ to the voltages at the outputs of N_r receive beamforming networks with coefficients from the columns of $\mathbf{U}$, the resulting channel matrix $\tilde{\mathbf{H}}$ is equal to $\mathbf{S}$. Since this matrix is diagonal, the parallel channels are independent, and capacity is maximized by allocating power to q of the independent channels according to the water filling solution.

This mathematical treatment motivates a method for implementing a MIMO communication system. At each symbol period, the transmitted signals are the sum of the columns of $\mathbf{V}$, weighted so that the average transmit power for each singular vector are given by the water-filling solution, and scaled by the transmit symbol. This could be realized using multiple analog beamforming networks, but in practice this is done in digital signal processing. At the receive side, the array outputs are weighted by several beamforming networks with coefficients taken from the columns of $\mathbf{U}$ (also in DSP), and the output of each receive beamforming network is sent to a decoder.

This provides a physical interpretation for the optimal MIMO spatial channel coding scheme. A time code is the modulation that encodes information over time in each symbol period. The temporal code is then multiplied by a column of $\mathbf{V}$, which represents a set of beamformer weights, or a spatial code. Multiple spatial codes are summed and sent simultaneously with transmit power levels for each spatial code determined by the water filling solution. The received signal is decoded with multiple receive spatial codes, given by columns of $\mathbf{U}$. The next major step in MIMO analysis is to consider the space and time codes together, to produce a space-time coding scheme.

12.5 Channel Matrix Normalization

One difficulty when working with MIMO systems is that capacity depends on the scaling of the channel matrix. If the elements of the channel matrix are large, then the transmit and receive arrays are close together, the propagation paths between the antennas are strong, or the transmitter power is high. In this case, the SNR at the receiver is high. Obstacles between the antennas or high path loss lead to low SNR and reduced capacity. When designing MIMO antennas or coding schemes, however, we do not typically care much about the absolute scaling of the channel matrix. To compare two different systems or codes on the same footing, we would like to scale the channel matrix so that the average SNR at the receive elements is set to a fixed value or can be easily adjusted as a parameter. This can be accomplished by properly normalizing the channel matrix.

Consider a transmit scheme with a constellation consisting of symbols a_k, $k = 1, 2, \ldots, N$ with probability p_k for each symbol. Let the transmit symbol vector $\mathbf{s}$ be a random vector with N_t elements taking on the values a_k. Assuming an uncoupled transmit array, the power radiated by the nth antenna is proportional to

$$\begin{aligned} P_s &= \mathrm{E}[|s_n|^2] \\ &= \sum_{k=1}^{N} p_k |a_k|^2 \end{aligned}$$

The received signal power at the mth receive element is proportional to

$$\begin{aligned} P_{r,m} &= \mathrm{E}[|x_m|^2] \\ &= \mathrm{E}\left[\left|\sum_{n=1}^{N_\mathrm{t}} H_{mn} s_n\right|^2\right] \\ &= \mathrm{E}\left[\sum_{n=1}^{N_\mathrm{t}} H_{mn} s_n \sum_{p=1}^{N_\mathrm{t}} H_{mp}^* s_p^*\right] \\ &= \mathrm{E}\left[\sum_{n=1}^{N_\mathrm{t}} |H_{mn}|^2 |s_n|^2\right] \qquad \text{(Assuming independent transmit signals)} \\ &= \sum_{n=1}^{N_\mathrm{t}} |H_{mn}|^2 \mathrm{E}[|s_n|^2] \\ &= P_s \sum_{n=1}^{N_\mathrm{t}} |H_{mn}|^2 \end{aligned}$$

Assuming an uncoupled receive array, averaging the power over all receive elements leads to

$$P_r = \frac{P_s}{N_\mathrm{r}} \sum_{m=1}^{N_\mathrm{r}} \sum_{n=1}^{N_\mathrm{t}} |H_{mn}|^2 \equiv \frac{P_s}{N_\mathrm{r}} \|\mathbf{H}\|_{\mathrm{Frobenius}}^2 \tag{12.44}$$

where the Frobenius norm is the RMS value of the elements of a matrix. Assuming spatially white noise, the average SNR over the elements of the receiving array is

$$\mathrm{SNR}_r = \frac{P_s}{\sigma_\nu^2 N_\mathrm{r}} \|\mathbf{H}\|_{\mathrm{Frobenius}}^2 \tag{12.45}$$

Using this expression in the log-det capacity formula for the case of an unknown channel with $\mathbf{R}_s = \sigma_\mathrm{s}^2\mathbf{I}$ and $\sigma_\mathrm{s}^2 = P_s$ leads to

$$\begin{aligned} C &= \log_2 \det \left[\mathbf{I} + \frac{\mathrm{SNR}_\mathrm{r} N_\mathrm{r}}{\|\mathbf{H}\|^2_\mathrm{Frobenius}} \mathbf{H}\mathbf{H}^H\right] \\ &= \log_2 \det \left[\mathbf{I} + \mathrm{SNR}_\mathrm{r} \mathbf{H}_n \mathbf{H}_n^H\right] \end{aligned} \quad (12.46)$$

where

$$\mathbf{H}_n = \frac{\sqrt{N_\mathrm{r}}}{\|\mathbf{H}\|_\mathrm{Frobenius}} \mathbf{H} \quad (12.47)$$

This is a convenient normalization if we want to vary the receive SNR as a parameter in a model rather than the transmit power. If the transmit symbols and receiver noise are correlated, then the normalization is only approximate, but can still be used to quantify the rough trends of performance as a function of SNR for different MIMO channels on an equal footing.

12.6 Space-Time Coding

With a SISO channel, we can use temporal coding to achieve high capacity and low bit error rates. With a MIMO channel, there is an additional dimension in the system, represented by the magnitudes and phases of the symbol vector across the transmitting array, as well as the magnitudes and phases of the receiving beamformer weight vectors used to separate the eigenchannels of the transmit and receive arrays and propagation environment, or the spatial codes. Combining the temporal and spatial codes into a unified treatment leads to the concept of space-time coding.

There are two major types of space-time codes: space-time trellis codes, which transmit symbols in serial using codes designed for multiple transmitters, and space-time block codes, which transmit a block of symbols in parallel using a transmission matrix to map the source symbol stream into transmit signals. We will consider in the next section a simple example of a block code.

12.6.1 Alamouti Code

The Alamouti code is a block code for a MIMO system with two transmitters and one receiver. The basic scheme is to transmit the symbols $[s_1 \;\; s_2]$ at time t and $[-s_2^* \;\; s_1^*]$ at the next symbol time, $t + T$. The transmission matrix is

$$\mathbf{S} = \begin{bmatrix} s_1 & s_2 \\ -s_2^* & s_1^* \end{bmatrix} \quad (12.48)$$

The first row is the pair of symbols to be transmitted at a given symbol period, and the second row is transmitted at the next symbol period. This matrix has the property that

$$\mathbf{S}\mathbf{S}^H = (|s_1|^2 + |s_2|^2)\mathbf{I} \quad (12.49)$$

which is an orthogonality condition for the matrix columns. The Alamouti code transmits two symbols in two time slots, so that the coding rate is one. Unlike other more complicated space-time codes, it is linear with respect to the transmit symbols, which simplifies analysis of the signaling scheme.

To analyze this code, we will assume that the variation of the propagation environment is slow enough that the channel does not change significantly from one symbol time to the next. The channel matrix for a 1 $\times$ 2 MIMO system is a row vector of the form $\mathbf{H} = [H_1 \;\; H_2]$. The received signals at two adjacent symbol times are

$$\begin{aligned} x_1 &= H_1 s_1 + H_2 s_2 + w_1 \\ x_2 &= H_1(-s_2^*) + H_2 s_1^* + w_2 \end{aligned}$$

where w_n represents samples of the noise voltage at the receiver. If we conjugate the received signal at the second symbol time and arrange the pair of received symbols into a vector, this becomes

$$\begin{bmatrix} x_1 \\ x_2^* \end{bmatrix} = \underbrace{\begin{bmatrix} H_1 & H_2 \\ H_2^* & -H_1^* \end{bmatrix}}_{\mathbf{A}} \begin{bmatrix} s_1 \\ s_2 \end{bmatrix} + \begin{bmatrix} w_1 \\ w_2^* \end{bmatrix} \tag{12.50}$$

The matrix $\mathbf{A}$ satisfies the same orthogonality property as $\mathbf{S}$. If we multiply both sides by $\mathbf{A}^H$, we obtain

$$\mathbf{y} = \mathbf{A}^H \mathbf{A} \mathbf{s} + \mathbf{A}^H \boldsymbol{\nu} = (|H_1|^2 + |H_2|^2)\mathbf{s} + \mathbf{A}^H \boldsymbol{\nu} \tag{12.51}$$

where we are using the notation $\mathbf{s}$ with a slight difference as compared to (12.3), because the elements of the vector represent adjacent symbols in time from the source rather than the index of the transmit element. The factor $|H_1|^2 + |H_2|^2$ is small only if both channel paths are small, so that Alamouti coding provides a two-level diversity gain. In order to use this detection scheme, the matrix $\mathbf{A}$ must be determined using a channel estimation algorithm.

The Alamouti code is optimal in that it achieves the log-det capacity formula for the case of channel unknown at the transmitter. The signal correlation matrix is

$$\mathbf{R}_\mathrm{s} = \frac{1}{2}\mathrm{E}\left\{ \begin{bmatrix} s_1 \\ s_2 \end{bmatrix} \begin{bmatrix} s_1^* & s_2^* \end{bmatrix} + \begin{bmatrix} -s_2^* \\ s_1^* \end{bmatrix} \begin{bmatrix} -s_2 & s_1 \end{bmatrix} \right\} = \sigma_\mathrm{s}^2 \mathbf{I} \tag{12.52}$$

where $\sigma_\mathrm{s}^2 = \mathrm{E}[|s_1|^2 + |s_2|^2]/2$. From (12.14), the log-det capacity formula for this signal correlation matrix is

$$C = \log_2 \left(1 + \frac{\sigma_\mathrm{s}^2}{\sigma_\nu^2} \mathbf{H}\mathbf{H}^H \right) \tag{12.53}$$

which is identical to (12.19).

While the Alamouti code (with known channel matrix) achieves the unknown-channel Shannon capacity bound, it does not achieve the known-channel capacity. The known-channel channel capacity bound from (12.40) is

$$C = \log_2 \left(1 + \frac{P_\mathrm{t}}{\sigma_\nu^2} \mathbf{H}\mathbf{H}^H \right) \tag{12.54}$$

Since $P_\mathrm{t} = \mathrm{tr}\, \sigma_\mathrm{s}^2 \mathbf{I} = 2\sigma_\mathrm{s}^2$ for the Alamouti code, the effective SNR in (12.53) is smaller by a factor of two than in (12.54), and the realized bit rate with Alamouti coding is smaller than the channel capacity. Also, as observed earlier receive diversity with $N_\mathrm{r} = 2$ and $N_\mathrm{t} = 1$ provides a lower bit error rate for a given total transmitted power. If the design goal is a simple receiver, then the Alamouti approach may be desirable since only one receive antenna is required.

The important properties of the Alamouti space-time block code are that it is rate one, since one symbol is transmitted per use of the channel, the processing at the receiver is a simple linear operation requiring little computation, and it is open-loop, since only the receiver needs to know the channel matrix. With differential encoding of the temporal transmitted bit sequence, the channel matrix estimation step can be eliminated as well.

The Alamouti code can be extended to a 2×2 MIMO system by expanding the matrix relationship developed above to a four by four matrix that relates a four element vector of symbols at both transmit elements and two symbol times to the received symbols. The 2×2 Alamouti code is not optimal, however, and it has been shown that there does not exist a rate-one STBC for more transmit and receive antennas. The Alamouti code is implemented in the IEEE 802.11n WiFi standard as well as other communication protocols.

12.7 Higher Order MIMO Architectures

The Alamouti code has been generalized to a theory of complex orthogonal space-time block codes which extend its mathematical elegance and optimal performance to systems with more than two antennas. The basic property of these signaling schemes is that blocks of symbols are converted to transmit symbol vectors using a transmission matrix with the number of rows given by the block length (time dimension) and the number of columns is the number of transmit elements N_t (space dimension). These codes require accurate estimates of the channel matrix. Differential space-time block codes can be used for systems that do not perform channel estimation.

Another family of MIMO architectures that offers low computational complexity is known as Bell Laboratories Layered Space-Time (BLAST). BLAST uses standard forward error correction codes on each MIMO branch and cyclic stepping of symbols across the array to achieve capacity approaching that of optimal space-time codes.

12.8 Narrowband Channel Estimation

For optimal MIMO performance, the channel matrix must be known at the transmitter to enable the system to exploit the multipath structure of the environment. Operationally, to obtain the water filling capacity, the transmitter and receiver must know the singular vectors of the channel matrix, and the transmitter must know the singular values in order to allocate power optimally to the eigenchannels. Here, we will consider a basic algorithm for channel state estimation.

If we transmit N sets of N_{t} known training symbols, the received signal is

$$\left[\begin{array}{c:c:c:c} \mathbf{x}(1) & \mathbf{x}(2) & \cdots & \mathbf{x}(N) \end{array}\right] = \mathbf{H}\left[\begin{array}{c:c:c:c} \mathbf{s}(1) & \mathbf{s}(2) & \cdots & \mathbf{s}(N) \end{array}\right] + \left[\begin{array}{c:c:c:c} \boldsymbol{\nu}(1) & \boldsymbol{\nu}(2) & \cdots & \boldsymbol{\nu}(N) \end{array}\right] \tag{12.55}$$

We can write this in the form

$$\mathbf{X} = \mathbf{H}\mathbf{S} + \mathbf{N} \tag{12.56}$$

The goal is to design $\mathbf{S}$ so that this relationship can be solved for the channel matrix, and to find an effect solution procedure.

Since $\mathbf{S}$ is typically not square, it does not have an inverse, but it does have a pseudo-inverse $\mathbf{S}^+$. Multiplying by $\mathbf{S}^+$ from the left leads to

$$\mathbf{X}\mathbf{S}^+ = \mathbf{H}\mathbf{S}\mathbf{S}^+ + \mathbf{N}\mathbf{S}^+ \tag{12.57}$$

There are various types of operator pseudo-inverses that could be used in (12.57). If $\mathbf{S}$ has full row rank, then the most common pseudo-inverse,

$$\mathbf{S}^+ = \mathbf{S}^H(\mathbf{S}\mathbf{S}^H)^{-1} \tag{12.58}$$

can be used. If we send enough training symbol vectors, the product $\mathbf{S}\mathbf{S}^+$ is close to the identity. Assuming a large SNR for the received training data, the channel matrix can be approximated as

$$\mathbf{H} \simeq \mathbf{X}\mathbf{S}^+ \tag{12.59}$$

This provides a way to estimate the channel matrix using a training symbol sequence.

The downsides of channel estimation is that data transmission must be interrupted to send the training sequence and to return the computed channel state information to the transmitter. A good deal of research has been done on effective methods for channel estimation, including methods for efficiently updating an estimate of a slowly varying channel and efficient ways to compress channel state information for transmission from the receiver to the transmitter.

12.9 Multipath Richness

When the focus of the analysis is on the propagation environment, it is helpful to have a measure of the amount of multipath in the environment. In free space, a MIMO system improves capacity simply by using the arrays to direct transmit power towards the receiver and reduce the receive array response to noise from other directions. Only one singular vector is used on the transmit and receive sides for beamforming, and the channel is effectively SISO. To achieve higher capacity, the propagation environment must have multiple scattering paths. For a single propagation path, the channel capacity increases as the log of SNR, whereas for a multipath environment, the channel capacity increases at a more rapid rate with SNR as more eigenchannels are energized in the waterfilling solution. Consequently, the multipath richness of a propagation environment can be quantified by how the bit error rate or channel capacity scales with SNR. We will discuss in this section two of the parameters that can be used to quantify multipath richness.

12.9.1 Diversity Order

The diversity order of a MIMO link is the asymptotic log-log slope of BER versus SNR in the high SNR limit:

$$\mathrm{DO} = -\lim_{\mathrm{SNR}\to\infty} \frac{\log[\mathrm{BER}(\mathrm{SNR})]}{\log(\mathrm{SNR})} \tag{12.60}$$

For a Rayleigh fading environment, the diversity order of maximum ratio combining is N_r, and the order for Alamouti coding is two. If the multipath environment is insufficiently rich, then the diversity order of a given communications link is smaller than that obtained with the Rayleigh model.

12.9.2 Effective Degrees of Freedom

Another measure of multipath richness is the effective degrees of freedom, which is defined to be

$$\mathrm{EDOF} = \left.\frac{\partial}{\partial\delta} C(2^\delta \rho)\right|_{\delta=0} \tag{12.61}$$

where the argument of the capacity is SNR. This is the slope of capacity on a linear scale relative to the SNR on a logarithmic scale.

We can gain some insight into EDOF using the uninformed channel capacity, which can be expressed using the singular values of the channel matrix as

$$C = \sum_{k=1}^{N} \log_2\left(1 + \frac{\rho}{N_\mathrm{t}}\sigma_k^2\right) \tag{12.62}$$

where $N = \min\{N_\mathrm{t}, N_\mathrm{r}\}$, ρ is the ratio of transmit power to receiver noise, and σ_k are the singular values. The effective degrees of freedom is

$$\begin{aligned}\mathrm{EDOF} &= \left.\frac{\partial}{\partial\delta}\sum_{k=1}^{N}\log_2\left(1+\frac{2^\delta\rho}{N_\mathrm{t}}\sigma_k^2\right)\right|_{\delta=0} \\ &= \left.\sum_{k=1}^{N}\frac{2^\delta\rho\sigma_k^2/N_\mathrm{t}}{1+2^\delta\rho\sigma_k^2/N_\mathrm{t}}\right|_{\delta=0} \\ &= \sum_{k=1}^{N}\frac{1}{1+N_\mathrm{t}/(\rho\sigma_k^2)}\end{aligned} \tag{12.63}$$

This measure of channel richness is independent of all system details except SNR, which essentially sets a threshold below which a given eigenchannel does not contribute to the capacity.

For a line of sight (LOS) channel, $\sigma_1 = 1$, $\sigma_2 = \sigma_3 = \cdots = 0$, and

$$\mathrm{EDOF} = \frac{1}{1+N_\mathrm{t}/\rho} \tag{12.64}$$

which goes to one as the SNR becomes large. If all eigenchannels have equal strength, then

$$\mathrm{EDOF} = \frac{N}{1+N_\mathrm{t}/\rho} \to N, \quad \rho\to\infty \tag{12.65}$$

If a rich channel changes to LOS, in order to maintain capacity we must increase the transmit power to raise the single channel capacity and compensate for the loss of channel degrees of freedom.