Chapter 8 Space-Time Coded-Modulation for Multiple Antenna Systems (/for MIMO Channels)

The advent of multiple-input–multiple-output (MIMO) space-time coded wireless systems has recently emerged as one of the most significant technical breakthroughs in modern communications. The research on MIMO systems, including the study of channel capacity and the design of communication schemes, demonstrates that MIMO systems have a potential to significantly increase spectral efficiency in wireless communications (without the requirement of increasing bandwidth and transmit power).

Note: The materials presented in Sections 8.3 and 8.6.3 are based largely on B. Vucetic's lecture notes.

8.1 Introduction

An MIMO system is an arbitrary wireless communication system in which the transmitter as well as the receiver is equipped with multiple antenna elements. A core idea in MIMO systems is space-time signal processing in which time is complemented with the spatial dimension inherent in the use of multiple spatially distributed antennas. MIMO effectively takes advantage of random fading and multipath delay spread for multiplying transfer rates. 图 8.1 给出了一个有 N 根发送天线、M 根接收天线的 MIMO 系统框图。下图是采用不同天线配置时的通信系统。

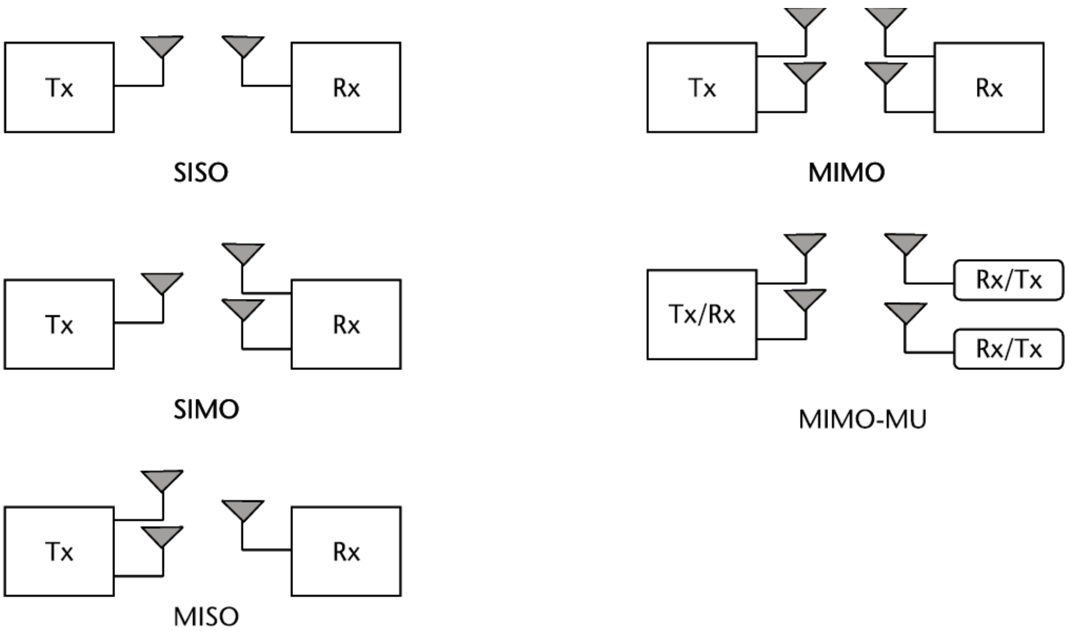


Figure 8.0 Different antenna configurations in space-time systems.

Winter、Foschini 和 Telatar 的开创性工作证明，多天线阵技术可极大的提高系统信道容量。若接收端可准确估计信道信息，并保证不同发射-接收天线对之间的信道衰落系数相互独立，则一个拥有 N 根发射天线和 M 根接收天线的 MIMO 系统的信道容量将随 min{M, N}线性增加。因此，信噪比、发射功率和信道带宽都相同时，多天线系统可提供的信道容量是单天线系统的 min(M, N)倍。在频率资源日益紧张的今天，多天线信道容量理论无疑给解决高速无线通信问题开辟了一条新思路，基于此发展起来的信道编码和信号处理技术正越来越受到人们的关注。

8.2 MIMO System Model

Consider a single-user MIMO communication system with N transmit and M receive antennas. (It will be called a (N, M) system.) The system block diagram is shown in Fig.8.1. The transmitted signal at time t is represented by an N×1 column vector $\mathbf{x} \in \mathbb{C}^N$, and the received signal is represented by an M×1 column vector $\mathbf{y} \in \mathbb{C}^M$ (For simplicity, we ignore the time index). The discrete-time MIMO channel can be described by

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n} \tag{8.1}$$

where $\mathbf{H}$ is an M×N complex matrix describing the channel and the element h_{ij} of $\mathbf{H}$ represents the channel gain from the transmit antenna j to the receive antenna i; and $\mathbf{n} \sim \mathcal{CN}\left(\mathbf{0},\ N_0\mathbf{I}_M\right)$ is a zero-mean complex Gaussian noise vector whose components are i.i.d. circularly symmetric complex Gaussian variables. The covariance matrix of the noise is given by $\mathbf{K}_n \equiv E[\mathbf{n}\mathbf{n}^H] = N_0\mathbf{I}_M = 2\sigma^2\mathbf{I}_M$, i.e., each of M receive antennas has identical noise power of N_0 (per complex dimension) (or, σ^2 per real dimension). The total transmitted power is constrained to P, regardless of the number of transmit antennas N. It can be represented as

$$\mathsf{E}\left[\|\mathbf{x}\|^2\right] = \mathsf{E}\left[\mathbf{x}^H\mathbf{x}\right] = \mathsf{E}\left[Tr(\mathbf{x}\mathbf{x}^H)\right] = Tr\left\{\mathsf{E}\left[\mathbf{x}\mathbf{x}^H\right]\right\} = \mathrm{tr}(\mathsf{K}_x) \le P\ ,$$

where $\mathbf{K}_x = \mathsf{E}\left[\mathbf{x}\mathbf{x}^H\right]$ is the covariance matrix of the transmitted signal $\mathbf{x}$. Furthermore, if the channel is unknown to the transmitter, we assume that the signals transmitted from individual antennas have equal powers of P/N. It means that

$$\mathbf{K}_x = \frac{P}{N}\mathbf{I}_N \tag{8.2}$$

The received signal covariance matrix, defined as $\mathbf{K}_y = \mathsf{E}\left[\mathbf{y}\mathbf{y}^H\right]$, is given by

$$\mathbf{K}_y \equiv \mathsf{E}\left[\mathbf{y}\mathbf{y}^H\right] = \mathbf{H}\mathbf{K}_x\mathbf{H}^H + N_0\mathbf{I}_M$$

Denoting by P_r the average signal power at the output of each receive antenna, the average

SNR at each receive antenna is given by

$$\text{SNR} = \gamma = \frac{P_r}{N_0}$$

which is independent of *N*. The total received signal power can be expressed as $\text{Tr}(\mathbf{K}_y)$.

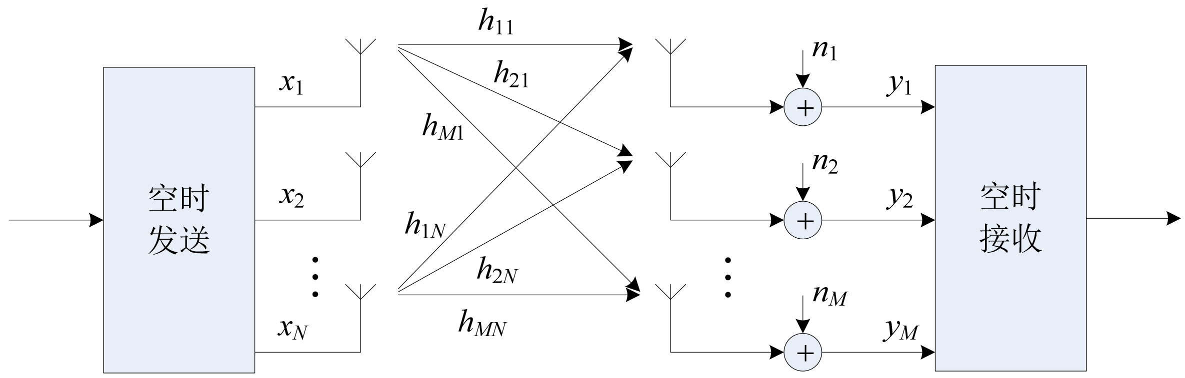


Fig. 8.1 An MIMO wireless system model

We will consider several possible scenarios for the matrix **H**:

1. **H** is deterministic.
2. **H** is a random matrix. Its entries are selected according a probability distribution at the beginning of each symbol interval *T* and are kept constant during one symbol interval. In other words, each use of the channel corresponding to an independent realization of **H**. Such a channel is called a *fast* (or independent) *fading channel*.
3. **H** is a random matrix. Its entries change randomly and are kept constant during a fixed number of symbol intervals, much shorter than a transmission block. Such a channel is called a *block fading channel*.
4. **H** is a random matrix but is fixed at the start of a transmission block and kept constant during a transmission block. Such a channel is called a *slow* or *quasi-static fading channel*.

For normalization purposes, we assume that the received power for each of *M* receive branches is equal to the total transmitted power. Thus, in the case when **H** is deterministic, we have

$$\sum_{n=1}^{N} |h_{mn}|^2 = N, \quad m = 1, 2, ..., M$$

When **H** is random, we will assume that its entries are i.i.d. zero-mean complex Gaussian variables, each with variance 1/2 per dimension. This case is usually referred to as a *rich scattering environment*. The normalization constraint for the elements of **H** is given by

$$\sum_{n=1}^{N} \mathrm{E}\left[|h_{mn}|^2 \right] = N, \quad m = 1, 2, ..., M$$

With the normalization constraint, the total received signal power per antenna is equal to the total transmitted power, and the average SNR at any receive antenna is $\gamma = P / N_0$.

In all cases, we will assume that the channel matrix is known to the receiver (i.e., perfect CSIR), equivalently, the channel output consists of the pair (**y**, **H**), and the distribution of **H** is known at the transmitter. In most situations, the realization of **H** (CSI) is assumed to be not known at the transmitter.

8.3 Fundamental Capacity Limits of MIMO Channels

Consider the case of deterministic **H**. The channel matrix **H** is assumed to be constant at all time and known to the receiver. The relation of (8.1) indicates a vector Gaussian channel. The Shannon capacity is defined as the maximum data rate that can be transmitted over the channel with arbitrarily small error probability. It is given in terms of the mutual information between vectors **x** and **y** as

$$C(\mathbf{H}) = \max_{p(\mathbf{x}):\mathsf{E}[\|\mathbf{x}\|^2]\le P} \mathcal{I}(\mathbf{x};\mathbf{y},\mathbf{H}) = \max_{p(x)} \mathcal{I}(\mathbf{x};\mathbf{H}) + \mathcal{I}(\mathbf{x};\mathbf{y}\,|\,\mathbf{H})$$

$$= \max_{p(x)} \mathcal{I}(\mathbf{x};\mathbf{y}\,|\,\mathbf{H}) = \max_{p(x)} \left[\mathcal{H}(\mathbf{y}\,|\,\mathbf{H}) - \mathcal{H}(\mathbf{y}\,|\,\mathbf{x},\mathbf{H})\right] \tag{8.3}$$

where $p(\mathbf{x})$ is the probability distribution of the vector **x**, $\mathcal{H}(\mathbf{y}\,|\,\mathbf{H})$ and $\mathcal{H}(\mathbf{y}\,|\,\mathbf{x},\mathbf{H})$ are the differential entropy and the conditional differential entropy of the vector **y**, respectively. Since the vectors **x** and **n** are independent, we have

$$\mathcal{H}(\mathbf{y}\,|\,\mathbf{x},\mathbf{H}) = \mathcal{H}(\mathbf{n}) = \log_2\left(\det(\pi e N_0 \mathbf{I}_M)\right)$$

which has fixed value and is independent of the channel input. Thus, maximizing the mutual information $\mathcal{I}(\mathbf{x};\mathbf{y}\,|\,\mathbf{H})$ is equivalent to maximize $\mathcal{H}(\mathbf{y}\,|\,\mathbf{H})$. From (8.1), the covariance matrix of **y** is

$$\mathbf{K}_y = \mathsf{E}\left[\mathbf{y}\mathbf{y}^H\right] = \mathbf{H}\mathsf{K}_x\mathbf{H}^H + N_0\mathbf{I}_M$$

Among all vectors **y** with a given covariance matrix K_y, the differential entropy $\mathcal{H}(\mathbf{y})$ is maximized when **y** is a zero-mean circularly symmetric complex Gaussian (ZMCSCG) random vector [Telatar99]. This implies that the input **x** must also be ZMCSCG, and therefore this is the optimal distribution on **x**. This yields the entropy $\mathcal{H}(\mathbf{y}\,|\,\mathbf{H})$ given by

$$\mathcal{H}(\mathbf{y}\,|\,\mathbf{H}) = \log_2\left(\det(\pi e \mathsf{K}_y)\right)$$

The mutual information then reduces to

$$\mathcal{I}(\mathbf{x};\mathbf{y}\,|\,\mathbf{H}) = \mathcal{H}(\mathbf{y}\,|\,\mathbf{H}) - \mathcal{H}(\mathbf{n})$$

$$= \log_2\left(\det\left(\mathbf{I}_M + \frac{1}{N_0}\mathbf{H}\mathsf{K}_x\mathbf{H}^H\right)\right) \tag{8.4}$$

where we have used the fact that $\det(\mathbf{AB}) = \det(\mathbf{A})\det(\mathbf{B})$ and $\det(\mathbf{A}^{-1}) = [\det(\mathbf{A})]^{-1}$. And the MIMO capacity is given by maximizing the mutual information (8.4) over all input covariance matrces K_x satisfying the power constraint:

$$C(\mathbf{H}) = \max_{\mathsf{K}_x : Tr(\mathsf{K}_x)=P} \log_2 \left(\det \left(\mathbf{I}_M + \frac{1}{N_0} \mathbf{H}\mathsf{K}_x \mathbf{H}^H \right) \right) \quad \text{bits per channel use} \qquad (8.5)$$

$$= \max_{\mathsf{K}_x : Tr(\mathsf{K}_x)=P} \log_2 \left(\det \left(\mathbf{I}_N + \frac{1}{N_0} \mathsf{K}_x \mathbf{H}^H \mathbf{H} \right) \right)$$

where the last equality follows from the fact that $\det(\mathbf{I}_m + \mathbf{AB}) = \det(\mathbf{I}_n + \mathbf{BA})$ for matrices $\mathbf{A}$ $(m \times n)$ and $\mathbf{B}$ $(n \times m)$.

Clearly, the optimization relative to K_x will depend on whether or not **H** is known at the transmitter. We now discuss this maximizing under different assumptions about transmitter CSI by decomposing the vector channel into a set of parallel, independent scalar Gaussian sub-channels.

8.3.1 Parallel Decomposition of the MIMO Channel

By the singular value decomposition (SVD) theorem, any $M \times N$ matrix $\mathbf{H} \in \mathbb{C}^{M \times N}$ can be written as

$$\mathbf{H} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^H \qquad (8.6)$$

where $\mathbf{\Lambda}$ is an $M \times N$ non-negative real and diagonal matrix, **U** and **V** are $M \times M$ and $N \times N$ unitary matrices, respectively. That is, $\mathbf{U}\mathbf{U}^H = \mathbf{I}_M$ and $\mathbf{V}\mathbf{V}^H = \mathbf{I}_N$, where the superscript "H" stands for the Hermitian transpose (or complex conjugate transpose). In fact, the diagonal entries of $\mathbf{\Lambda}$ are the non-negative square roots of the eigenvalues of matrix $\mathbf{HH}^H$, the columns of **U** are the eigenvectors of $\mathbf{HH}^H$ and the columns of **V** are the eigenvectors of $\mathbf{H}^H\mathbf{H}$.

Denote by λ the *eigenvalues* of $\mathbf{HH}^H$, which are defined by

$$\mathbf{HH}^H \mathbf{z} = \lambda \mathbf{z}, \quad \mathbf{z} \neq \mathbf{0} \qquad (8.7)$$

where **z** is an $M \times 1$ eigenvector corresponding to λ. The number of non-zero eigenvalues of matrix $\mathbf{HH}^H$ is equal to the rank of **H**. Let r be the rank of the matrix **H**. Since the rank of **H** cannot exceed the number of columns or rows of **H**, $r \le m = \min(M, N)$. If **H** is full rank, which is sometimes referred to as a *rich scattering environment*, then $r = m$. Equation (8.7) can be rewritten as

$$(\lambda \mathbf{I}_m - \mathbf{W})\mathbf{z} = 0, \quad \mathbf{z} \neq \mathbf{0} \qquad (8.8)$$

where **W** is the Wishart matrix defined to be

$$\mathbf{W} = \begin{cases} \mathbf{H}\mathbf{H}^H, & \text{if } M < N \\ \mathbf{H}^H\mathbf{H}, & \text{if } M \geq N \end{cases}$$

This implies that

$$\det(\lambda \mathbf{I}_m - \mathbf{W}) = 0 \tag{8.9}$$

The m nonzero eigenvalues of $\mathbf{W}$, λ_1, λ_2, …, λ_m, can be calculated by finding the roots of (8.9). The non-negative square roots of the eigenvalues of $\mathbf{W}$ are also referred to as *singular values* of $\mathbf{H}$.

Substituting (8.6) into (8.1), we have

$$\mathbf{y} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^H\mathbf{x} + \mathbf{n}$$

Let $\tilde{\mathbf{y}} = \mathbf{U}^H\mathbf{y}$, $\tilde{\mathbf{x}} = \mathbf{V}^H\mathbf{x}$, $\tilde{\mathbf{n}} = \mathbf{U}^H\mathbf{n}$. Note that $\mathbf{U}$ and $\mathbf{V}$ are invertible, $\tilde{\mathbf{n}}$ and $\mathbf{n}$ have the same distribution (i.e., zero-mean Gaussian with i.i.d. real and imaginary parts), and $E[\tilde{\mathbf{x}}^H\tilde{\mathbf{x}}] = E[\mathbf{x}^H\mathbf{x}]$. Thus the original channel defined in (8.1) is equivalent to the channel

$$\tilde{\mathbf{y}} = \mathbf{\Lambda}\tilde{\mathbf{x}} + \tilde{\mathbf{n}} \tag{8.10}$$

where $\mathbf{\Lambda} = diag\left(\sqrt{\lambda_1}, \sqrt{\lambda_2}, ..., \sqrt{\lambda_m}, 0, ..., 0\right)$ with $\sqrt{\lambda_i}, i = 1, 2, ..., m$ denoting the non-zero singular values of $\mathbf{H}$. The equivalence is summarized in Fig. 8.2. From (8.10), we obtain for the received signal components

$$\begin{aligned} \tilde{y}_i &= \sqrt{\lambda_i}\tilde{x}_i + \tilde{n}_i, \quad 1 \leq i \leq m \\ \tilde{y}_i &= \tilde{n}_i, \qquad \text{m+1} \leq i \leq M \end{aligned} \tag{8.11}$$

It is seen that received components $\tilde{y}_i$, $i > m$, do not depend on the transmitted signal. On the other hand, received components $\tilde{y}_i$, $i = 1, 2, ..., m$ depend only the transmitted component $\tilde{x}_i$. Thus the equivalent MIMO channel in (8.10) can be considered as consisting of m uncoupled parallel Gaussian sub-channels. Specifically,

- If N>M, (8.11) indicates that there will be at most M non-zero attenuation subchannels in the equivalent MIMO channel. See Fig. 8.3.
- If M>N, there will be at most N non-zero attenuation subchannels in the equivalent MIMO channel.

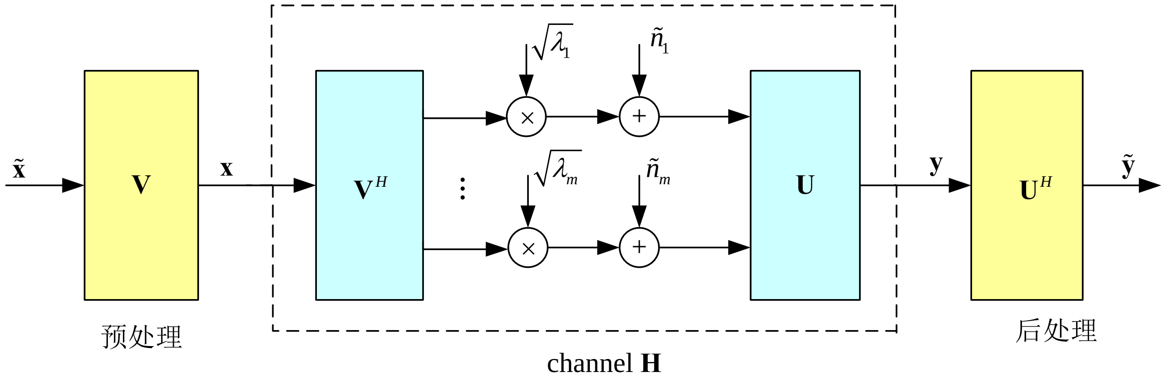


Figure 8.2 Converting the MIMO channel into a parallel channel through the SVD.

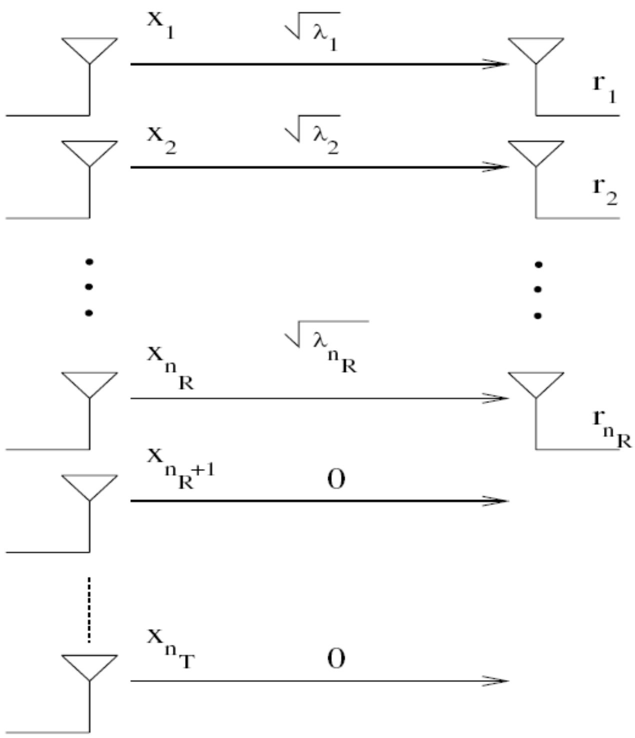


Fig. 8.3 Block diagram of an equivalent MIMO channel for $N>M$

With the above model (parallel orthogonal channels), the fundamental capacity of an MIMO channel can be calculated in terms of the positive eigenvalues of the matrix $\mathbf{HH}^H$ as follows.

8.3.2 Channel Known to the Transmitter

When the perfect channel knowledge is available at both the transmitter and the receiver, the transmitter can optimize its power allocation or input covariance matrix across antennas according to the "water-filling" rule (in space) to maximize the capacity formula (8.5). Substituting the matrix SVD (8.6) into (8.5) and using properties of unitary matrices, we get the MIMO capacity with CSIT and CSIR as

$$C = \max_{K_x : Tr(K_x) = P} \log_2 \left(\det \left(\mathbf{I}_N + \frac{1}{N_0} \mathbf{\Lambda} K_x \mathbf{\Lambda}^H \right) \right)$$

$$= \max_{P_i : \sum_i P_i \le P} \sum_{i=1}^{m} \log_2 \left(1 + \frac{P_i \lambda_i}{N_0} \right)$$

where P_i is the transmit power in the ith sub-channel. Solving the optimization leads to a water-filling power allocation over the parallel channels. The power allocated to channel i, $1 \le i \le m$, is given parametrically by

$$P_i = \left(\mu - \frac{N_0}{\lambda_i} \right)^+ \tag{8.12}$$

where a^+ denotes $max(0, a)$, and μ is a constant that is chosen to satisfy

$$\sum_{i=1}^{m} P_i = P \tag{8.13}$$

The resulting capacity is then

$$C_{\text{WF}} = \sum_{i=1}^{m} \log_2 \left[1 + \frac{1}{N_0} \left(\mu \lambda_i - N_0 \right)^+ \right] = \sum_{i=1}^{m} \log_2 \left(\frac{1}{N_0} \mu \lambda_i \right)^+ \quad \text{bits/channel use} \tag{8.14}$$

which is achieved by choosing each component $\tilde{x}_i$ according to an independent Gaussian distribution with power P_i. The covariance matrix of the capacity-achieving transmitted signal is given by

$$K_x = \mathbf{VPV}^H$$

where $\mathbf{P} = \text{diag}\left(P_1, P_2, \ldots, P_m, 0, \ldots, 0 \right)$ is an $N \times N$ matrix. Figure 8.4 depicts the SVD-based architechture for MIMO communication.

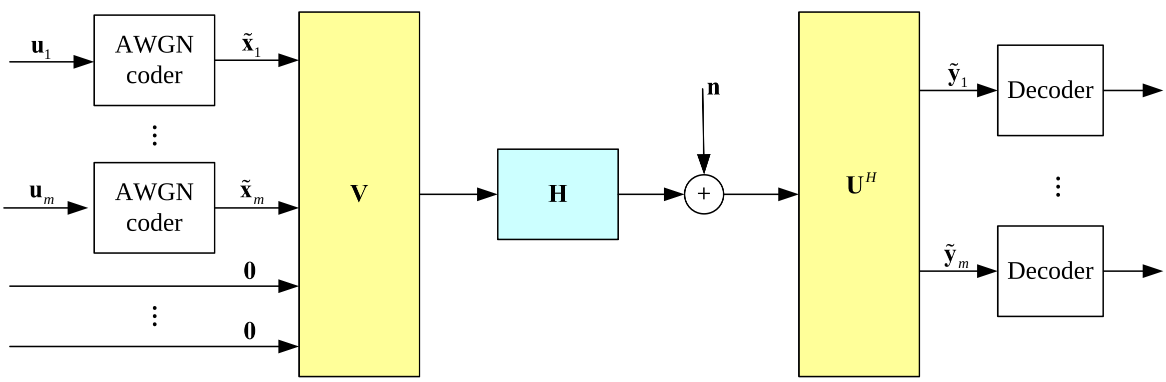


Figure 8.4 The SVD-based architechture for MIMO communication.

- Water-filling algorithm:
  The power allocation in (8.12) can be determined iteratively using the water-filling

algorithm. We now describe it.

We first set the iteration count p to 1 and assume that 所有(m-p+1)个并行子信道都使用。With this assumption, the constant μ is calculated (by substituting (8.12) into (8.13)) as

$$\sum_{i=1}^{m-p+1}\left(\mu-\frac{N_0}{\lambda_i}\right)=P$$

Then we have

$$\mu=\frac{1}{m-p+1}\left(P+N_0\sum_{i=1}^{m-p+1}\frac{1}{\lambda_i}\right) \tag{8.15a}$$

Using this value of μ, the power allocated to the ith subchannel is given by

$$P_i=\left(\mu-\frac{N_0}{\lambda_i}\right),\quad i=1,2,\ldots,m-p+1 \tag{8.15b}$$

If the power allocated to the channel with the lowest gain is negative (i.e., $P_{m-p+1}<0$), then we discard this channel by setting $P_{m-p+1}=0$ and return the algorithm with the iteration count p = p+1. 即迭代执行(8.15a)和(8.15b)，将总功率 P 在剩余的(m-p+1)个子信道之间进行分配。迭代计算直到获得的所有 $P_i \geq 0$ 或 p=m 为止。

8.3.3 Channel Unknown to the Transmitter

If the channel is known to the receiver, but not to the transmitter, then the transmitter cannot optimize its power allocation or input covariance structure across antennas. This implies that if the distribution of **H** follows the zero-mean spatially white (ZMSW) channel gain model, the signals transmitted from N antennas should be independent and the power should be equally divided among the transmit antennas, resulting an input covariance matrix $\mathsf{K}_x=\frac{P}{N}\mathbf{I}_N$. It is shown in [Telatar99] that this K_x indeed maximize the mutual information. Thus, the capacity in such a case is

$$C=\begin{cases}\log_2\left[\det\left(\mathbf{I}_M+\frac{\mathsf{SNR}}{N}\mathbf{H}\mathbf{H}^H\right)\right], & \text{if } M<N\\ \log_2\left[\det\left(\mathbf{I}_N+\frac{\mathsf{SNR}}{N}\mathbf{H}^H\mathbf{H}\right)\right], & \text{if } M\geq N\end{cases} \quad \text{bits per channel use} \tag{8.15}$$

where $\mathsf{SNR}=P/N_0$. Using the SVD of **H**, we can express this as

$$C=\sum_{i=1}^{m}\log_2\left(1+\frac{P_{ri}}{N_0}\right)=\sum_{i=1}^{m}\log_2\left(1+\frac{\lambda_i P}{NN_0}\right)$$

$$= \sum_{i=1}^{m} \log_2 \left(1 + \frac{\text{SNR}}{N} \lambda_i \right) \qquad \text{bits/channel use} \qquad (8.16)$$

where P_{ri} is the received signal power in the *i*th sub-channel. Equation (8.16) expresses the capacity of the MIMO channel as a sum of the capacities of m SISO channels, each having a power gain of λ_i and transmit power P/N. (Note: Because of the use of complex signals, the above unit is sometimes expressed in terms of "bits/sec/Hz".)

Since the nonzero eigenvalues of $\mathbf{HH}^H$ are the same as those of $\mathbf{H}^H\mathbf{H}$, the capacity of a channel with matrix **H** and $\mathbf{H}^H$ are the same. Furthermore, the capacity can be achieved by choosing independent $\{\tilde{x}_i, 1 \le i \le m\}$ with each $\tilde{x}_i$ having independent Gaussian, zero-mean real and imaginary parts.

8.3.4 MIMO Capacity Examples

Example 1 [Single antenna channel]. Consider a channel with N=M=1 and $\mathbf{H}$=h=1. The Shannon capacity of this channel is

$$C = \log_2 \left(1 + \frac{P\,|h|^2}{N_0} \right) = \log_2 \left(1 + \gamma\,|h|^2 \right) \qquad \text{bit/channel use} \qquad (8.17)$$

Example 2 [MIMO channel with coherent combining]. Consider a MIMO channel with h_{ij}=1 for all $1 \le i \le N$, $1 \le j \le M$. We can write **H** as

$$\mathbf{H} = \mathbf{UDV}^H = \begin{bmatrix} \sqrt{1/M} \\ \vdots \\ \sqrt{1/M} \end{bmatrix} \left(\sqrt{MN} \right) \left[\sqrt{1/N} \quad \cdots \quad \sqrt{1/N} \right]$$

and we see that the diagonal matrix **D** will have only one nonzero entry $\sqrt{MN}$. Thus, the Shannon capacity of this channel is

$$C = \log_2 \left(1 + MN \frac{P}{N_0} \right) \qquad (8.18)$$

The $\mathbf{x} = \mathbf{V}\tilde{\mathbf{x}}$ that achieves this capacity satisfies $E[x_i x_j^*] = P/N$ for all i, j; i.e., the *transmitters are all sending the same signal*. Thus, we can see that **H** corresponds to such a system in which the same signal x is transmitted from N transmit antennas and the receiver performs coherent maximum ratio combining (MRC) by M antennas. Then (8.18) can be interpreted as follows.

The received signal at antenna i is given by $y_i = Nx$ and the received signal power at antenna i is $P_{ri} = N^2 \cdot \frac{P}{N}$. Since each receiver sees the same signal, and the noises at the receivers are uncorrelated, the overall SNR is $MP_{ri}/\sigma^2 = M \cdot N \cdot P/\sigma^2$. Then (8.18) follows from (8.16).

We can see that this system achieves a diversity gain of *MN* relative to a single antenna link. However, the capacity grows logarithmically with the total number of antennas *MN*.

Note: If the signals transmitted from various antennas are different and all channel entries are equal to 1, the capacity is given by

$$C = \log_2\left(1 + M\frac{P}{\sigma^2}\right)$$

Example 3 [MIMO channel with orthogonal transmissions]. Consider a channel with $N=M=n$ and $\mathbf{H}=\mathbf{I}_n$. Since $\mathbf{H}\mathbf{H}^H=\mathbf{I}_n$, from (8.15) we obtain

$$C = \log_2\left[\det\left(\mathbf{I}_n + \frac{\gamma}{N}\mathbf{I}_n\right)\right]$$

$$= \log_2\left(1+\frac{\gamma}{N}\right)^n = n\log_2\left(1+\frac{P}{N\sigma^2}\right) \tag{8.19}$$

As the number of antennas $n\to\infty$, the capacity approaches

$$\lim_{n\to\infty} C = \frac{\gamma}{\ln 2}$$

In this case the capacity increases linearly with SNR. Equation (8.19) indicates that a MIMO channel gives a multiplexing gain of *n*. An implementation example of such system is to spread transmitted signals from various antennas by orthogonal spreading sequences.

For $\mathbf{x}$ that achieves the capacity given in (8.19), $E[x_i x_j^*] = \delta_{i-j} P/N$. Notice that, however, we cannot conclude that to achieve capacity one has to do independent coding for each transmitter. It is true that the capacity of this channel can be achieved by splitting the incoming data stream into *N* streams, coding and modulating these streams separately, and then transmitting the *N* modulated signals over the different antennas.

Example 4 [Receive diversity]. Consider a system with $N=1$ transmit and $M>1$ receive antennas. In this case, the channel matrix can be represented by the vector $\mathbf{H} = [h_1\ h_2\ \ldots\ h_M]^T$.

Since $\mathbf{H}^H\mathbf{H} = \sum_{i=1}^{M} |h_i|^2$, from (8.15), we obtain

$$C = \log_2\left(1+\gamma\sum_{i=1}^{M}|h_i|^2\right) = \log_2\left(1+\frac{P}{N_0}\sum_{i=1}^{M}|h_i|^2\right) \tag{8.20}$$

This capacity corresponds to linear MRC at the receiver. In the case when $|h_i|^2=1$ for all $1\le i\le M$, (8.20) becomes

$$C = \log_2\left(1 + M\frac{P}{N_0}\right) \tag{8.21}$$

This system achieves the diversity gain of M relative to a single antenna channel.

Example 5 [Transmit diversity]. In this system there are N transmit antennas and only one receive antenna. The channel matrix is given by the vector $\mathbf{H} = [h_1\ h_2\ \ldots\ h_N]$. Substituting $\mathbf{HH}^H = \sum_{j=1}^{N} |h_j|^2$ into (8.15) yields

$$C = \log_2\left(1 + \frac{\gamma}{N}\sum_{j=1}^{N}|h_j|^2\right) = \log_2\left(1 + \frac{P}{N\sigma^2}\sum_{j=1}^{N}|h_j|^2\right) \tag{8.22}$$

With the power normalization, i.e., $|h_1|^2=|h_2|^2=\ldots=|h_N|^2=1$, the capacity becomes

$$C = \log_2\left(1 + \frac{P}{\sigma^2}\right) \tag{8.23}$$

This equation applies to the case when the transmitter does not know the channel. When the transmitter knows the channel, we can apply the capacity formula in (8.14). Since m=min(M, N)=1, there is only one nonzero eigenvalue given by

$$\lambda = \sum_{j=1}^{N}|h_j|^2$$

Combining (8.12) and (8.13), we have $\mu = P + \frac{\sigma^2}{\lambda}$. So the capacity is given by

$$C_{\mathrm{WF}} = \log_2\left(1 + \frac{P}{\sigma^2}\sum_{j=1}^{N}|h_j|^2\right) \tag{8.24}$$

8.4 Capacity of MIMO Rayleigh Fading Channels (Random MIMO Channels)

We now proceed to consider the case when the channel matrix entries are (complex-valued) random variables (usually referred to as random MIMO channel). Assume that perfect CSI is available at the receiver but no CSI at the transmitter. Furthermore, we assume that that the entries of **H** have Rayleigh distributed magnitudes, uniform phases and expected magnitude squares equal to unity, $E[|h_{ij}|^2] = 1$. Equivalently, each entry of **H** is assumed to be zero-mean Gaussian with independent real and imaginary parts, each with variance 1/2. The antenna spacing is larger than one half of the carrier wavelength to ensure that the entries of **H** are independent. According to frequency of channel coefficient changes, we will distinguish three scenarios: namely, fast, slow and block fading channels.

Note: The probability density function (pdf) for a Rayleigh distributed random variable $z = \sqrt{z_c^2 + z_s^2}$, where z_c and z_s are two i.i.d Gaussian random variables each having zero mean and a variance σ_r^2, is given by $p(z) = \frac{z}{\sigma^2}\exp\left(-\frac{z^2}{2\sigma^2}\right)$, $z \geq 0$.

8.4.1 Capacity in Fast and Block Fading channels

We first consider the fast and flat fading channel. In other words, the channel is assumed to be memoryless: for each use of the channel an independent realization of **H** is drawn. In this case the ergodic capacity formula can be applied.

- For the simple single antenna link, the ergodic capacity is

$$C = E\left[\log_2\left(1+|h|^2\gamma\right)\right] = E\left[\log_2\left(1+\chi_2^2\gamma\right)\right]$$

where $\chi_2^2 = x_1^2 + x_2^2$ is a chi-squared distributed random variable with two degrees of freedom, and the expectation is performed with respect to the variable χ_2^2.

- The MIMO capacity in fast fading channels can be calculated as follows.
  The channel model is described by

$$\mathbf{y}[k] = \mathbf{H}[k]\mathbf{x}[k] + \mathbf{n}[k]$$

where k=1, 2,…, is the discrete time index.

By viewing the channel output as the pair (**y**, **H**), the mutual information between channel input and output is

$$I(\mathbf{x};(\mathbf{y},\mathbf{H})) = I(\mathbf{x};\mathbf{H}) + I(\mathbf{x};\mathbf{y}\,|\,\mathbf{H})$$

$$= I(\mathbf{x};\mathbf{y}\,|\,\mathbf{H})$$
$$= \mathsf{E}_{\mathbf{H}}\left[I(\mathbf{x};\mathbf{y}\,|\,\mathbf{H}=H)\right]$$

From (8.15), the rate achieved in a given channel state is

$$I\left(\mathbf{x};\mathbf{y}\,|\,\mathbf{H}=H\right) = \log_2\left[\det\left(\mathbf{I}_M + \frac{1}{N_0}H\mathsf{K}_x H^H\right)\right]$$

where $\mathbf{K}_x$ is the covariance matrix of the transmitted signals, and we have assumed M<N. As usual, by coding over many coherence time intervals of the channel, a long-term rate of reliable communication (Shannon capacity) is equal to

$$C = \max_{\mathsf{K}_x:\mathrm{tr}[\mathsf{K}_x]\le P} \mathsf{E}_{\mathbf{H}}\left\{\log_2\left[\det\left(\mathbf{I}_M + \frac{1}{N_0}\mathbf{H}\mathsf{K}_x\mathbf{H}^H\right)\right]\right\} \quad \text{bits/channel use} \qquad (8.25)$$

With the iid Rayleigh fading model, the capacity is achieved when **x** is a circularly symmetric zero-mean complex Gaussian vector with covariance matrix $\mathsf{K}_x = \left(\frac{P}{N}\right)\mathbf{I}_N$. With this equal powers, the resulting capacity is

$$C = \mathsf{E}\left\{\log_2\left[\det\left(\mathbf{I}_m + \frac{P}{NN_0}\mathbf{W}\right)\right]\right\}$$

$$= \mathsf{E}\left\{\log_2\left[\det\left(\mathbf{I}_m + \frac{\mathsf{SNR}}{N}\mathbf{W}\right)\right]\right\} \tag{8.26}$$

where $\mathbf{W} = \begin{cases} \mathbf{H}\mathbf{H}^H, & \text{if } M < N \\ \mathbf{H}^H\mathbf{H}, & \text{if } M \geq N \end{cases}$ is a matrix with Wishart distribution.

- For block fading channels, as long as the expected value with respect to the channel matrix **H** in (8.26) can be observed, we can calculate the channel capacity by using the same expression as in (8.26). This is usually the case of no delay constraint. Biglieri *et al*. discussed the block-fading channels with delay constraints in [Big2001].

Notice that the expectation in (8.26) is quite complex for larger values of *N* and *M*. With the aid of Laguerre polynominals, it can be evaluated as follows. (A detailed discussion can be found in [Telatar95]

$$C = \int_0^\infty \log_2\left(1 + \frac{\gamma}{N}\lambda\right)\sum_{k=0}^{m-1}\frac{k!}{(k+n+m)!}\left[L_k^{n-m}(\lambda)\right]^2 \lambda^{n-m}e^{-\lambda}d\lambda \tag{8.27}$$

where $m = \min(N, M)$, $n = \max(N, M)$, $L_k^{n-m}(x)$ is the associated Laguerre polynominal of order *k*, defined as

$$L_k^{n-m}(x) = \frac{1}{k!}e^x x^{m-n}\frac{d^k}{dx^k}(e^{-x}x^{n-m+k}) \tag{8.28}$$

Using limiting arguments, the capacity in (8.27) is upper and lower bounded by

$$C \leq m\log\frac{\gamma}{N} + \log(m!) + \log\left[L_m^{n-m}(-N/\gamma)\right]$$

$$C \geq m\log\frac{\gamma}{N} + \sum_{i=0}^{m=1}\psi(n-i)$$

where ψ is Euler's digamma function.

Example 6 [Fast fading channel with receive diversity-SIMO]. Consider a system with *N*=1 transmit and *M*>1 receive antennas in a fast Rayleigh fading channel. In this case, the channel matrix $\mathbf{H} = [h_1\ h_2\ \ldots\ h_M]^T$, and the capacity (for MRC) is given by

$$C = E\left[\log_2\left(1 + \frac{P}{\sigma^2}\chi_{2M}^2\right)\right] \tag{8.29}$$

where $\chi^2_{2M} = \sum_{i=1}^{M} |h_i|^2$ is a chi-squared random variable with 2M degrees of freedom. The capacity curves are shown in Fig. 8.5.

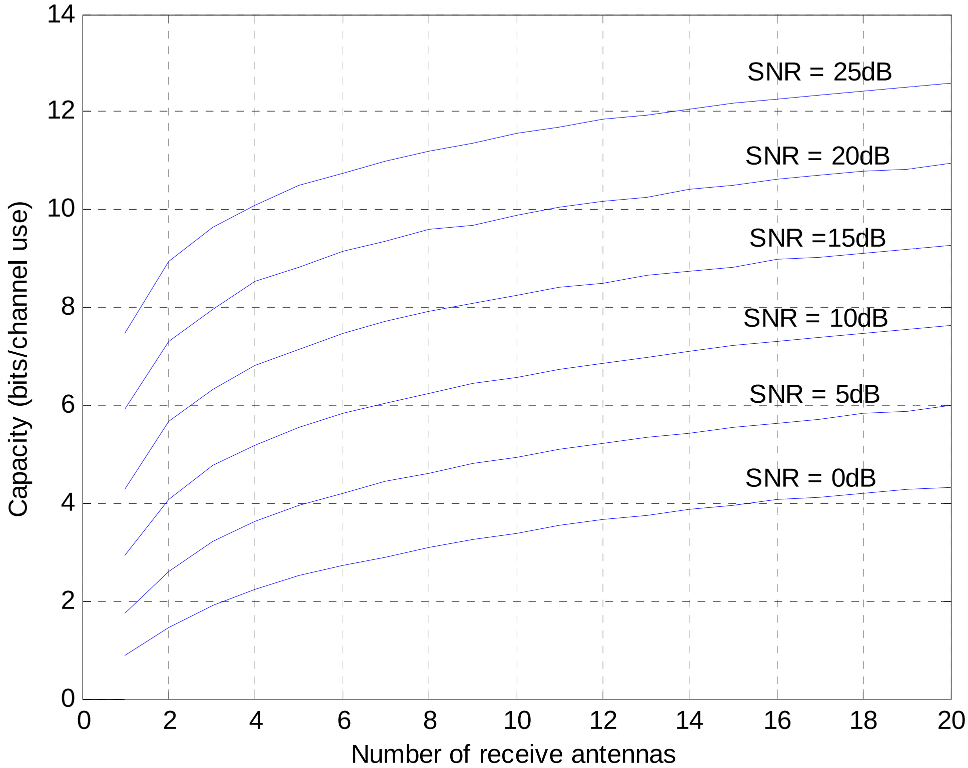


Fig.8.5 Capacity of SIMO (N=1) in a fast and block Rayleigh fading channel with MRC

Example 7 [Fast fading channel with transmit diversity-MISO]. Consider a system with N>1 transmit and M=1 receive antennas in a fast Rayleigh fading channel. In this case, the channel matrix is $\mathbf{H} = [h_1 \; h_2 \; ... \; h_N]$, and the Shannon capacity is given by

$$C = E\left[\log_2\left(1 + \frac{P}{N\sigma^2}\chi^2_{2N}\right)\right] \tag{8.30}$$

where $\chi^2_{2N} = \sum_{j=1}^{N} |h_j|^2$ is a chi-squared random variable with 2N degrees of freedom. The capacity curves are shown in Fig. 8.6.

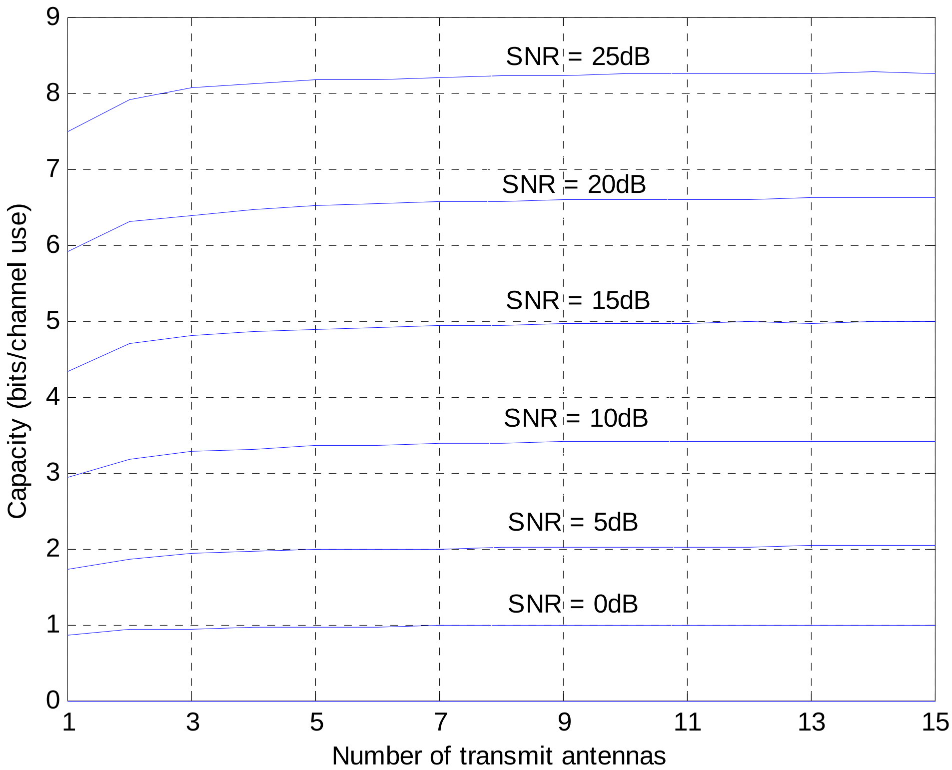


Fig.8.6(a) Capacity of MISO (for M=1) in a fast and block Rayleigh fading channel

As the number of transmit antennas increases, the capacity approaches the asymptotic value

$$\lim_{N\to\infty} C = \log_2\left(1+\frac{P}{\sigma^2}\right) \tag{8.31}$$

which indicates that ***the system behaves as if the total power is transmitted over a single unfaded channel***. In other words, the transmit diversity is able to remove the effect of fading for a large number of antennas. We can see from Fig. 8.6 that the capacity of the transmit diversity saturates for $N \geq 2$. That is, the capacity asymptotic value is achieved for the number of transmit antennas of 2 and there is no point in increasing it further.

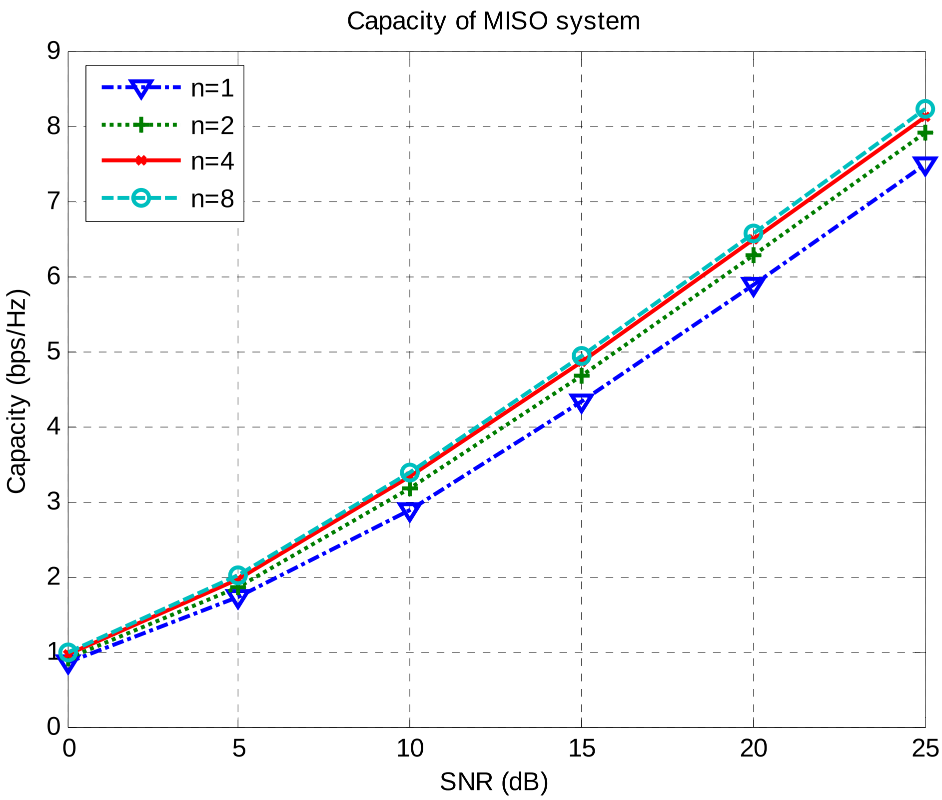


Figure 8.6(b) Capacity for MISO system with *N* transmit antennas

Example 8 [MIMO fast fading channel] We consider a system with N>1 transmit and M=N receive antennas in a fast Rayleigh fading channel. Assuming perfect channel knowledge at the receiver, but no channel knowledge at the transmitter.
It has been shown [Tse2005] that

- with $M = N = m = n$, the capacity can be approximated by $nc^*(\mathrm{SNR})$, where $c^*(\mathrm{SNR})$ is a constant given by

$$c^*(\mathrm{SNR}) = 2\log\left(1+\mathrm{SNR}-\frac{1}{4}F(\mathrm{SNR})\right)-\frac{\log e}{4\mathrm{SNR}}F(\mathrm{SNR}),$$

where

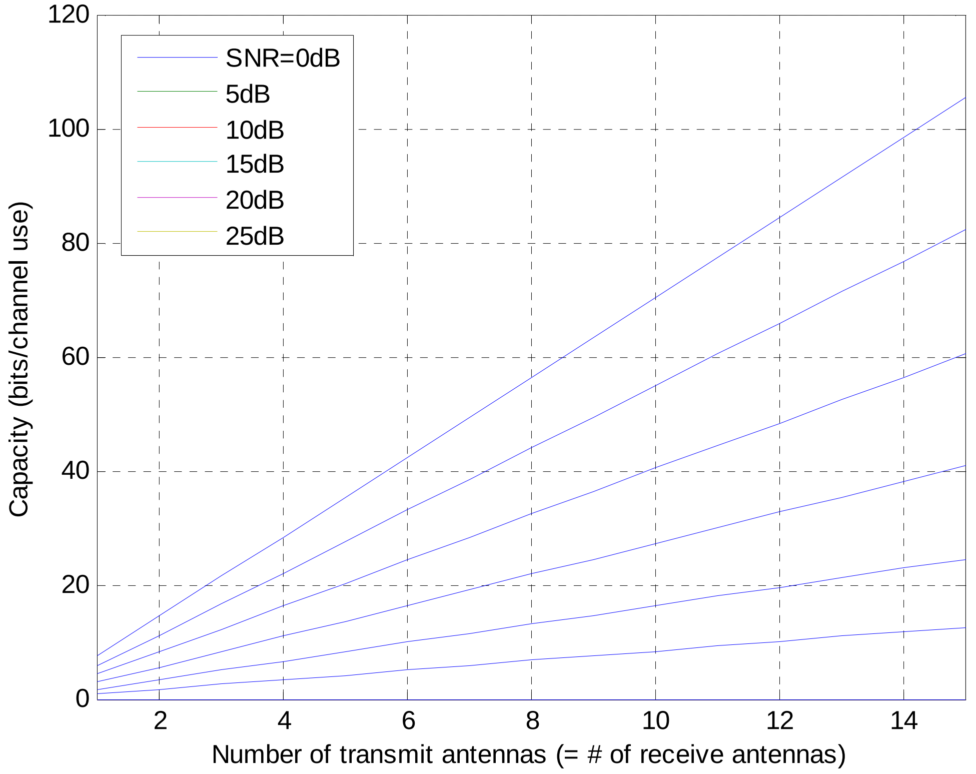
$$F(\mathrm{SNR}) \equiv \left(\sqrt{4\mathrm{SNR}+1}-1\right)^2$$

- At high SNR, the capacity is approximately equal (up to an additive constant) to $n\log\mathrm{SNR}$ bits/s/Hz.

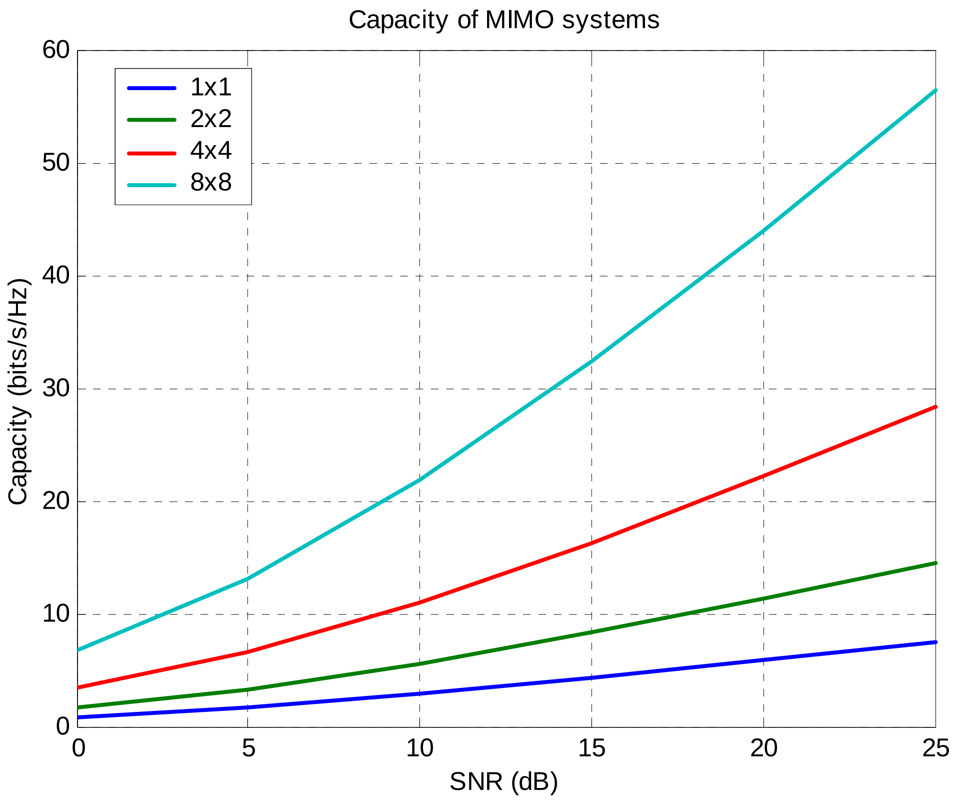
- At low SNR, the capacity is approximately equal to $n\mathrm{SNR}\log_2 e$ bits/s/Hz.

As a consequence, we can see that in a n×n MIMO channel, the capacity increases ***linearly*** with *n* over the entire SNR range.

Fig. 8.7 shows the capacity for different numbers of antennas in an iid Rayleigh fading channel.



(a) Capacity versus M for N=M, 0dB≤SNR≤25dB



(b)

Figure 8.7 Capacity for MIMO systems in an iid Rayleigh fading channel

The capacities, as a function of *n*, are plotted for the SIMO, MISO, and MIMO channels in Fig.8.8.

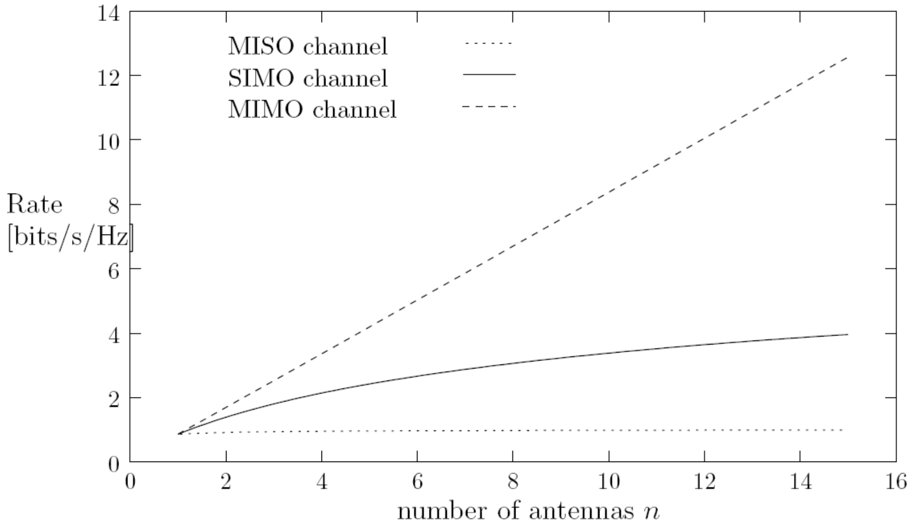


Figure 8.8 Capacities of the n×1 MISO channel, 1×n SIMO channel and the n×n MIMO channel, for SNR = 0dB.

8.4.2 The MIMO Capacity In Slow Rayleigh Fading Channels

In the case when **H** is chosen randomly at the beginning of a transmission block and held constant for all the uses of the channel, the maximum mutual information is in general not equal to the channel capacity. Specifically, for slow, flat fading channels, the model in () becomes

$$\mathbf{y}[k] = \mathbf{H}\mathbf{x}[k] + \mathbf{n}[k]$$

and each codeword, however long, experience only one channel state. This fading model is nonergodic.

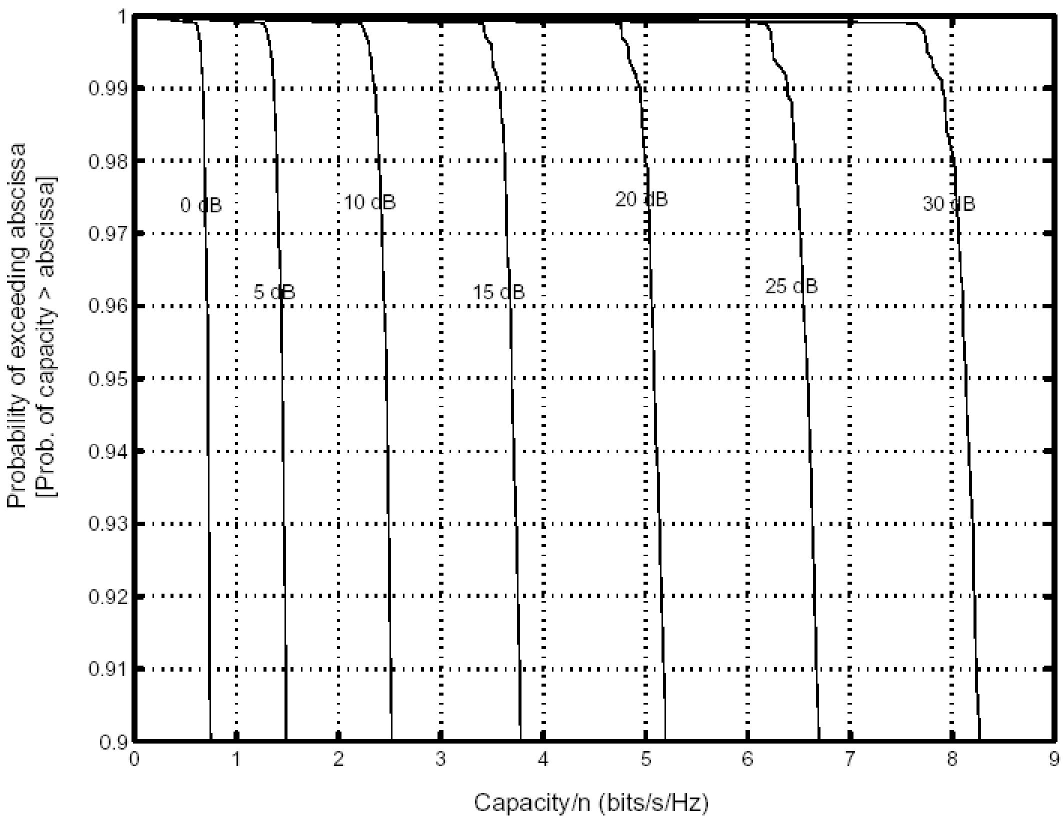
The capacity estimated by (8.15) is a random variable. We estimate the capacity complementary cumulative distribution function (CCDF), which results in the concept of *outage capacity*. The outage probability, denoted by P_{out}, specifies the probability of not achieving a certain level of capacity. It is equal to the capacity cumulative distribution function (CDF). Specifically, given a rate *R* and power *P*, the outage probability is defined by

$$P_{out}(R,P) = \Pr\left(C(\mathbf{H}) \le R\right)$$

where

$$C(\mathbf{H}) = \log_2\left[\det\left(\mathbf{I}_m + \frac{P}{NN_0}\mathbf{H}\mathbf{H}^H\right)\right]$$

Usually, the capacity curves are expressed in terms of “P_c = 1-P_{out} vs. *R*”; i.e., CCDF. Fig. 8.9 shows the simulation results for the CCDF capacity per antenna on a slow Rayleigh fading channel in an MIMO system with *N*=*M*=8.



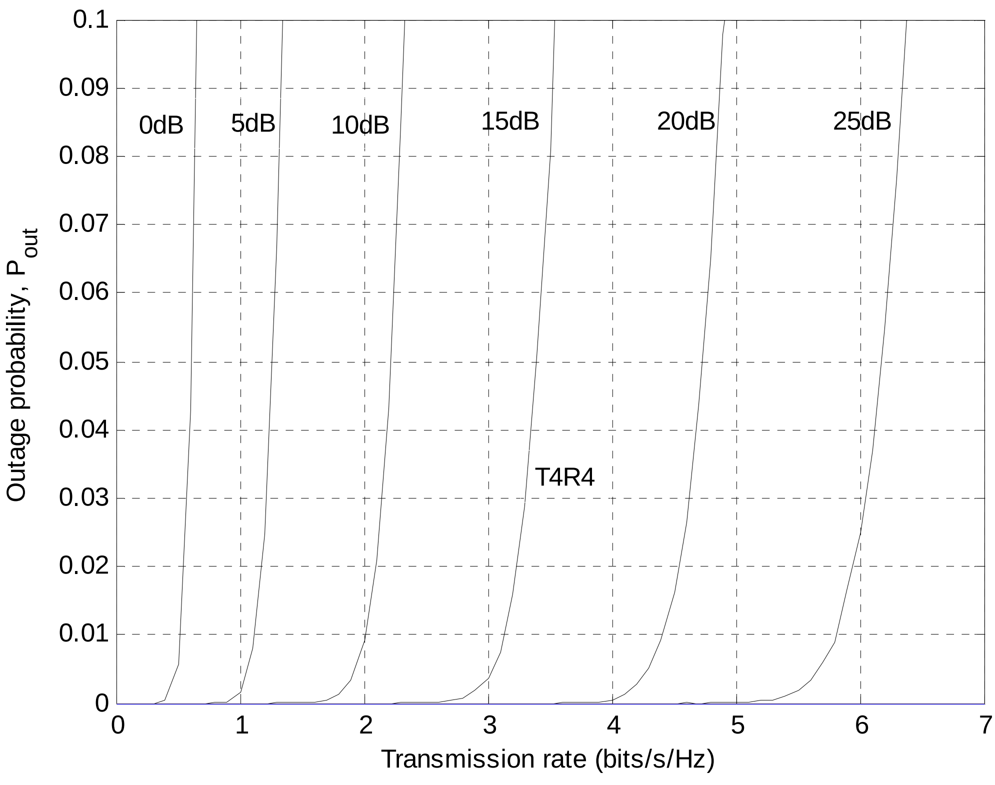
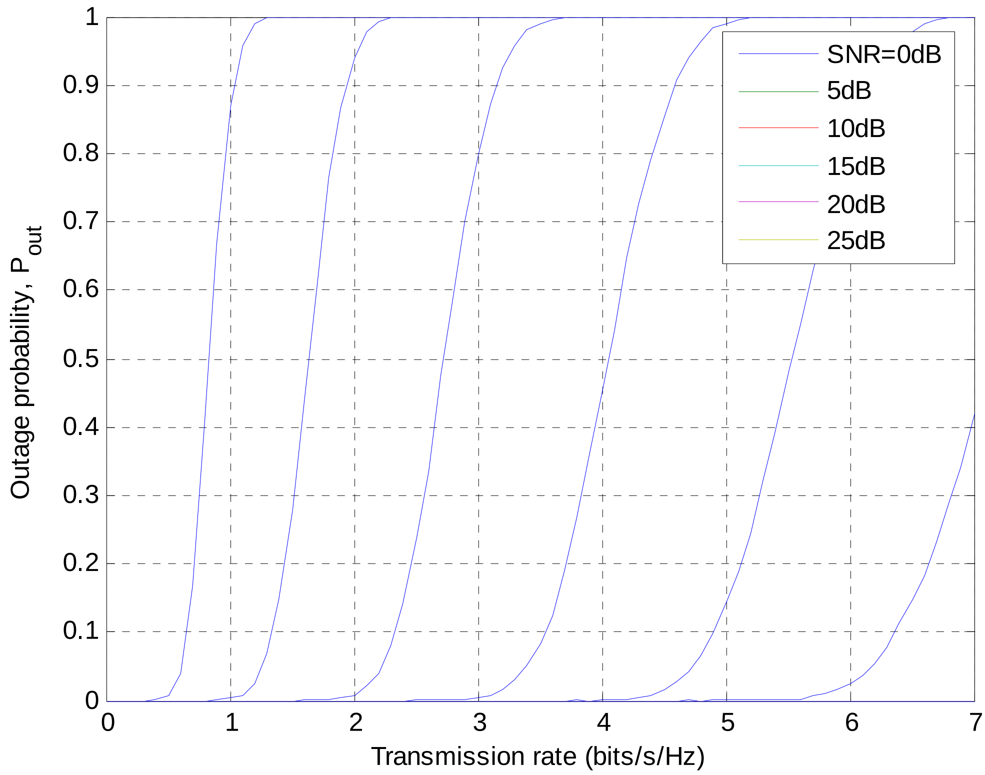


Fig. 8.9 Capacity per antenna CCDF curves for a constant number of antennas *N*=*M*=8 and a variable SNR



8.4.3 Constrained-Capacity in Fast Fading channels

For the discrete-time channel model

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n},$$

given in (8.1), let $\mathbf{x}$ be an $N\times1$ vector of symbols chosen from a signal constellation $\mathcal{A} \subset \mathbb{C}$ with $|\mathcal{A}|=2^b$ signal points. Let $E_s = P$ be the total transmitted power. We assume that the vector $x = [x_1,\ldots,x_n,\ldots,x_N]^T$ obeys the component-wise energy constraint $E[\|x_n\|^2] = E_s / N$. As mentioned above, assume that the components of the zero-mean complex Gaussian noise vector $\mathbf{n}$ have identical variance σ_n^2 per real dimension. Then, the SNR measured at the each receive antenna is

$$\gamma = \frac{E_s}{2\sigma_n^2} = \frac{E_s}{N_0}. \tag{8.32}$$

With the normalization constraint, the average signal energy per receive antenna is E_s. Hence, the M receive antennas collect total energy $M \cdot E_s$, carrying $N \cdot b$ coded bits or $R_c Nb$ information bits, where R_c is the rate of the used FEC code. Let R_m denote the number of information bits conveyed per transmitted symbol at each transmit antenna; i.e., R_m is the number of information bits transmitted by each x_n. We therefore have the average signal energy per information bit at the receiver $E_b = \frac{ME_s}{NR_c b} = \frac{ME_s}{NR_m} = \frac{ME_s}{\rho}$, where ρ is the spectral efficiency (in terms of b/s/Hz). The SNR can be expressed in terms of E_b/N_0 as

$$\frac{E_b}{N_0} = \frac{E_s}{N_0}\frac{M}{\rho} = \frac{M \cdot SNR}{\rho} \tag{8.33}$$

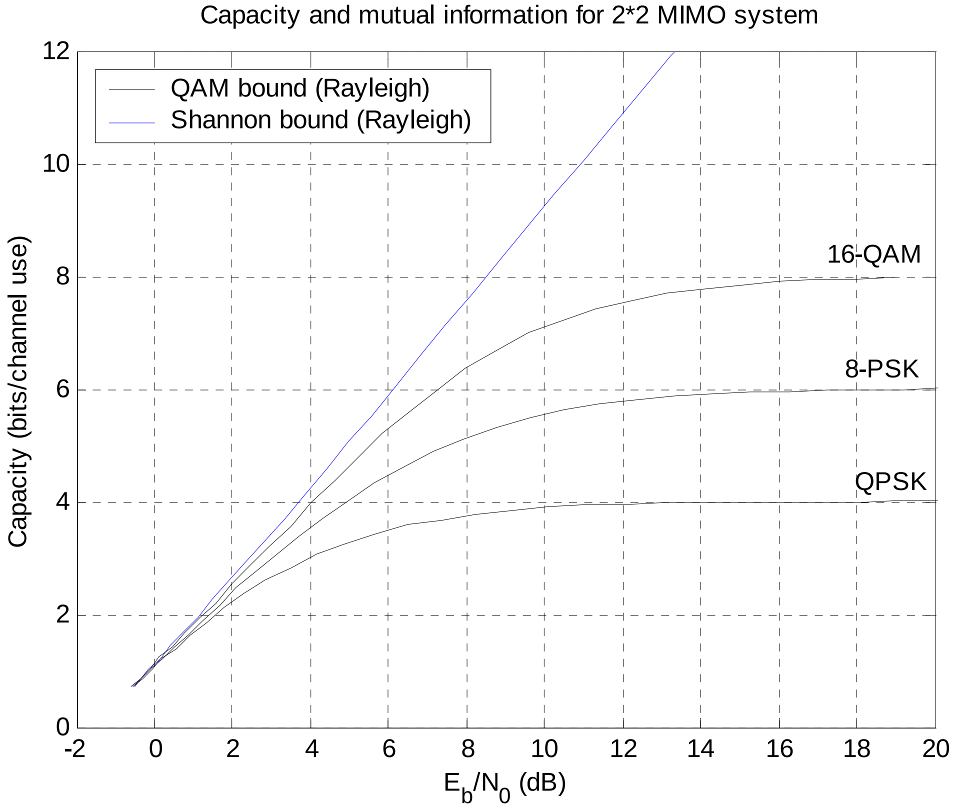


Figure 8.10 Constrained-capacity for MIMO system with N=M=2.

8.4.4 Influence of antenna correlation on the MIMO capacity

The capacity gain of MIMO channels, derived under the idealistic assumption that the channel matrix entries are independent complex Gaussian variables, might be reduced on real channels. Practical MIMO channel may not be iid. The effect of spatial fading correlation on the MIMO channel capacity has been addressed in []. Here we consider the separately correlated model. Assume that we model the correlation of the receive and the transmit array elements independently, their respective correlation matrices can be expressed as $\mathbf{R}_r$ and $\mathbf{R}_t$. The correlated MIMO channel matrix can be expressed in the form [Paulraj03]

$$\mathbf{H} = \mathbf{R}_r^{1/2}\mathbf{H}_w\mathbf{R}_t^{1/2}$$

where $\mathbf{H}_w$ is a $M \times N$ matrix of uncorrelated, circularly symmetric, zero mean, complex Gaussian r.v.'s. with unit variance, representing a Rayleigh i.i.d. spatially white MIMO channel, and $(\cdot)^{1/2}$ denotes matrix square root. Calculation of the ergodic capacity in this case is an open problem; its solution is known only in some special cases. Under the assumption that the entries of **H** are unknown at the transmitter, the MIMO capacity can be obtained with the independent signals across the transmit antennas, which can be written as

$$C = \mathsf{E}\left\{\log_2\left[\det\left(\mathbf{I}_r + \frac{\mathsf{SNR}}{N}\mathbf{R}_r^{1/2}\mathbf{H}_w\mathbf{R}_t\mathbf{H}_w^H\left(\mathbf{R}_r^{1/2}\right)^H\right)\right]\right\}$$

In the special case M=N and with the assumption that the receive and transmit correlation matrices are full rank, for high SNR the capacity can be approximated as []

$$C \approx \mathsf{E}\left\{\log_2\left[\det\left(\frac{\mathsf{SNR}}{N}\mathbf{H}_w\mathbf{H}_w^H\right)\right]\right\} + \log_2\left[\det\left(\mathbf{R}_t\right)\right] + \log_2\left[\det\left(\mathbf{R}_r\right)\right]$$

We note from above that both correlation matrixes have the same impact on the channel capacity. We now examine the conditions on $\mathbf{R}_t$ that maximize capacity. The same arguments apply to $\mathbf{R}_r$. Let λ_i, i=1,…,N denote the positive eigenvalues of $\mathbf{R}_t$, and recall that the power constraint $Tr\left(\mathbf{R}_t\right) = \sum_{i=1} \lambda_i = N$. We have

$$\det\left(\mathbf{R}_t\right)^{1/N} = \prod_i \lambda_i^{1/N} \leq \frac{1}{N}\sum_{i=1}\lambda_i = 1$$

This means that $\frac{1}{N}\log_2\left[\det\left(\mathbf{R}_t\right)\right] \leq 0$ with equality iff $\mathbf{R}_t = \mathbf{I}_N$. A similar result applies to $\mathbf{R}_r$. Therefore, antenna correlation does reduce the number of eigenvalues and thereby reduces the MIMO channel capacity. This loss in ergodic or outage capacity is given by $\log_2\left[\det\left(\mathbf{R}_t\right)\right] + \log_2\left[\det\left(\mathbf{R}_r\right)\right]$ bit/s/Hz.

Example: [Jankiraman04, pp.35]

If we assume an orthogonal channel where $M = N = 2$ and further assume that there is correlation only at the receiver, then we choose a receive correlation matrix as

$$\mathbf{R}_r = \begin{bmatrix} 1 & 0.8 \\ 0.8 & 1 \end{bmatrix}$$

We note from Figure 8.11 that there is a loss of 2.47 bit/s/Hz at high SNR compared with the case with no correlation.

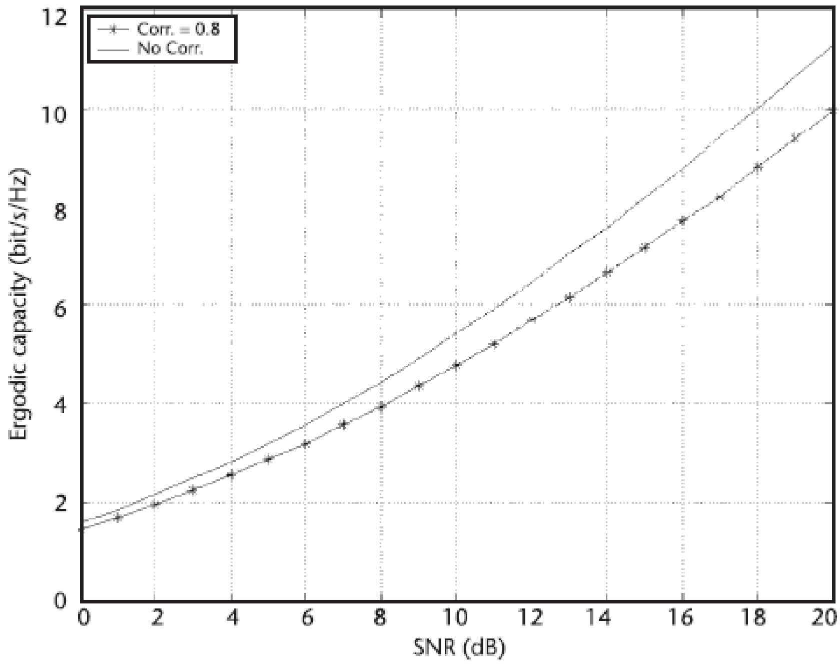


Figure 8.11

8.5 Influence of Channel-State Information [Biglieri05]

As we have seen, the aforementioned results on MIMO capacity are based on the assumption of perfect CSI available to the receiver, which can be viewed as a fundamental limit for coherent multiple-antenna systems. In practice, the channel knowledge may be imperfectly known or even not be available to the receiver. Biglieri and Taricco (2004) investigate effects that this imperfect estimation has on system performance. In the following, we will discuss the fundamental limits of noncoherent communication, where estimates of the fading coefficients are not available.

Consider a block fading channel model. Let L denote the block length (coherence time of the channel). To compute the capacity of this channel, we assume that coding is performed over mutilpe blocks, each of them consisting of NL elementary symbols to be transmitted by N antennas in L time instants. One block is represented by the $N \times L$ matrix $\mathbf{x}$. We further assume that the $M \times L$ noise matrix $\mathbf{n} \sim \mathcal{CN}\left(\mathbf{0},\ N_0\mathbf{I}_M\right)$. The received signal is the $M \times L$ matrix

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}$$

It has been shown [Marzetta and Hochwald, 1999] that the pdf of $\mathbf{y}$ can be expressed as

$$p(\mathbf{y} \mid \mathbf{x}) = \prod_{i=1}^{M} p(\mathbf{y}_i \mid \mathbf{x}) = \frac{1}{\pi^M \det^M\left[\mathbf{x}^H\mathbf{x} + \mathbf{I}_L\right]} \exp\left\{-\mathrm{Tr}\left(\left(\mathbf{x}^H\mathbf{x} + \mathbf{I}_L\right)^{-1}\mathbf{y}^H\mathbf{y}\right)\right\}$$

We observe the following:

(a) The pdf of $\mathbf{y}$ depends on its argument only through the product $\mathbf{y}^H\mathbf{y}$, which consequently plays the role of a sufficient statistic.

(b) The pdf of $\mathbf{y}$ depends on the transmitted signal $\mathbf{x}$ only through the $L \times L$ matrix $\mathbf{x}^H\mathbf{x}$.

Observation (b) above is the basis of the following theorem, which, in its essence, says that there is no increase in capacity if we have $N>L$; hence, there is no point in making the number of transmit antennas greater than L if there is no CSI. In particular, if $L=1$ (an independent fade occurs at each symbol interval), only one transmit antenna is useful. Note how this result contrasts sharply with its counterpart of CSI known at the receiver, where the capacity grows linearly with $\min\{N, M\}$.

Theorem 8.1: If the entries of **H** are i.i.d. (空间白 MIMO 信道), then (for any coherence interval of L symbols and any number of receive antennas) the channel capacity obtained with $N>L$ equals the capacity for $N=L$.

Marzetta and Hochwald also showed that the signal matrix that achieves capacity can be written in the form

$$\mathbf{x} = \mathbf{D}\mathbf{\Phi}$$

where $\mathbf{\Phi}$ is a $N \times L$ (unitary) matrix such that $\mathbf{\Phi}\mathbf{\Phi}^H = \mathbf{I}_N$, and **D** is a N×N real, non-negative, diagonal matrix independent of $\mathbf{\Phi}$. Moreover, $\mathbf{\Phi}$ has a pdf that is unchanged when the matrix is multiplied by a deterministic unitary matrix. The role of **D** is to scale **x** to meet the power constraint. In general, the optimal **D** is unknown; however, for the high-SNR regime, the following results are available (Marzetta and Hochwald, 1999; Zheng and Tse, 2002), but note that they depend critically on the assumed fading model (Lapidoth and Moser, 2003):

(a) If $L \gg N$ and $N \le \min\{L/2, M\}$, then capacity is attained when $\mathbf{D} = \sqrt{\mathsf{SNR} L N_0 / N}\,\mathbf{I}_N$, so that $\mathbf{x} = \sqrt{\mathsf{SNR} L N_0 / N}\,\mathbf{\Phi}$.

(b) For every 3 dB increase of SNR, the capacity increase is $N^*(1 - N^*/L)$, where $N^* = \min\left\{N, M, \lfloor L/2 \rfloor\right\}$.

(c) If $L \ge 2N$, there is no capacity increase by using $M>N$.

An obvious upper bound to capacity can be obtained if we assume that the receiver is provided with perfect knowledge of the realization of **H**. Hence, the bound to capacity per block of L symbols is

$$C \le L \cdot \log_2 \left[\det\left(\mathbf{I}_M + \frac{\mathsf{SNR}}{N} \mathbf{H}\mathbf{H}^H \right) \right]$$

8.6 Design of MIMO Systems

Many practical MIMO techniques have been developed to capitalize on the theoretical capacity gains predicted by Shannon theory. A major focus of such work is space-time coding. Other techniques for MIMO systems include space–time modulation, adaptive modulation and coding, space–time equalization, space–time signal processing, and space–time OFDM. An overview of the recent advances in these areas and other practical techniques along with their performance can be found in [4-10].

8.6.1 Two Usages of Multiple Antennas: Diversity and Multiplexing

Two key performance metrics associated with any communication system are the transmission rate and the frame-error rate (FER). In a MIMO system, multiple antennas can be utilized to achieve a higher transmission rate (as we have seen in capacity analysis), or provide a higher diversity gain, resulting in improved FER performance. Therefore, the advantage of a MIMO channel can be utilized in two ways: (1) to increase the diversity of the system and (2) to increase the number of transmitted symbols. This represents two way of the use of multiple antennas. It has been shown that there exists a fundamental trade-off between the transmission rate and FER. In [Zheng03], this trade-off is referred to as the diversity-multiplexing tradeoff (DMT) with diversity signifying the FER reduction and multiplexing signifying an increase in transmission rate. 它主要用来刻画空时传输方案在慢衰落信道上的性能。

Traditionally, multiple antennas have been used to increase *diversity* to combat channel fading. Each pair of transmit and receive antennas provides a signal path from the transmitter to the receiver. By sending signals that carry the same information through different paths, multiple independently faded replicas of the data symbol can be obtained at the receiver end; hence, more reliable reception is achieved. For example, in a slow Rayleigh-fading channel with one transmit and M receive antennas, the transmitted signal is passed through different paths. It is well known that if the fading is independent across antenna pairs, a maximal diversity gain of M can be achieved: the average error probability can be made to decay like $1/\mathrm{SNR}^M$ at high SNR, in contrast to the $1/\mathsf{SNR}$ for the single-antenna fading channel. *Space-time coding is such a method that uses multiple transmit antennas to get diversity*. In a system with N transmit and M receive antennas, assuming the path gains between individual antenna pairs are i.i.d. Rayleigh faded, the maximal diversity gain is MN, which is the total number of fading gains that one can average over. (For example, transmitting only one symbol per time slot)

A different line of thought suggests that in a MIMO channel, fading can in fact be beneficial, through increasing the *degrees of freedom* available for communication. Essentially, if the path gains between individual transmit–receive antenna pairs fade independently, the channel matrix is well conditioned with high probability, in which case multiple parallel *spatial channels* are created (as we have shown in SVD of **H**). By transmitting independent information streams in parallel through the spatial channels, the data rate can be increased. This effect is also called *spatial multiplexing*, and is particularly important in the high-SNR regime (where the system is degree-of-freedom limited). For a channel with N transmit and M receive antennas, from (8.16) and (8.26), the capacity in a fast fading scenario is given by

$$C(\mathsf{SNR}) = \sum_{i=1}^{m} \mathsf{E}\left[\log_2\left(1+\frac{\mathsf{SNR}}{N}\lambda_i\right)\right]$$

where m=min(M, N) is the number of spatial degrees of freedom in the channel. At high SNR, this ergodic capacity increases linearly with $\log_2$SNR according to the number of antennas; i.e.,

$$\lim_{\mathrm{SNR}\to\infty} \frac{C(\mathrm{SNR})}{\log_2 \mathrm{SNR}} = \min\{N, M\} \tag{8.34}$$

[$C = \min(M, N)\log SNR + O(1)$, which increases linearly with the number of antennas.]

The BLAST-type (Bell Labs space–time architecture) schemes are typical examples of such systems to exploit the spatial multiplexing (where m symbols are sent per time slot).

When we restrict our attention to communications over a slow fading channel, we have to back off from this ergodic capacity to achieve small error rates, but can still benefit from the increased degrees of freedom. To explore this, we consider rates that are a fixed fraction of this ergodic capacity at high SNR, i.e., rates of the form $R = r\log_2 \mathrm{SNR}$ with $0 < r \le \min\{N, M\}$, and evaluate the associate error behavior. The value of r can therefore be viewed as the spatial multiplexing gain.

In [Zheng03], it was shown that, for a given MIMO channel, both diversity gain and spatial multiplexing gain can be simultaneously obtained, but there is a fundamental tradeoff between them. 两种增益的最大值不能同时达到(For finite block lengths, it is not possible to achieve full diversity and full multiplexing gain)。

In the following discussions, we focus our attention on the non-ergodic fading channel with CSI available to the receiver only, and to a high-SNR situation. Denote by P_e and P_{out} the FER and the outage capacity of MIMO channels, respectively. In a situation where different data rates are involved, a sequence of codes with increasing rate, rather than a single code, must be considered.

■ Spatial multiplexing gain

A *spatial multiplexing gain* r that can be achieved over a MIMO channel is given by the asymptotic (in SNR) slope of the transmission rate (for fixed FER) plotted as a function of the SNR γ in a linear-log scale; i.e.,

$$r = \lim_{\gamma\to\infty} \frac{R(\gamma)}{\log_2 \gamma} \tag{8.35}$$

where R is the rate of the code (b/s/Hz), in bits per channel use. The main rational behind such a rate normalization is the fact that spatial multiplexing gain measures how far the rate R is from the capacity. For the Rayleigh iid spatially white MIMO channel with optimal transceiver design (i.e., Gaussian code books, asymptotically large frame length, ML detection, etc), $r_{\max} = \min\{M, N\}$ indicating that for a fixed FER, the transmission rate may be increased by min{M, N} b/s/Hz for every 3 dB increase in SNR.

■ Diversity gain

A *diversity gain d* that can be achieved over a MIMO channel is given by the negative of the asymptotic (in SNR) slope of the error probability for a fixed transmission rate, plotted as a function of the SNR in a log-log scale; i.e.,

$$d = -\lim_{\gamma \to \infty} \frac{\log P_e(\gamma)}{\log \gamma} \tag{8.36}$$

where $P_e(\gamma)$ is the error probability at an SNR equal to γ. The log in (8.36) can be in any base. For the Rayleigh iid spatially white MIMO channel with optimal transceiver design, $d_{\max} = M \cdot N$, indicating that for a fixed transmission rate, with every 3 dB increase in SNR, the FER decreases by a factor of 2^{-MN}. For convenience, we rewrite (8.36) as $P_e(\gamma) \doteq \gamma^{-d}$ with = denoting exponential equality.

■ Diversity-multiplexing tradeoff

A diversity gain *d* and a spatial multiplexing gain *r* are said to be simultaneously achievable if there exists a sequence of codes, satisfying

$$R = r \log_2 \mathrm{SNR} \quad \text{and} \quad P_{\text{out}}(R) \approx \mathrm{SNR}^{-d},$$

or more precisely,

$$\lim_{\mathrm{SNR} \to \infty} \frac{\log P_{\text{out}}(r \log_2 \mathrm{SNR})}{\log \mathrm{SNR}} = -d \tag{8.37}$$

The above tradeoff characterizes the slow fading performance limit of the channel. Similarly, we can formulate a diversity-multiplexing tradeoff for any space-time coding scheme, with outage probabilities replaced by FER [Tse05].

> A space-time coding scheme is a family of codes, indexed by the SNR. It attains a multiplexing gain *r* and a diversity gain *d* if the data rates *R* (b/s/Hz) and the error probability P_e satisfy
>
> $$R = r \log_2 \mathrm{SNR} \quad \text{and} \lim_{\mathrm{SNR} \to \infty} \frac{\log P_e}{\log \mathrm{SNR}} = -d$$

DMT indicates that an increase in SNR can be utilized for some combination of transmission rate increase and FER reduction.

■ Optimal tradeoff

The following theorem from [Zheng03] derives the optimal trade-off between multiplexing and diversity gains.

Theorem 8.2: Let us assume a code at the transmitter of a MIMO channel with N transmit antennas and M receive antennas. For a given spatial multiplexing gain r, where $r = 0,1,...,\min\{M, N\}$ is an integer, the maximum diversity gain is given by

$$d^*(r) = (N - r)(M - r) \tag{8.38}$$

If the block length of the code $L \geq N + M - 1$, the optimal tradeoff curve is achieved by connecting the points $\left(r, d^*(r)\right)$ by lines.

An example of the optimal tradeoff for N=4 transmit antennas and M=3 receive antennas is depicted in Fig. 8.12.

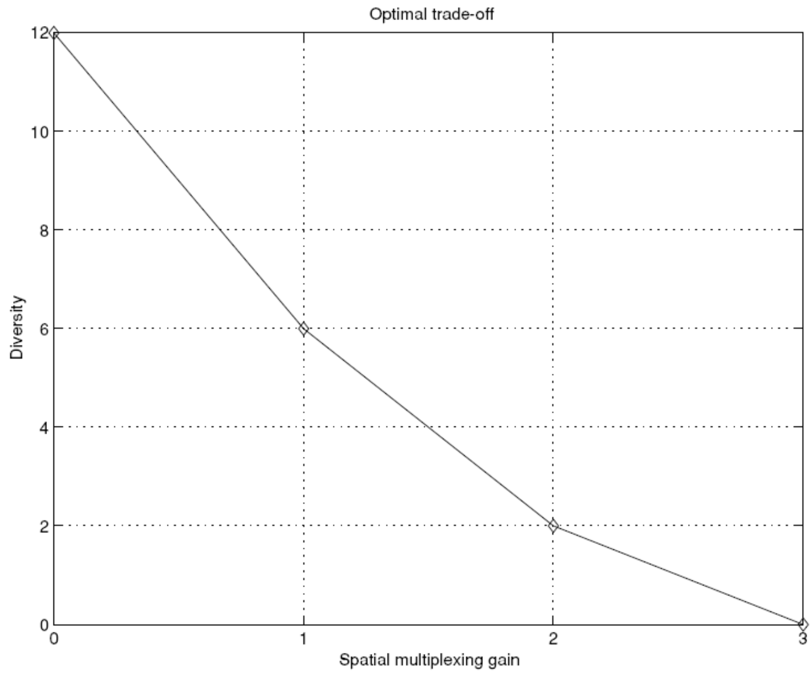


Figure 8.12

The main point here is that the maximum values of multiplexing gain and diversity gain cannot be achieved simultaneously. More generally, (8.38) shows that, out of the total number of N transmit and M receive antennas, r transmit and r receive antennas are allocated to increase the rate, and the remaining N-r and M-r create diversity, as shown in Fig.8.13. This diversity-multiplexing tradeoff curve can be used to compare different MIMO transmission schemes and to interpret their behavior.

STC设计目标：取得最佳互换。

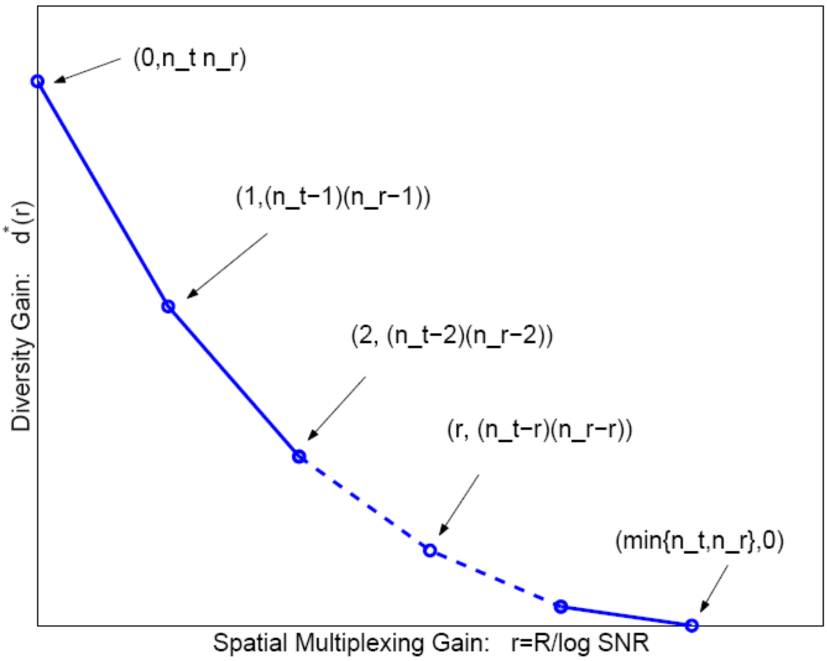


Figure 8.13 Diversity-multiplexing tradeoff, $d^*(r)$ for the i.i.d. Rayleigh fading channel.

8.6.2 Beamforming and Precoding (Known channels)

利用 Diversity 的方法: a) 发端已知信道：Precoding / beamforming; b) 发端未知信道：空时编码，靠时间来弥补。

8.6.3 Space-Time Coding (for Unknown channels at transmitter) – Transmit Diversity

Space-time coding is an effective and practical way to approach the capacity of MIMO wireless channels. It does not require CSI at the transmitter. Space-time codes (STCs) are designed based on combining error control coding and transmit diversity techniques, aiming at achieving both coding gains and diversity gains without sacrificing the bandwidth or total transmitted power. Coding is performed in both spatial and temporal domains to introduce correlation between signals transmitted from various antennas at various time periods.

The earliest form of spatial transmit diversity is the delay diversity scheme, where a signal is transmitted from one antenna, then delayed one time slot, and transmitted from the other antenna. Signal processing is used at the receiver to decode the superposition of the original and time-delayed signals. By viewing multiple-antenna diversity as independent information streams, more sophisticated coding schemes were proposed. Typical examples of STCs include space-time trellis codes (STTC), STBC, turbo STTC, etc. Fig. 8.14 shows a simple block diagram for STC.

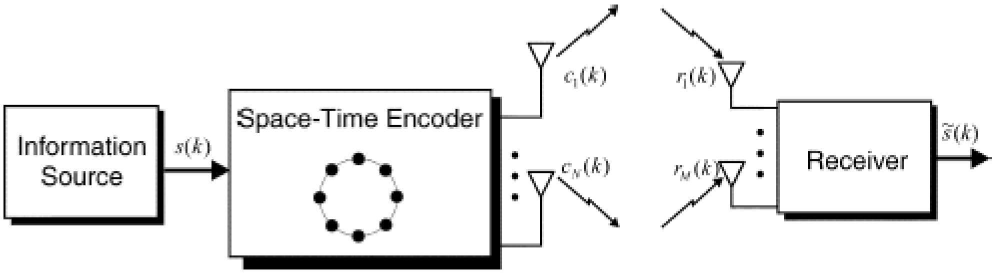


Fig. 8.14 Space-time coding

Now, STC are designed with the objective to achieve the optimal tradeoff. The tradeoff performance of specific coding schemes will be analyzed in Section x.

■ <u>Model of a space-time coded system</u>

As stated above, the MIMO channel with *N* transmit and *M* receive antennas can be represented by an (*M*×*N*) channel matrix **H**. At time *t*, the channel matrix is given by

$$\mathbf{H}_t = \begin{bmatrix} h_{1,1}^t & h_{1,2}^t & \cdots & h_{1,N}^t \\ h_{2,1}^t & h_{2,2}^t & \cdots & h_{2,N}^t \\ \vdots & \vdots & \ddots & \vdots \\ h_{M,1}^t & h_{M,1}^t & \cdots & h_{M,N}^t \end{bmatrix}$$

Assume that the fading coefficients $h_{j,i}^t$ are independent complex Gaussian random variables with mean $\mu_{j,i}$ and variance 1/2 per dimension.

Transmitter: Suppose that an input data sequence is encoded by a ST encoder into *N* parallel sequences of coded symbols. At each time instant *t*, the *N* parallel modulated symbols $x_t^1, x_t^2, \dots, x_t^N$ are simultaneously transmitted from *N* different antennas, where $x_t^i, 1 \le i \le N$, is transmitted from antenna *i*. Denote by *L* the length of output signal sequences. An *N*×*L* ST codeword matrix is given by

$$\mathbf{x} = \begin{bmatrix} x_1^1 & x_2^1 & \cdots & x_L^1 \\ x_1^2 & x_2^2 & \cdots & x_L^2 \\ \vdots & \vdots & \ddots & \vdots \\ x_1^N & x_2^N & \cdots & x_L^N \end{bmatrix}$$

Here, the *t*-th column $\mathbf{x}_t = (x_t^1, x_t^2, \dots, x_t^N)^T$ is the space-time symbol at time *t*. Denote by $\mathcal{C}$ the space-time code. In the following, we will use the space-time trellis coding (STTC) as an example to illustrate the design rule for STCs. For a STTC, the encoding process is represented by a trellis diagram.

Receiver: The received signal sequence is given by

$$\mathbf{y} = [\mathbf{y}_1, \dots, \mathbf{y}_t, \dots, \mathbf{y}_L]$$

where $\mathbf{y}_t = (y_t^1, y_t^2, ..., y_t^M)^T$, and

$$y_t^j = \sum_{i=1}^{N} h_{j,i}^t x_t^i + n_t^j, \quad \text{for j=1,2,…,M}$$

Here $n_t^j \sim \mathcal{CN}(0, N_0)$ is the noise component of receive antenna j at time t, which is assumed to be iid complex Gaussian variable with zero mean and variance $\sigma^2 = N_0 / 2$ per dimension.

Let $\mathbf{n}_t = (n_t^1, n_t^2, ..., n_t^M)^T$. In matrix form, $\mathbf{y}_t$ and $\mathbf{x}_t$ can be related by

$$\mathbf{y}_t = \mathbf{H}_t \mathbf{x}_t + \mathbf{n}_t$$

Assume that the Viterbi algorithm (VA) is used by the STTC decoder. The branch metric of the VA at time *t* is calculated as

$$\sum_{j=1}^{M} \left| y_t^j - \sum_{i=1}^{N} h_{j,i}^t x_t^i \right|^2$$

The path metric is given by

$$\sum_{t=1}^{L} \sum_{j=1}^{M} \left| y_t^j - \sum_{i=1}^{N} h_{j,i}^t x_t^i \right|^2$$

Assume that CSI is available to the receiver. Then the ML decoding corresponds to choosing the codeword **x** that minimizes the metric

$$\|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2 = \sum_{t=1}^{L} \sum_{j=1}^{M} \left| y_t^j - \sum_{i=1}^{N} h_{j,i}^t x_t^i \right|^2$$

8.6.3.1 Error Performance Analysis and Design Criteria for STCs

Problems: What is error probability of a multiple antenna system? How can we design ST codes matched to the channel structure?

■ <u>Pairwise Error Probability</u>

Assume that $\mathbf{x} = [\mathbf{x}_1, ..., \mathbf{x}_t, ..., \mathbf{x}_L]$ is the transmitted sequence, and perfect CSI is available at the receiver. A decoding error occurs if

$$\sum_{t=1}^{L} \sum_{j=1}^{M} \left| y_t^j - \sum_{i=1}^{N} h_{j,i}^t x_t^i \right|^2 \geq \sum_{t=1}^{L} \sum_{j=1}^{M} \left| y_t^j - \sum_{i=1}^{N} h_{j,i}^t \tilde{x}_t^i \right|^2$$

where $\{\tilde{\mathbf{x}}_t\} \in \mathcal{C}$ is another space-time codeword rather than **x**. The conditional pairwise error probability (i.e., the probability of transmitting **x** and deciding in favor of $\tilde{\mathbf{x}}$ is given by

$$P_2(\mathbf{x},\tilde{\mathbf{x}} \,|\, \mathbf{H}) = Q\left(\frac{d(\mathbf{x},\tilde{\mathbf{x}})}{2\sigma}\right) \le \frac{1}{2}\exp\left(-d^2(\mathbf{x},\tilde{\mathbf{x}})/4N_0\right) \tag{8.39}$$

where $N_0/2=\sigma^2$ is the noise variance per dimension, and

$$d^2(\mathbf{x},\tilde{\mathbf{x}}) = \sum_{t=1}^{L}\sum_{j=1}^{M}\left|\sum_{i=1}^{N} h_{j,i}^t\left(x_t^i-\tilde{x}_t^i\right)\right|^2 \tag{8.40}$$

- <u>Error Probability on Slow Fading Channels</u>
  On slow fading channels, the fading coefficients within each frame are constant; i.e.,

$$h_{j,i}^1 = h_{j,i}^2 = \ldots = h_{j,i}^L = h_{j,i}, \quad \text{for } 1\le i\le N, 1\le j\le M$$

Let $\mathbf{B}\left(\mathbf{X},\tilde{\mathbf{X}}\right)$ be an $N\times L$ *codeword difference matrix* defined by

$$\mathbf{B}\left(\mathbf{X},\tilde{\mathbf{X}}\right) = \mathbf{X}-\tilde{\mathbf{X}} = \begin{bmatrix} x_1^1-\tilde{x}_1^1 & x_2^1-\tilde{x}_2^1 & \cdots & x_L^1-\tilde{x}_L^1 \\ x_1^2-\tilde{x}_1^2 & x_2^2-\tilde{x}_2^2 & \cdots & x_L^2-\tilde{x}_L^2 \\ \vdots & \vdots & \ddots & \vdots \\ x_1^N-\tilde{x}_1^N & x_2^N-\tilde{x}_2^N & \cdots & x_L^N-\tilde{x}_L^N \end{bmatrix}$$

Then the *code distance matrix* $\mathbf{A}\left(\mathbf{X},\tilde{\mathbf{X}}\right)$ is an $N\times N$ matrix defined as

$$\mathbf{A}\left(\mathbf{X},\tilde{\mathbf{X}}\right) = \mathbf{B}\left(\mathbf{X},\tilde{\mathbf{X}}\right)\mathbf{B}^H\left(\mathbf{X},\tilde{\mathbf{X}}\right) \tag{8.41}$$

It is clear that $\mathbf{A}\left(\mathbf{X},\tilde{\mathbf{X}}\right)$ is Hermitian matrix, and the eigenvalues of $\mathbf{A}\left(\mathbf{X},\tilde{\mathbf{X}}\right)$ are nonnegative real numbers. Therefore, there exists an $N\times N$ unitary matrix $\mathbf{V}$ and an $N\times N$ real diagonal matrix Δ such that

$$\mathbf{V}\mathbf{A}\left(\mathbf{X},\tilde{\mathbf{X}}\right)\mathbf{V}^H = \Delta$$

Here, the rows of $\mathbf{V}$ are the eigenvectors of $\mathbf{A}\left(\mathbf{X},\tilde{\mathbf{X}}\right)$, and $\Delta = diag(\lambda_1,\lambda_2,\ldots,\lambda_N)$ with $\lambda_i\ge 0$ being the eigenvalues of $\mathbf{A}\left(\mathbf{X},\tilde{\mathbf{X}}\right)$.

Let $\mathbf{h}_j = \left(h_{j,1},h_{j,2},\ldots,h_{j,N}\right)$ and $\left(\beta_{j,1},\beta_{j,2},\ldots,\beta_{j,N}\right) = \mathbf{h}_j\mathbf{V}^H$. Then (8.40) can be written in matrix form as

$$d^2\left(\mathbf{x},\tilde{\mathbf{x}}\right) = \sum_{j=1}^{M}\mathbf{h}_j\mathbf{A}\left(\mathbf{x},\tilde{\mathbf{x}}\right)\mathbf{h}_j^H$$

$$= \sum_{j=1}^{M}\sum_{i=1}^{N}\lambda_i\,|\beta_{j,i}|^2 \tag{8.42}$$

Substituting (8.42) into (8.39) yields

$$P_2(\mathbf{x},\tilde{\mathbf{x}}\,|\,\mathbf{H}) \le \frac{1}{2}\exp\left(-\frac{1}{4N_0}\sum_{j=1}^{M}\sum_{i=1}^{N}\lambda_i\,|\beta_{j,i}|^2\right) \tag{8.43}$$

We now consider the evaluation of (8.43). First we determine the distribution of $|\beta_{j,i}|$. Since $h_{j,i} \sim \mathcal{CN}(0,1)$ are iid complex Gaussian random variables and **V** is an unitary matrix, it is obvious that $\beta_{j,i}$ are independent complex Gaussian random variables with zero mean and variance 1/2 per dimension. Thus, $|\beta_{j,i}|$ has a Rayleigh distribution with pdf

$$p(|\beta_{j,i}|) = 2|\beta_{j,i}|\exp\left(-|\beta_{j,i}|^2\right) \tag{8.44}$$

Using (8.44) and (8.43), and take expectation with respect to $|\beta_{j,i}|$, we have

$$P_2(\mathbf{x},\tilde{\mathbf{x}}) \le \int\cdots\int P_2(\mathbf{x},\tilde{\mathbf{x}}\,|\,\mathbf{H})\, p(|\beta_{1,1}|)\, p(|\beta_{1,2}|)\cdots p(|\beta_{M,N}|)\, d|\beta_{1,1}|\cdots d|\beta_{M,N}|$$

In the case of Rayleigh fading, it is given by

$$P_2(\mathbf{x},\tilde{\mathbf{x}}) \le \left(\prod_{i=1}^{N}\frac{1}{1+\lambda_i/4N_0}\right)^M$$

At high SNR's, the above upper bound can be simplified as

$$P_2(\mathbf{x},\tilde{\mathbf{x}}) \le \left(\prod_{i=1}^{r}\lambda_i\right)^{-M}\left(\frac{E_s}{4N_0}\right)^{-rM} \tag{8.45}$$

where $r=rank(\mathbf{A})$ is the rank of matrix **A**, and λ_i are the nonzero eigenvalues of matrix $\mathbf{A}(\mathbf{X},\tilde{\mathbf{X}})$, E_s is the symbol energy.

We define the *diversity gain* as an approximate measure of the gain of a system with space diversity over a reference system without space diversity. From (8.45) it is equal to rM. The diversity gain determines the slope of the PEP curve.

The *coding gain* is defined as the gain of the coded system over an uncoded system with the same diversity gain. From (8.45) the coding gain of a ST coded system is equal to $(\lambda_1\lambda_2...\lambda_r)^{1/r}$. The coding gain determines a horizontal shift of a PEP curve for a coded system relative to an uncoded system with the same diversity gain.

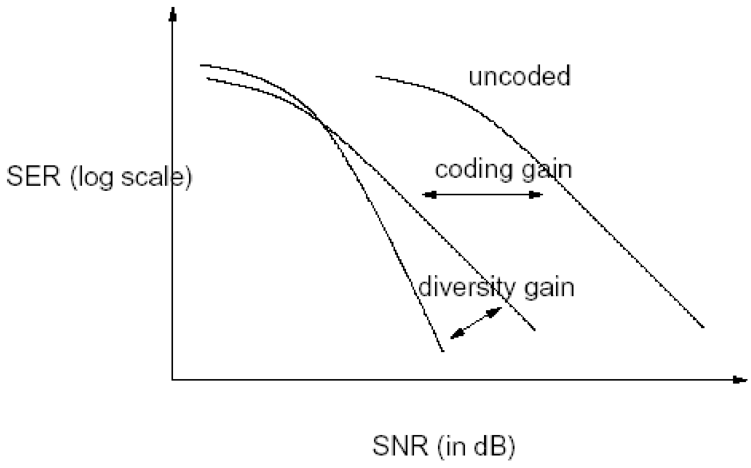


Fig. Diversity gain and coding gain

Applying a union bound on the basis of (8.45) and the code weight distribution, we may obtain bit (or frame) error probability bounds. Both bit and frame error probabilities are dominated by error paths corresponding to the code distance matrices with the minimum rank and the minimum product of eigenvalues.

Since the diversity gain is an exponent in the error probability upper bound (8.45), it is clear that achieving a large diversity gain is more important than achieving a high coding gain for systems with a small value of *rN*.

For large values of *rN*, the PEP is upper-bounded by

$$P_2(\mathbf{x},\tilde{\mathbf{x}}) \le \frac{1}{2}\exp\left(\frac{M\sum_{i=1}^{r}\lambda_i^2}{128\sigma^4} - \frac{M\sum_{i=1}^{r}\lambda_i}{8\sigma^2}\right) Q\left(\frac{\sqrt{M}\left(\sum_{i=1}^{r}\lambda_i^2 - 8\sigma^2\sum_{i=1}^{r}\lambda_i\right)}{8\sigma^2\sqrt{\sum_{i=1}^{r}\lambda_i^2}}\right) \tag{8.46}$$

- Space-Time Code Design Criteria for Slow Rayleigh Fading Channels *(for achieving coding gain & diversity gain)*

STC design criteria are based on minimizing the PEP bound stated above. We summarize them as follows.

a) Maximize the minimum rank r of the matrix $\mathbf{A}(\mathbf{x},\tilde{\mathbf{x}})$ *over all pairs of distinct codewords.*

b) Maximize the minimum product, $\prod_{i=1}^{r}\lambda_i$, *of the matrix* $\mathbf{A}(\mathbf{x},\tilde{\mathbf{x}})$ *along the pairs of distinct codewords with the minimum rank.*

Recall that $\prod_{i=1}^{r}\lambda_i$ *is the absolute value of the sum of determinants of all the principal* $r \times r$ *cofactors of matrix determinant* $\mathbf{A}(\mathbf{x},\tilde{\mathbf{x}})$, these criteria are referred to as *rank & determinant criteria*. The minimum rank of the matrix $\mathbf{A}(\mathbf{x},\tilde{\mathbf{x}})$ over all pairs of distinct codewords is called the minimum rank of the space-time code. Maximizing the minimum rank r means to make the matrix full rank such that $r=N$. However, the full rank is not always achievable due to the restriction of the trellis structure.

For large values of NM, in order to get an insight into the code design for systems of practical interest, we assume that the space-time code operates at a reasonably high SNR, which corresponds to

$$\frac{1}{4N_0} \geq \frac{\sum_{i=1}^{r}\lambda_i}{\sum_{i=1}^{r}\lambda_i^2}$$

By using the inequality $Q(x) \leq \frac{1}{2}e^{-x^2/2}, x \geq 0$, the bound (8.46) can be approximated by

$$P_2(\mathbf{x},\tilde{\mathbf{x}}) \leq \frac{1}{4}\exp\left(-\frac{N}{4N_0}\sum_{i=1}^{r}\lambda_i\right)$$

It is seen that in order to minimize the error probability, the minimum of the sum of all eigenvalues of the matrix $\mathbf{A}(\mathbf{x},\tilde{\mathbf{x}})$ among all pairs of distinct codewords should be maximized. Since, for a square matrix, the sum of all eigenvalues is equal to the trace of the matrix, we obtain

$$\mathrm{tr}\left(\mathbf{A}(\mathbf{x},\tilde{\mathbf{x}})\right) = \sum_{i=1}^{r}\lambda_i = \sum_{i=1}^{N}\sum_{t=1}^{L}|x_t^i - \tilde{x}_t^i|^2 \tag{8.47}$$

Equation (8.47) indicates that maximizing the minimum trace of the matrix $\mathbf{A}(\mathbf{x},\tilde{\mathbf{x}})$ is equivalent to maximizing the minimum Euclidean distance between all pairs of distinct codewords. This design criterion is called the *trace criterion*. In this case, the ST code design criteria can be stated as follows:

a) Make sure that the minimum rank r of the matrix $\mathbf{A}(\mathbf{x},\tilde{\mathbf{x}})$ over all pairs of distinct codewords is large enough, such that $rN \geq 4$.

b) Maximize the minimum trace $\sum_{i=1}^{r}\lambda_i$ of the matrix $\mathbf{A}(\mathbf{x},\tilde{\mathbf{x}})$ among all pairs of distinct codewords.

■ Error Probability on Fast Fading Channels

The method for PEP analysis for slow fading channels can be directly applied to fast fading channels.

A *code symbol distance matrix* $\mathbf{C}\left(\mathbf{x}_t,\tilde{\mathbf{x}}_t\right)$ at time t is an $N\times N$ matrix defined by

$$\mathbf{C}\left(\mathbf{x}_t,\tilde{\mathbf{x}}_t\right) = \left(\mathbf{x}_t - \tilde{\mathbf{x}}_t\right)\left(\mathbf{x}_t - \tilde{\mathbf{x}}_t\right)^H$$

$$= \begin{bmatrix} \left(x_t^1 - \tilde{x}_t^1\right)\left(x_t^1 - \tilde{x}_t^1\right)^* & \cdots & \left(x_t^1 - \tilde{x}_t^1\right)\left(x_t^N - \tilde{x}_t^N\right)^* \\ \vdots & & \vdots \\ \left(x_t^N - \tilde{x}_t^N\right)\left(x_t^1 - \tilde{x}_t^1\right)^* & \cdots & \left(x_t^N - \tilde{x}_t^N\right)\left(x_t^N - \tilde{x}_t^N\right)^* \end{bmatrix}$$

It is clear that the matrix $\mathbf{C}\left(\mathbf{x}_t, \tilde{\mathbf{x}}_t\right)$ is Hermitian, and there exists an unitary matrix **V**(*t*) and a real diagonal matrix **D**(*t*) such that

$$\mathbf{V}(t)\mathbf{C}\left(\mathbf{x}_t, \tilde{\mathbf{x}}_t\right)\mathbf{V}^H(t) = \mathbf{D}(t)$$

The diagonal elements of **D**(*t*), $D_t^i, i = 1, 2, ..., N$, are the eigenvalues of $\mathbf{C}\left(\mathbf{x}_t, \tilde{\mathbf{x}}_t\right)$. In the case that $\mathbf{x}_t \neq \tilde{\mathbf{x}}_t$, the matrix $\mathbf{C}\left(\mathbf{x}_t, \tilde{\mathbf{x}}_t\right)$ has only one nonzero eigenvalue and the other N-1 eigenvalues are zero. Let D_t^1 be the nonzero eigenvalue element that is equal to

$$D_t^1 = \left|\mathbf{x}_t - \tilde{\mathbf{x}}_t\right|^2 = \sum_{i=1}^{N} \left|x_t^i - \tilde{x}_t^i\right|^2$$

Let $\mathbf{h}_j(t) = \left(h_{j,1}^t, h_{j,2}^t, ..., h_{j,N}^t\right)$ and $\left(\beta_{j,1}(t), \beta_{j,2}(t), ..., \beta_{j,N}(t)\right) = \mathbf{h}_j(t)\mathbf{V}^H(t)$. Eq. (8.40) can be expressed as

$$d^2(\mathbf{x}, \tilde{\mathbf{x}}) = \sum_{t=1}^{L}\sum_{j=1}^{M}\sum_{i=1}^{N} \left|\beta_{j,i}(t)\right|^2 D_t^i$$

Since at each time t, there is only one nonzero eigenvalue, D_t^1, the above equation can be represented by

$$d^2(\mathbf{x}, \tilde{\mathbf{x}}) = \sum_{j=1}^{M}\sum_{i=1}^{N} \left|\beta_{j,i}(t)\right|^2 D_t^j$$

$$= \sum_{t \in \rho(\mathbf{x}, \tilde{\mathbf{x}})} \sum_{j=1}^{M} \left|\beta_{j,1}(t)\right|^2 \cdot \left|\mathbf{x}_t - \tilde{\mathbf{x}}_t\right|^2 \tag{8.48}$$

where $\rho(\mathbf{x}, \tilde{\mathbf{x}}) = \left\{t \middle| \left|\mathbf{x}_t - \tilde{\mathbf{x}}_t\right| \neq 0, 1 \leq t \leq L\right\}$ denotes the set of time indexes t=1,2,…L such that $\mathbf{x}_t \neq \tilde{\mathbf{x}}_t$.

Substituting (8.48) into (8.39), we obtain

$$P_2\left(\mathbf{x}, \tilde{\mathbf{x}} \mid \mathbf{H}\right) \leq \frac{1}{2}\exp\left(-\frac{1}{4N_0}\sum_{t \in \rho(\mathbf{x}, \tilde{\mathbf{x}})}\sum_{j=1}^{M} \mid \beta_{j,1}(t) \mid^2 \mid \mathbf{x}_t - \tilde{\mathbf{x}}_t \mid^2\right) \tag{8.49}$$

Note that $\beta_{j,i}^t$ are independent complex Gaussian random variables with zero mean and variance 1/2 per dimension. Therefore, $|\beta_{j,i}^t|$ follows a Rayleigh distribution. Take expectation of (8.49) with respect to $|\beta_{j,i}^t|$, the PEP at high SNR's is upper-bounded by

$$P_2(\mathbf{x},\tilde{\mathbf{x}}) \le \prod_{t\in\rho(\mathbf{x},\tilde{\mathbf{x}})} \left(\frac{1}{1+|\mathbf{x}_t-\tilde{\mathbf{x}}_t|^2/4N_0}\right)^M$$

$$\le \left(d_\rho^2\right)^{-M}\left(\frac{E_s}{4N_0}\right)^{-\delta_H M} \tag{8.50}$$

where $d_\rho^2=\prod_{t\in\rho}|\mathbf{x}_t-\tilde{\mathbf{x}}_t|^2$ is the product of the squared Euclidean distance between two space-time symbol sequences, and δ_H is the ST symbol Hamming distance which is defined as the number of ST symbols in which two codewords $\mathbf{x}$ and $\tilde{\mathbf{x}}$ differ.

■ Space-Time Code Design Criteria for Fast Rayleigh Fading Channels

Based on (8.50), we can summarize the code design criteria for fast Rayleigh fading below.

a) Maximize the minimum space-time symbol Hamming distance δ_H between all pairs of distinct codewords;

b) Maximize the minimum product distance d_ρ^2 along the path with the minimum symbol Hamming distance δ_H.

References

[1] E. Biglieri, J. Proakis, and S. Shamai (Shitz), "Fading channels: Information-theoretic and communication aspects," *IEEE Trans. Inform. Theory*, vol.44, pp.2619-2692, Oct. 1998.

[2] E. Telatar, "Capacity of multi-antenna Gaussian channels," *European. Trans. Telecomm.*, vol.10, no.6, pp.585-596, Nov. 1999.

[3] G. J. Foschini and M. J. Gans "On limits of wireless communications in a fading environment when using multiple antennas," Wireless Personal Commun., vol.6, pp.311-335, Mar. 1998.

[4] Special issue on MIMO systems and applications. Part I. *IEEE J-SAC*, vol.21, no.3, April 2003.

[5] Special issue on MIMO systems and applications. Part II. *IEEE J-SAC*, vol.21, no.5, June 2003.

[6] Special issue on Space-time transmission, reception, coding and signal processing. *IEEE Trans. Inform. Theory*, vol.49, no.10, Oct. 2003.

[7] Special issue on Space-time signal processing. *IEEE Trans. Signal Processing*, vol.50,

Oct. 2002.
[8] Special issue on Space-time signal processing. *IEEE Trans. Signal Processing*, vol.51, Nov. 2003.
[9] S. N. Diggavi, N. A.-Dhahie, A. Stamoulis and A. R. Calderbank, “Great expectations: The value of spatial diversity in wireless networks,” Proceedings of the IEEE, vol.92, no.2, pp.219-292, Feb. 2004.
[10] A. J. Paulraj, D. A. Gore, R. U. Nabar, and H. Bolcskel, “An Overview of MIMO Communications—A Key to Gigabit Wireless,” Proceedings of the IEEE, vol.92, no.2, pp. 198-218, Feb. 2004.
[11] B. Vucetic and Jinhong Yuan, *Space-Time Coding*. John Wiley & Sons, Ltd., 2003.
[12] L. Zheng and D. N. C. Tse, “Diversity and Multiplexing: A fundamental tradeoff in multiple-antenna channels,” *IEEE Trans. Inform. Theory*, vol.49, no.10, pp. 1073-1096, Oct. 2003.
[13] T. H. Liew and L. Hanzo, “Space-time codes and concatenated channel codes for wireless communications,” Proceedings of IEEE, vol.90, no.2, pp.187-219, Feb. 2002.
[14] V. Tarokh, N. Seshadri, and A. R. Calderbank, “Space-time codes for high data rate wireless communication: Performance criterion and code construction,” *IEEE Trans. Inform. Theory*, vol.44, no.2, pp. 744-765, Mar. 1998.
[15] M. Jankiraman, *Space-Time Codes and MIMO Systems*. Artech House, 2004.
[16] A. J. Paulraj, Rohit Nabar, and D. Gore, *Introduction to Space-Time Wireless Communications*, Cambridge, UK: Cambridge University Press, 2003.
[17] T. Marzetta, and B. Hochwald, ‘‘Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading,’’ *IEEE Trans. Inf. Theory*, 45(1), January 1999, pp. 139–157.