

MIMO Channel

1. 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.1. The transmitted signal at time t is represented by an $N \times 1$ column vector $\mathbf{x} \in \mathbb{C}^N$, and the received signal is represented by an $M \times 1$ column vector $\mathbf{y} \in \mathbb{C}^M$ (For simplicity, we ignore the time index). The discrete-time MIMO channel can be described by

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n} \quad (1)$$

where \mathbf{H} is an $M \times N$ 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 $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, N_0 \mathbf{I}_M)$ 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

$$E[\|\mathbf{x}\|^2] = E[\mathbf{x}^H \mathbf{x}] = E[\text{Tr}(\mathbf{x}\mathbf{x}^H)] = \text{Tr}\{E[\mathbf{x}\mathbf{x}^H]\} = \text{tr}(\mathbf{K}_x) \leq P,$$

where $\mathbf{K}_x = E[\mathbf{x}\mathbf{x}^H]$ is the covariance matrix of the transmitted signal \mathbf{x} .

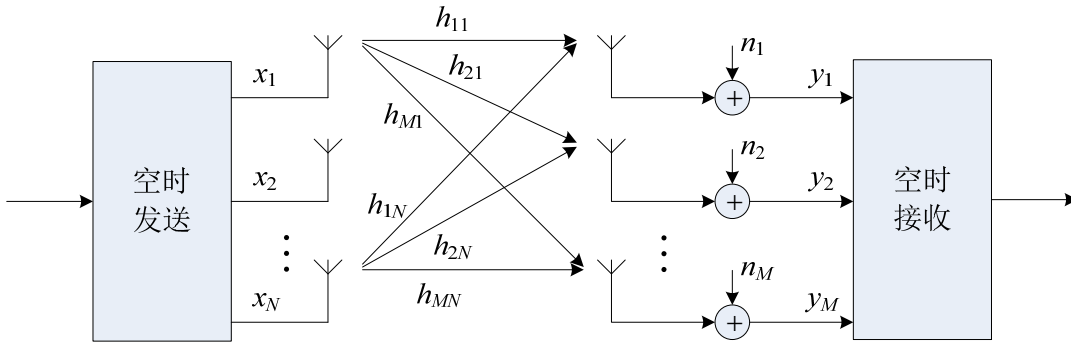


Fig. 1 An MIMO wireless system model

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 \mathbf{H} is deterministic, we have

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

When \mathbf{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 \mathbf{H} is given by

$$\sum_{n=1}^N \mathbb{E}[|h_{nm}|^2] = N, \quad m = 1, 2, \dots, M$$

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 $\text{SNR} = P/N_0$.

2. Fundamental Capacity Limits of MIMO Channels

Consider the case of deterministic \mathbf{H} . The channel matrix \mathbf{H} is assumed to be constant at all time and known to the receiver. The relation of (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 \mathbf{x} and \mathbf{y} as

$$\begin{aligned} C(\mathbf{H}) &= \max_{p(\mathbf{x}) : \mathbb{E}[\|\mathbf{x}\|^2] \leq P} \mathcal{I}(\mathbf{x}; \mathbf{y}, \mathbf{H}) = \max_{p(\mathbf{x})} \mathcal{I}(\mathbf{x}; \mathbf{H}) + \mathcal{I}(\mathbf{x}; \mathbf{y} | \mathbf{H}) \\ &= \max_{p(\mathbf{x})} \mathcal{I}(\mathbf{x}; \mathbf{y} | \mathbf{H}) = \max_{p(\mathbf{x})} [\mathcal{H}(\mathbf{y} | \mathbf{H}) - \mathcal{H}(\mathbf{y} | \mathbf{x}, \mathbf{H})] \end{aligned} \quad (0.1)$$

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

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

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 (1), the covariance matrix of \mathbf{y} is

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

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

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

The mutual information then reduces to

$$\begin{aligned} \mathcal{I}(\mathbf{x}; \mathbf{y} | \mathbf{H}) &= \mathcal{H}(\mathbf{y} | \mathbf{H}) - \mathcal{H}(\mathbf{n}) \\ &= \log_2 \left(\det \left(\mathbf{I}_M + \frac{1}{N_0} \mathbf{H} \mathbf{K}_x \mathbf{H}^H \right) \right) \end{aligned} \quad (0.2)$$

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 (0.2) over all input covariance matrices \mathbf{K}_x satisfying the power constraint:

$$\begin{aligned} C(\mathbf{H}) &= \max_{\mathbf{K}_x: \text{Tr}(\mathbf{K}_x) = P} \log_2 \left(\det \left(\mathbf{I}_M + \frac{1}{N_0} \mathbf{H} \mathbf{K}_x \mathbf{H}^H \right) \right) \text{ bits per channel use} \quad (0.3) \\ &= \max_{\mathbf{K}_x: \text{Tr}(\mathbf{K}_x) = P} \log_2 \left(\det \left(\mathbf{I}_N + \frac{1}{N_0} \mathbf{K}_x \mathbf{H}^H \mathbf{H} \right) \right) \end{aligned}$$

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 \mathbf{K}_x will depend on whether or not \mathbf{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.

2.3 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 \mathbf{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 $\mathbf{K}_x = \frac{P}{N} \mathbf{I}_N$. It is shown in [Telatar99] that this \mathbf{K}_x indeed maximize the mutual information. Thus, the capacity in such a case is

$$C = \begin{cases} \log_2 \left[\det \left(\mathbf{I}_M + \frac{\text{SNR}}{N} \mathbf{H} \mathbf{H}^H \right) \right], & \text{if } M < N \\ \log_2 \left[\det \left(\mathbf{I}_N + \frac{\text{SNR}}{N} \mathbf{H}^H \mathbf{H} \right) \right], & \text{if } M \geq N \end{cases} \text{ bits per channel use} \quad (0.4)$$

where $\text{SNR} = P / N_0$.

2.1 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

$$\mathbf{H} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^H \quad (0.5)$$

where $\mathbf{\Lambda}$ is an $M \times N$ non-negative real and diagonal matrix, \mathbf{U} and \mathbf{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 $\mathbf{H}\mathbf{H}^H$, the columns of \mathbf{U} are the eigenvectors of $\mathbf{H}\mathbf{H}^H$ and the columns of \mathbf{V} are the eigenvectors of $\mathbf{H}^H\mathbf{H}$.

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

$$\mathbf{H}\mathbf{H}^H \mathbf{z} = \lambda \mathbf{z}, \quad \mathbf{z} \neq \mathbf{0} \quad (0.6)$$

where \mathbf{z} is an $M \times 1$ eigenvector corresponding to λ . The number of non-zero eigenvalues of matrix $\mathbf{H}\mathbf{H}^H$ is equal to the rank r of \mathbf{H} . Since the rank of \mathbf{H} cannot exceed the number of columns or rows of \mathbf{H} , $r \leq m = \min(M, N)$. If \mathbf{H} is full rank, which is sometimes referred to as a *rich scattering environment*, then $r = m$. Equation (0.6) can be rewritten as

$$(\lambda \mathbf{I}_m - \mathbf{W})\mathbf{z} = \mathbf{0}, \quad \mathbf{z} \neq \mathbf{0} \quad (0.7)$$

where \mathbf{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

$$\det(\lambda \mathbf{I}_m - \mathbf{W}) = 0 \quad (0.8)$$

The m nonzero eigenvalues of \mathbf{W} , $\lambda_1, \lambda_2, \dots, \lambda_m$, can be calculated by finding the roots of (0.8). The non-negative square roots of the eigenvalues of \mathbf{W} are also referred to as *singular values* of \mathbf{H} .

Substituting (0.5) into (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 (1) is equivalent to the channel

$$\tilde{\mathbf{y}} = \mathbf{\Lambda} \tilde{\mathbf{x}} + \tilde{\mathbf{n}} \quad (0.9)$$

where $\mathbf{\Lambda} = \text{diag}(\sqrt{\lambda_1}, \sqrt{\lambda_2}, \dots, \sqrt{\lambda_m}, 0, \dots, 0)$ with $\sqrt{\lambda_i}, i = 1, 2, \dots, m$ denoting the non-zero

singular values of \mathbf{H} . The equivalence is summarized in Fig. 2. From (0.9), we obtain for the received signal components

$$\begin{aligned}\tilde{y}_i &= \sqrt{\lambda_i} \tilde{x}_i + \tilde{n}_i, \quad 1 \leq i \leq m \\ \tilde{y}_i &= \tilde{n}_i, \quad m+1 \leq i \leq M\end{aligned}\quad (0.10)$$

It is seen that received components $\tilde{y}_i, i > m$, do not depend on the transmitted signal. On the other hand, received components $\tilde{y}_i, i = 1, 2, \dots, m$ depend only the transmitted component \tilde{x}_i . Thus the equivalent MIMO channel in (0.9) can be considered as consisting of m uncoupled parallel Gaussian sub-channels. Specifically,

- If $N > M$, (0.10) indicates that there will be at most M non-zero attenuation subchannels in the equivalent MIMO channel. See Fig. 3.
- If $M > N$, there will be at most N non-zero attenuation subchannels in the equivalent MIMO channel.

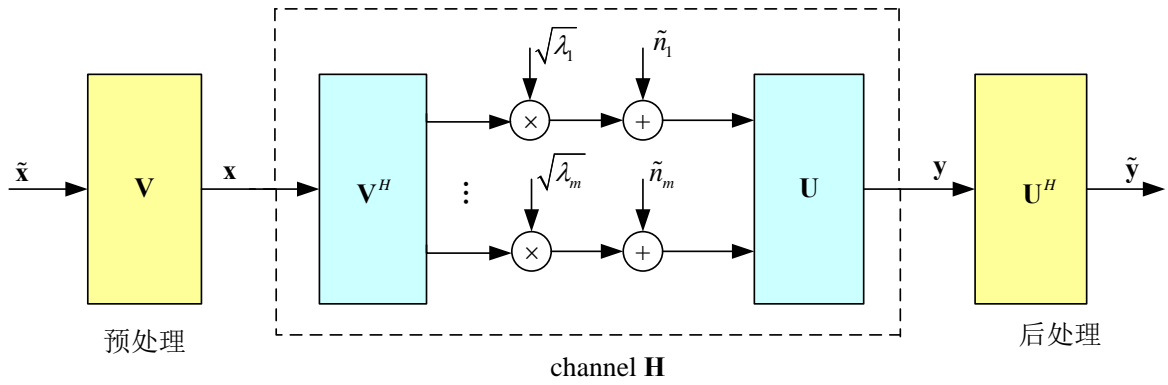


Figure 2 Converting the MIMO channel into a parallel channel through the SVD.

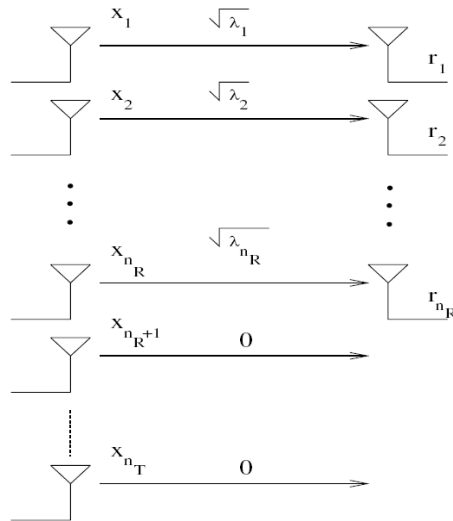


Fig. 3 Block diagram of an equivalent MIMO channel for $N > M$

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 $\mathbf{H}\mathbf{H}^H$ as follows.

2.2 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 (0.3). Substituting the matrix SVD (0.5) into (0.3) and using properties of unitary matrices, we get the MIMO capacity with CSIT and CSIR as

$$\begin{aligned} C &= \max_{\mathbf{K}_x: \text{Tr}(\mathbf{K}_x) = P} \log_2 \left(\det \left(\mathbf{I}_N + \frac{1}{N_0} \mathbf{\Lambda} \mathbf{K}_x \mathbf{\Lambda}^H \right) \right) \\ &= \max_{P_i: \sum_i P_i \leq P} \sum_{i=1}^m \log_2 \left(1 + \frac{P_i \lambda_i}{N_0} \right) \end{aligned}$$

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

$$P_i = \left(\mu - \frac{N_0}{\lambda_i} \right)^+ \quad (0.11)$$

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

$$\sum_{i=1}^m P_i = P \quad (0.12)$$

The resulting capacity is then

$$C_{\text{WF}} = \sum_{i=1}^m \log_2 \left[1 + \frac{1}{N_0} (\mu \lambda_i - N_0)^+ \right] = \sum_{i=1}^m \log_2 \left(\frac{1}{N_0} \mu \lambda_i \right)^+ \quad \text{bits/channel use} \quad (0.13)$$

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

$$\mathbf{K}_x = \mathbf{V} \mathbf{P} \mathbf{V}^H$$

where $\mathbf{P} = \text{diag}(P_1, P_2, \dots, P_m, 0, \dots, 0)$ is an $N \times N$ matrix.

■ Water-filling algorithm:

The power allocation in (0.11) 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 (0.11) into (0.12)) as

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

Then we have

$$\mu = \frac{1}{m-p+1} \left(P + N_0 \sum_{i=1}^{m-p+1} \frac{1}{\lambda_i} \right) \quad (0.15a)$$

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

$$P_i = \left(\mu - \frac{N_0}{\lambda_i} \right), \quad i = 1, 2, \dots, m-p+1 \quad (0.15b)$$

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(0.15a)和(0.15b)，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。