# MIMO Systems and Transmit Diversity

## 1 Introduction

So far we have investigated the use of antenna arrays in interference cancellation and for receive diversity. This final chapter takes a broad view of the use of antenna arrays in wireless communications. In particular, we will investigate the *capacity* of systems using multiple transmit and/or multiple receive antennas. This provides a fundamental limit on the data throughput in multiple-input multiple-output (MIMO) systems. We will also develop the use of transmit diversity, i.e., the use of multiple transmit antennas to achieve reliability (just as earlier we used multiple receive antennas to achieve reliability via receive diversity).

The basis for receive diversity is that each element in the receive array receives an *independent* copy of the same signal. The probability that all signals are in deep fade simultaneously is then significantly reduced. In modelling a wireless communication system one can imagine that this capability would be very useful on transmit as well. This is especially true because, at least in the near term, the growth in wireless communications will be asymmetric internet traffic. A lot more data would be flowing from the base station to the mobile device that is, say, asking for a webpage, but is receiving all the multimedia in that webpage. Due to space considerations, it is more likely that the base station antenna comprises multiple elements while the mobile device has only one or two.

In addition to providing diversity, intuitively having multiple transmit/receive antennas should allow us to transmit data faster, i.e., increase data throughput. The information theoretic analysis in this chapter will formalize this notion. We will also introduce a multiplexing scheme, transmitting multiple data streams to a single user with multiple transmit and receive antennas.

This chapter is organized as follows. Section 2 then presents a theoretical analysis of the capacity of MIMO systems. The following two sections, Sections 3 develops transmit diversity techniques for MIMO systems based on space-time coding. Section 4 then addresses the issue of maximizing data throughput while also providing reliability. We will also consider transmitting multiple data streams to a single user. This chapter ends in Section 5 with stating the fundamental tradeoff between data throughput (also called multiplexing) and diversity (reliability).

## 2 MIMO Capacity Analysis

Before investigating MIMO capacity, let us take a brief look at the capacity of single-input single-output (SISO) fading channels. We start with the original definition of capacity. This set of

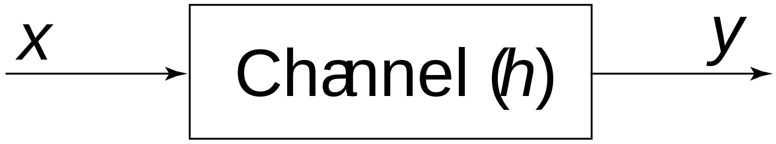


Figure 1: A single-input-single-output channel

notes assumes the reader knows the basics of information theory. See [1] for a detailed background. Consider the input-output system in Fig. 1. The capacity of the channel is defined as the maximum possible mutual information between the input ($x$) and output ($y$). The maximization is over the probability distribution of the input $f_X(x)$, i.e.

$$C = \max_{f_X(x)} [I(X;Y)] = \max_{f_X(x)} \left[h(Y) - h(Y/X)\right], \tag{1}$$

where $h(Y)$ is the entropy of the output $Y$.

For a SISO additive white gaussian noise (AWGN) channel, $y = x + n$, with $n \sim \mathcal{CN}(0, \sigma^2)$ and with limited input energy ($E\left\{|x|^2\right\} \leq E_s$), one can show that the capacity achieving distribution is Gaussian, i.e., $x \sim \mathcal{CN}(0, E_s)$ and $y \sim \mathcal{CN}(0, E_s + \sigma^2)$. It is not difficult to show that if $n$ is Gaussian and has variance $\sigma^2$, $h(N) = \log_2(\pi e \sigma^2)$. Therefore $h(Y) = \log_2(\pi e(E_s + \sigma^2))$. Also, $h(Y/X)$ is the residual entropy in $Y$ given the channel input $X$, i.e., it is the entropy in the noise term $N$. Therefore, $h(Y/X) = \log_2(\pi e \sigma^2)$ and the channel capacity, in bits/s/Hz, is given by

$$C = [h(Y) - h(Y/X)] = \log_2\left(\frac{E_s + \sigma^2}{\sigma^2}\right) = \log_2\left(1 + \rho\right), \tag{2}$$

where $\rho = E_s/\sigma^2$ is the signal-to-noise ratio (SNR).

In the case of a fading SISO channel, the received signal at the $k$-th symbol instant is $y[k] = h[k]x[k] + n[k]$. To ensure a compatible measure of power, set $\mathrm{E}\left\{|h[k]|^2\right\} = 1$ and $\mathrm{E}\left\{|x[k]|^2\right\} \leq E_s$. At this point there are two possibilities, a fixed fading channel with a random but unchanging channel gain and a slow, but fluctuating channel. In the first case, the capacity is given by

$$C = \log_2\left(1 + |h|^2 \rho\right), \tag{3}$$

where $\rho = E_s/\sigma^2$. An interesting aspect of this equation is that in a random, but fixed channel, the theoretical capacity may be zero. This is because, theoretically, the channel gain could be as close to zero making *guaranteeing* a data rate impossible. What is possible in this case is determining what are the chances a required capacity is available. This requires defining a new probability of outage, $P_{\text{out}}$, as the probability that the channel capacity is below a threshold rate $R_0$.

$$\begin{aligned} P_{\text{out}} &= P(C < R_0) = P\left(|h|^2 > \frac{2^{R_0} - 1}{\rho}\right), & (4) \\ &= 1 - \exp\left\{-\frac{2^{R_0} - 1}{\rho}\right\}, & (5) \end{aligned}$$

where the final equation is valid for Rayleigh fading. Note that in the high-SNR regime ($\rho \to \infty$),

$$P_{\text{out}} \propto \frac{1}{\rho}, \tag{6}$$

i.e., at high SNR, the outage probability falls off inversely with SNR.

In the case of a time varying channel, assuming sufficient interleaving that the channel is independent from one symbol to the next, the average capacity over $K$ channel realizations is

$$C_K = \frac{1}{K} \sum_{k=1}^{K} \left\{ \log_2 \left( 1 + |h_k|^2 \rho \right) \right\}. \tag{7}$$

Based on the law of large numbers, as $K \to \infty$ the term on the right converges to the average or expected value. Hence,

$$C = \mathrm{E}_h \left\{ \log_2 \left( 1 + |h|^2 \rho \right) \right\}, \tag{8}$$

where the expectation operation is taken over the channel values $h$. Note that this expression is non-zero and therefore with a *fluctuating* channel it is possible to guarantee the existence of an error-free data rate.

## 2.1 MIMO Systems

We now consider MIMO systems with the goal of evaluating the capacity of a system using $N$ transmit and $M$ receive antennas. We begin with the case of $N$ parallel channels - basically $N$ SISO channels operating in parallel. However, we will assume that the transmitter knows the $N$ channels and can therefore *allocate power* intelligently to maximize capacity.

### 2.1.1 Parallel Channels

The $N$ parallel channels are AWGN with a noise level of $\sigma^2$. The received data ($\mathbf{y}$) from input data $\mathbf{x}$ over $N$ channels is modelled as

$$\begin{aligned} \mathbf{y} &= \mathbf{x} + \mathbf{n} && (9) \\ \mathrm{E}\{\mathbf{n}\mathbf{n}^H\} &= \sigma^2 \mathbf{I}_N. && (10) \end{aligned}$$

The transmitter has an energy budget of $E_s$ which must be allocated across the $N$ channels. The capacity of this channel is

$$C = \max_{\{E_n\} \sum_{n=1}^{N} E_n \leq E_s, E_n \geq 0} \sum_{n=1}^{N} \log_2 \left( 1 + \frac{E_n |h_n|^2}{\sigma_n^2} \right), \tag{11}$$

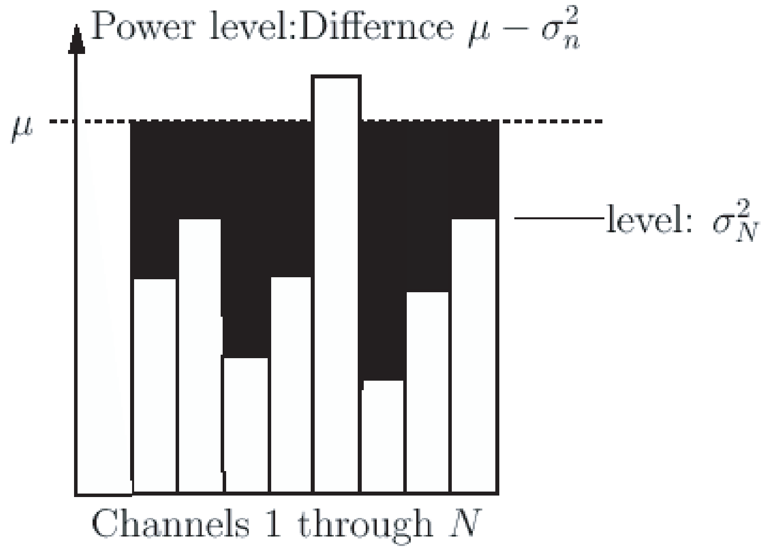


Figure 2: Illustrating Waterfilling.

where $E_n$ is the energy allocated to the $n^{\text{th}}$ channel. The equivalent Lagrange problem is[1]:

$$L(\{E_n\};\lambda) = \sum_{n=1}^{N}\log_2\left(1+\frac{E_n|h_n|^2}{\sigma^2}\right)+\lambda\left(\sum_{n=1}^{N}E_n - E_s\right) \tag{12}$$

$$\Rightarrow \frac{\partial L}{\partial E_n} = \frac{|h_n|^2}{\sigma^2}\frac{\log_2(e)}{\left(1+\frac{E_n|h_n|^2}{\sigma^2}\right)}+\lambda = 0, \tag{13}$$

$$\Rightarrow \forall n, \qquad \left(\frac{\sigma^2}{|h_n|^2}+E_n\right) = \mu \quad \text{(a constant)}. \tag{14}$$

Since $E_n \geq 0$,

$$E_n = \left(\mu - \frac{\sigma^2}{|h_n|^2}\right)^+, \tag{15}$$

where $(x)^+$ indicates only positive numbers are allowed, i.e. $(x)^+ = x$ if $x \geq 0$, else $(x)^+ = 0$. The constant $\mu$ is chosen to meet the total energy constraint. Equation (15) tells us how to allocate energy *given* knowledge of the channel attenuation through which the data must suffer.

Interestingly, the optimal power allocation scheme does *not* allocate all the power to the best channel. This is because the $\log_2(1+\rho)$ expression for capacity implies a diminishing marginal returns on adding signal power (the capacity grows only as $\log_2$ at high SNR, but linearly at low-SNR). So providing some power to weaker channels can actually increase overall sum capacity.

[1]Note that the Lagrange problem being set up ignores the constraint that $E_n \geq 0$ for now and that this constraint is "added" later. A formal proof that this is OK will take us into a detour. The proof uses the fact that if we were to add this constraint ($N$ of them), the associated Lagrange multiplier is either zero or the constraint is not met with equality.

This optimal scheme is known as *waterfilling.* An intuitive understanding of waterfilling (and why it is called so) may be obtained from Fig. 2, borrowed from Prof. Schlegel [2]. In the figure, $\sigma_n^2$ refers to the effective noise power at each time instant, $\sigma^2/\left|h_n\right|^2$. Waterfilling tells us that the optimal strategy is to 'pour energy' (allocate energy on each channel). In channels with lower noise power, more energy will be allocated. In channels with large noise power, the energy allocated is low . Some channels are so weak that the effective noise power becomes very large. Waterfilling tells us that transmitting any information on these channels is a waste of energy. If energy is allocated, the sum of the allocated energy and the effective noise power ($\sigma_n^2 = \sigma^2/|h_n^2|$) is a constant (the "water level", $\mu$). Finally, if the channel were all equal, i.e. $\sigma_n^2$ were a constant, waterfilling leads to an equal energy distribution. Determining the water level, $\mu$, is an iterative process.

The capacity on using the waterfilling approach is

$$C = \sum_{n=1}^{N} \log_2\left(1 + \frac{E_n\left|h_n\right|^2}{\sigma^2}\right). \tag{16}$$

*Aside:* The result also leads to an interesting observation: if one could only focus on the times that the channel is in a "good" condition one could get enormous gains in capacity. Of course, this may not always be possible. However, thinking of a *multiuser* situation, if the channel to each user is changing with time, it is likely that at any time instant, one user has a good channel. By transmitting energy on that channel, overall capacity can be achieved in a *multiuser situation.* This is a new form of diversity called "opportunistic beamforming" [3].

Finally, if the channel is *not available* at the transmitter, clearly the best distribution scheme is to spread the energy evenly between all transmitters, i.e. $E_n = E_s/N$ and

$$C = \sum_{n=1}^{N} \log_2\left(1 + \frac{E_s}{N\sigma^2}\right). \tag{17}$$

Note that since the log function increases significantly slower than the linear $N$ term, the overall capacity is significantly larger than that for the SISO case.

### 2.1.2 Known MIMO Channels

We now turn to the more practical MIMO situation with $N$ transmitters and $M$ receivers with a full $M \times N$ channel matrix $\mathbf{H}$ in between. We will assume we know the channel matrix $\mathbf{H}$ at both the transmitter and receiver. Also, we will set $M \leq N$, however, the results here are easily extended for $M > N$. To ensure no artificial amplification in the channel, we shall set $\mathrm{E}\{\left|h_{mn}\right|^2\} = 1$. The data received at the $M$ elements can be modelled as

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}, \tag{18}$$

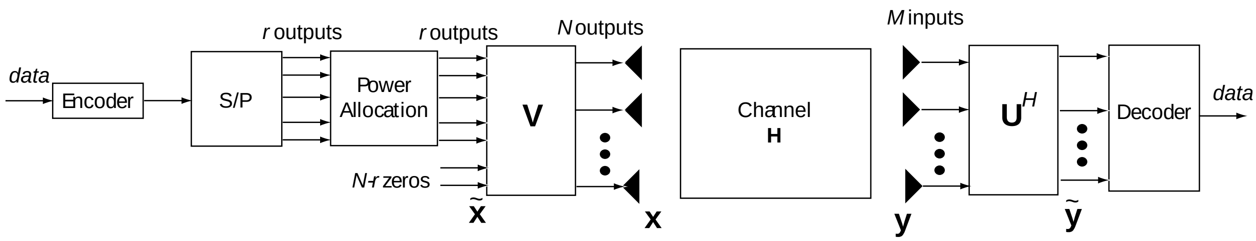


Figure 3: A communication system that achieves capacity.

where $\mathbf{H}$ is the full $M \times N$ channel matrix.

Based on the singular value decomposition[2] , one can decompose $\mathbf{H}$ as $\mathbf{H} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^H$, with $\boldsymbol{\Sigma} = [\mathrm{diag}(d_1, d_2, \ldots, d_M) \,|\, \mathbf{0}_{M \times N-M}]$, where $d_m \geq 0$ are the $M$ singular values of $\mathbf{H}$. Using Eqn. (18) and the fact that $\mathbf{U}^H\mathbf{U} = \mathbf{I}_M$,

$$\begin{aligned} \mathbf{y} &= \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^H\mathbf{x} + \mathbf{n}, && (19) \\ \Rightarrow \mathbf{U}^H\mathbf{y} &= \boldsymbol{\Sigma}\mathbf{V}^H\mathbf{x} + \mathbf{U}^H\mathbf{n}, && (20) \\ \Rightarrow \tilde{\mathbf{y}} &= \boldsymbol{\Sigma}\tilde{\mathbf{x}} + \tilde{\mathbf{n}}, && (21) \end{aligned}$$

where $\tilde{\mathbf{y}} = \mathbf{U}^H\mathbf{y}$ and $\tilde{\mathbf{x}} = \mathbf{V}^H\mathbf{x}$. *This transformed data in Eqn. (21) is equivalent to $M$ parallel channels* with effective noise powers of $\sigma_m^2 = \sigma^2/d_m^2$. Note that if $\mathrm{E}\{\mathbf{n}\mathbf{n}^H\} = \sigma^2\mathbf{I}$, $\mathrm{E}\{\tilde{\mathbf{n}}\tilde{\mathbf{n}}^H\} = \mathrm{E}\{\mathbf{U}^H\mathbf{n}\mathbf{n}^H\mathbf{U}\} = \sigma^2\mathbf{U}^H\mathbf{I}\mathbf{U} = \sigma^2\mathbf{I}$. Furthermore, since $\mathbf{V}^H\mathbf{V} = \mathbf{I}_N$, the energy constraint remains the same, i.e., $\sum_{n=1}^N \tilde{E}_n = E_s$. Since the last $(N-M)$ columns of $\boldsymbol{\Sigma}$ are all zero, the last $(N-M)$ entries in $\tilde{\mathbf{x}}$ are irrelevant. In fact, if the rank of $\mathbf{H}$ is $r$, the system is equivalent to $r$ parallel channels only. Note that $r \leq \min(N, M)$.

In the rotated (tilde) space MIMO communications is exactly the same as $r$ parallel channels. The optimal power allocation is, therefore, the same waterfilling scheme as with the $N$ parallel channels in Section 2.1.1. However, now the energy is spread over the eigen-channels, as opposed to physical channels. Figure 3 illustrates the communication system being considered. The data to be transmitted in encoded (if the encoder achieves capacity in an AWGN channel the overall scheme achieves channel capacity) and sent onto a serial-to-parallel converter with $r$ outputs, where $r$ is the rank of the channel matrix. *The waterfilling scheme is used to determine the powers of each element in these $r$ outputs.* The $r$ outputs are augmented with $(N-r)$ zeros to form the data vector $\tilde{\mathbf{x}}$. Multiplying with the right singular vector matrix $\mathbf{V}$ leads to the data vector $\mathbf{x}$ to be transmitted over the $N$ elements. This transmission suffers channel $\mathbf{H}$. At the receiver the

[2]Any $M \times N$ matrix $\mathbf{A}$ can be decomposed as $\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^H$. The columns of $\mathbf{U}$ are the $M$ eigenvectors of $\mathbf{H}\mathbf{H}^H$ and the columns of $\mathbf{V}$ are the $N$ eigenvectors of $\mathbf{H}^H\mathbf{H}$. The $M \times N$ matrix $\boldsymbol{\Sigma}$ is a diagonal matrix of singular values. If $M \leq N$, $\boldsymbol{\Sigma} = [\,\mathrm{diag}(\sigma_1, \sigma_2, \ldots, \sigma_M) \,|\, \mathbf{0}_{M\times N-M}]$ where $\sigma_m^2$ are the $M$ eigenvalues of $\mathbf{H}\mathbf{H}^H$. Note that this is for an arbitrary rectangular matrix $\mathbf{A}$ and these singular values should not be confused with the noise power. Since $\mathbf{H}\mathbf{H}^H$ and $\mathbf{H}^H\mathbf{H}$ are positive semi-definite matrices, $\mathbf{U}\mathbf{U}^H = \mathbf{U}^H\mathbf{U} = \mathbf{I}_M$, $\mathbf{V}\mathbf{V}^H = \mathbf{V}^H\mathbf{V} = \mathbf{I}_N$ and $\sigma_m \geq 0$. The matrix $\mathbf{U}$ ($\mathbf{V}$) is the matrix of left (right) singular vectors.

length-$M$ data vector $\mathbf{y}$ is multiplied by the left singular vectors ($\mathbf{U}^H$) resulting in the transformed vector $\tilde{\mathbf{y}}$. This transformed vector is used for decoding the original data symbols.

The optimal energy distribution $\tilde{E}_m$ on the $m$-th channel and overall capacity are given by

$$\tilde{E}_m = \left(\mu - \frac{\sigma^2}{d_m^2}\right)^+, \tag{22}$$

$$C = \sum_{m=1}^{R} \log_2\left(1 + \frac{\tilde{E}_m d_m^2}{\sigma^2}\right). \tag{23}$$

To illustrate the workings of this capacity formula, let us consider four examples:

*Case 1*: 1 transmitter and $M$ receivers, $\mathbf{H} = [h_1, h_2, \ldots, h_M]^T$, $\text{rank}(\mathbf{H}) = 1$.

Since $\text{rank}(\mathbf{H}) = 1$, only one singular value is non-zero and all the energy is allocated to this eigen-channel. This singular value and the resulting capacity are given by

$$d_1 = \sqrt{|h_1|^2 + |h_2|^2 + \ldots |h_M|^2}, \tag{24}$$

$$C = \log_2\left(1 + \frac{E}{\sigma^2}\sum_{m=1}^{M} |h_m|^2\right). \tag{25}$$

*Case 2*: $N$ transmitters and 1 receiver, $\mathbf{H} = [h_1, h_2, \ldots, h_N]$, $\text{rank}(\mathbf{H}) = 1$.

Since $\text{rank}(\mathbf{H}) = 1$, only one singular value is non-zero and all the energy is allocated to this eigen-channel. This singular value and the resulting capacity are given by

$$d_1 = \sqrt{|h_1|^2 + |h_2|^2 + \ldots |h_N|^2}, \tag{26}$$

$$C = \log_2\left(1 + \frac{E}{\sigma^2}\sum_{n=1}^{N} |h_n|^2\right), \tag{27}$$

Note that this result is valid *only if the channel is known at the transmitter.*

*Case 3*: $N$ transmitters and $M$ receivers with perfect line of sight (LOS), without multipath.

Let $d_t$ be the distance between the transmit elements and $d_r$ the distance between the receive elements. The transmitter transmits in direction $\phi_t$ with respect to its baseline while the receiver receives from angle $\phi_r$ with respect to its baseline. In this case,

$$h_{mn} = \exp(jkd_r(m-1)\cos\phi_r)\exp(jkd_t(n-1)\cos\phi_t). \tag{28}$$

Note that even though the channel matrix $\mathbf{H}$ is $M \times N$, it is still rank-1 and $d_1 = \sqrt{NM}$. The capacity is given by

$$C = \log_2\left(1 + NM\frac{E_s}{\sigma^2}\right), \tag{29}$$

i.e., in line-of-sight conditions, the arrays at the transmitter and receiver only provide a power gain of $NM$.

*Case 4*: $N = M$ and the channel has full rank with equal singular values.

Since the square of the singular values of $\mathbf{H}$ are the eigenvalues of $\mathbf{H}\mathbf{H}^H$,

$$\sum_{m=1}^{M} d_m^2 = \text{trace}\left(\mathbf{H}\mathbf{H}^H\right) = \sum_{n=1}^{N}\sum_{m=1}^{M} |h_{mn}|^2.$$

Since, on average, the each channel has unit power and we assume equal singular values, $d_m^2 = NM/M = N$, $\forall m$. Since all singular values are equal the energy allocation is clearly uniform ($E_m = E_s/N$) and

$$C = \sum_{m=1}^{M} \log_2\left(1 + \frac{E_s d_m^2}{N\sigma^2}\right) = \sum_{m=1}^{M} \log_2\left(1 + \frac{E_s}{\sigma^2}\right) = N \log_2\left(1 + \frac{E_s}{\sigma^2}\right). \tag{30}$$

Note the significant difference in the capacities described in Eqns. (29) and (30). Under perfect LOS conditions, the transmit and receive array only provide *power gain* and the capacity increases as the log of the number of elements. However, when the channel is set up such that each eigenchannel is *independent* and has equal power, the capacity gains *are linear*. The independent channels allow us to transmit *independent* data streams ($N$ in the final example above), thereby increasing capacity.

In summary, in this section we have shown that a system with $N$ transmitters and $M$ receivers can be reduced to a problem of $r$ parallel AWGN channels, where $r$ is the rank of the channel matrix. To achieve the greatest gains in capacity, the channels from two different transmitters to the receivers must be independent and have equal power. The maximum possible gain in the channel capacity (over the SISO case) is the minimum of the number of transmitters and receivers, i.e., $\min(N, M)$. We will address this final constraint on the linear growth in capacity again in Section 2.1.4.

### 2.1.3 Channel Unknown at Transmitter

The analysis in Section 2.1.2 assumes both the transmitter and receiver know the channel matrix $\mathbf{H}$. However, in the more practical case that the channel is not known at the transmitter, but is known at the receiver, the approach is not valid. In this case, channel capacity must be determined as the maximum possible mutual information between input $\mathbf{X}$ and output $\mathbf{Y}$.

The capacity is given by

$$C = \max_{f_{\mathbf{X}}(\mathbf{x})} I(X;Y) = \max_{f_{\mathbf{X}}(\mathbf{x})} \left[H(\mathbf{Y}) - H(\mathbf{Y}/\mathbf{X})\right], \tag{31}$$

where $H(\mathbf{X})$ is the entropy in $\mathbf{X}$ with probability density function $f_{\mathbf{X}}(\mathbf{x})$ and is not to be confused with the channel matrix $\mathbf{H}$. Assuming channel matrix $\mathbf{H}$ is *known at the receiver*, the entropy in $\mathbf{Y}$, given the input data $\mathbf{X}$, is clearly only due to the noise $\mathbf{N}$. Assuming the noise to be complex, white and Gaussian with variance $\sigma^2$,

$$H(\mathbf{Y}/\mathbf{X}) = H(\mathbf{N}) = M\log_2(\pi e\sigma^2) = \log_2(\pi e\sigma^2)^M, \tag{32}$$

Given the channel, the entropy is $\mathbf{Y}$ is determined by the distribution of $\mathbf{X}$. We invoke the fact that the input distribution required to achieve capacity is *Gaussian*, i.e., $\mathbf{X}$ must be Gaussian distributed with $\mathbf{X} \sim N(0, \mathbf{\Sigma}_x)$ where $\mathbf{\Sigma}_x$ is the covariance matrix of $\mathbf{X}$ and whose diagonal entries are such that they meet the criterion of limited transmit energy.

From Eqn. (18), given $\mathbf{H}$, $\mathbf{Y}$ is also Gaussian with $\mathbf{Y} \sim N(0, \mathbf{\Sigma}_y)$ where $\mathbf{\Sigma}_y = \sigma^2\mathbf{I}_M + \mathbf{H}\mathbf{\Sigma}_x\mathbf{H}^H$ and $\mathbf{I}_M$ is the $M \times M$ identity matrix. Using the entropy result for the Gaussian pdf [1],

$$\begin{aligned} H(\mathbf{Y}) &= \log_2\left[(\pi e)^M \det \mathbf{\Sigma}_y\right], &(33)\\ \Rightarrow C = \max_{f_{\mathbf{X}}(\mathbf{x})} I(\mathbf{X};\mathbf{Y}) &= \log_2\left[(\pi e)^M \det\left(\sigma^2\mathbf{I}_M + \mathbf{H}\mathbf{\Sigma}_x\mathbf{H}^H\right)\right] - \log_2\left(\pi e\sigma^2\right)^M, &(34)\\ &= \log_2\det\left(\mathbf{I}_M + \frac{1}{\sigma^2}\mathbf{H}\mathbf{\Sigma}_x\mathbf{H}^H\right), &(35)\end{aligned}$$

Based on an eigendecomposition of the covariance matrix of the input data, $\mathbf{\Sigma}_x$, one can show that the optimal covariance matrix is $\mathbf{\Sigma}_x = (E_s/N)\mathbf{I}_N$ [1, 2, 4] which corresponds to independent data streams and equal power distribution over all available channels. The capacity is therefore,

$$C = \log_2\det\left(\mathbf{I}_M + \frac{E_s}{N\sigma^2}\mathbf{H}\mathbf{H}^H\right). \tag{36}$$

Note that, as with SISO channels, for a fixed MIMO channel unknown at the transmitter, the true capacity is zero since we cannot guarantee any minimum channel quality.

### 2.1.4 Fading MIMO Channels

So far we have focused on fixed channels. In the most practical situation, the channels vary as a function of time. In this case, the channel can change from one time instant to the next. Assuming sufficient interleaving to make the channel independent from one symbol instant to the next, the average capacity over a block of $K$ data symbols is given by

$$\begin{aligned} C &= \frac{1}{K}\sum_{k=1}^{K}\max_{f_{\mathbf{X}}(\mathbf{x})} I(\mathbf{X}[\mathbf{k}];\mathbf{Y}[\mathbf{k}]) \\ &= \frac{1}{K}\sum_{k=1}^{K}\log_2\det\left(\mathbf{I}_M + \frac{E_s}{N\sigma^2}\mathbf{H}[\mathbf{k}]\mathbf{H}[\mathbf{k}]^H\right). &(37)\end{aligned}$$

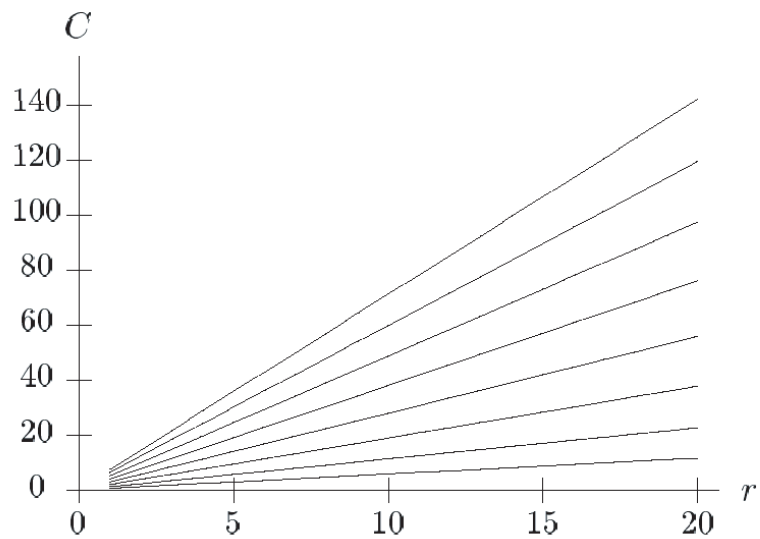


Figure 4: MIMO capacity in fading channels [5].

Based on the law of large numbers as $K \to \infty$ this approaches the expectation value of the right hand side in Eqn. (36) [4]

$$C = \mathrm{E}\left\{\log_2 \det\left(\mathbf{I}_M + \frac{E_s}{N\sigma^2}\mathbf{H}\mathbf{H}^H\right)\right\}. \tag{38}$$

If $\{d_m^2,\ m = 1, 2, \ldots, M\}$ are the $M$ eigenvalues of $\mathbf{H}\mathbf{H}^H$, the eigenvalues of $\left(\mathbf{I}_M + E_s/(N\sigma^2)\mathbf{H}\mathbf{H}^H\right)$ are $1 + E_s/(N\sigma^2)d_m^2$. The capacity in Eqn. (38) is then

$$C = \mathrm{E}\left\{\sum_{m=1}^{M} \log_2\left(1 + \frac{E_s}{N\sigma^2}d_m^2\right)\right\}, \tag{39}$$

where the expectation is taken over the $M$ eigenvalues.

This result in Eqns. (38) and (39) is valid for any type of fading. Specializing this to the case of completely independent Rayleigh fading from each transmit element to each receive element, each individual entry in $\mathbf{H}$ is an independent complex Gaussian random variable. In this case, the matrix $\mathbf{H}\mathbf{H}^H$ is Wishart distributed [6]. In addition, the pdf of its eigenvalues are known [5, 7].

$$f(d_1^2, \ldots, d_M^2) = \frac{1}{MK_{M,N}} e^{-\sum_{m=1}^{M} d_m^2} \prod_{m=1}^{M} (d_m^2)^{N-M} \prod_{m<n} \left(d_m^2 - d_n^2\right)^2, \tag{40}$$

where $K_{M,N}$ is a normalizing factor and

$$\begin{aligned} C &= \sum_{m=1}^{M} \mathrm{E}_{\{d_m^2\}}\left\{\log_2\left(1 + \frac{E_s d_m^2}{N\sigma^2}\right)\right\}, && (41) \\ \Rightarrow C &= M\left[\mathrm{E}_{\{d_1^2\}} \log_2\left(1 + \frac{E_s d_1^2}{N\sigma^2}\right)\right] && (42) \end{aligned}$$

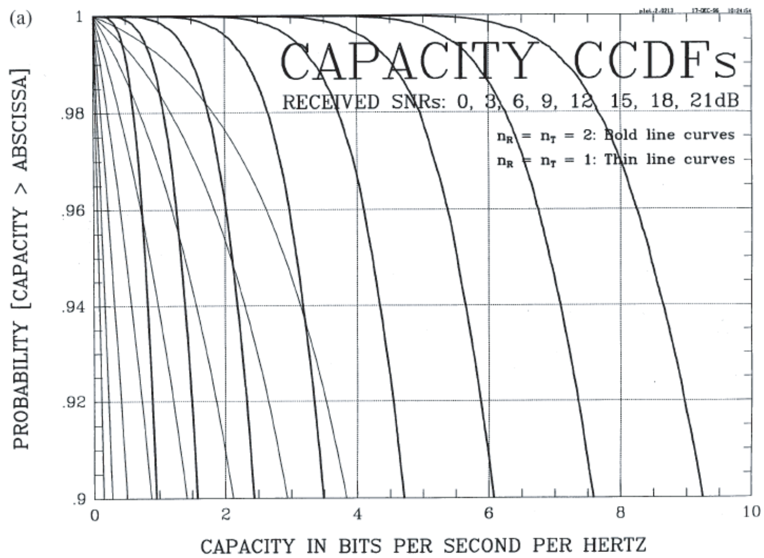


Figure 5: Outage probability in fading channels [8].

where the final expectation is taken over the pdf of an individual eigenvalue, found by marginalizing the multivariate pdf in Eqn. (40). The resulting capacity has been obtained by Telatar in [5] and shown in Fig. 4. The figure plots the capacity (in b/s/Hz) versus $M$ or $N$. The eight plots are for different SNRs between 0dB and 35dB in steps of 5dB. Note the linear relationship between the capacity and the number of transmit and receive channels.

There are two important results included in here, one positive, one cautionary. First, just as when the channel is known at the transmitter, if the channel is Rayleigh and independent, it is possible to have *linear* increases in capacity in fading channels as well. The second (cautionary) result is that the increase is proportional to the minimum of the number of transmitters and receivers, i.e., $\min(N, M)$. This has important implications in a cellular network - it is reasonable to assume multiple elements at the base station. But, it is unlikely one could have more than one or two elements in a handheld device. In this case, multiple antennas at one end will only provide power gains, but not the parallel channels that provide large capacity gains.

The result in Eqn. (42) and Fig. 4 present the true capacity of a MIMO channel with independent Rayleigh fading, i.e., it is theoretically possible to have error-free transmission with rate below this capacity. Another figure of merit is outage probability such as derived in Eqn. (5) for SISO channels. In [8], Foschini and Gans evaluate the outage probability under Rayleigh fading. One of their results is shown in Fig. 5. Note the huge improvement in outage probability (here they plot the cumulative distribution, which is $(1 - P_{\text{out}})$ by moving from a SISO channel to $N = M = 2$. With a SNR of 21dB, the capacity of a SISO channel is larger than approximately 2.5b/s/Hz 96% of the time, while for $N = M = 2$ the capacity is larger than approximately 8.5b/s/Hz 96% of the time.