MIMO and Spatial Multiplexing Gain

Multiple Input Multiple Output (**MIMO**) is one of the principal methods for exploiting multipath propagation. It has been employed in many cellular systems since 4G, and also appears in many contemporary WiFi standards.

In this lecture, we will explore some of the foundations behind MIMO, and see how it leads to spatial multiplexing gain.

System Model

In MIMO, we consider a system with N_t **transmitting antennas** and N_r **receiving antennas**. Each transmission is then modeled as a vector $\mathbf{s} \in \mathbb{R}^{N_t}$ of symbols across the transmitting antennas. On the receiver side, we let $\mathbf{r} \in \mathbb{R}^{N_r}$ be the vector of received symbols.

The relationship between $\mathbf{r}$ and $\mathbf{s}$ is given by the following channel model:

$$\mathbf{r} = \sqrt{\rho} \cdot \mathbf{H} \cdot \mathbf{s} + \mathbf{z}, \tag{1}$$

where $\mathbf{H} = [H_{ij}] \in \mathbb{R}^{N_r \times N_t}$ is the MIMO **channel matrix**, with H_{ij} being the gain from transmit antenna j to receive antenna i. $\mathbf{z} \in \mathbb{R}^{N_r}$ is the receiver noise, and $\rho > 0$. In general, $\mathbf{r}$, $\mathbf{s}$, $\mathbf{z}$, and $\mathbf{H}$ can be complex, but we will consider them in the space of reals for our discussion here. The following assumptions are common:

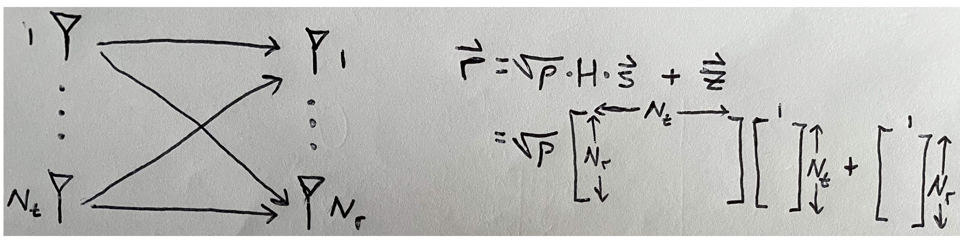
- $\mathbf{H}$ is assumed to be normalized such that $||\mathbf{H}||_F^2 = N_r \cdot N_t$, where $||\cdot||_F$ denotes the Frobenius norm of the matrix. For example, this would be true if $H_{ij} = 1$ for all elements, which would mean the gain between each transmitter-receiver pair is about the same.
- $\mathbf{s}$ is assumed to have unit power, i.e., $\mathbb{E}[||\mathbf{s}||^2] = 1$. This means that the total transmission power will not grow with N_t, i.e., as we add more

transmit antennas. For example, if the power was uniform across all antennas, we would have $\mathbb{E}[s_i^2] = 1/N_t$.

- Each element of $\mathbf{z}$ is assumed to be zero-mean, unit variance $\sigma^2 = 1$, and i.i.d. (independent and identically distributed). In general, having more antennas at the receiver leads to more noise being picked up.

As a result of these normalizations, ρ in (1) can be interpreted as the received **signal to noise ratio** (SNR), comparable to a SISO system.

The MIMO system and channel model are depicted below.



Eigen-Beamforming

To see how MIMO leads to spatial multiplexing gain, we need to consider methods for transmission and reception. We will assume here that both the sender and the receiver know the channel matrix $\mathbf{H}$ (this can be a big assumption!).

Using the SVD (singular value decomposition), $\mathbf{H}$ can be written as

$$\mathbf{H} = \mathbf{U}\Lambda\mathbf{V}^T = \mathbf{U} \begin{bmatrix} \sqrt{\lambda_1} & 0 & \cdots & 0 & \cdots & 0 \\ 0 & \sqrt{\lambda_2} & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{\lambda_N} & \cdots & 0 \end{bmatrix} \mathbf{V}^T. \tag{2}$$

- $\mathbf{U}$ is an $N_r \times N_r$ unitary matrix, containing the left singular vectors of $\mathbf{H}$.

- Λ is an $N_r \times N_t$ rectangular diagonal matrix, where $N = \min\{N_t, N_r\}$, drawn here for the case of $N = N_r$. The values $\sqrt{\lambda_1}, ..., \sqrt{\lambda_N}$ are the singular values of $\mathbf{H}$. Their squares $\lambda_1, ..., \lambda_N$ are the eigenvalues of $\mathbf{H}\mathbf{H}^T$.

- $\mathbf{V}$ is an $N_t \times N_t$ unitary matrix, containing the right singular vectors of $\mathbf{H}$.

A unitary matrix is one whose inverse is its transpose, i.e., $\mathbf{U}\mathbf{U}^T = \mathbf{U}^T\mathbf{U} = \mathbf{I}$, where $\mathbf{I}$ is the $N_r \times N_r$ identity matrix, and similarly, $\mathbf{V}\mathbf{V}^T = \mathbf{V}^T\mathbf{V} = \mathbf{I}$. Multiplying a vector $\mathbf{x}$ by a unitary matrix, e.g., $\mathbf{V}\mathbf{x}$, will rotate $\mathbf{x}$ but not change its length.

Suppose $\mathbf{x} \in \mathbb{R}^{N_t}$ is the vector of information that the transmitter wishes to send. In **eigen-beamforming**, the sender multiplies these information symbols by the matrix $\mathbf{V}$, i.e.,

$$\mathbf{s} = \mathbf{V}\mathbf{x} = \begin{bmatrix} \mathbf{V}_1 & \mathbf{V}_2 & \cdots & \mathbf{V}_{N_t} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_{N_t} \end{bmatrix} = \mathbf{V}_1 x_1 + \mathbf{V}_2 x_2 + \cdots + \mathbf{V}_{N_t} x_{N_t}, \quad (3)$$

where x_i is the ith information symbol and $\mathbf{V}_i \in \mathbb{R}^{N_t}$ is the ith column of $\mathbf{V}$. The information x_i is thus sent over all antennas as $\mathbf{V}_i x_i$. Doing so "rotates" $\mathbf{x}$ and forms a beam for a certain direction, leading to the concept of **beamforming**. Note that the total transmit power does not change, since $\mathbb{E}[||\mathbf{V}\mathbf{x}||_2^2] = \mathbb{E}[\mathbf{x}^T\mathbf{V}^T\mathbf{V}\mathbf{x}] = \mathbb{E}[\mathbf{x}^T\mathbf{x}] = \mathbb{E}[||\mathbf{x}||_2^2]$.

At the receiver end, we multiply $\mathbf{r}$ by $\mathbf{U}^T$ to "reverse" the effect of the channel. Based on the channel model in (1), this results in

$$\begin{aligned} \mathbf{y} = \mathbf{U}^T\mathbf{r} \quad &= \quad \sqrt{\rho} \cdot \underbrace{\mathbf{U}^T\mathbf{U}}_{\mathbf{I}} \cdot \Lambda \cdot \underbrace{\mathbf{V}^T\mathbf{V}}_{\mathbf{I}} \cdot \mathbf{x} + \mathbf{U}^T\mathbf{z} = \sqrt{\rho}\Lambda\mathbf{x} + \tilde{\mathbf{z}} \\ &= \quad \sqrt{\rho} \begin{bmatrix} \sqrt{\lambda_1} & 0 & \cdots & 0 & \cdots & 0 \\ 0 & \sqrt{\lambda_2} & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{\lambda_N} & \cdots & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_N \\ \vdots \\ x_{N_t} \end{bmatrix} + \tilde{\mathbf{z}}, \qquad (4) \end{aligned}$$

where the resulting noise vector $\tilde{\mathbf{z}}$ is simply a rotated version of $\mathbf{z}$ and thus each element still has unit variance $\sigma^2 = 1$.

Thus, we effectively end up with $N = \min\{N_t, N_r\}$ transmit-receive equations:

$$y_i = \sqrt{\rho}\sqrt{\lambda_i} x_i + \tilde{z}_i, \qquad i = 1, ..., N. \tag{5}$$

In other words, through eigen-beamforming, the channel can be viewed as equivalent to N separate channels, each of which has an SNR of

$$\frac{\rho \cdot \mathbb{E}[x_i^2] \cdot \lambda_i}{\sigma^2} = \rho \cdot \mathbb{E}[x_i^2] \cdot \lambda_i. \tag{6}$$

The total achievable rate R across these N channels is thus

$$R = \sum_{i=1}^{N} B \log_2 \left(1 + \rho \cdot \mathbb{E}[x_i^2] \cdot \lambda_i\right), \tag{7}$$

which is clearly very dependent on the eigenvalues!

The Eigenvalues $\{\lambda_i\}$

We seek to understand how the eigenvalues $\lambda_1, ..., \lambda_N$ impact (7), and whether $\mathbf{H}$ can be designed to maximize R.

For convenience, we will consider the case that $N_t = N_r = N$ (it is not difficult to generalize to $N_t \neq N_r$). Since $||\mathbf{H}||_F^2 = N^2$ by assumption, using the expression of the Frobenius norm in terms of the singular values, we have

$$||\mathbf{H}||_F^2 = \sum_{i=1}^{N} \lambda_i = N^2. \tag{8}$$

In other words, sum of the eigenvalues is always restricted to N^2, and grows as the number of antennas grows. We will consider two possibilities for $\mathbf{H}$:

(1) Equal values for H_{ij}: First suppose $H_{ij} = 1$ for all i, j. Since $\mathbf{H}$ is a square matrix, the SVD is the same as the eigen-decomposition, and thus we can work directly with the latter to find the values of λ_i:

$$
\mathbf{H} = \underbrace{\begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{bmatrix}}_{\substack{\text{all tran-rec pairs} \\ \text{highly correlated}}} = \begin{bmatrix} \frac{1}{\sqrt{N}} & \vdots & \cdots & \vdots \\ \frac{1}{\sqrt{N}} & \vdots & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{\sqrt{N}} & \vdots & \cdots & \vdots \end{bmatrix} \underbrace{\begin{bmatrix} N & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}}_{\substack{\text{only one non-zero} \\ \text{singular value}}} \begin{bmatrix} \frac{1}{\sqrt{N}} & \frac{1}{\sqrt{N}} & \cdots & \frac{1}{\sqrt{N}} \\ \cdots & \cdots & \cdots & \cdots \\ \vdots & \vdots & \ddots & \vdots \\ \cdots & \cdots & \cdots & \cdots \end{bmatrix} \tag{9}
$$

With one non-zero eigenvalue $\lambda_1 = \sigma_1^2 = N^2$, we really only have one effective channel. Based on (7), the total rate becomes

$$
R = B \log_2(1 + \rho \cdot N^2 \cdot \mathbb{E}[X_1^2]). \tag{10}
$$

The total rate does increase with N, so we do have some advantages of multiple antennas in this case. However, the growth is pretty slow (as $\log N^2$), so we prefer to have more effective channels.

(2) Equal values for λ_i: On the other hand, suppose $\mathbf{H}$ is defined in such a way that $\lambda_1 = \cdots = \lambda_N = N$, which satisfies (8). In this case, we have N effective channels, with the total rate becoming

$$
R = N \cdot B \log_2(1 + \rho \cdot N \cdot \mathbb{E}[X_1^2]), \tag{11}
$$

assuming a uniform transmit power $\mathbb{E}[X_1^2] = \cdots = \mathbb{E}[X_N^2]$ across the transmit antennas. This growth is more desirable as it includes a factor which is linear in N (more precisely, $N \log N$). This is called the **spatial multiplexing** gain that can be achieved with MIMO, exploiting multipath propagation properties.

So, when will all λ_i's be approximately equal to N? One such case is when each element of $\mathbf{H}$ is an i.i.d. zero-mean Gaussian random variable with variance 1. Indeed, when we form $\mathbf{H}\mathbf{H}^T$, we arrive at a roughly diagonal matrix as the number of transmit and receive antennas N grows large:

$$\mathbf{H}\mathbf{H}^T = \begin{bmatrix} H_{11} & H_{12} & \cdots & H_{1N} \\ \vdots & \vdots & \ddots & \vdots \\ H_{N1} & H_{N2} & \cdots & H_{NN} \end{bmatrix} \begin{bmatrix} H_{11} & \cdots & H_{N1} \\ H_{12} & \cdots & H_{N2} \\ \vdots & \ddots & \vdots \\ H_{1N} & \cdots & H_{NN} \end{bmatrix} \approx \underbrace{\begin{bmatrix} N & 0 & \cdots & 0 \\ 0 & N & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & N \end{bmatrix}}_{\substack{\text{N non-zero} \\ \text{eigenvalues}}} \tag{12}$$

For example, the $(1,1)$ element of $\mathbf{H}\mathbf{H}^T$ is $\sum_{i=1}^N H_{1i}^2 \approx N$, while the $(2,1)$ element is $\sum_{i=1}^N H_{2i} H_{i2} \ll N$ as N grows large. You will show this more formally in a homework problem.

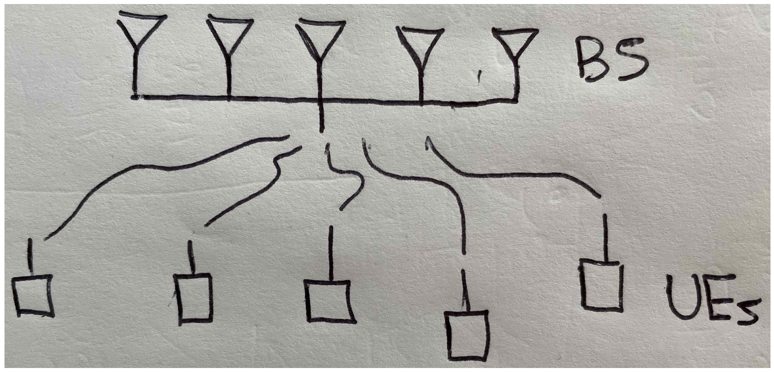
Therefore, to get the spatial multiplexing gain, it is desirable that the channel of each transmit-receive pair is independent of others. This provides sufficient "spatial diversity." It is more suitable when there are many multipaths, i.e., a "rich scattering" environment.

Unfortunately, this also means that there is more overhead to estimate the channel state information (**CSI**) of each transmit-receive pair. This overhead grows as N^2 in general. In addition to the eigen-beamforming formulation here, there are other expressions of MIMO capacity that assume CSI *only at the receiver*, which may also produce spatial multiplexing gains, though these will generally not be as good as in (11). There is (usually) always a tradeoff between control complexity and algorithm efficacy!

MU-MIMO

Our MIMO discussion so far has focused on a single-user setting (where the transmitter and receiver each have multiple antennas). In a wireless network, we will be concerned with many user devices, e.g., connected to a single base station (BS). This gives rise to the multi-user MIMO (**MU-MIMO**) setting. In MU-MIMO, on the downlink, the individual devices (receivers) often have less antennas than the BS (transmitter), but there are also many of them:

It is not feasible to apply eigen-beamforming in this setting, which we saw in (4) requires the multiplication by $\mathbf{U}^T$ at the receiver end. Instead, spatial multiplexing gain in MU-MIMO can be attained by the so-called **zero-forcing beamforming** method.



Formally, suppose there are $N_t = N$ transmit antennas at the BS, and N user devices, each with a single antenna. We can write the following channel model:

$$\mathbf{r} = \sqrt{\rho} \cdot \mathbf{H} \cdot \mathbf{s} + \mathbf{z} \tag{13}$$

where $\mathbf{r} \in \mathbb{R}^N$ is a vector of the received signal $r_1, ..., r_N$ at each user, $\mathbf{H} \in \mathbb{R}^{N \times N}$ is the channel matrix, and $\mathbf{s} \in \mathbb{R}^N$ is the vector of transmissions $s_1, ..., s_N$ placed on the BS antennas. Further, assume that $\mathbf{H}$ is invertible, i.e., $\mathbf{H}^{-1}$ exists, which we will call $\mathbf{V} \in \mathbb{R}^{N \times N}$:

$$\mathbf{H}\mathbf{H}^{-1} = \begin{bmatrix} \mathbf{H}_1^T \\ \vdots \\ \mathbf{H}_N^T \end{bmatrix} \begin{bmatrix} \mathbf{V}_1 & \cdots & \mathbf{V}_N \end{bmatrix} = \mathbf{I}, \tag{14}$$

where $\mathbf{H}_i^T$ is the ith row of $\mathbf{H}$ and $\mathbf{V}_j$ is the jth column of $\mathbf{H}^{-1}$. One useful equivalent condition for invertibility is that all eigenvalues of $\mathbf{H}$ are non-zero.

In zero-forcing beamforming, the sender (the BS) multiplies the information x_k of user k by $\mathbf{V}_k$, forming the following transmit vector:

$$\mathbf{s} = \sum_{k=1}^{N} \mathbf{V}_k \cdot x_k. \tag{15}$$

The signal received by user m is then

$$
\begin{aligned}
r_m &= \sqrt{\rho} \cdot \mathbf{H}_m^T \cdot (\mathbf{V}_k x_k) + \mathbf{z}_m \\
&= \sqrt{\rho} \cdot \underbrace{\mathbf{H}_m^T \mathbf{V}_m}_{1} \cdot x_m + \sqrt{\rho} \cdot \sum_{k \neq m}^{N} \underbrace{\mathbf{H}_m^T \mathbf{V}_k}_{0} \cdot x_k + \mathbf{z}_m = \sqrt{\rho} \cdot x_m + z_m. \quad (16)
\end{aligned}
$$

In other words, the beam $\mathbf{V}_m$ for user m will only produce a signal at user m, and will zero-out at all other users, giving the notion of "zero-forcing."

On the other hand, the effective transmit power $\mathbb{E}[||\mathbf{V}_m||^2]$ for each user m may be different, depending on the channel matrix.

Voice Systems vs. Data Systems

In the past few lectures, we have looked at different types of fading and different techniques for dealing with them. In closing, we should say a few words about the differences in techniques employed between voice systems and data systems.

Techniques for "working against fading," e.g., power control, are more commonly found in voice systems. On the other hand, techniques for "exploiting fading and diversity," e.g., opportunistic scheduling, OFDM, and MIMO, are more common for data systems. So, most of the techniques we have studied so far are geared towards data systems. The reason for this is that voice is a fixed rate, and is also more sensitive to delay than data.

One of the key challenges in OFDM and MIMO is the need for CSI collection, which can require high overhead. As the network gets larger and channel conditions change faster, as we see for IoT and machine-type communications, this overhead will continue to increase. This motivates techniques for improving the efficiency of CSI inference, as well as techniques for combating fading, in data systems.