

1 Antennas and Propagation

1.1 Introduction of Antenna

An antenna is a device that converts electrical signals into radio waves, and vice versa. It is used to transmit and receive electromagnetic waves in various applications.

Transmission line is a specialized cable or waveguide that connects the radio transmitter or receiver to the antenna. It carries the radio frequency (RF) signals between the transmitter/receiver and the antenna with minimal loss. There are several types of transmission lines, including *coaxial cables*, *parallel lines*, and *waveguides*. It ensures the power is efficiently delivered to the antenna for transmission and received signals are properly conveyed to the receiver.

1.2 Antenna Parameter

There are 3 main regions around an antenna: the reactive near-field region, the radiating near-field region, and the far-field region. Most antenna parameters are determined in the far-field region.

Reactive Near-Field Region is the area closest to the antenna, where the electromagnetic fields are highly reactive and non-radiating. It exists at

$$R < 0.62 \sqrt{\frac{D^3}{\lambda}}$$

where D is the largest dimension of the antenna, and λ is the wavelength. NFC, RFID antennas operate in this region.

Radiating Near-Field Region, or Fresnel region, is the region between the reactive near-field region and the far-field region. The radiation starts to form in this region, but the angular field distribution is still dependent on the distance from the antenna.

Far-field region, or Fraunhofer region, is the area furthest from the antenna where electromagnetic fields are radiated. In this region, the angular field distribution is independent of the distance from the antenna. Most antenna parameters are determined in this region. The radiation effect is greater in the far-field region. It exists at

$$R > \frac{2D^2}{\lambda}$$

Radiation Pattern is a graphical representation of the radiation properties of an antenna as a function of space coordinates. A transmitting (or receiving) antenna does not radiate (or receive) uniformly in all directions. The directional selectivity of an antenna is characterised by its radiation pattern. In most cases, the radiation pattern is determined in the far-field region.

The radiation pattern consists of lobes, which are regions where the radiation intensity is relatively high. The main types of lobes are:

- Major Lobe: The lobe containing the direction of maximum radiation.
- Minor Lobe: Any lobe other than the major lobe.
- Side Lobe: A minor lobe that is not in the direction of the major lobe.
- Back Lobe: The lobe opposite to the major lobe.

Directivity (D) is a far field parameter which describes the ability of antenna to focus energy to a particular direction. Higher the D , narrower the beam. There are 2 special types of antennas: *Isotropic antenna* is an idealized antenna that radiates power uniformly in all directions. And *omnidirectional antenna* radiates power uniformly in all directions of the horizontal plane but very limited in the vertical plane.

Gain (G) is a far field parameter which describes the ability of antenna to focus energy in a particular direction, taking into account the efficiency of the antenna. It is related to directivity by

$$G = \eta D$$

where η is the efficiency of the antenna. In dB,

$$G_{dB} = 10 \log_{10}(\eta) + D_{dB}$$

Use dBi to denote gain with respect to isotropic antenna, and dBd to denote gain with respect to dipole antenna.

Impedance (Z_A) is the ratio of voltage to current at the input terminals of the antenna.

$$Z_A = R_A + jX_A$$

For maximum efficiency, the antenna should be matched to the transmission line, i.e., $Z_A = Z_0$ where Z_0 is the characteristic impedance of the transmission line. (Usually $Z_0 = 50 \Omega$ or 75Ω)

When there is an impedance mismatch between the antenna and the transmission line, some of the power is reflected back toward the source. This is characterized by the **Reflection Coefficient (Γ)**:

$$\Gamma = \frac{Z_A - Z_0}{Z_A + Z_0}$$

And from this, we can calculate the **Return Loss (RL)**, which is a measure of how well the antenna is matched to the transmission line, expressed in dB:

$$RL = -10 \log_{10} |\Gamma|^2$$

The **Power reflected (P_r)** back due to impedance mismatch can be calculated as:

$$P_r = 10^{RL/10}$$

1.3 Types of Antenna

We have many types of antennas, here we review two common types: Half-Wave Dipole Antenna and Microstrip Antenna.

Half-Wave Dipole Antenna is a simple antenna consisting of two conductive elements, each approximately a quarter wavelength long, oriented end to end. Like this: -|. Some parameters are listed below.

- Total Length: $L = \lambda/2$
- Length of each arm: $L_{arm} = \lambda/4$
- Free Space Impedance: $Z_A \approx 73 \Omega$
- Ideal Directivity: $D \approx 1.64$ (≈ 2.16 dBi)
- Real Length: $L_r \approx 0.97 \cdot \lambda/2$

Microstrip Antenna is a type of antenna that consists of a flat rectangular sheet or "patch" of metal mounted over a larger sheet of metal called a ground plane. It contains the 3 main parts: Radiating Patch, Dielectric Substrate, and Ground Plane.

The Radiating Patch is the part that radiates electromagnetic waves. It is usually made of a conducting material such as copper or gold and has a thickness (t) much smaller than the wavelength ($t \ll \lambda$).

The Dielectric Substrate is a layer of insulating material that separates the radiating patch from the ground plane. It also has a thickness (h) much smaller than the wavelength ($h \ll \lambda$), usually $0.003\lambda \leq h \leq 0.05\lambda$.

The Ground Plane is located below the dielectric substrate. It helps to reflect the energy radiated from the patch to improve radiation efficiency.

Microstrip antennas are typically fed using a *Microstrip Feed Line*, which is a thin strip of metal that connects the patch to the transmitter or receiver. It is also used for impedance matching.

Unlike dipole antennas, microstrip antennas radiate primarily in the direction perpendicular to the surface of the patch, known as *broadside radiation*. This behaviour is mainly due to the reflective nature of the ground plane and other design factors.

Follow these steps to design a microstrip antenna:

1. Determine the desired resonant frequency (f_r) and dielectric constant of the substrate (ϵ_r) and height of the substrate (h). (Usually $2.2 \leq \epsilon_r \leq 12$)
2. Calculate the width (W) of the patch using:

$$W = \frac{c}{2f_r} \sqrt{\frac{2}{\epsilon_r + 1}}$$

3. Calculate the effective dielectric constant (ϵ_{eff}) using:

$$\epsilon_{eff} = \frac{\epsilon_r + 1}{2} + \frac{\epsilon_r - 1}{2} \left(1 + 12 \frac{h}{W} \right)^{-1/2}$$

4. Calculate the extended length (ΔL , due to fringing effects) using:

$$\Delta L = 0.412h \frac{(\epsilon_{eff} + 0.3) \left(\frac{W}{h} + 0.264 \right)}{(\epsilon_{eff} - 0.258) \left(\frac{W}{h} + 0.8 \right)}$$

5. Calculate the effective length (L_{eff}) using:

$$L_{eff} = \frac{c}{2f_r \sqrt{\epsilon_{eff}}}$$

6. Calculate the actual length (L) of the patch using:

$$L = L_{eff} - 2\Delta L$$

or

$$L \approx (0.47 \sim 0.49) \frac{c}{f_r \sqrt{\epsilon_r}} \approx \frac{\lambda_d}{2}$$

where λ_d is the wavelength in the dielectric substrate, $\lambda_d = \frac{\lambda_0}{\sqrt{\epsilon_r}}$ and λ_0 is the wavelength in free space.

7. Design the feed line to match the antenna impedance to the transmission line impedance (usually $Z_0 = 50 \Omega$). Calculate the width of the feed line (W_f) using:

$$W_f = \frac{7.48h}{e^{Z_0 \frac{\sqrt{\epsilon_r} + 1.41}{87}}} - 1.25t$$

8. Design the ground plane with length (L_g) and width (W_g) using:

$$L_g \geq 6h + L$$

$$W_g \geq 6h + W$$

Usually we take $6h + L$ and $6h + W$.

1.4 Antenna Propagation

Friis Transmission Formula is used to calculate the power received by an antenna under idealized conditions given the distance from the transmitter, the gain of the transmitting and receiving antennas, and the wavelength of the transmitted signal. Its equation is:

$$P_r = P_t G_t G_r \left(\frac{\lambda}{4\pi R} \right)^2$$

where:

- P_r : Power received by the receiving antenna (in W)
- P_t : Power transmitted by the transmitting antenna (in W)
- G_t : Gain of the transmitting antenna (unitless)
- G_r : Gain of the receiving antenna (unitless)
- λ : Wavelength of the transmitted signal (in m)
- R : Distance between the transmitting and receiving antennas (in m)

In dB, the Friis Transmission Formula can be expressed as:

$$P_{r,dB} = P_{t,dB} + G_{t,dB} + G_{r,dB} - 20 \log_{10} \left(\frac{4\pi R}{\lambda} \right)$$

1.5 Antennas for Satellite Communication

Satellite communication can be divided into 2 segments: Space Segment and Ground Segment.

Space Segment consists of the satellites themselves and facilities need to keep them operational. Satellites are equipped with transponders, which receive signals from the earth station, amplify and process them, and retransmit them back to earth. The main components of a transponder include:

- Receiving Antenna: Antenna on the satellite that receives signals from the earth station.

- Transponder electronics: Amplifies and processes the received signals before retransmitting them back to earth.
- Transmitting Antenna: Antenna on the satellite that transmits signals back to the earth station.

Different types of antennas are used in satellites, including Reflector (Dish) Antennas, Horn Antennas, Phased Array Antennas, Reflectarray Antennas, and Patch Antennas.

Reflector antennas are the most common type used in satellites, as they can produce high gain and directivity. *Horn antennas* are also used for their simplicity and circularly symmetric pattern. *Phased array antennas* allow for electronic steering of the radiation pattern without physically moving the antenna. *Reflectarray antennas* provide focused and highly directive radiation in a flat form factor. *Patch antennas* are low gain flat antennas mainly used in Low Earth Orbit (LEO) satellites.

Ground Segment consists of the transmit and receive earth stations/antennas. Earth station antennas are typically large parabolic reflector antennas that can be either fixed or mobile. They are designed to transmit and receive signals to and from the satellites in orbit. Different types of parabolic reflector antennas used in earth stations include Cassegrain Feed, Gregorian Feed, Front Feed, and Offset Feed antennas.

2 MIMO Antenna for 5G communication

2.1 Introduction of 5G

5G is the fifth generation of wireless cellular technology and is the new global wireless standard after 4G/LTE. 5G enables a new kind of network that is designed to connect virtually everyone and everything together including machines, objects, and devices. 5G wireless technology is meant to deliver *higher multi-Gbps peak data speeds, ultra low latency, more reliability, massive network capacity, increased availability, and a more uniform user experience to more users.*

There are three main concepts involved in 5G deployment: *Enhanced Mobile Broadband (eMBB)*, *Massive Machine Type Communications (mMTC)*, and *Ultra-Reliable Low Latency Communications (URLLC)*. eMBB provides significantly faster data speeds and larger data volumes for mobile communications. mMTC supports the connectivity of a vast number of devices, particularly in massive IoT applications. URLLC is designed for applications that require extremely low latency and high reliability, such as autonomous vehicles, remote real-time surgeries, etc.

5G spectrum is a range of radio frequencies in the sub-6 GHz range and the millimeter wave (mmWave) frequency range. It can be divided into three main frequency ranges: *Low-band spectrum (Sub-1 GHz)*, *Mid-band spectrum (1-6 GHz)*, and *High-band spectrum (Above 24 GHz, also known as mmWave)*. Low-band spectrum offers greater coverage with lower speeds and provides comprehensive coverage for mobile broadband and massive IoT applications. Mid-band spectrum offers a balance of both coverage and speeds and is mostly used bands worldwide for 5G cellular communications, e.g., eMBB. High-band spectrum offers super fast speed but a smaller coverage radius. Popular bands are 26 GHz, 28 GHz, 37-40 GHz, and 64-71 GHz; offering extreme bandwidth.

Advanced Antenna Systems (AAS) are a key component of 5G networks, utilizing multi-antenna techniques namely MIMO (Multiple Input Multiple Output) and beamforming to enhance network performance. A core component of the AAS is the active antenna system, where the active transceiver array and the passive antenna array are intelligently integrated into a single hardware unit.

2.2 MIMO antenna for 5G communication

Antenna has different types according to the number of antennas at the transmitter and receiver ends. They are:

SISO: Single antenna at both the transmitter and receiver ends (Single Input Single Output).

MISO: Multiple antennas at the receiver end and a single antenna at the transmitter end (Multiple Input Single Output).

SIMO: Multiple antennas at the transmitter end and a single antenna at the receiver end (Single Input Multiple Output). If there is a big distortion on one link, the probability of the other also being distorted is less. This is called *spatial diversity*.

MIMO: Multiple antennas at both the transmitter and receiver ends (Multiple Input Multiple Output). By using multiple antennas, MIMO can transmit multiple data streams simultaneously, increasing the overall throughput without needing additional bandwidth or higher power. It can direct signals toward the intended receiver, reducing interference and improving signal strength. This is called *beamforming*. It can transmit two (or more in case of more antennas) different data streams simultaneously by using the same timeslots and bandwidth, increasing the overall throughput. This is called *spatial reuse*. It also provides spatial diversity.

2.3 MIMO

We assume a MIMO system with M transmit antennas and N receive antennas, so:

- Transmit Vectors: $\mathbf{x} = [x_1, x_2, \dots, x_M]^T$
- Receive Vectors: $\mathbf{y} = [y_1, y_2, \dots, y_N]^T$

We can model the MIMO system as:

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

\mathbf{H} is the channel matrix, where the element $h_{i,j}$ represents the channel gain from the j -th transmit antenna to the i -th receive antenna. \mathbf{n} is the noise vector.

MIMO use two main techniques to improve the performance of wireless communication systems: *spatial diversity* and *spatial multiplexing*. In general, spatial diversity provides diversity gain to improve the reliability of the system, while spatial multiplexing provides multiplexing gain to improve the data rate of the system. Spatial diversity can be achieved by transmitting multiple copies of the same data across independently fading channels, while spatial multiplexing can be achieved by transmitting independent information on each channel.

Spatial diversity is used in SIMO and MIMO systems to improve the reliability of wireless communication systems. By using multiple antennas at the receiver end (SIMO) or both the transmitter and receiver ends (MIMO), spatial diversity can provide multiple independent paths for the transmitted signal to reach the receiver. This helps to mitigate the effects of fading and interference, as the probability of all paths being affected by deep fades or interference is significantly reduced.

Spatial multiplexing is used in MIMO systems to increase the data rate of wireless communication systems. By using multiple antennas at both the transmitter and receiver ends, spatial multiplexing can transmit multiple independent data streams simultaneously over the same frequency band. This effectively increases the capacity of the wireless channel without requiring additional bandwidth or higher transmit power.

2.4 Massive MIMO

Massive MIMO is an advanced form of MIMO technology that uses a large number of antennas at the base station to serve multiple users simultaneously. It is a key technology for 5G and beyond wireless communication systems, as it can significantly improve the spectral efficiency, energy efficiency, and reliability of wireless networks.

Massive MIMO works on the **beamforming** concept, where the base station uses multiple antennas to create narrow beams that can be directed toward specific users or devices, rather than broadcasting signals in all directions. It can provide better coverage of the user by concentrating signals in a particular direction. This enables operators to provide higher data rates, better coverage, and more efficient use of the available spectrum. Since the millimeter waves cannot penetrate through obstacles and do not propagate to longer distances due to a shorter wavelength, beamforming helps a user to receive a strong signal without interference with other users.

In a multi-antenna system, beamforming technology enables two beams to be superimposed with the best effect by pre-compensating the phases of transmit antennas. A spatial hole may occur if two beams have equal attenuation but opposite phases.

2.5 BDMA

Multiple access technique enables multiple users to simultaneously share a finite piece of radio spectrum. Each user device can access the available bandwidth at the same time without interfering with the other. Previous generations of wireless communication technologies used the following multiple access techniques: Code division multiple access, Time division multiple access, and Frequency division multiple access. 5G has its own set of multiple access techniques, including Orthogonal Frequency Division Multiple Access (OFDMA), single-carrier frequency-division multiple access (SC-FDMA), non-orthogonal multiple access (NOMA), and **Beam Division Multiple Access (BDMA)**.

BDMA works by allocating highly directive orthogonal beams to multiple mobile stations in 5G networks. It divides the base station antenna beam according to certain criteria and the location of subscribers within the cell. BDMA is independent of frequency, time and code constraints, helping it overcome the limitations of FDMA and orthogonal frequency division multiplexing (OFDM) in wireless communications. BDMA enables 5G networks to handle a high number of mobile users, while reducing multiuser interference. By implementing BDMA, 5G networks can more efficiently harness spectrum capacity even as they boost the number of channels.

Beam Steering is the process of dynamically changing the direction of the main lobe of a beam in an antenna array system. It is used to track or target a moving object or receiver. In an antenna array, multiple antennas work together to transmit or receive signals. By adjusting the phase and sometimes the amplitude of the signals at each individual antenna, the overall direction of the beam can be steered toward a specific location. This adjustment is typically done using electronic phase shifters in each antenna element, which allows antennas to change the direction of transmission or reception without moving physically. It is achieved through electronic adjustments. Applications: Radar systems to track aircraft, ships, or weather patterns, 5G and Wi-Fi networks to optimize signal quality by directing beams toward specific users, Satellite communication to maintain links with moving satellites or ground stations, etc.

2.6 Smart/Phased Array Antenna

An array antenna is a configuration of multiple individual antenna elements arranged in a specific pattern to work together as a single antenna system. A phased array antenna is a type of array antenna in which

the phase of the signal at each individual antenna element is controlled electronically. This allows to steer the beam in different directions by changing the relative phase of the signals across the antenna elements.

The Gain of the array antenna can be found by the following formula (in dB):

$$G_a = 10 \log_{10}(N) + G_e - L_{OHMIC} - L_{SCAN}$$

where N is the number of elements in the array, G_e is the gain of a single antenna element, L_{OHMIC} is the ohmic loss, and L_{SCAN} is the scan loss. Note that for every element added to an array, the gain increases by $10 \log_{10}(N)$. As adding new antenna element increases aperture size, but the noise figure stays constant. Phased array antenna provides higher gain and steerability.

3 MIMO Systems

3.1 Preview of Massive MIMO

Introduction of Massive MIMO has been discussed in the previous section. It has the following advantages. Massive MIMO can **enhance the capacity** of wireless networks by accommodating more users and devices in a single cell, which is crucial for dense urban environments. It also **enhances data rates** by boosting overall network throughput with the ability to serve multiple users simultaneously. Additionally, Massive MIMO **increases energy efficiency** by focusing the transmitted energy toward specific users through beamforming, reducing interference and power consumption. It also **enhances links reliability** by providing spatial diversity, which helps mitigate the effects of fading and interference in wireless channels.

Massive MIMO provides higher throughput and capacity gains promised by 5G by using smart antenna techniques like beamforming and beam steering. The beam is substantially narrower in massive MIMO because there are far more antennas than user equipment in the cell, allowing the base station to send RF energy to the UE more precisely and efficiently. The signal-to-noise ratio in the cell can be improved even further by installing a large number of antennas. Because 5G massive MIMO uses mmWave frequencies (often above 24 GHz), the antennas are tiny and simple to install and maintain. Using mmWave frequencies, the signal power drops quickly due to path loss. Higher number of antenna elements of massive MIMO provides greater gain and solves this issue.

3.2 Spectral Efficiency

Area throughput is the total data rate that can be transmitted over a given area, usually expressed in bits per second per hertz per square kilometer (bit/s/Hz/km²). It is a measure of how efficiently the available spectrum is utilized within a specific geographic area. The formula for calculating area throughput in a cellular network is:

$$\text{Area Throughput} = \text{Cell Density} \times \text{Bandwidth per Cell} \times \text{Spectral Efficiency}$$

To improve area throughput, network operators can increase cell density by deploying more base stations, allocate more bandwidth per cell, or enhance spectral efficiency.

A good design is using *hotspot tier* and *Coverage tier* together. In hotspot tier, base stations offer very high area throughput in small regions with high user density, which has very large cell density and large bandwidth per cell due to mmWave. Spectral efficiency is not very important here. In coverage tier, base stations provide wide area coverage with lower user density, which has low cell density and small bandwidth per cell due to sub-6 GHz. Spectral efficiency is very important here.

Spectral efficiency is a measure of how efficiently a given communication system utilizes the available frequency spectrum to transmit data. It is typically expressed in bits per second per hertz (bps/Hz) and indicates the amount of data that can be transmitted over a specific bandwidth in a given time period. To improve spectral efficiency, one of the most effective methods is to use advanced antenna techniques like Massive MIMO.

Massive MIMO has lots of antennas at the base station, which is far more than the number of user equipment in the cell. It has little interference leakage between different users and can provide strong directive signals to each other. Thus, it can support multiple users simultaneously in the same time-frequency resource, significantly increasing the overall data rate and spectral efficiency of the system. By using Massive MIMO, network operators can achieve higher spectral efficiency, leading to improved network performance and user experience.

3.3 Capacity of different systems

3.3.1 Single Input Single Output (SISO) System

The AWGN channel of a SISO system with a complex gain g can be modelled as:

$$y = gx + n$$

where $x \sim CN(0, q)$, $q = P/B$ is the energy per symbol, P is the transmit power, B is the bandwidth, and $n \sim CN(0, N_0)$ is the complex Gaussian noise with power spectral density N_0 . The capacity of this SISO system is given by:

$$C_{\text{SISO}} = \log_2 \left(1 + \frac{|g|^2 q}{N_0} \right)$$

3.3.2 Single Input Multiple Output (SIMO) System

The AWGN channel of a SIMO system with N receive antennas and a complex gain vector \mathbf{g} can be modelled as:

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

Using maximal ratio combining (MRC) or Matched Filtering at the receiver:

$$z = \frac{\mathbf{g}^H}{\|\mathbf{g}\|} \mathbf{y} = \|\mathbf{g}\| x + \frac{\mathbf{g}^H}{\|\mathbf{g}\|} \mathbf{n}$$

The capacity of this SIMO system per complex sample is given by:

$$C_{\text{SIMO}} = \log_2 \left(1 + \frac{\|\mathbf{g}\|^2 q}{N_0} \right)$$

where $\|\mathbf{g}\|^2 = \sum_{i=1}^N |g_i|^2$.

3.3.3 Multiple Input Single Output (MISO) System

The AWGN channel of a MISO system with M transmit antennas and a complex gain vector \mathbf{g} can be modelled as:

$$y = \mathbf{g}^T \mathbf{x} + n$$

Using maximal ratio transmission (MRT) or Conjugate Beamforming at the transmitter:

$$\mathbf{x} = \frac{\mathbf{g}^*}{\|\mathbf{g}\|} s$$

where $s \sim CN(0, q)$ is the information symbol. The capacity of this MISO system per complex sample is given by:

$$C_{\text{MISO}} = \log_2 \left(1 + \frac{\|\mathbf{g}\|^2 q}{N_0} \right)$$

3.4 Basic information of SVD

Consider an N order matrix \mathbf{A} , if there is a non-zero vector \mathbf{v} such that

$$\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$$

then λ is called an eigenvalue of \mathbf{A} and \mathbf{v} is called an eigenvector of \mathbf{A} . The rank of \mathbf{A} is equal to the number of non-zero eigenvalues of \mathbf{A} . To find the eigenvalues of \mathbf{A} , we can solve the characteristic equation:

$$\det(\mathbf{A} - \lambda\mathbf{I}) = 0$$

where \mathbf{I} is the identity matrix of the same order as \mathbf{A} . And eigenvectors can be found by solving the equation:

$$(\mathbf{A} - \lambda\mathbf{I})\mathbf{v} = 0$$

If \mathbf{A} has N non-zero eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_N$, then we can form the diagonal matrix Λ with these eigenvalues on the diagonal:

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_N \end{bmatrix}$$

The matrix \mathbf{A} has:

$$\mathbf{A} = \mathbf{V}\Lambda\mathbf{V}^{-1}$$

where \mathbf{V} is the matrix whose columns are the eigenvectors of \mathbf{A} .

If \mathbf{A} is symmetric ($\mathbf{A} = \mathbf{A}^H$), the \mathbf{V} is an unitary matrix, i.e., $\mathbf{V}^{-1} = \mathbf{V}^H$. Thus, we have:

$$\mathbf{A} = \mathbf{V}\Lambda\mathbf{V}^H$$

So, \mathbf{A} can be diagonalized by an unitary matrix \mathbf{V} :

$$\mathbf{V}^H \mathbf{A} \mathbf{V} = \Lambda$$

When \mathbf{A} is not a square matrix, we can use Singular Value Decomposition (SVD) to decompose it. The SVD of an $N \times M$ matrix \mathbf{A} is given by:

$$\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^H$$

where \mathbf{U} is an $N \times N$ unitary matrix whose columns are the left singular vectors of \mathbf{A} or the eigenvectors of $\mathbf{A}\mathbf{A}^H$; Σ is an $N \times M$ diagonal matrix with non-negative real numbers on the diagonal, known as the singular values s_i of \mathbf{A} ; \mathbf{V} is an $M \times M$ unitary matrix whose columns are the right singular vectors of \mathbf{A} or the eigenvectors of $\mathbf{A}^H\mathbf{A}$.

3.5 Point-to-Point MIMO System

The AWGN channel of a point-to-point MIMO system with M transmit antennas, N receive antennas, and a complex channel matrix \mathbf{H} can be modelled as:

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

where h_{ij} is the complex channel gain from the j -th transmit antenna to the i -th receive antenna. Using Singular Value Decomposition (SVD), we can decompose the channel matrix \mathbf{H} as:

$$\mathbf{H} = \mathbf{U}\Sigma\mathbf{V}^H$$

By applying the precoding matrix \mathbf{V} at the transmitter and the decoding matrix \mathbf{U}^H at the receiver, we can transform the MIMO channel into multiple parallel SISO channels:

$$\mathbf{r} = \mathbf{U}^H\mathbf{y} = \Sigma\tilde{\mathbf{x}} + \mathbf{w}$$

where $\tilde{\mathbf{x}} = \mathbf{V}^H\mathbf{x}$ is the transmitted signal after precoding, and $\mathbf{w} = \mathbf{U}^H\mathbf{n}$ is the noise after decoding. Let $S = \text{rank}(\mathbf{H}^H\mathbf{H}) \leq \min(M, N)$ be the number of non-zero singular values in Σ , then

$$r_i = \begin{cases} s_i \tilde{x}_i + w_i, & i = 1, 2, \dots, S \\ w_i, & i = S+1, S+2, \dots, N \end{cases}$$

where s_i is the effective channel gain for the i -th SISO channel. So the capacity of each SISO channel is:

$$C_i = \log_2 \left(1 + \frac{s_i^2 q_i}{N_0} \right)$$

where q_i is the power allocated to the i -th SISO channel. The total capacity of the MIMO system is given by:

$$C_{\text{MIMO}} = \max_{\{q_i\}: q_1 + \dots + q_S = q} \sum_{i=1}^S C_i$$

which means we need to find a set of $\{q_i\}$ that maximizes the total capacity under the total power constraint:

$$\sum_{i=1}^S q_i = q$$

This can be solved using the water-filling algorithm.

Water-filling algorithm is a power allocation strategy used to maximize the capacity of parallel channels under a total power constraint. From above we can find that

$$q_i = \max(0, \mu - \frac{N_0}{s_i^2})$$

where μ is the water level determined by the total power constraint. The larger the singular value s_i , the more power is allocated to that channel. Channels with very small singular values may receive no power at all if the water level is below the noise level for that channel. With a low SNR, the water level is low, and only a few channels with large singular values receive power. As the SNR increases, the water level rises, and more channels receive power, leading to a higher overall capacity.

When the SNR is very high, the water level becomes much higher than the noise levels of all channels. In this case, the power allocation becomes nearly uniform across all channels, and the capacity can be approximated as:

$$C_{\text{MIMO}} \approx \sum_{i=1}^S \log_2 \left(\frac{s_i^2 q}{N_0 S} \right) = \sum_{i=1}^S \log_2 \left(\frac{s_i^2}{S} \right) + S \log_2 \left(\frac{q}{N_0} \right)$$

When the SNR is very low, the water level is close to the noise levels of the channels. In this case, only the channel with the largest singular value receives power, and the capacity can be approximated as:

$$C_{\text{MIMO}} \approx \log_2 \left(1 + \frac{s_1^2 q}{N_0} \right)$$

where s_1 is the largest singular value.

Point-to-point MIMO systems has some problems. When applied in line-of-sight (LoS) scenarios (High SNR), the multiplexing gain $S \approx 1$, which means that the MIMO system behaves like a SISO system, and the capacity gain from using multiple antennas is limited. This is because only one dominant path exists between the transmitter and receiver, leading to high correlation between the channels of different antenna pairs. When applied in non-line-of-sight (NLoS) scenarios (Low SNR), the multiplexing gain S can be larger than 1, but the overall SNR is low, resulting in limited capacity improvement. And because $S \leq \min(M, N)$, the capacity gain is limited by the smaller number of antennas at either the transmitter or receiver. So the S cannot be very large due to the physical size constraints of the devices.

3.6 Multi-user MIMO System

In a multi-user MIMO (MU-MIMO) system, let's assume there is a base station serving 2 users and user 1 has αB bandwidth while user 2 has $(1 - \alpha)B$ bandwidth, the capacity for each user can be calculated as follows.

$$\begin{aligned} C_1 &= \alpha B \log_2 \left(1 + \frac{P\beta}{\alpha B N_0} \right) \\ C_2 &= (1 - \alpha) B \log_2 \left(1 + \frac{P\beta}{(1 - \alpha) B N_0} \right) \end{aligned}$$

where $\sqrt{\beta}$ is the average channel gain for each user.

For the uplink in MU-MIMO system, received signal can be normalized and modelled as:

$$\mathbf{y} = \sqrt{\rho_{ul}} \mathbf{H} \mathbf{x} + \mathbf{w}$$

where ρ_{ul} is the SNR, $E[x_k] \leq 1$, and $\mathbf{w} \sim \mathcal{CN}(0, \mathbf{I}_N)$.

The difference between point-to-point MIMO and MU-MIMO is that in MU-MIMO, users do not cooperate with each other (x_1, \dots, x_K are independent), each user cares about its own data rate and has its own power budget. The channel matrix \mathbf{H} is an $N \times K$ matrix, where each column \mathbf{h}_k represents the channel vector from user k to the base station.

3.7 Massive MIMO

Massive MIMO is similar to MU-MIMO, but still has some differences. In Massive MIMO, the base station is equipped with a very large number of antennas (typically hundreds or more) to serve multiple users

simultaneously ($N \gg K$). Massive MIMO has more directive signals, less randomness and larger beam-forming gain and less interference. Massive MIMO can stably reach the maximum capacity gain K while MU-MIMO cannot. MU-MIMO often used in LTE (4G) and WiFi, while Massive MIMO is a key technology for 5G and beyond wireless communication systems.

On uplink, assume we have 2 users sending signals $s_k, k \in 1, 2$ to the base station. The channel \mathbf{h}_k from user k is random, zero mean and unit variance. The noise $\mathbf{n} \sim CN(0, \mathbf{I}_N)$. The received signal at the base station is:

$$\mathbf{y} = \mathbf{h}_1 s_1 + \mathbf{h}_2 s_2 + \mathbf{n}$$

To detect s_1 , we can use the matched filter (MF) $v_1 = \mathbf{h}_1/N$:

$$\mathbf{v}_1^H \mathbf{y} = \frac{1}{N} \mathbf{h}_1^H \mathbf{h}_1 s_1 + \frac{1}{N} \mathbf{h}_1^H \mathbf{h}_2 s_2 + \frac{1}{N} \mathbf{h}_1^H \mathbf{n}$$

When N is very large, by law of large numbers we have $\mathbf{h}_1^H \mathbf{h}_1/N \rightarrow 1$, $\mathbf{h}_1^H \mathbf{h}_2/N \rightarrow 0$, $\mathbf{h}_1^H \mathbf{n}/N \rightarrow 0$. So the received signal becomes:

$$\mathbf{v}_1^H \mathbf{y} \rightarrow s_1$$

For a period of time T_c , the channel can be seen as a Linear time-invariant system. T_c is called the coherence time, usually $T_c = \lambda/2v$ or $T_c = \lambda/4v$, where v is the velocity of the user. And because the signal will pass through paths of different lengths, the received signal will be spread in time. So the frequency response of the channel is not flat. We can define the coherence bandwidth B_c as the bandwidth over which the channel can be considered flat-fading. Usually $B_c = c/|d_{\max} - d_{\min}|$ or $B_c = c/2|d_{\max} - d_{\min}|$, where d_{\max} and d_{\min} are the maximum and minimum path lengths respectively. So we can divide the signal into time-frequency coherence blocks of size $T_c \times B_c$ complex samples. In each coherence block, the channel can be considered flat-fading and constant.

To find out the capacity of Massive MIMO, we need to estimate the channel first. This can be done by using pilot signals. On uplink, the user send a single pilot signal to the base station. The received pilot signal at the base station is:

$$\mathbf{y} = \mathbf{h} s + \mathbf{n}$$

The base station can estimate the channel \mathbf{h} by:

$$\hat{\mathbf{h}} = \frac{s^*}{|s|^2} \mathbf{y}$$

Using one pilot signal, the transmitter get all the gains. So the number of pilots needed is equal to the number of users K . But on downlink, the base station sends a known pilot signal from each antenna. The received pilot signal at user is:

$$y_m = h_m s + n_m, m = 1, \dots, M$$

The user can estimate the channel \mathbf{h} by:

$$\hat{h}_m = \frac{s^*}{|s|^2} y_m$$

Then user feeds $\hat{\mathbf{h}}$ back to the base station. In total, M pilots (plus feedback) are needed to get all the gains for downlink.

There are 2 methods to transmit signals on downlink: *TDD (Time Division Duplexing)* and *FDD (Frequency Division Duplexing)*. In TDD, uplink and downlink transmissions occur at different times but share the same frequency band. The gain of the uplink channel can be used for the downlink channel due to channel

reciprocity. So the overhead per block is K . In FDD, uplink and downlink transmissions occur simultaneously but use different frequency bands. The uplink and downlink channels are not reciprocal, so the base station needs to estimate the downlink channel using M pilots on downlink, and users need to feed back the estimated channel information to the base station. The overhead per block is $M + K/2$ (K pilots + M feedback on uplink, M pilots on downlink). So TDD is more suitable for Massive MIMO systems.

The frame structure of TDD Massive MIMO system is as follows. Each coherence block has size of $\tau_c = T_c B_c$. The pilot length is τ_p and the data transmission length is $\tau_d = \tau_c - \tau_p$.

On uplink, the received signal at the base station is (normalized):

$$\mathbf{y} = \sqrt{\rho_{ul}} \mathbf{H} \mathbf{x} + \mathbf{w}$$

where ρ_{ul} is maximum power, $E[|x_k|^2] \leq 1$, $h_{ik} \sim CN(0, \beta_k)$ and $\mathbf{w} \sim CN(0, \mathbf{I}_N)$ (large scale fading coefficient β_k is not equal for different users). The maximum SNR is $\rho_{ul}\beta_k$ for user k , so:

$$\rho_{ul} = \frac{\text{Uplink radiated power} \cdot \text{Antenna gain}}{N_0 B}$$

Usually $B = 10$ MHz, Radiated power = 100 mW, Antenna gain = 0 dBi, $N_0 = 10^{-17}$. And for β_k :

$$\beta_k = 10^{-1.53} \left(\frac{r_k}{1}\right)^{-3.76}$$

where $r_k \geq 35$ m is the distance between user k and the base station in m. Usually $r_k = 35$ m $\rightarrow \beta_k = -73$ dB, $r_k = 1000$ m $\rightarrow \beta_k = -128$ dB.

Pilot contamination occurs when the same pilot sequences are reused in different cells, leading to interference during channel estimation. For example, consider two neighboring cells where user 1 in cell A and user 2 in cell B both use the same pilot sequence for channel estimation. When the base station in cell A receives the pilot signal from user 1, it also picks up interference from user 2 in cell B, which is using the same pilot sequence.

3.8 Applications

Ultra-reliable low-latency communications (URLLC) is one of the three main use cases defined for 5G networks, alongside massive machine-type communications (mMTC). URLLC is designed to support applications that require extremely low latency and high reliability, such as autonomous vehicles, remote surgeries, industrial automation, and mission-critical communications. mMTC focuses on connecting a large number of devices with low data rates and low power consumption, such as IoT sensors and smart meters.

A random channel $\mathbf{h} \sim CN(0, \mathbf{I}_N)$ is considered. Variations of effective channel gain $\|\mathbf{h}\|^2/M$ has the mean value 1 and variance $1/M$. When M is very large, by law of large numbers we have $\|\mathbf{h}\|^2/M \rightarrow 1$. This phenomenon is called **channel hardening**, which means that the effective channel gain becomes nearly deterministic as the number of antennas increases. It solve the problem of deep fading in wireless channels, improving the reliability of the communication link. And it also make all the subcarriers experience similar channel conditions, simplifying the design of communication systems (No need to schedule based on fading). Each user gets the whole bandwidth, whenever needed.

When BS uses M antennas to serve a single-antenna user, the received signal power is increased by a factor of M . This phenomenon is called **array gain**, which is usually expressed in decibels (dB) as $10 \log_{10}(M)$. There are 2 ways to exploit array gain: *Range extension* and *Low-power operation*. In range extension, the base station uses the array gain to increase the coverage area, while maintaining the power per antenna.

For the user already in the cell, it can get higher SNR and better performance. In low-power operation, the base station reduces the power per antenna while maintaining the same coverage area. This helps to reduce interference to other cells and save energy.

For mMTC applications, when not using Massive MIMO, each device needs high transmit power to overcome path loss and fading, leading to increased interference and reduced battery life.

$$\text{SNR} = \text{Transmit Power} + \text{Antenna Gain} - \text{Path Loss} - \text{Noise power}$$

Example of this is transmit power = 20 dBm, Antenna gain = 2.15 dBi for each, Path loss = 150 dB, Noise power = -120 dBm, so SNR = -5.7 dB. When using Massive MIMO, assume we have $M = 100$ antennas at the base station, the array gain is $10 \log_{10}(100) = 20$ dB. So the SNR becomes: $\text{SNR} = -5.7 \text{ dB} + 20 \text{ dB} = 14.3 \text{ dB}$. This means that each device can reduce its transmit power significantly while still achieving a good SNR at the base station. In the example, the transmit power can be reduced from 20 dBm to 10 dB.