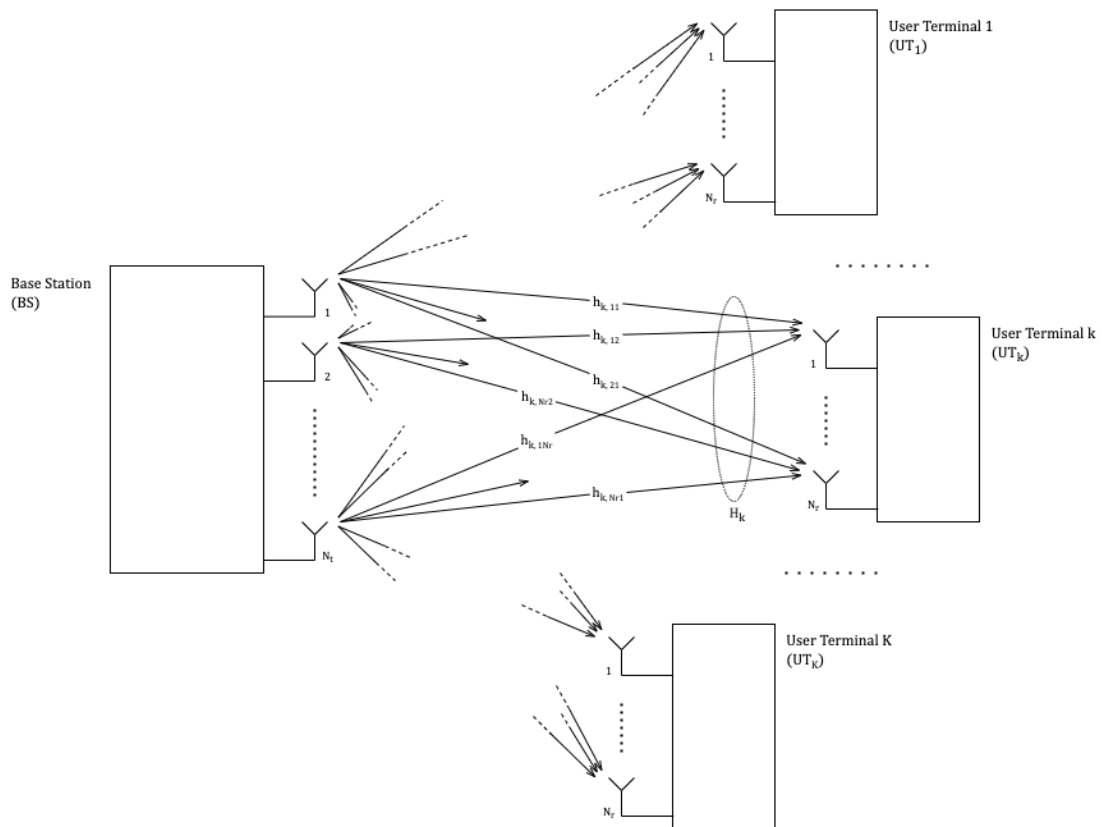# Downlink MU-MIMO

Wannes Baes

Figure 1: The downlink mu-mimo system.

1 Notations, Parameters & Symbols

1.1 System Parameters

K Number of user terminals

N_t Number of transmit antennas on the base station

N_r Number of receive antennas on each user terminal

N_s Number of data streams for each user terminal

We assume that the base station (BS) is equipped with more antennas than all user terminals (UTs) together, and that a UT cannot receive more data streams than it has antennas. This implies the relation $N_t \geq KN_r \geq KN_s$.

Without loss of generality, we assume that all UTs are equipped with the same number of receive antennas. In practical scenarios where this assumption does not hold, N_r can be defined as the maximum number of antennas among all UTs. For a UT_k with fewer antennas ($N_{r,k} < N_r$), no data streams can be assigned to the additional ($N_r - N_{r,k}$) antenna positions.

1.2 System Component Symbols

Power Allocator

- $p_{k,s} \in \mathbb{R}_{\geq 0}$ — power allocated to data stream s intended for UT_k

- $\mathbf{P}_k = \mathrm{diag}\left(\sqrt{p_{k,1}}, \ldots, \sqrt{p_{k,N_s}}\right) \in \mathbb{R}_{\geq 0}^{N_s \times N_s}$ — power allocation matrix for UT_k

- $\mathbf{P} = \mathrm{diag}(\mathbf{P}_1, \ldots, \mathbf{P}_K) \in \mathbb{R}_{\geq 0}^{KN_s \times KN_s}$ — compound power allocation matrix

Precoder

- $\mathbf{F}_k \in \mathbb{C}^{N_t \times N_s}$ — precoding matrix for UT_k

- $\mathbf{F} = [\mathbf{F}_1 \ \ldots \ \mathbf{F}_K] \in \mathbb{C}^{N_t \times KN_s}$ — compound precoding matrix

Channel

- $h_{k,n_r n_t} \in \mathbb{C}$ — channel gain between transmit antenna n_t at the BS and receive antenna n_r at UT_k

- $\mathbf{H}_k \in \mathbb{C}^{N_r \times N_t}$ — channel matrix between the BS and UT_k

- $\mathbf{H} = \left[\mathbf{H}_1^T \ \ldots \ \mathbf{H}_K^T\right]^T \in \mathbb{C}^{KN_r \times N_t}$ — compound channel matrix

Noise

- $n_{k,s} \sim \mathcal{N}_\mathcal{C}(\mathbf{0}, N_0) \in \mathbb{C}$ — noise sample at data stream s of UT_k

- $\mathbf{n}_k \sim \mathcal{N}_\mathcal{C}(\mathbf{0}, N_0\, \mathbf{I}_{N_r}) \in \mathbb{C}^{N_r \times 1}$ — noise vector at UT_k

- $\mathbf{n} = \begin{bmatrix} \mathbf{n}_1^T & \dots & \mathbf{n}_K^T \end{bmatrix}^T \sim \mathcal{N}_\mathcal{C}(\mathbf{0}, N_0\, \mathbf{I}_{KN_r}) \in \mathbb{C}^{KN_r \times 1}$ — compound noise vector

Combiner

- $\mathbf{W}_k \in \mathbb{C}^{N_s \times N_r}$ — combining matrix for UT_k

- $\mathbf{W} = \mathrm{diag}(\mathbf{W}_1, \dots, \mathbf{W}_K) \in \mathbb{C}^{KN_s \times KN_r}$ — compound combining matrix

1.3   System Signal Symbols

Data Symbols

- $a_{k,s} \in \mathcal{C}$ — data symbol in data stream s intended for UT_k

- $\mathbf{a}_k = [a_{k,1} \; \dots \; a_{k,N_s}]^T \in \mathcal{C}^{N_s \times 1}$ — data symbol vector intended for UT_k

- $\mathbf{a} = [\mathbf{a}_1^T \; \mathbf{a}_2^T \; \dots \; \mathbf{a}_K^T]^T \in \mathcal{C}^{KN_s \times 1}$ — compound data symbol vector

Transmitted Signal

- $x_{n_t} = \sum_{k=1}^{K} \cdot \sum_{s=1}^{N_s} (\mathbf{F}_k)_{(n_t,s)} \cdot p_{k,s} a_{k,s} \in \mathbb{C}$ — transmitted signal from transmit antenna n_t at the BS

- $\mathbf{x} = \begin{bmatrix} x_1 & \dots & x_{N_t} \end{bmatrix}^T = \mathbf{F}\,\mathbf{P}\mathbf{a} = \sum_{k=1}^{K} \mathbf{F}_k\, \mathbf{P}_k \mathbf{a}_k \in \mathbb{C}^{N_t \times 1}$ — transmitted signal vector

Received Signal

- $y_{k,n_r} = \sum_{n_t=1}^{N_t} h_{k,\,n_r\, n_t} x_{n_t} + n_{k,n_r} \in \mathbb{C}$ — received signal on antenna n_r of UT_k

- $\mathbf{y}_k = \mathbf{H}_k \mathbf{x} + \mathbf{n}_k \in \mathbb{C}^{N_r \times 1}$ — received signal vector at UT_k

- $\mathbf{y} = \begin{bmatrix} \mathbf{y}_1^T & \dots & \mathbf{y}_K^T \end{bmatrix}^T = \mathbf{H}\mathbf{x} + \mathbf{n} \in \mathbb{C}^{KN_r \times 1}$ — compound received signal vector

Scaled Decision Variable

- $z_{k,s} = \sum_{n_r=1}^{N_r} (\mathbf{W}_k)_{(s,n_r)}\, y_{k,n_r} \in \mathbb{C}$ — scaled decision variable in data stream s of UT_k

- $\mathbf{z}_k = \mathbf{W}_k \mathbf{y}_k \in \mathbb{C}^{N_s \times 1}$ — scaled decision variable vector at UT_k

- $\mathbf{z} = [\mathbf{z}_1^T \; \dots \; \mathbf{z}_K^T]^T = \mathbf{W}\mathbf{y} \in \mathbb{C}^{KN_s \times 1}$ — compound decision variable vector

2 System Equations

We focus on the data transmission at a certain time instant m only, hence we drop the discrete time index $[m]$ in the following equations for clarity. The received signal at UT_k at time instant m is expressed as:

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{x} + \mathbf{n}_k \tag{1}$$

The transmitted signal is constructed by allocating power to the data symbols vectors, followed by precoding: $\mathbf{x} = \sum_{k'=1}^{K} \mathbf{F}_{k'} \cdot \mathbf{P}_{k'} \mathbf{a}_{k'}$.

$$\mathbf{y}_k = \mathbf{H}_k \cdot \sum_{k'=1}^{K} \mathbf{F}_{k'} \cdot \mathbf{P}_{k'} \mathbf{a}_{k'} + \mathbf{n}_k \tag{2}$$

The received signal at UT_k is then combined to obtain the scaled decision variable $\mathbf{z}_k = \mathbf{W}_k \mathbf{y}_k$.

$$\begin{aligned}
\mathbf{z}_k &= \mathbf{W}_k \, \mathbf{H}_k \cdot \sum_{k'=1}^{K} \mathbf{F}_{k'} \cdot \mathbf{P}_{k'} \mathbf{a}_{k'} + \mathbf{W}_k \, \mathbf{n}_k \\
&= \underbrace{\mathbf{W}_k \, \mathbf{H}_k \, \mathbf{F}_k \cdot \mathbf{P}_k \, \mathbf{a}_k}_{\text{useful signal}} + \underbrace{\sum_{\substack{k'=1 \\ k' \neq k}}^{K} \mathbf{W}_k \, \mathbf{H}_k \, \mathbf{F}_{k'} \cdot \mathbf{P}_{k'} \, \mathbf{a}_{k'}}_{\text{inter-user interference}} + \underbrace{\mathbf{W}_k \, \mathbf{n}_k}_{\text{noise}}
\end{aligned} \tag{3}$$

The scaled decision variable $\mathbf{z}_k$ thus consists of three components: the useful signal intended for UT_k, the inter-user interference caused by the data streams intended for other UTs, and the noise component.

We express the scaled decision variable $z_{k,s}$ in data stream s of UT_k as follows:

$$\begin{bmatrix} z_{k,1} \\ \vdots \\ z_{k,N_s} \end{bmatrix} = (\mathbf{W}_k \mathbf{H}_k \mathbf{F}_k) \begin{bmatrix} \sqrt{p_{k,1}}\, a_{k,1} \\ \vdots \\ \sqrt{p_{k,N_s}}\, a_{k,N_s} \end{bmatrix} + \sum_{\substack{k'=1 \\ k' \neq k}}^{K} (\mathbf{W}_k \mathbf{H}_k \mathbf{F}_{k'}) \begin{bmatrix} \sqrt{p_{k',1}}\, a_{k',1} \\ \vdots \\ \sqrt{p_{k',N_s}}\, a_{k',N_s} \end{bmatrix} + \mathbf{W}_k \begin{bmatrix} n_{k,1} \\ \vdots \\ n_{k,N_r} \end{bmatrix}$$

$$\begin{aligned}
z_{k,s} &= \underbrace{(\mathbf{W}_k \, \mathbf{H}_k \, \mathbf{F}_k)_{(s,s)} \cdot \sqrt{p_{k,s}}\, a_{k,s}}_{\text{useful symbol}} + \underbrace{\sum_{\substack{s'=1 \\ s' \neq s}}^{N_s} (\mathbf{W}_k \, \mathbf{H}_k \, \mathbf{F}_k)_{(s,s')} \cdot \sqrt{p_{k,s'}}\, a_{k,s'}}_{\text{inter-stream interference}} \\
&\quad + \underbrace{\sum_{\substack{k'=1 \\ k' \neq k}}^{K} \sum_{s'=1}^{N_s} (\mathbf{W}_k \, \mathbf{H}_k \, \mathbf{F}_{k'})_{(s,s')} \cdot \sqrt{p_{k',s'}}\, a_{k',s'}}_{\text{inter-user interference}} + \underbrace{\sum_{n_r=1}^{N_r} (\mathbf{W}_k)_{(s,n_r)}\, n_{k,n_r}}_{\text{noise}}
\end{aligned} \tag{4}$$

We express the scaled decision variables $\mathbf{z}$ for all UTs in a compact form as follows:

$$\mathbf{z} = \mathbf{W}\,\mathbf{H}\,\mathbf{F} \cdot \mathbf{P}\,\mathbf{a} + \mathbf{W}\,\mathbf{n} \tag{5}$$

3 Problem Formulation

3.1 Achievable Rate & Rate Region

Achievable Rate   A data rate R is considered achievable if there exists a coding and modulation scheme that enables reliable communication at that rate. More formally, a rate is achievable if the probability of decoding error can be made arbitrarily small by increasing the codeword length without bound, while keeping all system parameters fixed.

Each UT_k receives N_s parallel data streams. Its total achievable rate is the sum of the rates of its individual streams:

$$R_k = \sum_{s=1}^{N_s} R_{k,s}, \tag{6}$$

where $R_{k,s}$ denotes the achievable rate of stream s for UT_k.

According to the Shannon–Hartley theorem [1], the maximum achievable rate corresponds to the channel capacity, i.e., the maximum mutual information between the transmitted and received signals. The achievable rate of stream s for UT_k is expressed as:

$$R_{k,s} = 2B \, \log_2\big(1 + \mathrm{SINR}_{k,s}\big), \tag{7}$$

where $2B$ is the allocated bandwidth to each data stream, and $\mathrm{SINR}_{k,s}$ denotes the signal-to-interference-plus-noise ratio of stream s for UT_k.

The SINR of data stream s for UT_k is defined as the ratio between the power of the useful signal and the power of the interference and noise components:

$$\mathrm{SINR}_{k,s} = \frac{p_{k,s} \left| (\mathbf{W}_k \mathbf{H}_k \mathbf{F}_k)_{(s,s)} \right|^2}{\sum\limits_{\substack{s'=1 \\ s' \neq s}}^{N_s} p_{k,s'} \left| (\mathbf{W}_k \mathbf{H}_k \mathbf{F}_k)_{(s,s')} \right|^2 + \sum\limits_{\substack{k'=1 \\ k' \neq k}}^{K} \sum\limits_{s'=1}^{N_s} p_{k',s'} \left| (\mathbf{W}_k \mathbf{H}_k \mathbf{F}_{k'})_{(s,s')} \right|^2 + N_0 \left\| (\mathbf{W}_k)_{(s,s)} \right\|^2} \tag{8}$$

Rate Region   The rate region $\mathcal{R}$ of a multi-user communication system is the set of all rate tuples $(R_1, R_2, \ldots, R_K)$ for the K user terminals that can be simultaneously achieved given the system constraints.
The rate region describes the trade-offs in multi-user performance: increasing the rate of one user may reduce the rates achievable by others due to interference and resource sharing. One should always aim to operate at the boundary of the rate region, known as the Pareto boundary, where no user's rate can be increased without decreasing another's.

3.2 Weighted Sum-Rate Optimization

We aim to optimize the performance of the downlink MU-MIMO system.
Let us define the weighted sum-rate (WSR) as the sum of the achievable rates of all UTs, each scaled by a weight factor ν_k. The weight represents the importance of the UT in the optimization process.

$$\text{WSR} = \sum_{k=1}^{K} \nu_k \, R_k \tag{9}$$

The WSR optimization problem with respect to the precoding, combining, and power allocation matrices ($\mathbf{F}$, $\mathbf{W}$, $\mathbf{P}$) under the total available transmit power P_t constraint is formulated as:

$$\max_{\mathbf{F}, \mathbf{W}, \mathbf{P}} \ \text{WSR} \tag{10}$$
$$\text{s.t. } \|\mathbf{F}\,\mathbf{P}\|_{\text{F}}^2 \leq P_t$$

4 Non-coördinated Beamforming with Optimal RSV Combiner and ZF Precoder

Non-coördinated   There is no feedforward message exchange between the BS and the UTs. Each UT estimates its own channel and computes its own combiner based on this channel estimate. The combiner is therefore heuristic instead of optimal from a global system perspective.

RSV Combiner   The rows of the combining matrix are chosen as the right singular vectors corresponding to the N_s largest singular values of the channel matrix:

$$\mathbf{W}_k = \mathbf{U}_k^H[: N_s], \tag{11}$$

where $\mathbf{U}_k$ is obtained from the SVD of the channel matrix $\mathbf{H}_k = \mathbf{U}_k \, \boldsymbol{\Sigma}_k \, \mathbf{V}_k^H$.

ZF Precoder [2]   The precoding matrix is computed at the BS such that all interference is completely eliminated.

Let us derive an expression for the ZF precoding matrix. No interference is allowed, which means the the system is decomposed into $K\,Ns$ parallel SISO channels:

$$\mathbf{z} = \mathbf{P}\,\mathbf{a} + \mathbf{W}\,\mathbf{n}$$

This implies that the following condition on the MIMO transfer matrix $\mathbf{T} = \mathbf{W}\mathbf{H}\mathbf{F}$ must hold:

$$\mathbf{T} = \mathbf{W}\,\mathbf{H}\,\mathbf{F} = \mathbf{I}_{KN_s}$$

Therefore, the ZF precoder is obtained by computing the pseudoinverse of the effective channel $\mathbf{H}_{\mathrm{BS}}$ seen by the BS (i.e., the compound channel followed by the compound combiner):

$$\mathbf{F} = (\mathbf{W}\,\mathbf{H})^{\dagger} = \mathbf{H}_{\mathrm{BS}}^{\dagger} \tag{12}$$

The pseudoinverse of $\mathbf{H}_{\mathrm{BS}}$ can be computed as follows:

$$\mathbf{H}_{\mathrm{BS}}^{\dagger} = \begin{cases} \mathbf{H}_{\mathrm{BS}}^{H} \left(\mathbf{H}_{\mathrm{BS}}\mathbf{H}_{\mathrm{BS}}^{H}\right)^{-1} & \text{if } \mathbf{H}_{\mathrm{BS}} \text{ is full rank} \\ \mathbf{V}_{\mathrm{eff}}\,\boldsymbol{\Sigma}_{\mathrm{eff}}^{\dagger}\,\mathbf{U}_{\mathrm{eff}}^{H} & \text{if } \mathbf{H}_{\mathrm{BS}} \text{ is rank-deficient} \end{cases}$$

where $\mathbf{H}_{\mathrm{BS}} = \mathbf{U}_{\mathrm{BS}}\,\boldsymbol{\Sigma}_{\mathrm{BS}}\,\mathbf{V}_{\mathrm{BS}}^{H}$ is the SVD of the effective channel and $\boldsymbol{\Sigma}_{\mathrm{eff}}^{\dagger}$ is obtained by taking the reciprocal of the non-zero singular values in $\boldsymbol{\Sigma}_{\mathrm{eff}}$ and transposing the resulting matrix.

Optimal The power allocation matrix is optimal in the sense that it maximizes the WSR for the given combiner and precoder, under the total transmit power constraint.

Let us derive an expression for the ZF precoding matrix. For a given channel, precoder and combiner, the transmit power constraint can be simplified as follows:

$$\begin{aligned} P_t &\leq \|\mathbf{F}\,\mathbf{P}\|_{\mathrm{F}}^2 \\ &= \left\|\left(\mathbf{H}_{\mathrm{BS}}^{H}\left(\mathbf{H}_{\mathrm{BS}}\mathbf{H}_{\mathrm{BS}}^{H}\right)^{-1}\right)\mathbf{P}\right\|_{\mathrm{F}}^2 \\ &= \mathrm{Tr}\left[\left(\left(\mathbf{H}_{\mathrm{BS}}^{H}\left(\mathbf{H}_{\mathrm{BS}}\mathbf{H}_{\mathrm{BS}}^{H}\right)^{-1}\right)\mathbf{P}\right)\left(\left(\mathbf{H}_{\mathrm{BS}}^{H}\left(\mathbf{H}_{\mathrm{BS}}\mathbf{H}_{\mathrm{BS}}^{H}\right)^{-1}\right)\mathbf{P}\right)^{H}\right] \\ &= \mathrm{Tr}\left[\mathbf{H}_{\mathrm{BS}}^{H}\left(\mathbf{H}_{\mathrm{BS}}\mathbf{H}_{\mathrm{BS}}^{H}\right)^{-1}\mathbf{P}\mathbf{P}^{H}\left(\left(\mathbf{H}_{\mathrm{BS}}\mathbf{H}_{\mathrm{BS}}^{H}\right)^{-1}\right)^{H}\mathbf{H}_{\mathrm{BS}}\right] \\ &= \mathrm{Tr}\left[\mathbf{P}\mathbf{P}^{H}\left(\left(\mathbf{H}_{\mathrm{BS}}\mathbf{H}_{\mathrm{BS}}^{H}\right)^{-1}\right)^{H}\mathbf{H}_{\mathrm{BS}}\mathbf{H}_{\mathrm{BS}}^{H}\left(\mathbf{H}_{\mathrm{BS}}\mathbf{H}_{\mathrm{BS}}^{H}\right)^{-1}\right] \\ &= \mathrm{Tr}\left[\mathbf{P}^2 \cdot \left(\mathbf{H}_{\mathrm{BS}}\mathbf{H}_{\mathrm{BS}}^{H}\right)^{-1}\right] \\ &= \sum_{k=1}^{K}\sum_{s=1}^{N_s} p_{k,s} \cdot \left(\left(\mathbf{H}_{\mathrm{BS}}\mathbf{H}_{\mathrm{BS}}^{H}\right)^{-1}\right)_{((k-1)N_s+s,\,(k-1)N_s+s)} \end{aligned} \tag{13}$$

The optimal power allocation problem has now become equivalent to the optimal power allocation in a SU-MIMO system with KN_s parallel eigenchannels with channel gains σ_i given by $\left(\left(\mathbf{H}_{\mathrm{BS}}\mathbf{H}_{\mathrm{BS}}^{H}\right)^{-1}\right)_{(i,i)}$, which can be solved using the waterfilling algorithm.

4.1   Simulation Results

Work In Progress...

References

[1] H. Steendam, *Syllabus Information Theory.* Ghent, Belgium: Ghent University, 2024-2025.

[2] A. Wiesel, Y. Eldar, and S. Shamai, "Zero-forcing precoding and generalized inverses," *Trans. Sig. Proc.*, vol. 56, p. 4409–4418, Sept. 2008.