Algorithms for Wireless Communications

Andreas Burg

Institute for Integrated Signal Processing Systems, RWTH Aachen

Telecommunications Circuits Laboratory, EPFL

Lecture: Spatial Multiplexing in MIMO Channels

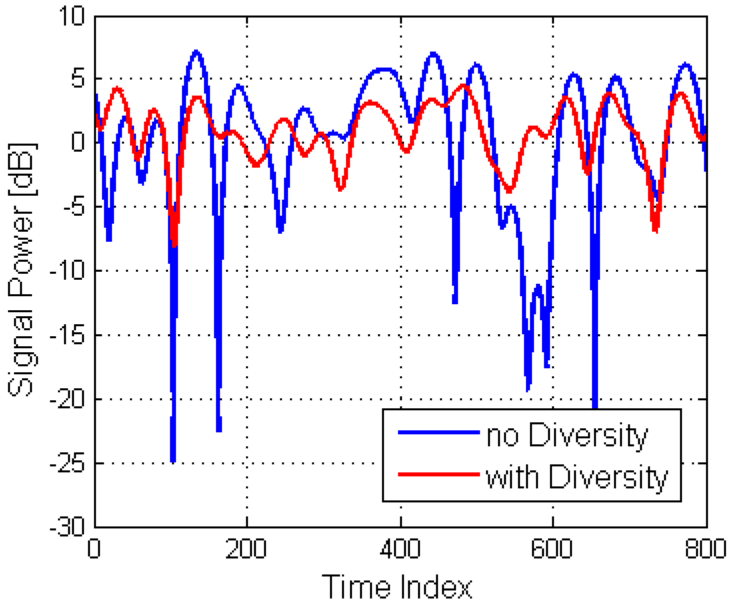
ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

RWTH AACHEN UNIVERSITY

Spatial Multiplexing with MIMO Systems

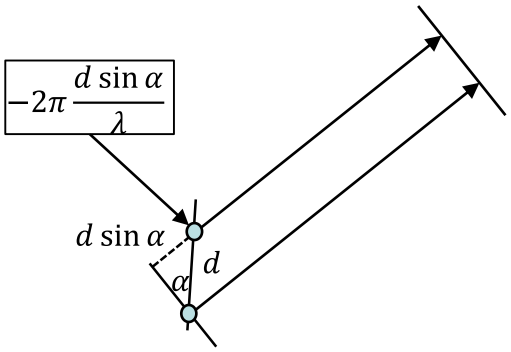
Spatial Diversity (e.g. Alamouti's scheme):

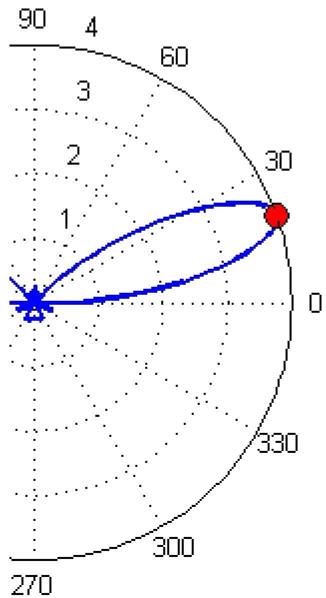
- Requires uncorrelated antenna elements (>> λ/2 distance)
- "Unlikely that all antennas are in deep fades"
- Reduces fast fading amplitude variations
- Makes the transmission more robust



Beamforming:

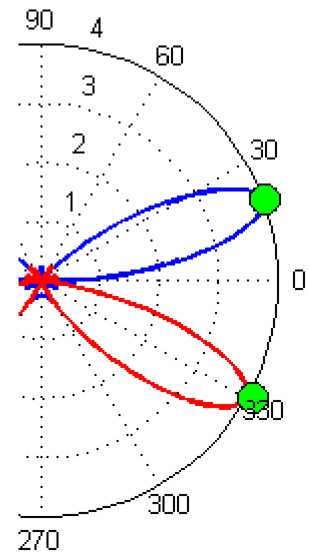
- Requires correlated antenna elements (λ/2 distance)
- Focusing of power into a desired direction
- Increases signal power in this direction
- Decreases interference power in other directions





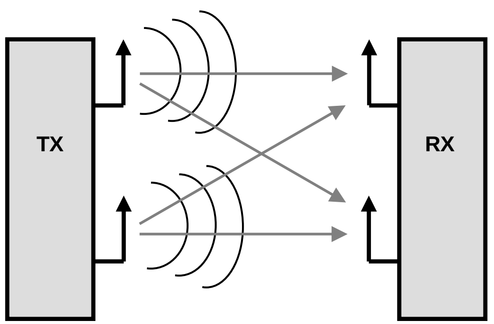
Space-Division Multiple Access:

- Users are separated by their spatial signature
- Multiple data streams, but only one per user
- Does not require multiple receive antennas
- Does not require complex receiver processing



Spatial Multiplexing

- Requires uncorrelated antenna elements
- Transmission of multiple code words along independent spatial modes
- Increases throughput for high link SNR
- Requires multiple receive antennas
- Requires complex receiver processing



- **Consider a system with**
  - n_T transmit antennas and
  - n_R receiveantennas and

- **Collect the transmitted symbols, the noise- and the received- samples in vectors**

$$\underline{\mathbf{x}} = \begin{pmatrix} x_1 \\ \vdots \\ x_{n_T} \end{pmatrix} \quad ; \quad \underline{\mathbf{w}} = \begin{pmatrix} w_1 \\ \vdots \\ w_{n_R} \end{pmatrix} \quad ; \quad \underline{\mathbf{y}} = \begin{pmatrix} y_1 \\ \vdots \\ Y_{n_R} \end{pmatrix}$$

and represent the channel as a n_R x n_T matrix

$$\underline{\mathbf{H}} = \begin{pmatrix} h_{1,1} & \cdots & \cdots & h_{1,n_T} \\ \vdots & & & \vdots \\ h_{n_R,1} & \cdots & \cdots & h_{n_R,n_t} \end{pmatrix}$$

yields

$$\underline{\mathbf{Y}} = \underline{\mathbf{H}}\,\underline{\mathbf{x}} + \underline{\mathbf{w}}$$

- **For OFDM, we can consider each tone as an independent MIMO system. For the discussion, we consider a single subcarrier only (flat fading). Thus, we can drop the subcarrier index.**

- **We further assume the channel is time invariant during a burst. Thus, we can drop the time indices in the channel matrix and the model is**

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

where x is the n_T-dimensional complex vector of transmit signals,
y is the n_R-dimensional complex vector of receive signals and
w the n_R-dimensional complex Gaussian noise vector with
w~*CN*(0, $N_0 \mathbf{I}_{n_R}$).
There is a total power constraint P' on the signals from the transmit antennas.

(Note that the underlining has been dropped for simplicity)

- **Let the singular value decomposition of H be**

$$\mathbf{H} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^{\mathrm{H}}$$

- **Recall that the first $n_{min}=\min(n_R,n_T)$ diagonal elements of Λ are the ordered positive square roots of the eigenvalues of the matrix HH^H. We can rewrite the SVD as**

$$\mathbf{H} = \sum_{i=1}^{n_{min}} \lambda_i \mathbf{u}_i \mathbf{v}_i^{\mathrm{H}}$$

where the u_i are the eigenvectors of HH^H and the v_i are the eigenvectors of H^HH.

- **The rank of H is precisely the number of non-zero singular values λ_i.**

- **If we define**

$$\tilde{\mathbf{x}} := \mathbf{V}^{\mathrm{H}}\mathbf{x} \Leftrightarrow \mathbf{x} := \mathbf{V}\tilde{\mathbf{x}} \quad , \quad \tilde{\mathbf{y}} := \mathbf{U}^{\mathrm{H}}\mathbf{y} \quad , \quad \tilde{\mathbf{w}} := \mathbf{U}^{\mathrm{H}}\mathbf{w}$$

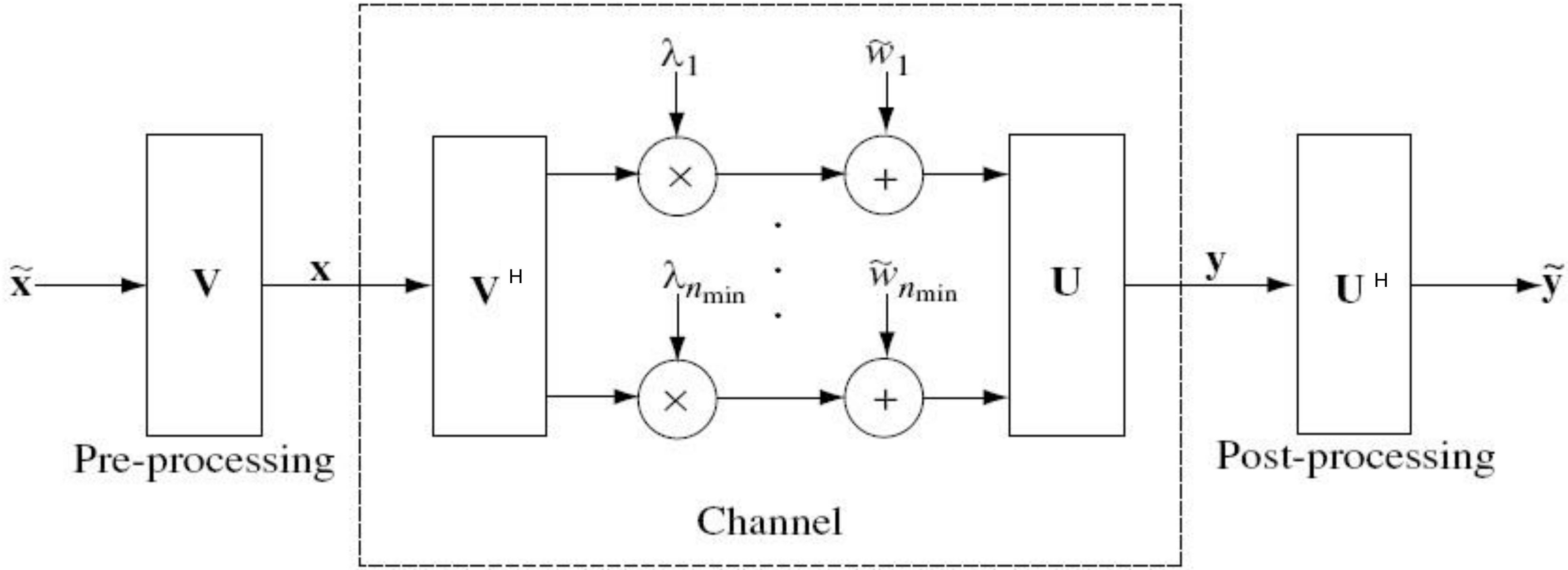
then we can rewrite the transmission equation

$$\mathbf{U}^{\mathrm{H}}\mathbf{y} = \mathbf{U}^{\mathrm{H}}\mathbf{H}\mathbf{x} + \mathbf{U}^{\mathrm{H}}\mathbf{w} = \mathbf{U}^{\mathrm{H}}\mathbf{H}\mathbf{V}\tilde{\mathbf{x}} + \mathbf{U}^{\mathrm{H}}\mathbf{w}$$

$$\Rightarrow \qquad \tilde{\mathbf{y}} = \boldsymbol{\Lambda}\tilde{\mathbf{x}} + \tilde{\mathbf{w}}$$

- **Since U is unitary, $\tilde{w} \sim \mathcal{CN}(0, N_0 I_{n_r})$ has the same distribution as w, and $\tilde{x}$ has the same power as x (i.e. $||x||^2 = ||\tilde{x}||^2$).**

- **The SVD can be interpreted as two coordinate transformations (two base transformations).**



$$\mathbf{x} := \mathbf{V}\tilde{\mathbf{x}} \qquad \mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{V}^H\mathbf{x} + \mathbf{w} \qquad \tilde{\mathbf{w}} := \mathbf{U}^H\mathbf{w} \quad , \quad \tilde{\mathbf{y}} := \mathbf{U}^H\mathbf{y}$$

$$= \mathbf{U}\left(\boldsymbol{\Lambda}\mathbf{V}^H\mathbf{x} + \tilde{\mathbf{w}}\right)$$

$$\tilde{\mathbf{y}} = \boldsymbol{\Lambda}\tilde{\mathbf{x}} + \tilde{\mathbf{w}}$$

Source: David Tse, Pramod Viswanath, „Fundamentals of Wireless Communication“, Cambridge Press 2005

Interpretation

- **The Singular Value Decomposition (SVD) can be interpreted as two coordinate transformations (or base transformations).**

  If
  - the input is expressed in terms of a coordinate system defined by the columns of V and
  - the output is expressed in terms of a coordinate system defined by the columns of U,
  - then the input/output relationship is very simple:

    we have n_{min} independent parallel channels.

Eigenmode transmission

MIMO Capacity without transmit Channel Knowledge

- **The maximum rate of reliable communication (achievable rate) for k parallel Gaussian channels is**

$$R = \sum_{n=0}^{k} \log_2 \left(1 + \frac{P_n' \lambda_n^{\ 2}}{N_0} \right) \textit{bits/channel use}$$

k equals the number of non-zero eigenvalues $k \le n_{\min}$

- **If transmitter has no channel knowledge: distribute power equally across dimensions**

$$P_n' = \frac{1}{k} P' \qquad , \quad n = 1, \ldots, k$$

the capacity is the maximum achievable rate for any power allocation meeting this constraint

$$R = \sum_{n=0}^{k} \log_2 \left(1 + \frac{1}{k} \frac{P'}{N_0} \lambda_n^{\ 2} \right) \textit{bits/channel use}$$

MIMO Capacity without transmit Channel Knowledge

- **Rewriting the capacity expression of the subchannels yields**

$$R = \log_2 \prod_{n=0}^{k} \left(1 + \frac{1}{k}\frac{P'}{N_0}\lambda_n{}^2\right) \textit{bits/channel use}$$

- **We use the following relationship that relates the product of the Eigenvalues of a matrix to its determinant**

$$\det(\mathbf{A}) = \prod_{n=0}^{rank} \left(\lambda_n{}^2\right)$$

- **We further use that $1+\lambda_n^2$ are the Eigenvalues of $1+Q$ if λ_n^2 are the Eigenvalues of Q**
- **Therefore:**

$$R = \log_2 \prod_{n=0}^{k} \left(1 + \frac{1}{k}\frac{P'}{N_0}\lambda_n{}^2\right) = \log_2\left(\det\left(I + \frac{1}{k}\frac{P'}{N_0}\mathbf{H}\mathbf{H}^H\right)\right) \textit{bits/channel use}$$

Derivation: Optimum Power Allocation in MIMO System (Water Filling)

- **The maximum rate of reliable communication (achievable rate) for k parallel Gaussian channels is**

$$R = \sum_{n=0}^{k} \log_2\left(1 + \frac{P_n' \lambda_n^2}{N_0}\right) \text{ bits/channel use}$$

k equals the number of non-zero eigenvalues $k \le n_{\min}$

- **Under a (common) power constraint**

$$\sum_{n=1}^{k} P_n' = P' \qquad P_n' \ge 0 \quad , \quad n = 1,\ldots,k$$

the capacity is the maximum achievable rate for any power allocation meeting this constraint

$$C = \max_{P_1',\ldots,P_k'} \sum_{n=1}^{k} \log_2\left(1 + \frac{P_n' |\lambda_n|^2}{N_0}\right) \quad \text{with} \quad \sum_{n=1}^{k} P_n' = P' \qquad P_n' \ge 0$$

- **We can determine the optimal power allocation using the method of Lagrange multiplier**

$$L(\lambda, P_1', \ldots, P'_k) = \sum_{n=1}^{k} \log_2\left(1 + \frac{P_n'|\lambda_n|^2}{N_0}\right) - \lambda\left(\sum_{n=1}^{k} P_n' - P'\right)\log_2 e$$

- **Since the individual power allocations must be nonnegative, the optimality condition is**

$$\frac{\partial \mathrm{L}}{\partial \mathrm{P_n}'} \begin{cases} = 0 & \text{if } \mathrm{P_n}' > 0 \\ \leq 0 & \text{if } \mathrm{P_n}' = 0 \end{cases}$$

i.e. using the maximizing value for positive allocations (first rule) or using a zero allocation if larger allocations would reduce L (due to a negative derivative, second rule)

- **Defining $x^+=\max(x,0)$, the resulting optimal power allocation is**

$$P_n' = \left(\frac{1}{\lambda} - \frac{N_0}{|\lambda_n|^2}\right)^+ \quad \text{with } \lambda \text{ such that} \quad P' = \sum_{n=1}^{k}\left(\frac{1}{\lambda} - \frac{N_0}{|\lambda_n|^2}\right)^+$$

- **We have already seen examples of parallel Gaussian channels (e.g. OFDM in frequency dimension). Here the spatial dimension plays the same role as the frequency dimension in OFDM. The capacity* is by now familiar:**

$$C'_{n_{min}} = \sum_{i=1}^{n_{min}} \log_2\left(1 + \frac{P_i'\lambda_i^2}{N_0}\right) \quad \textit{bits/s/Hz}$$

where the P_i' are the waterfilling power allocations:

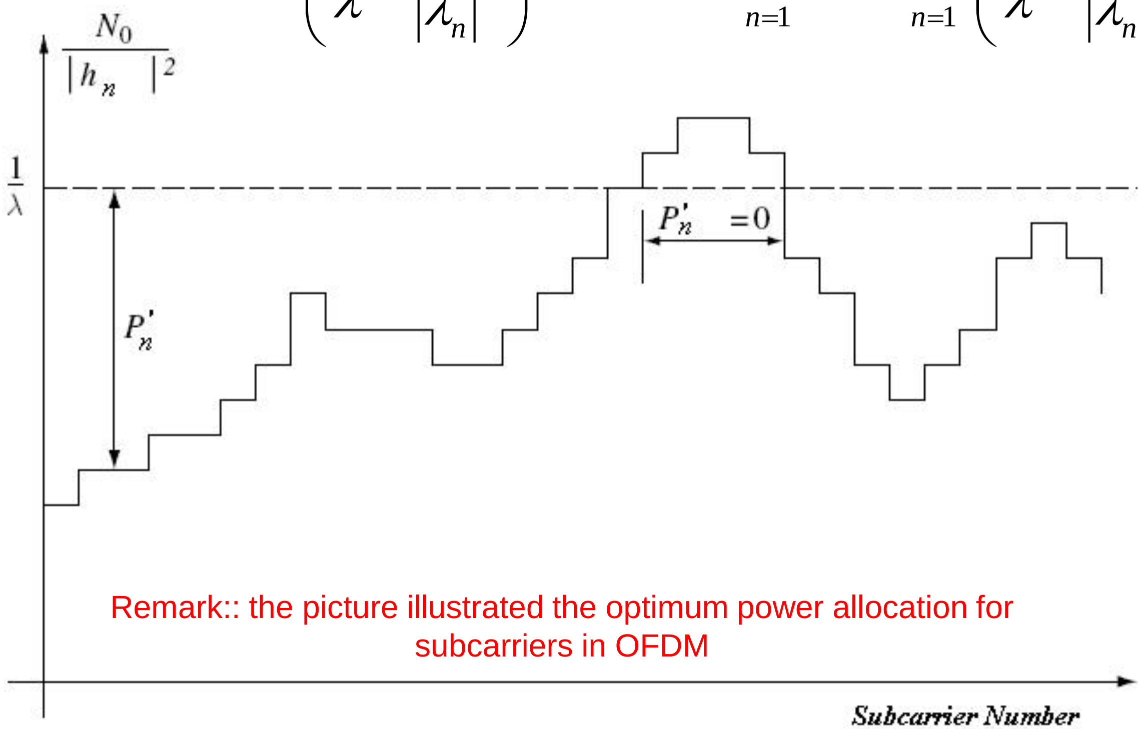
$$P_i' = \left(\mu - \frac{N_0}{\lambda_i^2}\right)^+ \qquad P' = \sum_{i=1}^{n_{min}} \left(\mu - \frac{N_0}{\lambda_i^2}\right)^+$$

Each λ_i corresponds to an eigenmode of the channel (also called eigenchannel). Each non-zero eigenchannel can support a data stream: thus, the MIMO-channel can support the spatial multiplexing of up to n_{min} data streams.

* Note that this again assumes that the channel is known at the transmitter

- **The optimal power allocation is**

$$P_n' = \left(\frac{1}{\lambda} - \frac{N_0}{|\lambda_n|^2} \right)^+ \qquad P' = \sum_{n=1}^{k} P_n' = \sum_{n=1}^{k} \left(\frac{1}{\lambda} - \frac{N_0}{|\lambda_n|^2} \right)^+$$

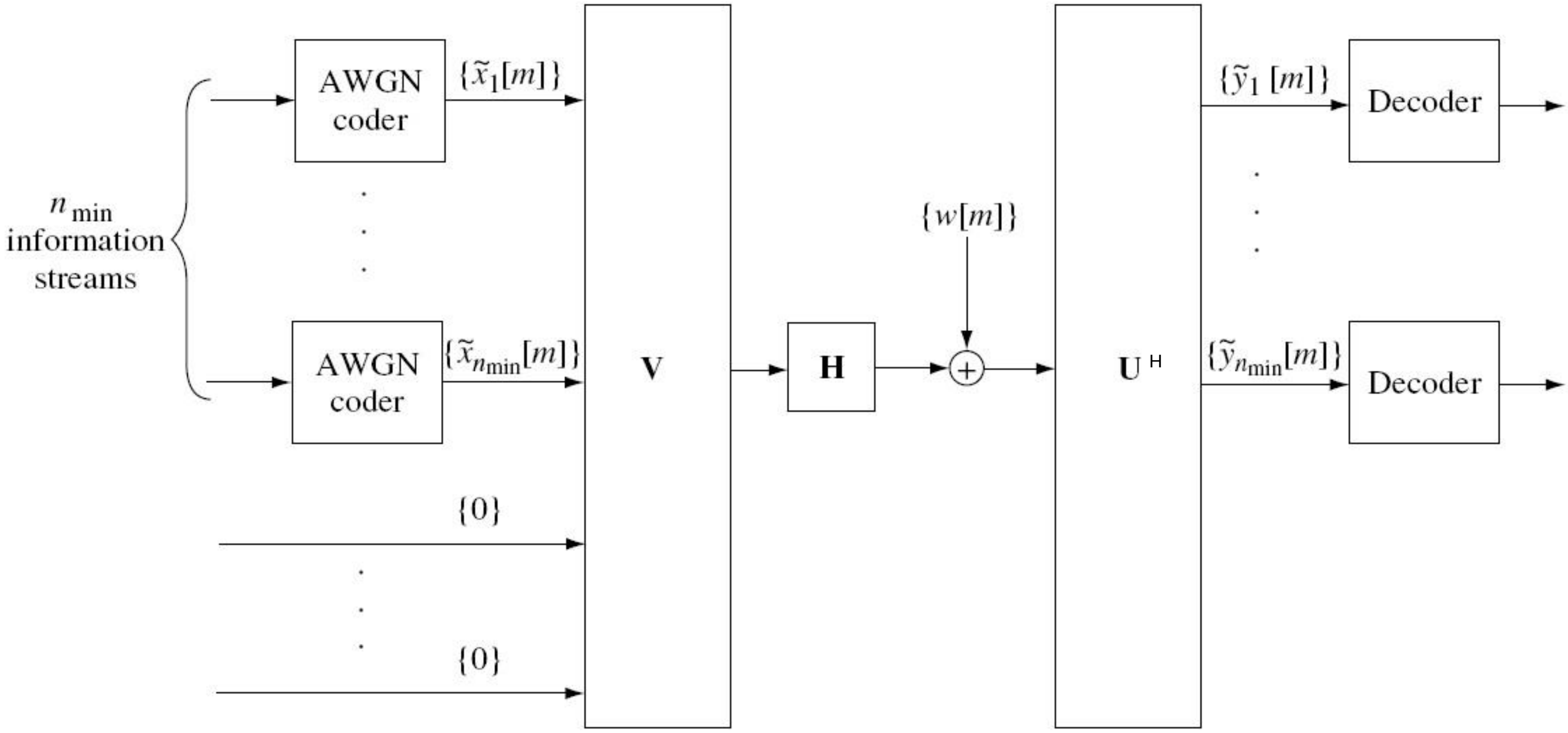


Remark:: the picture illustrated the optimum power allocation for subcarriers in OFDM

Source: David Tse, Pramod Viswanath, „Fundamentals of Wireless Communication“, Cambridge Press 2005

- **Note: more power is allocated to the already “strong” channels**

SVD Architecture for MIMO Communication



Source: David Tse, Pramod Viswanath, „Fundamentals of Wireless Communication", Cambridge Press 2005

- **What are the key parameters that determine performance?**
- **At high SNR**
  - „the water level is deep“, or, formally,

    for all **non-zero** singular values: $\frac{N_0}{\lambda_i^2} << \mu$

  - Thus, the optimum power allocations for the non-zero SV approach

    $$P_i' = \left(\mu - \frac{N_0}{\lambda_i^2}\right)^+ \approx \mu = \frac{P'}{k},$$

    which represents equal power allocation, and is P'_i=0 for the zero SV. (k ≤ n_{min} is the number of non-zero λ_i^2 or the rank of **H)**

    $$C' \approx \sum_{i=1}^{k} \log_2\left(1 + \frac{P'\lambda_i^2}{kN_0}\right)$$

  - In this case we get for the capacity

- Since $$\frac{N_0}{\lambda_i^2} << \mu = \frac{P'}{k} \quad \Leftrightarrow \quad \frac{P'\lambda_i^2}{kN_0} >> 1$$

  we can further approximate for the capacity (with SNR:=P'/N_0)

$$C' \approx \sum_{i=1}^{k} \log_2\left(1 + \frac{P'\lambda_i^2}{kN_0}\right) \approx \sum_{i=1}^{k} \log_2\left(\mathrm{SNR} \cdot \frac{\lambda_i^2}{k}\right)$$

$$\approx k \log_2 \mathrm{SNR} + \sum_{i=1}^{k} \log_2\left(\frac{\lambda_i^2}{k}\right) \mathrm{bits/s/Hz}$$

- The parameter k is the number of spatial degrees of freedom per second per Hertz. It represents the dimension of the transmitted signal as modified by the MIMO channel, i.e., the dimension of the image under the mapping by **H**. This is equal to the rank of **H** (or, equivalently, the number of non-zero singular values). With full rank $k=n_{min}$ and the MIMO channel provides n_{min} spatial degrees of freedom.

- A more refined picture is obtained by looking at the non-zero singular values themselves. By Jensen's inequality

$$C' \approx \sum_{i=1}^{k} \log_2\left(1+\frac{P'\lambda_i^2}{kN_0}\right)$$

$$= k\sum_{i=1}^{k}\frac{1}{k}\log_2\left(1+\frac{P'\lambda_i^2}{kN_0}\right) \le k\log_2\left(1+\frac{P'}{kN_0}\left(\frac{1}{k}\sum_{i=1}^{k}\lambda_i^2\right)\right)$$

- The sum of the eigenvalues of a matrix equals its trace (see, e.g. "Matrix Algebra"), thus, we have:

$$\sum_{i=1}^{k}\lambda_i^2 = \mathrm{Tr}\left[\mathbf{HH}^H\right] = \mathrm{Tr}\left[\mathbf{H}^H\mathbf{H}\right] = \sum_{i,j}\left|h_{i,j}\right|^2$$

This term can be interpreted as the total power gain of the matrix channel, if one spreads the energy equally between all transmit antennas.

For a real convex function φ, numbers x_i in its domain, and positive weights a_i, Jensen's inequality can be stated as:

$$\varphi\left(\frac{\sum a_i x_i}{\sum a_i}\right) \le \frac{\sum a_i \varphi(x_i)}{\sum a_i};$$

and the inequality is clearly reversed if φ is concave.

Source: Wikipedia

- **The inequality upper-bounds the capacity**
- **Consider channel realizations with the same total power gain and, thus, with the same capacity upper bound**
  - Since equality only holds for identical terms in the sum, the maximum capacity (under the above constraint!) is achieved if all singular values are identical
  - More generally, the less spread out the singular values, the larger the capacity in the high SNR regime
  - The ratio $\max_i(\lambda_i)/ \min_i(\lambda_i)$ is defined to be the condition number of the matrix **H**. The matrix is said to be well-conditioned if the condition number is close to 1

Well-conditioned channel matrices facilitate communication
in the high SNR regime

Rank and Condition Number (**Low SNR**)

- At ***low SNR*** only the strongest eigenmodes will be filled up by the waterfilling, thus, an approximately optimal strategy is to allocate all power to the strongest eigenmode. The resulting capacity is*

$$C' \approx \log_2\left(1+\frac{P'\max_i\left(\lambda_i^2\right)}{N_0}\right) \approx \frac{P'}{N_0}\max_i\left(\lambda_i^2\right)\log_2 e \quad \text{bits/s/Hz}$$

- The MIMO channel provides a power gain of $\max_i(\lambda_i^2)$. The rank and condition number of H are less relevant for low SNR. What matters is the maximum singular value λ_i.

* where we have made use of: log2(1+x)=ln(1+x)·log2(e)
and: ln(1+x)≈x for x<<1

Conclusions

- **The derivations so far assumed that the channel is known and constant**
- **This is a hypothetical scenario, but it indicates what the limits are and which transmitter architectures allow achieving the limits for some real scenarios:**
  - when the channel is fading slowly and
  - we have a burst transmission such that the channel does not change significantly during a burst and
  - we can estimate the channel before transmission starts.
- **Burst transmission in systems with high data rates and low mobility often meets the first two requirements**
- **In TDD (Time Division Duplex) systems the same carrier frequency is used in uplink and downlink. Because of the reciprocity of the link, the channel can be estimated for the downlink transmission from the received uplink signal.**
- **If the channel is constant over some bursts, we may estimate the channel in the downlink receiver and feed the channel state back through the uplink channel to use it in subsequent transmissions.**

Conclusions

- (Nearly)Error free transmission requires infinitely long code words. In practice, transmission with low error probability close to the capacity limit requires long codes. The derived limits are an average over all channel realizations and, thus, transmitting close to the capacity limit requires codes that span multiple time constants (coherence time) of the channel.
- There are practical limits for code word lengths and long code words cause long delays in decoding (which may not be acceptable).
- If the channel coherence time is much shorter than the code word length (in “Fundamentals of Wireless Communications” this is called fast fading) we may use the previously derived capacity limits as reference for a scenario where the transmitter at most has statistical information about the channel.
- If the channel coherence time is larger than the code word length (“slow fading”) we will get a time varying behaviour of the transmission. In this case it is advisable to feed back information to the transmitter to adjust the transmission scheme. This may range from full CSI (“initially studied scenario”) to simple quality information.

- Final remark: in the derivations we have assumed that that „CSI is known at the receiver“. In reality, the receiver must estimate the channel, i.e. it must extract information about the channel from the received signal. This reduces the mutual information between received signal and transmitted signal and, thus, reduces the capacity (recall the slides on achievable rates using pilöts)