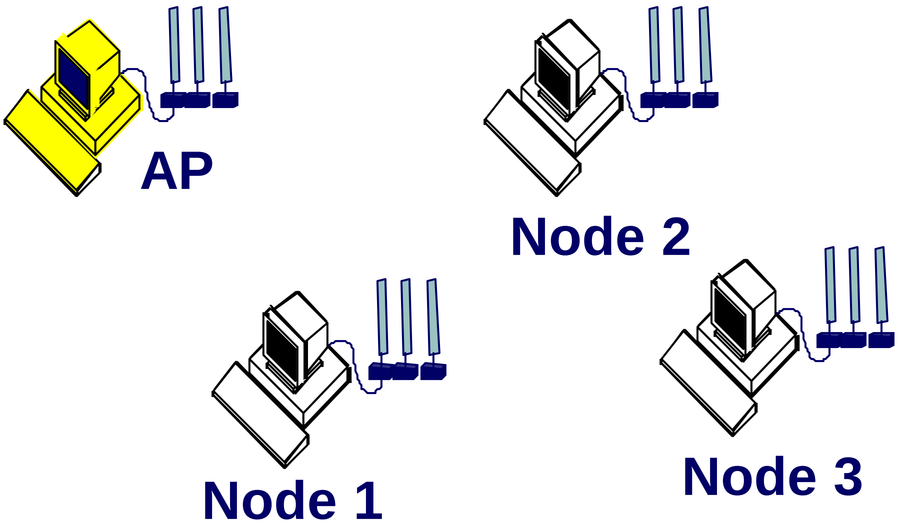
CAPITULO IX:

MIMO MULTIUSER COMMUNICATIONS



Miguel Angel Lagunas, Ana I. Perez-Neira

CAPITULO IX: MIMO MULTIUSER COMMUNICATIONS.

IX.1. INTRODUCTION.

Multiuser MIMO communications from the information theory point of view

Along the present section we have studied several specific multiuser multiantenna designs, but a natural question is: what are the "optimal" multiple access schemes? Information theory can be generalized from the point-to-point scenario, considered in chapter III, to the multiuser ones, providing limits to multiuser communications and suggesting optimal multiple access strategies.

In chapter III we have seen that in a Gaussian vector channel $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{z}$. When **cooperation** is possible both among the transmit terminals and among the receive terminals, the capacity of the vector channel under a power constraint is the solution to the following optimization problem:

$$\max_{\mathbf{R}_\mathbf{x}} \frac{1}{2}\log\frac{\left|\mathbf{H}\mathbf{R}_\mathbf{x}\mathbf{H}^\mathbf{T} + \mathbf{R}_z\right|}{\left|\mathbf{R}_\mathbf{z}\right|} = \max_{\mathbf{R}_x} \frac{1}{2}\log\left|\mathbf{H}^\mathbf{T}\mathbf{R}_z^{-1}\mathbf{H}\mathbf{R}_\mathbf{x} + \mathbf{I}\right|$$

$$s.t. \quad tr(\mathbf{R}_x) \le P$$

$$\mathbf{R}_x \ge 0$$

This leads to the well-known water-filling solution based on the singular-value decomposition of the equivalent channel correlation matrix $\mathbf{H}^T\mathbf{R}_z^{-1}\mathbf{H}$, as it was shown in chapter III. As the eigenvalues majorize the diagonal of a matrix, in order to maximize the determinant, the transmitter correlation matrix diagonalizes the equivalent channel correlation. Assume that $\mathbf{R}z$=**I**, then the optimum $\mathbf{R}$x must have its eigenvectors equal to the right singular vectors of **H** and its eigenvalues obeying the water-filling power allocation on the singular values of **H**. Further, the receive matrix can be chosen to match the left singular vectors of **H**, so that the vector Gaussian channel is diagonalized into a series of independent scalar channels onto which single-user codes can be used to collectively achieve the vector channel capacity. But this solution is only possible in a cooperative scheme.

When coordination is possible only among the receive terminals, but not among the transmit terminals, the vector channel becomes a Gaussian multiple-access channel or **MAC channel**. Although the sum capacity of a multiple-access channel is still a maximum mutual information , the transmit terminals of the multiple-access channel must be uncorrelated. Thus, the water-filling covariance, which is optimum for a coordinated vector channel, can no longer necessarily be synthesized. The optimum covariance matrix for the multiple-access channel must be found by solving an optimization problem that restricts the off-diagonal entries of the covariance matrix to zero

$$\max \frac{1}{2}\log\frac{\left|\mathbf{H}\mathbf{R}_x\mathbf{H}^T+\mathbf{R}_z\right|}{\left|\mathbf{R}_z\right|}$$
$$s.t.\quad tr(\mathbf{R}_x)\le P$$
$$\mathbf{R}_x(i,j)=0\quad \forall(i,\ j)\ \ uncoordinated$$
$$\mathbf{R}_x\ge 0$$

Thus, in terms of capacity, the value of cooperation at the transmitter side lies in the ability for the transmitters to send correlated signals. In addition, the lack of transmitter coordination makes the diagonalization of the vector channel impossible. Instead, the vector channel can only be triangularized . Such a triangularization decomposes a vector channel into a series of single-user subchannels each interfering with only subsequent subchannels. This enables a coding method based on the superposition of single-user codes and a decoding method based on successive decision feedback to be implemented. The optimal form of triangularization is a GDFE (General Decision Feedback Equalizer). If decisions on previous subchannels are assumed correct, GDFE achieves the sum capacity of a Gaussian vector
multiple-access channel. From an algebraic point of view, when only the transmitter or the receiver can face the superuser channel (i.e. BC or MAC channel), the channels is better viewed under a two matrix decomposition, as for instance the **QR** or **RQ**. Then either the transmitter or the receiver can perform $\mathbf{Q}^H$, thus, just leaving a triangular interference.

When coordination is possible only among the transmit terminals, but not among the receive terminals, the vector channel becomes a Gaussian vector broadcast channel or **BC channel**. We will see in this chapter that the sum capacity of a Gaussian vector broadcast channel is the saddle-point of a max-min problem

$$\max_{R_x}\min_{R_z} \frac{1}{2}\log\frac{\left|\mathbf{H}\mathbf{R}_x\mathbf{H}^T+\mathbf{R}_z\right|}{\left|\mathbf{R}_z\right|}$$
$$s.t.\quad tr(\mathbf{R}_x)\le P$$
$$\mathbf{R}_{z_i}\ known$$
$$\mathbf{R}_x,\mathbf{R}_z\ge 0$$

Although the actual noise distribution may not have the same joint distribution as the least favourable noise, because the marginal distributions $\mathbf{R}_{z_i}$ are the same (usually $\mathbf{R}_{z_i}=s_i^2\mathbf{I}$), in a broadcast channel a transmitter designed for the least favourable noise performs as well as with the actual noise. The key point is that because of the lack of coordination, the receivers can no longer distinguish between different noise correlations and the capacity is as if "nature" has chosen a least favourable noise correlation. In other words, the capacity in BC cannot be better than in any cooperative situation. Thus, from a capacity point of view, the value of cooperation at the receiver lies in the ability for the receivers to recognize and to take advantage of the true correlation among the noise and received signals.

Further, in the next section we will see that the structure of the sum-capacity achieving coding strategy for the Gaussian vector broadcast channel is a decision-feedback equalizer. The optimal coding strategy again decomposes the vector channel into independent scalar subchannels each interfering into subsequent subchannels, with the interference pre-subtracted using "writing on dirty paper" coding. When full coordination is not possible, GDFE has emerged as a unifying structure that is capable of achieving the sum capacities of both the multiple-access channel and the broadcast channel sum capacity.

Next figure plots the broadcast and mac channel structures and summarizes their relationship both, between them and among the point to point MIMO channel.

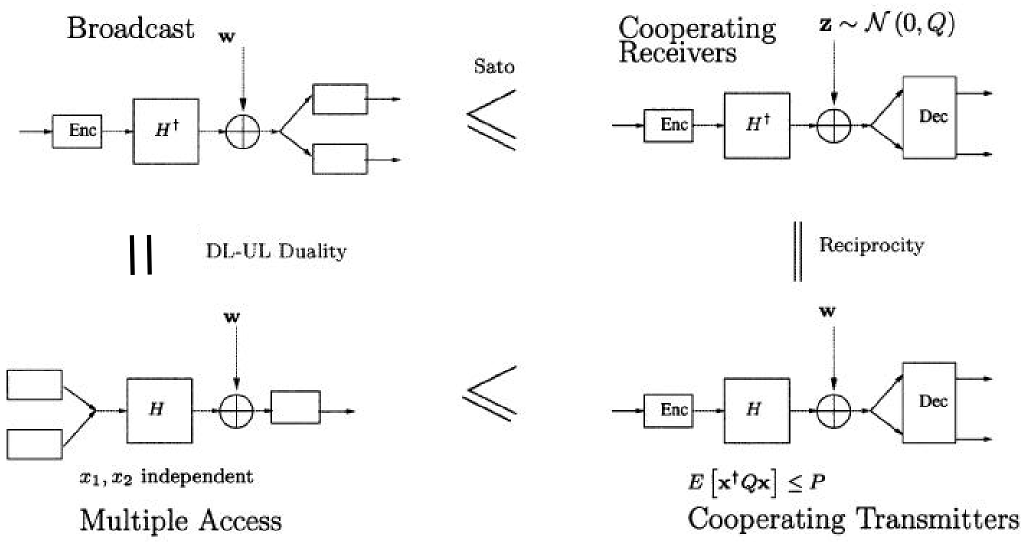


Figure: The four channels, multiple access, broadcast, and their corresponding point-to-point channels, depicted along with the relationship between their capacities.

The study of the multiuser channels is a topic which is not closed in the literature. This chapter differs from many of the existing works because it aims at presenting the basics of the topic under a perspective as much related as possible with signal processing or filtering. Thus, making the capacity results amenable to practical coding schemes, such as those presented in past Chapter VI. First the MAC channel is addressed, second the BC channel and finally considerations on practical multiuser schemes or schedulers are presented.

Along the chapter there are two main concepts that are widely used:

A.- Assume that y =H x + w then R=I(X,Y)

$$R = I(X,Y) = H(Y) - H(Y/X) = \log\frac{|R_y|}{|R_{y/x}|} = \log\frac{|R_y|}{|R_z|}$$

Then y is filtered to obtain an estimation of x, x' such that
x'=Ay
If x =x' + e with E{x'e}=0 then R=I(X,X')

$$R = I(X,Y) = H(X) - H(X/Y) = \log\frac{|R_x|}{|R_{x/y}|} = \log\frac{|R_x|}{|R_e|}$$

And filter A is **capacity lossless**. For instance the MMSE is capacity lossless.

Observe also that there are two possible expressions for I(X,Y). When computed as $I(X,Y)=H(Y)-H(Y\ /X\ )=\log\left|\mathbf{I}+\mathbf{R}_z^{1/2}\mathbf{H}\mathbf{R}_x\mathbf{H}^H\mathbf{R}_z^{1/2}\right|$

When computed as $I(X,Y)=H(X)-H(X/Y)=\log\left|\mathbf{I}+\mathbf{R}_x\mathbf{H}^H\mathbf{H}\right|$ if the MMSE is considered, where $\mathbf{R}_e=\left(\mathbf{H}^H\ \mathbf{H}+\mathbf{R}_x^{-1}\right)^{-1}$

B.- If $x=\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ and $|R_e|=|R_{e1}||R_{e2}|$ (i.e. the Hadamard's inequality fulfils with equality) then there is maximum rate transfer from x1 to x1' and from x2 to x2'

$$R=I(X,Y)=\log\frac{|R_x|}{|R_e|}\geq\log\frac{|R_{x_1}|}{|R_{e_1}|}+\log\frac{|R_{x_2}|}{|R_{e_2}|}$$

We note that equality is fulfilled when Re is diagonal, that means that E{e1.e2}=0, which is not equivalent to saying that E{x1'.x2'}=0.

IX.2. MIMO Multiuser MAC CHANNEL.

IX.2.1. MAC Capacity region

In the point to point case, the capacity of a channel provides the performance limit: reliable communication can be attained at any rate R<C; reliable communication is impossible at rates R>C. In the multiuser case, we should extend this concept to a capacity region C: this is the set of all pairs (R1,R2) (i.e. in the 2 user case) such that simultaneously user1 and user2 can reliably communicate at rate R1 and R2, respectively. Because signalling dimensions can be allocated to different users in an infinite number of different ways, multiuser channel capacity is defined by a rate region rather than a single number. This region describes all user rates that can be simultaneously supported by the channel with arbitrarily small error probability. From this capacity region, one can derive other scalar performance measures of interest.

For example, the symmetric capacity

$$C_{sym} = \max_{(R1,R2)\in C} R$$

Is the maximum common rate at which both the users can simultaneously reliably communicate.

The sum capacity

$$C_{sum} = \max_{(R1,R2)\in C} R_1 + R_2$$

Is the maximum total throughput that can be achieved.

With a single receive antenna at the Base station or access point, the capacity region of the two-user MAC or uplink channel is defined by the following equations

$$R_1 < \log\left(1+\frac{P_1}{N_o}\right) = I(x_1; y|x_2)$$

$$R_2 < \log\left(1+\frac{P_2}{N_o}\right) = I(x_2; y|x_1)$$

$$R_1 + R_2 < \log\left(1+\frac{P_1+P_2}{N_o}\right) = I(x_2, x_1; y) = I(x_1; y) + I(x_2; y|x_1) \neq I(x_1; y|x_2) + I(x_2; y|x_1)$$

Where P1 and P2 are the average power constraints on users 1 and 2 respectively. The individual rate constraints correspond to the maximum rate that each user can get if it has the entire channel to itself; the sum rate constraint is the total rate of a point-to-point channel with the two users acting as two transmit antennas of a single user, but sending independent signals. The three constraints define the pentagon of the figure, where, for instance, point C is obtained after matched filtering for each of the users.

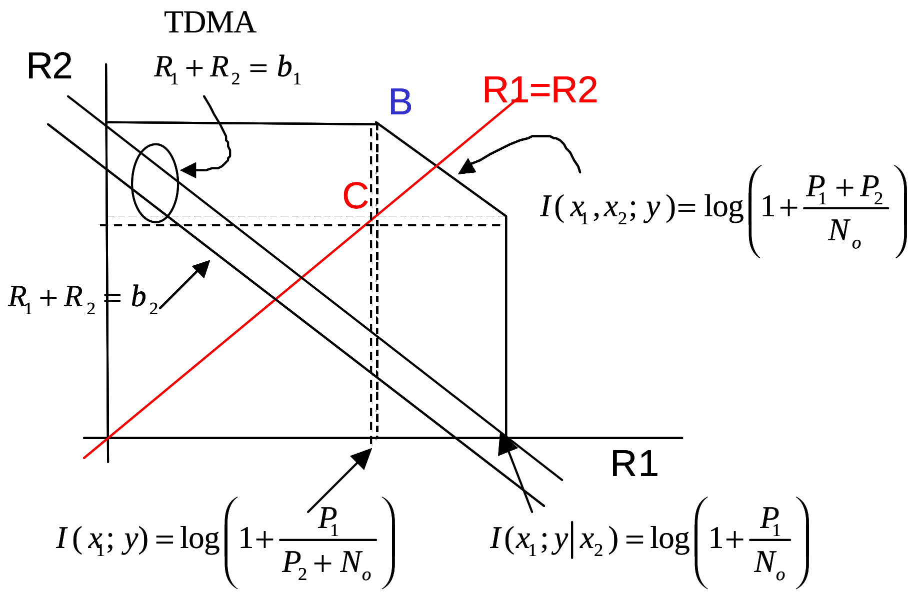


Figure: MAC capacity region for the single tx. antenna. For 2 users is the so-called Cover-Wyner pentagon

The capacity region of the multiple access channel is the convex hull[1] of the union of these pentagons over all possible independent input distributions subject to the appropriate individual average cost constraints, i.e.,

$$C = convex\ hull\ \ of\ \left(\bigcup_{P_{x1}, P_{x2}} C(P_{x1}, P_{x2})\right)$$

The convex hull operation means that we not only include points such C, but also all their convex combinations: $a R_1 + (1-a) R_2 \quad 0 \le a \le 1$ (i.e. the diagonal lines in the figure). This is natural since the convex combinations can be achieved by time-sharing. The diagonal lines in the figure correspond to time-sharing access. For the MAC channel with single transmit antennas, the capacity region is a pentagon, because there is a unique set of input distributions that simultaneously maximizes the different constraints for R1, R2 and R1+R2. With a single transmit antenna at each user, the transmitter architecture simplifies considerably: there is only one data stream and the entire power is allocated to it. By varying the power allocations (in the case of full CSIT), the users can communicate at rate pairs in the union of the pentagons, which is itself a pentagon. For example, 2 users SDMA with multiple antenna just at the base station is a natural extension of the single antenna case

[1] The convex hull of a set C is the set of all convex combinations of points in C:

$$convC = \{q_1 x_1 + \ldots + q_k x_k \mid x_i \in C, q_i \ge 0, i = 1,\ldots,k, q_1 + \ldots + q_k = 1\}$$

$$R_1 < \log\left(1+\frac{\|\mathbf{h}_1\|^2 P_1}{N_o}\right) = I(x_1; y | x_2)$$

$$R_2 < \log\left(1+\frac{\|\mathbf{h}_2\|^2 P_2}{N_o}\right) = I(x_2; y | x_1)$$

$$R_1 + R_2 < \log\det\left(\mathbf{I}_{nr} + \frac{1}{N_o}\mathbf{H}\mathbf{R}_\mathbf{x}\mathbf{H}^\mathbf{H}\right) \qquad \mathbf{R}_x = \begin{bmatrix} P_1 & 0 \\ 0 & P_2 \end{bmatrix}$$

Figure: MAC capacity region for single tx antenna. For 2 users is the so-called Cover-Wyner pentagon

The right hand side of the third inequality is the sum capacity and it is the total rate achieved in a point-to-point channel with the two users acting as two transmit antennas of one user with independent inputs at the antennas. Note that

$$\log\left|\mathbf{I} + \mathbf{H}\mathbf{R}_x\mathbf{H}^H\right| = \log\left|\mathbf{I} + \sum_i p_i \mathbf{h}_i \mathbf{h}_i^H\right|$$

and to get some insight this expression can be further work out and results that there is a gain in rate when using multiple antennas at reception if the channels of the users are not orthogonal

$$\log\left|\mathbf{I} + \mathbf{H}\mathbf{R}_x\mathbf{H}^H\right| = \log\left|\mathbf{I} + \sum_i p_i \mathbf{h}_i \mathbf{h}_i^H\right| = \log\left(1 + \frac{P_1|h_1|^2 + P_2|h_2|^2}{N_o} + \frac{P_1P_2}{N_o}\left|\mathbf{H}\mathbf{H}^H\right|\right)$$

$$\left|\mathbf{H}\mathbf{H}^H\right| = |\mathbf{h}_1|^2 |\mathbf{h}_2|^2 - \left|\mathbf{h}_1^H \mathbf{h}_2\right|^2$$

To define the capacity region in the general case of Ntot! users, and MIMO multiuser access, we can extend the previous region to

$$R_k \le \log\det\left(\mathbf{I}_{nt} + \frac{1}{N_o}\mathbf{H}\mathbf{R}_k\mathbf{H}^H\right) \quad k = 1 ... Ntot$$

$$\sum_{k=1}^{Ntot} R_k \le \log\det\left(\mathbf{I}_{nt} + \frac{1}{N_o}\sum_{k=1}^{Ntot}\mathbf{H}\mathbf{R}_k\mathbf{H}^H\right)$$

with constraints on each users power $tr\left[\mathbf{R}_k\right] = P_k$. The last inequality is upperbounded by the superuser capacity, which is a generalization of that obtained for the single transmitting antenna case.

This inequalities that define the MAC capacity region can be summarized in a single inequality and the MAC capacity region can be defined in the following more compact way

$$C_{MAC} = \bigcup_{\{Tr(\mathbf{R}_i)\le P_i\ \forall i\}} \left\{ (R_1 ... R_K) : \sum_{i\in S} R_i \le \frac{1}{2}\log\left| I + \sum_{i\in S} \mathbf{H}_i \mathbf{R}_i \mathbf{H}_i^H \right| \quad \forall S \subseteq \{1...Ntot\} \right\}$$

In the convex hull process, however, if there are multiple transmit antenna at the users, no single pentagon may dominate over the other pentagons (as it is shown in next figure).

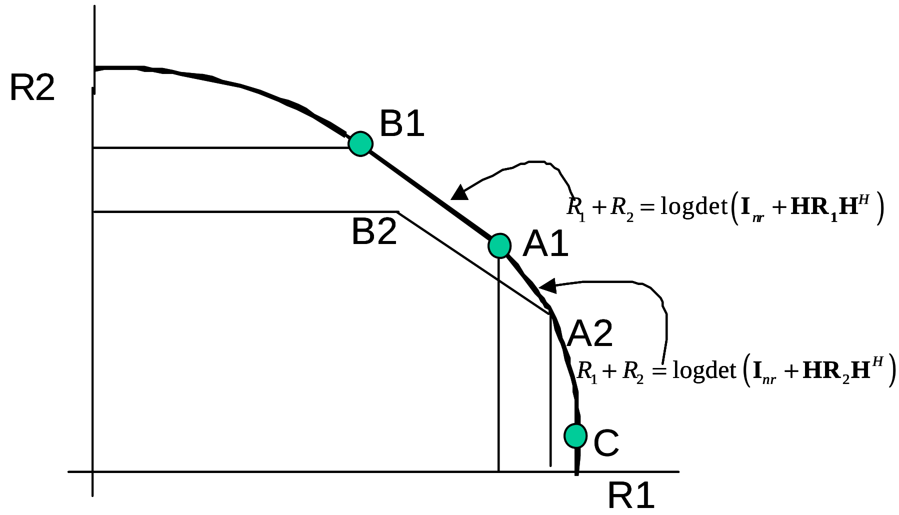


Figure: The achievable rate regions (pentagons) corresponding to two different input distributions may not fully overlap with respect to one another. Multiple transmitting and receiving antennas.

In that case, there is no single choice of covariance matrix that simultaneously maximize the constraints: the capacity region is the convex hull of the union of the pentagons created by all the possible covariance matrices (subject to the power constraints on the users). The global capacity region is generated by the union of such polyhedrons, each one corresponding to a specific power allocation satisfying the power constraints. The resulting boundary of the global capacity region is curved, except at the sum rate point, where the boundary is a straight line, and it is generated by the union of well-selected vertices. At point C, user1 is decoded last and achieves his single-user capacity by choosing R1 as a waterfill of the channel H1 (independent of H2 or R2). User 2 is decoded first, in the presence of interference from user 1, so R2 is chosen as a waterfill of the channel H2 and the interference from user1. The sum-rate corner points A1 and B1 are the two corner points of the pentagon corresponding to the sum-rate optimal covariance matrices R_1^{sum}, R_2^{sum} . We will see in IX.2.2. that at point A1 user 1 is decoded last, whereas at point B1 user 2 is decoded last. Thus, points A1 and B1 are achieved using the same covariance matrices but different decoding orders.

With multiple transmit antennas, we have a choice of power splits among the data streams and also the choice of the rotation **U** before sending the data streams out of the transmit antennas. We recall that in the time-invariant point-to-point MIMO channel, the rotation matrix **U** was chosen to correspond to the right rotation in the singular value decomposition of the channel and the powers allocated to the data streams correspond to the waterfilling allocations over the squared singular values of the channel matrix. In the MAC MIMO, in general, different choices of power splits and rotations lead to different pentagons and the capacity region in general is not a pentagon. This is because, unlike the single transmit antenna case, there are no covariance matrices **R**k, that simultaneously maximize the right hand of the all the inequalities. Depending on how one wants to trade off the performance of the two users, one would use different input strategies. In any case, note that the sum capacity is concave on **R**i, thus, in general there is no closed-form solution to the optimization problem considering sum rate, but efficient algorithms that arrive at numerical solutions exist. The obtained sum-rate maximizing covariance matrix of any user in the system should be the single-user water-filling covariance matrix of its won channel with noise equal to the actual noise plus the interference from the other K-1 transmitters [45].

The question to answer is what is the optimal receiver architecture that achieves sum capacity. How can corner points A1 and B1 be reached? Next section presents the Decision Feedback structure as the one that answers these questions.

IX.2.2. Decision Feedback equalizer and MAC sum capacity

The DFE structure achieves the entire capacity region of the multiple-access channel.

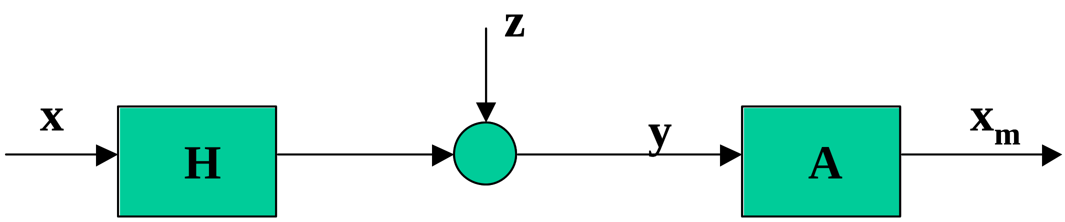
The DFE structure

Let us have $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w}$ $\quad \mathbf{x} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix}$ $\quad \mathbf{H} = \begin{bmatrix} \mathbf{H}_1 & \mathbf{H}_2 \end{bmatrix}$ $\quad \mathbf{R}_x = \begin{bmatrix} \mathbf{R}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_2 \end{bmatrix}$

where **x**i (i=1,2) are vectors and they are statistically independent, therefore, Rx is a block diagonal matrix (in contrast to the BC channel as we will see later on).

Next we show that the DFE decomposes the vector channel into two subchannels that can be independently encoded and decoded because of the error diagonalization that the DFE performs. The achievable rates of the two subchannels are R1=I(X1',X1) and R2=I(X2',X2), being xi' the output that estimates each subchannel xi at the output of the DFE, and the sum rate is R1+R2=I(X;Y). Thus, the DFE is capacity lossless). In order to prove it 2 key ideas are involved: 1) the MMSE filter is capacity lossless, in terms of sum capacity, 2) block Cholesky factorization of the minimum MMSE noise matrix.

Next consider the following figure, where **A** is the MMSE filter



1.- The MMSE filter is capacity lossless: I(X;Xm)=I(X;Y)

Let **x=xm+e**, where E{x_m,e}=0 thus $I(X;Xm)=I(Xm;X)=\frac{1}{2}\log\frac{|\mathbf{R}_x|}{|\mathbf{R}_e|}$

As

$$I(X;Y)=I(Y;X)=H(X)-H(X/Y)=\frac{1}{2}\log\frac{|\mathbf{R}_x|}{|\mathbf{R}_{x/y}|}=\frac{1}{2}\log\frac{|\mathbf{R}_x|}{|\mathbf{R}_e|}$$

However if we define $\mathbf{e}=\begin{bmatrix}\mathbf{e}_1\\ \mathbf{e}_2\end{bmatrix}=\begin{bmatrix}\mathbf{x}_1-\mathbf{x}_{1m}\\ \mathbf{x}_2-\mathbf{x}_{2m}\end{bmatrix}$, e1 and e2 are not necessarily uncorrelated.

$\mathbf{x_m}=\mathbf{AHx}+\mathbf{Az}$ with **e=Az.** Taking into account that the MMSE filter $\mathbf{A}=\mathbf{R_{xy}R_y^{-1}}=\mathbf{R_xH^T}\left(\mathbf{H^TS_xH}+\mathbf{I}\right)^{-1}=\left(\mathbf{H^TH}+\mathbf{S_x^{-1}}\right)^{-1}\mathbf{H^T}$ the correlation matrix of the error is

$$\mathbf{R}_e=E\left\{\mathbf{ee}^T\right\}=E\left\{(\mathbf{x}-\mathbf{x_m})(\mathbf{x}-\mathbf{x_m})^T\right\}=\mathbf{R_x}-\mathbf{R_xH^T(HR_xH^T+I)^{-1}HR_x}=\left(\mathbf{H^TH}+\mathbf{S_x^{-1}}\right)^{-1}$$

where $\left(\mathbf{H^TH}+\mathbf{S_x^{-1}}\right)^{-1}=\mathbf{R_x}-\mathbf{R_xH^T(HR_xH^T+I)^{-1}HR_x}$ is obtained by way of the matrix inversion lema.

So, by Hadamard's inequality, $|\mathbf{R_e}|\le|\mathbf{R_{e_1}}||\mathbf{R_{e_2}}|$. This implies

$$\frac{1}{2}\log\frac{|\mathbf{R}_x|}{|\mathbf{R}_e|}\ge\frac{1}{2}\log\frac{|\mathbf{R}_{x1}|}{|\mathbf{R}_{e1}|}+\frac{1}{2}\log\frac{|\mathbf{R}_{x2}|}{|\mathbf{R}_{e2}|}$$

That the independent decoding of x1 based on x1m and decoding of x2 based on x2m are capacity-lossy.

The goal of the DFE is to use a decision-feedback structure to enable the independent decoding of x1 and x2. This is accomplished by a diagonalization of the MMSE error e, while preserving the "information" in xm.

2) The diagonalization of the MMSE error can be done via a Block Cholesky factorization as follows

$$\mathbf{R_e}=\mathbf{G^{-1}?^{-1}G^{-T}}\qquad \mathbf{G}=\begin{bmatrix}\mathbf{I}&\mathbf{G}_{22}\\ \mathbf{0}&\mathbf{I}\end{bmatrix}\qquad \mathbf{?}=\begin{bmatrix}\mathbf{?}_{11}&\mathbf{0}\\ \mathbf{0}&\mathbf{?}_{22}\end{bmatrix}$$

then

$$\mathbf{x}=\mathbf{A}\,\mathbf{y}+\mathbf{e}=$$
$$=\mathbf{G^{-1}?^{-1}G^{-T}H}^T\mathbf{y}+\mathbf{e}$$

In order to decouple the error

$$\mathbf{Gx} = ?^{-1}\mathbf{G}^{-T}\mathbf{w}+\mathbf{Ge} = ?^{-1}\mathbf{G}^{-T}\mathbf{w}+\mathbf{Ge} \qquad (*)$$

$$\mathbf{e}'=\mathbf{Ge}=\begin{pmatrix}\mathbf{I} & \mathbf{G}_{22}\\ \mathbf{0} & \mathbf{I}\end{pmatrix}\begin{pmatrix}\mathbf{e}_1\\ \mathbf{e}_2\end{pmatrix}\rightarrow \mathbf{R}_{\mathbf{e}'} = ?^{-1}$$

$$\left|\mathbf{R}_{\mathbf{e}'}\right| = \left|\Delta_{11}^{-1}\right|\left|\Delta_{22}^{-1}\right|$$

Thus **e**’ is uncorrelated. From equation (*) we get

$$\mathbf{x} = ?^{-1}\mathbf{G}^{-T}\mathbf{w}+(\mathbf{I}-\mathbf{G})\mathbf{x}+\mathbf{e}'=\mathbf{x}'+\mathbf{e}'$$

Which gives the DFE structure of the new receiver shown in the figure, where the feedback filtering part can be implemented a successive interference cancellation due to the triangular structure of **G**. Note that in case **R**e were factorized following the SVD decomposition, then the successive interference cancellation interpretation is lost.

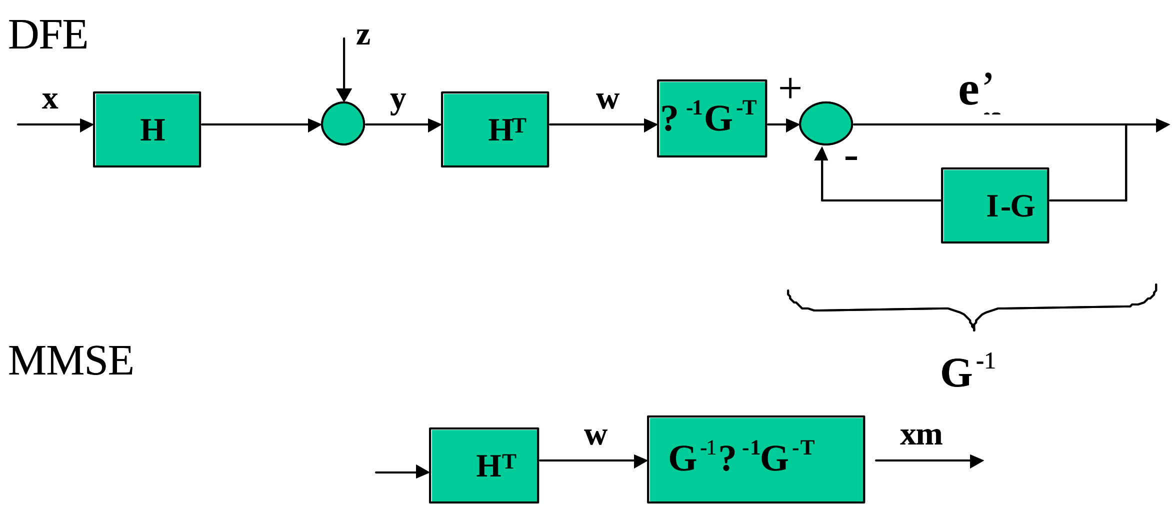
The achievable rates of the two subchannels are

$$R_1 = I(X_1^{'}; X_1) = \frac{1}{2}\log\frac{\left|\mathbf{R}_1\right|}{\left|\mathbf{R}_{\mathbf{e}'1}\right|}$$

$$R_2 = I(X_2^{'}; X_2) = \frac{1}{2}\log\frac{\left|\mathbf{R}_2\right|}{\left|\mathbf{R}_{\mathbf{e}'2}\right|}$$

$$R_1 + R_2 = I(X_1^{'}; X_1) + I(X_2^{'}; X_2) = \frac{1}{2}\log\frac{\left|\mathbf{R}_x\right|}{\left|\mathbf{R}_{\mathbf{e}}\right|} = I(X;Y)$$

Thus, proving that the GDFE is capacity lossless. Note that now **the independent decoding of x1 based on x1’ and decoding of x2 based on x2’ are capacity-lossless.** Therefore, trying to obtain a diagonal error matrix is useful both, in terms of BER, because there is no interference among branches, as seen in past chapters [tesis Daniel Palomar], and in terms of capacity.



In order to obtain the specific rates for each user when the DFE is used, the block Cholesky factorization[2] may be computed explicitly as

$$\left(\mathbf{R}_x^{-1}+\mathbf{H}^T\mathbf{H}\right)^{-1}=\begin{bmatrix}\mathbf{R}_1^{-1}+\mathbf{H}_1^T\mathbf{H}_1 & \mathbf{H}_1^T\mathbf{H}_2\\ \mathbf{H}_2^T\mathbf{H}_1 & \mathbf{R}_2^{-1}+\mathbf{H}_2^T\mathbf{H}_2\end{bmatrix}^{-1}=\mathbf{G}^{-1}?^{-1}\mathbf{G}^{-T}$$

where

$$\mathbf{G}=\begin{bmatrix}\mathbf{I} & \left(\mathbf{R}_x^{-1}+\mathbf{H}_1^T\mathbf{H}_1\right)^{-1}\mathbf{H}_1^T\mathbf{H}_2\\ \mathbf{0} & \mathbf{I}\end{bmatrix}^{-1}$$

$$?^{-1}=\begin{bmatrix}\left(\mathbf{R}_1^{-1}+\mathbf{H}_1^T\mathbf{H}_1\right)^{-1} & 0\\ 0 & \left(\mathbf{R}_2^{-1}+\mathbf{H}_2^T\mathbf{H}_2-\mathbf{H}_2^T\mathbf{H}_1\left(\mathbf{R}_1^{-1}+\mathbf{H}_1^T\mathbf{H}_1\right)^{-1}\mathbf{H}_2^T\mathbf{H}_2\right)^{-1}\end{bmatrix}$$

Therefore

$$R_1=I(X_1';X_1)=\frac{1}{2}\log\frac{\left|\mathbf{R}_1\right|}{\left|\left(\mathbf{R}_1^{-1}+\mathbf{H}_1^T\mathbf{H}_1\right)^{-1}\right|}=\frac{1}{2}\log\left|\mathbf{H}_1^T\mathbf{R}_1\mathbf{H}_1+\mathbf{I}\right|=I(X_1;Y/X_2)$$

$$R_2=I(X_2';X_2)=\frac{1}{2}\log\frac{\left|\mathbf{R}_2\right|}{\left|\left(\mathbf{R}_2^{-1}+\mathbf{H}_2^T\left(\mathbf{I}+\mathbf{H}_1\mathbf{R}_1\mathbf{H}_1^T\right)^{-1}\mathbf{H}_2\right)^{-1}\right|}=\frac{1}{2}\log\frac{\left|\mathbf{H}_1^T\mathbf{R}_1\mathbf{H}_1+\mathbf{H}_2^T\mathbf{R}_2\mathbf{H}_2+\mathbf{I}\right|}{\left|\mathbf{H}_1^T\mathbf{R}_1\mathbf{H}_1+\mathbf{I}\right|}=I(X_2;Y)$$

$$R_1+R_2=I(X_1,X_2;Y)=\frac{1}{2}\log\left|\mathbf{H}_1^T\mathbf{R}_1\mathbf{H}_1+\mathbf{H}_2^T\mathbf{R}_2\mathbf{H}_2+\mathbf{I}\right|=I(X_2;Y)+I(X_1;Y/X_2)$$

The third equation says that the total throughput cannot exceed the capacity of a point-to-point AWGN channel with the sum of the received powers of the 2 users. This is a valid constraint since the signals of the two users are independent. Without this third equation, the capacity region would have been a rectangle and each user could simultaneously transmit at the point-to-point AWGN channel with the sum of the received powers of the 2 users. This is a valid constraint since the signals of the two users are independent.

---

[2] Aside from the SVD decomposition, other matrix factorizations are going to be considered along this chapter : Cholesky factorization consists in $\mathbf{A}=\mathbf{B}\mathbf{B}^T$ where A is square and B is a lower triangular matrix; LU factorization consists in $\mathbf{A}=\mathbf{L}\mathbf{D}\mathbf{U}^H$ where A is square, L is lower triangular, D is diagonal and U^H is upper triangular; QR factorization $\mathbf{A}=\mathbf{Q}\mathbf{R}$ where A does not need to be square, Q is orthonormal and R is upper triangular.

Note that, in general, for Ntot users,

$$\sum_{i=1}^{N_{tot}} R_i = \log\left|\mathbf{I}+\sum_{i=1}^{Ntot}\mathbf{H_i^H R_i H_i}\right| = \log\left|\mathbf{I}+\mathbf{H}_1^{\mathbf{H}}\mathbf{R}_1\mathbf{H}_1\right| + \ldots + \log\frac{\left|\mathbf{I}+\sum_{i=j}^{1}\mathbf{H}_i^{\mathbf{H}}\mathbf{R}_i\mathbf{H}_i\right|}{\left|\mathbf{I}+\sum_{i=j-1}^{1}\mathbf{H}_i^{\mathbf{H}}\mathbf{R}_i\mathbf{H}_i\right|} + \ldots + \log\frac{\left|\mathbf{I}+\sum_{i=Ntot}^{1}\mathbf{H}_i^{\mathbf{H}}\mathbf{R}_i\mathbf{H}_i\right|}{\left|\mathbf{I}+\sum_{i=Ntot-1}^{1}\mathbf{H}_i^{\mathbf{H}}\mathbf{R}_i\mathbf{H}_i\right|}$$

Usually, the users are ordered $|\mathbf{h}_1| \le |\mathbf{h}_2| \le \ldots \le |\mathbf{h}_{Ntot}|$, in that case, the rate achieving sum MAC capacity is

$$R_j^{sum} = \log\frac{\left|\mathbf{I}+\sum_{i=j}^{Ntot}\mathbf{H}_i^{\mathbf{H}}\mathbf{R}_i\mathbf{H}_i\right|}{\left|\mathbf{I}+\sum_{i=j+1}^{Ntot}\mathbf{H}_i^{\mathbf{H}}\mathbf{R}_i\mathbf{H}_i\right|}$$

For only multi-antenna at reception, if Ntot=2 users, with the DFE

$$R_1 = I(X_1;\, Y/X_2) = \log\left(1+\frac{P_1\|\mathbf{h}_1\|^2}{N_o}\right)$$

$$R_2 = I(X_2;Y) = \log\left(1+P_2\mathbf{h}_2^H\left(N_o I + P_1\mathbf{h}_1\mathbf{h}_1^H\right)^{-1}\mathbf{h}_2\right)$$

In conclusion, the DFE achieves de sum capacity of the MAC channel.

<u>Other receiver structures</u>

We observe that the user rates Ri that achieve sum capacity require certain degree of interference among users in this way the transmission benefits from all the multiplexing gain that is offered by the multi-antenna channel, as we have commented before. This solution might cause, however, big difference among users depending on their channel. Any receiver structure that would try to null the interference (as for instance Zero Forcer), thus trying to make the reception more fair for all users, would decrease the rate. In any case, for low SNR or SNIR scenario, or if fairness is required, the DFE might not be realistic because, due to error propagation, the sum rates will not be obtained, thus requiring either a more sophisticated Maximum Likelihood detection or some interference cancellation scheme. Suboptimal theoretically speaking but more realistic if the actual throughput is the main concern, as indicated in next figure.

In general, there are Ntot! corner points on the boundary of the capacity region and each corner point is specified by an ordering of the Ntot users and the corresponding rates are achieved by an DFE receiver with that ordering of cancelling users.

IX.2.3. Fading channel

If the communication is over several coherence intervals of the user channels the new capacity region is

$$R_k \le E\left[\operatorname{logdet}\left(\mathbf{I}_{nr} + \frac{1}{N_o}\mathbf{H}\mathbf{R}_k\mathbf{H}^H\right)\right] \qquad k = 1...Ntot$$

$$\sum_{k=1}^{Ntot} R_k \le E\left[\operatorname{logdet}\left(\mathbf{I}_{nr} + \frac{1}{N_o}\sum_{k=1}^{Ntot}\mathbf{H}_k\mathbf{R}_k\mathbf{H}_k^{\mathbf{H}}\right)\right]$$

If only there are multi-antenna at reception, the regions are defined by

$$R_k < \log\left(1 + \frac{|\mathbf{h}_k|^2 P_k}{N_o}\right)$$

$$\sum_{k=1}^{Ntot} R_k < \operatorname{logdet}\left(\mathbf{I}_{nr} + \frac{1}{N_o}\mathbf{H}\mathbf{R}\mathbf{H}^H\right) = \operatorname{logdet}\left(\mathbf{I}_{nr} + \frac{1}{N_o}\sum_k P_k\mathbf{h}\mathbf{h}_k^H\right)$$

$$with \quad \mathbf{R} = diag\{P_i\}$$

With a sufficiently random and well-conditioned channel matrix **H**, the performance gain is significant.

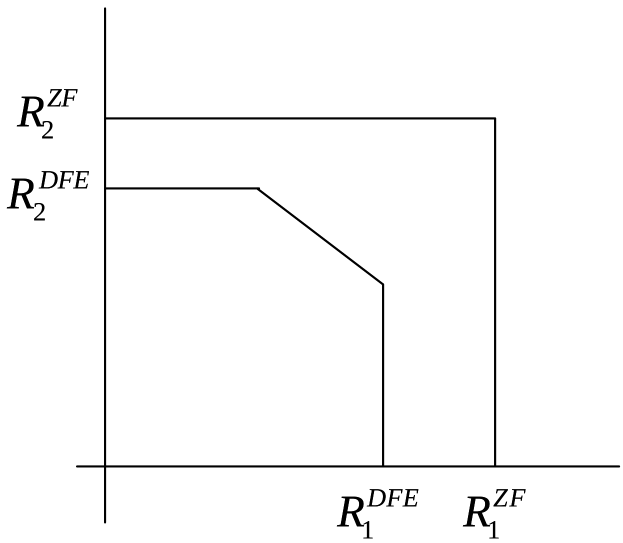


Figure: Throughput regions (that account for BER) for the ZF and DFE

IX.2.3.1. Only CSIR

In the case of i.i.d Rayleigh fading model, the capacity achieving power allocation is equal powers to the data streams (as in the point-to-point MIMO)

IX.2.3.2. Full CSIT and CSIR

In a MIMO MAC channel, the situation of full CSI is an unrealistic one due to the increase in number of parameters to feedback (so that the users can change their transmit strategies as a function of the time-varying channels). A more realistic situation is when only the receiver has multiple antenna.

Only multi-antenna at reception

Now the users can vary their transmit power as a function of the channel realizations; still subject to an average power constraint. If we assume 2 users $E\left[P_k(\mathbf{h}_1,\mathbf{h}_2)\right] \leq P \qquad k=1,2$

In the point-to-point channel, we have seen that the power variations are waterfilling over the channel states. To get some insight into how the power variations are done in the uplink with multiple receive antennas, let us focus on the sum capacity

$$C_{sum} = \max_{P_k} E\left[\operatorname{logdet}\left(\mathbf{I}_{nr} + \frac{1}{N_o}\mathbf{HRH}^H\right)\right] = \max_{P_k} E\left[\operatorname{logdet}\left(\mathbf{I}_{nr} + \frac{1}{N_o}\sum_{k=1}^{Ntot} P_k \mathbf{h}\mathbf{h}_k^H\right)\right]$$

where the power allocations are subject to the average constraints. Firstly, note that in the MAC channel with a single receive antenna at the base-station

$$\log\left(1+\frac{\sum_{k=1}^{Ntot} P_k \left|h_k\right|^2}{N_o}\right) \qquad \sum_k P_k = P$$

where, a global power constraint is considered, the power allocation that maximizes sum capacity allows only the best user to transmit

$$P_k = \begin{cases} \left(\dfrac{1}{l} - \dfrac{N_o}{\max_i \left|h_i\right|^2}\right)^+ & if\ \left|h_k\right|^2 = \max_i \left|h_i\right|^2 \\ 0 & \end{cases}$$

Thus, following a so-called **opportunistic strategy**.

In the MAC channel each user is received as a vector ($\mathbf{h}_k$ for user k) at the base-station and there is no natural ordering of the users to bring this argument forth here. Still the optimal allocation of powers can be found using the Lagrangian techniques, but there is no closed form solution. However, with both nr and Ntot large and comparable, there is a simple policy close to the optimal one: every user transmits and the power allocated is waterfilling over its own channel state

$$P_k(\mathbf{H}) = \left(\frac{1}{l} - \frac{I_o}{\left|\mathbf{h}_k\right|^2}\right)^+ \qquad k=1...Ntot$$

$$I_o = \frac{P_k^*(\mathbf{H})\left\|\mathbf{h}_k\right\|^2}{SINR_k}$$

where, Io itself is a function of the power allocations of the other users (which themselves depend on the power allocated to user k). However if Ntot and nr are large enough, Io converges to a constant in probability. As usual, the water level λ is chosen such that the average power constraint is met.

If we compare the last proposed waterfilling allocation with the opportunistic one with one receive antenna. The important difference is that when there is only one user transmitting, waterfilling is done over the channel quality with respect to the background noise. However, here all the users are simultaneously transmitting, using a similar waterfilling power allocation policy. The waterfilling is done over the channel quality (the receive beamforming gain) with respect to the background interference plus noise: this is denoted by the term Io. Note that the multiuser diversity gain is lost, which is called **hardening effect**. The traditional receive beamforming power gain is balanced by the loss of the benefit of the multiuser diversity gain (which is also a power gain) due to the "hardening" of the effective fading distribution: $\left\|\mathbf{h}_k\right\|^2 \approx n_r$. In particular, at high SNR the waterfilling policy simplifies to the constant power allocation at all times (if nr>Ntot).

A different result is obtained if nr is fixed and Ntot goes to infinite. In this case we can still talk of multiuser diversity gain, which is achieved by carrying out an opportunistic policy based on the users SNIR, and it is the basis of many practical schedules as we are going to see in the section dedicated to the BC channel.

IX.3. MIMO BROADCAST CHANNEL.

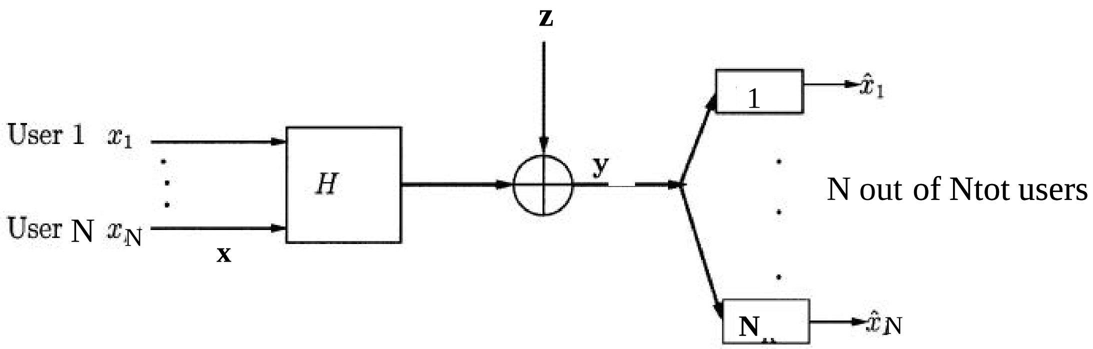
In our discussion of receiver architectures for point-to-point communications and the MAC channel, we boosted the performance of linear receivers by adding successive cancellation. Is there something analogous in the BC channel as well?

IX.3.1. Sum capacity in SISO BC channel

This section characterizes the sum capacity of a class of Gaussian vector broadcast channels where a single transmitter with multiple transmit terminals sends independent information to multiple receivers. Coordination is allowed among the transmit terminals, but not among the receive terminals. The sum capacity is shown to be a saddle-point of a Gaussian mutual information game, where a signal player chooses a transmit covariance matrix to maximize the mutual information and a fictitious noise player chooses a noise correlation to minimize the mutual information.

The sum capacity is achieved using a precoding strategy for Gaussian channels with additive side information noncausally known at the transmitter. The optimal precoding structure is shown to correspond to a decision-feedback equalizer that decomposes the broadcast channel into a series of single-user channels with interference pre-subtracted at the transmitter.

The figure illustrates a N user (out of Ntot users) BC channel, where independent messages are jointly encoded by the transmitter x, and the receivers are each responsible for decoding the messages, respectively.



Broadcast channel

The sum capacity result has also been obtained in simultaneous and independent work by D.Tse and A. Goldsmith These two separate pieces of work arrive at essentially the same result via a duality relation between the multiple-access channel capacity region and the dirty-paper precoding region for the broadcast channel. The proof technique contained in this section is different in that it reveals an equalization structure for the optimal broadcast strategy. This decision-feedback equalizer viewpoint leads directly to a path for implementation, thus connecting with the structures that have been given in chapter VI. It also makes the capacity result amenable to practical coding schemes.

Further, the result in this section is in fact more general than that of D. Tse and A. Goldsmith. The presented results apply to broadcast channels with arbitrary convex input constraints, while the results of Tse and Goldsmith appear to be applicable for broadcast channels with a total power constraint only. First we address the sum capacity optimization, the capacity region for the vector broadcast channel is addressed later on in this section.

The BC channel can be formulated as

$$\mathbf{y}_1 = \mathbf{H}_1\mathbf{x} + \mathbf{z}_1 = \mathbf{H}_1(\mathbf{x}_1 + \mathbf{x}_2) + \mathbf{z}_1$$
$$\mathbf{y}_2 = \mathbf{H}_2\mathbf{x} + \mathbf{z}_2 = \mathbf{H}_2(\mathbf{x}_1 + \mathbf{x}_2) + \mathbf{z}_2$$

or, under a superuser formulation (where the number of scheduled users N coincides with the total number of users Ntot)

$$\mathbf{y}_{n_r N x1} = \mathbf{H}_{(nrxN)x\ nt}\mathbf{x}_{ntx1} + \mathbf{z}_{n_r Nx1}$$
$$\mathbf{y} = \begin{bmatrix}\mathbf{y}_1\\ \mathbf{y}_2\end{bmatrix} \quad \mathbf{H} = \begin{bmatrix}\mathbf{H}_1\\ \mathbf{H}_2\end{bmatrix} \quad \mathbf{x} = \begin{bmatrix}\mathbf{x}_1\\ \mathbf{x}_2\end{bmatrix} \quad \mathbf{z} = \begin{bmatrix}\mathbf{z}_1\\ \mathbf{z}_2\end{bmatrix}$$

This section characterizes the maximum sum rate R1+R2. The development here is restricted to the two-user case for simplicity.

When a Gaussian broadcast channel has a scalar input and scalar outputs, it can be regarded as a degraded broadcast channel (see chapter VI) for which the capacity region is well established. Intuitively, this means that one user's signal is a noisier version of the other user's signal. The capacity region for a degraded broadcast channel is achieved using a superposition coding and interference subtraction scheme due to Cover. The "dirty paper" result by Costa gives us another way to derive the degraded Gaussian BC channel capacity. Let us go first for the Cover's scheme. Consider the Gaussian scalar broadcast channel

y1=x+z1
y2=x+z2

where x is the scalar transmitted signal subject to a power constraint P. Assume that $s_1 < s_2$. Then , z2 can be rewritten as z2=z1+z', where z' is $N(0, s_2^2 - s_1^2)$ is independent of z1. Since z2' has the same distribution as z2, y2 is now equivalent to y1+z'. Thus, y2 can be regarded as a degraded version of y1. The capacity region for a degraded broadcast channel si achieved by dividing the total power into P1=aP and P2=(1-a)P (0<a<1) and to construct two independent messages, one codeword is chosen from each codebook, and their sum is transmitted. Because y2 is a degraded version of y1, the codeword intended for y2 can also be decoded by y1. Thus, y1 can subtract the effect of the codeword intended for y2 and can effectively get a cleaner channel with noise power s_1^2 instead of $s_1^2 + P_2$.

Recalling chapter VI, the following rate pair is achievable:

$$R2 = \frac{1}{2}.\log\left(1+\frac{P_2}{s_2 + P_1}\right)$$

$$R1 = \frac{1}{2}.\log\left[1+\frac{P_1}{s_1}\right]$$

It can be shown that, when $s_1 > s_2$ the above rate region is smaller than the true capacity region formulated next

$$R1 = \frac{1}{2}.\log\left(1+\frac{P_1}{s_1 + P_2}\right)$$

$$R2 = \frac{1}{2}.\log\left[1+\frac{P_2}{s_2}\right]$$

Thus, the decoding order matters. This successive cancellation scheme can be carried out at the transmitter instead of doing it at the receiver, thus resulting the so-called Dirty Paper coding. It was proposed by Costa and it also achieves capacity in the BC SISO channel. In DP, the code for x2 is written on a paper got in dirt by the code of signal x1. In this way, the degraded channel is not necessary.

When a Gaussian BC channel has a vector input and vector outputs, it is no longer necessarily degraded, and superposition coding is no longer capacity achieving. For instance, if a linear superposition of signals is transmitted at the base-station

$$\mathbf{x}(m) = \sum_{k=1}^{Ntot} u_k(m)\,\mathbf{b}_k$$

Then each user's signal will be projected differently onto different users, and there is no guarantee that there is a single user who would have sufficient SIRN to decode everyone else's data. However, the "dirty-paper" result by Costa can be extended to the vector case to presubtract multiuser interference at the transmitter, again with no increase in transmit power. Next, the rest of the section is devoted to obtain the precoding scheme that emulates dirty paper coding, thus achieving the sum capacity in a broadcast channel:

- **First**, we establish the relationship between the BC sum capacity and that in a cooperative scheme, both for MIMO. As a result, the optimal transmitter correlation matrix is designed.

- **Second**, from the optimal **R**x, a capacity lossless precoder is designed in order to obtain independent reception process for each signal xi, thus suitable for BC. The final scheme is called DP precoder and is based on the GDFE.

- **Third**, the optimal power loading at the transmitter is obtained

- **Finally**, we obtain the sum capacity as an aggregation of each stream rate and comment its relationship with the MAC sum capacity thanks to the existing duality among them.

IX.3.2. BC sum capacity and cooperative sum capacity for MIMO

Let us consider a 2 user BC channel

$$\begin{bmatrix}\mathbf{y}_1\\ \mathbf{y}_2\end{bmatrix}=\begin{bmatrix}\mathbf{H}_1\\ \mathbf{H}_2\end{bmatrix}\begin{bmatrix}\mathbf{x}_1\\ \mathbf{x}_2\end{bmatrix}+\begin{bmatrix}\mathbf{z}_1\\ \mathbf{z}_2\end{bmatrix}$$

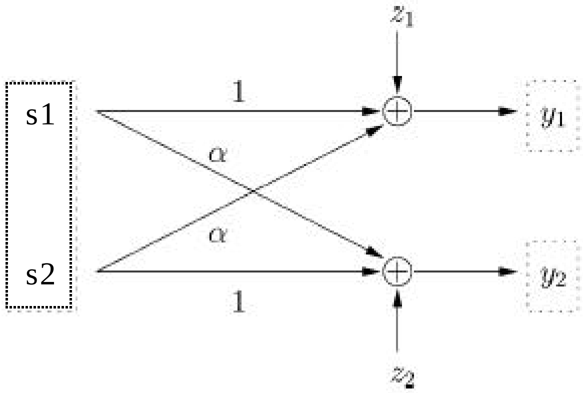
or

$$\mathbf{y}_i=\mathbf{H}_i\mathbf{x}+\mathbf{z}_i \qquad i=1,2$$

$$\mathbf{y}_1=\mathbf{H}_1(\mathbf{x}_1+\mathbf{x}_2)+\mathbf{z}_1=\mathbf{s}_1+a\mathbf{s}_2+\mathbf{z}_1$$
$$\mathbf{y}_2=\mathbf{H}_2(\mathbf{x}_1+\mathbf{x}_2)+\mathbf{z}_2=a\mathbf{s}_1+\mathbf{s}_2+\mathbf{z}_2$$

Because y1 and y2 cannot coordinate in a broadcast channel, the BC capacity does not depend on the joint noise distributions and only on the marginals. This is so because two broadcast channels with the same marginals but with different joint distribution can use the same encoder and decoders and maintain the same probability of error.

Let us first show an example that illustrates Sato's bound: the least favourable noise correlation depends on the structure of the channel (as α depends on the channel).



Simple BC channel

Consider the two-user two-terminal broadcast channel shown in above figure where the channel from s1 to y1 and the channel from s2 to y2 have unit gain, and the crossover channels have a gain a. Assume that s1 and s2 are independent Gaussian signals and z1 and z2 are Gaussian noises all with unit variance. The broadcast channel capacity is clearly bounded by I(X1,X2;Y1,Y2). This mutual information is a function of the crossover channel gain and the correlation coefficient r between z1 and z2. Consider the case α=0. In this case, the least favourable noise correlation is r=0 . This is because if z1 and z2 were correlated, decoding of y1 would reveal z1 from which z2 can be partially inferred. Such inference is possible, of course, only if y1 and y2 can cooperate. In a broadcast channel, where receivers y1 and y2 cannot take advantage of such correlation, the capacity with correlated and is the same as with uncorrelated z1 and z2. Thus, regardless of the actual correlation between z1 and z2 , the broadcast channel capacity is bounded by the mutual information evaluated assuming uncorrelated noise. Consider another case α=1 . The least favourable noise here is the perfectly correlated noise with r=1. This is because r=1 implies z1 and z2 equals. So, one of y1 and y2 is superfluous. If z1 and z2 were not perfectly correlated, collectively would reveal more information than y1 or y2 alone would. Since r=1 is the least favourable noise correlation, the broadcast channel sum capacity is bounded by the mutual information assuming r=1.

The previous explanation justifies that the cooperative capacity of the Gaussian vector channel with a least favourable noise bounds the capacity for the Gaussian broadcast channel

$$R_1 + R_2 \leq \min I(X;Y) = \min_{p(z)} I(X;HX+Z)$$

Sato's outer bound states that the broadcast channel sum capacity is bounded by (cannot be better than) the capacity of any discrete memoryless channel whose noise marginal distributions are equal to p(zi). The tightest outer bound is then the capacity of the channel with the least favourable noise correlation. Therefore,

$$
\begin{aligned}
& C_{sum}^{BC} \leq \min_{R_z} \max_{R_x} \frac{1}{2} \log \frac{\left| \mathbf{H R_x H^T} + \mathbf{R}_z \right|}{\left| \mathbf{R_z} \right|} = \min_{R_z} C_{sum}^{Coop} \\
& s.t. \quad tr(\mathbf{R}_x) \leq P \\
& \qquad \mathbf{R}_{z_i} \, known \\
& \qquad \mathbf{R}_x, \mathbf{R}_z \geq 0
\end{aligned}
$$

Additionally, although the actual noise distribution may not have the same joint distribution as the least favourable noise, because the marginal distributions are the same, a precoder designed for the worst noise and to require independent receivers, is oblivious of the correlation between zi's and performs as well as with the actual noise. Therefore,

$$
C_{sum}^{BC} \geq \max_{R_x} \min_{R_z} \frac{1}{2} \log \frac{\left| \mathbf{H R_x H^T} + \mathbf{R}_z \right|}{\left| \mathbf{R_z} \right|}
$$

If the function is convex-concave it has a saddle point and therefore $\min_x \max_y f(x, y) = \max_y \min_x f(x, y)$. This is our case[3], therefore

$$
C_{sum}^{BC} = \max_{R_x} \min_{R_z} \frac{1}{2} \log \frac{\left| \mathbf{H R_x H^T} + \mathbf{R}_z \right|}{\left| \mathbf{R_z} \right|}
$$

In order to obtain the optimal transmitter correlation matrix the max-min problem has to be solved.

The task of finding the least favourable noise correlation can be formulated as the following optimization problem.

$$
\begin{aligned}
& \min_{R_z} \frac{1}{2} \log \frac{\left| \mathbf{H R_x H^T} + \mathbf{R}_z \right|}{\left| \mathbf{R_z} \right|} \\
& s.t. \quad \mathbf{R}_z^{(i)} = \mathbf{I} \quad i = 1..N \\
& \qquad \mathbf{R}_z \geq 0
\end{aligned}
$$

To solve the problem we derive the Lagrangian

$$
\begin{aligned}
& L = \log \left| \mathbf{H R_x H^T} + \mathbf{R_z} \right| - \log \left| \mathbf{R_z} \right| + \sum_{i=1}^{2} tr\left(\mathbf{F}_i \left(\mathbf{R}_z^{(i)} - \mathbf{I} \right) \right) - tr\left(\mathbf{?} \, \mathbf{R}_z \right) \\
& \frac{\partial L}{\partial R_z} = 0 = \left(\mathbf{H R_x H^T} + \mathbf{R_z} \right)^{-1} - \mathbf{R}_z^{-1} + \begin{bmatrix} \mathbf{F}_1 & 0 \\ 0 & \mathbf{F}_2 \end{bmatrix} - \mathbf{?}
\end{aligned}
$$

and the least favourable noise is when its correlation matrix fulfils that

[3] The function $\frac{1}{2} \log \frac{\left| \mathbf{H R_x H^T} + \mathbf{R}_z \right|}{\left| \mathbf{R_z} \right|}$ is concave in Rz and $\min_{R_z} \frac{1}{2} \log \frac{\left| \mathbf{H R_x H^T} + \mathbf{R}_z \right|}{\left| \mathbf{R_z} \right|}$ is convex in Rx

$$\mathbf{R}_z^{-1} - \left(\mathbf{H}\mathbf{R}_x\mathbf{H}^T + \mathbf{R}_z\right)^{-1} = \begin{bmatrix} \mathbf{F}_1 & 0 \\ 0 & \mathbf{F}_2 \end{bmatrix}$$

or equivalently

$$\mathbf{R}_z + \mathbf{R}_z\left(\mathbf{H}\mathbf{R}_x\mathbf{H}^T\right)^{-1}\mathbf{R}_z = \begin{bmatrix} \mathbf{F}_1^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{F}_2^{-1} \end{bmatrix}.$$

Note that although the actual noise distribution may not have the same joint or cross distribution as the least favourable noise, the marginal distributions coincide. To interpret the obtained condition we can rewrite it in terms of the equivalent channel $\tilde{\mathbf{H}} = \mathbf{R}_z^{-1/2}\mathbf{H}$ as $\left(\mathbf{I} + \tilde{\mathbf{H}}\mathbf{R}_x\tilde{\mathbf{H}}^T\right)^{-1} = \begin{bmatrix} \mathbf{F}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{F}_2 \end{bmatrix}$. Then the condition can be seen as imposing block diagonal structure on the correlation matrix of the received signal after the whitening filter.

In order to find **R**x, note that the presented optimization problem suggests the following game-theoretical interpretation for the Gaussian vector broadcast channel. There are two players in the game. A signal player chooses an Rx to maximize the mutual information I(X;HX+Z) subject to the constraint tr(Rx)<=P . A noise player chooses a fictitious noise correlation Rz to minimize the mutual information subject to the constraint Rzi=I. A Nash equilibrium in the game is a set of strategies such that each player's strategy is the best response to the other player's strategy. The Nash equilibrium in this mutual information game exists, and the Nash equilibrium corresponds to the sum capacity of the Gaussian vector broadcast channel.

This max min scheme can be easily solved whenever a saddle point exist. Such a saddle point will be the Gaussian Broadcast channel sum capacity, and its calculations depend on both Rx and Rz. The saddle-point property of the Gaussian broadcast channel sum capacity implies that the capacity achieving is such that Rx is the water-filling covariance matrix for Rz ,and Rz is the least favourable noise covariance matrix for Rx. In fact, the converse is also true. This is because the mutual information is a concave–convex function, and the two KKT conditions, corresponding to the two optimization problems are, collectively, sufficient and necessary at the saddle-point. Thus, the computation of the saddle-point is equivalent to simultaneously solving the water-filling problem and the least favourable noise problem.

One might suspect that the following algorithm can be used to find a saddle-point numerically. The idea is to iteratively compute the best input covariance matrix for a given noise covariance, then compute the least favourable noise covariance matrix for the given input covariance. If the iterative process converges, both KKT conditions are satisfied, and the limit must be a saddle-point of $\frac{1}{2}\log\frac{\left|\mathbf{H}\mathbf{R}_x\mathbf{H}^T + \mathbf{R}_z\right|}{\left|\mathbf{R}_z\right|}$. Although such an iterative min-max procedure is not guaranteed to converge for a general game even when the payoff function is concave–convex, the iterative procedure appears to work well in practice for this particular problem. The convex–concave nature of the problem also suggests that general-purpose numerical convex programming algorithms can be used to solve for the saddle-point with polynomial complexity. Finally, the main sum capacity result can be easily generalized to broadcast channels with an arbitrary convex input constraint. This is so because the saddle-point for the mutual information expression is Gaussian as long as the input and noise constraints are convex.

Once Rx and Rz have been obtained by convex optimization, as for instance using the interior-point method, it just remain to make a connection between the transmitted symbols and the input data u, this is accomplished by the precoding matrix B.

IX.3.3. Precoder design

IX.3.3. 1. Towards non-cooperative receivers

Consider a Gaussian vector channel **y**=**Hx** + **z**. Assume that H is a square matrix. If the noise covariance matrix Rz is not block-diagonal, a noise whitening filter is required as a first step at reception. Suppose that the noise covariance matrix has an eigenvalue decomposition

$$\mathbf{R}_z = \mathbf{Q}^T ? \mathbf{Q}$$

If in addition, the transmitter covariance matrix Rx is also not block-diagonal, then a Gaussian source **u** and a transmit filter **B** can be created such that **R**u=**I** and **x=Bu**. Let the SVD of the optimal **R**x obtained in the previous section be

$$\mathbf{R}_x = \mathbf{V}\mathbf{S}\mathbf{V}^T = \mathbf{B}^{\mathbf{H}}\mathbf{B}$$

The appropriate transmit filter has the form

$$\mathbf{B} = \mathbf{V}\sqrt{\mathbf{S}}\mathbf{M}$$

where **M** is an arbitrary orthonormal matrix. It consists on a beamforming matrix **V**, a power allocation matrix S and a precoding matrix **M** whose whole is to carry out a proper interference cancellation in order to achieve capacity by decoupling the BC channels. The dimensions of **M** are the same dimension as $\mathbf{R}_z$, thus, number of receiving antennas. So, the rank of $\mathbf{R}_x$ is always equal to or lower than the rank of the superuser channel. When **R**x is of strictly lower rank, zeros can be padded in the channel to make the effective channel matrix a square matrix

$$\mathbf{R}_x = \mathbf{V}\mathbf{S}_o\mathbf{V}^T \qquad \mathbf{S} = [\sqrt{\mathbf{S}_0}\ 0]$$

In order to design **B**, let us take the configuration of the figure with MMSE reception, which is capacity lossless. The goal is to obtain a precoder **B** such that reception can be independently done by each receiver

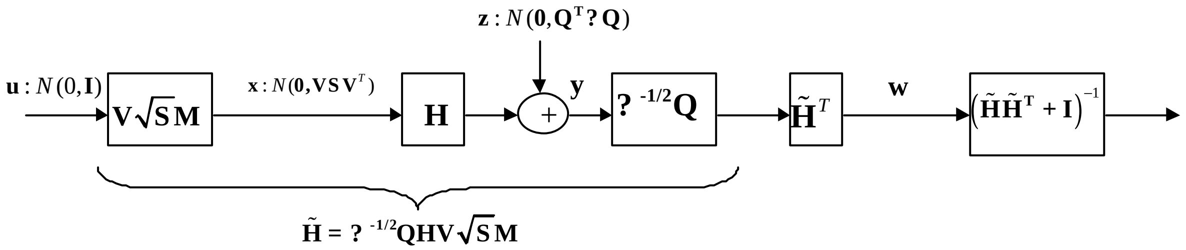


Figure: MMSE configuration

Note that the transmit filter and the noise whitening filter create the following effective channel

$$\tilde{\mathbf{H}} = ?^{-1/2}\mathbf{Q}\mathbf{H}\mathbf{V}\sqrt{\mathbf{S}}\mathbf{M}$$

Intuitively, the transmitting matrix $\mathbf{M}$ should transform this equivalent channel into a block diagonal channel so that each receiver could carry out independent reception. This is precisely the condition that we have obtained in order to solve the maxmin problem formulated before, where

$$\left(\mathbf{I}+\tilde{\mathbf{H}}\mathbf{R}_x\tilde{\mathbf{H}}^T\right)^{-1} = \begin{bmatrix} \mathbf{F}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{F}_2 \end{bmatrix}$$

being F i the dual variables of the optimization problem.

If the DFE receiver of the figure is considered because it allows maximal rate transfer to each parallel channel, then the whole system up to obtaining signal v is block diagonal (a more detailed solution of the problem is described in Appendix A).

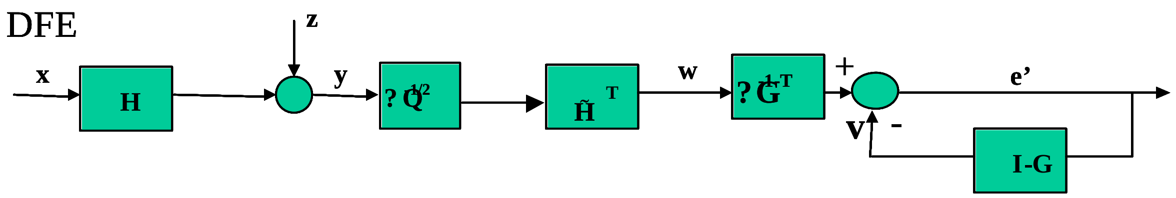


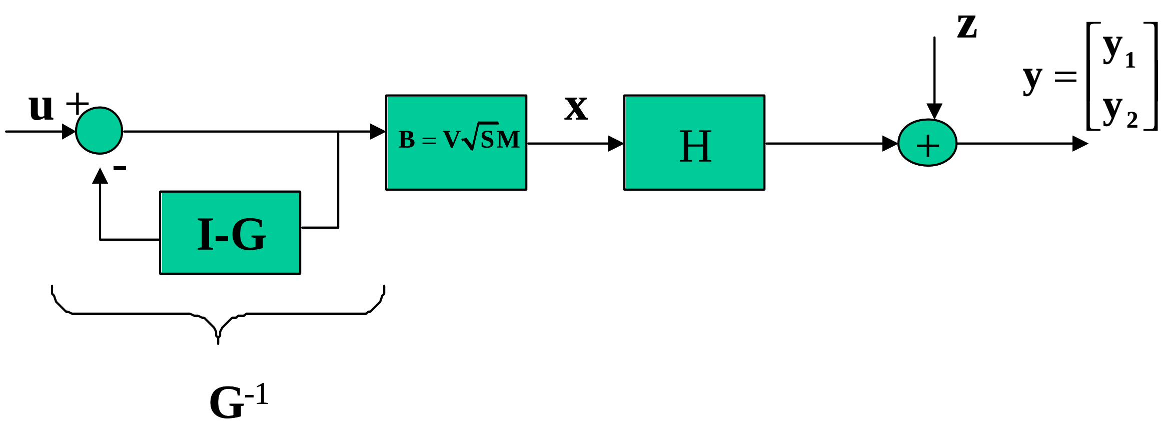
Figure: DFE receiver

There are still one question to answer before obtaining a complete precoding design:
- to complete the decoupled receiver design by transferring $\mathbf{G}^{-1}$ to the transmitter if possible.

<u>IX.3.3.2. Complete decoupled receiver design that achieves Csum: GDFE precoder versus DP precoder</u>

In order to finally get a decoupled receiver design, the feedback filtering at reception can be transferred to the transmitter side, so called **decision feedback precoder**, which results in a similar concept to DP coding.

The final design of the transmitter is then the one proposed in the figure



GDFE precoder and matrix **B** for achieving BC sum capacity

Therefore, matrix **B** is completed with $\mathbf{G}^{-1}$ implemented in a feedback way, in order to preserve capacity as appendix B shows. Thanks to the feedback implementation, the precoder follows a Dirty Paper philosophy. For instance, the transmitter first picks a codeword for receiver 2 with full (noncausal) knowledge of the codeword intended for receiver 1. Therefore, receiver 2 does not see the codeword intended for receiver 1 as interference. Similarly, the codeword for receiver 3 is chosen such that receiver 3 does not see the signals intended for receivers 1 and 2 as interference. This process continues for all K receivers. Receiver 1 subsequently sees the signals intended for all other users as interference, receiver 2 sees the signals intended for users 3 to K as interference, etc. Note that the ordering of the users clearly matters in such a procedure.

There is one aspect left and it is regarding the question whether the structure of the GDFE precoder is capacity lossless or not. The answer is yes, it is capacity loss-less and the prove can be found in the work done by Cioffi and also considered in the appendix B of this chapter.

To sum up, with the least favourable noise, there exists a GDFE structure with a block – diagonal feedforward filter. This, together with a precoder B, eliminates the need for coordination at the receiver. **Thus, the precoding GDFE achieves the BC sum capacity and the precoding matrix B has been obtained**.

Algorithm for achieving Csum at broadcast:

1. Solved the maxmin problem and get the best **R**x for the worst **R**z
2. From the convex optimization problem get the dual variables $\mathbf{F}_i$
3. From **R**x and $\mathbf{F}_i$, obtain $\mathbf{B} = \mathbf{V}\sqrt{\mathbf{S}}\mathbf{M}$ that decouples in an optimal way part of the lossless MMSE receiver

4. The rest of the receiver is $\mathbf{G}^{-1}$, which is moved to the transmitter. When implemented as a feedbackward filtering, the final transmitter structure emulates the DP precoding philosophy. Recall that:

$$\mathbf{G} = \begin{bmatrix} \mathbf{I} & \left(\mathbf{R}_{\mathbf{x}}^{-1} + \mathbf{H}_1^{\mathbf{T}}\mathbf{H}_1\right)^{-1} \mathbf{H}_1^{\mathbf{T}}\mathbf{H}_2 \\ \mathbf{0} & \mathbf{I} \end{bmatrix}^{-1}$$

IX.3.4. BC sum capacity and MAC-BC duality

The DFE transmit filter **B** designed for the least favourable noise also identifies the set of sum capacity-achieving **R**i. For example, for the 2 user case,

$$R_1 = \frac{1}{2}\log\frac{\left|\mathbf{H}_1\mathbf{R}_1\mathbf{H}_1^{\mathbf{T}} + \mathbf{H}_1\mathbf{R}_2\mathbf{H}_1^{\mathbf{T}} + \mathbf{R}_{z1}\right|}{\left|\mathbf{H}_1\mathbf{R}_2\mathbf{H}_1^{\mathbf{T}} + \mathbf{R}_{z1}\right|} = I(X_1;Y)$$

$$R_2 = \frac{1}{2}\log\frac{\left|\mathbf{H}_2\mathbf{R}_2\mathbf{H}_2^{\mathbf{T}} + \mathbf{R}_{z2}\right|}{\left|\mathbf{R}_{z2}\right|} = I(X_1;\, Y/\, X_2)$$

which are the rates that are obtained in the appendix B for DFE precoding (reversing the orders of user 1 and 2) and are also called dirty paper rates.

For the general case of K users, the rate for user i that achieves sum capacity is

$$R_i = \frac{1}{2}\log\frac{\left|\mathbf{H_i}\left(\sum_{k=i}^{K}\mathbf{R_k}\right)\mathbf{H_i^T} + \mathbf{I}\right|}{\left|\mathbf{H_i}\left(\sum_{k=i+1}^{K}\mathbf{R_k}\right)\mathbf{H_i^T} + \mathbf{I}\right|}$$

where users are ordered in increasing order. One important feature to notice about the dirty paper rate equations is that the rate equations are neither a concave nor a convex function of the covariance matrices. This makes finding the dirty paper region very difficult, because generally the entire space of covariance matrices which meet the power constraint must be searched over. This justifies our focus under a filtering perspective that departs from the MAC channel instead of directly obtaining the precoders for each user from the general maxmin problem.

The sum capacity is

$$C_{sum}^{BC} = \sum_{i=1}^{Ntot} R_i$$

resulting

$$C_{sum}^{BC} = \sum_{i=1}^{N_{tot}} R_i = \log\left|\mathbf{I}+\mathbf{H}_{Ntot}^{\mathbf{H}}\mathbf{R}_{Ntot}\mathbf{H}_{Ntot}\right| + \ldots + \log\frac{\left|\mathbf{I}+\mathbf{H}_1^{\mathbf{H}}\left(\sum_{i=1}^{Ntot}\mathbf{R}_i\right)\mathbf{H}_1\right|}{\left|\mathbf{I}+\mathbf{H}_1^{\mathbf{H}}\left(\sum_{i=1}^{Ntot-1}\mathbf{R}_i\right)\mathbf{H}_1\right|}$$

$$= \log\left|\mathbf{I}+\sum_{i=1}^{Ntot}\mathbf{H}_{\mathbf{i}}^{\mathbf{H}}\mathbf{R}_i^{MAC}\mathbf{H}_{\mathbf{i}}\right| \qquad with \quad \sum_i Tr\left[\mathbf{R}_i^{MAC}\right] \le P_T$$

Note also that the same rates sum capacity was obtained for the MAC channel with DFE reception. This is also called MAC-BC duality [44]: the capacity region of the BC channel can be obtained with the union of MAC capacity regions with equal average power constraints and not allowing to the transmitters to cooperate in the MAC. One key point is that to achieve the same rate vector in the BC and MAC, the decoding order must in general be reversed, i.e., if user 1 is decoded last in the BC then user 1 is decoded first in the MAC. Exploiting duality, in [43], Goldsmith proposes an alternative algorithm for the iterative design of the precoder in a MIMO BC channel.

In order to get some insight into the sum rate expression, note that in the case of only multiple antenna at the transmitter with MMSE decoding, the sum rate results in

$$R^{DP} = \sum_{i=1}^{Ntot}\log\left(1+\frac{\left|w_{ii}\right|^2}{1+\sum_{j>i}\left|w_{ij}\right|^2}\right) = \sum_{i=1}^{Ntot}\log\left(1+SNIR_i^{DP}\right) \qquad \mathbf{x}=\mathbf{B}\,\mathbf{u} \quad w_{ij} = \left[\mathbf{HB}\right]_{ij}$$

(**)

Being the last term of the equality very useful when designing practical precoding schemes for BC with the aim of sum rate maximization. Caire and Shamai proposed in [7], for the case of only antennas at transmission, that the components of u should be generated by successive dirty-paper encoding with Gaussian codebooks and B should maximized (**) over all precoding matrices B satisfying the trace or power constraint. For the 2 user case with 1 antenna per user they obtained that

$$R\begin{cases}\log(1+\left|\mathbf{h}_1\right|^2 A) & A \le A_1 \\ \log\dfrac{\left(A\left|\mathbf{HH}^H\right| + trace(\mathbf{HH}^H)\right)^2 - 4\left|\mathbf{h_2}\ \mathbf{h_1^H}\right|^2}{4\left|\mathbf{HH}^H\right|} & A > A_1\end{cases}$$

where $\left|\mathbf{h_1}\right|^2 \ge \left|\mathbf{h}_2\right|^2$ and $A_1 = \dfrac{\left|\mathbf{h}_1\right|^2 - \left|\mathbf{h}_2\right|^2}{\left|\mathbf{HH}^H\right|}$

As a conclusion,

$$C_{SUM} = R^{DP} = \max_{\substack{k=1..N \\ \sum_{k=1}^{N} Tr\Sigma_k \le P}} \operatorname{logdet}\left(\mathbf{I}+\sum_{k=1}^{N}\mathbf{H_k}\mathbf{S}_k\mathbf{H_k^H}\right)$$

$$\mathbf{S_k} = E\left\{\mathbf{x_k}\mathbf{x_k^H}\right\} = \mathbf{B_k}\mathbf{P}_k\mathbf{B}_k^{\mathbf{H}} \quad \mathbf{B}_k : DFE\ precoding(or\ DP)$$

or in the case of just multiple antenna at the transmitter

$$C_{SUM} = R^{DP} = \max_{p_1, p_2 \cdots \sum_{k=1}^{N} p_k \le NP} \log\det\left(\mathbf{I} + \sum_{k=1}^{N} \mathbf{h}_k p_k \mathbf{h}_\mathbf{k}^\mathbf{H}\right)$$

<u>Multiuser diversity</u>

If both the transmitter and receivers know the channel perfectly in a BC with Ntot single-antenna receivers with average transmit power of nt SNR and the transmitter has nt antennas, then for sufficiently large Ntot, the sum rate capacity scales like

$$C_{SUM}\big|_{Ntot\to\infty} \approx E\{C_{sum}\}_{Ntot\to\infty} \approx nt \log\log(Ntot\ SNR)$$

Where nt and SNR are fixed.

In order to prove it observe that

$$\begin{aligned}
& E\left\{ \max_{p_1, p_2 \cdots \sum_{k=1}^{N} p_k \le NP} \log\det\left(\mathbf{I} + \sum_{k=1}^{N_{tot}} \mathbf{h}_\mathbf{k} p_k \mathbf{h}_\mathbf{k}^\mathbf{H}\right)\right\} \\
& \le nt\ E\left\{ \max_{p_1, p_2 \cdots \sum_{k=1}^{N} p_k \le NP} \log\left(1 + \frac{\sum_{k=1}^{N_{tot}} Tr(\mathbf{h}_\mathbf{k}\mathbf{h}_\mathbf{k}^\mathbf{H} p_k)}{nt}\right)\right\} \\
& \le ntE\left\{ \max_{p_1, p_2 \cdots \sum_{k=1}^{N} p_k \le NP} \log\left(1 + \frac{\max_{1\le i\le Ntot} Tr(\mathbf{h}_\mathbf{k}\mathbf{h}_\mathbf{k}^\mathbf{H}) \sum_{k=1}^{N_{tot}} p_k}{nt}\right)\right\} \\
& = ntE\left\{\log\left(1 + P \max_{1\le i\le Ntot} Tr(\mathbf{h}_\mathbf{k}\mathbf{h}_\mathbf{k}^\mathbf{H})\right)\right\}
\end{aligned}$$

Where we have used the inequality $\det(\mathbf{A}_{MxM}) \le \left(\frac{tr(\mathbf{A})}{M}\right)^M$. As Ntot goes to inf. the max behaves like log Ntot+O(log log Ntot).

Compared to the single user capacity of nt log(1+SNR), we observe that the sum-rate increases double-logarithmically in Ntot.. Thus, the multiuser diversity gain increases SNR by a factor of logNtot. This is precisely the basis of the so-called opportunistic schemes, which just need SNIR feedback instead of the whole knowledge of **H**. Due to its importance in practical scheduler we devote later on a specific section to study these schemes.

Next low complexity alternatives to the optimal precoding are proposed. Note that if only the transmitter can be designed taking into account the superuser channel **H**, a two matrix decomposition of the channel is desirable, rather than a 3 matrix decomposition as the SVD. Cholesky, QR or LU decompositions are possible candidates that may lead to designs different from the one proposed. When full CSIR or CSIT is available, optimal structures in MAC or BC carry out a successive interference cancellation or signal encoding respectively. However, other solutions can be found in between that trade-off performance and complexity in implementation and in required CSI. When studying these alternative schemes, note that duality between MAC and BC can be carried out, thus transferring receiving filters to the transmitter and the other way round. When doing that one must be careful in not incurring in any capacity loss or transmitting power increase.

IX.3.5. Low complexity precoding structures

Chapter VI was precisely devoted to this subject. Low complexity precoding structures were obtained based on the idea extending the "dirty paper" concept of the degraded BC to the non-degraded one (that appears when either multiantenna are present at transmission or reception). Intuitively, this extension relates with moving the feedbackward part of the DFE at the transmitter. This implies a "triangular interference cancelling", that allows for interference suppression without increase in the transmitted power, in contrast to a straight forward Zero Forcing precoding. Matching intuition with theory, this section has shown the optimality, in terms of BC sum capacity, of moving the feedbackward filtering of the DFE to the transmitter and has given the optimal design procedure for the precoding.

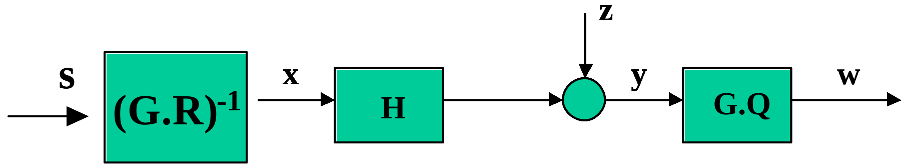
Due to the design complexity of the optimal precoding, in the literature there are various different practical implementation schemes. Although Chapter VI sums up the basics of these different schemes, we comment on some examples.

Tomlinson-Harashima

The basis of most of the practical precoders is the structure used by the Tomlinson-Harashima precoder. In that case, an alternative to design the DFE precoding is by considering a **QR decomposition of the channel**

$$\mathbf{H} = \mathbf{Q}^H\mathbf{R}$$

Being **R** lower triangular and modifying the receiver design accordingly to next figure where now the output is **w**



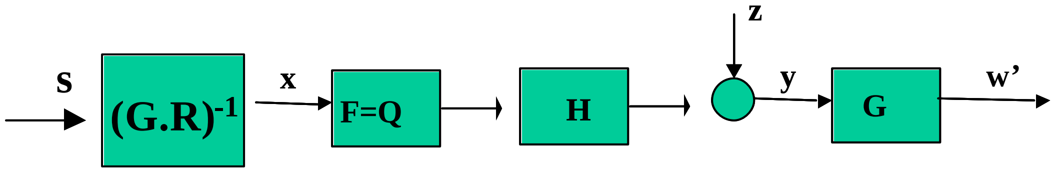
$$\mathbf{G}=\mathrm{diag}\{r_{ii}\}$$

$$\mathbf{w} = \mathbf{GQHx} + \mathbf{GQz} = \mathbf{GQQ^HRx} + \mathbf{GQz} = \mathbf{GRx} + \mathbf{GQz} = \mathbf{s} + \mathbf{GQz} \text{ if } \mathbf{B}=(\mathbf{GR})^{-1}$$

In order to study if it is capacity lossless the correlation of the error **e=w-s** should be studied. Observe that $\mathbf{R_e} = \mathbf{GQQ}^H\mathbf{G}^{\mathbf{H}} s^2$, thus diagonal and capacity lossless. However, if we want to obtain parallel receivers, the RQ decomposition of the channel should be considered, the lower triangular matrix R changes, and a different precoding structure is obtained, where the receivers are decoupled, thus suitable for the broadcast channel

$$\mathbf{H=RQ^H}$$

where **R** is a lower triangular matrix of dimension (Ntot x m) and $\mathbf{Q}^H$ is unitary of dimension m x nt (where m=rank(**H**))



$$\mathbf{G}=\text{diag}\{r_{ii}\}$$

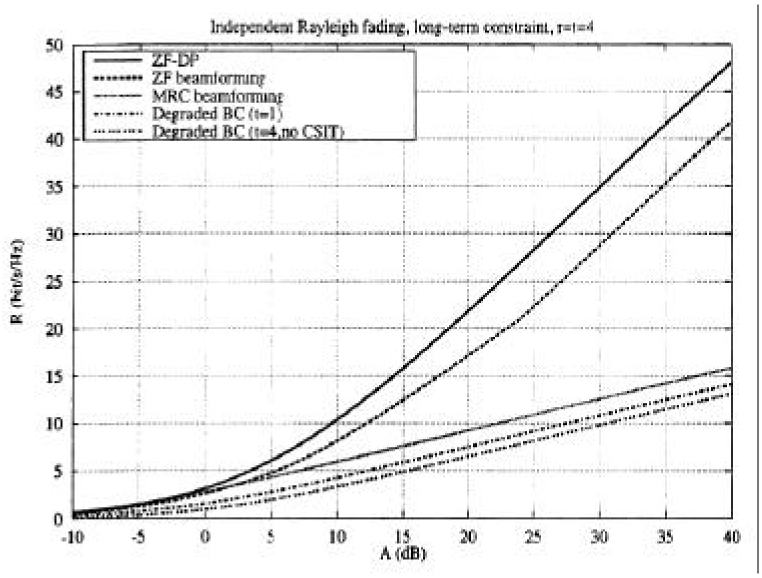
where now the output is **w'**

$\mathbf{w' = GHFx + Gz = GRQ^HQx + Gz = GRx + Gz = s + Gz}$ if $\mathbf{B=(GR)^{-1}}$

still being capacity lossless.

Zero Forcing precoder

Other examples are the studies of Caire and Shamai who compare different possibilities for the case of nt transmitting antennas and N users, with 1 antenna per user. In the precoding scheme, **x**=**Bu**, they obtained **u** by successive dirty-paper encoding so that nulling the interference terms i>j, while the remaining i<j terms are forced to zero by letting **B**=**Q**, where **Q** is the unitary matrix that results from the QR decomposition of the channel, **H**=**RQ** (where **R** is lower triangular). This suboptimal ZF-DP coding strategy is shown to achieve asymptotically optimal throughput for high SNR if the channel matrix has full row rank, while for vanishing SNR, it reduces to simple maximal ratio combining beamforming to the best user, which is shown to be also optimal in general, for low SNR. Next figure (figure 5 from [7]) plots some comparative results for nt=4 and N=4 users and depicts the importance of channel knowledge at the transmitter. Note that ZF beamforming consists of inverting the channel matrix at the transmitter in order to create orthogonal channels between the transmitter and the receivers without receivers' cooperation. Although it was one of the first BC architectures to be studied in the literature because of it simplicity, it is not optimal as the figure shows. We will come to the problem of transmit beamforming design later on.



Throughput versus SNR comparison

The practical implementation not only cares about the complexity and the tractability of the designed precoders/receivers, but also on the other system aspects as system fairness, users access control and system delay. For instance, next table (from [38]) compares ZF, DP-QR (with random ordering of users) and SVD precoding (called cooperative in the table). In all of them uniform power allocation is considered. The appendix compares the aforementioned techniques for the instantaneous channel case.

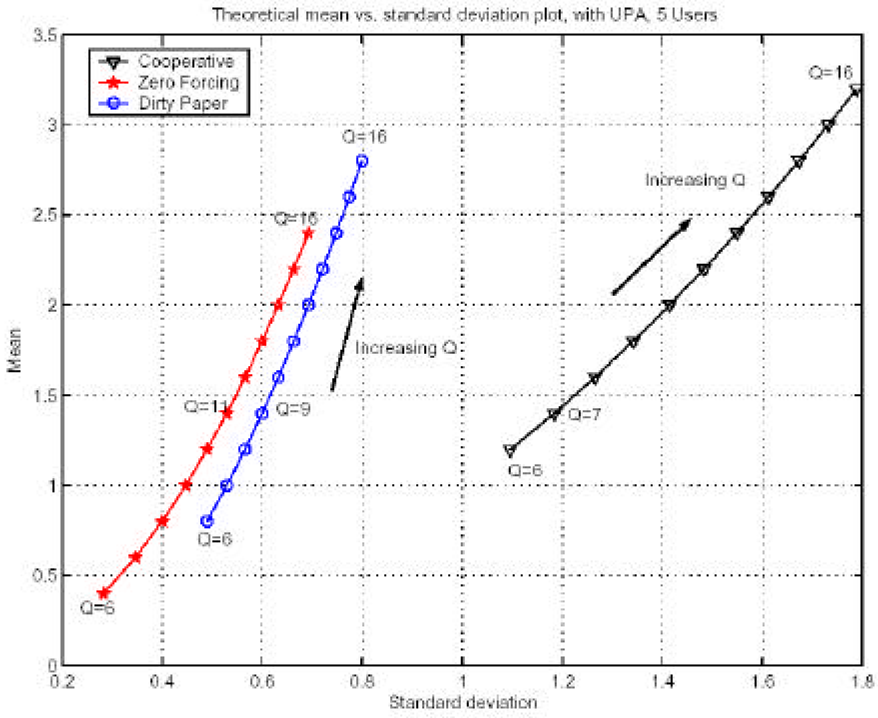
Technique	Gain	Mean	Standard Deviation	Asymptotic IF
Cooperative	λ_k^2/K	Q/K	$\sqrt{Q/K}$	$1/(1+\xi)$
Dirty Paper	d_k^2/K	$(2Q-K+1)/2K$	$\sqrt{Q+\frac{1}{12}(K-5)(K-1)}/K$	$(2-\xi)^2/\left[(2-\xi)^2+\xi^2/3\right]$
Zero Forcing	α_k^2/K	$(Q-K+1)/K$	$\sqrt{Q-K+1}/K$	1

Comparative mean SNR and variance among users, Q is the number of antenna and N the number of users

<u>Fairness</u>

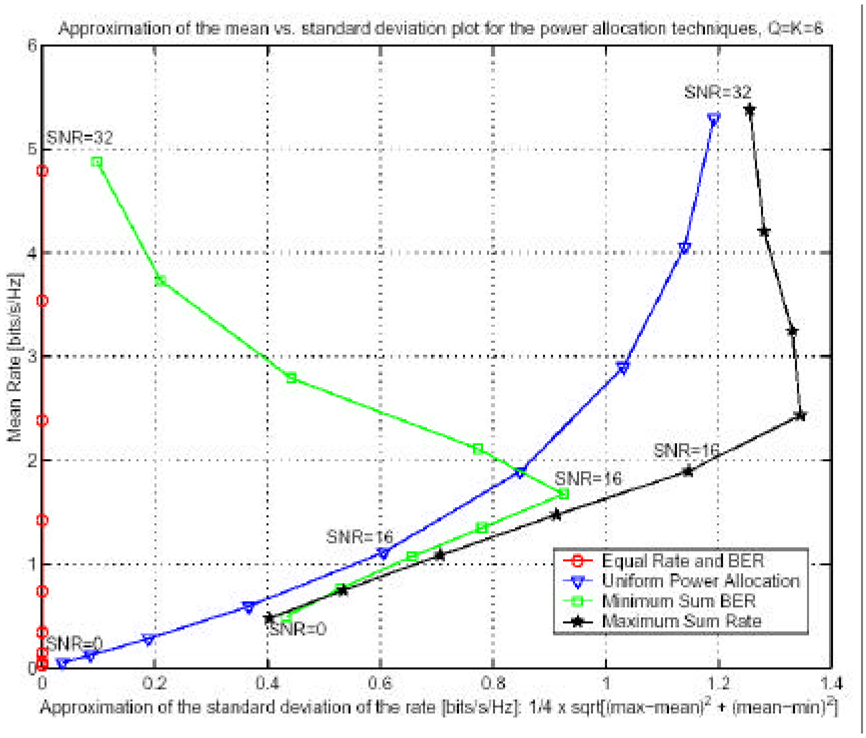
One of the aspects in the table is the index of fairness (IF), which is defined as the ratio between the mean and the variance. Although ZF is the precoder with worst performance it is the most fair among users. When getting into multiuser systems, a new aspect to study is the global performance of the system versus the individual performance. Depending on the desired fairness, different criteria design are to be chosen, which will not necessary coincide with the maximization of the sum rate.

Next figure is an example of the performance of the three precoders. Again, although the cooperative technique is the one that offers better performance for a specific SNR, it is also the most unfair among users. In the scheduling section, which takes into account system practical aspect, we come back to this fairness issue.



Comparative results for 5 users and varying number of antenna.

When considering practical precoders one aspect to design is the power allocated to each user. Next figure compares the sum rate results when a ZF precoder is used with different design criteria.



Outage mean rate vs. the approximation of the standard deviation at 90% of SNR

Note that the sum rate criterion is not necessarily the best one if fairness issues are to be taken into account. Note also that when designing the precoder in MIMO transmission the global system performance in terms of rate do not give the same results as if BER were the goal and it is important to know how to use the spatial degrees of freedom properly in terms of both, diversity and multiplexing (or rate) gain. In SISO transmission this aspect was not a concern because better BER implied better rate and the other way round.

<u>Access Control</u>

In a cooperative transmission, since for any unitary matrix **Q** (permutation matrices are unitary), the matrix **QH** has the same singular values of **H**, the sum capacity in a cooperative system is independent of the user ordering. On the other hand, ZF beamforming depends on the choice of the unordered active user set and ZF-DP depends on the ordered active user set. Note that in both cases, only m=rank(**H**) users can be served simultaneously. The access control or user selection problem is still to be solved in many practical situations and offers an important degree of freedom that can be exploited in order to improve the performance of suboptimal systems. Ideally, if the number of users goes to infinite, a subset of users with mutually orthogonal spatial signatures could be found and ZF beamforming would amount to a unitary transformation, thus involving no power penalty and perfect user separation. This can be seen as another manifestation of the ubiquitous principle of multiuser diversity, that plays a central role in multiuser channels with fading as we will seen later on. In [38;46, 47] the authors address the problem when ZF is applied. Note that the grouping of users requires an exhaustive search over the entire user set. This makes a low complexity implementation of the optimal ZF challenging and userful.

To finish this section, next, diversity optimizing techniques or transmit beamforming are presented.

Transmit beamforming

In the literature, practical schemes concentrate on the diversity advantage, that means, the increase of the effective SNIR at the receivers. Moreover, they are not based on information-theoretic issues, but rather strategies that could be currently implemented. In these schemes, the transmitted signal **x** has been substituted by a beamforming matrix **B**. The general SNIR for the kth user in the multiuser MISO system, γk, can be expressed as

$$g_k = \frac{\mathbf{b_k^H R_k b_k}}{\sum_{i \neq k} \mathbf{b_i^H R_k} \mathbf{b}_i + s_k^2}$$

The different strategies consist on transmit beamforming design and power assignment. The optimal beamforming strategy [46] in terms of rate is the one that

$$R_{BF} = \max_{b_k, P_k} \sum_{i=1}^{Ntot} \log(1 + SNIR^{BF})$$
$$s.t. \sum_{i=1}^{Ntot} \left|\mathbf{b}_k\right|^2 P_k \leq P$$

which is optimal for large number of users, as it achieves the same rate as DP coding [46]

$$E\{R_{BF}\} \approx nt \log(1 + \frac{P}{nt} \log Ntot)$$

but it is difficult to carry out in practice. Note that the SINR of each user in the BC depends in general on all the transmit signatures of the users. Hence, it is not meaningful to pose the problem of choosing the transmit beamformers to maximize each of the SINR separately. A more sensible formulation is to minimize the total transmit power needed to meet a given set of SINR requirements.

$$\min_{b_k} \sum_{k=1}^{K} \mathbf{b_k^H b_k}$$
$$s.t. g_k \geq g_{thres} \qquad k = 1 \ldots K$$

The optimal transmit signatures balance between focusing energy in the direction of the user of interest and minimizing the interference to other users. This transmit strategy can be thought of as performing transmit beamforming. Implicit in this problem formulation is also a problem of allocating powers to each of the users.

Taking into account the uplink-downlink duality (see appendix C), the transmit beamforming problem can be solved by looking at the uplink dual. Since for any choice of transmit signatures, the same SINR can be met in the uplink dual using the transmit beamformer as receive filters and the same total transmit power, the BC problem is solved if we can find receive filters that minimize the total transmit power in the uplink dual. The receive filters are always chosen to be the MMSE filters given the transmit powers of the users; the transmit powers are iteratively updated so that the SINR requirement of each user is just met. This MMSE beamformers can now be used as the optimal transmit beams in the BC and afterwards the optimal power allocation can be found. The optimal beamforming in MAC and BC are the same if power constraint is the same, the difference between both channels is the power that has to be allocated to each user. This duality is going to be used to obtain the BC capacity region in next section.

Note that if fairness issues come into play, other strategies are possible as that in [38; 48] the design is based on the maximization of the minimum SINR subject to a power constraint

$$\max_{b_k} \min_{k} g_k$$
$$\sum_{k=1}^{K} p_k \leq P$$

Conclusions

Along this section we have shown that aside from the theoretical analysis, practicality aspects are really important in multiuser MIMO systems. The main aspects are:

- The ordering of the users clearly matters in such a procedure and needs to be optimized in the capacity computation
- For latency and degrees of freedom reasons (N< Ntot). Then the throughput can be further optimized with respect to the active user set and SDMA (Spatial diversity multiple access) has to be combined with other strategies, as for instance TDMA. These aspects refer to access control policies
- The real situation is to have partial CSIT, which make optimal strategies such as DP unfeasible, thus requiring also for an access control policy.
- When not only optimality but users priority and QoS come into play, fairness has to be considered.
- Finally, in delay is also an important parameter to control in the network and come into play when users queues of finite length are considered. Although it is out of the scope of this chapter, the multiuser MIMO system design is narrowly related with the users buffer control. Some examples are shown in the last section of the chapter that considers the scheduling problem.

Basically, in the design of a scheduler or practical broadcast system 3 questions have to be answered:

- Transmitter architecture
- Power allocation
- Access control (i.e. number of users to give access and order)

In the section devoted to scheduling these aspects are going to be considered. Next, before getting into the problem of partial CSIT we study the BC capacity region.

IX.3.6. BC capacity region

When multiple users share the same channel, the channel capacity can no longer be characterized by single number. Since there is an infinite number of ways to divide the channel between many users, the multiuser channel capacity is characterized by a rate region.

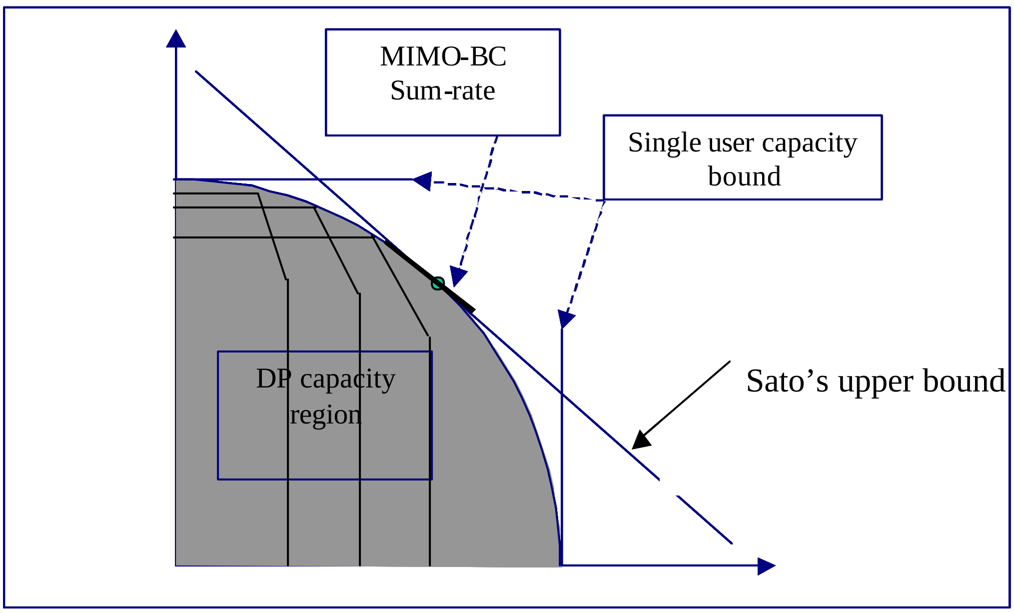
We have seen that in the MIMO BC, DP coding can be applied at the transmitter when choosing codewords for different users. The ordering of the users clearly matters in such a procedure, and needs to be optimized in the capacity calculation. Let $p(.)$ denote a permutation of the user indices and $\mathbf{R_x} = [\mathbf{R}_1 ... \mathbf{R_K}]$ denote a set of positive semi-definite covariance matrices with $Tr(\mathbf{R}_1 + \ldots + \mathbf{R_K}) \le P$. Under DP coding, if user $p(1)$ is encoded first, followed by user $p(2)$, etc., then the following rate vector is achievable

$$R(p, \mathbf{R}_i) : R_{p(k)} = \frac{1}{2} \log \frac{\left| \mathbf{H}_{p(k)} \left(\sum_{j=k}^{K} \mathbf{R}_{p(j)} \right) \mathbf{H}_{p(k)}^{\mathbf{T}} + \mathbf{I} \right|}{\left| \mathbf{H}_{p(k)} \left(\sum_{j=k+1}^{K} \mathbf{R}_{p(j)} \right) \mathbf{H}_{p(k)}^{\mathbf{T}} + \mathbf{I} \right|} \qquad k = 1,.., K$$

The capacity region C is then the convex hull of the union of all such rates vectors over all permutations and all positive semi-definite covariance matrices satisfying the average power constraints

$$C_{BC}(P, \mathbf{H}) = Co\left(\bigcup_{p, \mathbf{R}} R(p, \mathbf{R}) \right)$$

The transmitted signal is $\mathbf{x} = \mathbf{x}_1 + \ldots + \mathbf{x_K}$. The DP coding implies that $\mathbf{x}_1, ..., \mathbf{x_K}$ are uncorrelated, and thus $\mathbf{R_x} = \mathbf{R}_1 + \ldots + \mathbf{R_K} \le P$



Capacity region for the BC

As we have already commented, one important feature to notice about the rate equations is that they are neither a concave nor convex function of the covariance matrices. However, by exploiting the duality between the MIMO BC and the MIMO MAC that can be exploited to greatly simplify this calculation. The figure outlines the capacity region for the BC. In a separated way, Tse (uplink-downlink duality) proved that the DP or DFE achievable region achieves the sum rate capacity of the MIMO Gaussian BC. Also Shamai (Enhanced channel and Minkowski’s inequality) showed that DP coding or DFE precoding achievable rate region is the capacity region of the Gaussian MIMO BC.

The capacity region for the BC is obtained as the union of capacity regions of the dual MAC, where the union is taken over all individual power constraints that sum up to the BC power constraints

$$C_{BC}(\overline{P},\mathbf{H}) = \bigcup_{\{P_i\}_i^K : \sum_{i=1}^{K} P_i \le \overline{P}} C_{MAC}\left\{\mathbf{R}_1,...,\mathbf{R}_K;\mathbf{H}\right\}$$

This leads to the conclusion that the uplink (MAC) and downlink (BC) channels differ only due to the fact that power constraints are placed on each transmitter in the MAC instead of on all transmitters jointly. As shown in the figure, every point on the boundary of the BC capacity region is a corner point of the dual MAC for some set of powers with the same sum power. **Thus, the dirty paper BC achievable region equals the sum power MIMO MAC capacity region.** Successive encoding in the transmitter or successive decoding in the receiver is then required to separate the different signals that are superimposed at the channel output in a BC or MAC channel, respectively. The fundamental trade-off between users is here parameterized by two components: I) the poser allocation, which should be performed jointly, and ii) the encoding/decoding order, with Ntot! Possible orderings of the users. The figure depicts that dirty paper coding achieves the Sato upper bound, and therefore, equals the sumrate cpacity of the MIMO BC [48].

Interestingly, both capacity regions are exactly the same (duality property) **as soon as the power constraint is set on the total transmitted power**.

The boundary of the global capacity region can be traced out by means of a set of relative priority coefficients $\sum_k x_k = 1$, which control practical aspects such as fairness and priority aspects that appear in multiuser systems when sharing resources. Since the MAC capacity region is convex, it is well known from convex theory that the boundary of the capacity region can be fully characterized by maximizing the function $x_1 R_1 + ... + x_K R_K$ over all rate vectors in the capacity region and for all nonnegative priorities. Each boundary point of the capacity region maximizes the linear combination of the user rates $R_x = \sum_k x_k R_k$. The maximum aggregate rate is known to be

$$R_x = \sum_{k=1}^{Ntot} x_k \log_2\left(1 + SNIR_k\right)$$

For the BC capacity region, the same applies by duality.

For a fixed set of priorities, this is equivalent to finding the point on the capacity region boundary that is tangent to a line whose slope is defined by the priorities. The structure of the MAC capacity region implies that all boundary points of the capacity region are corner points of polyhedrons corresponding to different sets of correlation matrices. Furthermore, the corner point should correspond to successive decoding in order of increasing priority, i.e., the user with the highest priority should be decoded last and, therefore, sees no interference. Thus, the problem of finding the boundary point on the capacity region associated with priorities $x_1..x_K$ assumed to be in descending order.

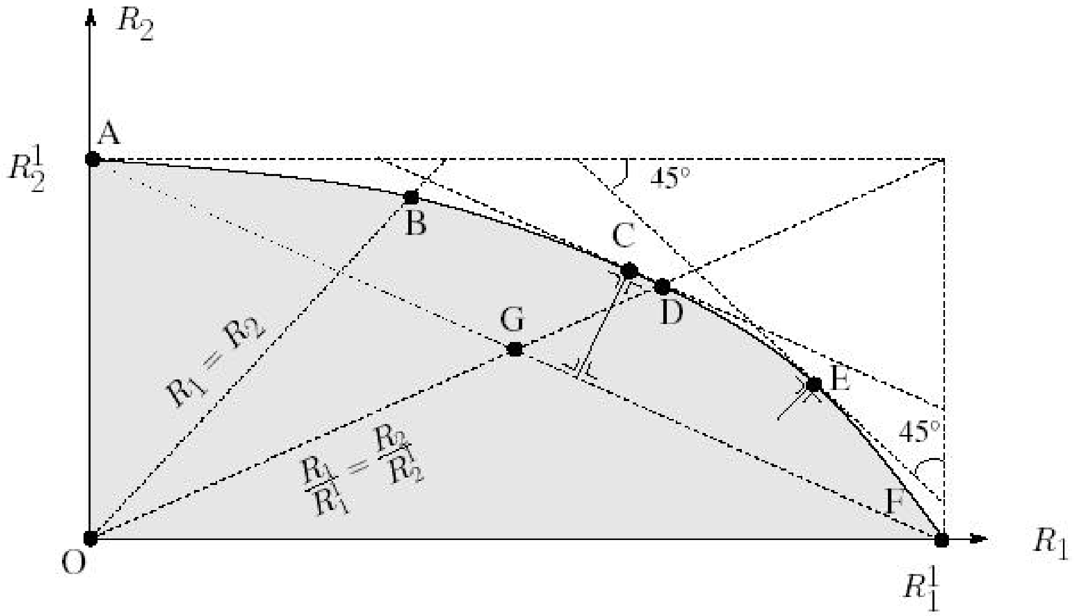


Figure: Two-user capacity region and specific points of the boundary

In the next figure, the boundary of the capacity region is the curve ABCDEF. The extreme points F and A, on this boundary, correspond to the single-user capacities R_{11} and R_{12} of users 1 and 2, respectively. Point E, with a local tangent at 45 degrees, gives the maximum sum-rate max (R1 + R2). This setting generally results in unfair situations where the users with the best channels have a much higher rate than the others, which is not desirable in practical applications. Point B, on the other hand, gives the maximum common rate or symmetric capacity. When the single-user rates are very different, the common rate constraint is generally a waste of resources as it forces the users with the best channels to lower their rate dramatically to reach the level of the weakest channels. The balanced capacity, given by point D, satisfies the relation $R1=R_{11} = R2=R_{12}$. It appears as a smart compromise between the symmetric capacity B and the maximum sum-rate E. The line AGF represents the rate distributions obtained by using Time-Division Multiple-Access (TDMA). Balanced rates $0.5R_{11}$ and $0.5R_{12}$ are obtained if time slots of equal duration are allocated to each user (point G). Higher balanced rates (point D) can be achieved by allowing a simultaneous transmission of signals by all users, with an appropriate power an spectrum allocation. In any case, the maximum balanced rates can be written

$$R_k = g \frac{R_k^1}{Ntot}$$

Additional requirements in terms of minimum throughput should be considered for some applications, where customers could pay for a minimum guaranteed service (e.g: a video connection), plus a best-effort service (e.g: Internet connection) with a variable rate that depends on network conditions. The balanced capacity criterion could then be applied on the variable rate only.

IX.3.7. The fading channel

With full CSI, both the base-station and the users track the channel fluctuations and, in this case, the extension of the linear beamforming strategies combined with Costa precoding to the fading channel is natural. Now we can vary the power and transmit signature allocations of the users, and the Costa precoding order as a function of the channel variations. Linear beamforming combined with Costa precoding achieves the capacity of the fast fading downlink channel with full CSI, just as in the time-invariant downlink channel.

Due to the duality, we have a connection between the strategies for the downlink channel and its dual uplink channel.

Next, aspects related with the practical implementation, as partial CSIT and real-time schedulers as fairness and access control are considered. Fairness comes to the scene when optimality has to be trade-off with users' priorities. Concerning access control, note that up to now we have not addressed the problem of real-time access and how users can be ordered in practice when partial CSIT is only available, thus dirty paper implementation is not realistic.

As we have already said, in the design of a scheduler or practical broadcast system 3 questions have to be answered:

- Transmitter architecture
- Power allocation
- Access control (i.e. number of users to give access and order)

IX.3.8. Sum capacity with partial CSIT

When studying the sum rate in the BC channel we obtained that when the number of users goes to infinite and only nt antennas are at the transmitter, the dirty paper precoding behaves as

$$E\left\{R^{DP}\right\}\Big|_{Ntot\to\infty} \approx nt \log\log r\, Ntot$$

Thus offering not only multiplexing diversity but also multiuser diversity.

Another comment on the asymptotic behaviour if Ntot goes to infinite is that the optimal transmitter structure that maximizes the sum capacity in multiuser MIMO is then the beamforming [46].

However, having full CSIT requires a lot of feedback and practically it is unrealistic. This motivates the question of how much partial side information is needed in the transmitter that provides us a linear scaling of the throughput with nt and reduces the amount of feedback. If we resort to the asymptotic analysis for Ntot going to infinite, it would be desirable that the new scheme would also take advantage of the multiuser diversity. This is precisely the basis of the so-called opportunistic schemes, which just need SNIR feedback instead of the whole knowledge of **H**.

Multiuser diversity and opportunistic transmission

R. Knopp obtained in [41] that the power control scheme that maximizes the information ergodic capacity of the MAC in a SISO multiuser communication, is such that only the user with best channel transmits at a time. Note that for a scalar BC channel the same result applies. In the conventional TDMA scheme, the base station transmits to only a single user at a time. In this case, the maximum sum-rate, achieved by sending to the user with the largest channel gain, is given by

$$R_{TDMA} = \max_{k\in\{1\ldots Ntot\}} \log\left(1+P\left|\mathbf{h}_k\right|^2\right)$$

and if Ntot goes to infinite, we observe that the sum-rate increases double-logarithmically in Ntot; thus, taking advantage of the Multiuser diversity. The multiuser diversity effect comes from the fact that when there are many users that fade independently, at any one time there is a high probability that one of the users will have a strong channel. By allowing only that user to transmit, the shared channel resource is used in the most efficient manner and the total system throughput is maximized (see fogire). The larger the number of users, the stronger tends to be the best channel, and the more the multiuser diversity gain.

The work by Knopp set the basis of the so-called opportunistic schemes: the sender opportunistically transmits only when its channel is near its peaks (“riding the peaks”). These schemes are the only implementable schedulers or BC schemes in the actual systems. Therefore, one section is devoted to them, which will comment on the main system issues when implementing them. Opportunistic schemes are the best proof of the benefit of letting PHY and MAC layer interact to face the problems of the mobile channel.

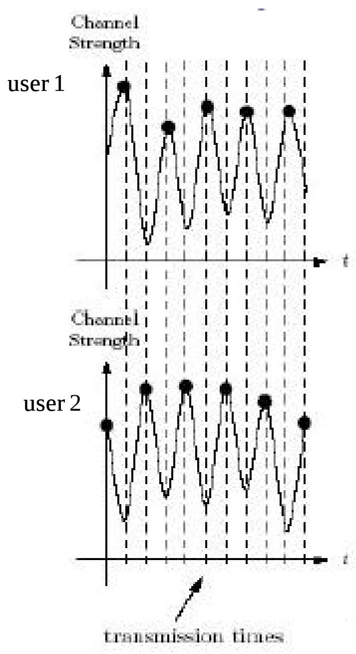


Figure: Opportunistic principle for 2 users

The question we pose now is if we can extend the opportunistic waterfilling to MIMO BC channels? Can the MIMO BC channel enjoy of both Multiuser diversity and multiplexing gain? The answer is given by Hassibi in the so-called opportunistic orthogonal SDMA, that extends the opportunistic results of SIMO to the MIMO case.

Opportunisitc SDMA and single antenna receivers

The conceptual idea is to have multiple beams, each orthogonal to one another, at the same time. Separate pilot symbols are introduced on each of the beams and the users feedback the SINR of each beam. Transmissions are scheduled to as many users as there are beams at each time slot. If there are enough users in the system, the user who is beamformed with respect to a specific beam (and orthogonal to the other beams) is scheduled on the specific beam.

Let us consider

$$\mathbf{y}_{Nx1} = \mathbf{A}^{H}_{N\,x(nrxN)}\mathbf{H}_{(nrxN)xnt}\mathbf{x}_{ntx1} + \mathbf{w}_{Nx1}$$

Let us consider Ntot>=nt, otherwise, we use only Ntot of the transmit antennas. Under this considerations we specify N as the number of served users. Note that usually, the number of served users is nt, smaller than Ntot, thus requiring an access control mechanism, which is provided in a natural and low complexity way by the opportunistic scheme that is going to be introduced. On the other hand, although in the DFE precoders Ntot users can be served simultaneously the user ordering stage implies a lot of computational effort.

At each time m, let $\mathbf{B}(m) = [\mathbf{b}_1(m),...,\mathbf{b}_{nt}(m)]$ be an nt x nt beamforming unitary matrix, with the columns or beams orthogonal. The vector signal sent out from the antenna array at time m is

$$\mathbf{x}_{ntxN} = \mathbf{B}(m)\mathbf{u} = \sum_{m=1}^{nt} \mathbf{b}_m(m)u_m$$

The unitary matrix **B**(m) is varied such that the individual components do not change abruptly in time. For simplicity we consider the scenario when the channel coefficients are not varying over the time-scale of communication (slow fading)

The ith rx knows (**H**i **b**m) m=1..nt (by training). Therefore, the ith rx can compute the following nt SINRs by assuming that the um is the desired signal and the other signals are interference as follows

$$SINR_{i,m} = \frac{\left|\mathbf{H_i b_m}\ u_{im}\right|^2}{1/SNR + \sum_{u \neq m} \left|\mathbf{H_i b_m}\ u_{im}\right|^2} \quad m = 1..nt$$

Note that on average the SINRs behave like

$$SINR_{i,n} \approx \frac{1}{\frac{1}{SNR} + (nt-1)} \approx \frac{1}{nt-1}$$

Therefore if the beams are assigned randomly, the rate or throughput will be

$$R = E\left\{\sum_{i=1}^{nt} \log\left(1 + SINR_{i,m}\right)\right\} \leq nt \log\left(1 + \frac{1}{nt-1}\right) \approx 1$$

Observe that there is no nt-fold in the system throughput and, therefore, CSIT is crucial. As an alternative Hassibi presented an scheme where nt orthogonal beams are assigned to nt users depending on the feedback SINR's. In that case the same asymptotic multiplexing and diversity gain as dirty paper is obtained when Ntot goes to infinite.

$$R = E\left\{\sum_{i=1}^{nt} \log\left(1 + \max_{1 \leq i \leq Ntot} SINR_{i,m}\right)\right\} \approx ntE\left\{\log\left(1 + \max_{1 \leq i \leq Ntot} SINR_{i,m}\right)\right\}$$

Intuitively, if the number of users is large the probability of finding nt users placed at the pointing directions of the nt orthogonal beams is high, thus almost nulling the interference among them

$$ntE\left\{\log\left(1 + \max_{1 \leq i \leq Ntot} SINR_{i,m}\right)\right\} \approx ntE\left\{\log\left(1 + P \max_{1 \leq i \leq Ntot} Tr(\mathbf{h_k h_k^H})\right)\right\}$$

Next figure compares the result of the opportunistic SMDA scheme with those of an opportunistic beamforming that serves one user at a time. Note that the opportunistic SDMA does not obtain higher throughput than the single beamformer if the number of users is small, this is due to the interference cased by the simultaneous nt beams.

There are some system requirements to support multiple beams. First, multiple pilot symbols have to be inserted (one for each beam) to enable coherent downlink reception;

thus, the fraction of pilot symbol power increases. Second, the receivers now track nt separate beams and feedback SINR of each on each of the beams. On a practical note, the receivers could feedback only the best SINR and the identification of the beam that yields this SINR; this restriction probably will not degrade the performance by much. Thus, with affordable feedback the proposed opportunistic SDMA utilizes all the spatial degrees of freedom.

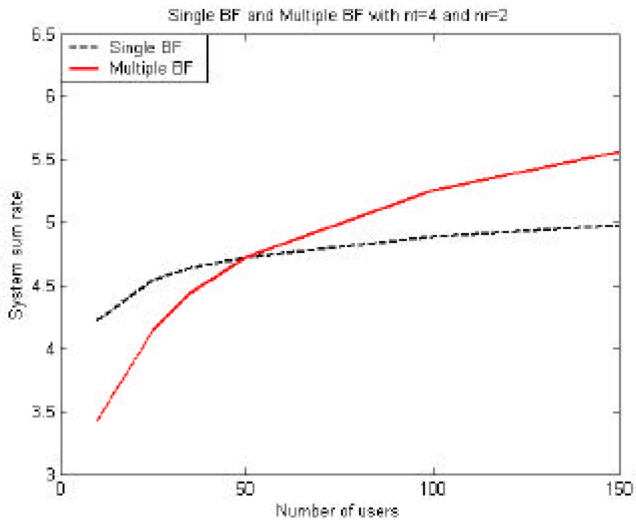


Figure: Opportunistic SDMA vs. opportunistic single beam.

The amount of multiuser diversity gain depends crucially on the tail of the fading distribution SNIRk: the heavier the tail, the more likely there is a user with a very strong channel, and the larger the multiuser diversity gain. For instance, because of the line-of-sight component, the Rician fading distribution is less “random” and has a lighter tail that the Rayleigh distribution with the same average channel gain. As a consequence, it can be seen that the multiuser diversity gain is significantly smaller in the Rician case compared to the Rayleigh case.

Finally, the opportunistic schemes that chose the best transmitting antenna would be optimal in those situations where maximal ratio combining transmission is optimal (e.g.if users are located in orthogonal positions) and for each user there is a predominant antenna in the equivalent channel response. Otherwise, these antenna selection schemes are suboptimal in multiuser scenario. However, due to its practicality and also due to the fact that they can help to diminish the hardening effect, they are an interesting option. In point to point MIMO antenna selection arise as optimal also when there is partial channel state information at the transmitter and rate is to be optimized. The same might happen in the multiuser case and it is a topic to be further researched.

Partial CSIT and Multi-antenna transmitters and receivers

In particular, we can ask what impact multiple receive antennas have on multiuser diversity, an important outcome. In general, any deterministic component that would be introduced in the system (e.g. receiving beamforming in a BC channel) reduces the multiuser gain and therefore, the benefit of using opportunistic schemes (the so-called hardening effect of the fading distribution).

With multiple transmit antennas at the base-station and multiple receive antennas at each of the users, with full CSI we split the information for user k into independent data streams, modulates them on different spatial signatures and then transmit them. The spatial signatures and power allocation to the users (and the further allocation among the data streams within a user) can be done as a function of the channel fluctuations. Linear strategies can be carried out or, if computational complexity is not a problem, Costa precoding (i.e., dirty paper or DFE precoding) can be incorporated.

Without CSI (i.e., only CSIR) the transmitter has no access to the channel fluctuations. One of the important conclusions is that time sharing among the users achieves the capacity region in the symmetric BC channel with CSIR alone. Note that since the statistics of the user channels are identical, if user k can decode its data reliably, then all the other users can also successfully decode user k's data, concluding that the sum of the rates at which the users are being simultaneously reliably transmitted is bounded by

$$\sum_{i=1}^{Ntot} R_i \le E\left[\log\left(1+\frac{SNR|\mathbf{h}|^2}{nt}\right)\right]$$

This implies that the total spatial degrees of freedom in the BC channel are restricted to one instead of the min (nt, Ntot) that can be reached with full CSIT. Thus lack of CSI at the base station causes a drastic reduction in the degrees of freedom of the channel.

With partial CSIT opportunistic schemes are an attractive alternative but it would be desirable to obtain intermediate solutions between those schemes and the optimal DP. To that purpose we recall the superuser transmitter formulation of precoding matrix **B**

$$\mathbf{B}_{ntxN} = \mathbf{V}_{ntxN}\sqrt{\mathbf{S}}_{ntxN}\mathbf{M}_{NxNtot}$$

Where we extend the design of each of the matrixes as follows: **V** is the beamforming matrix, build up with no channel state information. It can contain either nt orthogonal beams; thus, N=nt, or more than nt beams. In this case, Grassmanian manifold can be used to optimized the design of the quasiorthogonal beams. The role of more beams than antennas is when the low number of antennas hinders capacity and transmission to more users than antennas increase capacity if thanks to multiantennas at reception allow for interference cancellation at the receiver.

Matrix S accounts for the power allocation, which is a topic still unaddressed in opportunistic SDMA. Finally matrix **M** controls the selection of users and the association of each user with a transmitting beam. For instance, in an opportunistic system **M** will chose those users whose spatial channel is as much orthogonal as possible, thus achieving null interference.

Next figure plots a MIMO BC system with nt transmit antennas and Ntot>= nt users each with Nk receive antennas. The plot shows the proposed general scheduler.

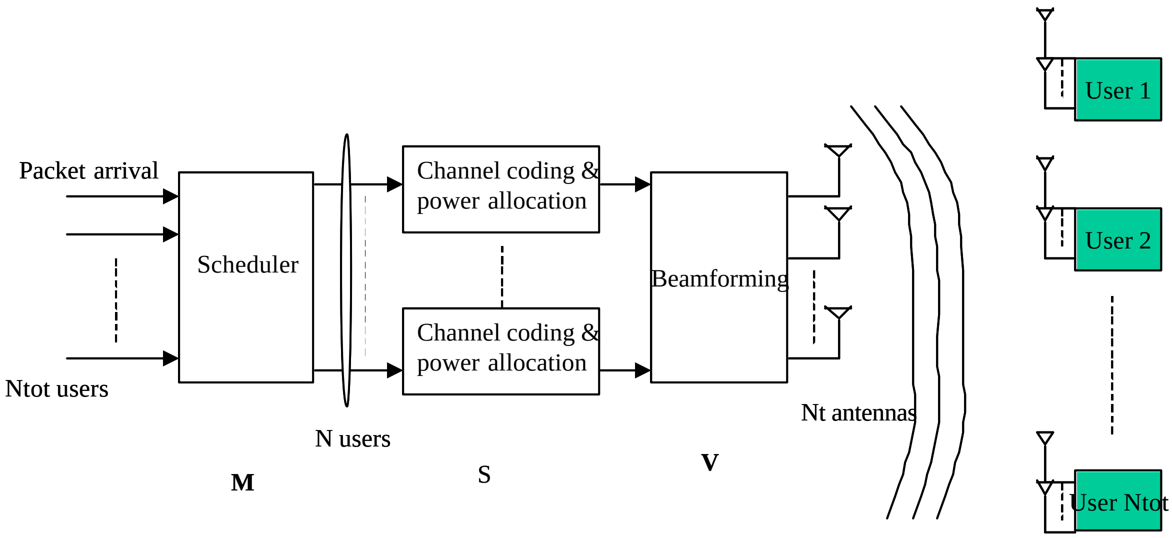


Figure: MIMO BC scheduler

IX.5. CONCLUSIONS

While the wealth of references and results gives the impression that the problem of MIMO multiuser is by now mature from both the information theory and communications aspects. We believe that such an interpretation is misleading. In fact the many unsolved problems, some strongly motivated by practical applications and implications remain and some, such as the role of the CSI at the transmitted and optimal MIMO BC strategies in the presence of partial CSI , are explicitly mentioned in the sequel. We believe that this topic still calls for intensive research addressing many fundamental problems, which are not yet fully understood.

APPENDIX A.
Detailed computation of the optimal precoding in BC

Departing from section IX.3.3.1 and before getting into the design, let us reformulate the MMSE noise correlation matrix $\left(\tilde{\mathbf{H}}^T\tilde{\mathbf{H}}+\mathbf{I}\right)^{-1}$. Let us consider the Cholesky factorization of the MMSE noise matrix

$$\mathbf{G}^{-1}?^{-1}\mathbf{G}^{-T} = \left(\tilde{\mathbf{H}}^T\tilde{\mathbf{H}}+\mathbf{I}\right)^{-1} = \left(\mathbf{M}^T\sqrt{\mathbf{S}}\mathbf{V}^T\mathbf{H}^T\mathbf{Q}^T?^{-1}\mathbf{Q}\mathbf{H}\mathbf{V}\sqrt{\mathbf{S}}\mathbf{M}+\mathbf{I}\right)^{-1}$$

Now, choose a square matrix **C**, such that

$$\mathbf{C}^T\mathbf{C} = \left(\sqrt{\mathbf{S}}\mathbf{V}^T\mathbf{H}^T\mathbf{Q}^T?^{-1}\mathbf{Q}\mathbf{H}\mathbf{V}\sqrt{\mathbf{S}}+\mathbf{I}\right)^{-1}$$

in general **C=UR**, where **U** is an orthonormal matrix and **R** an upper triangular one.

Then the Cholesky factorization can be written as

$$\mathbf{G}^{-1}?^{-1}\mathbf{G}^{-T} = \mathbf{M}^T\mathbf{R}^T\mathbf{U}^T\mathbf{U}\mathbf{R}\mathbf{M}$$

where **URM** is upper triangular.

Now the GDFE is used for the receiver implementation. Although, we show later in section IX.3.3. 2.and IX.3.3. 3. that GDFE has nice optimal features, note that the GDFE error e'i that is obtained when recovering each of the transmitted signals is decoupled, in contrast to the MMSE error. Therefore, we expect that GDFE would help in the decoupled receiver design better than the MMSE.

Recall that the feedforward filter, which we call **F** is

$$\mathbf{F} = ?^{-1}\mathbf{G}^{-T}\tilde{\mathbf{H}}^T?^{-1/2}\mathbf{Q}$$

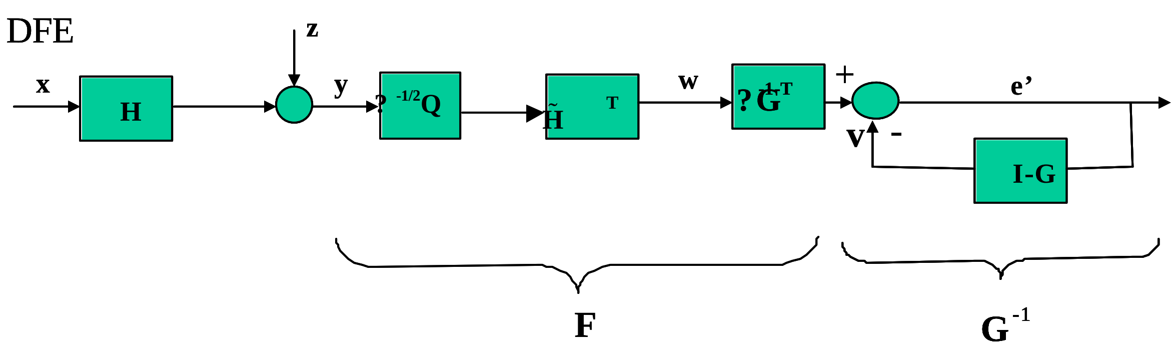


Figure: GDFE reception

Then, it can be further computed as follows

$$\mathbf{F} = ?^{-1/2}\mathbf{URMM}^{T}\sqrt{\mathbf{S}}\mathbf{V}^{T}\mathbf{H}^{T}\mathbf{Q}^{T}?^{-1}\mathbf{Q}$$
$$= ?^{-1/2}\mathbf{UR}\sqrt{\mathbf{S}}\mathbf{V}^{T}\mathbf{H}^{T}\mathbf{Q}^{T}?^{-1}\mathbf{Q}$$

Now a block diagonal structure is imposed on **F**

$$\mathbf{F} = ?^{-1/2}\mathbf{UR}\sqrt{\mathbf{S}}\mathbf{V}^{T}\mathbf{H}^{T}\mathbf{Q}^{T}?^{-1}\mathbf{Q} =$$
$$= ?^{-1/2}\begin{bmatrix} \mathbf{F}_1^{-1/2} & \mathbf{0} \\ \mathbf{0} & \mathbf{F}_1^{-1/2} \end{bmatrix}$$

where **UR** can be obtained. Finally, an appropriate transmit filter $\mathbf{B} = \mathbf{V}\sqrt{\mathbf{S}}\mathbf{M}$ is found by obtaining an **M** that makes **URM** block upper-triangular. This is possible by the following QR-factorization: $\mathbf{R}^{T}\mathbf{U}^{T} = \mathbf{MK}$, where **K** is lower triangular and **M** is orthogonal. Then $\mathbf{URM} = \mathbf{K}^{T}$ is upper-triangular. Note that $\mathbf{R}^{T}\mathbf{R} = \left(\sqrt{\mathbf{S}}\mathbf{V}^{T}\mathbf{H}^{T}\mathbf{Q}^{T}?^{-1}\mathbf{QHV}\sqrt{\mathbf{S}} + \mathbf{I}\right)^{-1}$

There are still the question to answer before obtaining a complete precoding design:

- the design of the block diagonal matrix

<u>Design of the block diagonal matrix</u>

Next we show that the condition under which there exists a suitable **UR** to make the feedforward filter **F** block-diagonal, and therefore suitable for non-cooperative receivers as in the BC channel, is the same as the diagonalization condition on the noise covariance matrix

$$\mathbf{R}_z^{-1} - \left(\mathbf{R}_z + \mathbf{H}\mathbf{R}_x\mathbf{H}^{T}\right)^{-1} = \begin{bmatrix} \mathbf{F}_1^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{F}_2^{-1} \end{bmatrix} \quad (**)$$

where $\mathbf{F}_i^{-1}$ are positive semi-definite matrices and are the dual variables associated with the block-diagonal constraints of the max-min problem.

In order to get some insight in the meaning of constraint (**), note that this conditions is equivalent to $\left(\mathbf{I} + \tilde{\mathbf{H}}\mathbf{R}_x\tilde{\mathbf{H}}^{T}\right)^{-1} = \begin{bmatrix} \mathbf{F}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{F}_2 \end{bmatrix}$, thus imposing block diagonal structure on the correlation matrix of the received signal after the whitening filter.

The condition under which there exists a suitable **UR** to make the feedforward filter **F** block-diagonal is the same as the diagonalization condition on the noise covariance matrix

$$\begin{bmatrix} \mathbf{F}_1^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{F}_2^{-1} \end{bmatrix} = \mathbf{R}_z^{-1} - \left(\mathbf{R}_z + \mathbf{H}\mathbf{R}_x\mathbf{H}^T\right)^{-1} = \mathbf{Q}^T ?^{-1}\mathbf{Q} - \left(\mathbf{Q}^T ? \mathbf{Q} + \mathbf{H}\mathbf{V}\mathbf{S}\mathbf{V}^T\mathbf{H}^T\right)^{-1} =$$

$$= \mathbf{Q}^T ?^{-1/2}\left(\mathbf{I} - \left(\mathbf{I} + ?^{-1/2}\mathbf{QHVSV}^T\mathbf{H}^T\mathbf{Q}^T ?^{-1/2}\right)^{-1}\right) ?^{-1/2}\mathbf{Q}^T =$$

$$= \mathbf{Q}^T ?^{-1}\mathbf{QHV}\sqrt{\mathbf{S}}\left(\mathbf{I} + \sqrt{\mathbf{S}}\mathbf{V}^T\mathbf{H}^T\mathbf{Q}^T ?^{-1}\mathbf{QHV}\sqrt{\mathbf{S}}\right)^{-1}\sqrt{\mathbf{S}}\mathbf{V}^T\mathbf{H}^T\mathbf{Q}^T ?^{-1}\mathbf{Q}$$

As $\mathbf{R}^T\mathbf{R} = \left(\sqrt{\mathbf{S}}\mathbf{V}^T\mathbf{H}^T\mathbf{Q}^T ?^{-1}\mathbf{QHV}\sqrt{\mathbf{S}} + \mathbf{I}\right)^{-1}$

Then

$$\mathbf{Q}^T ?^{-1}\mathbf{QHV}\sqrt{\mathbf{S}}\mathbf{R}^T\mathbf{R}\sqrt{\mathbf{S}}\mathbf{V}^T\mathbf{H}^T\mathbf{Q}^T ?^{-1}\mathbf{Q} = \begin{bmatrix} \mathbf{F}_1^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{F}_2^{-1} \end{bmatrix}$$

$$\mathbf{R}\sqrt{\mathbf{S}}\mathbf{V}^T\mathbf{H}^T\mathbf{Q}^T ?^{-1}\mathbf{Q} = \mathbf{U}\begin{bmatrix} \mathbf{F}_1^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{F}_2^{-1} \end{bmatrix}$$

Therefore,

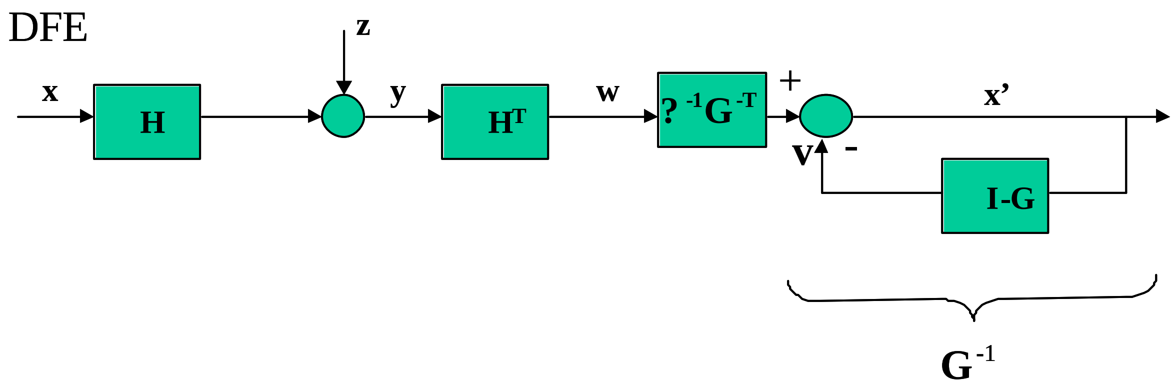
$$\mathbf{F} = ?^{-1/2}\mathbf{UR}\sqrt{\mathbf{S}}\mathbf{V}^T\mathbf{H}^T\mathbf{Q}^T ?^{-1}\mathbf{Q} =$$

$$= ?^{-1/2}\begin{bmatrix} \mathbf{F}_1^{-1/2} & \mathbf{0} \\ \mathbf{0} & \mathbf{F}_1^{-1/2} \end{bmatrix}$$

APPENDIX B

Decision Feedback precoder is capacity lossless

Recalling in the figure the DFE receiver structure that has been presented in the MAC section. Next we show that the feedforward filter can be implemented in the transmitter (if CSIT is available) without losing capacity with respect to the DFE.



Let us depart from the MAC channel model

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{z} = [\mathbf{H}_1\ \mathbf{H}_2]\begin{bmatrix}\mathbf{x}_1\\ \mathbf{x}_2\end{bmatrix} + \begin{bmatrix}\mathbf{z}_1\\ \mathbf{z}_2\end{bmatrix} \quad (*)$$

and consider the output of the feedforward filter, **v=[v1 v2]^T** . The goal is to compute the achievable rates of the two subchannels: form **x1** to **v1** and from **x2** to **v2**. If they are the same as the rates from **x1** to **x1'** and from **x2** to **x2'**, which are the rates of the lossless MMSE, no capacity loss is incurred by implementing a DFE precoder. In other words, we are interested in proving that:

$$R_1 = I(X_1;V_1) \overset{?}{=} I(X_1;X'_1)$$
$$R_2 = I(X_2;V_2) \overset{?}{=} I(X_2;X'_2)$$

To prove it note that

$$\mathbf{v} = ?^{-1}\mathbf{G}^{-T}\mathbf{H}^T\mathbf{H}\mathbf{x} + ?^{-1}\mathbf{G}^{-T}\mathbf{H}^T\mathbf{z} = ?^{-1}\mathbf{G}^{-T}\mathbf{H}^T(\mathbf{H}\mathbf{x}+\mathbf{z})$$

If **x** = **x'**+**e'** then $\mathbf{x}' = \mathbf{v} + (\mathbf{I}-\mathbf{G})\mathbf{x}$

Note that $\mathbf{x}'_2 = \mathbf{v}_2$ thus

$$R_2 = I(X_2;V_2) = I(X_2;X'_2) = I(X_2;Y)$$

Now, consider the subchannel from **x**1 to **v**1 with **x**2 available at the transmitter instead of at the receiver $I(X_1;V_1/X_2)$

By substituting for the transmission model in (*) as it was done for the MAC channel, we obtain

$$\mathbf{x}_1' = ?_{11}^{-1}\mathbf{w}_1 - \mathbf{G}_{22}\mathbf{x}_1$$

$$\mathbf{v}_1 = ?_{11}^{-1}\mathbf{w}_1 + \mathbf{e}_1' = \left(\mathbf{R}_1^{-1} + \mathbf{H}_1^T\mathbf{H}_1\right)^{-1}\mathbf{H}_1^T\left(\mathbf{H}_1\mathbf{x}_1 + \mathbf{H}_2\mathbf{x}_2 + \mathbf{z}_1\right)$$

being **w** the output to the matched filter to the channel

$$\mathbf{w} = \mathbf{H}^T\mathbf{H}\mathbf{x} = \begin{bmatrix} \mathbf{H}_1^T\mathbf{H}_1\mathbf{x}_1 + \mathbf{H}_1^T\mathbf{H}_2\mathbf{x}_2 \\ \mathbf{H}_2^T\mathbf{H}_1\mathbf{x}_1 + \mathbf{H}_2^T\mathbf{H}_2\mathbf{x}_2 \end{bmatrix}$$

and we obtain

$$\mathbf{x}_1' = \left(\mathbf{R}_1^{-1} + \mathbf{H}_1^T\mathbf{H}_1\right)^{-1}\mathbf{H}_1^T(\mathbf{H}_1^T\mathbf{x}_1 + \mathbf{z}_1)$$

Observe that $\mathbf{v_1=x_1'+H_2\,x_2}$ as $\mathbf{x_1'}$ and $\mathbf{x_2}$ are independent (since, x1, x2 and z1 are jointly independent)

$$R_1 = I(X_1;V_1/X_2) = I(X_1;X_1') = I(X_1;Y/\ X_2)$$

where, the last equality was shown in the MAC channel section and proves that a precoder designed in order to obtain $R_2=I(X_2;Y)$ and $R_1=I(X_1;Y/\ X_2)$ achieves the same capacity as a DFE equalizer: $R_2=I(X_2;X_2')$ and $R_1=I(X_1;X_1')$

Therefore,

$$I(X_1\ X_2;Y) = I(X_2;Y) + I(X_1;Y/\ X_2) = I(X_2;\ X_2') + I(X_1;\ X_1')$$

or, depending on the precoding order

$$I(X_1\ X_2;Y) = I(X_1;Y) + I(X_2;Y/\ X_1) = I(X_1;\ X_1') + I(X_2;\ X_2')$$

Interference cancellation may occur at the transmitter by pre-subtracting x2 form x1. Pre-subtraction achieves the exact same capacity as a decision feedback equalizer. In conclusion, the DFE precoding as shown in the next figure is capacity lossless.

APPENDIX C

Uplink-downlink SINR duality

Restricting to linear beamforming strategies for the downlink and to linear detection strategies for the uplink, let **F** denote the linear transmitter matrix and $\mathbf{F}^H$ denote the linear receiver matrix, respectively. Without loss of generality, we can assume the normalization $\left[\mathbf{F}\mathbf{F}^H\right]_{jj}=1$ for all j. With this normalization, the input power constraint is given simply by $\sum_k P_k^{bc} \le P$. The SINR for user i in the BC channel is

$$SINR_{bc,i} = \frac{P_i^{bc} f_{ii}}{1+\sum_{j\neq i} P_j^{bc} f_{ij}} \qquad f_{ij} = \left|\mathbf{HF}\right|_{ij}^2$$

In the dual MAC, the output of the linear detector of user i is given by the i-th element of the vector $\mathbf{F}^H\mathbf{y}_{MAC}$. The SINR for user i is given by

$$SINR_{mac,i} = \frac{P_i^{mac} f_{ii}}{1+\sum_{j\neq i} P_j^{mac} f_{ij}}$$

Suppose that target SINRs γ1…γm are required in both the BC and its dual MAC. The system of equations $SINR_i \ge g_i$ can be written in compact matrix form as follows.

Let $\mathbf{a} = \left[a_1 \cdots a_1\right]^T \qquad a_i = \frac{g_i}{(1+g_i)f_{ii}}$

Then the SINR equations for the BC take on the form

$$\left[\mathbf{I} - diag(\mathbf{a})\mathbf{F}\right]\mathbf{p}^{bc} \ge \mathbf{a}$$

For the MAC take on the form

$$\left[\mathbf{I} - diag(\mathbf{a})\mathbf{F}^T\right]\mathbf{p}^{MAC} \ge \mathbf{a}$$

The SINR vector g is feasible for both BC and MAC with linear processing matrix **F** if and only if the non-negative matrix diag(**a**)F has Perron-Frobenious eigenvalue r(diag(**a**)F)<1. In this case, the solutions

$$\mathbf{p}_{opt}^{bc} = \left[\mathbf{I} - diag(\mathbf{a})\mathbf{F}\right]^{-1}\mathbf{a}$$

and

$$\mathbf{p}_{opt}^{mac} = \left[\mathbf{I} - diag(\mathbf{a})\mathbf{F}^T\right]^{-1}\mathbf{a}$$

of the BC and MAC power allocation equations are the componentwise minimal power allocation that meets the g with equality. Moreover $\sum_i p_{opt,i}^{bc} = \sum_i p_{opt,i}^{mac}$

AS an immediate corollary we get that, a SINR vector g is feasible if and only if

$$r(diag(\mathbf{a})\mathbf{F}) < 1$$
and
$$\sum_i p_{opt\,i}^{bc} = \mathbf{1}^T \left[\mathbf{I} - diag(\mathbf{a})\mathbf{F}\right]^{-1} \mathbf{a} \le P$$

REFERENCES

[1] T. M. Cover, "Broadcast channels," IEEE Trans. Inform. Theory, vol. IT-18, pp. 2–14, Jan. 1972.
[2] R. G. Gallager, "Coding for degraded broadcast channels," Probl. Inform. Transm., vol. X, pp. 3–14, Sept. 1974.
[3] P. Bergman, "A simple converse for broadcast channels with additive white Gaussian noise," IEEE Trans. Inform. Theory, vol. IT-20, pp. 279–280, Mar. 1974.
[4] K. Marton, "A coding theorem for the discrete memoryless broadcast channel," IEEE Trans. Inform. Theory, vol. IT-25, pp. 306–311, May 1979.
[5] A. El Gamal and E. C. van der Meulen, "A proof of Marton's coding theorem for the discrete memoryless broadcast channel," IEEE Trans. Inform. Theory, vol. IT-27, pp. 120–122, Jan. 1981.
[6] T. M. Cover, "Comments on broadcast channels," IEEE Trans. Inform. Theory, vol. 44, pp. 2524–2530, Oct. 1998.
[7] G. Caire and S. Shamai (Shitz), "On the achievable throughput of a multi-antenna Gaussain broadcast channel," IEEE Trans. Inform. Theory, vol. 49, pp. 1691–1706, July 2003, submitted for publication.
[8] M. Costa, "Writing on dirty paper," IEEE Trans. Inform. Theory, vol. IT-29, pp. 439–441, May 1983.
[9] H. Sato, "An outer bound on the capacity region of broadcast channels," IEEE Trans. Inform. Theory, vol. IT-24, pp. 374–377, May 1978.
[10] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, achievable rates and sum-rate capacity of Gaussian MIMO broadcast channels," IEEE Trans. Inform. Theory, vol. 49, no. 10, pp. 2658–2668, Oct. 2003.
[11] P. Viswanath and D. N. C. Tse, "Sum capacity of the vector Gaussian broadcast channel and uplink-downlink duality," IEEE Trans. Inform. Theory, vol. 49, pp. 1912–1921, Aug. 2003.
[12] R. Zamir, S. Shamai (Shitz), and U. Erez, "Nested linear/lattice codes for structured multiterminal binning," IEEE Trans. Inform. Theory, vol. 48, pp. 1250–1276, June 2002.
[13] W. Yu and J. M. Cioffi, "Trellis precoding for the broadcast channel," in Proc. IEEE GLOBECOM, Sna Antonio, TX, 2001.
[14] S. Vishwanath, G. Kramer, S. Shamai, S. Jafar, and A. Goldsmith, "Capacity bounds for Gaussian vector broadcast channels," in Proc. DIMACS Workshop on Signal Processing for Wireless Transmission, Pisactaway, NJ, Oct. 2002.
[15] D. N. C. Tse and P. Viswanath, "On the capacity of the multiple antenna broadcast channel," in Proc. DIMACS Workshop on Signal Processing for Wireless Transmission, Piscataway, NJ, Oct. 2002.
[16] T. M. Cover and J. A. Thomas, Elements of Information Theory. New

York: Wiley, 1991.
[17] S. I. Gel'fand and M. S. Pinsker, "Coding for channel with random parameters," Probl. Control Inform. Theory, vol. 9, no. 1, pp. 19–31, 1980.
[18] C. Heegard and A. El Gamal, "On the capacity of computer memories with defects," IEEE Trans. Inform. Theory, vol. IT-29, pp. 731–739, Sept. 1983.
[19] A. J. Goldsmith and M. Effros, "The capacity region of broadcast channels with intersymbol interference and colored Gaussian noise," IEEE Trans. Inform. Theory, vol. 47, pp. 219–240, Jan. 2001.
[20] A. Cohen and A. Lapidoth, "The Gaussian watermarking game: Part I," IEEE Trans. Inform. Theory, vol. 48, pp. 1639–1667, June 2002.
[21] W. Yu, A. Sutivong, D. Julian, T. M. Cover, and M. Chiang, "Writing on colored paper," in Proc. IEEE Int. Symp. Information Theory, Washington, DC, June 2001, p. 302.
[22] G. Ginis and J. M. Cioffi, "Vectored transmission for digital subscriber line systems," IEEE J. Select. Areas Comm., vol. 20, pp. 1086–1104, June 2002.
[23] J. M. Cioffi and G. D. Forney, "Generalized decision-feedback equalization for packet transmission with ISI and Gaussian noise," in Communications, Computation, Control and Signal Processing: A Tribute to Thomas Kailath, A. Paulraj, V. Roychowdhury, and C. D. Shaper, Eds. Norwell, MA: Kluwer, 1997.
[24] J. M. Cioffi, G. P. Dudevoir,M. V. Eyuboglu, and G. D. Forney, "MMSE decision feedback equalizers and coding: Part I and II," IEEE Trans. Commun., vol. 43, pp. 2582–2604, Oct. 1995.
[25] M. K. Varanasi and T. Guess, "Optimum decision feedback multiuser equalization with successive decoding achieves the total capacity of the Guassian multiple-access channel," in Proc. Asilomar Conf. Signal System Computers, 1997, pp. 1405–1409.
[26] T. Kailath, A. Sayed, and B. Hassibi, State-Space Estimation. Englewood Cliffs, NJ: Prentice Hall, 1999.
[27] S. Ihara, "On the capacity of channels with additive non-Gaussian noise," Inform. Contr., vol. 37, pp. 34–39, 1978.
[28] S. N. Diggavi and T. M. Cover, "Worst additive noise under covariance constraints," IEEE Trans. Inform. Theory, vol. 47, pp. 3072–3081, Nov. 2001.
[29] R. A. Horn and C. R. Johnson, Matrix Analysis. Cambridge, U.K.: Cambridge Univ. Press, 1990.
[30] K. Fan, "Minimax theorems," Proc. Nat. Acad. Sci., vol. 39, pp. 42–47, 1953.
[31] W. Yu, J.M.Cioffi, "Sum capacity of Gaussian Vector Broadcast channels," IEEE Trans. IT, vol. 50, no.9, September 2004
[32] R. T. Rockafellar, Convex Analysis. Princeton, NJ: Princeton Univ. Press, 1970.
[33] , "Saddle-points and convex analysis," in Differential Games and Related Topics, H. W. Kuhn and G. P. Szegö, Eds. Amsterdam, The Netherlands: North-Holland, 1971.
[34] S. Zakovic and C. Pantelides, "An interior point algorithm for computing saddle points of constrained continuous minimax," Ann. Oper. Res., vol. 99, pp. 59–77, 2000.

[35] Y. Nesterov and A. Nemirovskii, Interior-Point Polynomial Algorithms in Convex Programming. Philadelphai, PA: SIAM, 1994.
[36] S. Kasturia, J. Aslanis, and J. M. Cioffi, “Vector coding for partial-response channels,” IEEE Trans. Inform. Theory, vol. 36, pp. 741–62, July 1990.
[37] W. Yu, “The structure of lease-favorable noise in the Gaussian vector broadcast channels,” in Proc. DIMACS Workshop on Network Information Theory, Piscataway, NJ, Mar. 2003.
[38] D. Bartolome, “Fairness analysis of wiureless beamforming schedulers,” PhD at UPC, November 2004.
[39] M.Bengtsson and B. Ottersten, “Optimal and suboptimal transmit beamforeming,” in Handbook of antennas in wireless communications, L.Godara, Ed. CRC Press, 2001.
[40] N.Jindal, S.Vishwanath, A.Goldsmith, “On the duality of Gaussian Multiple-access and broadcast channel,” IEEE Trans. On IT, vol.50, May 2004
[41] R.Knopp, P.A.Humblet, “Inforamtion capacity and power control in single-cell multiuser communications,” Globecom 1995
[42] M.Sharif, B.Hassibi, “On the capacity of MIMO BC with partial side information,” IEEE Trans. On IT, February 2005.
[43] N.Jindal et al., « Sum power iterative water-filling for multi-antenna gaussian broadcast channels, » IEEE Trans. On IT, vol.51, no.4, April 2005
[44] S.Vishwanath, et al., « Duality, achievable rates, and sum-rate capacity of Gaussian MIMO broadcast channels,” IEEE Trans. On Information theory, vol.49, no.10, cotober 2003.
[45] A.Goldsmith et al., “Capacity Limits of MIMO Channels,” IEEE JSAC, vol. 21, no.5, June 2003.
[46] T.Yoo, A.Goldsmith, “on the optimality of multi-antenna broadcast scheduling using Zero-Forcing Beamforming,” to be published.
[47] “Space-time wireless systems: from array processing to MIMO communications,” Chap. On MIMO BC channel: a short overview, G.Caire et al., Edited by H.Bölcskei et al. To be published
[48] “Multiantenna Channels: apacity, coding and signal processing,” DIMACS: Series in discrete mathematics and theoretical computer science, vol. 62. Edited by G.J.Foschini, S.Verdu, American Mathematic Society