OPTIMAL POWER ALLOCATION FOR MSE AND BIT-LOADING IN MIMO SYSTEMS AND THE IMPACT OF CORRELATION

Thomas Haustein and Holger Boche

Heinrich-Hertz-Institut für Nachrichtentechnik Einsteinufer 37, D-10587 Berlin, Germany,
{haustein, boche}@hhi.de

ABSTRACT

In this paper, the optimal power allocation to minimize the mean square error (MSE) in MIMO systems with and without channel knowledge at the transmitter (Tx) is investigated. Furthermore the impact of correlation is discussed. We show that for MIMO systems with no channel knowledge at the Tx correlation will increases the MSE at the receiver . This does not hold in general when channel knowledge is available at the Tx. The optimum transmit solution can then be obtained by solving a MSE minimization problem. Based on this solution we present a bit loading algorithm which matches the optimum power allocation with a set of finite modulation alphabets with the constraint to certain bit error rate requirements.

1. INTRODUCTION

The growing communication market demands always more of the limited resource bandwidth. Transmission systems using a multiple-input multiple-output (MIMO) structure have been shown to achieve a very high spectral efficiency. Therefore much work was already done to increase mainly capacity. Considering practical aspects towards implementation [1] discusses optimal sequences and power control when using linear MMSE multiuser receivers. Optimal sequences for pilot based CDMA and OFDM systems are investigated in [2], [3] and recent work from [4] discusses the effect of correlation on the ergodic and outage capacity without channel state information (CSI) at the transmitter (Tx).

The first part of this paper focuses on the mean square error (MSE) as a measure for the two cases of CSI at the receiver (Rx) only and CSI also available at the Tx. We start with the theory from [1] and show that for the Tx without CSI, correlation will increase the MSE at the Rx. Furthermore we derive the optimal power allocation for the case of the Tx with CSI which minimizes the MSE at the Rx. This solution totally differs from the classical "waterfilling" solution which maximizes the ergodic capacity.

In the second part of this paper we investigate bit-loading with discrete modulation alphabets. We choose the optimal power allocation solution from above as the starting point for practical bit-loading, which we are primarily interested in from an application point of view. The presented bit-loading strategy finds the maximum data throughput under the constraint of a maximum average bit error rate (BER). Similar bit loading problems were already addressed by e.g. [5] or [6] but their approaches start either with a uniform power allocation or achieve a bit loading solution step by step over a rate-and-power optimization. In this paper we transform the rather abstract optimal solution into an applicable recipe-like bit-loading strategy which is easy to implement. After only a few iterations we achieve the closest to optimum bit and power allocation. A simple example may illustrate our approach.

This work was supported in part by the Bundesministerium für Bildung und Forschung (BMBF) under grant 01BU150

2. SYSTEM MODEL

We assume a single user MIMO system with M Tx and N Rx antennas with $M \leq N$ and up to L data streams are transmitted over the MIMO channel H. At the Rx we assume a Minimum Mean Square Error (MMSE) detector at the Rx with perfect CSI. The transmitter has either no CSI (2.1) or perfect CSI (2.2) like the Rx and we always consider a total sum power constraint at the Tx.

The MIMO transmission model in matrix form reads

$$\vec{y} = H \cdot \vec{x} + \vec{n} \tag{1}$$

with $\vec{y}$ the receive vector of length N, $\vec{x}$ the transmitted vector of size M, $\vec{n}$ is the additive Gaussian noise vector of size N.

2.1. CSI only at the Receiver

If no CSI is available at the Tx then uniform power allocation and one data stream per antenna is optimal: $E[\vec{x}\vec{x}^H] = \frac{P}{M} \cdot I_M$ where $E[.]$ means the expectation, $[.]^H$ means Hermitean conjugate, P is the total sum power and I_M is the identity matrix of size $M \times M$.

The data symbol estimate by the linear MMSE receiver is

$$\hat{\vec{x}} = \frac{P}{M} H^* \left[\sigma^2 I_N + \frac{P}{M} H H^*\right]^{-1} \vec{y}$$

with σ^2 the noise variance at the Rx. The covariance matrix K_ε

$$\begin{aligned} K_\varepsilon &= E[(\hat{\vec{x}} - \vec{x})(\hat{\vec{x}} - \vec{x})^H] \\ &= \frac{P}{M} I_M - \frac{P}{M} I_M H^* [\sigma^2 I_N + \frac{P}{M} H H^*]^{-1} H \frac{P}{M} I_M \end{aligned}$$

yields with a normalization

$$\frac{M}{P} K_\varepsilon = I_M - \sqrt{\frac{P}{M}} I_M H^* [\sigma^2 I_N + \frac{P}{M} H H^*]^{-1} H \sqrt{\frac{P}{M}} I_M.$$

trace (K_ε) gives the normalized MSE at the Rx.

$$\text{tr}\left(\frac{M}{P} K_\varepsilon\right) = M - \text{tr}\left([\sigma^2 I_N + \frac{P}{M} H H^*]^{-1} \frac{P}{M} H H^*\right). \tag{2}$$

We consider the singular value decomposition (SVD) of $H = U\Lambda_H^{1/2}V^*$ where U and V are unitary matrices and $\Lambda^{1/2}$ is a diagonal matrix with the square root of the ordered eigenvalues of H on its diagonal. We now decompose HH^*

$$\begin{aligned} HH^* = U\Lambda_H^{1/2}V^*V\Lambda_H^{1/2}U^* &= U\Lambda_H U^* \\ \text{therefore } \sigma^2 I_N + \frac{P}{M}HH^* &= U\left(\sigma^2 I_N + \frac{P}{M}\Lambda_H\right)U^*. \end{aligned}$$

We define $D = \sigma^2 I_N + \frac{P}{M}\Lambda_H$ then

$$[\sigma^2 I_N + \frac{P}{M}HH^*]^{-1} = UD^{-1}U^*. \tag{3}$$

We now apply (3) to substitute the last part in (2)

$$\begin{aligned} & \text{tr }([\sigma^2 I_N + \frac{P}{M}HH^*]^{-1}\frac{P}{M}HH^*) \\ = & \text{ tr }\left(\frac{P}{M}UD^{-1}U^*U\Lambda_H U^*\right) = \text{tr }\left(\frac{P}{M}UD^{-1}\Lambda_H U^*\right) \\ = & \sum_{l=1}^{N}\frac{\frac{P}{M}\lambda_H(l)}{\sigma^2 + \frac{P}{M}\lambda_H(l)} = N - \sigma^2\sum_{l=1}^{N}\frac{1}{\sigma^2 + \frac{P}{M}\lambda_H(l)} \end{aligned}$$

and the normalized MSE of (2) (also see Fig.2) reads like

$$\frac{M}{P}\text{tr }(K_\varepsilon) = M - N + \sigma^2\sum_{l=1}^{N}\frac{1}{\sigma^2 + \frac{P}{M}\lambda_H(l)} = M - N +$$

$$+\sigma^2\left(\sum_{l=1}^{M}\frac{1}{\sigma^2 + \frac{P}{M}\lambda_H(l)} + \sum_{l=M+1}^{N}\frac{1}{\sigma^2}\right) = \sum_{l=1}^{M}\frac{1}{1 + \frac{P}{M}\frac{\lambda_H(l)}{\sigma^2}}.$$

The right term of the MSE in (4) is a Schur-convex function which leads to the following theorem.

Theorem 1: For trace(HH^*) = constant, rising correlation[1] in H increases the normalized MSE at the MMSE receiver.

proof: Let trace$(HH^*) = \sum_{l=1}^{N}\lambda_H(l) \doteq 1$. H_1 and H_2 be two channel matrices and H_1 has more correlation than H_2 which we write $\sum_{l=1}^{m}\lambda_{H_1}(l) \geq \sum_{l=1}^{m}\lambda_{H_2}(l) \quad m = 1, ..., M$. The MSE is of the form MSE$= \sum_{l=1}^{m} f(x)$ with the Schur-convex function $f(x) = \frac{1}{1+x}$. According to theorem C1 from chapter 3 in [7] also MSE$= \sum_{l=1}^{m} f(x)$ is Schur-convex. Therefore always holds

$$\sum_{l=1}^{m}\frac{1}{\sigma^2 + \frac{P}{M}\lambda_{H_1}(l)} \geq \sum_{l=1}^{m}\frac{1}{\sigma^2 + \frac{P}{M}\lambda_{H_2}(l)} \qquad \square \tag{4}$$

2.2. CSI at the Transmitter and the Receiver

Let L be fixed and perfect CSI is available at the Tx, the data symbol vector $\vec{s}\epsilon\mathbb{C}^L$ is preprocessed and then $\vec{x}\epsilon\mathbb{C}^M$ is emitted from the M Tx antennas. $M - L$ data streams are switched off. The transmission scheme is depicted in Fig.1 and $\vec{x} = WD\vec{s}$, where $D = diag(\sqrt{P_1}, ..., \sqrt{P_L})$ is the power allocation matrix and W is a unitary beamforming matrix of size $M \times L$. Now (1) reads

$$\vec{y} = H\vec{x} + \vec{n} = HWD\vec{s} + \vec{n}$$

[1]Correlation is used here in the sense of the distribution of the ordered eigenvalues (EW)[4]. Uncorrelated - best case, when all EW are the same, fully correlated - worst case, when there is only one EW bigger then zero.

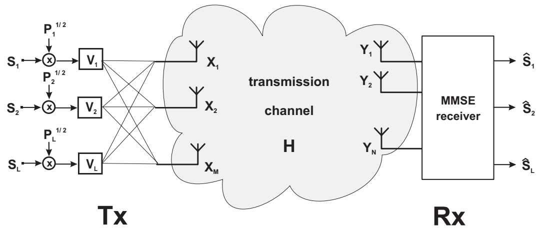

Fig. 1. MIMO transmission setup with channel knowledge at Tx.

We define $\vec{s_p} = D\vec{s}$ then $E[\vec{s}\vec{s}^H] = I_L$ and $E[\vec{s_p}\vec{s_p}^H]$=D with $\sum_{l=1}^{L} P_l \leq P$. The estimated data $\hat{\vec{s}}\epsilon\mathbb{C}^L$ at the MMSE Rx is then

$$\hat{\vec{s}} = DW^*H^*[\sigma^2 I_N + HWDW^*H^*]^{-1}\vec{y}.$$

The covariance matrix K_ε is

$$K_\varepsilon = D - DW^*H^*[\sigma^2 I_N + HWDW^*H^*]^{-1}HWD$$

or in normalized form

$$D^{-\frac{1}{2}}K_\varepsilon D^{-\frac{1}{2}} = I_L - D^{\frac{1}{2}}W^*H^*[\sigma^2 I_N + HWDW^*H^*]^{-1}HWD^{\frac{1}{2}}.$$

With $H = U\Lambda_H^{1/2}V^*$, $V = [V_1, ..., V_M]$ and $W = [V_1, ..., V_L]$ the normalized MSE is given by

$$\text{tr }(D^{-\frac{1}{2}}K_\varepsilon D^{-\frac{1}{2}}) = L - \text{tr }\left([\sigma^2 I_N + HWDW^*H^*]^{-1}HWDW^*H^*\right)$$

$$\begin{aligned} = & \; L - \left(N - \sigma^2\sum_{l=1}^{N}\frac{1}{\sigma^2 + \lambda_H(l)P_l}\right) = L - N + \\ + & \; \sigma^2\left(\sum_{l=1}^{L}\frac{1}{\sigma^2 + \lambda_H(l)P_l} + \sum_{l=L+1}^{N}\frac{1}{\sigma^2}\right) = \sum_{l=1}^{L}\frac{1}{1 + \frac{\lambda_H(l)P_l}{\sigma^2}}. \end{aligned}$$

In order to minimize the sum of the MSE's for all data streams we solve the following minimization problem

$$\min_{\substack{\sum_{l=1}^{L} P_l \leq P \\ P_l \geq 0}} \lim \sum_{l=1}^{L}\frac{1}{1 + \frac{\lambda_H(l)P_l}{\sigma^2}}. \tag{5}$$

We find the Lagrange function $\mathcal{L}(\vec{P}, \mu, \vec{\omega})$

$$\mathcal{L}(\vec{P}, \mu, \vec{\omega}) = \sum_{l=1}^{L}\frac{1}{1 + \frac{\lambda_H(l)P_l}{\sigma^2}} + \mu\left(\sum_{l=1}^{L}P_l - P\right) - \sum_{l=1}^{L}\omega_l P_l \tag{6}$$

where μ is the Lagrange multiplier to satisfy $\sum_{l=1}^{L} P_i \leq P$ and $\vec{\omega}$ guarantees all $P_i \geq 0$. Partial differentiation of (6) gives

$$\frac{\partial\mathcal{L}}{\partial P_r} = -\frac{\frac{\lambda_H(r)}{\sigma^2}}{\left(1 + \frac{\lambda_H(r)P_r}{\sigma^2}\right)^2} + \mu - \omega_r = 0. \tag{7}$$

With a closer look at (5) we see that the sub-channels have different impact on the MSE. We expect a "waterfilling"-like solution, which means that with little sum power, sub-channels corresponding to smaller eigenvalues λ_i are switched off.

$$\text{if } P_l^{opt} = 0, \text{ then } \omega_l \geq 0 \text{ and if } P_l^{opt} > 0, \text{ then } \omega_l = 0$$

There exists a maximum index L for which holds: if $l>L$, then $P_l^{opt}=0$ and $\frac{\lambda_H(l)}{\sigma^2}=\mu-\omega_l$. With (6) and for $l\leq L$ the optimal solution is then given by

$$P_l^{opt}=\left[\sqrt{\frac{\sigma^2}{\mu\lambda_H(l)}}-\frac{\sigma^2}{\lambda_H(l)}\right]^+ \tag{8}$$

where μ satisfies $\sum_{l=1}^{L}P_l^{opt}=P$. This leads to:

Theorem 2: In case of perfect CSI at the Tx and Rx and a MMSE receiver, then the optimal transmit strategy is given by transmitting L data streams with the transmit vector $\vec{x}$:

$$\vec{x}=WD\vec{s}.$$

The unitary beamforming matrix W is given by the first L columns of V obtained from SVD of $H=V\Lambda^{1/2}U^*$ and the power allocation matrix $D=diag(\sqrt{P_1},...,\sqrt{P_L})$ with P_r in (8) from the solution of the minimization problem formulated in (5).

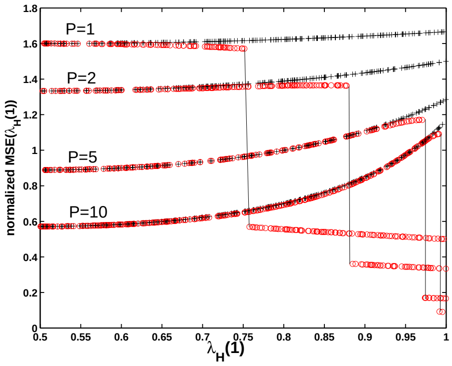

Fig. 2. MSE as a function of the eigenvalues $\lambda_H(1)$ with $\lambda_H(1)=(1-\lambda_H(2))$ and $\sum_{i=1}^{2}P_i$ = P = 1, 2, 5, 10 and $\sigma^2=1.0$

Fig. 2 shows the MSE functions for the ordered 2 Eigenvalues example. The upper curves(+) belong to the system with no CSI at the Tx and are Schur-convex. The lower curves ($\circ$) of the 4 sets are the MSE's with optimum power allocation. These functions are not Schur-convex, in general, which we show in the following. On the right hand side from the jump discontinuity only the effective MSE of the remaining data stream is depicted because P_2=0. Let us assume σ and P to be fixed and $\lambda_H(1)$ be the parameter for the MSE like in Fig. 2 ($\lambda_H(1)+\lambda_H(2)=1$). The solution of the minimization task be $L=2$ for $\lambda_H(1)=\lambda_H(2)$. With rising correlation ($\lambda_H(1)\uparrow$) we find a $\tilde{\lambda}_H(1)$, so that $L=1$. Now we consider $\lambda_H^{(1)}\geq\lambda_H^{(2)}\geq\tilde{\lambda}_H(1)$ then

$$\text{MSE}^{(1)}=\frac{1}{1+\frac{\lambda_H^{(1)}P}{\sigma^2}}\leq\frac{1}{1+\frac{\lambda_H^{(2)}P}{\sigma^2}}=\text{MSE}^{(2)}\leq\text{MSE}(\tilde{\lambda}_H)$$

which is Schur-concave for $\lambda_H^{(i)}\geq\tilde{\lambda}_H$. The more general case reads:

$$\sum_{l=1}^{M}\lambda_l^{(1)}=\sum_{l=1}^{M}\lambda_l^{(2)}\text{ and }\lambda^{(1)}\geq\lambda^{(2)}.$$

We substitute P_l in (5) with (8) and find for the L best channels in use $\text{MSE}(L)=\sum_{l=1}^{L}\frac{1}{\sqrt{\mu\lambda(l)}}$. Since we know that $\sum_{l=1}^{L}\frac{1}{\sqrt{\lambda^{(1)}(l)}}\geq\sum_{l=1}^{L}\frac{1}{\sqrt{\lambda^{(2)}(l)}}$ a comparison of $\text{MSE}^{(1)}=\sum_{l=1}^{L}\frac{1}{\sqrt{\mu^{(1)}\lambda(l)}}$ and $\text{MSE}^{(2)}=\sum_{l=1}^{L}\frac{1}{\sqrt{\mu^{(2)}\lambda(l)}}$ depends on $\mu^{(1)}$ and $\mu^{(2)}$. Therefore it can not be generally stated whether $\text{MSE}^{(1)}\geq\text{MSE}^{(2)}$ or vice versa. This complex behaviour is to be seen in Fig. 2.

We now find the critical power when a channel has to be switched off, assuming a fixed correlation and noise. We consider the two eigenvalue example. We assume P to be the sum power, so that $P_2^{opt}>0$. We choose a $\hat{P}$, with $P\geq\hat{P}$, which holds $P_1^{opt}=\hat{P}$ and $P_2^{opt}=0$. We find the function $f(P_1^{opt},P_2^{opt})$ and parameterize it with $P_1^{opt}=P-\varepsilon$ and $P_2^{opt}=\varepsilon$.

$$\begin{aligned} f(P_1^{opt},P_2^{opt}) &= \frac{1}{1+\frac{\lambda_H(1)P_1^{opt}}{\sigma^2}}+\frac{1}{1+\frac{\lambda_H(2)P_2^{opt}}{\sigma^2}} \\ f(P-\varepsilon,\varepsilon) &= \frac{1}{1+\frac{\lambda_H(1)(P-\varepsilon)}{\sigma^2}}+\frac{1}{1+\frac{\lambda_H(2)\varepsilon}{\sigma^2}} \end{aligned}$$

Now we look at the point where the derivative becomes positiv

$$\frac{df(P-\varepsilon,\varepsilon)}{d\varepsilon}\,|_{\varepsilon=0}\geq 0.$$

$$\frac{df}{d\varepsilon}\,|_{\varepsilon=0}=\frac{\frac{\lambda_H(1)}{\sigma^2}}{\left(1+\frac{\lambda_H(1)(P-\varepsilon)}{\sigma^2}\right)^2}-\frac{\frac{\lambda_H(2)}{\sigma^2}}{\left(1+\frac{\lambda_H(2)\varepsilon}{\sigma^2}\right)^2}\geq 0.$$

Theorem 3: A necessary and sufficient condition for beamforming to be optimum is given by

$$P\leq\frac{\sigma^2}{\lambda_H(1)}\left(\sqrt{\frac{\lambda_H(1)}{\lambda_H(2)}}-1\right)=P^{crit}.$$

3. BIT-LOADING STRATEGIES

Based on the results from above we propose a bit-loading strategy (BLS) which uses finite modulation alphabets. Under the constraint of a minimum transmission quality (e.g. maximum BER) the BLS finds the best match of the available SNRs per data stream and the finite data symbols. In this way we achieve the highest data throughput under the given constrains.

1.) The BLS starts with the above given optimum solution and computes the SNR for every sub-stream. Each of the sub-streams is then given the highest modulation satisfying the BER constraint.
2.) The allocated Tx power per data stream is reduced to achieve just the necessary SNR for each modulation.
3.) Test, if taking one bit/Hz/s from one data stream and giving it to another saves power. If not, proceed to next step.
4.) Sum up the remaining power and give it to the stream which can support one more bit/Hz/s (a higher modulation scheme) while it needs least from the rest power.
5.) Continue from 1. to 4. until no positiv rest power is available after an intended modulation step.
6.) Distribute the remaining power in a way that all channels have the same SNR increase.

4. SIMULATION

The simulation shall illustrate the described bit-loading-strategy. Without loss of generality we choose a transmit matrix H with random real entries instead of complex i.i.d. values. M=4, N=4 and L=1,...,4.

We first compute the optimal power allocation solution depending on L and give the MSE's and SNR's for the sub-streams. Next we perform optimal bit-loading with the aim of maximizing the data throughput under a maximum BER constraint.

4.1. example for the minimization task solution

We set $\sigma^2 = 0.1$ and $P = 15$. We choose a random matrix H

$$H = \begin{pmatrix} 0.733 & -0.777 & 0.314 & 0.398 \\ 0.059 & 1.55 & 1.419 & -0.073 \\ 0.149 & 1.055 & 0.327 & 1.315 \\ 1.596 & -0.166 & 0.475 & 0.978 \end{pmatrix}$$

with the following Eigenvalues: Λ_H=(6.45; 4.69; 1.28; 0.12).

Now we show the optimal power allocations $D_{L=i}$ derived from (8) for this example and $L = 1, ..., 4$:

$$D_{L=4} = \begin{pmatrix} 1.347 \\ 1.577 \\ 2.982 \\ 9.094 \end{pmatrix} \quad \text{SNR[dB]} = \begin{pmatrix} 19.39 \\ 18.69 \\ 15.82 \\ 10.45 \end{pmatrix}$$

$$D_{L=3} = \begin{pmatrix} 3.405 \\ 3.992 \\ 7.603 \end{pmatrix} \quad \text{SNR[dB]} = \begin{pmatrix} 23.42 \\ 22.72 \\ 19.88 \end{pmatrix}$$

$$D_{L=2} = \begin{pmatrix} 6.904 \\ 8.096 \end{pmatrix} \quad \text{SNR[dB]} = \begin{pmatrix} 26.49 \\ 25.79 \end{pmatrix}$$

$$D_{L=1} = (15.00) \quad \text{SNR[dB]} = (29.86)$$

The sum of the MSE's is then given by:

L	$\sum MSE's$
4	$1.3293E-1$
3	$2.0006E-2$
2	$4.8639E-3$
1	$1.0315E-3$

4.2. Bit-Loading example

In the following we consider a set of discrete modulation schemes available for bit loading. To limit complexity we assume the same error protection coding for all data streams on bit level before modulation otherwise adaptive coding would be one more parameter. Fig. 3 shows spectral efficiency, modulation and the required SNR per sub-channel to achieve an uncoded BER better than 10^{-3} [8].

Bits/s/Hz	Modulation	E_b/N_0[dB] BER$\approx 10^{-3}$	SNR[dB] BER$\approx 10^{-3}$
1	BPSK	6.8	6.8
2	QPSK	6.8	9.8
3	8QAM	9.0	13.7
4	16QAM	10.6	16.6
5	32QAM	12.7	19.7
6	64QAM	15.0	22.7
7	128QAM	17.1	25.5

Fig. 3. Spectral efficiency, Modulation and SNR (BER < 10^{-3})

We now start with the optimal MMSE solution from above and initially load all channels according to their actual SNR's.

$$\text{SNR[dB]} = \begin{pmatrix} 19.39 \\ 18.69 \\ 15.82 \\ 10.45 \end{pmatrix} \quad \text{MOD}^{1'} = \begin{pmatrix} 16QAM \\ 16QAM \\ 8QAM \\ QPSK \end{pmatrix}$$

We reduce the emitted power that just the chosen modulation is supported and sum up the remaining power. We check if taking one bit/Hz/s from one channel and giving it to another saves power. If no gain can be achieved we add the remaining power to every sub-channel and compute the rest power after this modulation step. The channel which would needs the least power to transmit 1 more bit/s/Hz is upgraded. Next we compute the minimum necessary Tx power to satisfy the new modulation scheme and s.o.

After 6 iterations no modulation upgrade can be performed with the remaining power and the BLS terminates with:

$$\text{MOD}^{6'} = \begin{pmatrix} 64QAM \\ 64QAM \\ 32QAM \\ 0 \end{pmatrix} \quad D^{6'}_{L=4} = \begin{pmatrix} 2.884 \\ 3.968 \\ 7.288 \\ 0 \end{pmatrix}$$

where $D^{6'}_{L=4}$ means the power allocation matrix after 6 iterations. The remaining power (5.7%) is distributed over the 3 active channels that their SNR's are improved equally or just power is saved. The exemplary modulation scheme would support a spectral efficiency of 17 bits/s/Hz with a BER better than 10^{-3}. The other extreme is to allocate all power to the best channel. The $\text{SNR}_{L=1}$=29.8 dB can support a 256QAM which has a 8 bits/s/Hz efficiency. If we start with the highest possible L we quickly find the maximum data throughput supporting a targeted BER.

5. CONCLUSION

We showed that correlation increases the MSE if no CSI is available at the Tx. This does not hold in general if the Tx has CSI and the data signals are preprocessed prior to transmission. We showed that preprocessing with a unitary beamforming matrix and a power allocation matrix can minimize the MSE at the Rx. The solution was obtained by solving the minimization problem from (5).

Based on this optimum power allocation we proposed a bit loading algorithm which matches the optimum power distribution over the sub-channels with finite modulation alphabets under the constraint of a certain BER requirement.

6. REFERENCES

[1] P.Viswanath, V.Anantharam and D.N.C. Tse, "Optimal Sequences, Power Control and User Capacity of Synchronous CDMA Systems with Linear MMSE Multiuser Receivers," *on their website*, March 1999.

[2] G.Wunder, S.Stanczak and H.Boche, "Channel Estimation for the Uplink of CDMA Systems with Linear MMSE Estimators: Lower Bounds and Optimal Sequences," in *ISSSTA 02*, Prag, Sept. 2002, pp. 34–38.

[3] G.Wunder and H.Boche, "Performance Bounds and Optimal Pilot Signals in OFDM-MIMO Systems," in 10^{th} *Aachen Symposium on Signal Theory*, Sept. 2001, pp. 100–106.

[4] H.Boche and E.Jorswieck , "On the Schur-concavity of the Ergodic and Outage Capacity with respect to Correlation in Multi-antenna Systems with no CSI at the Transmitter," in *Allerton 02*, October 2002.

[5] P.S.Chow, J.M.Cioffi, and J.A.C.Bingham, "A Practical Discrete Multitone Transceiver Loading Algorithm for Data Transmission over Spectrally Shaped Channels," *IEEE Trans. on Communications*, vol. 43, no. 2/3/4, pp. 773–775, Feb./ Mar./ April 1995.

[6] B.S.Krongold, K.Ramchandran, and D.L.Jones, "Computationally Efficient Optimal Power Allocation Algorithms for Multicarrier Communication Systems," *IEEE Trans. on Communications*, vol. 48, no. 1, pp. 23–27, 2000.

[7] A.W.Marshall and I.Olkin, *Inequalities: Theory of Majorization and Its Application*, vol. Vol. 143 of *Mathematics in Science and Engineering*, Academic Press, Inc. (London) Ltd., 1979.

[8] Fuqin Xiong, *Digital modulation techniques*, II. Artech House, Inc., Norwood, MA 02062, 2000.