Achievable Rate of MIMO Channels with Data-Aided Channel Estimation

Jens Baltersee, Gunnar Fock, Heinrich Meyr
Integrated Signal Processing Systems Lab
Aachen University of Technology (RWTH)
Templergraben 55, 52056 Aachen, Germany
e-mail: Jens.Baltersee@iss.rwth-aachen.de

Abstract — **The achievable rate of a coherent coded modulation (CM) digital communication system with data-aided channel estimation and a discrete, equiprobable symbol alphabet is derived under the assumption that the system operates on a flat fading MIMO channel and uses an interleaver to combat the bursty nature of the channel. It is shown that linear minimum mean square error (LMMSE) channel estimation directly follows from the derivation, and links average mutual information to the channel dynamics. Based on the assumption that known training symbols are transmitted, the achievable rate of the system is optimized with respect to the amount of training information needed.**

I. Introduction

Due to complexity constraints, virtually all of todays digital wireless communication systems follow the principle of synchronized detection for which a channel estimate is formed and subsequently used for detection as if it were the true known channel [1]. Furthermore, we assume that known pilot symbols are transmitted in order to estimate the channel in a data-aided (DA) fashion. We are interested in the question of how DA channel estimation affects the average mutual information of MIMO communication systems. Similar problems have been treated in the literature and [7] gives a good overview of the area. In [4], for example, the capacity of a system with a fixed, modified nearest-neighbor decoding rule is analyzed with respect to errors in the estimation of the channel fading process. Here, since we make no assumption on the decoding rule, optimal decoding is implied. Similar problems are also treated in [5] and [6]. However, even if these two papers consider a similar problem, the approach taken is entirely different. Here, the average mutual information of a flat fading MIMO system with perfect interleaving is computed, whereas [5] derives capacity bounds for a channel without interleaver, and [6] uses these bounds in the framework of MIMO block-fading channels. Furthermore, it is shown here that the MMSE channel estimator directly follows from the derivation, whereas in [5] channel estimation is introduced in an ad-hoc fashion and not an outcome of the derivation. The quantitative results presented here assume a time-varying (Rayleigh fading) channel model unlike the block-fading model used in [6]. In this paper, average mutual information is computed for a typical coded modulation (CM) transmission system operating on flat fading MIMO channels and using an interleaver to combat the bursty nature of the channel. The interleaver is an integral part of the system, since most well-known codes have been devised to combat statistically independent channel realizations. Virtually all of todays communication systems are based on that assumption, and therefore this type of channel is of greatest practical interest. The results of the derivation are then used to optimize the achievable rate with respect to the amount of pilot information needed for a given scenario in terms of channel dynamics and the SNR via Monte Carlo simulations.

II. Channel Model

In the transmitter of a CM transmission system, the signal is encoded, interleaved, and pilot symbols are inserted.The inner receiver performs DA channel estimation and delivers the channel estimates and the received symbols to the outer receiver. The outer transmission system thus comprises channel coding, modulation, interleaving/deinterleaving, and decoding. The interleaving/deinterleaving is employed to transform the bursty channel into an independently distributed channel. This is necessary, since most well-known codes have been devised to combat statistically independent channel realizations. We assume a flat fading channel which is characterized by the number of transmit antennas T and the number of receive antennas R. In the following such systems are referred to as $R \times T$ MIMO systems. For the encoded and interleaved data symbols $\mathbf{a}_k^{'}$ a linear transmission model results where at time instance k a received signal vector $\mathbf{z}_k^{'}$ depends on a $R \times T$ channel matrix $\mathbf{H}_k^{'}$. Equivalently, one can also use a $R \times RT$ transmit signal matrix $\mathbf{A}_k^{'}$ and a $RT \times 1$ channel vector $\mathbf{h}_k^{'}$. The second representation is better suited for the subsequent derivations.

$$\mathbf{z}_k^{'} = \mathbf{H}_k^{'}\mathbf{a}_k^{'} + \mathbf{m}_k^{'} = \mathbf{A}_k^{'}\mathbf{h}_k^{'} + \mathbf{m}_k^{'} \tag{1}$$

The additive complex Gaussian noise at the different receive antenna elements is assumed to be independent, i.e. we have $E\{\mathbf{m}_k^{'}\mathbf{m}_k^{'H}\} = N_0 \cdot \mathbf{I}$. Furthermore, the entries $h_{l;k}^{'}$ of the channel vector $\mathbf{h}_k^{'}$ are modeled as stationary, zero mean, circularly symmetric complex Gaussian processes of variance σ_h^2 and Doppler spectrum $S(e^{j\omega})$. This choice models a Rayleigh fading environment with enough separation of the receiving antenna elements such that the fades for each transmit-receive antenna pair are independent. Furthermore, unity symbol power is assumed throughout, i.e. $E\{\mathbf{a}_k^{'H}\mathbf{a}_k^{'}\} = 1$. For purposes of DA channel estimation, it is assumed that, on average, N_P known pilot symbol vectors $\{\mathbf{a}_{P;k}\}$ per N_D data symbol vectors are multiplexed into the interleaved data stream. We call the period, with which we introduce one pilot symbol vector per each transmit antenna into the data stream, the channel sampling period L. For the received pilot symbol vectors, the

0-7803-7119-4/01/$10.00 ©2001 IEEE

following transmission model results

$$\mathbf{z}_{P;k} = \mathbf{H}_{P;k}\mathbf{a}_{P;k} + \mathbf{m}_{P;k} = \mathbf{A}_{P;k}\mathbf{h}_{P;k} + \mathbf{m}_{P;k} \quad (2)$$

Based on the received pilot symbols, the DA channel estimation produces estimates $\{\hat{\mathbf{h}}_k'\}$ of the fading process $\{\mathbf{h}_k'\}$. Both, the channel estimates and the received data symbols are passed to the de-interleaver which maps $\{\mathbf{z}_k'\} \to \{\mathbf{z}_k\}$ and $\{\hat{\mathbf{h}}_k'\} \to \{\hat{\mathbf{h}}_k\}$. If we assume an ideal interleaving/de-interleaving operation which produces independent received data symbols then the mapping $\{\mathbf{z}_k'\} \to \{\mathbf{z}_k\}$ results in the following channel model

$$\mathbf{z}_k = \mathbf{H}_k\mathbf{a}_k + \mathbf{m}_k = \mathbf{A}_k\mathbf{h}_k + \mathbf{m}_k \quad (3)$$

where now the individual fading processes $h_{l;k}$ of the channel vector $\mathbf{h}_k$ are spectrally white. The entire sequence of N_P received pilot symbol vectors, denoted with the $R \cdot N_P \times 1$ vector $\mathbf{z}_P$, is given by

$$\mathbf{z}_P = \mathbf{A}_P\mathbf{h}_P + \mathbf{m}_P \quad (4)$$

Similarly, the received data symbol vectors are stacked in $\mathbf{z}$, and the corresponding transmitted data, channel, and noise are written as $\mathbf{A}$, $\mathbf{h}$, and $\mathbf{m}$, respectively.

III. DERIVATION OF THE ACHIEVABLE RATE

Since the transmission model assumes the usage of pilot symbols in order to estimate the channel, the channel is said to be partially known to the receiver. We denote this channel P-CSI. In contrast, if complete (perfect) channel knowledge is available, we choose to use C-CSI. The P-CSI channel described in the previous Section, with inputs $\{\mathbf{A}\}$, output $\{\mathbf{z}, \mathbf{z}_P\}$, and known parameter $\{\mathbf{A}_P\}$ is completely characterized by the distribution $p(\mathbf{z}, \mathbf{z}_P|\mathbf{A}, \mathbf{A}_P)$. Hence, for a given symbol constellation the capacity of this channel is given by the average mutual information $I(\mathbf{z}, \mathbf{z}_P; \mathbf{A}|\mathbf{A}_P)$. We will denote this achievable rate with C^* as compared to C which is reserved for the true channel capacity that requires the maximization over the input symbol distribution. Mutual information is measured in bits per channel use, where one channel use is defined as one second per hertz. If we consider a block of $N = N_D + N_P$ transmitted symbol vectors, of which N_D are usable data symbols vectors, the achievable rate over such a channel per channel use is given by

$$C^*_{P-CSI} = \frac{1}{N} I(\mathbf{z}, \mathbf{z}_P; \mathbf{A}|\mathbf{A}_P) \quad (5)$$

Now, according to the chain rule for mutual information we can rewrite $I(\mathbf{z}, \mathbf{z}_P; \mathbf{A}, \mathbf{A}_P)$ as follows

$$\begin{aligned} I(\mathbf{z}, \mathbf{z}_P; \mathbf{A}|\mathbf{A}_P) &= I(\mathbf{z}_P; \mathbf{A}|\mathbf{A}_P) + I(\mathbf{z}; \mathbf{A}|\mathbf{A}_P, \mathbf{z}_P) \\ &= I(\mathbf{z}; \mathbf{A}|\mathbf{A}_P, \mathbf{z}_P) \end{aligned} \quad (6)$$

where $I(\mathbf{z}_P; \mathbf{A}|\mathbf{A}_P) = 0$, since $\mathbf{z}_P$ does not convey any information about $\mathbf{A}$. Having that in mind, it is possible to write

$$I(\mathbf{z}, \mathbf{z}_P; \mathbf{A}|\mathbf{A}_P) = I(\mathbf{z}; \mathbf{A}|\mathbf{A}_P, \mathbf{z}_P) = E_{\mathbf{z}, \mathbf{z}_P, \mathbf{A}}\left\{ \log \frac{p(\mathbf{z}|\mathbf{A}, \mathbf{A}_P, \mathbf{z}_P)}{\sum_{\mathbf{A}} p(\mathbf{z}|\mathbf{A}, \mathbf{A}_P, \mathbf{z}_P) \cdot p(\mathbf{A})} \right\} \quad (7)$$

The channel characterized by the distribution $p(\mathbf{z}|\mathbf{A}, \mathbf{A}_P, \mathbf{z}_P)$ is not memoryless, because the fading co-efficients are not perfectly known to the receiver. Therefore, in general, the received data symbol vectors $\{\mathbf{z}_k\}$ are not independent. However, with an ideal interleaving/de-interleaving operation that completely breaks up the channel memory, and for a finite-index set $\mathcal{N}$ we can write [2, 3]

$$p(\mathbf{z}|\mathbf{A}, \mathbf{A}_P, \mathbf{z}_P) = \prod_{k \in \mathcal{N}} p(\mathbf{z}_k|\mathbf{A}_k, \mathbf{A}_P, \mathbf{z}_P) \quad (8)$$

This is a key assumption in our derivation. It must be mentioned that, exactly because the correlation between received data symbol vectors is not used, some information is actually "thrown away". However, it is also true at the same time that the better the channel is known, the less information is contained in these correlations. This is plausible, since for a perfectly known channel, the channel is indeed memoryless. A system with a reasonably good channel estimation scheme will therefore nevertheless exploit almost all the available information. Now, remembering that our transmission model is given by $\mathbf{z}_k = \mathbf{A}_k\mathbf{h}_k + \mathbf{m}_k$, it is obvious that $\mathbf{z}_k$ and $\mathbf{z}_P$, conditioned on $\mathbf{A}_k$ and $\mathbf{A}_P$, are jointly Gaussian. Therefore, $p(\mathbf{z}_k|\mathbf{A}_k, \mathbf{A}_P, \mathbf{z}_P)$ is also normally distributed and completely described by its conditional mean and covariance matrix. From estimation theory we know that the conditional mean $E\{\mathbf{z}_k|\mathbf{A}_k, \mathbf{A}_P, \mathbf{z}_P\}$ is the estimator $\hat{\mathbf{z}}_h$ of $\mathbf{z}_k$ in the minimum mean square error (MMSE) sense. Since the channel model is linear and all associated quantities are Gaussian, the corresponding estimator is itself linear. The conditional mean $\hat{\mathbf{z}}_k$ computes as

$$\begin{aligned} \hat{\mathbf{z}}_k &= E\{\mathbf{z}_k|\mathbf{A}_k, \mathbf{A}_P, \mathbf{z}_P\} \\ &= \mathbf{A}_k \cdot E\{\mathbf{h}_k|\mathbf{A}_P, \mathbf{z}_P\} \\ &= \mathbf{A}_k\hat{\mathbf{h}}_k \end{aligned} \quad (9)$$

where it is recognized that $p(\mathbf{h}_k \mid \mathbf{A}_P, \mathbf{z}_P)$ is also Gaussian distributed and therefore $E\{\mathbf{h}_k \mid \mathbf{A}_P, \mathbf{z}_P\}$ is the optimal linear minimum mean square error (LMMSE) channel estimator $\hat{\mathbf{h}}_k$. The covariance matrix of $p(\mathbf{z}_k|\mathbf{A}_k, \mathbf{A}_P, \mathbf{z}_P)$ is given by

$$\begin{aligned} \mathbf{C}_{\hat{\mathbf{z}}_k} &= E\{|\mathbf{z}_k - \hat{\mathbf{z}}_k|^2 \mid \mathbf{A}_k, \mathbf{A}_P, \mathbf{z}_P\} \\ &= \mathbf{A}_k E\{|\mathbf{h}_k - \hat{\mathbf{h}}_k|^2\}\mathbf{A}_k^H + \mathbf{I} \cdot N_0 \\ &= \mathbf{A}_k \mathbf{C}_{\varepsilon;k} \mathbf{A}_k^H + \mathbf{I} \cdot N_0 \end{aligned} \quad (10)$$

The distribution $p(\mathbf{z}_k|\mathbf{A}_k, \mathbf{A}_P, \mathbf{z}_P)$ is thus normal according to

$$p(\mathbf{z}_k|\mathbf{A}_k, \mathbf{A}_P, \mathbf{z}_P) = p(\mathbf{z}_k|\mathbf{A}_k, \hat{\mathbf{h}}_k) \sim \mathcal{N}\left(\hat{\mathbf{z}}_k, \mathbf{A}_k \mathbf{C}_{\varepsilon;k} \mathbf{A}_k^H + \mathbf{I} \cdot N_0\right) \quad (11)$$

It is noticed, that this distribution is a function of the error covariance matrix $\mathbf{C}_{\varepsilon;k}$ of the channel estimate which depends on the time index k for which we wish to estimate the channel. It is therefore concluded that the capacity of an interleaved channel is a function of the channel dynamics via LMMSE channel estimation. Since $\hat{\mathbf{h}}_k$ is a linear combination of the Gaussian variables of $\mathbf{z}_P$, it is too Gaussian. The corresponding mean and covariance matrix are given by

$$\begin{aligned} E\{\hat{\mathbf{h}}_k\} &= E\{\mathbf{h}_k\} - E\{\mathbf{h}_k - \hat{\mathbf{h}}_k\} = \mathbf{0} \\ E\left\{|\hat{\mathbf{h}}_k|^2\right\} &= \sigma_h^2 \cdot \mathbf{I} - \mathbf{C}_{\varepsilon;k} \end{aligned} \quad (12)$$

The first result is due to the fact that for the LMMSE estimator the mean of the error is zero, and the latter result is a consequence of the orthogonality principle. In summary, two effects influence C^*_{P-CSI}: Firstly, the Gaussian process $p(\mathbf{z}_k|\mathbf{A}_k,\hat{\mathbf{h}}_k)$ has a higher variance than the channel AWGN $\{\mathbf{m}_k\}$, which leads to an effective SNR loss. Secondly, the optimal LMMSE estimator $\hat{\mathbf{h}}_k$ delivers channel estimates that are orthogonal to the estimation error. Hence, the estimated channel has a lower MSE than the true channel which, again, leads to an additional effective SNR loss. The quality of the LMMSE channel estimate is characterized by $\mathbf{C}_{\varepsilon;k}$. If pilot vectors are used that are orthogonal with respect to time, we have $\mathbf{C}_{\varepsilon;k} = \sigma^2_{\varepsilon;k} \cdot \mathbf{I}$. It can then be shown that for a rectangular Doppler spectrum $S(e^{j\omega})$, and assuming, in an information theoretic framework, an infinite length pilot symbol vector $\mathbf{a}_P$, we get [3]

$$\begin{aligned}\sigma^2_{\varepsilon;k} &= E\{|h_{l;k} - \hat{h}_{l;k}|^2\} \quad \text{for } l = 0,\ldots,RT-1 \\ &= \frac{2\,F\,L\,N_0\,\sigma_h^2}{2\,F\,L\,N_0 + \sigma_h^2}\end{aligned} \tag{13}$$

where F is the normalized cutoff frequency of the Doppler spectrum, and L is the channel sampling period (i.e. the period with which we introduce pilot symbols into the data stream).

IV. CONSTRAINED CAPACITY ANALYSIS & CONCLUSION

In the following we would like to examine how DA channel estimation affects the achievable rate of flat fading MIMO systems. We recall that the average mutual information is a function of the channel dynamics through LMMSE channel estimation which followed directly from the derivation. Specifically, it was shown that the total achievable rate is given by $C^*_{P-CSI} = \frac{1}{N} I(\mathbf{z};\mathbf{A}|\mathbf{z}_P,\mathbf{A}_P)$ which can be calculated by evaluating eq. (7) via Monte Carlo simulation. Since each channel coefficient is sampled with a rate $1/L$, the fraction of the channel capacity available to data transmission is given by $(N-N_P)/N = (L-T)/L$. The assumption that the channel must be sampled at least with Nyquist rate, in order to guarantee that the channel can be reconstructed properly, results in the constraint $L \leq \frac{0.5}{F}$. Figure 1 shows the result of the simulations for 1×1, and 4×4 systems, plotting channel capacity versus channel sampling period L for several typical SNR's and Doppler frequencies. The simulations presented here assume an M-ary communication system with QAM modulation ($M = 16$ was used), and equally probable symbol vectors, i.e. $p(\mathbf{A}_k) = 1/M^T$. Capacity curves are only shown for channel sampling periods L up to the Nyquist rate. Note that results for the 1×1 system are scaled by a factor of 4 for ease of comparison to the 4×4 system. Inspection of the results reveals that all capacity curves for the 1×1 system exhibit a distinct maximum for channel sampling periods strictly smaller than Nyquist sampling. Two opposing effects are at work here: For very small L, the channel can be estimated very well and the channel estimates have a very small MSE. In order to achieve such a small MSE, however, a large portion of the data rate is sacrificed to the insertion of pilot symbols. For very large L approaching Nyquist sampling, the capacity is reduced not primarily due to inserted pilot symbols, but the increased MSE of the channel estimates now reduces $I(\mathbf{z};\mathbf{A}|\mathbf{z}_P,\mathbf{A}_P)$. Further

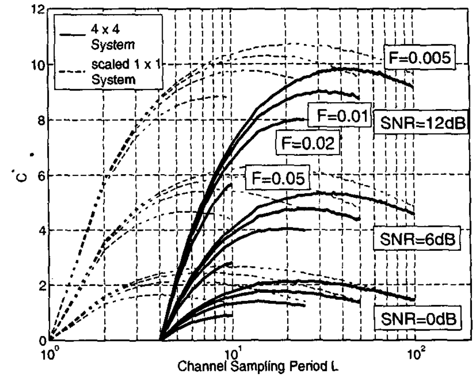

Figure 1: Capacities of 1×1, and 4×4 Systems.

inspection of the simulation results indicates that for smaller Doppler frequencies the capacity difference between the optimal L and Nyquist rate sampling is much larger than for the larger Doppler frequencies. In that sense, an optimal choice of L is much more important for a slow fading channel than for a fast fading channel, because potential capacity losses are larger. For C-CSI channels, it is well known that the capacity of a MIMO system scales linearly with the number of antenna pairs. The Figure demonstrates that for small Doppler frequencies capacity scales almost linearly, whereas for increasingly larger Doppler frequencies capacity eventually degrades substantially. It is noted that in general the optimal capacities are achieved for the 4×4 system at significantly higher channel sampling periods L than for the 1×1 system. This is plausible, since for larger numbers of transmit antennas T, more pilot symbol vectors are required in order to sample the channel adequately. We conclude that, considering that the percentage of the data rate which is dedicated to pilots grows linearly with the number of transmit antennas, it becomes clear that, for a large amount of antennas and a rapidly varying channel, too many pilots are needed in order to adequately estimate the channel, and in such cases significant capacity losses are inevitable.

REFERENCES

[1] Heinrich Meyr, Marc Moeneclaey and Stefan Fechtel. *Digital Communication Receivers : Synchronization, Channel Estimation and Signal Processing*, John Wiley and Sons, New York, 1998.

[2] Guiseppe Caire, Giorgio Taricco, Ezio Biglieri. "Bit-Interleaved Coded Modulation", *IEEE Trans. Inf. Theory*, Vol. 44, No. 3, pp. 927–946, May 1998

[3] Jens Baltersee, Gunnar Fock, Heinrich Meyr. "Achievable Rate of MIMO Channels with DA Channel Estimation", *submitted to IEEE JSAC*

[4] Amos Lapidoth, and Shlomo Shamai (Shitz). "Fading Channels: How Perfect need 'Perfect Side Information' be?", *Proceedings of Information Theory and Communications Workshop*, pp. 36–38, 1999

[5] Muriel Medard. "The Effect upon Channel Capacity in Wireless Communications of Perfect and Imperfect Knowledge of the Channel", *IEEE Trans. Inf. Theory*, Vol. 46, No. 3, pp. 933–946, May 2000

[6] Ashutosh Sabharwal, Elza Erkip, Behnaam Aazhang. "On Channel State Information in Multiple Antenna Block Fading Channels", *Proceedings of International Symposium on Information Theory and Its Applications*, Honululu, Hawaii, November 2000

[7] Ezio Biglieri, John Proakis, and Shlomo Shamai (Shitz) "Fading Channels: Information-Theoretic and Communications Aspects, *IEEE Trans. Inf. Theory*, Vol. 44, No. 6, pp. 2619–2692, October 1998