PERFORMANCE EVALUATION OF MIMO COMMUNICATION SYSTEMS BASED ON SUPERIMPOSED PILOTS

Patrik Bohlin and Mikael Tapio

Department of Signals & Systems
Chalmers University of Technology, Sweden
Email: {pbohlin,mito}@s2.chalmers.se

ABSTRACT

In the recent past, a number of publications have suggested superimposed pilots (SIP) for channel estimation in MIMO systems. However, the performance gain achieved by SIP compared to conventional (time-multiplexed) training is still questionable. In this paper we introduce a framework for modeling a general SIP scheme, of which the conventional scheme is a special case. Utilizing this framework, we derive a lower bound on the channel capacity to compare the performance between the two systems. It is found that in certain scenarios (high SNR, many receive antennas and short coherence times), it is beneficial to also transmit data during the training mode (i.e., use SIP). The main conclusion though, is that for most cases where these kind of schemes would be realistic, i.e. for systems with coherence lengths of more than a few symbols, the general SIP-scheme reduces to the conventional scheme. Hence, rendering the same capacity.

1. INTRODUCTION

In order to detect the transmitted data at the receiver, an estimate of the channel is often used. One common technique to estimate the channel involves transmission of a known training sequence. The obtained channel estimate is then used for detection of the distorted data. Conventionally, these pilot symbols are time-multiplexed with the data. However, recent results suggest that the pilot can be embedded in (added to) the data, so called superimposed pilots (SIP). See e.g., [1, 2, 3]. This allows transmission of data also during training. We generalize this scheme by allowing different power levels for the data, as illustrated in Fig. 1. In this more general setting, conventional training becomes a special case of SIP (by simply letting $\sigma_{dt}^2 = 0$), which gives a useful framework for comparing the performance of the two methods.

The advantage of SIP, as compared to conventional training, is that data is transmitted in all time slots – no slots are purely dedicated to known pilots. This can lead to higher spectral efficiency. The drawback of SIP is a degradation in the channel estimate quality, since the received signal, which is used for channel estimation, also contains additive unknown data. The unknown data is typically incorporated in the noise term, in essence rendering a lower SNR.

Based on a block fading channel model, we find the optimal training parameters in a SIP-based MIMO system employing an LMMSE channel estimator, and compare the performance to the conventional system. We achieve this by deriving a lower bound on the ergodic channel capacity and maximize this bound over the time allocated for training and data transmission, the number of transmit antennas and the power allocated for training and data symbols. That is, we extend the results in [4] to cover also the more general SIP case. The results suggest that the optimal SIP-based system gain in capacity, compared to conventional training-based systems, for short channel coherence times and increasing numbers of receive antennas.

2. DATA MODEL

We employ a block-fading model for the MIMO channel in which the channel matrix $\mathbf{H}$ is assumed to be constant during T (channel coherence time) symbol periods and then changes to a new independent realization. This model can be seen as an approximation of, e.g., a TDMA or frequency hopping system. The $N \times M$ channel matrix is assumed to have independent complex Gaussian $\mathcal{CN}(0, 1)$ entries. M is the number of transmit antennas and N is the number of receive antennas.

According to the SIP concept, the transmission of a $M \times T$ data block is split into two modes; 1) transmission of a data sub-block of length T_t symbols added with pilot symbols, followed by, 2) only unknown data sub-block of length T_d symbols. The length of the total block (SIP block and data block) equals the channel coherence time, i.e. $T = T_t + T_d$. Due to the fact that we assume the channel to be constant over the coherence interval, there is no loss of generality by placing the training symbols in the beginning of each block. See Fig. 1.

1) **Training mode:** Let $\mathbf{S}_t$ and $\mathbf{S}_{dt}$ denote the normalized transmitted $M \times T_t$ complex-valued known pilot symbol and random data (coded) symbol matrices, respectively. The

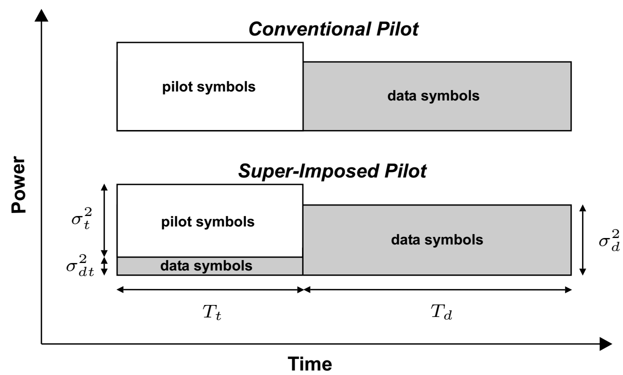

Fig. 1. Conventional pilot insertion (top) and superimposed pilot (bottom).

signal matrices also have the following constraints; $\mathrm{tr}\{\mathbf{R}_{\mathbf{S}_t}\} = MT_t$ and $\mathrm{tr}\{\mathbf{R}_{\mathbf{S}_{dt}}\} = MT_t$, where $\mathbf{R}_{\mathbf{S}_t} = \mathbf{S}_t\mathbf{S}_t^H$ and $\mathbf{R}_{\mathbf{S}_{dt}} = \mathrm{E}\{\mathbf{S}_{dt}\mathbf{S}_{dt}^H\}$. The power allocated for the training and data symbols are denoted by σ_t^2 and σ_{dt}^2, respectively, i.e. the total transmitted energy during training is $(\sigma_t^2 + \sigma_{dt}^2)T_t$. The received $N \times T_t$ signal matrix during training is then

$$\mathbf{X}_t = \mathbf{H}\left(\sqrt{\frac{\sigma_t^2}{M}}\mathbf{S}_t + \sqrt{\frac{\sigma_{dt}^2}{M}}\mathbf{S}_{dt}\right) + \mathbf{V}_t, \tag{1}$$

where $\mathbf{V}_t$ is the $N \times T_t$ additive noise matrix with independent $\mathcal{CN}(0,1)$ entries and is both channel and data independent.

2) **Data mode:** During data only transmission we transmit the $M \times T_d$ normalized data matrix $\mathbf{S}_d$. Thus, the received $N \times T_d$ signal matrix becomes

$$\mathbf{X}_d = \sqrt{\frac{\sigma_d^2}{M}}\mathbf{H}\mathbf{S}_d + \mathbf{V}_d,\ \mathrm{tr}\{\mathbf{R}_{\mathbf{S}_d}\} = \mathrm{Etr}\{\mathbf{S}_d\mathbf{S}_d^H\} = MT_d, \tag{2}$$

where $\mathbf{V}_d$ is the $N \times T_d$ noise matrix with independent $\mathcal{CN}(0,1)$ entries. $\mathrm{Etr}\{\cdot\}$ denotes the expected value of the trace. The power allocated for data symbols is denoted by σ_d^2 and the total energy during data mode is $\sigma_d^2 T_d$.

The LMMSE principle is used to obtain the channel estimate. Due to its orthogonality property, the estimate $\hat{\mathbf{H}}$ is uncorrelated with the estimation error $\tilde{\mathbf{H}} = \mathbf{H} - \hat{\mathbf{H}}$. Treating the unknown data as noise, the LMMSE channel estimate can be written as

$$\begin{aligned}\hat{\mathbf{H}} &= N\sqrt{\frac{\sigma_t^2}{M}}\mathbf{X}_t\mathbf{R}_{\mathbf{X}_t^H}^{-1}\mathbf{S}_t^H \\ &= \sqrt{\frac{M}{\sigma_t^2}}\mathbf{X}_t\mathbf{S}_t^H\left(\frac{M(\sigma_{dt}^2+1)}{\sigma_t^2}\mathbf{I} + \mathbf{S}_t\mathbf{S}_t^H\right)^{-1}. \end{aligned}\tag{3}$$

In the second step we have used $\mathbf{R}_{\mathbf{S}_{dt}^H} = \mathrm{E}\{\mathbf{S}_{dt}^H\mathbf{S}_{dt}\} = M\mathbf{I}$. It is later argued that choosing the data to be spatio-temporally uncorrelated maximizes the mutual information between the transmitted and the received matrices. After removing the pilots and using the channel estimate as the true channel, we will have the following signal model in the training mode

$$\begin{aligned}\mathbf{X}'_t &= \mathbf{X}_t - \sqrt{\frac{\sigma_t^2}{M}}\hat{\mathbf{H}}\mathbf{S}_t = \sqrt{\frac{\sigma_{dt}^2}{M}}\hat{\mathbf{H}}\mathbf{S}_{dt} + \\ &\quad \underbrace{\sqrt{\frac{\sigma_t^2}{M}}\tilde{\mathbf{H}}\mathbf{S}_t + \sqrt{\frac{\sigma_{dt}^2}{M}}\tilde{\mathbf{H}}\mathbf{S}_{dt} + \mathbf{V}_t}_{\mathbf{V}'_t}, \end{aligned}\tag{4}$$

where the noise $\mathbf{V}'_t$ has variance the $\sigma^2_{\mathbf{V}'_t} \triangleq \frac{1}{NT_t}\mathrm{Etr}\{\mathbf{V}'^H_t\mathbf{V}'_t\} = \sigma_t^2\sigma^2_{\tilde{\mathbf{H}}\mathbf{S}_t} + \sigma_{dt}^2\sigma^2_{\tilde{\mathbf{H}}} + 1$. Moreover, $\sigma^2_{\tilde{\mathbf{H}}\mathbf{S}_t}$ is given by $\sigma^2_{\tilde{\mathbf{H}}\mathbf{S}_t} \triangleq \frac{1}{MNT_t}\mathrm{Etr}\{\tilde{\mathbf{H}}^H\tilde{\mathbf{H}}\mathbf{R}_{\mathbf{S}_t}\}$ and $\sigma^2_{\tilde{\mathbf{H}}} \triangleq \frac{1}{MN}\mathrm{Etr}\{\tilde{\mathbf{H}}^H\tilde{\mathbf{H}}\}$.

Again using the channel estimate as the true channel, we have the following signal model in the data mode

$$\mathbf{X}_d = \sqrt{\frac{\sigma_d^2}{M}}\hat{\mathbf{H}}\mathbf{S}_d + \underbrace{\sqrt{\frac{\sigma_d^2}{M}}\tilde{\mathbf{H}}\mathbf{S}_d + \mathbf{V}_d}_{\mathbf{V}'_d}, \tag{5}$$

where the noise variance is $\sigma^2_{\mathbf{V}'_d} \triangleq \frac{1}{NT_d}\mathrm{Etr}\{\mathbf{V}'^H_d\mathbf{V}'_d\} = \sigma_d^2\sigma^2_{\tilde{\mathbf{H}}} + 1$.

Remark: In the SIP case, the LMMSE estimate is not the MMSE estimate. The noise $\mathbf{V}'_t$ has been colored by the channel and this information is ignored by the LMMSE estimator. Hence, conditioned on $\mathbf{X}_t$ and $\mathbf{S}_t$, the noises $\mathbf{V}'_t$ and $\mathbf{V}'_d$ are correlated with the respective signals $\mathbf{S}_{dt}$ and $\mathbf{S}_d$. Since the SIP scheme ignores this information, we can simply replace the noises $\mathbf{V}'_t$ and $\mathbf{V}'_d$ with other noises that have the same variance but that also are uncorrelated with the respective signals.

3. CAPACITY AND OPTIMIZATION

In this section a lower bound on the capacity for the SIP-based system described above is derived. We want to point out that the capacity bound of the SIP-based system is derived in a way to be comparable with the conventional scheme [4]. That is, we extend the results in [4] to SIP operation.

The capacity bound is found by maximizing a lower bound on the mutual information. Such an approach is applicable since we model the channel as time discrete and memoryless. Let $\hat{\mathbf{H}} = f(\mathbf{X}_t, \mathbf{S}_t)$ be the channel estimate, formed by treating the data $\mathbf{S}_{dt}$ as noise. The mutual information can then be written as

$$\begin{aligned} I &= \mathrm{I}[(\mathbf{X}_d, \mathbf{X}_t, \mathbf{S}_t, \hat{\mathbf{H}}); (\mathbf{S}_d, \mathbf{S}_{dt})] \\ &= \underbrace{\mathrm{I}[\mathbf{X}_d; (\mathbf{S}_d, \mathbf{S}_{dt}) \mid (\mathbf{X}_t, \mathbf{S}_t, \hat{\mathbf{H}})]}_{I_d} + \\ &\quad \underbrace{\mathrm{I}[\mathbf{X}_t; (\mathbf{S}_d, \mathbf{S}_{dt}) \mid (\mathbf{S}_t, \hat{\mathbf{H}})]}_{I_t} + \\ &\quad \mathrm{I}[(\hat{\mathbf{H}}, \mathbf{S}_t); (\mathbf{S}_d, \mathbf{S}_{dt})]. \end{aligned}\tag{6}$$

The idea of using a training (or SIP) based system is to use the channel estimate as if it was the true channel. This means that we can bound the mutual information I_t and I_d by only condition on the channel estimate i.e.

$$I_d \geq \mathrm{I}[\mathbf{X}_d; (\mathbf{S}_d, \mathbf{S}_{dt}) \mid \hat{\mathbf{H}}] \quad (7)$$

$$I_t \geq \mathrm{I}[\mathbf{X}_t; (\mathbf{S}_d, \mathbf{S}_{dt}) \mid \hat{\mathbf{H}}]. \quad (8)$$

Clearly, these measures are bounds since we have disregarded information by using the channel estimate as if it was correct. By conditioning on the channel, the data symbols will only affect the mutual information in their respective modes, which means that $\mathrm{I}[\mathbf{X}_d; (\mathbf{S}_d, \mathbf{S}_{dt})|\hat{\mathbf{H}}] = \mathrm{I}[\mathbf{X}_d; \mathbf{S}_d|\hat{\mathbf{H}}]$ and $\mathrm{I}[\mathbf{X}_t; (\mathbf{S}_d, \mathbf{S}_{dt})|\hat{\mathbf{H}}] = \mathrm{I}[\mathbf{X}_t; \mathbf{S}_{dt}|\hat{\mathbf{H}}]$. The last term in (6) will not contribute to the capacity, since the idea of the SIP-scheme is to treat $\mathbf{S}_{dt}$ as noise when estimating the channel. This means that data transmitted during the training mode is optimized only to maximize the mutual information for the known channel case. Finally, the SIP-based system capacity is given by the maximum, with respect to the distributions of $\mathbf{S}_d$ and $\mathbf{S}_{dt}$, of this bound on the mutual information

$$C_{\mathrm{sip}} = \max \frac{1}{T} \left\{ \mathrm{I}(\mathbf{X}_d; \mathbf{S}_d \mid \hat{\mathbf{H}}) + \mathrm{I}(\mathbf{X}_t; \mathbf{S}_{dt} \mid \hat{\mathbf{H}}) \right\}. \quad (9)$$

Starting from (4)–(5) and noting that the conditions of Theorem 1 in [4] are fulfilled, we can conclude that the noise with a given covariance matrix that will yield the lowest mutual information is Gaussian distributed. On the other hand, the distribution of the transmitted signal that maximizes the mutual information is also Gaussian distributed with no space or time correlation. That is, both $\mathbf{R}_{\mathbf{S}_{dt}}$ and $\mathbf{R}_{\mathbf{S}_d}$ should be multiples of the identity matrix. A lower bound on the capacity will then be composed of two terms; one for the training mode and one for the data mode and is given by

$$C_{\mathrm{sip}} \geq \mathrm{E}\frac{T_t}{T} \log\det\left(\mathbf{I} + \rho_t \frac{\bar{\mathbf{H}}\bar{\mathbf{H}}^H}{M}\right) + \mathrm{E}\frac{T_d}{T} \log\det\left(\mathbf{I} + \rho_d \frac{\bar{\mathbf{H}}\bar{\mathbf{H}}^H}{M}\right). \quad (10)$$

The elements of the normalized channel $\bar{\mathbf{H}} = \frac{\hat{\mathbf{H}}}{\sigma_{\hat{\mathbf{H}}}}$ will be uncorrelated with zero mean and unit variance and have a distribution that is approximately Gaussian. The normalization constant $\sigma_{\hat{\mathbf{H}}}$ is given by $\sigma^2_{\hat{\mathbf{H}}} = \frac{1}{NM}\mathrm{E}\mathrm{tr}\{\hat{\mathbf{H}}\hat{\mathbf{H}}^H\}$ and the effective SNRs are given by

$$\rho_t = \frac{\sigma^2_{dt}\sigma^2_{\hat{\mathbf{H}}}}{\sigma^2_{\mathbf{V}'_t}} = \frac{\sigma^2_{dt}\sigma^2_{\hat{\mathbf{H}}}}{\sigma^2_t\sigma^2_{\tilde{\mathbf{H}},\mathbf{R}_{\mathbf{S}_t}} + \sigma^2_{dt}\sigma^2_{\tilde{\mathbf{H}}} + 1} \quad (11)$$

$$\rho_d = \frac{\sigma^2_d\sigma^2_{\hat{\mathbf{H}}}}{\sigma^2_{\mathbf{V}'_d}} = \frac{\sigma^2_d\sigma^2_{\hat{\mathbf{H}}}}{\sigma^2_d\sigma^2_{\tilde{\mathbf{H}}} + 1}. \quad (12)$$

To maximize the capacity, (10) has to be optimized with respect to the following parameters: $\mathbf{S}_t, T_t, T_d, \sigma^2_t, \sigma^2_{dt}, \sigma^2_d$ and M subject to the constraints $T_t + T_d = T$, $(\sigma^2_{dt} + \sigma^2_t)T_t + \sigma^2_d T_d = \sigma^2 T$, where T is the coherence time of the channel. It might seem counter-intuitive to optimize over the number of transmit antennas, since when the channel is known the capacity is known to be an increasing function of the number of transmit antennas [5]. The reason why there will be an optimum is that the more transmit antennas we use, the more time we have to spend on training, since we need at least as many training symbols as there are transmit antennas for identifiability. Even if data is transmitted during the training mode, the effective SNR is much lower than during the data mode, so we want to keep the training period as short as possible, as we will later see.

Firstly, the criterion function (10) is concentrated with respect to the training sequence, $\mathbf{S}_t$. If the same $\mathbf{S}_t$ maximizes both ρ_t and ρ_d, that $\mathbf{S}_t$ will clearly also maximize the capacity. Using the orthogonality of the LMMSE estimate, we can write $\sigma^2_{\hat{\mathbf{H}}} = 1 - \sigma^2_{\tilde{\mathbf{H}}}$. Hence, from (12) we can see that maximizing ρ_d is equivalent to minimizing $\sigma^2_{\tilde{\mathbf{H}}}$, which in turn corresponds to following minimization problem

$$\min_{\mathbf{S}_t, ||\mathbf{S}_t||^2_F = MT_t} \mathrm{tr}\left[\mathbf{I} + \frac{\sigma^2_t}{M(1+\sigma^2_{dt})}\left(\mathbf{S}_t\mathbf{S}_t^H\right)^T\right]^{-1}, \quad (13)$$

and the solution $\mathbf{S}_t\mathbf{S}_t^H = T_t\mathbf{I}$. This choice of training sequence will also maximize the effective SNR during training mode. By rewriting (11) as

$$\rho_t = \frac{\sigma^2_{dt}(1-\sigma^2_{\tilde{\mathbf{H}}})}{\underbrace{\frac{1+\sigma^2_{dt}}{NT_t}}_{K_1} + 1 + \underbrace{\left(\sigma^2_{dt} - \frac{M(1+\sigma^2_{dt})}{T_t}\right)\sigma^2_{\tilde{\mathbf{H}}}}_{K_2}}, \quad (14)$$

it is obvious that maximizing ρ_t is the same as minimizing $\sigma^2_{\tilde{\mathbf{H}}}$, since $T_t \geq M$ and hence $K_1 + K_2 > 0$. Therefore, the solution is again given by $\mathbf{S}_t\mathbf{S}_t^H = T_t\mathbf{I}$.

Inserting this training sequence into (11) and (12) the effective SNRs become

$$\rho_t = \frac{\sigma^2_{dt}\sigma^2_t T_t}{(\sigma^2_{dt} + \sigma^2_t + 1)(\sigma^2_{dt}+1)M + \sigma^2_t T_t} \quad (15)$$

$$\rho_d = \frac{\sigma^2_d\sigma^2_t T_t}{(\sigma^2_d + 1)(\sigma^2_{dt}+1)M + \sigma^2_t T_t}. \quad (16)$$

Numerical optimization of (10) is used to find the optimal values of the remaining parameters.

4. NUMERICAL EXAMPLES

To illustrate the theories described above, the capacity lower bound has been calculated for a typical case where it can be beneficial to use the SIP scheme. In this paper a scenario

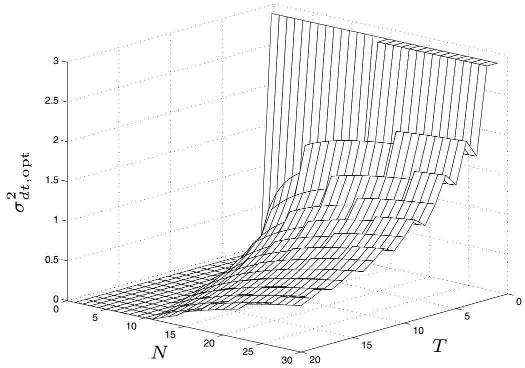

Fig. 2. Optimal power in data symbols during training, $\sigma^2_{dt,\mathrm{opt}}$, vs. channel coherence time T and number of receivers N.

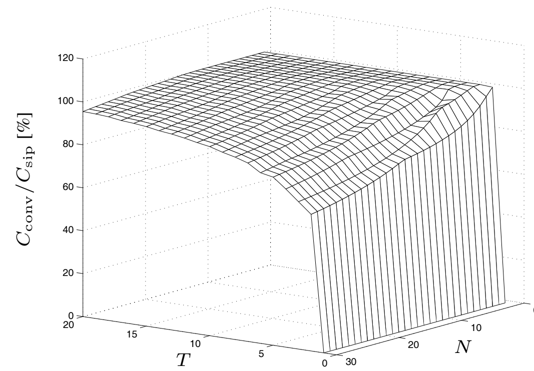

Fig. 3. Relative capacity in percent between SIP and conventional training, $C_{\mathrm{conv}}/C_{\mathrm{sip}}$, vs. channel coherence time T and number of receivers N. 10 dB SNR.

with an average SNR of 10 dB is considered, but similar results can be obtained also at other SNR values. We have evaluated the bound for the optimal number of transmit antennas, which depends on the number of receive antennas and the coherence length of the channel. Only channels with relatively short coherence time have been studied, since the longer the coherence time is, the lower the gain provided by the SIP scheme will be. This is due to the fact that the effective SNR during the data mode will be lower because of the extra noise in the estimation. Also, the fraction of time which is dedicated solely for training is reduced in the training based scheme.

In all scenarios we have found that the optimal number of training symbols always was equal to the number of transmit antennas, i.e. the smallest possible. This is in agreement with [4] and intuitively appealing since during the training mode the effective SNR will become much lower than during the data mode because of two things: First the data power σ^2_{dt} will be relatively small in order not to introduce a lot of extra noise during channel estimation, and second that there will be an extra noise term from the known training symbols due to the channel estimation error.

As can be seen in Fig. 2 the power of the data in the training mode increases as the number of receive antennas, N, increases. Most power should of course be allocated when the training mode takes up the whole coherence interval. We want to point out that it is not very often that it is beneficial to transmit data during the training mode. From Fig. 2 it can be concluded that there is only a small set of systems, especially with many receive antennas and short coherence times, where it is suboptimal to have $\sigma^2_{dt} = 0$.

In order to compare the performance of a conventional training-based scheme to that of a scheme based on SIP, the performance gain $C_{\mathrm{conv}}/C_{\mathrm{sip}}$ is shown in Fig. 3. It is found that the conventional training-based scheme will reach the SIP-based capacity as the coherence time increases or the number of receive antennas decreases.

5. CONCLUSIONS

In this paper we have extended the results in [4] to also cover superimposed pilot (SIP) based systems. The framework we set up for SIP also includes the conventional training scheme as a special case. Hence, we can compare the performances of the two systems by optimizing over their respective parameters. It is shown that in certain scenarios (many receive antennas and short coherence times) it is beneficial to also transmit data during the training mode (i.e. use SIP). The main conclusion, though, is that for most cases where a training based scheme would be realistic, i.e. for systems with coherence length of more than just a few symbols, the optimal SIP-scheme reduces to the optimal conventional scheme.

6. REFERENCES

[1] F. Mazzenga. “Channel Estimation and Equalization for M-QAM Transmission with a Hidden Pilot Sequence”. *IEEE Trans. on Broadcasting*, 46(2):170–176, June 2000.

[2] G.T. Zhou, M. Viberg, and T. McKelvey. “A first-order statistical method for channel estimation”. *IEEE SP Letters*, 10:57–60, March 2003.

[3] H. Zhu and B. Farhang-Boroujeny. “Pilot Embedding for Joint Channel Estimation and Data Detection in MIMO Communication Systems”. *IEEE Communications Letters*, 7(1):30–32, January 2003.

[4] B. Hassibi and B. M. Hochwald. “How Much Training is Needed in Multiple-Antenna Wireless Links?”. *IEEE Trans. on Information Theory*, 49(4):951–963, April 2003.

[5] I.E. Telatar. “Capacity of Multi-Antenna Gaussian Channels”. *Eur. Trans. Telecom.*, 10:585–595, November 1999.