Linear Precoding for MIMO Channels with Non-zero Mean and Transmit Correlation in Orthogonal Space-Time Coded Systems

Mai Vu, Arogyaswami Paulraj

Information Systems Laboratory, Department of Electrical Engineering
Stanford University, Stanford, CA 94305-9510, USA
E-mails: {mhv, apaulraj}@stanford.edu

***Abstract*– We study the problem of exploiting MIMO wireless transmitter channel knowledge in form of the channel mean and transmit correlation from a coding perspective. A linear channel precoder is designed to capture the channel information, which is used in concatenation with an orthogonal space-time block code. Based on the pair-wise error probability criterion, the optimal precoding matrix is derived analytically. Depending on the channel mean and correlation, and the SNR, the solution may require mode-dropping, which resembles the water-filling principle. The difference here is that both the "water level" and the mode directions change with each water-fill iteration, hence it is termed dynamic water-filling. Efficient binary search algorithms are proposed to carry out the dynamic water-filling process in solving for the optimal precoder. Numerical examples show that significant gain can be obtained using this linear channel precoder.**

1. INTRODUCTION

Since the seminal studies by Tarokh et.al [1, 2], there has been a wealth of space-time code (STC) designs to realize the potential performance gain in MIMO wireless channels. Due to the random nature of a wireless link, a common assumption for STC design is that the transmitter does not have any knowledge of the channel. The channel is therefore assumed to be Rayleigh fading, where the coefficients are i.i.d. zero-mean complex Gaussian random variables.

In practice, MIMO wireless channels may exhibit some conditions and statistics that are different from the common channel model above. For example, there can be a correlation between the transmit antennas [3], the channel can exhibit a Rician non-zero mean [4], or have a high K factor. While exact channel knowledge is difficult to get at the transmitter, some partial channel information can usually be obtained. This channel information can be in various forms, such as channel statistics (mean, correlation or both), channel estimates with error, or channel parameters (condition number, K factor, SNR...). Transmitter channel knowledge in a MIMO wireless system has been shown to provide valuable gain in system performance, in term of both mutual information [5, 8] and coding gain [6, 7], and at the same time it can help to significantly simplify the system complexity.

To exploit the channel information at the transmitter in a MIMO system, the information can either be incorporated into the design of a new STC, or it can be used to design a channel precoder separately from the STC. The role of a channel precoder is to capture transmitter channel knowledge, such that the same STC that was designed for i.i.d channels can be used for other channel conditions with a precoder. Hence precoder provides the flexibility of adapting to various channel knowledge conditions without having to change the STC. In this study, we examine the precoder design scenario, where a linear precoder is used in concatenation with a space-time block code. Linear precoder can be viewed as a matrix or multi-modal beamformer, where the modes (or "beam directions") are the left singular vectors of the precoder matrix, and the powers on the modes are the square singular values.

In this communication, we study a model of transmitter channel knowledge that includes both a non-zero channel mean and a transmit correlation matrix. Channel information of only transmit correlation has been studied in [7], where the optimal precoder matrix has mode directions given by the eigenvectors of the correlation matrix, and the mode powers are obtained by water-filling over the eigenvalues of the correlation matrix. The inclusion of the channel mean matrix changes the solution significantly in the sense that, now the mode directions depend on both the mean and correlation matrices and the SNR, and they can not be pre-determined. If the channel condition and the operating SNR are such that mode-dropping is required, the mode directions and powers change dynamically through each "water-fill" iteration. This gives raise to the term "dynamic water-filling".

The paper is organized as follows: In the next Section we set up the channel and signal models, and discuss the pair-wise error probability (PEP) criterion. Section 3 introduces the precoder and the corresponding optimization problem. In section 4, we solve for the precoder solution in both cases of full-rank precoder (no mode dropping) and precoder with mode dropping, and explain the concept of dynamic water-filling. Section 5 presents some numerical performance examples which illustrate the gain obtained by channel precoding. We give some concluding remarks in Section 6.

2. MODEL AND PEP CRITERION

2.1. Channel and signal model

We consider a MIMO wireless communication system with N transmit and M receive antennas. The channel is frequency flat quasi-static fading which is represented by matrix $\mathbf{H}$ of size $M \times N$. Assuming a non-zero mean channel with transmit antenna correlation, the channel matrix can be written in the form

$$\mathbf{H} = \mathbf{H}_m + \mathbf{H}_w \mathbf{R}_t^{1/2} , \quad (1)$$

where $\mathbf{H}_m$ is the channel mean, $\mathbf{R}_t$ is the transmit correlation matrix, and $\mathbf{H}_w$ has elements which are independent zero-mean complex Gaussian random variables with unit variance, i.e. $\mathbf{H}_w \in$

Mai Vu would like to acknowledge the support by the Rambus Stanford Graduate Fellowship.

0-7803-8521-7/04/$20.00 © 2004 IEEE

$\mathcal{N}(0, \mathbf{I})$. We assume that the transmit correlation matrix $\mathbf{R}_t$ is full rank and therefore is invertible.

The receiver is assumed to know the channel perfectly (i.e. it knows the channel realization $\mathbf{H}$), whereas the transmitter only knows the channel mean $\mathbf{H}_m$ and transmit correlation $\mathbf{R}_t$.

Let $\mathbf{X}$ be the transmit signal block over T symbols, then the receive signal block is

$$\mathbf{Y} = \mathbf{H}\mathbf{X} + \mathbf{V} \ ,$$

where $\mathbf{V} \in \mathcal{N}(0, \mathbf{I}\sigma^2)$ is the additive complex white Gaussian noise with σ^2 being the noise power. $\mathbf{X}$ can be a codeblock or a signaling block over which we perform detection. Note that there is no channel error correction code involved in the system under study, and we are interested in the performance in term of the uncoded block error rate.

2.2. PEP Criterion

We consider the pair-wise error probability (PEP), which is the probability that a transmitted signal block $\mathbf{X}$ is erroneously decoded as a signal block $\hat{\mathbf{X}}$. While this is not the system block error rate, it is a measurement that is strongly related to the system performance.

We assume ML decoder using the Euclidean distance decoding metric

$$\hat{\mathbf{X}} = \arg\min_{\mathbf{X}\in\mathcal{X}} ||\mathbf{Y} - \mathbf{H}\mathbf{X}||_F^2 \ .$$

The subscript F here denotes the Frobenius norm. Applying the Chernoff bound, similar to [1], the PEP can be upper bounded by

$$P(\mathbf{X} \to \hat{\mathbf{X}}) \leq \exp\left(-\frac{||\mathbf{H}(\mathbf{X} - \hat{\mathbf{X}})||_F^2}{4\sigma^2}\right) \ . \quad (2)$$

Taking the expectation of (2) over the channel statistics, we obtain the following bound on the average PEP

$$\bar{P}_e \leq \frac{\exp\left[\mathrm{tr}(\mathbf{H}_m\mathbf{W}^{-1}\mathbf{H}_m^*)\right]}{\det(\mathbf{W})^M} \det(\mathbf{R}_t)^M \exp\left[-\mathrm{tr}(\mathbf{H}_m\mathbf{R}_t^{-1}\mathbf{H}_m^*)\right], \quad (3)$$

where

$$\mathbf{W} = \mathbf{R}_t\frac{\Delta\Delta^*}{4\sigma^2}\mathbf{R}_t + \mathbf{R}_t \quad (4)$$

and $\Delta = \mathbf{X} - \hat{\mathbf{X}}$ is the codeword difference matrix. The superscript $(.)^*$ denotes the matrix conjugate transpose operation. Since the exact error expression is very complex, we will aim to minimize this upper bound on the average PEP (3). It is equivalent to minimizing the logarithm of the bound, and by ignoring the constant terms, this leads to the following objective function

$$J(\mathbf{W}) = \mathrm{tr}(\mathbf{H}_m\mathbf{W}^{-1}\mathbf{H}_m^*) - M\log\det(\mathbf{W}). \quad (5)$$

This objective function is matrix convex in $\mathbf{W}$.

3. LINEAR CHANNEL PRECODING IN SPACE-TIME CODED SYSTEMS

3.1. General linear precoding structure

We study a MIMO system where there have been a space-time block code (STBC) with an appropriate code rate in place to capture diversity in the channel. The STBC is designed assuming no knowledge of the channel at the transmitter, i.e. it targets channels with zero-mean i.i.d. complex Gaussian fading coefficients. To capture the channel information at the transmitter, we use a linear precoding matrix $\mathbf{F}$, connected in concatenation with the STBC as depicted in Figure 1. This gives the flexibility of adapting to various transmit channel knowledge conditions without changing the STBC that is already implemented. To maintain the total average transmit power, the precoding matrix needs to satisfy the power constraint

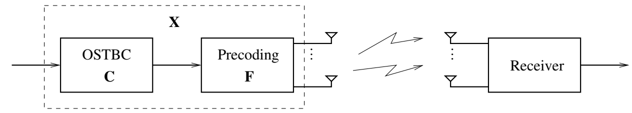

Fig. 1. MIMO channel precoding in OSTBC systems.

$$\mathrm{tr}(\mathbf{F}\mathbf{F}^*) = 1. \quad (6)$$

Let $\mathbf{C}$ be the codeblock in the STBC, the transmit signal block $\mathbf{X}$ then has the following structure

$$\mathbf{X} = \mathbf{F}\mathbf{C}.$$

The codeword difference matrix becomes

$$\Delta = \mathbf{X} - \hat{\mathbf{X}} = \mathbf{F}(\mathbf{C} - \hat{\mathbf{C}}).$$

In general, the codeword difference matrix depends on both the precoder and the specific codeword pair.

3.2. Linear precoder with OSTBC

In this analysis we will assume the use of orthogonal STBC [2] specifically. Due to orthogonality, we have

$$(\mathbf{C} - \hat{\mathbf{C}})(\mathbf{C} - \hat{\mathbf{C}})^* = \lambda P\mathbf{I},$$

where P is the average total transmit power and λ represents the codeword distance which depends on the specific pair of codewords.

Let λ_0 be the minimum distance over all pairs of codewords. The minimum codeword distance is the term that dominates the error probability exponent and hence is a reasonable indicator to the system performance. We aim to minimize this worst case average pair-wise error probability. In the later simulations, it is verified that the error bound obtained via minimizing the worst case average PEP has an almost constant gap to the actual error rate curve, hence validating the choice for the optimization objective.

Using the minimum distance codeword pair, then $\Delta\Delta^* = \lambda_0 P\mathbf{F}\mathbf{F}^*$, and the expression for $\mathbf{W}$ in (4) becomes a function of the precoding matrix only

$$\mathbf{W} = \frac{\lambda_0 P}{4\sigma^2}\mathbf{R}_t\mathbf{F}\mathbf{F}^*\mathbf{R}_t + \mathbf{R}_t \ . \quad (7)$$

With the objective function (5) and the power constraint on the precoder (6), we arrive at the following convex optimization problem

$$\min_{\mathbf{W}} \quad J = \mathrm{tr}(\mathbf{H}_m\mathbf{W}^{-1}\mathbf{H}_m^*) - M\log\det(\mathbf{W}) \quad (8)$$
$$\text{s.t.} \quad \mathrm{tr}(\mathbf{R}_t^{-1}\mathbf{W}\mathbf{R}_t^{-1} - \mathbf{R}_t^{-1}) = \tfrac{1}{4}\lambda_0\rho$$

where $\rho = P/\sigma^2$ is the SNR. Note that this problem in term of $\mathbf{F}$ is non-convex, but by the change of variable to $\mathbf{W}$, the problem becomes convex and can be solved analytically.

4. OPTIMAL LINEAR PRECODER

Problem (8) can be solved analytically using the Lagrange multiplier technique [10]. Form the Lagrangian function and differentiate with respect to $\mathbf{W}$ to get

$$-\mathbf{W}^{-1}\mathbf{H}_m^*\mathbf{H}_m\mathbf{W}^{-1} - M\mathbf{W}^{-1} + \nu\mathbf{R}_t^{-2} = 0 \ ,$$

where ν is the Lagrange multiplier. Define

$$\mathbf{\Psi} = M^2\mathbf{I}_N + 4\nu\mathbf{R}_t^{-1}\mathbf{H}_m^*\mathbf{H}_m\mathbf{R}_t^{-1} \ , \qquad (9)$$

then the solution for $\mathbf{W}$ is given as

$$\mathbf{W} = \frac{1}{2\nu}\mathbf{R}_t\Big(M\mathbf{I}_N + \mathbf{\Psi}^{\frac{1}{2}}\Big)\mathbf{R}_t \ .$$

We will need to find the Lagrange multiplier ν so that the precoder satisfies the transmit power constraint (6).

There is another implicit constraint on the positive semi-definite (PSD) property of $\mathbf{FF}^*$, which translates to the following condition on $\mathbf{W}$

$$\mathbf{R}_t^{-1}\mathbf{W}\mathbf{R}_t^{-1} - \mathbf{R}_t^{-1} = \mathbf{FF}^* \geq 0.$$

We will first solve for ν without taking into account this PSD constraint. If the solution does not satisfy the constraint, mode-droppings are then required, which means that the precoder will not distribute power in certain directions due to weak channel condition. This mode-dropping idea is analogous to the water-filling process but the actual implementation is different and will be discussed in the following sections.

4.1. Full-rank precoder solution

In this section, we solve for the Lagrange multiplier ν using the power constraint (6), which is equivalent to

$$\mathrm{tr}\Big(\frac{1}{2\nu}(M\mathbf{I}_N + \mathbf{\Psi}^{\frac{1}{2}}) - \mathbf{R}_t^{-1}\Big) = \frac{1}{4}\lambda_0\rho \ . \qquad (10)$$

Let λ_i ($i = 1 \dots N$) be the eigenvalues of $\mathbf{R}_t^{-1}\mathbf{H}_m^*\mathbf{H}_m\mathbf{R}_t^{-1}$, and let $\beta = 2[\mathrm{tr}(\mathbf{R}_t^{-1}) + \lambda_0\rho/4]$, the above equation becomes

$$MN + \sum_{i=1}^{N}\sqrt{M^2 + 4\nu\lambda_i} = \beta\nu \ . \qquad (11)$$

In the general case, this equation does not appear to have a close-form solution. However, solving for ν can be done efficiently using binary search, which we call the *inner algorithm*.

Inner algorithm for solving ν

The following lower and upper bounds on ν are established

$$\nu_{\text{lower}} = \frac{MN}{\beta} \ , \qquad \nu_{\text{upper}} = \frac{1}{\beta^2}\big(4N\sum_{i=1}^{N}\lambda_i + 2\beta MN\big) \ .$$

The lower bound is obtained directly from (11) by ignoring the summation of square-roots term, while the upper bound is obtained from applying the Cauchy-Schwartz inequality to the summation term. Binary numerical search can then be carried out to find solution for (11) between these bounds up to a desired precision. The number of iterations depends on the problem parameters, but usually the convergence happens very fast since this is an one dimensional binary search.

Precoder solution

Once ν is found, we can form the matrix $\mathbf{\Psi}$ as in (9) and establish the precoding matrix product as

$$\mathbf{FF}^* = \frac{1}{2\nu}(M\mathbf{I}_N + \mathbf{\Psi}^{\frac{1}{2}}) - \mathbf{R}_t^{-1} \ . \qquad (12)$$

From this product expression, an optimal precoder can be derived. The optimal precoder is not unique. In term of singular value decomposition of the precoder matrix, the left singular vectors and its singular values are the eigenvectors and square root of the eigenvalues of $\mathbf{FF}^*$ respectively. The right singular vectors however can be any unitary matrix. This linear precoder represents the idea of a multi-modal beamformer, where the beam directions (modes) are the eigenvectors of $\mathbf{FF}^*$, and the power on each mode is the corresponding eigenvalue of $\mathbf{FF}^*$.

4.2. Precoder solution with mode-dropping

For expression (12) to be a valid precoding solution, it has to be positive semidefinite (PSD). If the SNR is weak such that this expression is not PSD, we need to drop the weakest mode of $\mathbf{FF}^*$ and solve for ν again. This is analogous to the water-filling process. Now the total power will be distributed on the $N-1$ largest eigenvalues of $\mathbf{FF}^*$ and the power constraint (10) changes to

$$\sum_{i=2}^{N}\lambda_i\Big(\frac{1}{2\nu}(M\mathbf{I}_N + \mathbf{\Psi}^{1/2}) - \mathbf{R}_t^{-1}\Big) = \frac{1}{4}\lambda_0\rho \ , \qquad (13)$$

where $\lambda_i(\cdot)$ is the ith eigenvalue of the matrix in the brackets, sorted in increasing order ($\lambda_1 \leq \cdots \leq \lambda_N$). The sum of any k largest eigenvalues of a Hermitian matrix is convex. Again the above equation does not have a close form solution for ν. We design an algorithm to numerically solve for ν efficiently in this case based on two-fold binary search, which is termed the *outer algorithm*.

Outer algorithm for solving ν

There is no explicit function that relates the individual eigenvalue in the expression in (13) to ν. Fortunately, again we can derive upper bound and lower bound values on ν and then use binary search to find the solution efficiently. Using inequalities on eigenvalues of a sum of Hermitian matrices [9], we obtain the following bounds on the left-hand-side expression of (13)

$$f_{\text{upper}} = \frac{M(N-k)}{2\nu} + \sum_{i=k+1}^{N}\frac{\sqrt{M^2+4\nu\lambda_i}}{2\nu} - \sum_{i=1}^{N-k}\lambda_i(\mathbf{R}_t^{-1})$$

$$f_{\text{lower}} = \frac{M(N-k)}{2\nu} + \sum_{i=k+1}^{N}\frac{\sqrt{M^2+4\nu\lambda_i}}{2\nu} + k\lambda_1(\mathbf{R}_t^{-1}) - \mathrm{tr}(\mathbf{R}_t^{-1})$$

Here the bounds are given for the general case when k modes are dropped. Equating each expression to $\rho\lambda_0/4$ and solve for ν, using the inner algorithm mentioned in the previous section. The solutions of these two equations then become the upper bound and lower bound on the solution for ν in (13). Noting that the sum of eigenvalues is monotonous in ν, a binary search can then be carried out to solve equation (13) efficiently up to a desired numerical precision.

Dynamic water-filling

The mode-dropping process above is similar to the water-filling process in that at each outer iteration, a mode is dropped (the weakest amongst the active modes) and the total transmit power is redistributed over the rest of the modes. There is however a significant difference between this and the conventional water-filling process. In conventional water-filling process, the mode directions (i.e. the eigenvectors of $\mathbf{F}\mathbf{F}^*$) remain the same over the water-fill iterations, and only the water level changes after each iteration. In our problem, the mode directions also change at each iteration. This is due to the interaction between the channel mean and transmit correlation matrices. To see this effect more clearly, rewrite the expression for $\mathbf{F}\mathbf{F}^*$ in the following form

$$\mathbf{F}\mathbf{F}^* = \frac{M}{2\nu}\mathbf{I}_N + \left(\frac{1}{2\nu}\mathbf{\Psi}(\nu)^{\frac{1}{2}} - \mathbf{R}_t^{-1}\right), \qquad (14)$$

where the notation $\mathbf{\Psi}(\nu)$ illustrates the dependence of $\mathbf{\Psi}$ on ν. The total power (which is normalized to one in this case) is distributed over the positive eigenvalues of $\mathbf{F}\mathbf{F}^*$, and all the other eigenvalues are set to zero. In expression (14), the "water-level" is given by $M/2\nu$ and the mode directions are determined by the eigenvectors of the expression in the bracket. Thus when ν changes at each iteration, both the water-level and the mode directions change. For this reason, we call this a "dynamic water-filling" process.

5. NUMERICAL RESULTS

In this section we present examples of some system performance via simulation. The first system is a 2×2 MIMO using Alamouti SBTC. The channel mean and correlation matrices are generated arbitrarily and the optimal precoder is obtained using the procedure outlined in the previous section. The system performance with and without the channel precoding is shown in Figure 2 for QPSK modulation. It shows a performance gain in the system using precoder. The exact precoding gain depends on the channel parameters and the error rate of interest (or the operating SNR). At the block error rate of 10% for example, the gain is 2.5dB. At high SNRs, the gain gets smaller. The same trend has been observed on capacity gain of transmitter channel knowledge, which shows that the advantage of having channel knowledge at the transmitter diminishes at high SNRs for system with equal number of transmit and receive antennas [5]. Also shown are the worst-case PEP error bound that was used as the optimization criterion. It can be seen that the bounds track the actual performance closely with an almost constant gap, hence validating the optimization objective.

In Figure 3, we shows the error performance of a 4×1 system using rate 1 OSTBC with BPSK modulation [2]. The precoding gain is larger than the previous 2×2 system in Figure 2. For example, the gain at 10% block error rate is 4dB in this case. Generally the precoding gain increases with a larger number of transmit antennas.

Another point to mention here is that the gain of the precoder over non-precoded systems, while depending on the operating SNR, also depends on the channel parameters, that is the specific mean and transmit correlation matrices. Some channel condition favors precoding (i.e. the precoder can achieve a large performance gain) while for some others, the precoding gain is less significant. There can be various factors contributing to this effect, for example, how closely the eigenvectors of the channel mean and transmit correlation matrices align. This is a subject of further research.

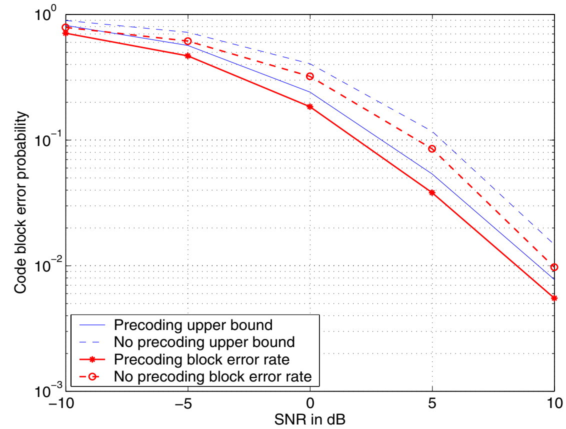

Fig. 2. Performance of a 2×2 system using Alamouti code and QPSK modulation.

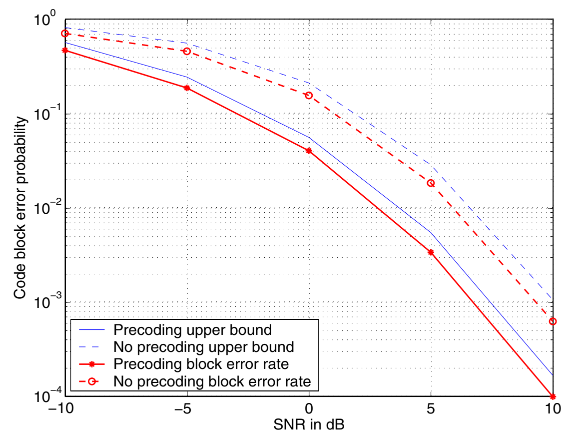

Fig. 3. Performance of a 4×1 system using rate 1 OSTBC and BPSK modulation.

6. CONCLUSION

We have derived analytically the optimal solution for a linear precoder that exploits the channel mean and transmit correlation information at the transmitter. The optimal precoder is a function of both the channel mean and transmit correlation matrices, and the operating SNR. If no mode-dropping is required, the precoder matrix is full rank and distributes power in all directions. The calculation of the precoder matrix requires the computation of a Lagrange multiplier, which can be found efficiently via a binary search. When mode dropping is needed, a two-fold binary search algorithm is proposed to calculate the Lagrange multiplier and arrive at the precoder solution efficiently. The mode-dropping process resembles the water-filling principle. The difference here is that both the water-level and the mode directions change with each iteration and hence, it is called dynamic water-filling. Examples of performance gain using the precoder are given via simulation for 2×2 and 4×1 systems. It shows that significant gains can be obtained by the precoder. The exact precoder gain depends on the channel parameters and the operating SNR. The gain generally increases with an increasing number of transmit and/or receive antennas.

REFERENCES

[1] V.Tarokh, N.Seshadri, .R.Calderbank, *"Space-Time Codes for High Data Rate Wireless Communication: Performance Criterion and Code Construction"*, IEEE Trans. on Info. Theory, Vol. 44, No. 2, pp 744-765, Mar 1998.

[2] V.Tarokh, H.Jafakhani and A.R.Calderbank, *"Space-Time Block Codes from Orthogonal Designs"*, IEEE Trans. on Info. Theory, Vol. 45, No. 5, pp. 1456-1467, July 1999.

[3] D.Gesbert, H.Blcskei, D.A. Gore, and A.J.Paulraj, *"Outdoor MIMO Wireless Channels: Models and Performance Prediction"*, IEEE Trans. on Comm., Vol. 50, No. 12, pp. 1926-1934, Dec 2002.

[4] D.S.Baum, et. al., *"Measurement and Characterization of Broadband MIMO Fixed Wireless Channels at 2.5GHz"*, Int. Conf. Per. Wireless Comm., pp.203-206, Dec 2000.

[5] M.Vu and A.Paulraj, *"Some asymptotic capacity results for MIMO wireless with and without channel knowledge at the transmitter"*, Proc. 37th Asilomar Conf. Sig., Sys. and Comp., pp. 258-262, Nov. 2003.

[6] G.Jongren, M.Skoglund and B.Ottersten, *"Combining Beamforming and Orthogonal Space-Time Block Coding"*, IEEE Trans. on Info. Theory, Vol 48, pp. 611-627, Mar 2002.

[7] H.Sampath and A.Paulraj, *"Linear Precoding for Space-Time Coded Systems with Known Fading Correlations"*, IEEE Comm. Lett., Vol. 6, pp. 239-241, June 2002.

[8] M.Vu and A.Paulraj, *"Optimum Transmission Scheme for a MISO Wireless System with Partial Channel Knowledge and Infinite K factor"*, Proc. ICC, June 2004.

[9] R.A.Horn and C.R.Johnson, *"Matrix Analysis"*, Cambridge University Press, 1985.

[10] S.Boyd and L.Vandenberghe, *"Convex Optimization"*, Cambridge University Press 2003. Avail. at http://www.stanford.edu/ boyd/cvxbook.html.