This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the ICC 2008 proceedings.

Optimal Precoding for Digital Subscriber Lines

Fernando Pérez-Cruz
Department of Electrical Engineering
Engineering Quadrangle
Princeton University
Princeton, New Jersey 08544
Email: fp@princeton.edu

Miguel R. D. Rodrigues
Instituto de Telecomunicações
Computer Science Department
University of Porto
4169 - 007 Porto (Portugal)
Email: mrodrigues@dcc.fc.up.pt

Sergio Verdú
Department of Electrical Engineering
Engineering Quadrangle
Princeton University
Princeton, New Jersey 08544
Email: verdu@princeton.edu

***Abstract*—We determine the linear precoding policy that maximizes the mutual information for general multiple-input multiple-output (MIMO) Gaussian channels with arbitrary input distributions, by capitalizing on the relationship between mutual information and minimum mean squared error (MMSE). The optimal linear precoder can be computed by means of a fixed-point equation as a function of the channel and the input constellation. We show that diagonalizing the channel matrix does not maximize the information transmission rate for nonGaussian inputs. A full precoding matrix may significantly increase the information transmission rate, even for parallel non-interacting channels. We illustrate the application of our results to typical Gigabit DSL systems.**

I. Introduction

Digital subscriber line (DSL) technology divides the telephone line into parallel non-interacting subchannels with flat frequency response [1]. A capacity-achieving strategy allocates power to each subchannel following the well-known waterfilling algorithm [2] and uses Gaussian inputs on each subchannel. To accommodate for practical structures, such as PSK and QAM input distributions, several authors have proposed a modification of the waterfilling algorithm, where a power gap is added to the base level prior to the water-pouring phase [3]-[6].

Recently proposed Gigabit DSL systems [7] use all the available copper wires in the last distribution-area (pedestal-drop segment) [8]. These additional copper-wire pairs allow using MIMO diversity techniques in each DSL subchannel, boosting the transmission rate, whose ultimately limit is the capacity found in [9]. To maximize the information transmission rate with nonGaussian inputs, Lee et al. [7] propose to diagonalize each MIMO subchannel, amend the base level using a constant power gap and allocate power using the waterfilling algorithm.

In [10] the power allocation for parallel channels with arbitrary inputs is revisited exploiting the mutual information and the minimum mean squared error (MMSE) relation [11]. The mercury/waterfilling algorithm proposed in [10] can be interpreted as a generalization of the waterfilling power-allocation policy, in which the base level is modified by adding mercury to account for the suboptimal input distribution. The mercury level depends on the gain of every subchannel as well as on the input distributions. Thus, the mercury level amends the base level in a very different way from power gap approaches, thereby significantly increasing the transmission rate for multitone DSL systems [10].

The results in [10] illustrate that capacity-achieving strategies for Gaussian inputs may be quite suboptimal for discrete input constellations. The natural belief is that since the channels are noninterfering, to maximize mutual information is sufficient to take into account the nonGaussiannes of the input constellations by means of the mercury/waterfilling power allocation. This paper challenges this natural belief, and shows that linear precoding techniques that introduce correlation among the channel inputs may achieve substantially higher information transmission rates.

The optimization of a linear precoder and equalizer for MIMO or mutually interfering channels has been typically addressed from the MMSE viewpoint [12]-[16], with the optimal solution being the diagonalization of the channel matrix. In [17], a unifying approach using different criteria, such as the MMSE, the signal to interference-plus-noise ratio (SINR) and the bit error rate (BER), leads to identical results. We propose instead to optimize a linear precoder to maximize the input-output mutual information. We do not restrict the optimization of the precoder to diagonal channel matrices and we derive the optimal linear precoder for general (not necessarily diagonal) channel matrices. The solution is expressed as a fixed-point equation in terms of the relation between the MMSE and mutual information [11], [18]. In contrast to the situation for Gaussian inputs [19], [20], the optimal linear precoder for arbitrary inputs does not diagonalize the channel and, in particular, it does not reduce to a diagonal matrix for parallel non-interfering channels. We also show that in the regime of high snr the precoder that minimizes the MMSE and BER asymptotically maximizes mutual information.

The rest of the paper is organized as follows. We obtain the optimal precoding matrix in Section II. The asymptotic regime of high SNR is considered in Section III, and Section IV applies the results to typical DSL noninterfering channels.

II. Optimal precoding policy

Consider the deterministic vector channel complex-valued model:

$$\mathbf{y} = \sqrt{\mathsf{snr}}\mathbf{H}\mathbf{P}\mathbf{x} + \mathbf{w} \quad (1)$$

where the n-dimensional vector $\mathbf{y}$ and the m-dimensional vector $\mathbf{x}$ represent, respectively, the received vector and the independent zero-mean unit-variance transmitted information vector. The distributions of the components of $\mathbf{x}$ are fixed, and not necessarily Gaussian or equal for different dimensions. The $n \times m$ complex-valued matrix $\mathbf{H}$ corresponds to the deterministic channel gains (known to both encoder and decoder) and $\mathbf{w}$ is the n-dimensional complex Gaussian noise with independent zero-mean unit-variance components[1]. The snr is a scaling factor that accounts for the total transmitted power. The optimization of the mutual information is carried out over all precoding matrices $\mathbf{P}$ that do not increase the transmitted power. The precoding problem can be cast as a constrained nonlinear optimization problem:

$$\max_{\mathbf{P}} I(\mathbf{x};\mathbf{y}) \tag{2}$$

subject to:

$$\mathsf{Tr}\left\{\mathbb{E}[\mathbf{P}\mathbf{x}\mathbf{x}^\dagger\mathbf{P}^\dagger]\right\} = \mathsf{Tr}\{\mathbf{P}\mathbf{P}^\dagger\} \leq m \tag{3}$$

Theorem 1: The optimum precoding matrix $\mathbf{P}^*$ that solves (2) subject to (3) satisfies:

$$\mathbf{P}^* = \lambda^{-1}\mathbf{H}^\dagger\mathbf{H}\mathbf{P}^*\mathbf{E} \tag{4}$$

with $\lambda = \|\mathbf{H}^\dagger\mathbf{H}\mathbf{P}^*\mathbf{E}\|/\sqrt{m}$. The matrix

$$\mathbf{E} = \mathbb{E}\left[\left(\mathbf{x} - \mathbb{E}\left[\mathbf{x}|\mathbf{y}\right]\right)\left(\mathbf{x} - \mathbb{E}\left[\mathbf{x}|\mathbf{y}\right]\right)^\dagger\right] \tag{5}$$

is known as the MMSE matrix [21][2].

The proof relies on the KKT conditions [23] and the relation between the mutual information and the MMSE [11], [18]. The KKT conditions are satisfied by any critical point (minimum, maximum or saddle point). There is a unique $\mathbf{P}^*$ that satisfies the KKT conditions when the problem is strictly concave, corresponding to the global maximum. In general, the power allocation as stated in (2)-(3) is nonconcave. It becomes concave in some specific cases, e.g. for Gaussian input distributions with arbitrary channel matrices [24]-[19]; it is also concave for low snr.

Consequently, to obtain a global optimum for any snr we can first obtain the unique solution for a low enough snr to ensure that the problem is concave. Then, we increase the snr by small amounts until the desired snr is achieved. For each increased snr value, a new precoding matrix is obtained, using as starting point the precoding matrix for the previous snr. Alternatively we could also choose several starting points for $\mathbf{P}$, find a local maxima, and keep the precoding matrix with largest mutual information.

The optimal precoder is given by the fixed-point equation (4). For discrete constellations, the MMSE matrix cannot be computed analytically, because we need to sum over all the possible transmitted vectors, which grows exponentially with m. For large dimensions we typically estimate the MMSE matrix $\mathbf{E}$ using Monte Carlo methods and we use the following iterative procedure to determine the optimal precoding matrix:

$$\mathbf{P}_{k+1} = \lambda^{-1}(\mathbf{P}_k + \mu\mathsf{snr}\mathbf{H}^\dagger\mathbf{H}\mathbf{P}_k\mathbf{E})$$

where μ is a small constant, the MMSE matrix depends on $\mathbf{P}_k$ and λ^{-1} ensures that $\mathsf{Tr}\{\mathbf{P}_{k+1}\mathbf{P}_{k+1}^\dagger\} = m$.

[1]It is also possible to consider non-identity noise covariance matrices, since we could add a whitening filter at the receiver that only affects the definition of the channel matrix.

[2]Full proofs of all the results can be found in [22].

III. High snr approximation for discrete inputs

We now consider the optimal precoding policy for MIMO Gaussian channels with arbitrary discrete input distributions in the high snr regime. We prove upper and lower bounds to the MMSE, and consequently, upper and lower bounds to the mutual information by exploiting the relationship between mutual information and MMSE [11], [18]. Then we use the upper and lower bounds to derive the form of the optimal policy. These results rest upon Theorem 3 in [10], where the MMSE for single-input single-output Gaussian channels with QPSK input distribution is expanded for large snr. These results represent a generalization and tightening of the bound in Theorem 4 in [10], since we consider general channel matrices, rather than diagonal ones.

Theorem 2: For the channel model in (1) we can bound its MMSE as:

$$\frac{1}{M}\frac{e^{-d_{min}^2\mathsf{snr}/4}}{d_{min}\sqrt{\mathsf{snr}}}\left(\sqrt{\pi} - \frac{4.37}{d_{min}^2\mathsf{snr}}\right) \leq \mathsf{mmse}(\mathsf{snr}) \leq (M-1)\frac{e^{-d_{min}^2\mathsf{snr}/4}}{d_{min}\sqrt{\mathsf{snr}}}\sqrt{\pi} \tag{6}$$

where

$$\mathsf{mmse}(\mathsf{snr}) = \mathbb{E}\left[||\mathbf{H}\mathbf{P}\mathbf{x} - \mathbf{H}\mathbf{P}\mathbb{E}[\mathbf{x}|\mathbf{y}]||^2\right] \tag{7}$$

and M is the product of the constellations cardinality and d_{min} is the minimum distance in the received lattice:

$$d_{min} = \min_{\substack{\overline{\mathbf{x}},\mathbf{x} \\ \mathbf{x}\neq\overline{\mathbf{x}}}} ||\mathbf{H}\mathbf{P}\mathbf{x} - \mathbf{H}\mathbf{P}\overline{\mathbf{x}}|| \tag{8}$$

The lower bound is proven assuming there is only a pair of points at minimum distance and the upper bound is proven assuming that every pair of points is at minimum distance. In most practical cases, these bounds are loose because for each received point there are several pairs of points at minimum distance, so we can approximate the MMSE for large snr by:

$$\mathsf{mmse}(\mathsf{snr}) \approx \pi\frac{K}{M}Q\left(d_{min}\sqrt{\mathsf{snr}/2}\right) \tag{9}$$

where K is the number of pairs of points at minimum distance. This expression is π times the symbol error rate (SER) of the received constellation. Therefore minimizing the MMSE for large snr is equivalent to minimizing the SER for discrete inputs.

We can use the upper and lower bound of the MMSE and the relation between the MMSE and the mutual information to bound the mutual information for large snr and discrete input distributions.

Theorem 3: The mutual information between $\mathbf{x}$ and $\mathbf{y}$ for the channel model in (1) can be bounded by:

$$\log M - \frac{2(M-1)\sqrt{\pi}}{d_{min}^3\sqrt{\mathsf{snr}}}e^{-d_{min}^2\mathsf{snr}/4} \leq I(\mathbf{x};\mathbf{y}) \leq \log M - \frac{2e^{-d_{min}^2\mathsf{snr}/4}}{Md_{min}^3\sqrt{\mathsf{snr}}}\left(\sqrt{\pi} - \frac{4.37+2\sqrt{\pi}}{d_{min}^2\mathsf{snr}}\right) \quad (10)$$

As we did for the MMSE, we can also approximate the mutual information for large snr as:

$$I(\mathbf{x};\mathbf{y}) \approx \log M - 2\frac{K\pi}{Md_{min}^2}Q\left(d_{min}\sqrt{\mathsf{snr}/2}\right) \quad (11)$$

These results show that for a MIMO channel model with discrete inputs, the precoding matrix $\mathbf{P}$ that maximizes the information transmission rate corresponds to the matrix that maximizes the minimum distance in the received lattice, in the high snr regime. This matrix also ensures the minimization of the MMSE and the SER. This result is a direct generalization of a result in [10] for non-interfering channels. This is an appealing result as it tells us that for discrete inputs in high snr regime, minimizing the symbol error rate is equivalent to both minimizing the mean squared error and maximizing the mutual information, thereby linking three standard criteria typically used for power allocation in the all-important case of discrete input distributions.

In many communication systems, diagonal channel models are typically encountered or sought [20]-[28] with diagonal power allocation strategies commonly proposed to maximize the mutual information. Power allocation for arbitrary inputs is often sought based on the fact that power-allocation strategies are capacity achieving for parallel independent channels with independent Gaussian inputs [2].

We can illustrate with a simple example why linear precoders achieve larger information transmission rates than diagonal power allocation matrices. Let us assume a simple real-valued communication system with a non-interfering channel matrix

$$\mathbf{H} = \begin{bmatrix}\sqrt{3} & 0\\ 0 & 1\end{bmatrix}, \quad (12)$$

in which both inputs are BPSK distributions. We first set $\mathbf{P} = \mathbf{I}$ and the minimum distance in the received constellation is 2. The optimal power allocation matrix corresponding to the mercury/waterfilling [10] solution is:

$$\mathbf{P} = \begin{bmatrix}1/\sqrt{2} & 0\\ 0 & \sqrt{3/2}\end{bmatrix} \quad (13)$$

and the minimum distance increases to $\sqrt{6}$. We can further compute the optimal linear precoder:

$$\mathbf{P} = \begin{bmatrix}1/\sqrt{2} & 1/\sqrt{2}\\ -1/\sqrt{2} & 1/\sqrt{2}\end{bmatrix}, \quad (14)$$

and obtain the largest minimum distance of $2\sqrt{2}$. In Figure 1 we have plotted the constellations for the optimal precoder and power allocation matrices to highlight the difference between the received constellations. The difference between the transmission rates for precoding and power allocation strategies are quite substantial, as illustrated in the next section.

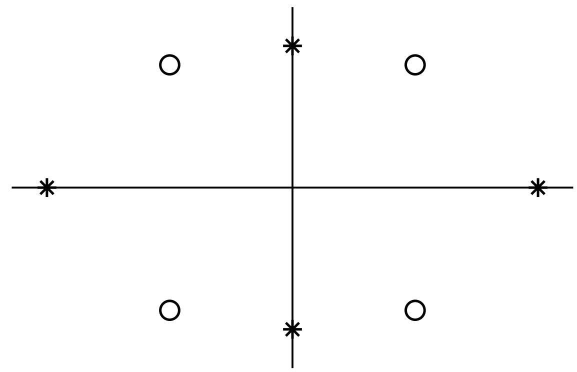

Fig. 1. '$\circ$' represents the received constellation for the diagonal power allocation solution and '$*$' represents the output constellation achieved with the optimum linear precoder.

IV. APPLICATION: GIGABIT DSL

In this section we show the benefit of using a full precoding matrix instead of a diagonal power allocation matrix for maximizing the mutual information in a MIMO communication channel with Gaussian noise when the inputs are QPSK. We first deal with a diagonal channel matrix, encountered in many different scenarios as detailed in [20]-[28]. We compare the information transmission rate obtained using the optimal mercury/waterfilling power allocation proposed in [10], corresponding to a diagonal real-valued matrix $\mathbf{P}$, to a non-diagonal precoding matrix. The results show that even for diagonal channel matrices a full precoding matrix significantly increases the information transmission rate in the communication channel, compared to a diagonal power allocation matrix.

In the second example we use a full four-by-four channel matrix. This matrix is typically encountered in each subchannel of Gigabit DSL systems [7]. However, rather than performing diagonalization of the channel matrix followed by waterfilling power allocation over the subchannels, as proposed in [7], we directly optimize the precoding matrix which significantly increases the input-output mutual information.

A. Diagonal channel matrix

For our first example we have used the following diagonal channel matrix:

$$\mathbf{H} = \begin{bmatrix}2e^{-j2.51} & 0 & 0 & 0 & 0\\ 0 & 1.5e^{j1.60} & 0 & 0 & 0\\ 0 & 0 & e^{-j0.79} & 0 & 0\\ 0 & 0 & 0 & 0.8e^{-j3.05} & 0\\ 0 & 0 & 0 & 0 & 0.6e^{-j1.67}\end{bmatrix} \quad (15)$$

and QPSK inputs. In this matrix the ratio between the norm of the largest and smallest element is less than 4.

For this diagonal channel model, we have computed the optimal power-allocation policy corresponding to a diagonal real-valued $\mathbf{P}$ [10] (denoted as MWF) and the optimal non-diagonal precoding matrix. We depicted the mutual information for these precoding matrices in Figure 2, together with the waterfilling solution for Gaussian inputs, which serves as an upper bound to the mutual information. It can be seen that the mutual information for the full precoding matrix is substantially higher than that for the diagonal power allocation matrix, which is the best power-allocation matrix as shown in [10], yielding gains in snr higher than 2dB. The lower bound in Figure 2 corresponds to the precoder that maximizes the minimum distance. It represents a lower bound, because maximum minimum distance precoders are equivalent to maximum mutual information precoders only for infinite snr. However, the maximum minimum distance precoder achieves substantially higher information transmission rates than the optimal power allocation, for the snr range of practical interest even when the high snr assumption no longer holds.

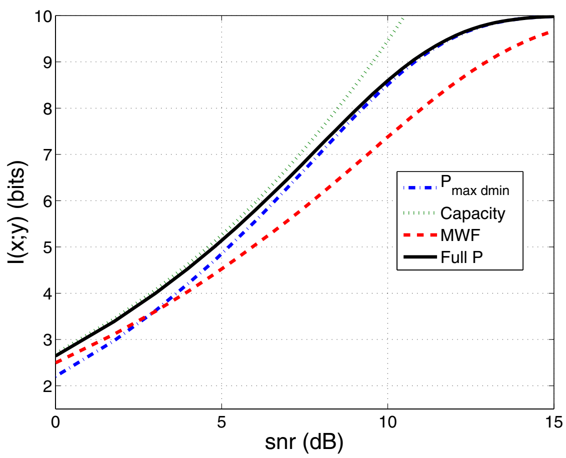

Fig. 2. Information transmission rates for the 5×5 matrix in (15).

As the snr increases, the optimal diagonal power-allocation matrix tends to:

$$\mathbf{P}_1 = \begin{bmatrix} 0.20 & 0 & 0 & 0 & 0 \\ 0 & 0.27 & 0 & 0 & 0 \\ 0 & 0 & 0.41 & 0 & 0 \\ 0 & 0 & 0 & 0.51 & 0 \\ 0 & 0 & 0 & 0 & 0.68 \end{bmatrix}$$

while the full precoding matrix tends to:

$$\mathbf{P}_2 = \begin{bmatrix} 0.28e^{j2.5} & 0.26e^{j1.8} & 0.27e^{-j0.1} & 0.16e^{j1.5} & 0.23e^{j1.8} \\ 0.09e^{j1.9} & 0.21e^{j0.9} & 0.16e^{-j3.1} & 0.34e^{-j2.5} & 0.25e^{j2.7} \\ 0.26e^{j0.7} & 0.09e^{-j2.6} & 0.15e^{j1.6} & 0.17e^{j0.1} & 0.27e^{j3.0} \\ 0.07e^{j0.6} & 0.23e^{j1.4} & 0.21e^{j1.8} & 0.02e^{-j0.1} & 0.17e^{-j0.7} \\ 0.18e^{j0.8} & 0.17e^{-j2.9} & 0.15e^{-j0.5} & 0.21e^{-j1.0} & 0.12e^{-j1.0} \end{bmatrix}$$

For the diagonal power-allocation matrix, the assigned power is inversely proportional to norm of the subchannel. Hence, this policy for power allocation ensures that $\mathbf{HP}_1 = c\mathbf{I}$ and all the subchannels are equally powerful at the receiver end.

The non-diagonal precoding matrix approximately assigns the same power for all inputs, as the norms of the columns of $\mathbf{P}_2$ are almost identical. However, more power is assigned to the stronger channels, as the norm of the rows of $\mathbf{P}_2$ decreases with the channel gain. Therefore, using a full precoding matrix, we are able significantly increase the mutual information for the snr range of interest.

B. Full channel matrix: The Gigabit DSL scenario

In this second example, we illustrate for a nondiagonal channel matrix $\mathbf{H}$ that a full precoder provides substantially higher information transmission rates than standard channel diagonalization followed by mercury/waterfilling power allocation. We consider a situation encountered in Gigabit DSL systems with four copper-wire pairs typically available in the last distribution area [8]. In particular, we use the following channel matrix:

$$\mathbf{H} = \begin{bmatrix} e^{j2.80} & 0.5e^{-j0.97} & 0.01e^{-j1.02} & 10^{-3}e^{-j1.06} \\ 0.5e^{-j0.43} & e^{-j2.98} & 0.1e^{j0.12} & 0.01e^{-j3.12} \\ 0.01e^{j2.93} & 0.1e^{j1.27} & e^{-j2.00} & 0.5e^{-j1.83} \\ 10^{-3}e^{j2.63} & 0.01e^{-j2.06} & 0.5e^{j2.38} & e^{j2.69} \end{bmatrix} \quad (16)$$

in which we have assume that each channel has unit gain and it is interfered by the other pairs.

In Figure 3, we show the mutual information as a function of the snr for the optimal precoding as well as for the optimal power allocation policy using the mercury/waterfilling algorithm [10] performed after diagonalization of the channel matrix (denoted as MWF). As in the previous example, the full precoder significantly increases the information transmission rate with respect to the optimal power allocation strategy.

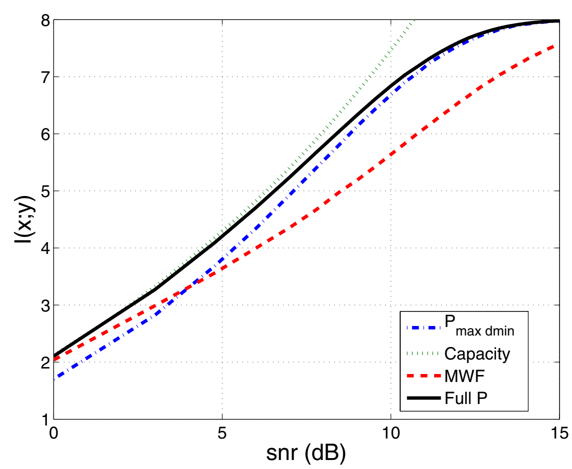

Fig. 3. Information transmission rates for the 4×4 matrix in (16).

V. Conclusions and Further Work

In this paper we have shown that when the inputs are not Gaussian, information transmission rates are maximized when a full complex-valued precoder is used instead of a power allocation strategy coupled to channel diagonalization. The optimal precoder matrix is expressed using the relation between the MMSE and the mutual information and it is computed as a fixed-point equation. We have also shown that for asymptotically high snr the optimal precoder matrix for discrete inputs achieves maximal minimum distance and minimal mean squared error, linking three common criteria for designing digital communication systems.

The complexity of the proposed approach has not been addressed in this paper. For example, a 256 subchannel DSL system would need to optimize 65536 elements in the precoder matrix $\mathbf{P}$. To reduce the complexity of optimizing a full $\mathbf{P}$ we could constraint it to be block diagonal, Toeplitz or unitary. Hence reducing the number of components we need to optimize to obtain the optimal precoder, but reducing the possible gains from using a full precoder. For block diagonal or Toeplitz $\mathbf{P}$ matrices, we would also need to sort the channels in a way that we ensure a maximal information transmission rate.

ACKNOWLEDGMENT

We are sincerely grateful to Albert Gillén for his comments and suggestions.

Fernando Pérez-Cruz is supported by AI-COM Marie Curie Fellowship from the 6^{th} Research Framework Programme of the European Union. Fernando Pérez-Cruz is also a faculty member at the Signal Theory and Communications Department in Carlos III University (Spain).

This work was partially funded through collaborative participation in the Communications and Networks Consortium sponsored by the U.S. Army Research Laboratory under the Collaborative Technology Alliance Program, Cooperative Agreement DAAD19-01-2-0011. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation thereon.

REFERENCES

[1] J. Lechleider, "High bit rate digital subscriber lines: A review of HDSL progress," *IEEE Journal on Selected Areas in Communzcations*, vol. 9, no. 6, pp. 769–784, 8 1991.

[2] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York, USA: Wiley, 1991.

[3] J. A. C. Bingham, P. S. Chow, and J. M. Cioffi, "A practical discrete multitone transceiver loading algorithm for data transmission over spectrally shaped channels," *IEEE Transactions on Communications*, vol. 43, no. 2/3/4, pp. 773–775, Feb./Mar./Apr. 1995.

[4] G. D. Forney and G. Ungerboeck, "Modulation and coding for linear Gaussian channels," *IEEE Transactions on Information Thoery*, vol. 44, no. 6, pp. 2384–2415, 10 1998.

[5] B. S. Krongold, K. Ramchandran, and D. L. Jones, "Computationally efficient optimal power allocation algorithms for multicarrier communication systems," *IEEE Transactions on Communications*, vol. 48, no. 1, pp. 23–27, 1 2000.

[6] T. Starr, J. M. Cioffi, and P. J. Silverman, *Understanding Digital Subscriber Line Technology*. Prentice-Hall, 1999.

[7] B. Lee, J. Cioffi, S. Jagannathan, and M. Mohseni, "Gigabit DSL," *IEEE Transactions on Communications*, p. Accepted, 2007.

[8] J. Cioffi, S. Jagannathan, M. Mohseni, and G. Ginis, "CuPON: The copper alternative to PON 100gb/s DSL networks," *IEEE Communications Magazine*, vol. 45, no. 6, pp. 132–139, 6 2007.

[9] L. H. Brandenburg and A. D. Wyner, "Capacity of the Gaussian channel with memory: the multivariate case," *Bell Sys. Tech. J.*, vol. 53, no. 5, pp. 745–778, May-June 1974.

[10] A. Lozano, A. M. Tulino, and S. Verdú, "Optimum power allocation for parallel Gaussian channels with arbitrary input distributions," *IEEE Transactions on Information Theory*, vol. 52, no. 7, pp. 3033–3051, 7 2006.

[11] D. Guo, S. Shamai, and S. Verdú, "Mutual information and minimum mean-square error in Gaussian channels," *IEEE Transactions on Information Theory*, vol. 51, no. 4, pp. 1261–1282, 4 2005.

[12] K. H. Lee and D. P. Petersen, "Optimal linear coding for vector channels," *IEEE Transactions on Communications*, vol. 24, pp. 1283–1290, 12 1976.

[13] A. Scaglione, G. B. Giannakis, and S. Barbarossa, "Redundant filterbank precoders and equalizers part I: Unification and optimal designs," *IEEE Transactions on Signal Processing*, vol. 47, no. 7, pp. 1988–2006, 7 1999.

[14] A. Scaglione, P. Stoica, S. Barbarossa, G. B. Giannakis, and H. Sampath, "Optimal designs for space-time linear precoders and decoders," *IEEE Transactions on Signal Processing*, vol. 50, no. 5, pp. 1051–1064, 5 2002.

[15] J. Yang and S. Roy, "On joint transmitter and receiver optimization formultiple-input-multiple-output (MIMO) transmission systems," *IEEE Transactions on Communications*, vol. 42, pp. 3221–3231, 12 1994.

[16] ——, "Joint transmitter-receiver optimization for multiinput multi-output systems with decision feedback," *IEEE Transactions on Information Thoery*, vol. 40, pp. 1334–1347, 9 1994.

[17] D. P. Palomar, J. M. Cioffi, and M. A. Lagunas, "Joint tx-rx beamforming design for multicarrier mimo channels: A unified framework for convex optimization," *IEEE Transactions on Signal Processing*, vol. 51, no. 9, pp. 2381–2401, 9 2003.

[18] D. P. Palomar and S. Verdú, "Gradient of mutual information in linear vector Gaussian channels," *IEEE Transactions on Information Theory*, vol. 52, no. 1, pp. 141–154, 1 2006.

[19] W. Yu, W. Rhee, S. Boyd, and J. M.Cioffi, "Iterative water-filling for Gaussian vector multiple-access channels," *IEEE Transactions on Information Theory*, vol. 50, no. 1, pp. 145–152, 1 2004.

[20] I. E. Telatar, "Capacity of multiantenna Gaussian channels," *European Transactions on Communications*, vol. 10, no. 6, pp. 585–595, 6 1999.

[21] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1993.

[22] F. Pérez-Cruz, M. R. D. Rodrigues, and S. Verdú, "Optimal precoding for multiple-input multiple-output Gaussian channels with arbitrary inputs," *preprint*, 2008.

[23] R. Fletcher, *Practical Methods of Optimization*, 2nd ed. New York: Wiley, 1987.

[24] R. G. Gallager, *Information Theory and Reliable Communication*. New York, USA: Wiley, 1968.

[25] B. S. Tsybakov, "Capacity of a discrete-time Gaussian channel with a filter," *Problems of Information Transmission*, vol. 6, no. 3, pp. 253–256, July-September 1970.

[26] J. A. C. Bingham, "Multicarrier modulation for data transmission: An idea whose time has come," *IEEE Communication Magazine*, vol. 28, no. 5, pp. 5–14, 5 1990.

[27] G. Raleigh and J. M. Cioffi, "Spatio-temporal coding for wireless communications," *IEEE Transactions on Communications*, vol. 46, no. 3, pp. 357–366, 3 1998.

[28] A. J. Goldsmith and P. Varaiya, "Capacity of fading channels with channel side information," *IEEE Transactions on Information Thoery*, vol. 43, no. 6, pp. 1986–1992, 11 1997.