Optimal Transmit-Diversity Precoders for Random Fading Channels†

Georgios B. Giannakis and Shengli Zhou
Dept. of ECE, Univ. of Minnesota
200 Union Street SE, Minneapolis, MN 55455
Emails: {georgios,szhou}@ece.umn.edu

Abstract— **Optimal transmitter-receiver designs obeying the water-filling principle are well-known and widely applied when the propagation channel is deterministically known and frequently updated at the transmitter. Because the latter may be costly or impossible to acquire in rapidly varying wireless environments, we develop in this paper statistical water-filling-like criteria for stationary random fading channels. The resulting optimal designs require only knowledge of the channel's correlation matrix that does not require frequent updates and can be easily acquired. Applied to a multiple transmit-antenna paradigm, the optimal precoder and loading algorithm outperform not only the conventional equal-power allocation across all antennas but also their deterministic water-filling counterparts in fast fading scenarios.**

I. Introduction

Antenna diversity is well motivated for wireless communications through fading channels. In many cases however (e.g., in cellular downlink), receive antennas may be expensive or impractical, which endeavors diversity gain through multiple transmit antennas [11], [10], [4]. The diversity gain at the transmitter reaches within 0.1dB of that at the receiver with the same number of antennas [11].

With deterministically known channel state information (CSI) at the transmitter, optimal transmitter design based on the water filling principle [3] has been proposed in [1],[7],[8]. However, when the channel is fast fading, it is costly yet not accurate to acquire CSI at the transmitter and the optimal design based on previously acquired information becomes outdated quickly. Therefore, it is meaningful to design optimal transmitters by modeling the channel explicitly as a stationary random process.

So long as the channel remains stationary, it has invariant statistics. Through field measurements, the transmitter can acquire such statistical CSI a priori. Alternatively, the receiver can estimate channel statistics and feed them back to the transmitter on line. In some applications such as Time Division Duplex (TDD) systems, the transmitter can obtain channel statistics directly since the forward and backward channels share the same physical channel during different time slots. Based on channel covariance information, optimal transmitter design is proposed in [2] to minimize the system Symbol Error Rate (SER). However, [2] is only applicable to a fixed constellation (BPSK), a fixed modulation (differential encoding), and specific random channels of known probability density function (Rayleigh p.d.f).

†This work was supported by NSF Wireless Initiative grant no. 99-79443.

In this paper, we formulate general criteria to design optimal transmitter precoders based only on the channel's second-order statistics. The resulting optimal precoders are thus applicable to any constellation, any modulation and channel types (e.g., Rayleigh, Rician, Nakagami [6]).

This paper is organized as follows: Section II describes the discrete time baseband equivalent system model and Section III develops optimal precoders derived under two different criteria. Performance is analyzed in Section IV with numerical results. Conclusions are then drawn on Section V.

II. System model

Fig. 1 depicts the block diagram of a transmit diversity system with L transmit antennas. In the ith ($i \in [1, L]$)

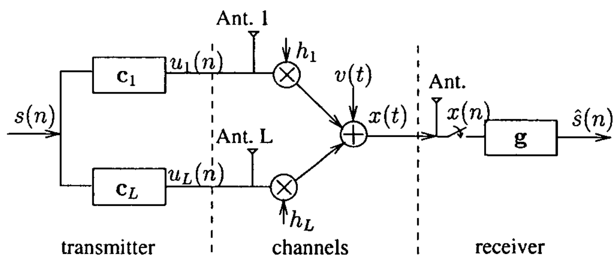

Fig. 1. discrete time model

branch, the information-bearing signal $s(n)$ is first spread by the code $\mathbf{c}_i := [c_i(0), \dots, c_i(P-1)]^T$ of length P to obtain: $u_i(n) = \sum_{k=-\infty}^{\infty} s(k)c_i(n-kP)$. After pulse shaped by filter $\varphi(t)$ (not shown in Fig. 1), the continuous signal $u_i(t) = \sum_{n=\infty}^{\infty} u(n)\varphi(t-nT_c)$ is transmitted through the ith antenna, where T_c is the chip duration. We here assume that the channels are flat faded (frequency non-selective) with complex fading coefficients h_i, $i = 1, \dots, L$. The received signal in the presence of additive Gaussian noise $v(t)$ is then given by $x(t) = \sum_{i=1}^{L} h_i u_i(t) + v(t)$. After chip level filtering with $\bar{\varphi}(t)$ that is matched to $\varphi(t)$, $x(t)$ is sampled at $t = nT_c$ to yield the discrete time signal $x(n) := x(t)|_{t=nT_c}$. Selecting $\varphi(t)$ to possess the Nyquist-T_c property avoids inter symbol interference, and allows us to express $x(n)$ as:

$$x(n) = \sum_{k=-\infty}^{\infty} \sum_{i=1}^{L} h_i s(k) c_i(n-kP) + v(n), \qquad (1)$$

where $v(n) := v(t)|_{t=nT_c}$. To cast (1) into a convenient matrix-vector form, we let ^{T} denote transpose and define the $P \times 1$ vectors $\mathbf{x}(n) := [x(nP+0), \ldots, x(nP+P-1]^T$, and $\mathbf{v}(n) := [v(nP+0), \ldots, v(np+P-1)]^T$; the $L \times 1$ vector $\mathbf{h} := [h_1, \ldots, h_L]^T$, and the $P \times L$ code matrix $\mathbf{C} := [\mathbf{c}_1, \ldots, \mathbf{c}_L]$. We then can rewrite (1) as $\mathbf{x}(n) = \mathbf{C}\mathbf{h}s(n) + \mathbf{v}(n)$. Because we only focus on symbol by symbol detection, we omit the time index n and subsequently obtain:

$$\mathbf{x} = \mathbf{C}\mathbf{h}s + \mathbf{v}. \tag{2}$$

At the receiver, the channel estimate $\hat{\mathbf{h}}$ is obtained first to enable maximum ratio combining (MRC) using

$$\hat{\mathbf{g}}_{opt} := [g(0), \ldots, g(P-1)]^T = \mathbf{C}\hat{\mathbf{h}}. \tag{3}$$

MRC is known to maximize the signal to noise ratio (SNR) [6] at its output that yields the symbol estimate: $\hat{s} = \hat{\mathbf{g}}_{opt}^{\mathcal{H}}\mathbf{x} = \hat{\mathbf{h}}^{\mathcal{H}}\mathbf{C}\mathbf{x}$, where $(\cdot)^{\mathcal{H}}$ denotes Hermitian transpose.

Given a precoder $\mathbf{C}$, (3) specifies the optimal receiver $\mathbf{g}$ in the sense of maximizing output SNR. It also shows that the effectiveness of MRC hinges critically on the quality of the channel estimate $\hat{\mathbf{h}}$; i.e., the better $\hat{\mathbf{h}}$ is, the better $\hat{\mathbf{g}}_{opt}$ will perform. The question that arises is how to select the precoder $\mathbf{C}$. Once the goodness criterion of $\hat{\mathbf{h}}$ has been specified, we propose to optimize it with respect to $\mathbf{C}$ in order to obtain the optimum precoder.

III. Optimal precoder designs

For simplicity, in this paper we adopt the following assumptions:

a0) channel $\mathbf{h}$ and noise $\mathbf{v}$ are uncorrelated; i.e., $\mathrm{E}[\mathbf{h}\mathbf{v}^{\mathcal{H}}] = \mathbf{0}$, where $\mathbf{0}$ denotes the all-zero matrix.

a1) noise $\mathbf{v}$ is white; i.e., $\mathbf{R}_{vv} := \mathrm{E}[\mathbf{v}\mathbf{v}^{\mathcal{H}}] = \sigma_v^2\mathbf{I}$, where $\mathbf{I}$ denotes identity matrix.

To obtain the optimal transceivers $(\mathbf{C}, \mathbf{g})$ for random channels, one may be tempted to minimize the mean square error of symbol estimates: $\mathrm{E}\left[|s-\hat{s}|^2\right]$, with respect to all possible channel realizations. However, it turns out that minimizing $\mathrm{E}\left[|s-\hat{s}|^2\right]$ only yields a trivial solution for random channels (recall that [8] models the channel as deterministic). Once the precoder $\mathbf{C}$ is specified, the optimal receiver is given by (3). Therefore, the more accurate $\hat{\mathbf{h}}$ is, the better the overall system performance. This observation motivates us to search for an optimal $\mathbf{C}$ that minimizes channel estimation error variance.

A. Mean-Square Channel Error Criterion

Letting $\mathbf{R}_{hx} := \mathrm{E}[\mathbf{h}\mathbf{x}^{\mathcal{H}}]$ and $\mathbf{R}_{xx} := \mathrm{E}[\mathbf{x}\mathbf{x}^{\mathcal{H}}]$, the linear MMSE estimator for $\mathbf{h}$ given $\mathbf{x}$ is:

$$\hat{\mathbf{h}} = \mathrm{E}[\mathbf{h}] + \mathbf{R}_{hx}\mathbf{R}_{xx}^{-1}(\mathbf{x} - \mathrm{E}[\mathbf{x}]), \tag{4}$$

and has a variance matrix:

$$\mathrm{E}\left[|\hat{\mathbf{h}} - \mathbf{h}|^2\right] = \mathbf{R}_{hh} - \mathbf{R}_{hx}\mathbf{R}_{xx}^{-1}\mathbf{R}_{hx}^{\mathcal{H}}. \tag{5}$$

Without loss of generality, we set $\mathbf{h}$ and $\mathbf{x}$ to have zero mean in our subsequent derivation. Indeed, if $\mathrm{E}[\mathbf{h}] \neq 0$, $\mathrm{E}[\mathbf{x}] \neq 0$, it suffices to substitute $\mathbf{h}, \mathbf{x}$ by their zero mean counterparts: $\mathbf{h}_0 := \mathbf{h} - \mathrm{E}[\mathbf{h}]$ and $\mathbf{x}_0 := \mathbf{x} - \mathrm{E}[\mathbf{x}]$, respectively. Simplifying (4) to $\hat{\mathbf{h}} = \mathbf{R}_{hx}\mathbf{R}_{xx}^{-1}\mathbf{x}$, we find using a0) that $\mathbf{R}_{hx} = \mathbf{R}_{hh}\mathbf{C}^{\mathcal{H}}s^*$, $\mathbf{R}_{xx} = |s|^2\mathbf{C}\mathbf{R}_{hh}\mathbf{C}^{\mathcal{H}} + \mathbf{R}_{vv}$, which allow us to write (4) as:

$$\hat{\mathbf{h}} = \mathbf{R}_{hh}\mathbf{C}^{\mathcal{H}}s^*\left(|s|^2\mathbf{C}\mathbf{R}_{hh}\mathbf{C}^{\mathcal{H}} + \mathbf{R}_{vv}\right)^{-1}\mathbf{x}. \tag{6}$$

We will select precoder $\mathbf{C}$ to minimize the mean-square error of the channel estimator [c.f. (4),(5)]:

$$\begin{aligned}\mathcal{E}(\mathbf{C}) &= \mathrm{tr}\left\{\mathbf{R}_{hh} - \mathbf{R}_{hx}\mathbf{R}_{xx}^{-1}\mathbf{R}_{hx}^{\mathcal{H}}\right\} \\ &= \mathrm{tr}\left\{\mathbf{R}_{hh} - |s|^2\mathbf{R}_{hh}\mathbf{C}^{\mathcal{H}}(|s|^2\mathbf{C}\mathbf{R}_{hh}\mathbf{C}^{\mathcal{H}} + \mathbf{R}_{vv})^{-1}\mathbf{C}\mathbf{R}_{hh}\right\} \\ &= \mathrm{tr}\left\{(\mathbf{R}_{hh}^{-1} + |s|^2\mathbf{C}^{\mathcal{H}}\mathbf{R}_{vv}^{-1}\mathbf{C})^{-1}\right\},\end{aligned} \tag{7}$$

where the last step results from the matrix inversion lemma [5, p. 565] and $\mathrm{tr}\{\cdot\}$ denotes the trace operator. Without any constraint, minimizing $\mathcal{E}$ leads to the trivial solution that requires infinite power to be transmitted ($\|\mathbf{C}\| = \infty$). A reasonable constraint that takes into account limited budget resources is the transmitted power, which is expressed as $\mathcal{P}_0 = \mathrm{tr}\{(s\mathbf{C})^{\mathcal{H}}(s\mathbf{C})\} = \mathrm{tr}\left\{|s|^2\mathbf{C}^{\mathcal{H}}\mathbf{C}\right\}$. With the transmit-power constraint, our objective becomes:

$$\min_{\mathbf{C}} \mathcal{E}(\mathbf{C}) \quad \text{subject to} \quad \mathcal{C} := \mathrm{tr}\left\{|s|^2\mathbf{C}^{\mathcal{H}}\mathbf{C}\right\} - \mathcal{P}_0 = 0. \tag{8}$$

To simplify $\mathcal{E}$ in (7), we diagonalize $\mathbf{R}_{hh}$ using its spectral decomposition:

$$\mathbf{R}_{hh} = \mathbf{U}\mathbf{D}_h\mathbf{U}^{\mathcal{H}}, \quad \mathbf{D}_h = \mathrm{diag}(\lambda_{11}, \ldots, \lambda_{LL}), \tag{9}$$

where $\mathbf{U}$ is unitary, $\lambda_{ii} \geq 0$ denotes the ith eigenvalue of $\mathbf{R}_{hh}$, and $\mathrm{diag}(\cdot)$ stands for a diagonal matrix with specified diagonal entries. If the h_i's are uncorrelated, then $\lambda_{ii} = \mathrm{E}\left[|h_i|^2\right]$. Based on (9) and a1), we can rewrite (7) as:

$$\begin{aligned}\mathcal{E}(C) &= \mathrm{tr}\left\{(\mathbf{U}\mathbf{D}_h^{-1}\mathbf{U}^{\mathcal{H}} + \frac{|s|^2}{\sigma_v^2}\mathbf{C}^{\mathcal{H}}\mathbf{C})^{-1}\right\} \\ &= \mathrm{tr}\left\{(\mathbf{D}_h^{-1} + \frac{|s|^2}{\sigma_v^2}\mathbf{U}^{\mathcal{H}}\mathbf{C}^{\mathcal{H}}\mathbf{C}\mathbf{U})^{-1}\right\} \\ &:= \mathrm{tr}\left\{(\mathbf{D}_h^{-1} + \frac{|s|^2}{\sigma_v^2}\mathbf{F}^{\mathcal{H}}\mathbf{F})^{-1}\right\},\end{aligned} \tag{10}$$

where $\mathbf{F} := \mathbf{C}\mathbf{U}$ and $\mathbf{C} = \mathbf{F}\mathbf{U}^{\mathcal{H}}$ is uniquely determined by $\mathbf{F}$. Because $\mathbf{U}$ is unitary, the power constraint is equivalently written as

$$\mathrm{tr}\left\{|s|^2\mathbf{F}^{\mathcal{H}}\mathbf{F}\right\} = \mathcal{P}_0. \tag{11}$$

Equations (10) and (11) imply that only the eigenvalues of $\mathbf{F}^{\mathcal{H}}\mathbf{F}$, denoted by $f_{11}^2, \ldots, f_{LL}^2$, need to be designed in our constrained minimization problem. Specifically, (8) can be rewritten as:

$$\begin{aligned}&\min_{\mathbf{F}} \mathcal{E}(\mathbf{F}) = \sum_{i=1}^{L}\left(\frac{1}{\lambda_{ii}} + \frac{|s|^2}{\sigma_v^2}f_{ii}^2\right)^{-1} \\ &\quad\text{subject to} \quad \mathcal{C} := |s|^2\sum_{i=1}^{L}f_{ii}^2 - \mathcal{P}_0 = 0.\end{aligned} \tag{12}$$

Applying the Lagrange multiplier method, we form the Lagrangian

$$\mathcal{E} + \mu\mathcal{C} = \sum_{i=1}^{L}(\frac{1}{\lambda_{ii}} + \frac{|s|^2}{\sigma_v^2} f_{ii}^2)^{-1} + \mu(|s|^2 \sum_{i=1}^{L} f_{ii}^2 - \mathcal{P}_0), \quad (13)$$

and take derivatives with respect to f_{ii}^2 to obtain:

$$\frac{\partial(\mathcal{E} + \mu\mathcal{C})}{\partial f_{ii}^2} = \frac{-|s|^2}{\sigma_v^2}(\frac{1}{\lambda_{ii}} + \frac{|s|^2}{\sigma_v^2} f_{ii}^2)^{-2} + \mu |s|^2 = 0. \quad (14)$$

Solving for f_{ii}^2 from (14), we find:

$$f_{ii}^2 = \frac{\sqrt{\sigma_v^2}}{|s|^2 \sqrt{\mu}} - \frac{\sigma_v^2}{|s|^2 \lambda_{ii}}, \quad (15)$$

and plugging f_{ii}^2 back to the power constraint in (12), we obtain μ as:

$$\sqrt{\frac{\sigma_v^2}{\mu}} = \frac{1}{L}(\mathcal{P}_0 + \sum_{i=1}^{L} \frac{\sigma_v^2}{\lambda_{ii}}). \quad (16)$$

Substituting (16) into (15), we reach the optimal $\mathbf{F}$ with its eigenvalues given by:

$$f_{ii}^2 = \frac{1}{|s|^2}\left(\frac{\mathcal{P}_0}{L} + \frac{1}{L}\sum_{i=1}^{L}\frac{\sigma_v^2}{\lambda_{ii}} - \frac{\sigma_v^2}{\lambda_{ii}}\right). \quad (17)$$

The only constraint on matrix $\mathbf{F}$ (thus $\mathbf{C}$) so far is that $\mathbf{F}^{\mathcal{H}}\mathbf{F}$ should have eigenvalues as in (17). Without affecting the constrained optimization in (8) or (12), we can assume that $\mathbf{F}^{\mathcal{H}}\mathbf{F}$ is a diagonal matrix:

$$\mathbf{F}^{\mathcal{H}}\mathbf{F} = \mathbf{D}_f^2, \quad \text{where } \mathbf{D}_f := \text{diag}(f_{11}, \cdots, f_{LL}), \quad f_{ii} \geq 0, \forall i \in [1, L]. \quad (18)$$

However, if $\mathbf{h}$ is also complex Gaussian distributed with zero mean (Rayleigh) or non zero mean (Rician), maximization of mutual information between the received vector $\mathbf{x}$ and the channel $\mathbf{h}$ will constrain $\mathbf{F}^{\mathcal{H}}\mathbf{F}$ to have an exact diagonal form, as we describe next.

B. Conditional Mutual Information Criterion

In this section, we will further assume that:
a2) channel $\mathbf{h}$ is complex Gaussian distributed.
Recalling (2) and interchanging the roles of $\mathbf{h}$ and s let us view $\mathbf{h}$ as the input to the channel $s\mathbf{C}$. We then seek precoders $\mathbf{C}$ that minimize the mutual information $I(\mathbf{x}, \mathbf{h}|s)$ between the Gaussian input $\mathbf{h}$ and the output $\mathbf{x}$ conditioned on s. Mutual information between Gaussian vectors is well known (see e.g., [1, Thm. 1], [7], and references therein).

Lemma 1: *Consider the finite-dimensional vector model* $\mathbf{x} = s\mathbf{C}\mathbf{h} + \mathbf{v}$, *where* $\mathbf{h}$ *and* $\mathbf{x}$ *satisfy a0) and a1). The conditional mutual information between* $\mathbf{x}$ *and* $\mathbf{h}$, $I(\mathbf{x}, \mathbf{h}|s)$, *is maximized when* $\mathbf{h}$ *is Gaussian as per a2), and is given by* ($|\cdot|$ *denotes matrix determinant):*

$$I(\mathbf{x}, \mathbf{h}|s) = \log_2 \left|(\mathbf{R}_{hh}^{-1} + (s\mathbf{C})^{\mathcal{H}}\mathbf{R}_{vv}^{-1}(s\mathbf{C}))\mathbf{R}_{hh}\right|. \quad \square \quad (19)$$

Based on (9) and a1), we can simplify (19) to:

$$\begin{aligned} I(\mathbf{x}, \mathbf{h}|s) &= \log_2 \left|(\mathbf{U}\mathbf{D}_h^{-1}\mathbf{U}^{\mathcal{H}} + \mathbf{C}^{\mathcal{H}}\mathbf{C}\,|s|^2/\sigma_v^2)\mathbf{U}\mathbf{D}_h\mathbf{U}^{\mathcal{H}}\right| \\ &= \log_2 \left|(\mathbf{D}_h^{-1} + \mathbf{U}^{\mathcal{H}}\mathbf{C}^{\mathcal{H}}\mathbf{C}\mathbf{U}\,|s|^2/\sigma_v^2)\mathbf{D}_h\right| \\ &= \log_2 \left|(\mathbf{D}_h^{-1} + \mathbf{F}^{\mathcal{H}}\mathbf{F}\,|s|^2/\sigma_v^2)\mathbf{D}_h\right|. \end{aligned} \quad (20)$$

According to Hadamard's inequality [3, p. 502], maximum $I(\mathbf{x}, \mathbf{h}|s)$ is achieved when the matrix $(\mathbf{D}_h^{-1} + \mathbf{F}^{\mathcal{H}}\mathbf{F}\,|s|^2/\sigma_v^2)\mathbf{D}_h$ is diagonal. Therefore, $\mathbf{F}^{\mathcal{H}}\mathbf{F}$ should be exactly diagonal and thus it can be written as in (18).

We then maximize (20) under our transmit-power constraint as:

$$\max_{\mathbf{D}_f} I(\mathbf{x}, \mathbf{h}|s) = \log_2 \left|\mathbf{I} + \frac{|s|^2}{\sigma_v^2}\mathbf{D}_f^2\mathbf{D}_h\right| = \sum_{i=1}^{L}\log_2(1 + \frac{|s|^2}{\sigma_v^2} f_{ii}^2 \lambda_{ii})$$
$$\text{subject to } \mathcal{C} := |s|^2 \sum_{i=1}^{L} f_{ii}^2 - \mathcal{P}_0 = 0. \quad (21)$$

Differentiating the Lagrangian $I(\mathbf{x}, \mathbf{h}|s) + \mu\mathcal{C}$ with respect to f_{ii}^2, and equating it to zero, we obtain:

$$\frac{\partial I(\mathbf{x}, \mathbf{h}|s) + \mu\mathcal{C}}{\partial f_{ii}^2} = \frac{\lambda_{ii}|s|^2/\sigma_v^2}{\ln 2(1 + f_{ii}^2\lambda_{ii}|s|^2/\sigma_v^2)} + \mu|s|^2 = 0. \quad (22)$$

From (22), we can solve for

$$f_{ii}^2 = -1/(\mu|s|^2 \ln 2) - \sigma_v^2/(|s|^2 \lambda_{ii}), \quad (23)$$

and plug into (11) to obtain μ as:

$$-\frac{1}{\mu \ln 2} = \frac{1}{L}(\mathcal{P}_0 + \sum_i \frac{\sigma_v^2}{\lambda_{ii}}). \quad (24)$$

Substituting μ into (23), we arrive at the optimal loading:

$$f_{ii}^2 = \frac{1}{|s|^2}\left(\frac{\mathcal{P}_0}{L} + \frac{1}{L}\sum_{i=1}^{L}\frac{\sigma_v^2}{\lambda_{ii}} - \frac{\sigma_v^2}{\lambda_{ii}}\right), \quad (25)$$

which is surprisingly identical to (17) that has been derived under a different criterion. Therefore, objective (12) also leads to the maximum $I(\mathbf{x}, \mathbf{h}|s)$ in (21), under assumption a2).

C. Loading Algorithm

From Sections III-A and III-B, we see that the optimal $\mathbf{F}^{\mathcal{H}}\mathbf{F}$ should be the diagonal matrix $\mathbf{D}_f^2$, with elements f_{ii}^2 specified in (17) (or (25)). Diagonal $\mathbf{F}^{\mathcal{H}}\mathbf{F}$ implies that the columns of $\mathbf{F}$ are orthogonal and allows one to factor $\mathbf{F}$ and $\mathbf{C}$ as:

$$\mathbf{F} = \mathbf{\Phi}\mathbf{D}_f, \quad \mathbf{C} = \mathbf{\Phi}\mathbf{D}_f\mathbf{U}^{\mathcal{H}}, \quad (26)$$

where the columns of $\mathbf{\Phi}$ are orthonormal (note that $\mathbf{D}_f$ takes care of power loading).

Equation (26) provides a general optimal precoder for random channels for a given transmit-power budget. We summarize this result as follows:

Theorem 1: *Suppose a0) and a1) hold true and $\mathbf{R}_{hh}$ and $\mathcal{P}_0$ be available. The optimum receive-filter $\mathbf{g}_{opt}$ is given by* (3) *and the optimum precoding matrix $\mathbf{C}_{opt} = \mathbf{\Phi D}_f \mathbf{U}^{\mathcal{H}}$ has $\mathbf{U}$ and $\mathbf{D}_f$ formed as in* (9), (17) *and* (18) *with $\mathbf{\Phi}$ an arbitrary $P \times L$ matrix with orthonormal columns. Optimality in $\mathbf{g}_{opt}$ refers to maximum-SNR while optimality in $\mathbf{C}_{opt}$ pertains to either minimizing the random channels' estimation error, or, maximizing the conditional mutual information under a2).*

Note that [2] arrived at the same power loading as in (17) by minimizing the system SER. However, [2] is only applicable to differential BPSK under Rayleigh fading channels, in which case a simple closed-form expression of SER can be obtained. However, Theorem 1 holds for any constellation and adopted modulation. Under the channel MMSE criterion, Theorem 1 holds regardless of the channel p.d.f. Therefore, [2] falls into the general class of Theorem 1. On the other hand, we can infer from [2] that the optimal precoder $\mathbf{C}$ leads to minimum SER for differential BPSK as well.

The entry $f_{ii}^2 \geq 0$ in (17) imposes the following lower bound on $\mathcal{P}_0$:

$$\mathcal{P}_0 > \frac{L\sigma_v^2}{\min(\lambda_{ii})} - \sum_{i=1}^{L} \frac{\sigma_v^2}{\lambda_{ii}} := \bar{\mathcal{P}}_{th}. \qquad (27)$$

If $\mathcal{P}_0$ is not large enough to afford optimal power allocation, i.e., $\mathcal{P}_0 < \bar{\mathcal{P}}_{th}$, we will set the minimum f_{ii} to zero, and load power on the remaining $L-1$ summands in (17).

The power loading algorithm is summarized in the following steps:

1) Arrange λ_{ii} in decreasing order $\lambda_{11} \geq \lambda_{22} \cdots \geq \lambda_{LL}$. For $\bar{L} = 1, \ldots, L$, calculate $\bar{P}_{\bar{L}} := \bar{P}_{th}$ from (27) based only on the first $\bar{L}$ channel eigenvalues: $\lambda_{11}, \ldots, \lambda_{\bar{L}\bar{L}}$.

2) With the power budget $\mathcal{P}_0$ in the interval $[\bar{P}_i, \bar{P}_{i+1}]$, set $f_{i+1,i+1}, \ldots, f_{LL} = 0$, and obtain $f_{11}, \ldots, f_{ii}$ according to (17) based only on $\lambda_{11}, \ldots, \lambda_{ii}$.

We now examine several special cases of the general precoder form in (26).

Special Case 1: If the channel coefficients are uncorrelated, then $\mathbf{U} = \mathbf{I}$ and $\mathbf{C} = \mathbf{\Phi D}_f$; thus, any orthogonal matrix $\mathbf{\Phi}$ together with the optimal power loading matrix $\mathbf{D}_f$ are equivalent in terms of optimizing (8). This intuitive fact is widely applied in practice through orthogonal spreading sequences. To fully exploit diversity gains offered by L antennas, $P \geq L$ is required. However, we observe that the choice $P - L \geq 0$ gains nothing more in terms of optimizing of (8) (or (20)) than the minimum choice of $P = L$ which minimizes bandwidth requirements, or, increases information rate. The simplest $\mathbf{\Phi}$ will be the identity matrix, which in practice corresponds to allowing only one antenna transmission per time slot. If $f_{ii} = 0$, the ith physical antenna will be turned off to avoid power loss.

Special Case 2: If we equi-distribute power among all branches, then $\mathbf{D}_f = f_{11}\mathbf{I}$ and the precoder $\mathbf{C} = f_{11}\mathbf{\Phi U}^{\mathcal{H}}$ will have orthogonal columns with identical norms. The practical design with equal power allocated to each antenna and orthogonal spreading codes falls into this category, and results in a diagonal $\mathbf{F}^{\mathcal{H}}\mathbf{F}$ with identical diagonal elements.

Special Case 3: If the transmit power is high enough to have $\mathcal{P}_0 \gg \sigma_v^2/\lambda_{ii}, \forall i$, then $f_{ii}^2 \approx f_{11}^2, \forall i$. In this case, power is equally distributed to each branch and thus to each antenna. Therefore, equal power distribution is only optimal when the power is sufficiently high.

IV. Performance Analysis

To obtain a closed-form SER, we here assume that the channel estimates at the receiver are error-free. Since the received vector $\mathbf{x}$ can be expressed as $\mathbf{x} = \mathbf{\Phi D}_f \mathbf{U}^{\mathcal{H}}\mathbf{h}s + \mathbf{v}$, we can always multiply $\mathbf{x}$ with $\mathbf{\Phi}^{\mathcal{H}}$ to obtain:

$$\tilde{\mathbf{x}} := \mathbf{\Phi}^{\mathcal{H}}\mathbf{x} = \mathbf{D}_f \mathbf{U}^{\mathcal{H}}\mathbf{h}s + \mathbf{\Phi}^{\mathcal{H}}\mathbf{v} := \tilde{\mathbf{h}}s + \tilde{\mathbf{v}}, \qquad (28)$$

where $\tilde{\mathbf{h}}$ and $\tilde{\mathbf{v}}$ henceforth denote equivalent channel and noise vectors. Matrix $\mathbf{R}_{\tilde{h}\tilde{h}} = \mathbf{D}_f \mathbf{U}^{\mathcal{H}}\mathbf{R}_{hh}\mathbf{U}\mathbf{D}_f = \mathbf{D}_f^2 \mathbf{D}_h$ implies that the entries of $\tilde{\mathbf{h}}$ are independent, while $\tilde{\mathbf{v}}$ is still white since $\mathbf{R}_{\tilde{v}\tilde{v}} = \sigma_v^2\mathbf{I}$. For MRC symbol estimates $\hat{s} = \tilde{\mathbf{h}}^{\mathcal{H}}\tilde{\mathbf{x}}$, it is possible to obtain a closed form SER expression for MPSK signals [9, eq. (44)] as:

$$P_s(E) = \frac{1}{\pi}\int_0^{(M-1)\pi/M} \prod_{i=1}^{L} I_i(\lambda_{ii} f_{ii}^2/\sigma_v^2, \gamma_{PSK}, \theta) d\theta, \quad (29)$$

where $\gamma_{PSK} := \sin^2(\pi/M)$, and $I_i(x, \gamma_{PSK}, \theta)$ is the moment of the p.d.f of $\tilde{h}_i$ evaluated at $-\gamma_{PSK}/\sin^2(\theta)$ (see [9, eq. (24)]). SER for other constellations such as QAM can be easily carried out as in [9].

Unlike [2] that relies on differential BPSK signaling to obtain a simple SER closed form, our approach provides a general framework regardless of the constellation involved or the modulation adopted. Due to space limitations, we only present simulations for QPSK ($M = 4$) and adopt the same 3-channel set up as in [2], i.e., $\lambda_{11} = 1$, $\lambda_{22} = 0.05$, $\lambda_{33} = 0.01$ with normalized noise power $\sigma_v^2 = 1$. If h_i's are independent, the physical channel will have $\sigma_{h_i}^2 = \lambda_{ii}, \forall i$. However, this setting corresponds also to correlated channels with the same variance and correlation coefficients $\rho_{12} = \rho_{23} = 0.94$, and $\rho_{13} = 0.86$ [2].

We define the SNR as the total transmitted power divided by noise power: $\mathcal{P}_0/\sigma_v^2$. Fig. 2 shows the optimal power allocation among the different $\tilde{h}_i$'s. In low power, the transmitter prefers to null certain $\tilde{h}_i$'s, while it approximately equates power to all antennas in high power to benefit from diversity. The SER curve in Fig. 3 confirms that the optimal allocation outperforms the conventional equal power allocation as well as the selective power allocation which corresponds to simply transmitting on a few strong $\tilde{h}_i$'s.

Next, we check robustness of the loading algorithm with respect to finite-sample effects that introduce estimation error in the channel covariance. At the receiver (or at the transmitter in a TDD mode), $\mathbf{R}_{xx}$ can be estimated by the sample average $\hat{\mathbf{R}}_{xx} = \sum_{n=1}^{N} \mathbf{x}(n)\mathbf{x}(n)^{\mathcal{H}}/N$, from which we obtain

$$\hat{\mathbf{R}}_{hh} = \mathbf{C}^{\dagger}(\hat{\mathbf{R}}_{xx} - \sigma_v^2\mathbf{I})(\mathbf{C}^{\mathcal{H}})^{\dagger}, \qquad (30)$$

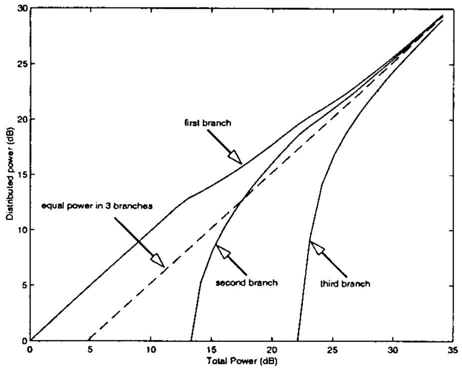

Fig. 2. Optimal loading vs. equal power allocation

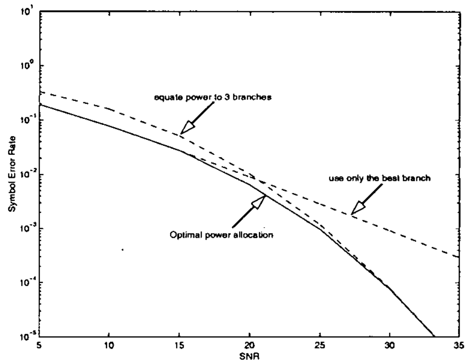

Fig. 3. Symbol Error Rate vs SNR

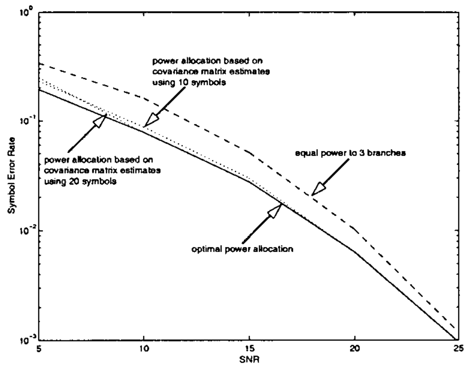

Fig. 4. Power allocation using $\hat{\mathbf{R}}_{hh}$

where $\dagger$ stands for matrix pseudo-inverse. At the beginning, we assume that the transmitter does not have CSI and employs equal power allocation to all 3 antennas. We then estimate the channel covariance matrix at the receiver using N symbols. This statistical CSI is then fed back to the transmitter to initialize power allocation. Here we use the same channel setting as before, and rely on the simplest precoding matrix $\mathbf{\Phi} = \mathbf{I}_{L\times L}$. To obtain SER in closed form, we again assume that the channel is known at the receiver and apply (29). In Fig. 4, we observe that the channel covariance estimates based on only $N = 10$ symbols lead to a system performance close to the ideal case (perfectly known $\mathbf{R}_{hh}$), and outperform the equal power loading considerably. As N increases to 20, the power allocation algorithm based on $\hat{\mathbf{R}}_{hh}$ is indistinguishable from the ideal case. Therefore, power allocation based on channel covariance estimates is robust to estimation errors. The latter implies that the optimal transceivers designed with statistical CSI offer an attractive choice in fast fading applications.

V. Conclusions

In this paper, we have proposed optimal precoder designs for multiple transmit-antennas and a single receive-antenna system based only on channel covariance information. The optimal designs are applicable to any signal constellation and adopted modulation in random fading channels of unknown p.d.f. In rapidly fading environments they outperform existing approaches that require exact channel information. Conventional equal power allocation falls into our general precoder class and is approximately optimal when the transmitted power goes to infinity. Simulation results confirm the superiority of optimal loading with respect to conventional equal power distribution among transmit-antennas.

Acknowledgments

The authors would like to thank professor M.-S. Alouini of the ECE Dept. at the Univ. of Minnesota for pointing their attention to [2] and [9].

References

[1] N. Al-Dhahir and J. M. Cioffi, "Block transmission over dispersive channels: Transmit filter optimization and realization, and MMSE-DFE receiver performance," *IEEE Trans. on Information Theory*, vol. 42, no. 1, pp. 137–160, Jan. 1996.

[2] J. K. Cavers, "Optimized use of diversity modes in transmitter diversity systems," in *Proc. of the VTC.*, 1999, pp. 1768–1773.

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

[4] J. C. Guey, M. P. Fitz, M. R. Bell, and W. Y. Kuo, "Signal design for transmitter diversity wireless communication systems over rayleigh fading channels," *IEEE Trans. on Comm.*, vol. 47, no. 4, pp. 527–537, Apr. 1999.

[5] S. Haykin, *Adaptive Filter Theory*, Prentice-Hall, Inc., 3rd edition, 1996.

[6] J. Proakis, *Digital Communications*, McGraw-Hill, 1989.

[7] A. Scaglione, S. Barbarossa, and G. B. Giannakis, "Filterbank transceivers optimizing information rate in block transmissions over dispersive channels," *IEEE Trans. on Information Theory*, vol. 45, Apr. 1999.

[8] A. Scaglione, G. B. Giannakis, and S. Barbarossa, "Redundant filterbank precoders and equalizers Part I: Unification and optimal designs," *IEEE Trans. on SP*, pp. 1988–2006, July 1999.

[9] M. K. Simon and M.-S. Alouini, "A unified approach to the performance analysis of digital communication over generalized fading channels," *Proc. of the IEEE*, pp. 1860–1877, Sept. 1998.

[10] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: performance criterion and code construction," *IEEE Trans. on Information Theory*, vol. 44, no. 2, pp. 744–765, Mar. 1998.

[11] J. H. Winters, "The diversity gain of transmit diversity in wireless systems with rayleigh fading," *IEEE Trans. on Vehicular Tech.*, vol. 47, no. 1, pp. 119–123, Feb. 1998.