Hindawi Publishing Corporation
EURASIP Journal on Wireless Communications and Networking
Volume 2006, Article ID 27694, Pages 1–11
DOI 10.1155/WCN/2006/27694

Energy-Efficient Channel Estimation in MIMO Systems

Sarod Yatawatta,[1] Athina P. Petropulu,[1] and Charles J. Graff[2]

[1] *Electrical and Computer Engineering Department, Drexel University, Philadelphia, PA 19104, USA*
[2] *US Army RDECOM CERDEC STCD, Fort Monmouth, NJ 07703, USA*

Received 14 February 2005; Revised 10 June 2005; Accepted 5 December 2005

Recommended for Publication by Stefan Kaiser

The emergence of MIMO communications systems as practical high-data-rate wireless communications systems has created several technical challenges to be met. On the one hand, there is potential for enhancing system performance in terms of capacity and diversity. On the other hand, the presence of multiple transceivers at both ends has created additional cost in terms of hardware and energy consumption. For coherent detection as well as to do optimization such as water filling and beamforming, it is essential that the MIMO channel is known. However, due to the presence of multiple transceivers at both the transmitter and receiver, the channel estimation problem is more complicated and costly compared to a SISO system. Several solutions have been proposed to minimize the computational cost, and hence the energy spent in channel estimation of MIMO systems. We present a novel method of minimizing the overall energy consumption. Unlike existing methods, we consider the energy spent during the channel estimation phase which includes transmission of training symbols, storage of those symbols at the receiver, and also channel estimation at the receiver. We develop a model that is independent of the hardware or software used for channel estimation, and use a divide-and-conquer strategy to minimize the overall energy consumption.

1. INTRODUCTION

The use of multiple-input multiple-output (MIMO) channels formed using multiple transmit/receive antennas has been demonstrated to have great potential for achieving high data rates [1]. Of concern, however, is the increased complexity associated with multiple transmit/receive antenna systems. First, increased hardware cost is required to implement multiple RF chains and adaptive equalizers. Second, increased complexity and energy is required to estimate large-size MIMO channels.

Energy conservation in MIMO systems has been considered in different perspectives. In [2], for instance, hardware-level optimization is done to minimize energy. On the other hand, in [3, 4], energy consumption is minimized at the receiver by using low-rank equalization. In [5], reducing the order of MIMO systems by selection of antennae is given as a viable option to minimize energy consumption both at the receiver and transmitter, without degrading the system performance. In [6], the transmission and circuit energy consumption per bit of information transmitted is analyzed. The authors claim in [6] that single-input single-output (SISO) (1×1) systems gives best performance over MIMO (2×2) systems for short-range transmission.

In this paper, we focus on MIMO channel estimation subject to delay and error constraints. We propose an antenna selection scheme for channel estimation that can minimize energy consumed both at the transmitter and the receiver. Note that antenna selection for data transmission [5] requires at least partial knowledge of the full channel matrix. Hence, the proposed scheme can be applied before the antenna selection is done for data transmission.

We can summarize the novelty of the proposed scheme as follows: (i) we concentrate exclusively on the channel estimation phase unlike in [6] where the authors have considered the data transmission phase; (ii) we propose an antenna selection scheme to minimize energy during channel estimation unlike [5] where information-theoretic performance (channel capacity) during data transmission is considered for antenna selection; (iii) the proposed method can be applied independent of the hardware or software used for channel estimation. In fact, the hardware and software can be optimized independently of the proposed method as in [2].

The rest of the paper is organized as follows. First, we study the channel estimation error and the cost of computation of the MIMO system under consideration. Next, we describe the generalized energy reduction scheme. After this, we focus on minimizing energy at the transmitter and the receiver separately. Next, we consider joint transmitter and receiver energy minimization. To illustrate our method, we consider a MIMO system with flat-fading channels of arbitrary size and give comparisons of energy and error variation for different channel estimation schemes obtained by varying the number of active transmit/receive antennas under a fixed delay and error constraint.

2. CHANNEL ESTIMATION IN MIMO FLAT-FADING ENVIRONMENTS USING TRAINING

Before we proceed to formulate our problem, we need to have a valid model of channel estimation. The basic equation of the flat-fading MIMO system in concern is given in (1):

$$\underbrace{\mathbf{y}^i}_{N\times 1} = \underbrace{\mathbf{H}}_{N\times M}\underbrace{\mathbf{x}^i}_{M\times 1} + \underbrace{\mathbf{v}^i}_{N\times 1}, \quad i = 1, 2, \ldots, \tag{1}$$

where we consider a MIMO system with M transmitters and N receivers. The received data vector is $\mathbf{y}^i$, the transmitted data vector is $\mathbf{x}^i$, while the noise vector is $\mathbf{v}^i$ at the ith time interval. The channel matrix is $\mathbf{H}$ of size $N \times M$. Let the noise variance be σ^2 and let the signal power level be P. By transmitting J data blocks, we form the augmented matrix equation (2):

$$\underbrace{\mathbf{Y}}_{N\times J} = \underbrace{\mathbf{H}}_{N\times M}\underbrace{\mathbf{X}}_{M\times J} + \underbrace{\mathbf{V}}_{N\times J}, \tag{2}$$

where

$$\mathbf{Y} = [\mathbf{y}^1, \ldots, \mathbf{y}^J], \qquad \mathbf{X} = [\mathbf{x}^1, \ldots, \mathbf{x}^J], \qquad \mathbf{V} = [\mathbf{v}^1, \ldots, \mathbf{v}^J] \tag{3}$$

and we form the least squares estimation of the channel as [7]

$$\hat{\mathbf{H}} = \mathbf{Y}\mathbf{X}^\dagger. \tag{4}$$

The matrix pseudoinverse $\mathbf{X}^\dagger$ is formed as

$$\mathbf{X}^\dagger = \mathbf{X}^H(\mathbf{X}\mathbf{X}^H)^{-1}. \tag{5}$$

2.1. Channel estimation error

The channel estimation error is obtained from (2) and (4) as

$$\boldsymbol{\xi} = \hat{\mathbf{H}} - \mathbf{H} = \mathbf{V}\mathbf{X}^\dagger \tag{6}$$

and the average squared error (χ) is

$$\chi \triangleq \frac{1}{NM}\text{trace}(\boldsymbol{\xi}\boldsymbol{\xi}^H) = \frac{1}{NM}\text{trace}(\mathbf{X}^\dagger\mathbf{X}^{\dagger H}\mathbf{V}^H\mathbf{V}). \tag{7}$$

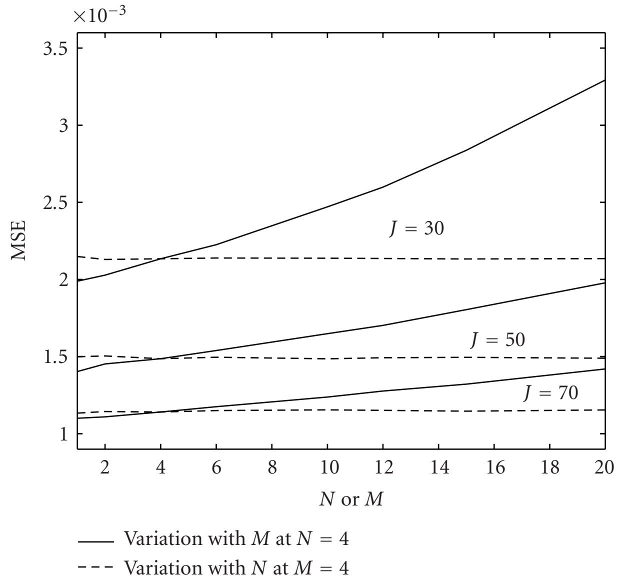

Figure 1: MSE variation with N, M, and J. We see that the MSE is independent of N, has a linear variation with M, and is inversely proportional to J.

We can find a lower bound to χ as follows [7]. We assume the noise to be additive white and Gaussian. Then, taking expectation of χ, we get the mean-squared error

$$\begin{aligned}\text{MSE} = E\{\chi\} &\geq \frac{1}{NM}\text{trace}(\mathbf{X}^\dagger\mathbf{X}^{\dagger H}E\{\mathbf{V}^H\mathbf{V}\})\\ &\geq \frac{1}{NM}\text{trace}(\mathbf{X}^\dagger\mathbf{X}^{\dagger H}N\sigma^2\mathbf{I})\\ &\geq \frac{\sigma^2}{M}\text{trace}\left((\mathbf{X}\mathbf{X}^H)^{-1}\right).\end{aligned} \tag{8}$$

From [7], we see that

$$\text{trace}\left((\mathbf{X}\mathbf{X}^H)^{-1}\right) \geq \frac{M^2}{\text{trace}(\mathbf{X}\mathbf{X}^H)} \tag{9}$$

and under optimal training, with $\mathbf{X}\mathbf{X}^H = JP\mathbf{I}$, we get

$$\text{MSE} \geq \frac{\sigma^2}{PJ}, \tag{10}$$

which is the result derived in [7].

However, the above MSE is not always achievable in practice. First, the above derivation assumes the noise covariance to be identity, which is only feasible if the training length is infinitely large. Moreover, it is not always possible to design an optimal training sequence. Hence, we need to choose a more pragmatic, worst-case error formula to model our system. If we consider our channel estimation scheme, we know that the channel estimation error is inversely proportional to the SNR and the data block length while it is directly proportional to the interference, that is, the number of transmitters.

In order to investigate this behavior, we have simulated random channels and have given the result in Figure 1. In order to model this behavior as nearly as possible, we formulate the error as

$$\text{MSE} = c\sigma^2 \frac{M}{JP}, \tag{11}$$

where σ^2 is the noise power, P is the signal power, and c is a real, positive constant of proportionality. We should note that (11) is purely a heuristic formula that seems to model the behavior seen in Figure 1 very well. Note also that it is possible for us to calculate the error more precisely in terms of $\mathbf{X}$ because the training is known. However, since $\mathbf{X}$ is also a function of M, it is not possible to formulate the error in explicit form and our analysis becomes more complicated. Thus, we limit our analysis to (11) in the remainder of this paper.

2.2. Channel estimation cost

It should be kept in mind that the variables in (1) are complex numbers in general. However, we prefer to study the number of computations required to obtain (4) in terms of real floating-point operations.

Before proceeding in our analysis, we make some generalized assumptions.

- (i) We assume the computations are done in a sequential manner. However, in real systems, most computations are done in a blockwise manner [8, 9]. Moreover, more than one floating-point operation can be performed simultaneously. However, the analysis of such schemes is beyond the scope of this paper.
- (ii) The cost of multiplication is higher than the cost of addition. However, how much higher this is dependent on exact hardware implementation. For instance, in [10, 11], it is given as 4 to 1. Moreover, the cost of division is higher than multiplication. In order to make our analysis simpler, we consider cost of additions and multiplications separately and we take the cost of division to be equal to two multiplications (reciprocal operation and multiplication).

We should stress that for a given hardware model, a more detailed and more accurate set of assumptions can be made.

We use the following basic rules, coupled with our assumptions, as in [12]. Let us denote ν_m and ν_a as the energy cost for one real floating-point multiplication and addition, respectively.

- (i) One complex addition requires two real additions, so the cost is $2\nu_a$.
- (ii) One complex multiplication, that is, $(\alpha_1 + j\beta_1)(\alpha_2 + j\beta_2)$, where $(\alpha_1 + j\beta_1)$ and $(\alpha_2 + j\beta_2)$ are the complex numbers with real and imaginary parts, costs three real multiplications and five real additions $3\nu_m + 5\nu_d$. This is done as $\alpha_1\alpha_2 - \beta_1\beta_2 + j((\alpha_1 + \beta_1)(\alpha_2 + \beta_2) - \alpha_1\alpha_2 - \beta_1\beta_2)$. The naive multiplication method requires one more multiplication, $4\nu_m + 4\nu_a$.
- (iii) One division with complex numerator and real denominator is equivalent to the cost of two real multiplications, $2\nu_m$.
- (iv) One division with complex numerator and denominator $(\alpha_1 + j\beta_1)/(\alpha_2 + j\beta_2)$ requires 7 real multiplications and 6 real additions, that is, $(\alpha_1 + j\beta_1)(\alpha_2 - j\beta_2)/(\alpha^2 + \beta^2)$.

Next, let us consider the calculation of (5) in detail. The steps involved in this are as follows.

(1) Calculation of $\mathbf{X}\mathbf{X}^H$. The resulting matrix is of size $M\times M$. Each element of this matrix is an inner product of two vectors of size $1 \times J$. This inner product requires J complex multiplications and $J-1$ complex additions. Hence, the total computation is M^2J complex multiplications and $M^2(J-1)$ complex additions. We can cut this almost by half by noticing that $\mathbf{X}\mathbf{X}^H$ is Hermitian. Finally, we have $M(M+1)J/2$ complex multiplications and $M(M+1)(J-1)/2$ complex additions. (Note that we can reduce this even more by noticing that the main diagonal is real. However, we ignore this fact for simplicity in our analysis.)

(2) Calculation of $\mathbf{X}^H(\mathbf{X}\mathbf{X}^H)^{-1}$. In terms of complexity as well as numerical stability, it is not advisable to compute $\mathbf{X}^\dagger$ via explicitly computing the matrix inverse $(\mathbf{X}\mathbf{X}^H)^{-1}$ [13]. Instead, we do this by solving a system of linear equations as follows:

$$\mathbf{X}^\dagger = \mathbf{X}^H(\mathbf{X}\mathbf{X}^H)^{-1}, \tag{12}$$

$$(\mathbf{X}^\dagger)^T = \left((\mathbf{X}\mathbf{X}^H)^T\right)^{-1}(\mathbf{X}^H)^T, \tag{13}$$

$$(\mathbf{X}^\dagger)^T = (\mathbf{L}\mathbf{L}^H)^{-1}(\mathbf{X}^H)^T, \tag{14}$$

$$\mathbf{L}\mathbf{L}^H(\mathbf{X}^\dagger)^T = (\mathbf{X}^H)^T, \tag{15}$$

where $\mathbf{L}\mathbf{L}^H$ is the Cholesky decomposition of $\mathbf{A} = (\mathbf{X}\mathbf{X}^H)^T$. $(\mathbf{X}^\dagger)^T$ are the unknowns that need to be found by solving the linear system. Each column of (15) can be written as

$$\mathbf{L}\mathbf{L}^H\widetilde{\mathbf{x}}_i = \overline{\mathbf{x}}_i, \quad i = 1, 2, \ldots, J. \tag{16}$$

We need to solve a system as given in (16), J times to obtain the matrix $\mathbf{X}^\dagger$. First, forward elimination is used to solve for $\mathbf{z}_i$ in (17):

$$\mathbf{L}\mathbf{z}_i = \overline{\mathbf{x}}_i, \quad i = 1, 2, \ldots, J. \tag{17}$$

Next, back substitution is used to solve for $\widetilde{\mathbf{x}}_i$ in (18):

$$\mathbf{L}^H\widetilde{\mathbf{x}}_i = \mathbf{z}_i, \quad i = 1, 2, \ldots, J. \tag{18}$$

Since we have described the basic steps to be followed, let us consider the complexity of each operation.

(i) The Cholesky decomposition can be given in pseudocode as [14] in Algorithm 1.

From Algorithm 1, we see that for each j, there are $i-1$ complex multiplications and additions and 1 real division. Since for fixed i, j varies from $i+1$ to M, the number of iterations of j is $M-i$. Moreover, in order to calculate L_{ii}, we need $i-1$ multiplications and additions and one square root operation. We consider the cost of square root to be equivalent to the cost of division.

To summarize, the total number of operations for each value of i is $(i-1+(i-1)(M-i))$ complex multiplications,

Table 1: Computational cost of channel estimation.

Operation	Complex ×	Complex +	Divisions
$\mathbf{X} \times \mathbf{X}^H$	$M(M+1)J/2$	$M(M+1)(J-1)/2$	—
Cholesky decomposition	$M(M+1)(M-1)/6$	$M(M+1)(M-1)/6$	$M(M+1)/2$
Forward elimination	$JM(M-1)/2$	$JM(M-1)/2$	JM
Back substitution	$JM(M-1)/2$	$JM(M-1)/2$	JM
$\mathbf{Y} \times \mathbf{X}^\dagger$	NMJ	$NM(J-1)$	—
Total	$1/6(M^3 - M)$ $+J/2(3M^2 - M + NM)$	$1/6(M^3 - 3M^2 - 4M - 6NM)$ $+J/2(3M^2 - M + 2NM)$	$1/2(M^2 + M)$ $+2JM$

for $i := 1, \ldots, M\{$

$$L_{ii} := \left(A_{ii} - \sum_{k=1}^{i-1} |L_{ik}|^2\right)^{1/2}$$

for $j := i+1, \ldots, M\{$

$$L_{ij} := (1/L_{ii})\left(A_{ij} - \sum_{k=1}^{i-1} L_{ik}L_{jk}^*\right)$$

$\}$

$\}$

Algorithm 1: Pseudocode for Cholesky decomposition.

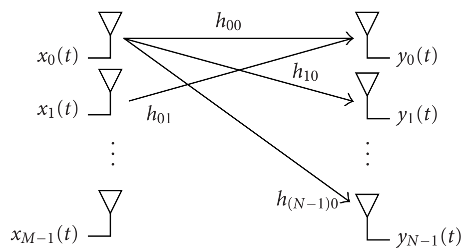

Figure 2: MIMO channel.

$(i-1+(i-1)(M-i))$ complex additions, and $(1+M-i)$ divisions(square root). Accumulating this from $i = 1, \ldots, M$, we get $M(M+1)(M-1)/6$ complex multiplications, $M(M+1)(M-1)/6$ complex additions, and $M(M+1)/2$ divisions.

(ii) The forward elimination involves the step

$$z_j = \frac{1}{L_{jj}}\left(\overline{x}_j - \sum_{k=1}^{j-1} L_{jk}z_k\right), \quad j = 1, \ldots, M. \tag{19}$$

For fixed j, this involves $j-1$ complex multiplications, $j-1$ complex additions, and 1 division. This sums up to the final cost of $M(M-1)/2$ multiplications, $M(M-1)/2$ additions and M divisions.

The forward elimination has to be done for each column of $\mathbf{X}^\dagger$, that is, J times. Thus, the final cost is $JM(M-1)/2$ complex multiplications, $JM(M-1)/2$ complex additions, and JM divisions.

(iii) The back substitution has the same complexity of forward elimination. Thus, we have the same final cost of $JM(M-1)/2$ complex multiplications, $JM(M-1)/2$ complex additions, and JM divisions.

Using the above calculations, we can compute the total cost of forming $\mathbf{X}^\dagger$. This is given is Table 1.

(3) Calculation of product $\mathbf{Y}\mathbf{X}^\dagger$. Once again, we have a matrix product where each element of the resultant matrix is an inner product of vectors whose dimensions are $1 \times J$. Hence, we have NMJ complex multiplications and $NM(J-1)$ complex additions.

A summary of our analysis is given in Table 1. From Table 1, we can derive the total number of real floating-point operations to be $J((9/2)M^2+(5/2)M+3NM)+(1/2)M^3+M^2+(1/2)M$ multiplications and $J((21/2)M^2-(7/2)M+7NM)+(7/6)M^3-(13/6)M-M^2-2NM$ additions.

3. GENERAL METHODOLOGY

In this section, we describe the proposed method in a general sense. The fundamental property that we assume in our scheme is the modularity of hardware. For instance, when a complex hardware system is built, it is done in a modular way by assembling less complex blocks. Hence, a MIMO system can be considered as a collection of SISO systems, with respect to hardware. For instance, we assume that a 4-by-4 MIMO system can operate as a 2 by 2 system by turning off some modules.

The MIMO system in concern, with M transmitters and N receivers, can be given as in Figure 2. We call the set of transmitters $\mathbf{T}$ and the set of receivers $\mathbf{R}$. Their cardinalities, $|\mathbf{T}|$ and $|\mathbf{R}|$, are M and N, respectively. The objective is to estimate the channels h_{ij}, $0 \le i \le N-1$, $0 \le j \le M-1$, in an energy-efficient manner. The channel estimation requires the consumption of energy and time.

We make the following assumptions.

(A1) We first ignore electromagnetic interaction between antenna elements. Thus, if we estimate h_{ij} by having active only a subset of transmitters/receivers, the estimate will be the same as the estimate we would get for the same channel if all transmitters/receivers were active. However, we refine our method taking correlation into account at a later section.

(A2) The channels are frequency flat fading and during the training phase, the channels remain time invariant.

We propose the following divide-and-conquer strategy. Instead of estimating the full channel matrix at once (which we call the *naive* method), we propose to estimate the full channel matrix in K steps. On the kth step ($k \in [1, K]$), we select the transmitters given by the set $\mathbf{T}_k$ ($\subseteq \mathbf{T}$) and the receivers given by the set $\mathbf{R}_k$ ($\subseteq \mathbf{R}$) and estimate the channels between those transmitters and receivers. Let P_k be the power level of each transmitter at the kth step, and let l_k denote the length of training data to be used in channel estimation. Moreover, let the noise power level at the receiver be σ^2. Hence, at the kth step, the average SNR at the receiver will be proportional to P_k/σ^2. We assume all transmitters have the same path loss, that is, each transmitter is approximately at the same distance from the receiver and the noise power level is the same on all paths.

We will focus on minimizing the total energy consumption, both at the receiver and transmitter. We define the following functions. Let g_T be the energy spent by all the transmitters. At the receivers, the energy consumption can be broken down into two components: the energy required to perform data acquisition and storage, which we denote by g_I, and the energy needed to perform channel estimation or computations, which we denote by g_C. In our formulation, g_T, g_I, and g_C are functions of the variables K, $\mathbf{T}_k$, $\mathbf{R}_k$, l_k, P_k, $k = 1, \dots, K$. For notational convenience, this dependence is not shown in the sequel.

The total energy consumed can be given as

$$g = g_T + g_I + g_C. \tag{20}$$

Our objective is to minimize g. Next we consider the constraints involved.

(i) *Avoiding trivial solutions*. In order to estimate all the channels, we need

$$\bigcup_{k=1,\dots,K} \mathbf{T}_k \otimes \mathbf{R}_k = \mathbf{T} \otimes \mathbf{R}, \tag{21}$$

where $\otimes$ is the Cartesian product. In order to avoid trivial solutions, we need

$$\mathbf{T}_k \neq \phi, \quad \mathbf{R}_k \neq \phi, \quad k \in [1, K], \tag{22}$$

where ϕ is the null set.

(ii) *Satisfying a channel MSE constraint*. For acceptable performance, the mean channel estimation error (MSE) at each step ϵ_k should be below a minimum threshold,

$$\epsilon_k = \epsilon_k(\mathbf{T}_k, \mathbf{R}_k, P_k, l_k) \leq \epsilon, \quad k \in [1, K]. \tag{23}$$

The exact expression for ϵ_k is dependent on the channel estimation method. If we consider the power level at each step, it should be lower than the maximum allowed by the transmitter P:

$$P_k \leq P, \quad k \in [1, K]. \tag{24}$$

(iii) *Satisfying a transmission delay constraint*. The training length at step k should be above a certain threshold $\underline{l_k}$ for the channel estimation to work (i.e., to have full rank $\mathbf{X}$) and the total data length would be below the maximum delay allowed L:

$$\underline{l_k} \leq l_k, \quad k \in [1, K], \quad \sum_{k=1}^{K} l_k \leq L. \tag{25}$$

Our objective is to find $\mathbf{T}_k$, $\mathbf{R}_k$, P_k, and l_k for $k = 1, \dots, K$ subject to the above constraints (21), (22), (23), (24), and (25) that minimize g given in (20). This is a typical set partitioning problem, where the objective is to find the optimal partition of the sets $\mathbf{T}$ and $\mathbf{R}$. In general, solving such problems would have to consider every possible partition in order to find the optimal one. The complexity of such an approach would be exponential in the set size. However, we pursue simplified solutions in the following sections.

Before we proceed, let us consider the feasibility of the problem. We see that all the parameters are bounded. Hence, the feasibility region is bounded and in order to find feasible solutions, we should choose the limits ϵ and L in a suitable manner. For instance, if we choose $\epsilon = 0$ or $L = 0$, it is obvious that no solutions exist. Hence, by increasing either or both of these values, we can increase the feasibility region. In other words, we can trade off energy with channel estimation error and delay.

3.1. Mutual coupling of antennas

In the preceding discussion under assumption (A1), we have assumed the antennas to be uncorrelated. However, in real life, this is far from the truth. In this section, we consider mutual coupling between antennas at the transmitter and the receiver and examine its effect on the channel estimate. First, we break down the effective channel matrix into components due to mutual coupling and fading, as given in (26):

$$\mathbf{H} = \mathbf{F}\widetilde{\mathbf{H}}\mathbf{G}. \tag{26}$$

The receiver mutual coupling is given by the $N \times N$ matrix $\mathbf{F}$ while the transmitter mutual coupling is given by the $M \times M$ matrix $\mathbf{G}$. We assume a rich scattering environment where the fading matrix $\widetilde{\mathbf{H}}$ (dimension $N \times M$) has full rank and thus $\mathbf{H}$ is full rank. Let us consider the effective channel formed between the transmitters $\mathbf{T}_k$ and the receivers $\mathbf{R}_k$. We assume arbitrary ordering of the transmitters and receivers such that $\mathbf{T}_k$ and $\mathbf{R}_k$ can be grouped together. Then we can partition the matrices in (26) into a 3×3 partition, in an arbitrary manner as

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_{11} & \mathbf{H}_{12} & \mathbf{H}_{13} \\ \mathbf{H}_{21} & \mathbf{H}_{22} & \mathbf{H}_{23} \\ \mathbf{H}_{31} & \mathbf{H}_{32} & \mathbf{H}_{33} \end{bmatrix},$$
$$\mathbf{F} = \begin{bmatrix} \mathbf{F}_{11} & \mathbf{F}_{12} & \mathbf{F}_{13} \\ \mathbf{F}_{21} & \mathbf{F}_{22} & \mathbf{F}_{23} \\ \mathbf{F}_{31} & \mathbf{F}_{32} & \mathbf{F}_{33} \end{bmatrix},$$
$$\widetilde{\mathbf{H}} = \begin{bmatrix} \widetilde{\mathbf{H}}_{11} & \widetilde{\mathbf{H}}_{12} & \widetilde{\mathbf{H}}_{13} \\ \widetilde{\mathbf{H}}_{21} & \widetilde{\mathbf{H}}_{22} & \widetilde{\mathbf{H}}_{23} \\ \widetilde{\mathbf{H}}_{31} & \widetilde{\mathbf{H}}_{32} & \widetilde{\mathbf{H}}_{33} \end{bmatrix}, \tag{27}$$
$$\mathbf{G} = \begin{bmatrix} \mathbf{G}_{11} & \mathbf{G}_{12} & \mathbf{G}_{13} \\ \mathbf{G}_{21} & \mathbf{G}_{22} & \mathbf{G}_{23} \\ \mathbf{G}_{31} & \mathbf{G}_{32} & \mathbf{G}_{33} \end{bmatrix}.$$

The dimensions of the submatrices are such that the product in (26) holds. In particular, $\mathbf{F}_{22}$ is size $|\mathbf{T}_k| \times |\mathbf{T}_k|$, $\widetilde{\mathbf{H}}_{22}$ is size $|\mathbf{T}_k| \times |\mathbf{R}_k|$, and $\mathbf{G}_{22}$ is size $|\mathbf{R}_k| \times |\mathbf{R}_k|$. The dimensions of the other matrices are irrelevant to this discussion.

Next, we can express the channel between $\mathbf{T}_k$ and $\mathbf{R}_k$ as

$$\mathbf{H}_{22} = \sum_{j=1}^{3} \left(\sum_{i=1}^{3} \mathbf{F}_{2i} \widetilde{\mathbf{H}}_{ij} \right) \mathbf{G}_{j2}. \tag{28}$$

However, if we apply the divide-and-conquer scheme, with all transmitters and receivers except $\mathbf{T}_k$ and $\mathbf{R}_k$ being turned off, the channel that we estimate in the kth step is

$$\widehat{\mathbf{H}}_k = \mathbf{F}_{22} \widetilde{\mathbf{H}}_{22} \mathbf{G}_{22}. \tag{29}$$

Thus, we see that there is an error in the channel estimation due to mutual coupling. However, we can correct this error provided we know the mutual coupling matrices $\mathbf{F}$ and $\mathbf{G}$ perfectly and have full rank. This is not an unreasonable requirement because $\mathbf{F}$ and $\mathbf{G}$ are constant for a given antenna configuration.

The procedure to correct the error due to mutual coupling is as follows. At the kth step, after getting the estimate $\widehat{\mathbf{H}}_k$, we solve (29) to obtain $\widetilde{\mathbf{H}}_{22}$.

The number of computations required to solve this linear system of equations can be calculated as follows. We first solve the system $\widetilde{\mathbf{H}}_{22}\mathbf{G}_{22} = \mathbf{F}_{22}^{-1}\widehat{\mathbf{H}}_k$ and next solve the system $\widetilde{\mathbf{H}}_{22} = \mathbf{F}_{22}^{-1}\widehat{\mathbf{H}}_k\mathbf{G}_{22}^{-1}$. This is similar to the analysis done in Section 2.2. However, the matrices $\mathbf{F}_{22}$ and $\mathbf{G}_{22}$ are not Hermitian in general. So we need to use LU decomposition in this case. Let us consider the solution of $\widetilde{\mathbf{H}}_{22}\mathbf{G}_{22} = \mathbf{F}_{22}^{-1}\widehat{\mathbf{H}}_k$ first. The LU decomposition of $\mathbf{LU} = \mathbf{F}_{22}$ can be given as [14] in Algorithm 2.

The LU decomposition as given in Algorithm 2 requires $T(T-1)(2T-1)/6$ complex multiplications, $T(T-1)(2T-1)/6$ complex additions, and $T(T+1)/2$ complex divisions, where $T \triangleq |\mathbf{T}_k|$. The cost of forward elimination and back substitution can be deduced from Section 2.2. Note that the forward elimination requires no divisions because the main diagonal of the lower triangular matrix consists of 1. We have given the total number of computations required in Table 2, where $R \triangleq |\mathbf{R}_k|$.

```
for i := 1,..., |T_k|
    L_ii := 1
for j := 1,..., |T_k| {
    for i := 1,..., j
            U_ij := L_ij − Σ_{k=1}^{i−1} L_ik U_kj
    for i := j + 1,..., |T_k|
            L_ij := (1/U_jj)(L_ij − Σ_{k=1}^{j−1} L_ik U_kj)
}
```

Algorithm 2: Pseudocode for LU decomposition.

The cost of $\widetilde{\mathbf{H}}_{22} = \mathbf{F}_{22}^{-1}\widehat{\mathbf{H}}_k\mathbf{G}_{22}^{-1}$ can be calculated in a similar manner. We have given the result in Table 3.

In the above result, the division can include a complex divisor. Thus, the cost of complex division which is 7 real multiplications and 6 real additions has to be taken into account. Finally, we get the total cost as $(T^3+2T^2+4T+3RT^2+8RT+3R^2T+4R+2R^2+R^3)$ real multiplications and $(7/3T^3-1/2T^2+25/6T+7RT^2-2RT+7R^2T+25/6R-1/2R^2+7/3R^3)$ real additions, where $T = |\mathbf{T}_k|$ and $R = |\mathbf{R}_k|$. This is the additional cost due to mutual coupling of antennas that appear in the proposed divide-and-conquer method.

After steps $k = 1,\dots,K$, we would have formed the entire matrix $\widetilde{\mathbf{H}}$. Finally, we form the product $\mathbf{F}\widetilde{\mathbf{H}}\mathbf{G}$ to obtain the actual channel matrix. The complexity of this operation is $3(N^2M + NM^2)$ real multiplications and $(7N^2M + 7NM^2 - 4NM)$ real additions.

4. MINIMIZING ENERGY AT THE TRANSMITTER

We make the following assumptions.

(B1) We assume the receiver has no constraints on energy because we only minimize energy at the transmitter. This allows us to always make $\mathbf{R}_k = \mathbf{R}$. In other words, we use all receivers at all steps.

(B2) We assume the antennas to be uncorrelated, so that the channel estimate will not change with the selection of $\mathbf{T}_k$ and $\mathbf{R}_k$. Moreover, we assume the only variable affecting the channel estimation error to be the sizes of $\mathbf{T}_k$ and $\mathbf{R}_k$ and not the individual elements in them.

(B3) We assume retransmissions to be costly and hence select disjoint sets of transmitters, that is, $\mathbf{T}_k$ are disjoint. In other words, each transmitter only transmits at one step k.

From (11), the channel estimation error at the kth step is

$$\epsilon_k = c_1 \frac{\sigma^2}{P_k l_k} |\mathbf{T}_k|, \tag{30}$$

where σ^2 is the noise variance, and c_1 is a real, positive constant. The transmitter power level and the training length are given by P_k and l_k, respectively. The cardinality of the set $\mathbf{T}_k$ is given as $|\mathbf{T}_k|$.

Table 2: Computational cost of $\widetilde{\mathbf{H}}_{22}\mathbf{G}_{22} = \mathbf{F}_{22}^{-1}\widehat{\mathbf{H}}_k$.

Operation	Complex ×	Complex +	Divisions
LU decomposition	$T(T-1)(2T-1)/6$	$T(T-1)(2T-1)/6$	$T(T+1)/2$
Forward elimination	$RT(T-1)/2$	$RT(T-1)/2$	—
Back substitution	$RT(T-1)/2$	$RT(T-1)/2$	RT
Total	$1/6(2T^3-3T^2+T)$ $+RT(T-1)$	$1/6(2T^3-3T^2+T)$ $+RT(T-1)$	$1/2(T^2+T)$ $+RT$

Table 3: Computational cost of $\widetilde{\mathbf{H}}_{22} = \mathbf{F}_{22}^{-1}\widehat{\mathbf{H}}_k\mathbf{G}_{22}^{-1}$.

Operation	Complex ×	Complex +	Divisions
LU decomposition	$R(R-1)(2R-1)/6$	$R(R-1)(2R-1)/6$	$R(R+1)/2$
Forward elimination	$RT(R-1)/2$	$RT(R-1)/2$	—
Back substitution	$RT(R-1)/2$	$RT(R-1)/2$	RT
Total	$1/6(2R^3-3R^2+R)$ $+RT(R-1)$	$1/6(2R^3-3R^2+R)$ $+RT(R-1)$	$1/2(R^2+R)$ $+RT$

The energy expenditure at the transmitter occurs mainly due to transmission of training symbols. This energy is proportional to the transmitter power level, the duration (or length) of training, and the number of active transmitters. Thus, total energy spent by all the transmitters can be given as

$$g_T = \sum_{k=1}^{K} c_2 P_k l_k \left|\mathbf{T}_k\right|, \tag{31}$$

where c_2 is a real, positive constant of proportionality. Due to $\mathbf{R}_k = \mathbf{R}$, and $\mathbf{T}_k$ being disjoint, we can simplify (21) further. We can leave $\mathbf{R}_k$ from the Cartesian product because $\mathbf{R}_k = \mathbf{R}$. This reduces (21) to

$$\bigcup_{k=1,\dots,K} \mathbf{T}_k = \mathbf{T} \tag{32}$$

and since all $\mathbf{T}_k$ are disjoint, we get

$$\sum_{k=1}^{K} \left|\mathbf{T}_k\right| = |\mathbf{T}| = M. \tag{33}$$

Solving for $|\mathbf{T}_k|$ is a standard integer partition problem. For instance, if $M = 4$, the ways we can select the number of transmitters during the K steps are $\{4\}(K = 1)$, $\{3,1\}(K = 2)$, $\{2,2\}(K = 2)$, $\{2,1,1\}(K = 3)$, and $\{1,1,1,1\}(K = 4)$. Thus, there are 5 possible ways in this case. If the number of possible ways of selecting $|\mathbf{T}_k|$ is $p(M)$ for $|\mathbf{T}| = M$, we have [15]

$$p(M) \approx \frac{1}{4\sqrt{3}}\left(\frac{e^{\pi\sqrt{(2/3)M}}}{M}\right). \tag{34}$$

We first select a partitioning scheme and keeping it fixed, we solve the problem

$$\min_{P_i, l_i, i\in[1,K]} \sum_{k=1}^{K} c_2 P_k l_k \left|\mathbf{T}_k\right| \tag{35}$$

subject to (23), (24), and (25), where $\mathbf{T}_k$, $\mathbf{R}_k$ and K are constants. We find the minimum cost associated with the solution. For small values of M, that is, $M \le 10$, we can try all possible partitions to find the best one with the minimum cost.

Proposition 1. *Under assumptions* (A1)-(A2) *and* (B1)-(B3), *the channel estimation scheme that minimizes transmitter energy is to reduce the MIMO channel into a set of single-input multiple-output (SIMO) channels and transmit using one transmitter only at a time. Thus, each time a SIMO channel is estimated. The minimum energy is*

$$\underline{g}_T = c_1 c_2 \frac{\sigma^2}{\epsilon} M \tag{36}$$

as opposed to the energy of the naive method

$$\overline{g}_T = c_1 c_2 \frac{\sigma^2}{\epsilon} M^2. \tag{37}$$

The proof is given in Appendix A.

This result agrees with intuition since in this case there is reduced interference from other transmitters. However, under different assumptions and different channel estimation schemes, we might get different results.

5. MINIMIZING ENERGY AT THE RECEIVER

In contrast to the transmitter, the energy consumption at the receiver is due to data acquisition and computation. From Section 2.2, we see that the computational energy required is

$$\begin{aligned} g_{C,k} = \nu_m\Big(& l_k\Big(\frac{9}{2}|\mathbf{T}_k|^2 + \frac{5}{2}|\mathbf{T}_k| + 3|\mathbf{R}_k||\mathbf{T}_k|\Big) \\ &+ \frac{1}{2}|\mathbf{T}_k|^3 + |\mathbf{T}_k|^2 + \frac{1}{2}|\mathbf{T}_k|\Big) \\ + \nu_a\Big(& l_k\Big(\frac{21}{2}|\mathbf{T}_k|^2 - \frac{7}{2}|\mathbf{T}_k| + 7|\mathbf{R}_k||\mathbf{T}_k|\Big) \\ &+ \frac{7}{6}|\mathbf{T}_k|^3 - \frac{13}{6}|\mathbf{T}_k| - |\mathbf{T}_k|^2 - 2|\mathbf{R}_k||\mathbf{T}_k|\Big), \end{aligned} \tag{38}$$

where ν_m and ν_a are constants. The energy required for data acquisition and storage is proportional to the amount of data received (and processed). This is proportional to the number of active receivers $|\mathbf{R}_k|$ and the length of training l_k. Hence, by conservation of energy, we have

$$g_{I,k} = c_4|\mathbf{R}_k|l_k, \tag{39}$$

where c_4 is a real, positive constant of proportionality and the total energy is

$$g_R = \sum_{k=1}^{K} g_{C,k} + g_{I,k}. \tag{40}$$

Our objective is to minimize g_T subject to the constraints (23), (24), and (25).

Proposition 2. *Under assumptions* (A1)-(A2) *and* (B2), *the channel estimation scheme that minimizes the energy consumption at the receiver is to estimate each SIMO channel individually by using one transmitter and all receivers at each step. The minimum energy is*

$$\underline{g}_R = M\big(\nu_m(l(7+3N)+2)+\nu_a(l(7+7N)-2-2N)+c_4Nl\big) \tag{41}$$

as opposed to the energy of the naive method

$$\begin{aligned} \overline{g}_R = M\Big(&\nu_m\Big(l\Big(\frac{9}{2}M^2 + \frac{5}{2}M + 3NM\Big) + \frac{1}{2}M^2 + M + \frac{1}{2}\Big) \\ &+ \nu_a\Big(l\Big(\frac{21}{2}M^2 - \frac{7}{2}M + 7NM\Big) - \frac{13}{6} - M - 2N\Big) \\ &+ c_4Nl\Big), \end{aligned} \tag{42}$$

where $l = c_1\sigma^2/P\epsilon$.

The proof is given in Appendix B.

5.1. Minimizing energy with correlated antennas

With mutual coupling between antennas, at each step, there will be an additional cost as given in Section 3.1 of

$$\begin{aligned} g'_{C,k} = \nu_m\Big(& |\mathbf{T}_k|^3 + 2|\mathbf{T}_k|^2 + 4|\mathbf{T}_k| + 3|\mathbf{R}_k||\mathbf{T}_k|^2 \\ &+ 8|\mathbf{R}_k||\mathbf{T}_k| + 3|\mathbf{R}_k|^2|\mathbf{T}_k| + 4|\mathbf{R}_k| \\ &+ 2|\mathbf{R}_k|^2 + |\mathbf{R}_k|^3\Big) \\ + \nu_a\Big(& \frac{7}{3}|\mathbf{T}_k|^3 - \frac{1}{2}|\mathbf{T}_k|^2 + \frac{25}{6}|\mathbf{T}_k| \\ &+ 7|\mathbf{R}_k||\mathbf{T}_k|^2 - 2|\mathbf{R}_k||\mathbf{T}_k| + 7|\mathbf{R}_k|^2|\mathbf{T}_k| \\ &+ \frac{25}{6}|\mathbf{R}_k| - \frac{1}{2}|\mathbf{R}_k|^2 + \frac{7}{3}|\mathbf{R}_k|^3\Big) \end{aligned} \tag{43}$$

and the final cost needs to be modified as

$$\begin{aligned} g'_R = {}& 3\nu_m(N^2M + NM^2) + \nu_a(7N^2M + 7NM^2 - 4NM) \\ &+ \sum_{k=1}^{K} g_{C,k} + g_{I,k} + g'_{C,k}. \end{aligned} \tag{44}$$

In this case, the proposed optimal method in Proposition 2 has higher cost than the naive method because of the added computations as well as due to the appearance of higher power terms of $|\mathbf{R}_k|$ in the cost expression. Due to the same reasons, we cannot derive an optimal solution analytically in this case.

6. MINIMIZING ENERGY BOTH AT THE TRANSMITTER AND RECEIVER

From Propositions 1 and 2, we can conclude that the optimal scheme of channel estimation for a MIMO system (with no mutual coupling) that minimizes both transmitter and receiver energy consumption is to reduce the system into a set of SIMO channels and estimate each SIMO channel individually. In other words, instead of transmitting the training symbols from all transmitters simultaneously, we have to transmit them in a sequential manner by activating only one transmitter at a time. In order to satisfy the delay requirement, each transmitter will be active only for a fraction of the time it would have been active if all transmitters were transmitting simultaneously.

6.1. Numerical example

In order to illustrate the result stated in the previous paragraph, we consider an 8×8 MIMO system with flat fading channels. In Table 4 and Figures 3, 4, and 5, we show results of some possible schemes for channel estimation. Our constraints are, maximum error $\epsilon = 10^{-3}$ and delay $L = 56$. In *scheme 1*, we used 56 symbols per each transmitter (total 448) and employed the naive method to estimate the 8×8 system. In *scheme 2*, we used Proposition 1 and used 7 symbols per each transmitter (total 56) to estimate the 8×1 SIMO systems (8 times). In *scheme 3*, we transmitted 7 symbols from

TABLE 4: Energy consumption for different channel estimation schemes for 50 random 8-by-8 channels.

Scheme	Total energy
1	$448c_2P + 448c_4 + 28324\nu_m + 61540\nu_a$
2	$56c_2P + 448c_4 + 1752\nu_m + 3384\nu_a$
3	$56c_2P + 56c_4 + 3824\nu_m + 8032\nu_a$
4	$112c_2P + 448c_4 + 4012\nu_m + 8108\nu_a$

each transmitter (total 56) and again used the naive method. Finally, in *scheme 4*, we used 14 symbols per each transmitter but reduced the system into four 8×2 systems (total 112) to estimate the channel in 4 steps. In Figure 3, we see that scheme 1 has the lowest error but highest energy consumption. Either scheme 2 or 3 has the lowest energy consumption. Scheme 3 has the lowest delay, but the channel cannot be estimated because the training matrix does not have full row rank. Thus we have omitted the results of scheme 3 from figures. Schemes 2 and 4 have intermediate performance in terms of error and energy. This is also illustrated in Figure 4, where we have considered the number of estimation steps to be performed (K) to be the independent variable. Thus, we see that a tradeoff can be accomplished between error and energy.

Although in this example schemes 2 and 4 have higher channel estimation error, a better conclusion about performance can be drawn only after numerical evaluation of the performance in terms of the bit error rate. In order to investigate this, we have simulated transmission of 4-QAM data through the same set of channels. An MMSE equalizer was used at the receiver, constructed from the channel estimate obtained as described in the previous paragraph. For each channel, 10 Monte Carlo simulations were performed. We have given the uncoded BER results in Figure 5 and we see acceptable performance of the proposed scheme.

The constants P, c_1, c_2, ν_m, ν_a, c_4 can be calculated given the specific hardware, or can be experimentally measured. We should stress that although we have assumed them to be constants, for better results, some of them (such as c_1) might be taken as variables of M, N, and SNR and can be incorporated into the optimization.

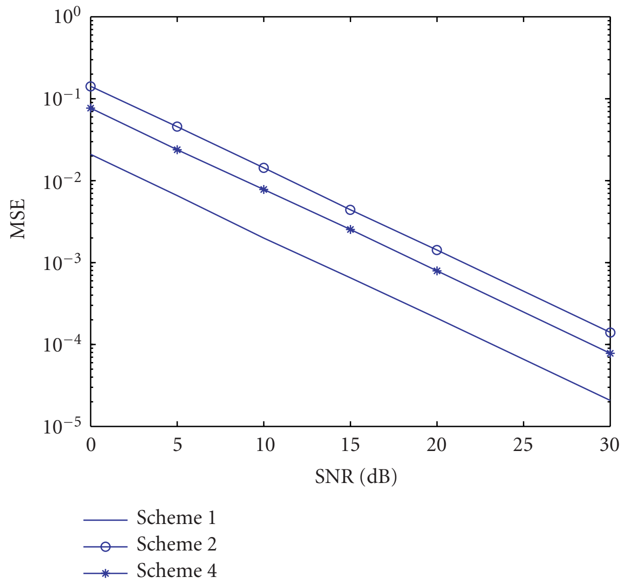

FIGURE 3: MSE variation with SNR for different channel estimation schemes for 50 random 8-by-8 channels.

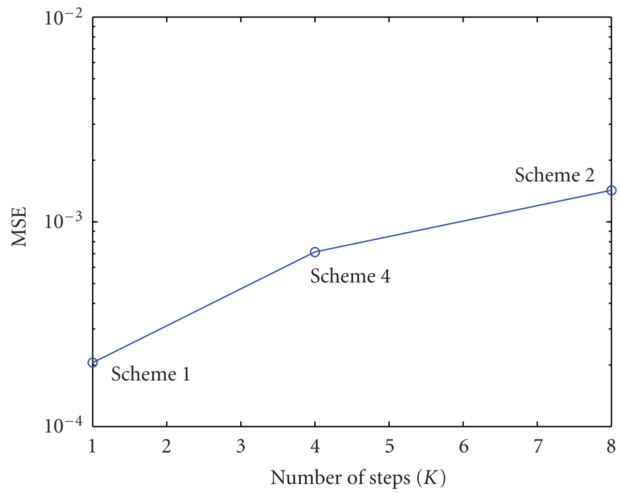

FIGURE 4: MSE variation with the number of steps taken (K) by different schemes at SNR 20 dB.

6.2. Minimizing energy with correlated antennas

With mutual coupling, the total cost to be minimized becomes

$$g' = g_T + g'_R, \tag{45}$$

where g_T and g'_R are given in (31) and (44), respectively. Minimization of each term in the summation of (45) individually would lead us to two different answers. We know that the scheme proposed in Propositions 1 and 2 would minimize g_T but would increase g'_R with respect to the naive solution. However, if the cost saving in g_T is greater than the loss in g'_R, we still might reduce the overall cost. The derivation of an analytical solution for this problem is impossible due to the fact that the associated costs are disparate in nature. However, given the exact cost functions, numerical optimization is possible for a specific hardware setup. This will be tackled in future work.

7. CONCLUSIONS

Using a generic model for channel estimation error and energy consumption of a MIMO system, we have shown that under flat-fading and least-squares channel estimation, the optimal channel estimation scheme in terms of minimizing energy consumption is to convert the MIMO system into a set of SIMO channels by activating each transmitter

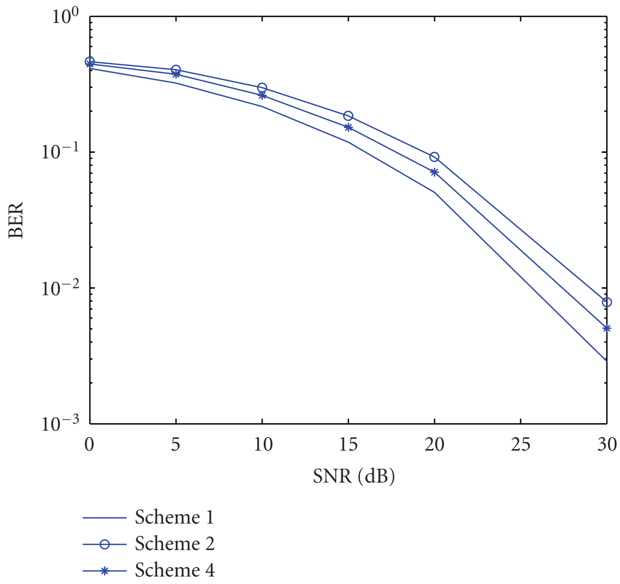

Figure 5: BER variation SNR for different channel estimation schemes for 50 random 8-by-8 channels.

individually and performing channel estimation on each SIMO system. However, the energy reduction comes at an increase in estimation error. In our formulation, we have assumed a homogeneous, isotropic, uncorrelated set of transmitters and receivers. There is room in this area for future work on adapting this method to a MIMO channel formed by a disparate set of transmitters and receivers with different power, computation, and storage capabilities and different radiation patterns. Future work will consider application of the proposed scheme to actual hardware.

APPENDICES

A. PROOF OF PROPOSITION 1

If we assume the partition $\mathbf{T}_k$, $\mathbf{R}_k$ is already given for all k, the problem at hand reduces to a typical nonlinear continuous optimization problem. In order to solve this, we first find the extremal points and see if they satisfy the Karush-Kuhn-Tucker (KKT) conditions [16].

The Lagrangian (ignoring the lower bounds for l_k) is

$$\begin{aligned}\mathcal{L} = &\sum_{k=1}^{K} c_2 P_k l_k \left|\mathbf{T}_k\right| + \sum_{k=1}^{K} \lambda_{1,k}\left(c_1 \frac{\sigma^2}{P_k l_k}\left|\mathbf{T}_k\right| - \epsilon\right) \\ &+ \sum_{k=1}^{K} \lambda_{2,k}\left(P_k - P\right) + \lambda_3\left(\sum_{k=1}^{K} l_k - L\right),\end{aligned} \tag{A.1}$$

where $\lambda_{1,k}$, $\lambda_{2,k}$, λ_3 are the multipliers. For optimality, we need

$$\frac{\partial\mathcal{L}}{\partial l_k} = c_2 P_k \left|\mathbf{T}_k\right| - \lambda_{1,k} c_1 \frac{\sigma^2}{P_k l_k^2}\left|\mathbf{T}_k\right| + \lambda_3 = 0, \tag{A.2}$$

$$\frac{\partial\mathcal{L}}{\partial P_k} = c_2 l_k \left|\mathbf{T}_k\right| - \lambda_{1,k} c_1 \frac{\sigma^2}{P_k^2 l_k}\left|\mathbf{T}_k\right| + \lambda_{2,k} = 0. \tag{A.3}$$

We select a solution as follows. From (24), we select $P_k = P$. We select

$$l_k = c_1 \frac{\sigma^2}{P\epsilon}\left|\mathbf{T}_k\right|, \tag{A.4}$$

where ϵ is the maximum error limit we need. If we substitute this into (30), we get the error at kth step $\epsilon_k = \epsilon$, thus satisfying (23) as well.

If $\mathbf{T}_k = \mathbf{T}$, $|\mathbf{T}_k| = M$ and from (25), we need

$$l_1 = L \geq c_1 \frac{\sigma^2}{P\epsilon} M. \tag{A.5}$$

Hence,

$$\sum_{k=1}^{K} l_k = c_1 \frac{\sigma^2}{P\epsilon} \sum_{k=1}^{K}\left|\mathbf{T}_k\right| = c_1 \frac{\sigma^2}{P\epsilon} M \leq L \tag{A.6}$$

and we see that by selecting l_k according to (A.4), (25) is automatically satisfied. Hence, we have a feasible solution. Next we check its optimality. Since (25) is satisfied and active, we have $\lambda_3 > 0$. From (A.2) we have

$$\lambda_{1,k} = \left(\lambda_3 + c_2 P\left|\mathbf{T}_k\right|\right)\frac{P l_k^2}{c_1 \sigma^2 \left|\mathbf{T}_k\right|}, \tag{A.7}$$

which is positive. Next from (A.3) we have $\lambda_{2,k} = \lambda_3(l_k/P)$, which is again positive. Hence, the solution is optimal. By substitution of $P = P_k$ and (A.4) in (31), we get (A.8). The minimum transmitter energy given the partition of $\mathbf{T}$ is

$$\underline{g_T} = c_1 c_2 \frac{\sigma^2}{\epsilon} \sum_{k=1}^{K}\left|\mathbf{T}_k\right|^2. \tag{A.8}$$

Thus, we have solved the nonlinear continuous optimization problem. The next step is to find the partition that minimizes the cost. We see that the partition that minimizes (A.8) consists of all ones, that is, $\{1, 1, \ldots, 1\}$. In other words, in order to minimize transmission energy, we should estimate channels selecting each transmitter individually. Substituting $|\mathbf{T}_k| = 1$ into (A.8), we get (36); and substituting $K = 1$, $|\mathbf{T}_k| = M$ into (A.5), we get (37).

B. PROOF OF PROPOSITION 2

Note that there is no transmitter power term P_k in (40) and we can select $P_k = P$. Next we select the data length as in (A.4). Next we make the following observations.

(i) Suppose we partition $\mathbf{R}$ such that $|\mathbf{R}_{k'}| < |\mathbf{R}|$. This implies we have turned off some receivers at some point and thus we need some transmitters to transmit more than once. Let the partition scheme where transmitters transmit more than one be given as $\mathbf{T}_{k'}$, as opposed to the partition scheme $\mathbf{T}_k$ where they transmit only once. Then we have

$$\sum_{k'=1}^{K'}\left|\mathbf{T}_{k'}\right|^i \geq \sum_{k=1}^{K}\left|\mathbf{T}_k\right|^i, \quad i = 1, 2, \ldots, \tag{B.1}$$

because we need to estimate all the channels. Thus, if we select a partition scheme where not all receivers are active, we get an increase in cost.

(ii) If we keep all receivers active at each step, $\mathbf{R}_k = \mathbf{R}$, we can select a partition for transmitters such that $\sum_{k=1}^{K} |\mathbf{T}_k| = |\mathbf{T}|$. In this case, the partition scheme that minimizes the cost is $\mathbf{T} = \{1, 1, \ldots\}$ because in this case

$$\sum_{k=1}^{M} \left|\mathbf{T}_k\right|^i = M, \quad i = 1, 2, \ldots. \tag{B.2}$$

Thus, in this case, we need to use all the receivers and the only possible partition for $\mathbf{R}_k$ is $\mathbf{R}$. Hence, we can conclude that the channel estimation scheme that minimizes receiver energy consumption is to estimate each SIMO channel individually. Substituting $|\mathbf{T}_k| = 1$, $K = M$, $|\mathbf{R}_k| = |\mathbf{R}|$ into (40) we get (41) and substituting $K = 1$, $|\mathbf{T}_k| = M$, into (A.5) we get (42).

ACKNOWLEDGMENTS

The authors would like to thank the editor and reviewers for their careful review and helpful comments that enhanced the quality of this paper. This work has been supported by ONR under Grant ONR-N00014-03-1-0123, and the US Army Research Office under Contract DAAD19-02-D-0001. Preliminary results of this work were presented at the IEEE ICASSP 2005.

REFERENCES

[1] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Communications*, vol. 6, no. 3, pp. 311–335, 1998.

[2] S. Rajagopal, S. Bhashyam, J. R. Cavallaro, and B. Aazhang, "Efficient VLSI architectures for multiuser channel estimation in wireless base-station receivers," *The Journal of VLSI Signal Processing*, vol. 31, no. 2, pp. 143–156, 2002.

[3] G. Dietl and W. Utschick, "On reduced-rank approaches to matrix Wiener filters in MIMO systems," in *Proceedings of the 3rd IEEE International Symposium on Signal Processing and Information Technology (ISSPIT '03)*, pp. 82–85, Darmstadt, Germany, December 2003.

[4] Y. Sun, M. L. Honig, and V. Tripathi, "Adaptive, iterative, reduced-rank equalization for MIMO channels," in *Proceedings of Military Communications Conference (MILCOM '02)*, vol. 2, pp. 1029–1033, Anaheim, Calif, USA, October 2002.

[5] A. F. Molisch and M. Z. Win, "MIMO systems with antenna selection," *IEEE Microwave Magazine*, vol. 5, no. 1, pp. 46–56, 2004.

[6] S. Cui, A. J. Goldsmith, and A. Bahai, "Energy-efficiency of MIMO and cooperative MIMO techniques in sensor networks," *IEEE Journal on Selected Areas in Communications*, vol. 22, no. 6, pp. 1089–1098, 2004.

[7] H. Viswanathan and J. Balakrishnan, "Space-time signaling for high data rates in EDGE," *IEEE Transactions on Vehicular Technology*, vol. 51, no. 6, pp. 1522–1533, 2002.

[8] R. C. Whaley and J. J. Dongarra, "Automatically tuned linear algebra software," in *Proceedings of 10th Anniversary. International Conference on High Performance Computing and Communications (SC '98)*, p. 33, Orlando, Fla, USA, November 1998.

[9] R. C. Whaley, A. Petitet, and J. J. Dongarra, "Automated empirical optimization of software and the ATLAS project," ATLAS project, 2000, http://netlib.org/atlas/.

[10] R. Henning and C. Chakrabarti, "A quality/energy tradeoff approach for IDCT computation in MPEG-2 video decoding," in *Proceedings of IEEE Workshop on Signal Processing Systems (SiPS '00)*, pp. 90–99, Lafayette, La, USA, October 2000.

[11] F. Catthoor, E. de Greef, and S. Suytack, *Custom Memory Management Methodology: Exploration of Memory Organisation for Embedded Multimedia System Design*, Kluwer Academic, Norwell, Mass, USA, 1998.

[12] D. Waters, "Complexity analysis of MIMO detectors," on line publication, 2003, http://users.ece.gatech.edu/~deric/Projects/MIMO/Complexity.pdf.

[13] G. H. Golub and C. F. Van Loan, *Matrix Computations*, Johns Hopkins University Press, Baltimore, Md, USA, 3rd edition, 1996.

[14] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes in C: The Art of Scientific Computing*, Cambridge University Press, Cambridge, UK, 1992.

[15] G. H. Hardy and S. Ramanujan, "Asymptotic formulae in combinatory analysis," *Proceedings of the London Mathematical Society*, vol. 17, no. 2, pp. 75–115, 1918.

[16] J. Nocedal and S. J. Wright, *Numerical Optimization*, Springer, New York, NY, USA, 1999.

Sarod Yatawatta obtained his B.S. degree in electrical and electronic engineering from the University of Peradeniya, Sri Lanka, in 2000. He obtained his Doctorate degree in electrical engineering from Drexel University in 2004. Currently he is working for the LOFAR Project in The Netherlands, developing software for next-generation radio telescopes. His research interests include signal processing, scientific computing, and wireless communications.

Athina P. Petropulu received the Diploma in electrical engineering from the National Technical University of Athens, Greece, in 1986, the M.S. degree in electrical and computer engineering in 1988, and the Ph.D. degree in electrical and computer engineering in 1991, both from Northeastern University, Boston, Mass. In 1992, she joined the Department of Electrical and Computer Engineering at Drexel University where she is now a Professor. Her research interests span the area of statistical signal processing, wireless communications and networking, and ultrasound imaging. She is the recipient of the 1995 Presidential Faculty Fellow Award. She is the coauthor (with C.L. Nikias) of the textbook entitled *Higher-Order Spectra Analysis: A Nonlinear Signal Processing Framework* (Prentice-Hall, Inc., Englewood Cliffs, NJ, 1993). She has served as an Associate Editor for the IEEE Transactions on Signal Processing, the IEEE Signal Processing Letters, the IEEE Signal Processing Magazine, and the EURASIP Journal on Wireless Communications and Networking. She is IEEE Signal Processing Society Vice-President-Conferences. She was the General Chair of the 2005 International Conference on Acoustics, Speech, and Signal Processing (ICASSP-05, Philadelphia, Pa).

Charles J. Graff received a BSEE and MSEE degrees from Drexel University, an M.S. degree in systems from Fairleigh Dickinson University, and Ph.D. degree in computer science from Stevens Institute of Technology. Since 1972, he has been a Senior Electronics Engineer involved with packet networking and radio-based communications for the US Army at Fort Monmouth, NJ.