Training-Based and Semiblind Channel Estimation for MIMO Systems With Maximum Ratio Transmission

Chandra R. Murthy, Aditya K. Jagannatham, and Bhaskar D. Rao, *Fellow, IEEE*

***Abstract*—This paper is a comparative study of training-based and semiblind multiple-input multiple-output (MIMO) flat-fading channel estimation schemes when the transmitter employs maximum ratio transmission (MRT). We present two competing schemes for estimating the transmit and receive beamforming vectors of the channel matrix: a training-based conventional least-squares estimation (CLSE) scheme and a closed-form semiblind (CFSB) scheme that employs training followed by information-bearing spectrally white data symbols. Employing matrix perturbation theory, we develop expressions for the mean-square error (MSE) in the beamforming vector, the average received signal-to-noise ratio (SNR) and the symbol error rate (SER) performance of both the semiblind and the conventional schemes. Finally, we describe a weighted linear combiner of the CFSB and CLSE estimates for additional improvement in performance. The analytical results are verified through Monte Carlo simulations.**

***Index Terms*—Beamforming, channel estimation, constrained estimation, Cramer–Rao bound, least squares, maximum ratio transmission (MRT), multiple-input multiple-output (MIMO), semiblind.**

I. Introduction

MULTIPLE-INPUT multiple-output (MIMO) and smart antenna systems have gained popularity due to the promise of a linear increase in achievable data rate with the number of antennas, and because they inherently benefit from effects such as channel fading. Maximum ratio transmission (MRT) is a particularly attractive beamforming scheme for MIMO communication systems because of its low implementation complexity. It is also known that MRT coupled with maximum ratio combining (MRC) leads to signal-to-noise ratio (SNR) maximization at the receiver and achieves a performance close to capacity in low-SNR scenarios. However, in order to realize these benefits, an accurate estimate of the channel is necessary. One standard technique to estimate the channel is to transmit a sequence of training symbols (also called pilot symbols) at the beginning of each frame. This training symbol sequence is known at the receiver and thus the channel is estimated from the measured outputs to training symbols. Training-based schemes usually have very low complexity making them ideally suited for implementation in systems (e.g., mobile stations) where the available computational capacity is limited.

However, the above training-based technique for channel estimation in MRT-based MIMO systems is transmission scheme agnostic. For example, channel estimation algorithms when MRT is employed at the transmitter only need to estimate $\mathbf{v}_1$ and $\mathbf{u}_1$, where $\mathbf{v}_1$ and $\mathbf{u}_1$ are the dominant eigenvectors of $H^H H$ and HH^H respectively, H is the $r \times t$ channel transfer matrix, and r/t are the number of receive/transmit antennas. Hence, techniques that estimate the entire H matrix from a set of training symbols and use the estimated H to compute $\mathbf{v}_1$ and $\mathbf{u}_1$ may be inefficient, compared to techniques designed to use the training data specifically for estimating the beamforming vectors. Moreover, as r increases, the mean-square error (MSE) in estimation of $\mathbf{v}_1 \in \mathbb{C}^t$ remains constant since the number of unknown parameters in $\mathbf{v}_1$ does not change with r, while that of H increases since the number of elements, rt, grows linearly with r. Added to this, the complexity of reliably estimating the channel increases with its dimensionality. The channel estimation problem is further complicated in MIMO systems because the SNR per bit required to achieve a given system throughput performance decreases as the number of antennas is increased. Such low-SNR environments call for more training symbols, lowering the effective data rate.

For the above reasons, semiblind techniques can enhance the accuracy of channel estimation by efficiently utilizing not only the known training symbols but also the unknown data symbols. Hence, they can be used to reduce the amount of training data required to achieve the desired system performance, or equivalently, achieve better accuracy of estimation for a given number of training symbols, thereby improving the spectral efficiency and channel throughput. Work on semiblind techniques for the design of fractional semiblind equalizers in multipath channels has been reported earlier by Pal in [2] and [3]. In [4] and [5] error bounds and asymptotic properties of blind and semiblind techniques are analyzed. In [6]–[8], an orthogonal pilot-based maximum-likelihood (OPML) semiblind estimation scheme is proposed, where the channel matrix H is factored into the product of a whitening matrix W and a unitary rotation matrix Q. W is estimated from the data using a blind algorithm, while Q is estimated exclusively from the training data using the OPML algorithm. However, feedback-based transmission schemes such as MRT pose new challenges for semiblind estimation, because employment of the precoder (beamforming vector) corresponding to an erroneous channel estimate precludes the use of the received data symbols to improve the channel estimate. This necessitates the development of new

Manuscript received February 23, 2005; revised August 23, 2005. This work was supported in part by UC Discovery grant nos. Com 02-10105, Com 02-10119, Intersil Corp. and the UCSD Center for Wireless Communications. A conference version of this work appears in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing* (ICASSP), vol. 3, Mar. 2005, pp. 585–588. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Luc Vandendorpe.

The authors are with the Department of Electrical and Computer Engineering, University of California, San Diego, La Jolla, CA 92093-0407 USA (e-mail: cmurthy@ucsd.edu; ajaganna@ucsd.edu; brao@ucsd.edu).

Digital Object Identifier 10.1109/TSP.2006.874780

transmission schemes to enable implementation of semiblind estimation, as shown in Section II-C. Furthermore, the proposed techniques specifically estimate the MRT beamforming vector, and hence can potentially achieve better estimation accuracy compared with techniques that are independent of the transmission scheme.

The contributions of this paper are as follows. We describe the training-only-based conventional least-squares estimation (CLSE) algorithm, and derive analytical expressions for the MSE in the beamforming vector, the mean received SNR and the symbol error rate (SER) performance. For improved spectral efficiency (reduced training overhead), we propose a closed-form semiblind (CFSB) algorithm that estimates $\mathbf{u}_1$ from the data using a blind algorithm, and estimates $\mathbf{v}_1$ exclusively from the training. This necessitates the introduction of a new signal transmission scheme that involves transmission of information-bearing spectrally white data symbols to enable semiblind estimation of the beamforming vectors. Expressions are derived for the performance of the proposed CFSB scheme. We show that given perfect knowledge of $\mathbf{u}_1$ (which can be achieved when there are a large number of white data symbols), the error in estimating $\mathbf{v}_1$ using the CFSB scheme asymptotically achieves the theoretical Cramer–Rao lower bound (CRB), and thus the CFSB scheme outperforms the CLSE scheme. However, there is a tradeoff in transmission of white data symbols in semiblind estimation, since the SER for the white data is frequently greater than that for the beamformed data. Thus, we show that there exist scenarios where for a reasonable number of white data symbols, the gains from beamformed data for this improved estimate in CFSB outweigh the loss in performance due to transmission of white data. As a more general estimation method when a given number of blind data symbols are available, we propose a new scheme that judiciously combines the above described CFSB and CLSE estimates based on a heuristic criterion. Through Monte Carlo simulations, we demonstrate that this proposed linearly combined semiblind (LCSB) scheme outperforms the CLSE and CFSB scheme in terms of both estimation accuracy as well as SER and thus achieves good performance.

The rest of this paper is organized as follows. In Section II, we present the problem setup and notation. We also present both the CFSB and CLSE schemes in detail. The MSE and the received SNR performance of the CLSE scheme are derived using a first-order perturbation analysis in Section III and the performance of the CFSB scheme is analyzed in Section IV. In Section V, to conduct an end-to-end system comparison, we derive the performance of Alamouti space–time coded data with training-based channel estimation, and present the proposed LCSB algorithm. We compare the different schemes through Monte Carlo simulations in Section VI and present our conclusions in Section VII.

II. Preliminaries

A. System Model and Notation

Fig. 1 shows the MIMO system model with beamforming at the transmitter and the receiver. We model a flat-fading channel by a complex-valued channel matrix $H \in \mathbb{C}^{r\times t}$. We assume that

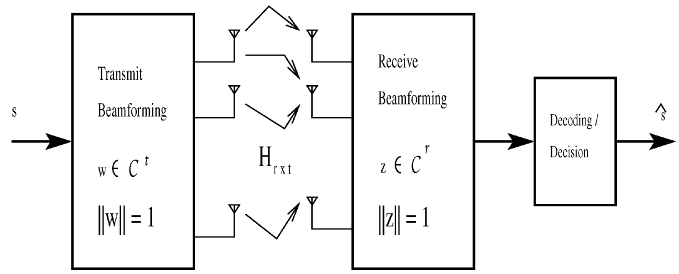

Fig. 1. MIMO system model, with beamforming at the transmitter and receiver.

H is quasi-static and constant over the period of one transmission block. We denote the singular value decomposition (SVD) of H by $H = U\Sigma V^H$, and $\Sigma \in \mathbb{R}^{r\times t}$ contains singular values $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_m > 0$, along the diagonal, where $m = \mathrm{rank}(H)$. Let $\mathbf{v}_1$ and $\mathbf{u}_1$ denote the first columns of V and U, respectively.

The channel input–output relation at time instant k is

$$\mathbf{y}_k = H\mathbf{x}_k + \mathbf{n}_k, \tag{1}$$

where $\mathbf{x}_k \in \mathbb{C}^t$ is the channel input, $\mathbf{y}_k \in \mathbb{C}^r$ is the channel output, and $\mathbf{n}_k \in \mathbb{C}^r$ is the spatially and temporally white noise vector with independent and identically distributed (i.i.d.) zero-mean circularly symmetric complex Gaussian (ZMCSCG) entries. The input $\mathbf{x}_k$ could denote data or training symbols. Also, we let the noise power in each receive antenna be unity, that is, $\mathbf{E}\{\mathbf{n}_k\mathbf{n}_k^H\} = I_r$, where $\mathbf{E}\{\cdot\}$ denotes the expectation operation, and I_r is the $r \times r$ identity matrix.

Let L training symbol vectors be transmitted at an average power P_{T} per vector (T stands for "training"). The training symbols are stacked together to form a training symbol matrix $X_{\mathrm{p}} \in \mathbb{C}^{t\times L}$ as $X_{\mathrm{p}} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_L]$ (p stands for "pilot"). We employ orthogonal training sequences because of their optimality properties in channel estimation [9]. That is, $X_{\mathrm{p}}X_{\mathrm{p}}^H = \gamma_{\mathrm{p}}I_{\mathrm{t}}$, where $\gamma_{\mathrm{p}} \triangleq LP_{\mathrm{T}}/t$, thus maintaining the training power of P_{T}. The data symbols $\mathbf{x}_k$ could either be spatially white (i.e., $\mathbf{E}\{\mathbf{x}_k\mathbf{x}_k^H\} = (P_{\mathrm{D}}/t)I_{\mathrm{t}}$), or it could be the result of using beamforming at the transmitter with unit-norm weight vector $\mathbf{w} \in \mathbb{C}^{t\times 1}$ (i.e., $\mathbf{E}\{\mathbf{x}_k\mathbf{x}_k^H\} = P_{\mathrm{D}}\mathbf{w}\mathbf{w}^H$), where the data transmit power is $\mathbf{E}\{\mathbf{x}_k^H\mathbf{x}_k\} = P_{\mathrm{D}}$ (D stands for "data"). We let N denote the number of spatially white data symbols transmitted, that is, a total of $N + L$ symbols are transmitted prior to transmitting beamformed data. Note that the N white data symbols carry (unknown) information bits, and hence are not a waste of available bandwidth.

In this paper, we restrict our attention to the case where the transmitter employs MRT to send data, that is, a *single* data stream is transmitted over t transmit antennas after passing through a beamformer $\mathbf{w}$. Given the channel matrix H, the optimum choice of $\mathbf{w}$ is $\mathbf{v}_1$ [10]. Thus, MRT only needs an accurate estimate of $\mathbf{v}_1$ to be fed back to the transmitter. We assume that $t \geq 2$, since when $t = 1$, estimation of the beamforming vector has no relevance. Finally, we will compare the performance of different estimation techniques using several different measures, namely, the MSE in the estimate of $\mathbf{v}_1$, the gain (rather, the power amplification/attenuation), and

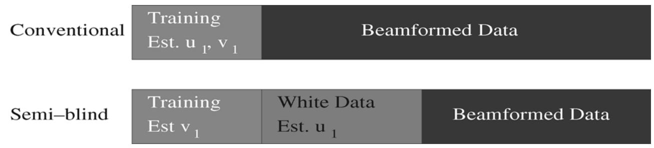

Fig. 2. Comparison of the transmission scheme for conventional least-squares (CLSE) and closed-form semiblind (CFSB) estimation.

the symbol error rate (SER) of the one-dimensional channel resulting from beamforming with the estimated vector $\hat{\mathbf{v}}_1$ (assuming uncoded M-ary QAM transmission). The performance of a practical communication would also be affected by factors such as quantization error in $\hat{\mathbf{v}}_1$, errors in the feedback channel, feedback delay in time-varying environments, etc., and a detailed study of these factors warrant separate treatment.

B. Conventional Least-Squares Estimation

Conventionally, a maximum-likelihood (ML) estimate of the channel matrix, $\hat{H}_c$, is first obtained from the training data as the solution to the following least-squares problem:

$$\hat{H}_c = \arg \min_{G \in \mathbb{C}^{r \times t}} \|Y_{\mathrm{p}} - GX_{\mathrm{p}}\|_F^2 \tag{2}$$

where $\|\cdot\|_F$ represents the Frobenius norm, Y_{p} is the $r \times L$ matrix of received symbols given by $Y_{\mathrm{p}} = HX_{\mathrm{p}} + \eta_{\mathrm{p}}$, where $\eta_{\mathrm{p}} \in \mathbb{C}^{r \times L}$ is the set of additive white Gaussian noise ((AWGN) spatially and temporally white) vectors. From [11], the solution to this least-squares estimation problem can be shown to be $\hat{H}_c = Y_{\mathrm{p}} X_{\mathrm{p}}^{\dagger}$, where $X_{\mathrm{p}}^{\dagger}$ is the Moore–Penrose generalized inverse of X_{p}. Since orthogonal training sequences are employed, we have $X_{\mathrm{p}}^{\dagger} = (1/\gamma_{\mathrm{p}}) X_{\mathrm{p}}^H$, and consequently

$$\hat{H}_c = \frac{1}{\gamma_{\mathrm{p}}} Y_{\mathrm{p}} X_{\mathrm{p}}^H. \tag{3}$$

The ML estimates of $\mathbf{v}_1$ and $\mathbf{u}_1$, denoted $\hat{\mathbf{v}}_c$ and $\hat{\mathbf{u}}_c$ respectively, are now obtained via an SVD of the estimated channel matrix $\hat{H}_c$. Since $\hat{H}_c$ is the ML estimate of H, from properties of ML estimation of principal components [12], the $\hat{\mathbf{v}}_c$ obtained by this technique is also the ML estimate of $\mathbf{v}_1$ given only the training data.

C. Semiblind Estimation

In the scenario that the transmitted data symbols are spatially-white, the ML estimate of $\mathbf{u}_1$ is the dominant eigenvector of the output correlation matrix $\hat{R}_y$, which is estimated as $\hat{R}_y = \sum_{i=1}^{N} \mathbf{y}_i \mathbf{y}_i^H$. Now, the estimate of $\mathbf{u}_1$ is obtained by computing the following SVD:

$$\hat{U} \hat{\Sigma}^2 \hat{U}^H = \hat{R}_y. \tag{4}$$

Note that it is possible to use the entire received data to compute $\hat{R}_y$ in (4) rather than just the data symbols, in this case, N should be changed to $N + L$. The estimate of $\mathbf{u}_1$, denoted $\hat{\mathbf{u}}_{\mathrm{s}}$ (the subscript "s" stands for semiblind), is thus computed blind from the received data as the first column of $\hat{U}$. As N grows, a near perfect estimate of $\mathbf{u}_1$ can be obtained.

In order to estimate $\mathbf{u}_1$ as described above, it is necessary that the transmitted symbols be *spatially-white*. If the transmitter uses any (single) beamforming vector $\mathbf{w}$, the expected value of the correlation at the receiver is $H\mathbf{w}(H\mathbf{w})^H = H\mathbf{w}\mathbf{w}^H H^H \neq HH^H$, and hence, the estimated eigenvector will be a vector proportional to $H\mathbf{w}$ instead of $\mathbf{u}_1$. Fig. 2 shows a schematic representation of the CLSE and the CFSB schemes. Thus, the CFSB scheme involves a two-phase data transmission: spatially white data followed by beamformed data. White data transmission could lead to a loss of performance relative to beamformed data, but this performance loss can be compensated for by the gain obtained from the improved estimate of the MRT beamforming vector. Thus, the semiblind scheme can have an overall better performance than the CLSE scheme. Section V presents an overall SER comparison in a practical scenario, after accounting for the performance of the white data as well as for the beamformed data.

Having obtained the estimate of $\mathbf{u}_1$ from the white data, the training symbols are now exclusively used to estimate $\mathbf{v}_1$. Since the vector $\mathbf{v}_1$ has fewer real parameters $(2t-1)$ than the channel matrix H $(2rt)$, it is expected to achieve a greater accuracy of estimation for the same number of training symbols, compared to the CLSE technique which requires an accurate estimate of the full H matrix in order to estimate $\mathbf{v}_1$ accurately. If $\mathbf{u}_1$ is estimated perfectly from the blind data, the received training symbols can be filtered by $\mathbf{u}_1^H$ to obtain

$$\mathbf{u}_1^H Y_{\mathrm{p}} = \sigma_1 \mathbf{v}_1^H X_{\mathrm{p}} + \mathbf{u}_1^H \eta_{\mathrm{p}}. \tag{5}$$

Since $\|\mathbf{u}_1\| = 1$, (here $\|\cdot\|$ represents the 2-norm) the statistics of the Gaussian noise η_{p} are unchanged by the above operation. We seek the estimate of $\mathbf{v}_1$ as the solution to the following least-squares problem:

$$\hat{\mathbf{v}}_{\mathrm{s}} = \arg \min_{\mathbf{v} \in \mathbb{C}^t, \|\mathbf{v}\|=1} \left\| \mathbf{u}_1^H Y_{\mathrm{p}} - \mathbf{v}^H X_{\mathrm{p}} \sigma_1 \right\|^2 \tag{6}$$

where $\hat{\mathbf{v}}_{\mathrm{s}}$ denotes the semiblind estimate of $\mathbf{v}_1$. The following lemma establishes the solution.

Lemma 1: If X_{p} satisfies $X_{\mathrm{p}} X_{\mathrm{p}}^H = \gamma_{\mathrm{p}} I_{\mathrm{t}}$, the least-squares estimate of $\mathbf{v}_1$ (under $\|\mathbf{v}_1\| = 1$) given perfect knowledge of $\mathbf{u}_1$ is

$$\hat{\mathbf{v}}_{\mathrm{s}} = \frac{X_{\mathrm{p}} Y_{\mathrm{p}}^H \mathbf{u}_1}{\left\| X_{\mathrm{p}} Y_{\mathrm{p}}^H \mathbf{u}_1 \right\|}. \tag{7}$$

Proof: See Section A of the Appendix. ■

Closed-Form Semiblind Estimation Algorithm: Based on the above observations, the proposed CFSB algorithm is as follows. First, we obtain $\hat{\mathbf{u}}_{\mathrm{s}}$, the estimate of $\mathbf{u}_1$, from (4). Then, we estimate $\mathbf{v}_1$ from the L training symbols by substituting $\hat{\mathbf{u}}_{\mathrm{s}}$ for $\mathbf{u}_1$ in (7). This requires $L + N$ symbols to actually estimate $\mathbf{v}_1$, however, N of these symbols are data symbols (which carry information bits). Hence, we can potentially achieve the desired accuracy of estimation of $\mathbf{v}_1$ using fewer training symbols compared to the CLSE technique.

An alternative to employing $\hat{\mathbf{u}}_1$ at the receiver is employ maximum ratio combining (MRC), i.e., to use an estimate of $H\hat{\mathbf{v}}_1/\|H\hat{\mathbf{v}}_1\|$ (which can be accurately estimated as the dominant eigenvector of the sample covariance matrix of the beamformed data). The performance of such a scheme is summarized in [1], and the analysis can be carried out along the lines presented in this paper.

III. Conventional Least-Squares Estimation

A. Perturbation of Eigenvectors

We recapitulate a result from matrix perturbation theory [13] that we will use frequently in the sequel. Consider a first-order perturbation of a Hermitian symmetric matrix R by an error matrix ΔR to get $\hat{R}$, that is, $\hat{R} = R + \Delta R$. Then, if the eigenvalues of R are distinct, for small perturbations, the eigenvectors $\hat{\mathbf{s}}_k$ of $\hat{R}$ can be approximately expressed in terms of the eigenvectors $\mathbf{s}_k$ of R as

$$\hat{\mathbf{s}}_k = \mathbf{s}_k + \sum_{\substack{r=1 \\ r\neq k}}^{n} \frac{\mathbf{s}_r^H \Delta R \mathbf{s}_k}{\lambda_k - \lambda_r} \mathbf{s}_r \tag{8}$$

where n is the rank of R, λ_k is the kth eigenvalue of R, and $\lambda_k \neq \lambda_j$, $k \neq j$.

When $k = 1$, we have $\hat{\mathbf{s}}_1 = S\check{\mathbf{d}}$, where $S = [\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_n]$ is the matrix of eigenvectors and $\check{\mathbf{d}} = [1, \mathbf{s}_2^H \Delta R \mathbf{s}_1/(\lambda_1 - \lambda_2), \ldots, \mathbf{s}_n^H \Delta R \mathbf{s}_1/(\lambda_1 - \lambda_n)]^T$. One could scale the vector $\hat{\mathbf{s}}_1$ to construct a unit-norm vector as $\tilde{\mathbf{s}}_1 = \hat{\mathbf{s}}_1/\|\hat{\mathbf{s}}_1\|$. Then, $\tilde{\mathbf{s}}_1 = S\mathbf{d}$, where $\mathbf{d} = \check{\mathbf{d}}/\|\check{\mathbf{d}}\| = [1 + \Delta d_1, \Delta d_2, \ldots, \Delta d_n]^T$. Following an approach similar to [14], if Δd_i are small, since $\|\mathbf{d}\| = 1$, the components Δd_i are approximately given by

$$\Delta d_i \simeq \frac{\mathbf{s}_i^H \Delta R \mathbf{s}_1}{\lambda_1 - \lambda_i}, \quad i = 2, \ldots, n$$
$$\Delta d_1 \simeq -\frac{1}{2} \sum_{i=1}^{n} |\Delta d_i|^2. \tag{9}$$

Note that Δd_1 is real, and is a higher-order term compared to Δd_i, $i \geq 2$. We will use this fact in our first-order approximations to ignore terms such as $|\Delta d_1|^2, |\Delta d_1|^3, \ldots$ and $|\Delta d_i|^3, |\Delta d_i|^4, \ldots, i \geq 2$. In the sequel, we assume that the dominant singular value of H is distinct, so the conditions required for the above result are valid.

B. MSE in $\hat{\mathbf{v}}_c$

To compute the MSE in $\hat{\mathbf{v}}_c$, we use (3) to write the matrix $\hat{H}_c^H \hat{H}_c$ as a perturbation of $H^H H$ and use the above matrix perturbation result to derive the desired expressions.

$$\hat{H}_c^H \hat{H}_c = V \Sigma^2 V^H + E_{\mathrm{T}} \tag{10}$$

where $E_{\mathrm{T}} \approx [V \Sigma U^H E_{\mathrm{p}} + E_{\mathrm{p}}^H U \Sigma V^H]$ with $E_{\mathrm{p}} = (1/\gamma_{\mathrm{p}}) \eta_{\mathrm{p}} X_{\mathrm{p}}^H$. Here, we have ignored the $E_{\mathrm{p}}^H E_{\mathrm{p}}$ term in writing the expression for E_{T}, since it is a second-order term due to the $(1/\gamma_{\mathrm{p}})$ factor in E_{p}. Now, we can regard E_{T} as a perturbation of the matrix $H^H H$. As seen in Section II-B, $\hat{\mathbf{v}}_c$ is estimated from the SVD of $\hat{H}_c$. Since the basis vectors V span $\mathbb{C}^t$, we can let $\hat{\mathbf{v}}_c = V\mathbf{d}$, and write $\mathbf{d} = [1 + \Delta d_1, \Delta d_2, \ldots, \Delta d_{\mathrm{t}}]^T$ as a perturbation of $[1, 0, \ldots, 0]^T$.

For $i \geq 2$, Δd_i is obtained from (9) as

$$\Delta d_i = \frac{\mathbf{v}_i^H E_{\mathrm{T}} \mathbf{v}_1}{\sigma_1^2 - \sigma_i^2} = \frac{\sigma_i \mathbf{u}_i^H E_{\mathrm{p}} \mathbf{v}_1 + \sigma_1 \mathbf{v}_i^H E_{\mathrm{p}}^H \mathbf{u}_1}{\sigma_1^2 - \sigma_i^2}. \tag{11}$$

Note that, if $r < t$, we have $\sigma_i = 0$, $i > r$, hence, $\Delta d_i = \mathbf{v}_i^H E_{\mathrm{p}}^H \mathbf{u}_1/\sigma_1$, for $i > r$. Therefore, to simplify notation, we can define $\mathbf{u}_i \triangleq \mathbf{0}_{r\times 1}$ and $\mathbf{v}_j \triangleq \mathbf{0}_{t\times 1}$, for $i > r$ and $j > t$ respectively. The following result is used to find $\mathbf{E}\{|\Delta d_i|^2\}$.

Lemma 2: Let $\mu_1, \mu_2 \in \mathbb{C}$ be fixed complex numbers. Let $\sigma_{\mathrm{p}}^2 = (1/\gamma_{\mathrm{p}})$ denote the variance of one of the elements of E_{p}. Then

$$\mathbf{E}\left\{\left|\mu_1 \mathbf{u}_i^H E_{\mathrm{p}} \mathbf{v}_j + \mu_2 \mathbf{v}_i^H E_{\mathrm{p}}^H \mathbf{u}_j\right|^2\right\} = \sigma_{\mathrm{p}}^2 \left(|\mu_1|^2 + |\mu_2|^2\right) \tag{12}$$

for any $1 \leq i \leq r, 1 \leq j \leq t$.

Proof: Let $a \triangleq \mathbf{u}_i^H E_{\mathrm{p}} \mathbf{v}_j$ and $b \triangleq \mathbf{v}_i^H E_{\mathrm{p}}^H \mathbf{u}_j$. Then, from lemma in Section E of the Appendix, a and b are circularly symmetric random variables. Since E_{p} is circularly symmetric $(\mathbf{E}\{E_{\mathrm{p}}(i,j) E_{\mathrm{p}}(k,l)\} = 0, \forall i,j,k,l)$ and a and b^* are both linear combinations of elements of E_{p}, we have $\mathbf{E}\{ab^*\} = 0$. Finally, since $\|\mathbf{u}_i\| = \|\mathbf{v}_j\| = 1$, the variance of a and b are equal, and $\sigma_a^2 = \sigma_b^2 = \sigma_{\mathrm{p}}^2$. Substituting, we have

$$\begin{aligned}\mathbf{E}\left\{\left|\mu_1 \mathbf{u}_i^H E_{\mathrm{p}} \mathbf{v}_j + \mu_2 \mathbf{v}_i^H E_{\mathrm{p}}^H \mathbf{u}_j\right|^2\right\} &= |\mu_1|^2 \sigma_a^2 + |\mu_2|^2 \sigma_b^2 \\ &= \sigma_{\mathrm{p}}^2 \left(|\mu_1|^2 + |\mu_2|^2\right).\end{aligned}$$

■

Using the above lemma with $\mu_1 = \sigma_i$, $\mu_2 = \sigma_1$ and $j = 1$, we get, for $i \geq 2$,

$$\mathbf{E}\left\{|\Delta d_i|^2\right\} = \sigma_{\mathrm{p}}^2 \frac{\sigma_1^2 + \sigma_i^2}{\left(\sigma_1^2 - \sigma_i^2\right)^2} \tag{13}$$

where the expectation is taken with respect to the AWGN term η_{p}. The following lemma helps simplify the expression further. We omit the proof, as it is straightforward.

Lemma 3: If $\hat{\mathbf{v}}_c = V\mathbf{d}$, then

$$\|\hat{\mathbf{v}}_c - \mathbf{v}_1\|^2 = 2\left(1 - \mathrm{Re}(d_1)\right) = -\left(\Delta d_1 + \Delta d_1^*\right) \tag{14}$$

where $d_1 = 1 + \Delta d_1$ is the first element of $\mathbf{d}$.

Using (13) in (9) and substituting into in (14), the final estimation error is

$$\mathbf{E}\left\{\|\hat{\mathbf{v}}_c - \mathbf{v}_1\|^2\right\} = \frac{1}{\gamma_{\mathrm{p}}} \sum_{i=2}^{t} \frac{\sigma_1^2 + \sigma_i^2}{\left(\sigma_1^2 - \sigma_i^2\right)^2}. \tag{15}$$

C. Received SNR and Symbol Error Rate

In this section, we derive the expression for the received SNR when beamforming using $\hat{\mathbf{v}}_c$ at the transmitter and filtering using $\hat{\mathbf{u}}_c$ at the receiver. Since the unitary matrices V and U span $\mathbb{C}^t$ and $\mathbb{C}^r$, $\hat{\mathbf{v}}_c$ and $\hat{\mathbf{u}}_c$ can be expressed as $\hat{\mathbf{u}}_c = U\mathbf{c}$ and $\hat{\mathbf{v}}_c = V\mathbf{d}$, respectively. Borrowing notation from Section III-B, let $\mathbf{c} = [1 + \Delta c_1, \Delta c_2, \ldots, \Delta c_r]^T \in \mathbb{C}^r$ and $\mathbf{d} = [1 + \Delta d_1, \Delta d_2, \ldots, \Delta d_{\mathrm{t}}]^T \in \mathbb{C}^t$ respectively. Then, $\mathbf{c}$ can be derived by a perturbation analysis on $\hat{H}_c\hat{H}_c^H$ analogous to that in (10) in Section III-B. We obtain

$$\Delta c_i = \frac{\sigma_i \mathbf{v}_i^H E_{\rm p}^H \mathbf{u}_1 + \sigma_1 \mathbf{u}_i^H E_{\rm p} \mathbf{v}_1}{\sigma_1^2 - \sigma_i^2}$$

where, as before, we define $\sigma_i = 0$, and $\mathbf{u}_i \triangleq \mathbf{0}_{r\times 1}, \mathbf{v}_j \triangleq \mathbf{0}_{t\times 1}$, for $i > r$ and $j > t$ respectively, so that $\Delta c_i = 0$; $i > r$, as expected. The channel gain is given by

$$\hat{\mathbf{u}}_c^H H \hat{\mathbf{v}}_c = \mathbf{c}^H \Sigma \mathbf{d} = \sigma_1(1+\Delta d_1)(1+\Delta c_1^*) + \sum_{i=2}^{t} \sigma_i \Delta c_i^* \Delta d_i.$$

Ignoring higher-order terms (cf. Section II-A), the power amplification $\rho_c \triangleq \mathbf{E}\{|\hat{\mathbf{u}}_c^H H \hat{\mathbf{v}}_c|^2\}$ is

$$\rho_c \approx \sigma_1^2 \mathbf{E}\left\{ 1 + (\Delta d_1 + \Delta d_1^*) + (\Delta c_1 + \Delta c_1^*) + \sum_{i=2}^{t} \frac{\sigma_i}{\sigma_1} (\Delta c_i \Delta d_i^* + \Delta c_i^* \Delta d_i) \right\}. \quad (16)$$

From (9) and (13), we have

$$\mathbf{E}\{\Delta d_1 + \Delta d_1^*\} = -\frac{1}{\gamma_{\rm p}} \sum_{i=2}^{t} \frac{\sigma_1^2 + \sigma_i^2}{(\sigma_1^2 - \sigma_i^2)^2}$$
$$\mathbf{E}\{\Delta c_1 + \Delta c_1^*\} = -\frac{1}{\gamma_{\rm p}} \sum_{i=2}^{r} \frac{\sigma_1^2 + \sigma_i^2}{(\sigma_1^2 - \sigma_i^2)^2}.$$

Now, $\Delta c_i \Delta d_i^*$ can be written as

$$\begin{aligned}\mathbf{E}\{\Delta c_i \Delta d_i^*\} &= \left\{ \mathbf{E}\left(\frac{\sigma_i \mathbf{v}_i^H E_{\rm p}^H \mathbf{u}_1 + \sigma_1 \mathbf{u}_i^H E_{\rm p} \mathbf{v}_1}{\sigma_1^2 - \sigma_i^2} \right) \times \left(\frac{\sigma_i \mathbf{v}_1^H E_{\rm p}^H \mathbf{u}_i + \sigma_1 \mathbf{u}_1^H E_{\rm p} \mathbf{v}_i}{\sigma_1^2 - \sigma_i^2} \right) \right\} \\ &= \mathbf{E}\left\{ \frac{\sigma_1\sigma_i \left(\left|\mathbf{u}_i^H E_{\rm p} \mathbf{v}_1\right|^2 + \left|\mathbf{v}_i^H E_{\rm p}^H \mathbf{u}_1\right|^2 \right)}{(\sigma_1^2 - \sigma_i^2)^2} \right\} \\ &= \frac{2\sigma_{\rm p}^2 \sigma_1 \sigma_i}{(\sigma_1^2 - \sigma_i^2)^2} = \frac{2\sigma_1\sigma_i}{\gamma_{\rm p}(\sigma_1^2 - \sigma_i^2)^2}.\end{aligned}$$

And likewise for $\Delta c_i^* \Delta d_i$. Denoting $m \triangleq \mathrm{rank}(H)$, the power amplification is

$$\begin{aligned}\rho_c &= \sigma_1^2 \left(1 - \frac{1}{\gamma_{\rm p}} \sum_{i=2}^{t} \frac{\sigma_1^2 + \sigma_i^2}{(\sigma_1^2 - \sigma_i^2)^2} - \frac{1}{\gamma_{\rm p}} \sum_{i=2}^{r} \frac{\sigma_1^2 + \sigma_i^2}{(\sigma_1^2 - \sigma_i^2)^2} + \frac{4}{\gamma_{\rm p}} \sum_{i=2}^{m} \frac{\sigma_i^2}{(\sigma_1^2 - \sigma_i^2)^2} \right) \\ &= \sigma_1^2 - \frac{2}{\gamma_{\rm p}} \sum_{i=2}^{m} \frac{\sigma_1^2}{\sigma_1^2 - \sigma_i^2} - \frac{1}{\gamma_{\rm p}}(r + t - 2m). \quad (17)\end{aligned}$$

In obtaining (17), we have used the fact that $\sigma_i = 0$ for $i > m$, where $m = \mathrm{rank}(H)$. Finally, the received SNR is

$$\mathrm{SNR} = \rho_c P_{\rm D} \quad (18)$$

where $P_{\rm D}$ is the power per data symbol. The power amplification with perfect knowledge of H at the transmitter and the receiver is $\rho_{\rm p} \triangleq \sigma_1^2$. As $\gamma_{\rm p} = LP_{\rm T}/t$ increases, ρ_c approaches $\rho_{\rm p}$. Note that, when $r = 1$, the above expression simplifies to $\rho_c = \rho_{\rm p} - (t-1)/\gamma_{\rm p}$. Also, $\sum_{i=2}^{m} \sigma_1^2(\sigma_1^2 - \sigma_i^2) \geq (m-1)$ since $\sigma_1 \geq \sigma_i$. Hence, if $r = t$, the CLSE performs best when the channel is spatially single dimensional (for example, in keyhole channels or highly correlated channels), that is, $\sigma_i = 0$, $i \geq 2$. In this case, we have $\rho_c = \rho_{\rm p} - (2/\gamma_{\rm p})(t-1)$. At the other extreme, if the dominant singular values are very close to each other such that $(\sigma_1^2 - \sigma_2^2) < 2/\gamma_{\rm p}$, the analysis is incorrect because it requires that the dominant singular values of H be sufficiently separated. For Rayleigh fading channels, i.e., H has i.i.d. ZMCSCG entries of unit variance, we can numerically evaluate the probability $Pr\{\sigma_1^2 - \sigma_2^2 < 2/\gamma_{\rm p}\}$ to be approximately 1.7×10^{-4}, with $r = t = 4$ and a typical value of $\gamma_{\rm p} =$ 10 dB. Thus, the above analysis is valid for most channel instantiations and practical SNRs.

Having determined the expected received SNR for a given channel instantiation, assuming uncoded M-ary QAM transmission, the corresponding SER P_M is given as [15]

$$P_{\sqrt{M}}(\rho_c) = 2\left[1 - \frac{1}{\sqrt{M}}\right] Q\left(\sqrt{\frac{3\rho_c P_{\rm T}}{M-1}}\right) \quad (19)$$
$$P_M(\rho_c) = 1 - \left(1 - P_{\sqrt{M}}(\rho_c)\right)^2 \quad (20)$$

where $Q(\cdot)$ is the Gaussian Q-function, and ρ_c is given by (17). The above expression can now be averaged over the probability density function of σ_i^2 through numerical integration.

IV. Closed-Form Semiblind Estimation

First, recall that the first-order Taylor expansion of a function of two variables $g(x, y)$ is given by

$$\begin{aligned}g(x + \Delta x, y + \Delta y) - g(x, y) &= \frac{\partial g(x,y)}{\partial x}\Delta x + \frac{\partial g(x,y)}{\partial y}\Delta y + O(\Delta x^2) + O(\Delta y^2) \\ &\approx [g(x + \Delta x, y) - g(x, y)] + [g(x, y + \Delta y) - g(x, y)].\end{aligned} \quad (21)$$

Now, in CFSB, the error in $\mathbf{v}_1$ (or loss in SNR) occurs due to two reasons: first, the noise in the received training symbols, and second, the use of an imperfect estimate of $\mathbf{u}_1$ (from the noise in the data symbols and availability of only a finite number N of unknown white data). More precisely, let the estimator of $\mathbf{v}_1$ be expressed as a function $\hat{\mathbf{v}}_{\rm s} = f(Y_{\rm p}, \hat{\mathbf{u}}_{\rm s})$ of the two variables $Y_{\rm p}$ and $\hat{\mathbf{u}}_{\rm s}$. Using the above expansion, we have

$$\begin{aligned}f(Y_{\rm p}, \hat{\mathbf{u}}_{\rm s}) - f(HX_{\rm p}, \mathbf{u}_1) \approx{}& [f(Y_{\rm p}, \mathbf{u}_1) - f(HX_{\rm p}, \mathbf{u}_1)] \\ &+ [f(HX_{\rm p}, \hat{\mathbf{u}}_{\rm s}) - f(HX_{\rm p}, \mathbf{u}_1)]\end{aligned} \quad (22)$$

where $\hat{\mathbf{v}}_{\rm s} = f(Y_{\rm p}, \hat{\mathbf{u}}_{\rm s})$ and from (7), $\mathbf{v}_1 = f(HX_{\rm p}, \mathbf{u}_1)$. Since the training noise $\eta_{\rm p}$ and the error in the estimate $\hat{\mathbf{u}}_{\rm s}$ are mutually independent, we get

$$\mathbf{E}\left\{\|\hat{\mathbf{v}}_{\rm s} - \mathbf{v}_1\|^2\right\} \approx \underbrace{\mathbf{E}\left\{\|f(Y_{\rm p}, \mathbf{u}_1) - f(HX_{\rm p}, \mathbf{u}_1)\|^2\right\}}_{T_1} + \underbrace{\mathbf{E}\left\{\|f(HX_{\rm p}, \hat{\mathbf{u}}_{\rm s}) - f(HX_{\rm p}, \mathbf{u}_1)\|^2\right\}}_{T_2}. \quad (23)$$

Note that the term T_1 represents the MSE in $\hat{\mathbf{v}}_{\rm s}$ as if the receiver had perfect knowledge of $\hat{\mathbf{u}}_{\rm s}$ (i.e., $\hat{\mathbf{u}}_{\rm s} = \mathbf{u}_1$), and the term T_2 represents the MSE in $\hat{\mathbf{v}}_{\rm s}$ when the training symbols are noise-free (i.e., $Y_{\rm p} = HX_{\rm p}$). Hence, the error in $\hat{\mathbf{v}}_{\rm s}$ can be thought of as the sum of two terms: the first one being the error due to the noise in the white (unknown) data, and the second being the error due to the noise in the training data. A similar decomposition can be used to express the loss in channel gain (relative to σ_1).

A. MSE in $\hat{\mathbf{v}}_{\rm s}$ With Perfect $\hat{\mathbf{u}}_{\rm s}$

In this section we consider the error arising exclusively from the training noise, by setting $\hat{\mathbf{u}}_{\rm s} = \mathbf{u}_1$. Let $\tilde{\mathbf{v}}_{\rm s}$ be defined as $\tilde{\mathbf{v}}_{\rm s} \triangleq X_{\rm p} Y_{\rm p}^H \mathbf{u}_1/(\sigma_1 \gamma_{\rm p})$. Then, from (5)

$$\tilde{\mathbf{v}}_{\rm s} = \mathbf{v}_1 + \frac{E_{\rm p}^H \mathbf{u}_1}{\sigma_1}$$

where, $E_{\rm p} \triangleq \eta_{\rm p} X_{\rm p}^H/\gamma_{\rm p}$, as before. Recall from (7) that $\hat{\mathbf{v}}_{\rm s} = \tilde{\mathbf{v}}_{\rm s}/\|\tilde{\mathbf{v}}_{\rm s}\|$. Now, $\|\tilde{\mathbf{v}}_{\rm s}\|$ can be simplified as $\|\tilde{\mathbf{v}}_{\rm s}\|^2 \simeq 1 + (\mathbf{u}_1^H E_{\rm p} \mathbf{v}_1 + \mathbf{v}_1^H E_{\rm p}^H \mathbf{u}_1)/\sigma_1$, whence we get

$$\hat{\mathbf{v}}_{\rm s} \simeq \left(\mathbf{v}_1 + \frac{E_{\rm p}^H \mathbf{u}_1}{\sigma_1}\right)\left[1 - \frac{1}{2\sigma_1}\left(\mathbf{u}_1^H E_{\rm p} \mathbf{v}_1 + \mathbf{v}_1^H E_{\rm p}^H \mathbf{u}_1\right)\right].$$

Ignoring terms of order $E_{\rm p}^H E_{\rm p}$ and simplifying, the MSE in $\hat{\mathbf{v}}_{\rm s}$ is

$$\hat{\mathbf{v}}_{\rm s} - \mathbf{v}_1 \simeq \frac{E_{\rm p}^H \mathbf{u}_1}{\sigma_1} - \frac{1}{2\sigma_1}\left(\mathbf{u}_1^H E_{\rm p} \mathbf{v}_1 + \mathbf{v}_1^H E_{\rm p}^H \mathbf{u}_1\right)\mathbf{v}_1$$
$$\|\hat{\mathbf{v}}_{\rm s} - \mathbf{v}_1\|^2 = \frac{\left\|E_{\rm p}^H \mathbf{u}_1\right\|^2}{\sigma_1^2} - \frac{1}{4\sigma_1^2}\left|\mathbf{u}_1^H E_{\rm p} \mathbf{v}_1 + \mathbf{v}_1^H E_{\rm p}^H \mathbf{u}_1\right|^2. \quad (24)$$

Taking expectation and simplifying the above expression using Lemma 2, we get

$$\mathbf{E}\left\{\|\hat{\mathbf{v}}_{\rm s} - \mathbf{v}_1\|^2\right\} = \frac{1}{2\gamma_{\rm p}\sigma_1^2}(2t-1). \quad (25)$$

Interestingly, the above expression is the CRB for the estimation of $\mathbf{v}_1$ assuming perfect knowledge of $\mathbf{u}_1$, which we prove in the following theorem.

Theorem 1: The error given in (25) is the CRB for the estimation of $\mathbf{v}_1$ under perfect knowledge of $\mathbf{u}_1$.

Proof: From (37), the effective SNR for estimation of $\mathbf{v}_1$ is $\gamma_{\rm s} = \gamma_{\rm p}\sigma_1^2$. From the results derived for the CRB with constrained parameters [7], [16], and since $\tilde{X}_{\rm p}\tilde{X}_{\rm p}^H = I_{\rm t}/\gamma_{\rm p}$, the estimation error in $\mathbf{v}_1$ is proportional to the number of parameters, which equals $2t-1$ as $\mathbf{v}_1$ is a t-dimensional complex vector with one constraint ($\|\mathbf{v}_1\| = 1$). The estimation error is given by

$$\mathbf{E}\left\{\|\hat{\mathbf{v}}_{\rm s} - \mathbf{v}_1\|^2\right\} = \frac{1}{2\gamma_{\rm s}}\{\text{Num. Parameters}\} = \frac{1}{2\gamma_{\rm p}\sigma_1^2}(2t-1) \quad (26)$$

which agrees with the ML error derived in (25). ■

B. Received SNR With Perfect $\hat{\mathbf{u}}_{\rm s}$

We start with the expression for the channel gain when using $\hat{\mathbf{u}}_{\rm s}$ and $\hat{\mathbf{v}}_{\rm s}$ as the transmit and receive beamforming vectors. When we have perfect knowledge of $\mathbf{u}_1$ at the receiver, $\hat{\mathbf{u}}_{\rm s} = \mathbf{u}_1$ and $\hat{\mathbf{v}}_{\rm s} = \tilde{\mathbf{v}}_{\rm s}/\|\tilde{\mathbf{v}}_{\rm s}\|$, where $\tilde{\mathbf{v}}_{\rm s} = \mathbf{v}_1 + E_u \mathbf{u}_1$ and $E_u \triangleq E_{\rm p}^H/\sigma_1$. The power amplification with perfect knowledge of $\mathbf{u}_1$, denoted by $\rho_u \triangleq \mathbf{E}\{|\mathbf{u}_1^H H \hat{\mathbf{v}}_{\rm s}|^2\} = \mathbf{E}\{|\mathbf{u}_1^H H \tilde{\mathbf{v}}_{\rm s}|^2/\|\tilde{\mathbf{v}}_{\rm s}\|^2\}$. As shown in the Section B of the Appendix, this can be simplified to

$$\rho_u = \sigma_1^2 - \frac{t-1}{\gamma_{\rm p}}. \quad (27)$$

Finally, the received SNR is given by $P_{\rm D}\rho_u$, as before. Comparing the above expression with the power amplification with CLSE (17), we see that when $r = t$, even in the best case of a spatially single-dimensional channel $\rho_c = \rho_{\rm p} - (2/\gamma_{\rm p})(t-1) < \rho_u$. Next, when $r = 1$, CLSE and CFSB techniques perform exactly the same: $\rho_c = \rho_u = \sigma_1^2 - (t-1)/\gamma_{\rm p}$ since $\mathbf{u}_1 = 1$ (that is, no receive beamforming is needed). Thus, if perfect knowledge of $\mathbf{u}_1$ is available at the receiver, CFSB is guaranteed to perform as well as CLSE, regardless of the training symbol SNR.

C. MSE in $\hat{\mathbf{v}}_{\rm s}$ With Noise-Free Training

We now present analysis to compute the second term in (23), the MSE in $\hat{\mathbf{v}}_{\rm s}$ solely due to the use of the erroneous vector $\hat{\mathbf{u}}_{\rm s}$ in (7), and hence let $\eta_{\rm p} = 0$, or $Y_{\rm p} = HX_{\rm p}$. As in Section III-C, we can express $\hat{\mathbf{u}}_{\rm s}$ as a linear combination of the columns of U as $\hat{\mathbf{u}}_{\rm s} = U\mathbf{c}$. We slightly abuse notation from Section IV-A and redefine $\tilde{\mathbf{v}}_{\rm s}$ as $\tilde{\mathbf{v}}_{\rm s} \triangleq X_{\rm p} Y_{\rm p}^H \hat{\mathbf{u}}_{\rm s}/\gamma_{\rm p} = V\Sigma\mathbf{c}$. Hence

$$\|\tilde{\mathbf{v}}_{\rm s}\|^2 = \mathbf{c}^H \Sigma^2 \mathbf{c}.$$

Thus, from (7), we have, $\hat{\mathbf{v}}_{\rm s} = V\tilde{\mathbf{c}}$, where $\tilde{\mathbf{c}} = \Sigma\mathbf{c}/\sqrt{\mathbf{c}^H\Sigma^2\mathbf{c}}$. From Lemma (3)

$$\|\hat{\mathbf{v}}_{\rm s} - \mathbf{v}_1\|^2 = 2\left(1 - \text{Re}(\tilde{c}_1)\right). \quad (28)$$

Let $\mathbf{c} = [1 + \Delta c_1, \Delta c_2, \ldots, \Delta c_r]^T$. Then, as shown in the Section C of the Appendix, $\tilde{c}_1$, the first element of $\tilde{\mathbf{c}}$, is given by

$$\tilde{c}_1 \simeq 1 - \frac{1}{2}\sum_{i=2}^{r}\frac{\sigma_i^2}{\sigma_1^2}|\Delta c_i|^2 \quad (29)$$

and hence $\|\hat{\mathbf{v}}_s - \mathbf{v}_1\|^2 = \sum_{i=2}^{r}(\sigma_i^2/\sigma_1^2)|\Delta c_i|^2$. Let γ_D be defined as $\gamma_D \triangleq NP_D/t$. Then, from Section C of the Appendix, $\mathbf{E}\{|\Delta c_i|^2\}$ is given by

$$\mathbf{E}\left\{|\Delta c_i|^2\right\} = \frac{1}{(\sigma_1^2 - \sigma_i^2)^2}\left(\frac{\sigma_1^2\sigma_i^2}{N} + \frac{\sigma_i^2 + \sigma_1^2}{\gamma_D} + \frac{N}{\gamma_D^2}\right). \tag{30}$$

Substituting, we get the final expression for the MSE as

$$\mathbf{E}\left\{\|\hat{\mathbf{v}}_s - \mathbf{v}_1\|^2\right\} = \sum_{i=2}^{r}\frac{\sigma_i^2}{\sigma_1^2(\sigma_1^2 - \sigma_i^2)^2} \times \left(\frac{\sigma_1^2\sigma_i^2}{N} + \frac{\sigma_i^2 + \sigma_1^2}{\gamma_D} + \frac{N}{\gamma_D^2}\right). \tag{31}$$

Note that the above expression decreases as $O(1/N)$ (since γ_D depends linearly on N), and therefore the MSE asymptotically (as $N \to \infty$) approaches the bound in (25).

D. Received SNR With Noise-Free Training

The power amplification with noise-free training, denoted ρ_w, is given by $\rho_w = |\hat{\mathbf{u}}_s^H H \hat{\mathbf{v}}_s|^2$. We also have $\hat{\mathbf{u}}_s = U\mathbf{c}$ and $\hat{\mathbf{v}}_s = V\tilde{\mathbf{c}}$, where $\tilde{\mathbf{c}} = \Sigma\mathbf{c}/\sqrt{\mathbf{c}^H\Sigma^2\mathbf{c}}$. Then, $\hat{\mathbf{u}}_s^H H \hat{\mathbf{v}}_s = \mathbf{c}^H\Sigma\tilde{\mathbf{c}} = \sqrt{\mathbf{c}^H\Sigma^2\mathbf{c}}$, and thus

$$\begin{aligned}\rho_w &= \mathbf{c}^H\Sigma^2\mathbf{c} = \sigma_1^2(1 + \Delta c_1^*)(1 + \Delta c_1) + \sum_{i=2}^{r}\sigma_i^2\Delta c_i^*\Delta c_i \\ &\simeq \sigma_1^2(1 + \Delta c_1 + \Delta c_1^*) + \sum_{i=2}^{r}\sigma_i^2|\Delta c_i|^2.\end{aligned}$$

Substituting for Δc_1 from (9) and Δc_i from (30), we obtain the power amplification with noise-free training as

$$\rho_w = \sigma_1^2 - \sum_{i=2}^{r}\frac{1}{(\sigma_1^2 - \sigma_i^2)}\left(\frac{\sigma_1^2\sigma_i^2}{N} + \frac{\sigma_1^2 + \sigma_i^2}{\gamma_D} + \frac{N}{\gamma_D^2}\right). \tag{32}$$

As before, the received SNR is given by $P_D\rho_w$. Note that ρ_w approaches $\rho_p = \sigma_1^2$ for large values of length N and SNR γ_D.

E. Semiblind Estimation: Summary

Recall that $\gamma_p = LP_T/t$ and $\gamma_D = NP_D/t$. The final expressions for the MSE in $\hat{\mathbf{v}}_s$ and the power amplification, from (23), are

$$\mathbf{E}\left\{\|\mathbf{v}_1 - \hat{\mathbf{v}}_s\|^2\right\} = \frac{(2t-1)}{2\gamma_p\sigma_1^2} + \sum_{i=2}^{r}\frac{\sigma_i^2}{\sigma_1^2(\sigma_1^2 - \sigma_i^2)^2} \times \left(\frac{\sigma_1^2\sigma_i^2}{N} + \frac{\sigma_i^2 + \sigma_1^2}{\gamma_D} + \frac{N}{\gamma_D^2}\right) \tag{33}$$

$$\rho_s = \sigma_1^2 - \frac{t-1}{\gamma_p} - \sum_{i=2}^{r}\frac{1}{(\sigma_1^2 - \sigma_i^2)} \times \left(\frac{\sigma_1^2\sigma_i^2}{N} + \frac{\sigma_1^2 + \sigma_i^2}{\gamma_D} + \frac{N}{\gamma_D^2}\right). \tag{34}$$

The SER with semiblind estimation is given by $P_M(\rho_s)$, with $P_M(\cdot)$ defined as in (19).

V. Comparison of CLSE and Semiblind Schemes

In order to compare the CFSB and CLSE techniques, one needs to account for the performance of the white data versus beamformed data, an issue we address now. Generic comparison of the semiblind and conventional schemes for any arbitrary system configuration is difficult, so we consider an example to illustrate the tradeoffs involved. We consider the 2×2 system with the Alamouti scheme [17] employed for white data transmission, and with uncoded 4-QAM symbol transmission. The choice of the Alamouti scheme enables us to present a fair comparison of the two estimation algorithms since it has an effective data rate of one bit per channel use, the same as that of MRT. In addition, it is possible to employ a simple receiver structure, which makes the performance analysis tractable.

Let the beamformed data and the white data be statistically independent, and a zero-forcing receiver based on the conventional estimate of the channel (3) be used to detect the white data symbols. In Section D of the Appendix, we derive the average SNR of this system as

$$\rho_w = \frac{\left(\|H\|_F^4 + \frac{\|H\|_F^2}{\gamma_p}\right)P_x}{\frac{\|H\|_F^2}{\gamma_p}P_x + \|H\|_F^2 + \frac{2r}{\gamma_p}} \tag{35}$$

where $\|\cdot\|_F$ is the Frobenius norm, P_x is the per-symbol transmit power and $\gamma_p = LP_T/t$ as defined before. From (35), we can also obtain the symbol error rate performance of the Alamouti coded white data by using (19) with $\rho_c P_D$ replaced by ρ_w. The resulting expression can be numerically averaged over the pdf of $\|H\|_F^2$, which is Gamma distributed with $2rt$ degrees of freedom, to obtain the SER. The analysis of the beamformed data with the CFSB estimation when the Alamouti scheme is employed to transmit spatially white data remains largely the same as that presented in the previous section, where we had assumed that X_D satisfies $\mathbf{E}\{X_D X_D^H\} = \gamma_D I_t$. With Alamouti white-data transmission, we have that $X_D X_D^H = \gamma_D I_t$, which causes the E_χ term to drop out in (41) of Section C of the Appendix.

A. Performance of a 2 × 2 System With CLSE and CFSBl

In order to get a more concrete feel for the expressions obtained in the preceding, let us consider a 2×2 system with $L = 2$, $N = 8$, $P_D =$ 6 dB and 110 total symbols per frame, i.e., two training symbols, eight white data symbols, and 100 beamformed data symbols in the semiblind case, and two training symbols and 108 beamformed data symbols in the conventional case. The average channel power gain ρ versus training symbol SNR (P_T), obtained under different CSI and signal transmission conditions are shown in Fig. 3. When the receiver has perfect channel knowledge (labeled perfect $\mathbf{u}_1$, $\mathbf{v}_1$), the average power gain ρ is $\mathbf{E}\{\sigma_1^2\} =$ 5.5 dB, independent of the training symbol SNR. The ρ with CLSE as well as the semiblind techniques asymptotically tend to this gain of 5.5 dB as the SNR becomes large, since the loss due to estimation error becomes negligible. The channel power gain with only white (Alamouti) data transmission asymptotically approaches 3 dB (the gain per symbol of the 2×2 system with Alamouti encoding).

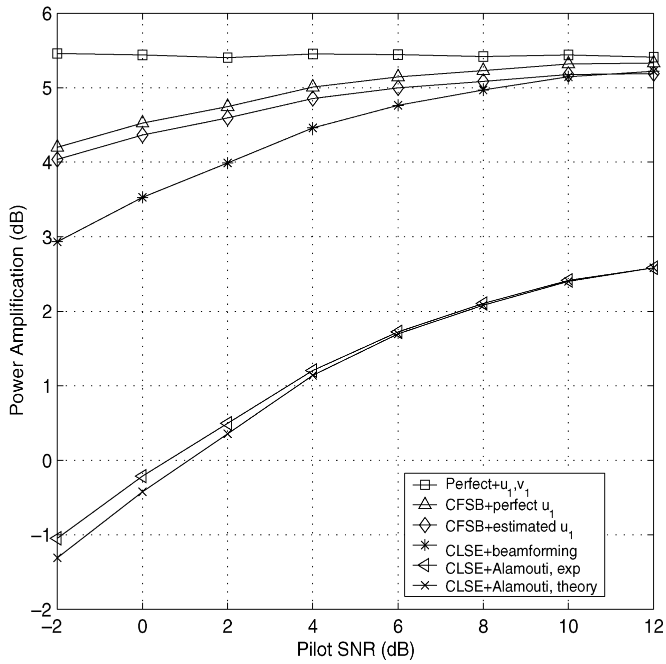

Fig. 3. Average channel gain of a $t = r = 2$ MIMO channel with $L = 2$, $N = 8$, and $P_{\rm D} = 6$ dB, for the CLSE and beamforming, CFSB and beamforming (with and without knowledge of $\mathbf{u}_1$), CLSE and white data (Alamouti-coded), and perfect beamforming at transmitter and receiver. Also plotted is the theoretical result for the performance of Alamouti-coded data with channel estimation error, given by (35).

The channel power gain at any $P_{\rm T}$ is given by (35), which is validated in Fig. 3 through simulation. Observe that at a given training SNR, there is a loss of approximately $P_a = -3$ dB in terms of the channel gain performance for the Alamouti scheme compared to the beamforming with conventional estimation. The results of the channel power gain obtained by employing the CFSB technique with $N = 8$ Alamouti-coded data symbols are shown in Fig. 3, and show the improved performance of CFSB. By transmitting a few Alamouti-coded symbols, the CFSB scheme obtains a better estimate of $\mathbf{v}_1$, thereby gaining about $P_{sb} = 0.8$ dB per symbol over the CLSE scheme, at a training SNR of 2 dB.

If the frame length is 110 symbols, we have $L_{\rm D} = 100$ beamformed data symbols in the semiblind case and $L_{\rm D} + N = 108$ beamformed data symbols in the conventional case. Using the beamforming vectors estimated by the CFSB algorithm, we then have a net power gain ρ_g given by $\rho_g = (L_{\rm D} + N)/(L_{\rm D}/P_{sb} + N/P_a)$, or about 0.4 dB per frame. Thus, this simple example shows that CFSB estimation can potentially offer an overall better performance compared to the CLSE. Although we have considered uncoded modulation here, in more practical situations a channel code will be used with interleaving both between the white and beamformed symbols as well as across multiple frames. In this case, burst errors can be avoided and the errors in the white data symbols corrected. Furthermore, the performance of the white data symbols can also be improved by employing an MMSE receiver or other more advanced multiuser detectors rather than the zero-forcing receiver, leading to additional improvements in the CFSB technique.

B. Discussion

We are now in a position to discuss the merits of the conventional estimation and the semiblind estimation. Clearly, the CLSE enjoys the advantages of being simple and easy to implement. As with any semiblind technique, CFSB being a second-order method requires the channel to be relatively slowly time-varying. If not, the CLSE can still estimate the channel quickly from a few training symbols, whereas the CFSB may not be able to converge to an accurate estimate of $\mathbf{u}_1$ from the second-order statistics computed using just a few received vectors. Another disadvantage of the CFSB is that it requires the implementation of two separate receivers, one for detecting the white data and the other for the beamformed data. However, the CFSB estimation could outperform the CLSE in channels where the loss due to the transmission of spatially white data is not too great, i.e., in full column-rank channels. Given the parameters N, L, $P_{\rm T}$ and $P_{\rm D}$, the theory developed in this paper can be used to decide if the CFSB technique would offer any performance benefits versus the CLSE technique. If the CFSB technique is to perform comparably or better than the CLSE, the following two things need to be satisfied.

1) The *estimation* performance of CLSE and CFSB should be comparable, i.e., the number of white data symbols N and the data power $P_{\rm D}$ should be large enough to ensure that the estimate $\hat{\mathbf{u}}_{\rm s}$ is accurate, so that the resulting $\hat{\mathbf{v}}_{\rm s}$ can perform comparably to the conventional estimate. For example, since the channel gain with semiblind estimation is given by (33), N should be chosen to be of the same order as $\gamma_{\rm D}$; and both N and $\gamma_{\rm D}$ should be of the order higher than $\gamma_{\rm p}$. With such a choice, the $(t-1)/\gamma_{\rm p}$ term will dominate the SNR loss in the CFSB, thus enabling the beamformed data with CFSB estimation to outperform the beamformed data with CLSE.
2) The block length should be sufficiently long to ensure that after sending $L + N$ symbols, there is sufficient room to send as many beamformed symbols as is necessary for the CFSB technique to be able to make up for the performance lost during the white data transmission. In the above example, after having obtained the appropriate value of N, one can use (35) to determine the loss due to the white data symbols (for the $t = 2$ case), and then finally determine whether the block length is long enough for the CFSB to be able to outperform the CLSE method.

In Section VI, we demonstrate through additional simulations that the CFSB technique does offer performance benefits relative to the CLSE, for an appropriately designed system.

C. Semiblind Estimation: Limitations and Alternative Solutions

The CFSB algorithm requires a sufficiently large number of spatially-white data (N) to guarantee a near perfect estimate of $\mathbf{u}_1$ and this error cannot be overcome by increasing the white-data SNR. It is therefore desirable to find an estimation scheme that performs at least as well as the CLSE algorithm, regardless of the value of N and L. Formal fusion of the estimates obtained from the CLSE and CFSB techniques is difficult, hence we adopt an intuitive approach and consider a simple weighted linear combination of the estimated beamforming vectors as follows:

$$\hat{\mathbf{u}}_1 = \frac{\beta_{\mathbf{u}}\gamma_{\mathrm{P}}\hat{\mathbf{u}}_c + \gamma_{\mathrm{D}}\hat{\mathbf{u}}_{\mathrm{s}}}{\|\beta_{\mathbf{u}}\gamma_{\mathrm{P}}\hat{\mathbf{u}}_c + \gamma_{\mathrm{D}}\hat{\mathbf{u}}_{\mathrm{s}}\|_2}, \quad \hat{\mathbf{v}}_1 = \frac{\beta_{\mathbf{v}}\gamma_{\mathrm{P}}\hat{\mathbf{v}}_c + \gamma_{\mathrm{D}}\hat{\mathbf{v}}_{\mathrm{s}}}{\|\beta_{\mathbf{v}}\gamma_{\mathrm{P}}\hat{\mathbf{v}}_c + \gamma_{\mathrm{D}}\hat{\mathbf{v}}_{\mathrm{s}}\|_2}. \tag{36}$$

The above estimates will be referred to as the *linear combination semiblind* (LCSB) estimates. The weights $\gamma_{\mathrm{P}} = LP_{\mathrm{T}}/t$ and $\gamma_{\mathrm{D}} = NP_{\mathrm{D}}/t$ are a measure of the accuracy of the vectors estimated from the CLSE and CFSB schemes respectively. The scaling factor of $\beta_{\mathbf{u}}$ and $\beta_{\mathbf{v}}$ is introduced because $\hat{\mathbf{u}}_c$ obtained from known training symbols is more reliable than the blind estimate $\hat{\mathbf{u}}_{\mathrm{s}}$ when $L = N$ and $\gamma_{\mathrm{P}} = \gamma_{\mathrm{D}}$. In our simulations, for $t = r = 4$, the choice $\beta_{\mathbf{u}} = \beta_{\mathbf{v}} = 4$ was found to perform well. Analysis of the impact of $\beta_{\mathbf{u}}$ and $\beta_{\mathbf{v}}$ is a topic for future research.

VI. Simulation Results

In this section, we present simulation results to illustrate the performance of the different estimation schemes. The simulation setup consists of a Rayleigh flat fading channel with 4 transmit antennas and 4 receive antennas ($t = r = 4$). The data (and training) are drawn from a 16-QAM constellation. 10 000 random instantiations of the channel were used in the averaging.

Measuring the Error Between Singular Vectors: In the simulations, $\mathbf{v}_1$ and $\hat{\mathbf{v}}_1$ are obtained by computing the SVD of two different matrices H and $\hat{H}$ respectively. However, the SVD involves an unknown phase factor, that is, if $\mathbf{v}_1$ is a singular vector, so is $\mathbf{v}_1 e^{j\phi}$ for any $\phi \in (-\pi, \pi]$. Hence, for computational consistency in measuring the MSE in $\mathbf{v}_1$, we use the following dephased norm in our simulations, similar to [18]: $\|\mathbf{v}_1 - \hat{\mathbf{v}}_1\|_{DN}^2 \triangleq 2(1 - |\mathbf{v}_1^H\hat{\mathbf{v}}_1|)$, which satisfies $\|\mathbf{v}_1 - \hat{\mathbf{v}}_1\|_{DN}^2 = \min_{\phi\in(-\pi,\pi]} \|\mathbf{v}_1 - \hat{\mathbf{v}}_1 e^{j\phi}\|^2$. The norm considered in our analysis is implicitly consistent with the above dephased norm. For example, the norm in (14) is the same as the dephased norm, since the perturbation term Δd_1 is real (as noted in Section III-A). Also, for small additive perturbations, it can easily be shown that (for example) in (24), the dephased norm reduces to the Euclidean norm.

Experiment 1: In this experiment, we compute the MSE of conventional estimation and the MSE of the semiblind estimation with perfect $\mathbf{u}_1$, which serves as a benchmark for the performance of the proposed semiblind scheme. Fig. 4 shows the MSE in $\hat{\mathbf{v}}_1$ versus L, for two different values of pilot SNR (or γ_{P}), with perfect $\mathbf{u}_1$. CFSB performs better than the CLSE technique by about 6 dB, in terms of the training symbol SNR for achieving the same MSE in $\hat{\mathbf{v}}_1$. The experimental curves agree well with the theoretical curves from (15), (25). Also, the results for the performance of the semiblind OPML technique proposed in [8] are plotted in Fig. 4. In the OPML technique, the channel matrix H is factored into the product of a whitening matrix W ($= U\Sigma$) and a unitary rotation matrix Q. A blind algorithm is used to estimate W, while the training data is used exclusively to estimate Q. Thus, the OPML technique outperforms the CFSB because it assumes perfect knowledge of the entire U and Σ matrices (and is computationally more expensive). The CFSB technique, on the other hand, only needs an accurate estimate of $\mathbf{u}_1$ from the spatially-white data.

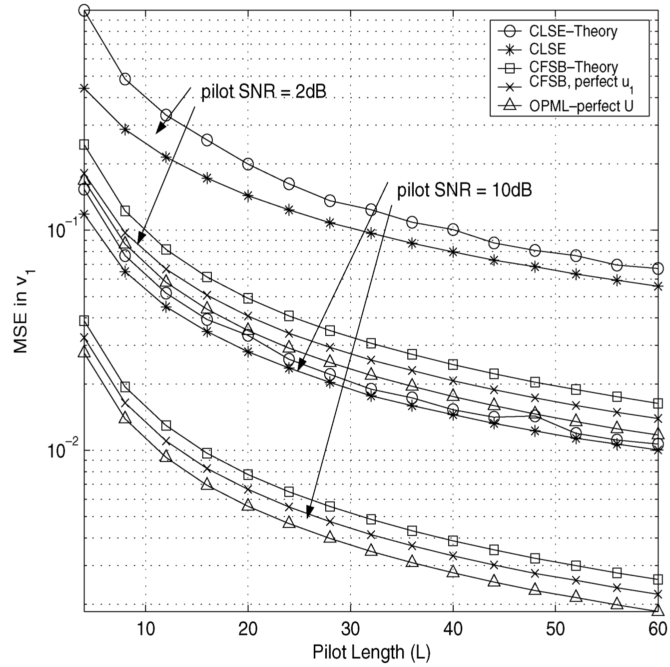

Fig. 4. MSE in $\mathbf{v}_1$ versus training data length L, for a $t = r = 4$ MIMO system. Curves for CLSE, CFSB and OPML with perfect $\mathbf{u}_1$ are plotted. The top five curves correspond to a training symbol SNR of 2 dB, and the bottom five curves 10 dB.

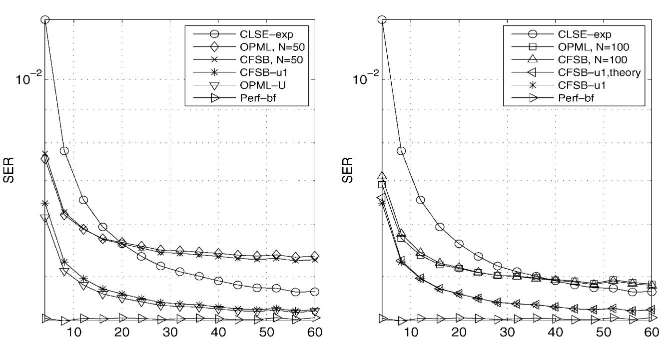

Fig. 5. SER of beamformed data versus number of training symbols L, $t = r = 4$ system, for two different values of white-data length N, and data and training symbol SNR fixed at $P_{\mathrm{T}} = P_{\mathrm{D}} = 6$ dB. The two competing semiblind techniques, OPML and CFSB, are plotted. CFSB marginally outperforms OPML for $N = 50$, as it only requires an accurate estimate of $\mathbf{u}_1$ from the blind data.

Experiment 2: Next, we relax the perfect $\mathbf{u}_1$ assumption. Fig. 5 shows the SER performance of the CLSE, OPML and the CFSB schemes at two different values of N, as well as the $N = \infty$ (perfect knowledge of U) case. At $N = 50$ white data symbols, the CLSE technique outperforms the CFSB for $L \geq 24$, as the error in $\mathbf{u}_1$ dominates the error in the semiblind technique. As white data length increases, the CFSB performs progressively better than the CLSE. Also, in the presence of a finite number (N) of white data, the CFSB marginally outperforms the OPML scheme as CFSB only requires an accurate estimate of the dominant eigenvector $\mathbf{u}_1$ from the white data. In Fig. 6, we plot both the theoretical and experimental curves for the CFSB scheme when $N = 100$, as well as the simulation result for the LCSB scheme defined in Section V-C. The LCSB outperforms the CLSE and the CFSB technique at both

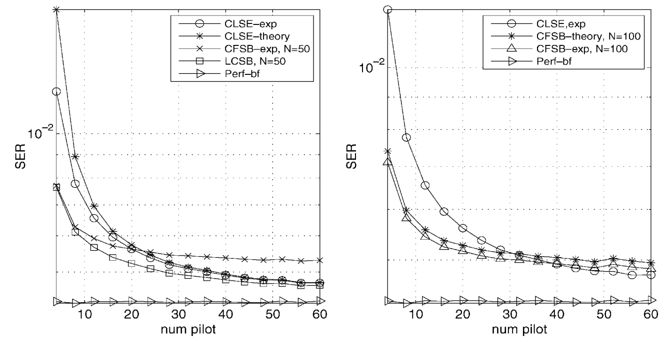

Fig. 6. SER versus $L, t = r = 4$ system, for two different values of N, and data and training symbol SNR fixed at $P_{\mathrm{T}} = P_{\mathrm{D}} =$ 6 dB. The theoretical and experimental curves are plotted for the CFSB estimation technique. Also, the LCSB technique outperforms both the conventional (CLSE) and semiblind (CFSB) techniques.

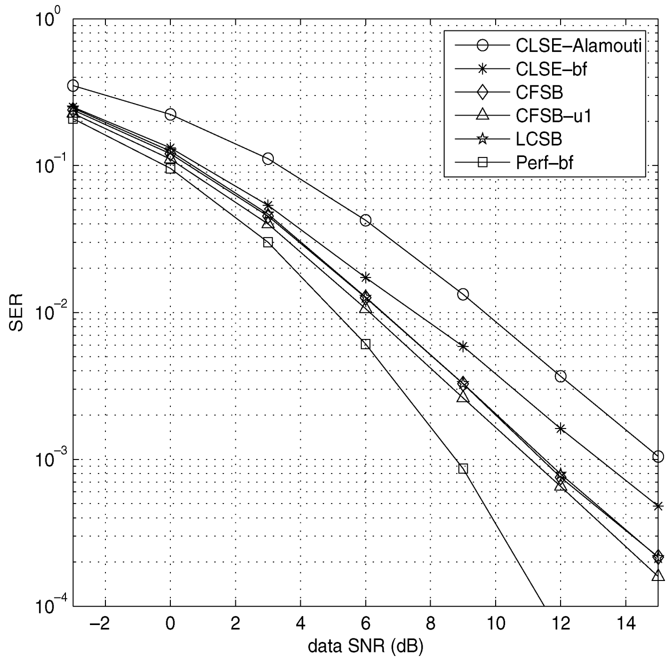

Fig. 7. SER versus data SNR for the $t = r = 2$ system, with $L = 2, N = 16$, $\gamma_{\mathrm{P}} =$ 2 dB. "CLSE-Alamouti" refers to the performance of the spatially-white data with conventional estimation, "CLSE-bf" is the performance of the beamformed data with $\hat{\mathbf{v}}_c$, "CFSB" and "LCSB" refer to the performance of the corresponding techniques after accounting for the loss due to the white data. "CFSB-u1" is the performance of CFSB with perfect-$\mathbf{u}_1$, and "Perf-bf" is the performance with the perfect $\mathbf{u}_1$ and $\mathbf{v}_1$ assumption.

$N = 50$ and $N = 100$. Thus, the theory developed in this paper can be used to compare the performance of CFSB and CLSE techniques for any choice of N and L.

Experiment 3: Finally, as an example of overall performance comparison, Fig. 7 shows the SER performance versus the data SNR of the different estimation schemes for a 2×2 system, with uncoded 4-QAM transmission, $L = 2$ training symbols, $N = 16$ white data symbols (for the semiblind technique) and a frame size $L_{\mathrm{D}} = 500$ symbols. The parameter values are chosen for illustrative purposes, and as L and P_{T} increase, the gap between the CLSE and CFSB reduces. From the graph, it is clear that the LCSB scheme outperforms the CLSE scheme in terms of its SER performance, including the effect of white data transmission.

VII. Conclusion

In this paper, we have investigated training-only and semiblind channel estimation for MIMO flat-fading channels with MRT, in terms of the MSE in the beamforming vector $\mathbf{v}_1$, received SNR and the SER with uncoded M-ary QAM modulation. The CFSB scheme is proposed as a closed-form semiblind solution for estimating the optimum transmit beamforming vector $\mathbf{v}_1$, and is shown to achieve the CRB with the perfect $\mathbf{u}_1$ assumption. Analytical expressions for the MSE, the channel power gain and the SER performance of both the CLSE and the CFSB estimation schemes are developed, which can be used to compare their performance. A novel LCSB algorithm is proposed, which is shown to outperform both the CFSB and the CLSE schemes over a wide range of training lengths and SNRs. We have also presented Monte Carlo simulation results to illustrate the relative performance of the different techniques.

Appendix

A. Proof of Lemma 1

Let $\tilde{Y}_{\mathrm{p}} \triangleq \mathbf{u}_1^H Y_{\mathrm{p}}/(\sigma_1\gamma_{\mathrm{p}})$, $\tilde{X}_{\mathrm{p}} \triangleq X_{\mathrm{p}}/\gamma_{\mathrm{p}}$, and $\tilde{n} \triangleq \mathbf{u}_1^H \eta_{\mathrm{p}}/(\sigma_1\gamma_{\mathrm{p}})$. Then, since the training sequence is orthogonal, $X_{\mathrm{p}}X_{\mathrm{p}}^H = I_{\mathrm{t}}$ holds. Substituting into (5), we have

$$\tilde{Y}_{\mathrm{p}} = \mathbf{v}_1^H \tilde{X}_{\mathrm{p}} + \tilde{n}. \tag{37}$$

Thus, we seek the estimate of $\mathbf{v}_1$ as the solution to the following least-squares problem

$$\hat{\mathbf{v}}_{\mathrm{s}} = \arg \min_{\mathbf{v}\in\mathbb{C}^t, \|\mathbf{v}\|=1} \|\tilde{Y}_{\mathrm{p}} - \mathbf{v}^H \tilde{X}_{\mathrm{p}}\|^2. \tag{38}$$

Note that

$$\begin{aligned}\arg &\min_{\mathbf{v}_1:\|\mathbf{v}_1\|=1} \left\|\tilde{Y}_{\mathrm{p}} - \mathbf{v}_1^H \tilde{X}_{\mathrm{p}}\right\|^2 \\ &= \arg \min_{\mathbf{v}_1:\|\mathbf{v}_1\|=1} \left(\tilde{Y}_{\mathrm{p}}\tilde{Y}_{\mathrm{p}}^H + \frac{\|\mathbf{v}_1\|^2}{\gamma_{\mathrm{p}}} - \tilde{Y}_{\mathrm{p}}\tilde{X}_{\mathrm{p}}^H\mathbf{v}_1 - \mathbf{v}_1^H\tilde{X}_{\mathrm{p}}\tilde{Y}_{\mathrm{p}}^H\right) \\ &= \arg \max_{\mathbf{v}_1:\|\mathbf{v}_1\|=1} \left(\tilde{Y}_{\mathrm{p}}\tilde{X}_{\mathrm{p}}^H\mathbf{v}_1 + \mathbf{v}_1^H\tilde{X}_{\mathrm{p}}\tilde{Y}_{\mathrm{p}}^H\right).\end{aligned}$$

The $\mathbf{v}_1$ that maximizes the above expression is readily found to be $\hat{\mathbf{v}}_1 = \tilde{X}_{\mathrm{p}}\tilde{Y}_{\mathrm{p}}^H/\|\tilde{X}_{\mathrm{p}}\tilde{Y}_{\mathrm{p}}^H\|$. Substituting for $\tilde{X}_{\mathrm{p}}$ and $\tilde{Y}_{\mathrm{p}}$, the desired result is obtained.

B. Received SNR With Perfect $\hat{\mathbf{u}}_{\mathrm{s}}$

Here, we derive the expression in (27). For notational simplicity, define $x \triangleq \mathbf{v}_1^H E_u \mathbf{u}_1$ and $y \triangleq \mathbf{u}_1^H E_u^H E_u \mathbf{u}_1$. Then, we have

$$\begin{aligned}\rho_u &= \mathbf{E}\left\{\frac{\sigma_1^2(1+x)(1+x^*)}{1+x+x^*+y}\right\} \\ &\simeq \sigma_1^2 \mathbf{E}\left\{(1+x)(1+x^*)\right. \\ &\qquad \left.\times\left(1-(x+x^*+y)+(x+x^*+y)^2\right)\right\} \\ &\simeq \sigma_1^2\left(1+\mathbf{E}\{xx^*-y\}\right)\end{aligned} \tag{39}$$

where x^* is the complex conjugate of x. Also, $\mathbf{E}\{xx^*\} = \sigma_{\mathrm{p}}^2/\sigma_1^2 = 1/(\gamma_{\mathrm{p}}\sigma_1^2)$, and $\mathbf{E}\{y\} = \mathbf{E}\{\mathbf{u}_1^H E_u^H E_u \mathbf{u}_1\} =$

$t/(\gamma_{\rm p}\sigma_1^2)$. Thus, the power amplification for perfect $\mathbf{u}_1$ is given by $\rho_u = \sigma_1^2 - (t-1)/\gamma_{\rm p}$.

C. Proof for Equations (29) and (30)

In order to derive an expression for $\tilde{c}_1$, we write $\mathbf{c} = [1 + \Delta c_1, \Delta c_2, \ldots, \Delta c_{\rm t}]^T$ as a perturbation of $[1, 0, \ldots, 0]^T$. Since $\tilde{\mathbf{c}} = \Sigma \mathbf{c}/\sqrt{\mathbf{c}^H \Sigma^2 \mathbf{c}}$, equating components, we have

$$\begin{aligned}\tilde{c}_1 &= \frac{\sigma_1(1+\Delta c_1)}{\sqrt{\sigma_1^2|1+\Delta c_1|^2 + \sum_{i=2}^r \sigma_i^2 |\Delta c_i|^2}} \\ &\simeq (1+\Delta c_1)\left[1 - \frac{1}{2}\left(2\Delta c_1 + \sum_{i=2}^r \frac{\sigma_i^2}{\sigma_1^2}|\Delta c_i|^2\right)\right] \\ &\simeq 1 - \frac{1}{2}\sum_{i=2}^r \frac{\sigma_i^2}{\sigma_1^2}|\Delta c_i|^2.\end{aligned}$$

Substituting in (28), we get

$$\|\mathbf{v}_1 - \tilde{\mathbf{v}}_{\rm s}\|^2 = \sum_{i=2}^r \frac{\sigma_i^2}{\sigma_1^2}|\Delta c_i|^2. \tag{40}$$

It now remains to compute Δc_i. Recall that $\hat{\mathbf{u}}_{\rm s}$ is computed from the SVD in (4). Stacking the transmitted and received data vectors into matrices $X_{\rm D} \in \mathbb{C}^{t\times N}$ and $Y_{\rm D} \in \mathbb{C}^{r\times N}$ and the noise vectors into $\eta_{\rm D} \in \mathbb{C}^{r\times N}$, with appropriate scaling we can rewrite (4) as

$$\hat{U}\hat{\Sigma}^2\hat{U}^H = HH^H + E_{\rm s}$$

where

$$E_{\rm s} \triangleq HE_\chi H^H + HE_{\chi\eta} + E_{\chi\eta}^H H^H + E_\eta$$

and $E_\chi \triangleq (1/\gamma_{\rm D})(X_{\rm D}X_{\rm D}^H - \gamma_{\rm D} I_{\rm t})$, $E_{\chi\eta} \triangleq X_{\rm D}\eta_{\rm D}^H/\gamma_{\rm D}$, $E_\eta \triangleq (1/\gamma_{\rm D})(\eta_{\rm D}\eta_{\rm D}^H - NI_r)$, and finally $\gamma_{\rm D} = (NP_{\rm D}/t)$, as before.

Observe that, since the white data $X_{\rm D}$ and AWGN are mutually independent, the elements of E_χ, $E_{\chi\eta}$ and E_η are pairwise uncorrelated. Also, $\mathbf{E}\{|E_\chi(i,j)|^2\} = (P_{\rm D}/t)^2/(N(P_{\rm D}/t)^2) = 1/N$, $\mathbf{E}\{|E_{\chi\eta}(i,j)|^2\} = (P_{\rm D}/t)/(N(P_{\rm D}/t)^2) = 1/\gamma_{\rm D}$, and $\mathbf{E}\{|E_\eta(i,j)|^2\} = 1/(N(P_{\rm D}/t)^2) = N/\gamma_{\rm D}^2$. Thus, from the first-order perturbation analysis (8), $\Delta c_i = \mathbf{u}_i^H E_{\rm s}\mathbf{u}_1/(\sigma_1^2 - \sigma_i^2)$, and therefore

$$\begin{aligned}&\mathbf{E}\left\{|\Delta c_i|^2\right\} \\ &= \frac{1}{(\sigma_1^2-\sigma_i^2)^2}\left(\mathbf{E}\left\{\left|\mathbf{u}_i^H HE_\chi H^H\mathbf{u}_1\right|^2\right\}\right. \\ &\quad + \mathbf{E}\left\{\left|\mathbf{u}_i^H HE_{\chi\eta}\mathbf{u}_1\right|^2\right\} + \mathbf{E}\left\{\left|\mathbf{u}_i^H E_{\chi\eta}^H H^H\mathbf{u}_1\right|^2\right\} \\ &\quad \left. + \mathbf{E}\left\{\left|\mathbf{u}_i^H E_\eta \mathbf{u}_1\right|^2\right\}\right).\end{aligned} \tag{41}$$

Simplifying the different components in the above expression, we have $\mathbf{E}\{|\mathbf{u}_i^H HE_\chi H^H\mathbf{u}_1|^2\} = \sigma_1^2\sigma_i^2/N$, $\mathbf{E}\{|\mathbf{u}_i^H E_\eta\mathbf{u}_1|^2\} = N/\gamma_{\rm D}^2$ and $\mathbf{E}\{|\mathbf{u}_i^H HE_{\chi\eta}\mathbf{u}_1|^2\} = \sigma_i^2/\gamma_{\rm D}$. Substituting into (41), we get (30).

D. Performance of Alamouti Space–Time-Coded Data With Conventional Estimation

In this section, we determine the performance of Alamouti space–time-coded data for a general $r \times 2$ matrix channel with estimation error and a zero-forcing receiver. Similar results for other specific cases can be found in [19], [20]. Denote the $r\times 2$ channel matrix H in terms of its columns as $H = [\mathbf{h}_1, \mathbf{h}_2]$. Also, let the $2 \times L$ orthogonal training symbol matrix $X_{\rm p}$ be defined in terms of its rows as $X_{\rm p}^T = [X_{p1}^T, X_{p2}^T]^T$. Thus, from (3), the channel is estimated conventionally as

$$\begin{aligned}\hat{H}_c &= \frac{1}{\gamma_{\rm p}}\left[Y_{\rm p}X_{p1}^H, Y_{\rm p}X_{p2}^H\right] \\ [\hat{\mathbf{h}}_1, \hat{\mathbf{h}}_2] &= \left[\mathbf{h}_1 + \frac{\eta_{\rm p}X_{p1}^H}{\gamma_{\rm p}}, \mathbf{h}_2 + \frac{\eta_{\rm p}X_{p2}^H}{\gamma_{\rm p}}\right]\end{aligned} \tag{42}$$

The effective channel with Alamouti-coded data transmission can be represented by stacking two consecutively received $r\times 1$ vectors $\mathbf{y}_1$ and $\mathbf{y}_2^*$ vertically as follows:

$$\begin{bmatrix}\mathbf{y}_1\\ \mathbf{y}_2^*\end{bmatrix} = \begin{bmatrix}\mathbf{h}_1 & \mathbf{h}_2\\ -\mathbf{h}_2^* & \mathbf{h}_1^*\end{bmatrix}\begin{bmatrix}x_1\\ x_2^*\end{bmatrix} + \begin{bmatrix}\mathbf{n}_{w1}\\ \mathbf{n}_{w2}^*\end{bmatrix} \tag{43}$$

where $\mathbf{n}_{wi}$, $i = 1, 2$ is the AWGN affecting the white data symbols. When a zero-forcing receiver based on the estimated channel is employed, the received vectors are decoded using $[\hat{\mathbf{h}}_1, \hat{\mathbf{h}}_2]$ as

$$\begin{bmatrix}\hat{x}_1\\ \hat{x}_2^*\end{bmatrix} = \begin{bmatrix}\hat{\mathbf{h}}_1^H & -\hat{\mathbf{h}}_2^T\\ \hat{\mathbf{h}}_2^H & \hat{\mathbf{h}}_1^T\end{bmatrix}\begin{bmatrix}\mathbf{y}_1\\ \mathbf{y}_2^*\end{bmatrix}. \tag{44}$$

It is clear from symmetry that the performance of $\hat{x}_1$ and $\hat{x}_2$ will be the same; hence, we can focus on determining the SER performance of $\hat{x}_1$. Now, $\hat{x}_1$ contains three components, the signal component coming from x_1, and a leakage term coming from the symbol x_2 and the noise term coming from the white noise term $\mathbf{n}_w$ as follows:

$$\begin{aligned}\hat{x}_1 = \underbrace{\left(\hat{\mathbf{h}}_1^H\mathbf{h}_1 + \mathbf{h}_2^H\hat{\mathbf{h}}_2\right)}_{\xi_{x1}} x_1 + \underbrace{\left(\hat{\mathbf{h}}_1^H\mathbf{h}_2 - \mathbf{h}_1^H\hat{\mathbf{h}}_2\right)}_{\xi_{x2}} x_2^* \\ + \underbrace{\hat{\mathbf{h}}_1^H\mathbf{n}_{w1} - \mathbf{n}_{w2}^H\hat{\mathbf{h}}_2}_{\xi_n}.\end{aligned} \tag{45}$$

The coefficient of the x_1 term, denoted ξ_{x1} is

$$\xi_{x1} = \left(\mathbf{h}_1 + \frac{\eta_{\rm p}X_{p1}^H}{\gamma_{\rm p}}\right)^H \mathbf{h}_1 + \mathbf{h}_2^H\left(\mathbf{h}_2 + \frac{\eta_{\rm p}X_{p2}^H}{\gamma_{\rm p}}\right) \tag{46}$$

$$= \|H\|_F^2 + \frac{X_{p1}\eta_{\rm p}^H\mathbf{h}_1 + \mathbf{h}_2^H\eta_{\rm p}X_{p2}^H}{\gamma_{\rm p}}. \tag{47}$$

From the above equation, it is clear that the performance of the x_1 symbol is dependent on the training noise instantiation $\eta_{\rm p}$.

However, we can consider the average power gain, averaged over the training noise, as follows:

$$\begin{aligned}\mathbf{E}\left\{|\xi_{x1}|^2\right\} &= \|H\|_F^4 + \frac{1}{\gamma_\mathrm{p}^2}\mathbf{E}\left\{X_{p1}\eta_\mathrm{p}^H\mathbf{h}_1\mathbf{h}_1^H\eta_\mathrm{p}X_{p1}^H\right. \\ &\qquad \left. +\mathbf{h}_2^H\eta_\mathrm{p}X_{p2}^H X_{p2}\eta_\mathrm{p}^H\mathbf{h}_2\right\}, \quad (48)\\ &= \|H\|_F^4 + \frac{1}{\gamma_\mathrm{p}^2}\left(\gamma_\mathrm{p}\|\mathbf{h}_1\|^2 + \gamma_\mathrm{p}\|\mathbf{h}_2\|^2\right), \quad (49)\\ &= \|H\|_F^4 + \frac{\|H\|_F^2}{\gamma_\mathrm{p}} \quad (50)\end{aligned}$$

where, in (48), the cross terms disappear since the noise η_p is zero-mean and due to the orthogonality of the training X_p. Similarly, the coefficient of the x_2^* term, denoted ξ_{x2}, can be simplified as

$$\xi_{x2} = \frac{X_{p1}\eta_\mathrm{p}^H\mathbf{h}_2 - \mathbf{h}_1^H\eta_\mathrm{p}X_{p2}^H}{\gamma_\mathrm{p}}. \quad (51)$$

We will assume for simplicity that the x_2 term is an additive white Gaussian noise impairing the estimation of x_1, i.e., we do not perform joint detection. This noise term is independent of the AWGN component $\mathbf{n}_w$. Similar to the coefficient of x_1, we can consider the average power gain of the x_2 term, which can be obtained after a little manipulation as

$$\mathbf{E}\left\{|\xi_{x2}|^2\right\} = \frac{\|H\|_F^2}{\gamma_\mathrm{p}}. \quad (52)$$

Finally, the noise term, denoted ξ_n, is

$$\xi_n = \mathbf{h}_1^H\mathbf{n}_{w1} - \mathbf{n}_{w2}^H\mathbf{h}_2 + \frac{X_{p1}\eta_\mathrm{p}^H\mathbf{n}_{w1} - \mathbf{n}_{w2}^H\eta_\mathrm{p}X_{p2}^H}{\gamma_\mathrm{p}} \quad (53)$$

from which we can obtain the noise power as

$$\mathbf{E}\left\{|\xi_n|^2\right\} = \|H\|_F^2 + \frac{2r}{\gamma_\mathrm{p}}. \quad (54)$$

Thus, the SNR for detection of a white data symbol is given by

$$\rho_w = \frac{\left(\|H\|_F^4 + \frac{\|H\|_F^2}{\gamma_\mathrm{p}}\right)P_x}{\frac{\|H\|_F^2}{\gamma_\mathrm{p}}P_x + \|H\|_F^2 + \frac{2r}{\gamma_\mathrm{p}}}. \quad (55)$$

E. Other Useful Lemmas

In this section, we present three useful lemmas without proof for the sake of brevity.

Lemma 4: Let $X_\mathrm{p} \in \mathbb{C}^{t\times L}$ be an orthogonal set of vectors (i.e., $X_\mathrm{p}X_\mathrm{p}^H = \gamma_\mathrm{p}I_\mathrm{t}$), and let $\eta_\mathrm{p} \in \mathbb{C}^{r\times L}$ contain i.i.d. ZMCSCG entries with mean $\mu = 0$ and variance $\sigma_n^2 = 1$. Then, the elements of $E_\mathrm{p} = X_\mathrm{p}\eta_\mathrm{p}^H$ are uncorrelated, and the variance of each element of E_p is $\sigma_\mathrm{p}^2 = \gamma_\mathrm{p}$.

Lemma 5: A transformation of E_p (defined in Lemma 4) by any orthogonal matrix $V \in \mathbb{C}^{t\times t}$ (i.e., $VV^H = V^HV = I_\mathrm{t}$) to get $\hat{E} = VE_\mathrm{p}$, leaves the second-order statistics of E_p unaltered, that is

$$\begin{aligned}\mathbf{E}\left\{E(i,j)\right\} &= \mathbf{E}\left\{\hat{E}(i,j)\right\} = 0\\ \mathbf{E}\left\{E(i,j)E^*(k,l)\right\} &= \mathbf{E}\left\{\hat{E}(i,j)\hat{E}^*(k,l)\right\}\\ &= \sigma_\mathrm{p}^2\delta(i-k, j-l), \quad \forall i,j,k,l\end{aligned}$$

where $\delta(p,q) = 1$ when $p = q = 0$, and 0 otherwise.

Lemma 6: If the random vector $X_\mathrm{p} \in \mathbb{C}^{t\times L}$ has zero-mean circularly symmetric i.i.d. entries, then so does $\mathbf{v}^H X_\mathrm{p}$, where $\mathbf{v} \in \mathbb{C}^{t\times 1}$. Further, if $\mathbf{v}$ satisfies $\|\mathbf{v}\| = 1$, then the variance of an element of X_p is the same as that of $\mathbf{v}^H X_\mathrm{p}$.

References

[1] A. K. Jagannatham, C. R. Murthy, and B. D. Rao, "A semi-blind MIMO channel estimation scheme for MRT," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, vol. 3, Philadelphia, PA, Mar. 2005, pp. 585–588.

[2] D. Pal, "Fractionally spaced semi-blind equalization of wireless channels," in *Proc. 26th Asilomar Conf.*, vol. 2, 1992, pp. 642–645.

[3] ——, "Fractionally spaced equalization of multipath channels: a semi-blind approach," in *Proc. 1993 Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, vol. 3, 1993, pp. 9–12.

[4] E. de Carvalho and D. T. M. Slock, "Cramer–Rao bounds for semi-blind and training sequence based channel estimation," in *Proc. 1st IEEE Workshop Signal Processing Advances in Wireless Communications (SPAWC)*, 1997, pp. 129–32.

[5] ——, "Asymptotic performance of ML methods for semi-blind channel estimation," in *Proc. 31st Asilomar Conf.*, vol. 2, 1998, pp. 1624–1628.

[6] A. Medles, D. T. M. Slock, and E. de Carvalho, "Linear prediction based semi-blind estimation of MIMO FIR channels," in *Proc. 3rd IEEE Workshop Signal Processing Advances in Wireless Communications (SPAWC)*, Taiwan, R.O.C., Mar. 2001, pp. 58–61.

[7] A. K. Jagannatham and B. D. Rao, "A semi-blind technique for MIMO channel matrix estimation," in *Proc. IEEE Workshop Signal Processing Advances in Wireless Communications (SPAWC)*, Rome, Italy, Jun. 2003, pp. 304–308.

[8] ——, "Whitening rotation based semi-blind MIMO channel estimation," *IEEE Trans. Signal Process.*, vol. 54, no. 3, pp. 861–869, Mar. 2006.

[9] T. Marzetta, "BLAST training: Estimating channel characteristics for high-capacity space–time wireless," in *Proc. 37th Annu. Allerton Conf. Communications, Control, Computing*, Monticello, IL, Sep. 1999, pp. 22–24.

[10] T. K. Y. Lo, "Maximum ratio transmission," *IEEE Trans. Commun.*, vol. 47, no. 10, pp. 1458–1461, Oct. 1999.

[11] S. M. Kay, *Fundamentals of Statistical Signal Processing, Vol. I: Estimation Theory*, 1st ed. Englewood Cliffs, NJ: Prentice-Hall PTR, 1993.

[12] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*. New York: Wiley, 1971, ch. 11.

[13] J. H. Wilkinson, *The Algebraic Eigenvalue Problem*, 1st ed. Oxford, U.K.: Oxford Univ. Press, 1965.

[14] M. Kaveh and A. J. Barabell, "The statistical performance of the MUSIC and the minimum-norm algorithms in resolving plane waves in noise," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 34, no. 2, pp. 331–341, Feb. 1986.

[15] J. G. Proakis, *Digital Communications*. New York: McGraw-Hill Higher Education, 2001.

[16] A. K. Jagannatham and B. D. Rao, "Complex constrained CRB and its application to semi-blind MIMO and OFDM channel estimation," presented at the IEEE Sensor Array and Multi-Channel Signal Processing (SAM) Workshop, Sitges, Barcelona, Spain, 2004.

[17] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 16, pp. 1451–1458, Oct. 1998.

[18] D. J. Love and R. W. Heath Jr., "Equal gain transmission in multiple-input multiple-output wireless systems," *IEEE Trans. Commun.*, vol. 51, no. 7, pp. 1102–1110, Jul. 2003.

[19] D. Gu and C. Leung, "Performance analysis of transmit diversity scheme with imperfect channel estimation," *IEE Electron. Lett.*, vol. 39, pp. 402–403, Feb. 2003.

[20] T. Baykas and A. Yongacoglu, "Robustness of transmit diversity schemes with multiple receive antennas at imperfect channel state information," in *Proc. IEEE Canadian Conf. Electrical Computer Engineering (CCECE) 2003*, vol. 1, May 2003, pp. 191–194.

Chandra R. Murthy received the B.Tech. degree in electrical engineering from the Indian Institute of Technology, Madras, in 1998 and the M.S. degree in electrical and computer engineering at Purdue University, West Lafayette, IN, in 2000. He is currently working towards the Ph.D. degree in electrical engineering at the University of California, San Diego.

From August 2000 to August 2002, he was an Engineer for Qualcomm, Inc., where he worked on WCDMA baseband transceiver design and 802.11b baseband receivers. His research interests are primarily in the areas of digital signal processing, information theory, estimation theory, and their applications in the optimization of MIMO, OFDM, and CDMA wireless communication systems.

Aditya K. Jagannatham received the B.Tech. degree in electrical engineering from the Indian Institute of Technology, Bombay, in 2001 and the M.S. degree from the Department of Electrical and Computer Engineering at the University of California, San Diego, in 2003. He is currently working toward the Ph.D. degree at the University of California, San Diego.

From June 2003 to September 2003, he was an intern at Zyray Wireless (now Broadcom), San Diego, CA, where he contributed to the WCDMA effort and the IEEE 802.11n high throughput WLAN standard. His research interests are in the area of digital signal processing, statistical estimation theory and its applications in multiple-input multiple-output, OFDM, and CDMA wireless communications.

Mr. Jagannatham was awarded the Bhavesh Gandhi Memorial award for best seminar from the Indian Institute of Technology, Bombay.

Bhaskar D. Rao (S'80–M'83–SM'91–F'00) received the B.Tech. degree in electronics and electrical communication engineering from the Indian Institute of Technology, Kharagpur, India, in 1979 and the M.S. and Ph.D. degrees from the University of Southern California, Los Angeles, in 1981 and 1983, respectively.

Since 1983, he has been with the University of California at San Diego, La Jolla, where he is currently a Professor with the Electrical and Computer Engineering Department. His interests are in the areas of digital signal processing, estimation theory, and optimization theory, with applications to digital communications, speech signal processing, and human–computer interactions.

Dr. Rao has been a member of the Statistical Signal and Array Processing technical committee and the Signal Processing Theory and Methods technical committee of the IEEE Signal Processing Society. He is currently a member of Signal Processing for Communications technical committee.