

Downlink Channel Estimation in Cellular Systems with Antenna Arrays at Base Stations Using Channel Probing with Feedback

Mehrzhad Biguesh

*Department of Communication Systems, University of Duisburg-Essen, Bismarckstrasse 81, 47057 Duisburg, Germany
Email: biguesh@sent5.uni-duisburg.de*

Alex B. Gershman

*Department of Communication Systems, University of Duisburg-Essen, Bismarckstrasse 81, 47057 Duisburg, Germany
Email: gershman@sent5.uni-duisburg.de*

Received 21 May 2003; Revised 4 December 2003

In mobile communication systems with multisensor antennas at base stations, downlink channel estimation plays a key role because accurate channel estimates are needed for transmit beamforming. One efficient approach to this problem is channel probing with feedback. In this method, the base station array transmits probing (training) signals. The channel is then estimated from feedback reports provided by the users. This paper studies the performance of the channel probing method with feedback using a multisensor base station antenna array and single-sensor users. The least squares (LS), linear minimum mean square error (LMMSE), and a new scaled LS (SLS) approaches to the channel estimation are studied. Optimal choice of probing signals is investigated for each of these techniques and their channel estimation performances are analyzed. In the case of multiple LS channel estimates, the best linear unbiased estimation (BLUE) scheme for their linear combining is developed and studied.

Keywords and phrases: antenna array, downlink channel, channel estimation, training sequence.

1. INTRODUCTION

In recent years, transmit beamforming has been a topic of growing interest [1, 2, 3, 4, 5]. The aim of transmit beamforming is to send desired information signals from the base station array to each user and, at the same time, to minimize undesired crosstalks, that is, to satisfy a certain quality of service constraint for each user. This task becomes very complicated if the transmitter does not have precise knowledge of the downlink channel information for each user. Therefore, the beamforming performance severely depends on the quality of channel estimates and an accurate downlink channel estimation plays a key role in transmit beamforming [6, 7, 8, 9]. One of the most popular approaches to downlink channel estimation is channel probing with user feedback [1, 2]. This approach suggests to probe the downlink channel by transmitting training signals from the base station to each user and then to estimate the channel from feedback reports provided by the users.

In this paper, we study the performance of channel probing with feedback in the case of a multisensor base station antenna array and single-sensor users [2]. We develop three channel estimators which offer different tradeoffs in

terms of performance and a priori required knowledge of the channel statistical parameters. First of all, the traditional least squares (LS) method is considered which does not require any knowledge of the channel parameters. Then, a refined version of the LS estimator is proposed (which is referred to as the scaled LS (SLS) estimator). The SLS estimator offers a substantially improved performance relative to the LS method but requires that the trace of the channel covariance matrix and the receiver noise powers be known a priori. Finally, the linear minimum mean square error (LMMSE) channel estimator is developed and studied. The latter technique is able to outperform both the LS and SLS estimators, but it requires the full a priori knowledge of the channel covariance matrix and the receiver noise powers. For each of the aforementioned techniques, the optimal choices of probing signal matrices for downlink channel measurement are studied and channel estimation errors are analyzed. Moreover, in the case of multiple LS channel estimates, an optimal scheme for their linear combining is proposed using the so-called best linear unbiased estimation (BLUE) approach. The effect of such a combining on the performance of downlink channel estimation is investigated.

2. BACKGROUND

We assume a base station array of L sensors and arbitrary geometry and consider the case of flat block fading¹ [2]. In this case, the signal received by the i th mobile user can be expressed as follows:

$$r_i(k) = s(k)\mathbf{w}^H\mathbf{h}_i + n_i(k), \quad (1)$$

where $s(k)$ is the transmitted signal, \mathbf{w} is the $L \times 1$ downlink weight vector, \mathbf{h}_i is the $L \times 1$ vector which describes an unknown complex vector channel from the array to the i th user, $n_i(k)$ is the user zero-mean white noise, and $(\cdot)^H$ stands for the Hermitian transpose.

In order to measure the vector channel for each user, the method of [2] suggests to use the so-called *probing mode* to transmit $N \geq L$ training signals $s(1), \dots, s(N)$ from the base station antenna array using the beamforming weight vectors $\mathbf{w}_1, \dots, \mathbf{w}_N$, respectively. The received signals at the i th mobile can be expressed as follows:

$$\mathbf{r}_i = \mathbf{W}^H\mathbf{h}_i + \mathbf{n}_i, \quad (2)$$

where

$$\mathbf{W} = [s^*(1)\mathbf{w}_1, s^*(2)\mathbf{w}_2, \dots, s^*(N)\mathbf{w}_N] \quad (3)$$

is the $L \times N$ *probing matrix*, $\mathbf{r}_i = [r_i(1), \dots, r_i(N)]^T$, $\mathbf{n}_i = [n_i(1), \dots, n_i(N)]^T$, and $(\cdot)^*$ and $(\cdot)^T$ stand for the complex conjugate and the transpose, respectively.

Then, each receiver (mobile user) employs the *information mode* to feed the data received in the probing mode back to the base station where these data are used to estimate the downlink vector channels. Alternatively (to decrease the amount of feedback bits), channel estimation can be done directly at each receiver. In the latter case, receivers feed the corresponding channel estimates back to the base station.

3. LS CHANNEL ESTIMATION

Knowing \mathbf{r}_i , the downlink vector channel between the base station and the i th user can be estimated using the least LS approach as [2]

$$\hat{\mathbf{h}}_i = \mathbf{W}^\dagger \mathbf{r}_i, \quad (4)$$

where $\mathbf{W}^\dagger = (\mathbf{W}\mathbf{W}^H)^{-1}\mathbf{W}$ is the pseudoinverse of \mathbf{W}^H . Assume that the transmitted power in the probing mode is constrained as:

$$\|\mathbf{W}\|_F^2 = P, \quad (5)$$

where P is a given power constant. We find \mathbf{W} which minimizes the channel estimation error for the i th user subject to the transmitted power constraint (5). This is equivalent to

the optimization problem

$$\min_{\mathbf{W}} E\{\|\mathbf{h}_i - \hat{\mathbf{h}}_i\|^2\} \quad \text{subject to } \|\mathbf{W}\|_F^2 = P, \quad (6)$$

where $E\{\cdot\}$ is the statistical expectation. Using (2) and (4), we have that $\mathbf{h}_i - \hat{\mathbf{h}}_i = \mathbf{W}^\dagger \mathbf{n}_i$ and, hence, the objective function in (6) can be rewritten as

$$\begin{aligned} J_{LS} &= E\{\|\mathbf{h}_i - \hat{\mathbf{h}}_i\|^2\} = E\{\|\mathbf{W}^\dagger \mathbf{n}_i\|^2\} \\ &= \sigma_i^2 \text{tr}\{\mathbf{W}^\dagger \mathbf{W}^{\dagger H}\} = \sigma_i^2 \text{tr}\{(\mathbf{W}\mathbf{W}^H)^{-1}\}, \end{aligned} \quad (7)$$

where we use the fact that $E\{\mathbf{n}_i \mathbf{n}_i^H\} = \sigma_i^2 \mathbf{I}$. Here, σ_i^2 is the noise power of the i th user, \mathbf{I} is the identity matrix, and $\text{tr}\{\cdot\}$ denotes the trace of a matrix.

Using (7), the optimization problem (6) can be equivalently written in the following form:

$$\min_{\mathbf{W}} \text{tr}\{(\mathbf{W}\mathbf{W}^H)^{-1}\} \quad \text{subject to } \text{tr}\{\mathbf{W}\mathbf{W}^H\} = P. \quad (8)$$

We obtain the solution to this problem using the Lagrange multiplier method, that is, via minimizing the function

$$L(\mathbf{W}, \lambda) = \text{tr}\{(\mathbf{W}\mathbf{W}^H)^{-1}\} + \lambda(\text{tr}\{\mathbf{W}\mathbf{W}^H\} - P), \quad (9)$$

where λ is the Lagrange multiplier.

To compute $\partial L(\mathbf{W}, \lambda) / \partial \mathbf{W}^H$, the following lemma will be useful.

Lemma 1. *If a square matrix \mathbf{F} is a function of another square matrix $\mathbf{G} = \mathbf{A} + \mathbf{B}\mathbf{X} + \mathbf{X}^H\mathbf{C}\mathbf{X}$, then the following chain rule is valid:*

$$\frac{\partial \text{tr}\{\mathbf{F}\}}{\partial \mathbf{X}} = \frac{\partial \text{tr}\{\mathbf{G}\}}{\partial \mathbf{X}} \frac{\partial \text{tr}\{\mathbf{F}\}}{\partial \mathbf{G}}, \quad (10)$$

where \mathbf{A} , \mathbf{B} , and \mathbf{C} are constant matrices and the dimensions of all the matrices in (10) are assumed to match.

Proof. See Appendix A. □

Furthermore, the following expressions for the matrix derivatives of traces will be used [10]:

$$\frac{\partial \text{tr}\{\mathbf{X}\mathbf{X}^H\}}{\partial \mathbf{X}^H} = \mathbf{X}^T, \quad (11)$$

$$\frac{\partial \text{tr}\{\mathbf{X}^{-1}\}}{\partial \mathbf{X}} = -\mathbf{X}^{-2T}. \quad (12)$$

Inserting $\mathbf{F} = (\mathbf{W}\mathbf{W}^H)^{-1}$, $\mathbf{X} = \mathbf{W}^H$, and $\mathbf{G} = \mathbf{W}\mathbf{W}^H$ into (10), we have

$$\frac{\partial \text{tr}\{(\mathbf{W}\mathbf{W}^H)^{-1}\}}{\partial \mathbf{W}^H} = \frac{\partial \text{tr}\{\mathbf{W}\mathbf{W}^H\}}{\partial \mathbf{W}^H} \frac{\partial \text{tr}\{(\mathbf{W}\mathbf{W}^H)^{-1}\}}{\partial \mathbf{W}\mathbf{W}^H}. \quad (13)$$

¹The flat fading assumption is valid for narrowband communication systems.

Applying (11) and (12) to (13), we can transform the latter equation as

$$\frac{\partial \text{tr} \{ (\mathbf{W}\mathbf{W}^H)^{-1} \}}{\partial \mathbf{W}^H} = -\mathbf{W}^T (\mathbf{W}\mathbf{W}^H)^{-2T}. \quad (14)$$

Using (14) and applying (11) to compute $\partial \text{tr} \{ \mathbf{W}\mathbf{W}^H \} / \partial \mathbf{W}^H$ in the second term of (9), we have that

$$\frac{\partial L(\mathbf{W}, \lambda)}{\partial \mathbf{W}^H} = \mathbf{W}^T (\lambda \mathbf{I} - (\mathbf{W}\mathbf{W}^H)^{-2T}). \quad (15)$$

Setting (15) to zero, we obtain that any probing matrix is the optimal one if it satisfies the equation

$$(\mathbf{W}\mathbf{W}^H)^{-2} = \lambda \mathbf{I}. \quad (16)$$

Since $\mathbf{W}\mathbf{W}^H$ is Hermitian and positive definite, we can write its eigendecomposition as

$$\mathbf{W}\mathbf{W}^H = \mathbf{Q}\mathbf{\Gamma}\mathbf{Q}^H, \quad (17)$$

where $\mathbf{\Gamma}$ is a diagonal matrix with positive eigenvalues on the main diagonal. Using the positiveness of the eigenvalues of $\mathbf{W}\mathbf{W}^H$ and taking into account that \mathbf{Q} is a unitary matrix ($\mathbf{Q}^H\mathbf{Q} = \mathbf{Q}\mathbf{Q}^H = \mathbf{I}$), we have from (16) that

$$\mathbf{Q}\mathbf{\Gamma}^{-2}\mathbf{Q}^H = \lambda \mathbf{I} \quad (18)$$

and, therefore,

$$\mathbf{\Gamma} = \frac{1}{\sqrt{\lambda}} \mathbf{I}. \quad (19)$$

Inserting (19) into (17) and using the identity $\mathbf{Q}\mathbf{Q}^H = \mathbf{I}$, we obtain that \mathbf{W} is an optimal probing matrix if

$$\mathbf{W}\mathbf{W}^H = \frac{1}{\sqrt{\lambda}} \mathbf{I}. \quad (20)$$

Using the power constraint (5), we can rewrite (20) as

$$\mathbf{W}\mathbf{W}^H = \frac{P}{L} \mathbf{I}. \quad (21)$$

Therefore, any probing matrix with *orthogonal rows* of the same norm $\sqrt{P/L}$ is an optimal one. Note that the similar fact has been earlier discovered from different points of view in [11, 12]. With such optimal probing, the LS estimator reduces to the simple decorrelator-type estimator.

According to (21), there is an infinite number of choices of the optimal probing matrix. It is also worth noting that each optimal choice of \mathbf{W} is *user independent*. Therefore, any probing matrix that satisfies (21) is optimal for *all users*.

It should be stressed that additional constraints on \mathbf{W} may be dictated by particular implementation issues. For example, the peak transmitted power per antenna may be limited. In this case, we have to distribute the transmitted power uniformly over the antennas and, therefore, the additional constraint is that all the elements of the optimal probing matrix should have the same magnitude. To satisfy this con-

straint, a properly normalized submatrix of the DFT matrix can be used, that is,

$$\mathbf{W} = \sqrt{\frac{P}{NL}} \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & W_N & W_N^2 & \cdots & W_N^{N-1} \\ 1 & W_N^2 & W_N^4 & \cdots & W_N^{2(N-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & W_N^{L-1} & W_N^{2(L-1)} & \cdots & W_N^{(L-1)(N-1)} \end{bmatrix}, \quad (22)$$

where $W_N = e^{j2\pi/N}$.

Using (21) along with (7), we obtain that the minimum downlink channel mean-square estimation error becomes

$$\min_{\mathbf{W}} J_{LS} = \frac{\sigma_i^2 L^2}{P}. \quad (23)$$

We stress that the error in (23) is proportional to the square of the number of transmit antennas and this may lead to a certain restriction of the dimension of the transmit array. However, one can compensate for this effect by increasing the total transmitted power in the probing mode.

Another interesting observation is that the error in (23) is independent of the channel realization \mathbf{h}_i and the array geometry.

4. SCALED LS CHANNEL ESTIMATION

Obviously, the LS estimate (4) does not necessarily minimize the channel estimation error because its objective is to minimize the signal estimation error rather than the channel estimation error. Therefore, it may be possible to use an additional scaling factor γ to further reduce this error. Using this idea, applying (2) and (4), and dropping the user index i for the sake of simplicity, we can write the channel estimation error in the following form:

$$\begin{aligned} E \{ \|\mathbf{h} - \gamma \hat{\mathbf{h}}_{LS}\|^2 \} &= \text{tr} \{ E \{ (\mathbf{h} - \gamma \hat{\mathbf{h}}_{LS})(\mathbf{h} - \gamma \hat{\mathbf{h}}_{LS})^H \} \} \\ &= (1 - \gamma)^2 \text{tr} \{ \mathbf{R}_h \} + \gamma^2 \sigma^2 \text{tr} \{ (\mathbf{W}\mathbf{W}^H)^{-1} \} \\ &= (J_{LS} + \text{tr} \{ \mathbf{R}_h \}) \left(\gamma - \frac{\text{tr} \{ \mathbf{R}_h \}}{J_{LS} + \text{tr} \{ \mathbf{R}_h \}} \right)^2 \\ &\quad + \frac{J_{LS} \text{tr} \{ \mathbf{R}_h \}}{J_{LS} + \text{tr} \{ \mathbf{R}_h \}}, \end{aligned} \quad (24)$$

where $\hat{\mathbf{h}}_{LS}$ is the LS channel estimate of (4), $\mathbf{R}_h = E \{ \mathbf{h}\mathbf{h}^H \}$ is the channel correlation matrix, and J_{LS} is given by (7). Clearly, (24) is minimized with

$$\gamma = \frac{\text{tr} \{ \mathbf{R}_h \}}{J_{LS} + \text{tr} \{ \mathbf{R}_h \}} \quad (25)$$

and the minimum of (24) with respect to γ is given by

$$\begin{aligned} J_{SLS} &= \min_{\gamma} E \{ \|\mathbf{h} - \gamma \hat{\mathbf{h}}_{LS}\|^2 \} \\ &= \frac{J_{LS} \text{tr} \{ \mathbf{R}_h \}}{J_{LS} + \text{tr} \{ \mathbf{R}_h \}} < J_{LS}. \end{aligned} \quad (26)$$

Note that the optimal γ in (25) is a function of the trace of the channel correlation matrix \mathbf{R}_h and the noise variance σ^2 . Therefore, these values have to be known (or preliminary estimated) when using the SLS approach. In practice, the estimate of $\text{tr}\{\mathbf{R}_h\}$,

$$\widehat{\text{tr}\{\mathbf{R}_h\}} = \hat{\mathbf{h}}_{\text{LS}}^H \hat{\mathbf{h}}_{\text{LS}}, \quad (27)$$

can be used in (25) in lieu of $\text{tr}\{\mathbf{R}_h\}$. Assuming that the values of $\text{tr}\{\mathbf{R}_h\}$ and σ^2 are given in advance, defining the SLS channel estimate as

$$\hat{\mathbf{h}}_{\text{SLS}} = \gamma \hat{\mathbf{h}}_{\text{LS}}, \quad (28)$$

and using (4) and (25), we have

$$\hat{\mathbf{h}}_{\text{SLS}} = \frac{\text{tr}\{\mathbf{R}_h\}}{\sigma^2 \text{tr}\{(\mathbf{W}\mathbf{W}^H)^{-1}\} + \text{tr}\{\mathbf{R}_h\}} \mathbf{W}^\dagger \mathbf{r}. \quad (29)$$

The optimal probing matrix for channel estimation using the SLS method can be found by means of solving the following optimization problem:

$$\min_{\mathbf{W}} J_{\text{SLS}} \quad \text{subject to } \text{tr}\{\mathbf{W}\mathbf{W}^H\} = P. \quad (30)$$

Since $\text{tr}\{\mathbf{R}_h\} > 0$, we see from (26) that J_{SLS} is a monotonically increasing function of J_{LS} . Note that $\text{tr}\{\mathbf{R}_h\}$ is not a function of \mathbf{W} , and, therefore, J_{LS} is the only term in (26) which depends on \mathbf{W} . This means that the optimization problems (6) and (30) are *equivalent*. Therefore, the optimal choice of probing matrix for the SLS channel estimation technique is the same as for the LS approach.

5. LMMSE CHANNEL ESTIMATION

In this section, we consider the LMMSE estimator of \mathbf{h} which is given by [13]

$$\begin{aligned} \hat{\mathbf{h}}_{\text{LMMSE}} &= \mathbf{R}_h \mathbf{W} (\mathbf{W}^H \mathbf{R}_h \mathbf{W} + \sigma^2 \mathbf{I})^{-1} \mathbf{r} \\ &= \sigma^{-2} (\mathbf{R}_h^{-1} + \sigma^{-2} \mathbf{W}\mathbf{W}^H)^{-1} \mathbf{W}\mathbf{r}. \end{aligned} \quad (31)$$

The performance of this estimator is characterized by the error $\mathbf{e} = \mathbf{h} - \hat{\mathbf{h}}_{\text{LMMSE}}$ whose mean is zero, and the covariance matrix is given by [13]

$$\mathbf{R}_e = \mathbf{E}\{\mathbf{e}\mathbf{e}^H\} = (\mathbf{R}_h^{-1} + \sigma^{-2} \mathbf{W}\mathbf{W}^H)^{-1}. \quad (32)$$

The LMMSE estimation error is given by

$$J_{\text{LMMSE}} = \mathbf{E}\left\{\|\mathbf{h} - \hat{\mathbf{h}}_{\text{LMMSE}}\|^2\right\} = \text{tr}\{\mathbf{R}_e\}. \quad (33)$$

To minimize (33) subject to the transmitted power constraint $\text{tr}\{\mathbf{W}\mathbf{W}^H\} = P$, we can use the Lagrange multiplier method. The problem can be written as follows:

$$L = \text{tr}\left\{(\mathbf{R}_h^{-1} + \sigma^{-2} \mathbf{W}\mathbf{W}^H)^{-1}\right\} + \lambda \text{tr}\{\mathbf{W}\mathbf{W}^H\}. \quad (34)$$

Using the chain rule (10), it can be readily shown that the optimal probing must satisfy

$$\mathbf{W}\mathbf{W}^H = \frac{\sigma^2}{\sqrt{\lambda}} \mathbf{I} - \sigma^2 \mathbf{R}_h^{-1}. \quad (35)$$

Using the constraint $\text{tr}\{\mathbf{W}\mathbf{W}^H\} = P$, (35) can be rewritten as follows:

$$\mathbf{W}\mathbf{W}^H = \frac{1}{L} (P + \sigma^2 \text{tr}\{\mathbf{R}_h^{-1}\}) \mathbf{I} - \sigma^2 \mathbf{R}_h^{-1}. \quad (36)$$

Interestingly, in the high signal-to-noise ratio (SNR) case ($\sigma^2 \rightarrow 0$), (36) transforms to (21). Therefore, in the high SNR domain, the LS, SLS, and LMMSE approaches all have the same condition on optimal probing matrices.

Using (36), we obtain that in the optimal probing case,

$$\mathbf{R}_e = \frac{\sigma^2 L}{P + \sigma^2 \text{tr}\{\mathbf{R}_h^{-1}\}} \mathbf{I}. \quad (37)$$

Therefore,

$$\min_{\mathbf{W}} J_{\text{LMMSE}} = \frac{\sigma^2 L^2}{P + \sigma^2 \text{tr}\{\mathbf{R}_h^{-1}\}}. \quad (38)$$

If the channel coefficients are all i.i.d. random variables, we have $\mathbf{R}_h = \xi^2 \mathbf{I}$, where ξ^2 can be viewed as the channel attenuation parameter. In this case, (36) transforms to (21) and, therefore, the optimal probing matrix for the LS estimator is also optimal for the LMMSE estimator. Furthermore, in such a situation, the minimum of the channel estimation error is given by

$$\min_{\mathbf{W}} J_{\text{LMMSE}} = \frac{\xi^2 \sigma^2 L^2}{\xi^2 P + \sigma^2 L}. \quad (39)$$

Interestingly, if $\mathbf{R}_h = \xi^2 \mathbf{I}$, then (26) and (39) are identical which means that the performances of the SLS and LMMSE estimators are similar in this case.

6. COMBINING OF MULTIPLE LS CHANNEL ESTIMATES

In Sections 3, 4, and 5, the specific case of a single channel estimate has been considered. In this section, we extend the optimal probing approach to the case of *multiple* LS channel estimates. If there are multiple probing periods available within the channel coherency time, it may be inefficient from the computational and buffering viewpoints to store and process dynamically long amounts of data that are formed by accumulation of multiple received data blocks corresponding to different probing periods. A good alternative here is to obtain a particular channel estimate for each probing period and then to store these estimates dynamically rather than storing the data itself, and to compute the final channel estimate based on a proper combination of such (previously obtained) particular estimates.

Let us have K estimates $\hat{\mathbf{h}}_{i,1}, \dots, \hat{\mathbf{h}}_{i,K}$ of the downlink channel corresponding to the i th user. Let each estimate

be computed using (4) based on some probing matrices $\mathbf{W}_1, \dots, \mathbf{W}_K$, respectively. The channel is assumed to be quasistatic (fixed) at the interval of K probeings, and $P_k = \|\mathbf{W}_k\|_F^2$ is the transmitted power during the k th probing.

We aim to improve the performance of downlink channel estimation by combining the estimated values $\hat{\mathbf{h}}_{i,k}$ for $k = 1, \dots, K$ in a linear way as follows:

$$\hat{\mathbf{h}}_i = \sum_{k=1}^K \alpha_{i,k} \hat{\mathbf{h}}_{i,k}, \quad (40)$$

where $\alpha_{i,k}$ are unknown weighting coefficients.

Let us obtain the optimal weighting coefficients by means of minimizing the error in (40). Then, these coefficients can be found by solving the following optimization problem:

$$\min_{\alpha_{i,1}, \dots, \alpha_{i,K}} E \left\{ \left\| \mathbf{h}_i - \sum_{k=1}^K \alpha_{i,k} \hat{\mathbf{h}}_{i,k} \right\|^2 \right\} \quad \text{subject to} \quad \sum_{k=1}^K \alpha_{i,k} = 1, \quad (41)$$

where the constraint in (41) guarantees the unbiasedness of the final channel estimate. This problem corresponds to the so-called BLUE estimator [13].

The solution to (41) is given by the following lemma.

Lemma 2. *The optimal weights $\{\alpha_{i,k}\}_{k=1}^K$ for the i th user are given by*

$$\alpha_{i,k} = \frac{1}{\text{tr} \left\{ (\mathbf{W}_k \mathbf{W}_k^H)^{-1} \right\} \sum_{n=1}^K 1 / \text{tr} \left\{ (\mathbf{W}_n \mathbf{W}_n^H)^{-1} \right\}}. \quad (42)$$

Proof. See Appendix B. \square

It is worth noting that the optimal weighting coefficients $\alpha_{i,k}$ are user independent (i.e., they are the same for each user).

Choosing optimal orthogonal weighting matrices in each probing, we have

$$\begin{aligned} \text{tr} \left\{ (\mathbf{W}_k \mathbf{W}_k^H)^{-1} \right\} &= \frac{L^2}{P_k}, \\ \sum_{n=1}^K \frac{1}{\text{tr} \left\{ (\mathbf{W}_n \mathbf{W}_n^H)^{-1} \right\}} &= \frac{P_{\text{tot}}}{L^2}, \end{aligned} \quad (43)$$

where

$$P_{\text{tot}} = \sum_{k=1}^K P_k \quad (44)$$

is the total transmitted power during the K probeings.

Inserting (43) into (42), we obtain that in the case of using optimal orthogonal weighting matrices, the expression for optimal weighting coefficients can be simplified to

$$\alpha_{i,k} = \frac{P_k}{P_{\text{tot}}}. \quad (45)$$

In this case, the downlink channel estimation error is given by

$$\begin{aligned} E \left\{ \|\mathbf{h}_i - \hat{\mathbf{h}}_i\|^2 \right\} &= E \left\{ \left\| \mathbf{h}_i - \sum_{k=1}^K \frac{P_k}{P_{\text{tot}}} \hat{\mathbf{h}}_{i,k} \right\|^2 \right\} \\ &= E \left\{ \left\| \sum_{k=1}^K \frac{P_k}{P_{\text{tot}}} (\mathbf{h}_i - \hat{\mathbf{h}}_{i,k}) \right\|^2 \right\} \\ &= \frac{L^2}{P_{\text{tot}}^2} E \left\{ \left\| \sum_{k=1}^K \mathbf{W}_k \mathbf{n}_{i,k} \right\|^2 \right\} \\ &= \frac{\sigma_i^2 L^2}{P_{\text{tot}}^2} \text{tr} \left\{ \sum_{k=1}^K \mathbf{W}_k \mathbf{W}_k^H \right\} = \frac{\sigma_i^2 L^2}{P_{\text{tot}}}, \end{aligned} \quad (46)$$

where $\mathbf{n}_{i,k}$ is the zero-mean white noise vector of the i th user in the k th probing. When deriving (46), we have used the property $E \{ \mathbf{n}_{i,k} \mathbf{n}_{i,l}^H \} = \sigma_i^2 \delta_{k,l} \mathbf{I}$, where $\delta_{k,l}$ is the Kronecker delta.

We observe that, similar to (23), the error in (46) is independent of the channel realization and the array geometry. Comparing (46) with (23), we see that the optimal linear combining of multiple estimates reduces the estimation error by a factor of P_{tot}/P . For example, if each probing has the same power ($P_k = P$, $K = 1, 2, \dots, K$), then $P_{\text{tot}} = KP$ and the estimation error is reduced by a factor of K .

7. NUMERICAL EXAMPLES

In our simulations, we compare the performance of the LS, SLS, and LMMSE channel estimators in the cases of optimal and nonoptimal choices of probing matrices. Throughout all our simulation examples, we assume that $N = L$. The channel coefficients and the receiver noise are assumed to be circular complex Gaussian random variables with the unit variance.

We assume that the base station has a uniform linear array and the downlink channel correlation matrix \mathbf{R}_h has the following structure:

$$[\mathbf{R}_h]_{n,m} = r^{|n-m|}, \quad 0 \leq r < 1, \quad (47)$$

where n and m are the indices of the array sensors. This model of the array covariance is frequently used in the literature; see [14, 15, 16] and references therein.

The elements of $L \times L$ probing matrices \mathbf{W} in the case of nonoptimal probing have been drawn independently from a zero-mean complex Gaussian random generator in each simulation run. However, to avoid possible computational inaccuracy of the LS estimator, we have ignored all probing matrices that have resulted in a condition number of $\mathbf{W}\mathbf{W}^H$ greater than 10^4 . Each simulated point is obtained by averaging 5000 independent simulation runs.

In Figure 1, we display the mean of the norm squared of the channel estimation error (MNSE) of the LS channel estimator in the optimal and nonoptimal probing matrix cases. In this figure, MNSEs are plotted versus the probing power

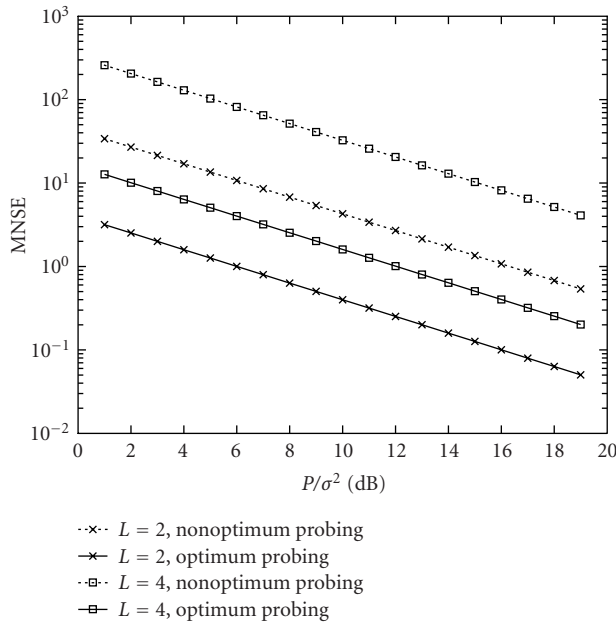


FIGURE 1: MNSEs versus P/σ^2 for the LS estimator.

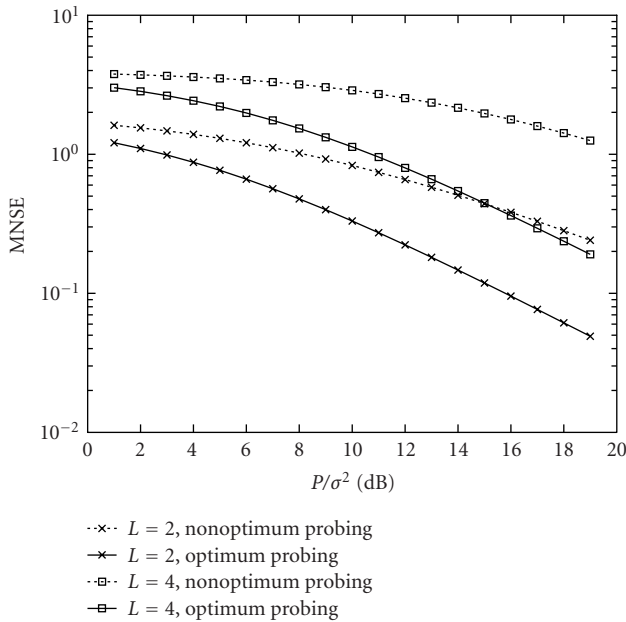


FIGURE 2: MNSEs versus P/σ^2 for the SLS estimator.

P/σ^2 . Note that the performance of the LS estimator is independent of the parameter r . The parameter L is varied in this figure.

In Figure 2, the performance of the SLS estimator is tested under the similar conditions. Similar to the LS method, the performance of the LS estimator is independent of the parameter r .

Figures 3 and 4 display the performance of the LMMSE estimator in the cases of $r = 0$ and $r = 0.25$, respectively.

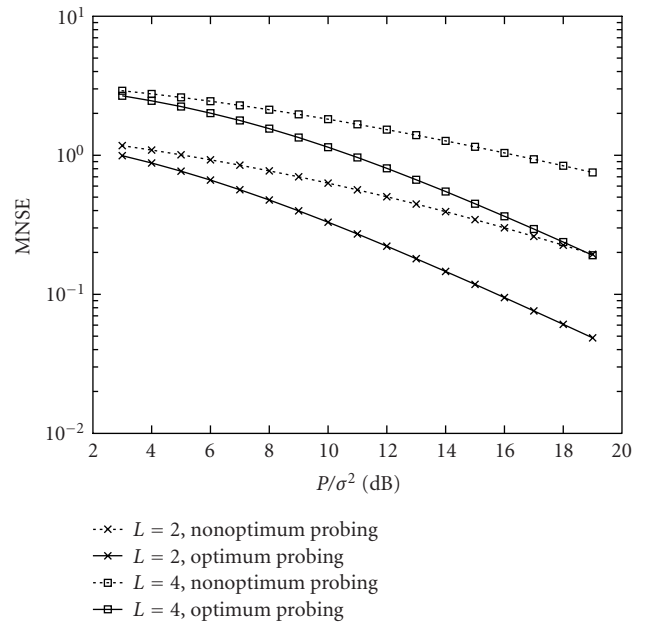


FIGURE 3: MNSEs versus P/σ^2 for the LMMSE estimator in the case of uncorrelated channel coefficients ($r = 0$).

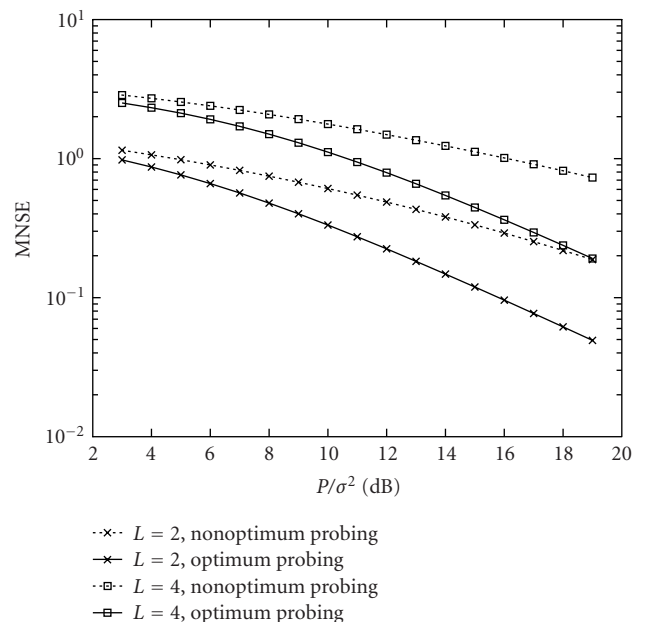
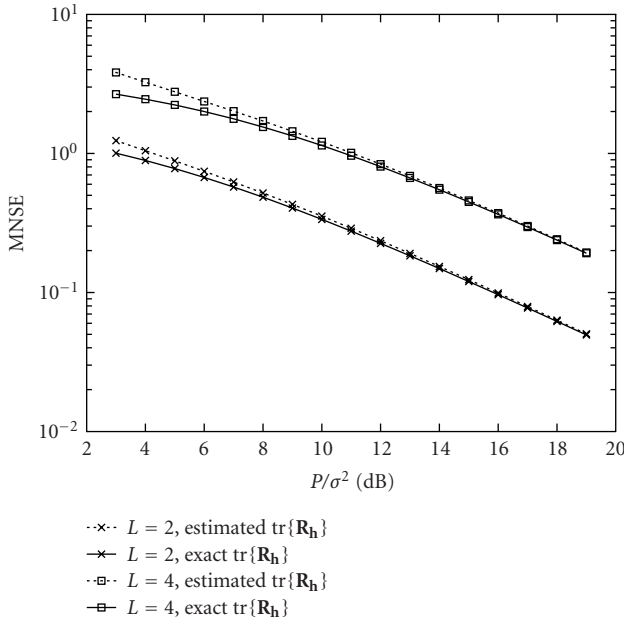


FIGURE 4: MNSEs versus P/σ^2 for the LMMSE estimator in the case of correlated channel coefficients ($r = 0.25$).

In both figures, the channel covariance matrix \mathbf{R}_h is assumed to be known exactly. Other conditions are similar to that of Figures 1 and 2.

From Figures 1, 2, 3, and 4, it can be seen that the optimal probing improves the quality of channel estimation substantially for all methods. Note that this improvement is especially pronounced for large values of P/σ^2 if the SLS or LMMSE method is used. Comparing Figures 3 and 4, we also see that these figures give nearly the same results. This means

FIGURE 5: MNSEs versus P/σ^2 for the SLS estimator.

that moderate correlation of the channel coefficients does not affect the LMMSE approach.

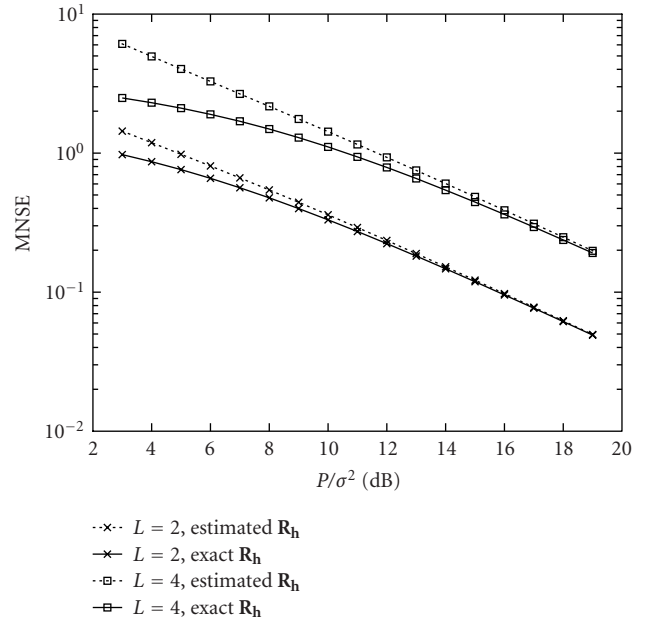
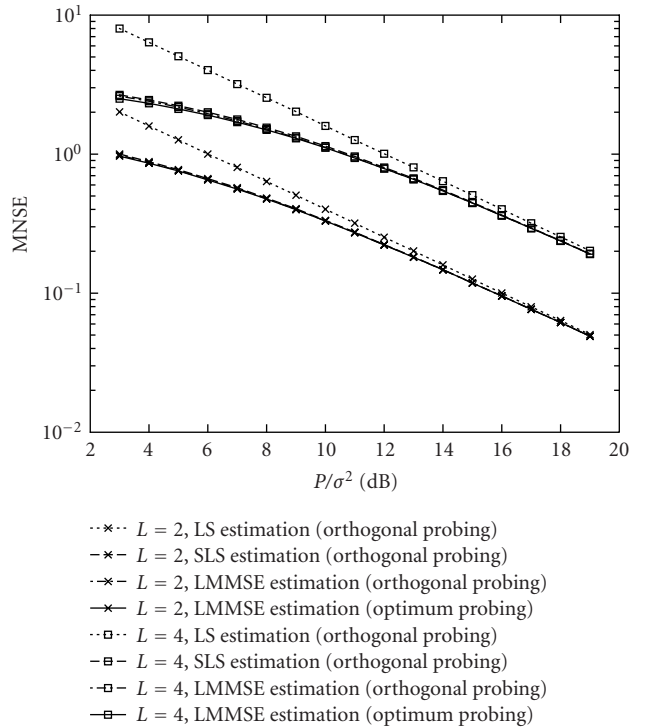
As it has been mentioned before, the SLS channel estimator requires the knowledge of $\text{tr}\{\mathbf{R}_h\}$. However, note that the LS estimator can be applied to estimate this parameter using (27). In Figure 5, the MNSEs of the SLS estimator with optimal probing are plotted versus P/σ^2 in the cases when the exact and estimated values of $\text{tr}\{\mathbf{R}_h\}$ are used. In the latter case, the LS method is applied to obtain the estimate of $\text{tr}\{\mathbf{R}_h\}$ which is then inserted into the SLS estimator. All other conditions are similar to that of the previous figures.

In the LMMSE method, the full knowledge of the channel correlation matrix \mathbf{R}_h is required either at the base station or at the mobile station to estimate the channel (depending on where the channel estimation is done). Also, the base station transmitter has to know this matrix in order to compute the optimal probing matrix. However, one may use the following rank-one estimate of this matrix:

$$\hat{\mathbf{R}}_h = \hat{\mathbf{h}}_{LS} \hat{\mathbf{h}}_{LS}^H. \quad (48)$$

In Figure 6, the performance of the LMMSE channel estimator is tested versus P/σ^2 in the cases when \mathbf{R}_h is known exactly and when its estimate (48) is used. In the latter case, the optimal LS probing is used (note, however, that in the general case, such a probing is nonoptimal for the LMMSE approach). The value of L is varied in this figure and $r = 0.25$ is taken.

From Figures 5 and 6, we see that there are only small performance losses caused by using the estimated values of $\text{tr}\{\mathbf{R}_h\}$ and \mathbf{R}_h in the SLS and LMMSE estimators, respectively, in lieu of the exact values of $\text{tr}\{\mathbf{R}_h\}$ and \mathbf{R}_h . Also, from Figure 6, we see that the optimal LS probing becomes nearly

FIGURE 6: MNSE versus P/σ^2 for the LMMSE estimator in the case of correlated channel coefficients ($r = 0.25$).FIGURE 7: Comparison of the performances of the LS, SLS, and LMMSE estimators versus P/σ^2 in the case of correlated channel coefficients ($r = 0.25$).

optimal for the LMMSE approach starting from moderate values of SNR. This observation supports theoretical results of Section 5.

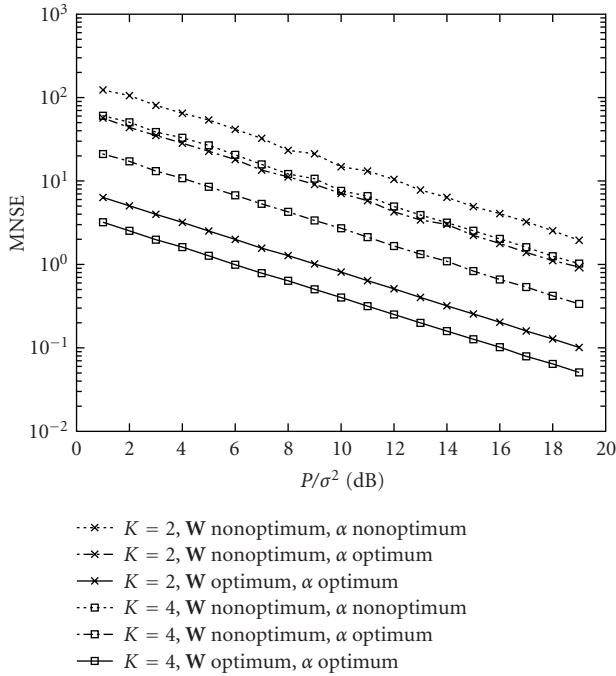


FIGURE 8: MNSE versus P/σ^2 for the case of multiple LS channel estimates (the BLUE estimator).

Figure 7 compares the performances of the LS, SLS, and LMMSE estimators versus P/σ^2 . In this figure, we assume that $r = 0.25$, and two variants of the LMMSE estimator are considered. Both these variants assume that the estimator knows \mathbf{R}_h exactly, but the first variant uses the optimal probing signal that satisfies (36), while the second one employs the matrix which satisfies (21) and, therefore, is optimal only for LS and SLS estimators and/or for the uncorrelated channel case ($r = 0$). From this figure, we observe that the difference in performance between the first and second variants of the LMMSE estimator is negligible at all the tested values of SNR. Therefore, the LS/SLS probing appears to be suboptimal for the LMMSE estimator.

In the last example, the case of multiple LS channel estimates are assumed. In Figure 8, the parameter $L = 4$ is chosen and the performance of the BLUE estimator is compared for $K = 2$ and $K = 4$. Three cases are considered in this figure:

- (i) both the probing matrices and the coefficients $\alpha_{i,k}$ are optimal;
- (ii) the probing matrices are nonoptimal but the coefficients $\alpha_{i,k}$ are optimal;
- (iii) both the probing matrices and the coefficients $\alpha_{i,k}$ are nonoptimal.

In the third case, the coefficients $\alpha_{i,k} = 1/K$ are assumed for all i and k .

Figure 8 demonstrates substantial improvements which can be achieved when the BLUE estimator is used in the case

of multiple channel estimates. This figure also shows that the choice of optimal probing matrices and coefficients $\alpha_{i,k}$ is critical for the estimator performance as nonoptimal choices of one or both of these parameters may cause a severe performance degradation.

8. CONCLUSIONS

We have studied the performance of the channel probing method with feedback using a multisensor base station antenna array and single-sensor users. Three channel estimators have been developed which offer different tradeoffs in terms of performance and a priori required knowledge of the channel statistical parameters. First of all, the traditional LS method has been considered. The LS estimator does not require any knowledge of the channel parameters. Then, a new (refined) version of the LS estimator has been proposed. This refined technique is referred to as the SLS estimator. It has been shown to offer a substantially improved channel estimation performance relative to the LS method but requires that the trace of the channel covariance matrix and the receiver noise powers be known a priori. Finally, the LMMSE channel estimator is developed and studied. The latter technique has been shown to potentially outperform both the LS and SLS estimators, but it requires the full a priori knowledge of the channel covariance matrix and the receiver noise powers.

For each of the above mentioned techniques, the optimal choices of probing signal matrices for downlink channel measurement have been studied and channel estimation errors have been analyzed. In the case of multiple LS channel estimates, the BLUE scheme for their linear combining has been developed.

Simulation examples have demonstrated substantial performance improvements that can be achieved using optimal channel probing.

APPENDICES

A. PROOF OF LEMMA 1

First of all, we prove the chain rule for the particular case when $\mathbf{G} = \mathbf{B}\mathbf{X}$. Writing this equation elementwise, we have $g_{i,l} = \sum_k b_{i,k}x_{k,l}$ and, therefore,

$$\frac{\partial g_{i,l}}{\partial x_{m,n}} = \delta_{i,n}b_{i,m}, \quad (\text{A.1})$$

where the Wirtinger derivatives for complex variables are used, $\delta_{i,n}$ is the Kronecker delta, and

$$b_{i,m} = \frac{\partial \text{tr}\{\mathbf{G}\}}{\partial x_{m,i}}. \quad (\text{A.2})$$

Since \mathbf{F} is a function of \mathbf{G} , then $\text{tr}\{\mathbf{F}\}$ can be a function of all elements of \mathbf{G} . Thus, applying the extended derivative chain

rule ([17, page 99]) and (A.1)-(A.2), we have

$$\begin{aligned} \left[\frac{\partial \text{tr}\{\mathbf{F}\}}{\partial \mathbf{X}} \right]_{m,n} &= \frac{\partial \text{tr}\{\mathbf{F}\}}{\partial x_{m,n}} = \sum_i \sum_l \frac{\partial \text{tr}\{\mathbf{F}\}}{\partial g_{i,l}} \frac{\partial g_{i,l}}{\partial x_{m,n}} \\ &= \sum_i \frac{\partial \text{tr}\{\mathbf{F}\}}{\partial g_{i,n}} b_{i,m} = \sum_i \frac{\partial \text{tr}\{\mathbf{G}\}}{\partial x_{m,i}} \frac{\partial \text{tr}\{\mathbf{F}\}}{\partial g_{i,n}} \quad (\text{A.3}) \\ &= \left[\frac{\partial \text{tr}\{\mathbf{G}\}}{\partial \mathbf{X}} \frac{\partial \text{tr}\{\mathbf{F}\}}{\partial \mathbf{G}} \right]_{m,n} \end{aligned}$$

and the proof for the particular case $\mathbf{G} = \mathbf{B}\mathbf{X}$ is completed.

To extend the proof to the general case $\mathbf{G} = \mathbf{A} + \mathbf{B}\mathbf{X} + \mathbf{X}^H\mathbf{C}\mathbf{X}$, we notice that this equation can be rewritten as $\mathbf{G} = \mathbf{A} + (\mathbf{B} + \mathbf{X}^H\mathbf{C})\mathbf{X}$ and, therefore, the established result for the particular case $\mathbf{G} = \mathbf{B}\mathbf{X}$ can be applied taking into account that the matrix \mathbf{A} is constant and that $\partial \text{tr}\{\mathbf{B} + \mathbf{X}^H\mathbf{C}\}/\partial \mathbf{X} = 0$. In other words, replacing the matrix \mathbf{B} by the matrix $\mathbf{B} + \mathbf{X}^H\mathbf{C}$, we straightforwardly extend our proof to the general case.

B. PROOF OF LEMMA 2

To solve (41), we insert (4) into the objective function of (41) and, using (2), rewrite it as

$$\begin{aligned} &E \left\{ \text{tr} \left\{ \left(\sum_{m=1}^K \alpha_{i,m} \mathbf{W}_m^\dagger \mathbf{n}_{i,m} \right) \left(\sum_{n=1}^K \alpha_{i,n} \mathbf{W}_n^\dagger \mathbf{n}_{i,n} \right)^H \right\} \right\} \\ &= \text{tr} \left\{ \left(\sum_{m=1}^K \sum_{n=1}^K \alpha_{i,m} \alpha_{i,n}^* \mathbf{W}_m^\dagger \mathbf{W}_n^H E \{ \mathbf{n}_{i,m} \mathbf{n}_{i,n}^H \} \right) \right\} \quad (\text{B.1}) \\ &= \text{tr} \left\{ \sigma_i^2 \sum_{n=1}^K |\alpha_{i,n}|^2 (\mathbf{W}_n \mathbf{W}_n^H)^{-1} \right\}, \end{aligned}$$

where $\mathbf{n}_{i,m}$ is the noise vector of the i th user during the m th probing interval and the property $E \{ \mathbf{n}_{i,m} \mathbf{n}_{i,n}^H \} = \delta_{mn} \mathbf{I}$ is used.

To minimize (B.1) subject to the constraint $\sum_{k=1}^K \alpha_{i,k} = 1$, we have to find the minimum of the Lagrangian

$$L(\boldsymbol{\alpha}, \lambda) = \text{tr} \left\{ \sigma_i^2 \sum_{k=1}^K |\alpha_{i,k}|^2 (\mathbf{W}_k \mathbf{W}_k^H)^{-1} \right\} - \lambda \left(\sum_{k=1}^K \alpha_{i,k} - 1 \right), \quad (\text{B.2})$$

where the vector $\boldsymbol{\alpha}$ captures all the coefficients $\alpha_{i,k}$.

The gradient of (B.2) is given by

$$\frac{\partial L(\boldsymbol{\alpha}, \lambda)}{\partial \alpha_{i,k}} = 2\sigma_i^2 \alpha_{i,k} \text{tr} \left\{ (\mathbf{W}_k \mathbf{W}_k^H)^{-1} \right\} - \lambda. \quad (\text{B.3})$$

Setting it to zero, we have

$$\alpha_{i,k} = \frac{\lambda}{2\sigma_i^2 \text{tr} \left\{ (\mathbf{W}_k \mathbf{W}_k^H)^{-1} \right\}}. \quad (\text{B.4})$$

Noting that $\sum_{k=1}^K \alpha_{i,k} = 1$, we obtain (42).

ACKNOWLEDGMENTS

A. B. Gershman is on research leave from the Department of Electrical and Computer Engineering, McMaster University, Canada. This work was supported in part by the Wolfgang Paul Award Program, the Alexander von Humboldt Foundation; Premier's Research Excellence Award Program, the Ministry of Energy, Science and Technology (MEST) of Ontario; Natural Sciences and Engineering Research Council (NSERC), Canada; and Communications and Information Technology Ontario (CITO).

REFERENCES

- [1] D. Gerlach and A. Paulraj, "Adaptive transmitting antenna methods for multipath environments," in *Proc. IEEE Global Telecommunications Conference (GLOBECOM '94)*, vol. 1, pp. 425–429, San Francisco, Calif, USA, November–December 1994.
- [2] D. Gerlach and A. Paulraj, "Adaptive transmitting antenna arrays with feedback," *IEEE Signal Processing Letters*, vol. 1, no. 10, pp. 150–152, 1994.
- [3] D. Gerlach and A. Paulraj, "Base station transmitting antenna arrays for multipath environments," *Signal Processing*, vol. 54, no. 1, pp. 59–73, 1996.
- [4] F. Rashid-Farrokhi, K. J. R. Liu, and L. Tassiulas, "Transmit beamforming and power control for cellular wireless systems," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 8, pp. 1437–1450, 1998.
- [5] C. Farsakh and J. A. Nossek, "Spatial covariance based downlink beamforming in an SDMA mobile radio system," *IEEE Trans. Communications*, vol. 46, no. 11, pp. 1497–1506, 1998.
- [6] S. Bhashyam, A. Sabharwal, and U. Mitra, "Channel estimation for multirate DS-CDMA systems," in *Proc. 34th Asilomar Conference on Signals, Systems & Computers*, vol. 2, pp. 960–964, Pacific Grove, Calif, USA, October–November 2000.
- [7] A. Arredondo, K. R. Dandekar, and G. Xu, "Vector channel modeling and prediction for the improvement of downlink received power," *IEEE Trans. Communications*, vol. 50, no. 7, pp. 1121–1129, 2002.
- [8] Y.-C. Liang and F. P. S. Chin, "Downlink channel covariance matrix (DCCM) estimation and its applications in wireless DS-CDMA systems," *IEEE Journal on Selected Areas in Communications*, vol. 19, no. 2, pp. 222–232, 2001.
- [9] M. Biguesh, S. Shahbazpanahi, and A. B. Gershman, "Robust power adjustment for transmit beamforming in cellular communication systems," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP '03)*, vol. 5, pp. 105–108, Hong Kong, April 2003.
- [10] H. Lütkepohl, *Handbook of Matrices*, John Wiley & Sons, New York, NY, USA, 1996.
- [11] S. D. Silverstein, "Application of orthogonal codes to the calibration of active phased array antennas for communication satellites," *IEEE Transactions on Signal Processing*, vol. 45, no. 1, pp. 206–218, 1997.
- [12] T. L. Marzetta, "BLAST training: estimating channel characteristics for high capacity space-time wireless," in *Proc. 37th Annual Allerton Conference on Communication, Control, and Computing*, pp. 958–966, Monticello, Ill, USA, September 1999.
- [13] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, Prentice-Hall, Englewood Cliffs, NJ, USA, 1993.
- [14] A. B. Gershman, C. F. Mecklenbräuker, and J. F. Böhme, "Matrix fitting approach to direction of arrival estimation with

- imperfect spatial coherence of wavefronts," *IEEE Transactions on Signal Processing*, vol. 45, no. 7, pp. 1894–1899, 1997.
- [15] A. B. Gershman, P. Stoica, M. Pesavento, and E. G. Larsson, "Stochastic Cramer-Rao bound for direction estimation in unknown noise fields," *IEEE Proceedings-Radar, Sonar and Navigation*, vol. 149, no. 1, pp. 2–8, 2002.
- [16] J. Ringelstein, A. B. Gershman, and J. F. Böhme, "Direction finding in random inhomogeneous media in the presence of multiplicative noise," *IEEE Signal Processing Letters*, vol. 7, no. 10, pp. 269–272, 2000.
- [17] G. A. Korn and T. M. Korn, *Mathematical Handbook for Scientists and Engineers*, Dover Publications, Mineola, NY, USA, 2000.

Mehrzad Biguesh was born in Shiraz, Iran. He received the B.S. degree in electronics engineering from Shiraz University in 1991, and the M.S. and Ph.D. degrees in telecommunications (with honors) from Sharif University of Technology (SUT), Tehran, Iran, in 1994 and 2000, respectively. During his Ph.D. studies, he was appointed at Guilan university and SUT as a Lecturer. From November 1998 to August 1999,



he was with INRS-Telecommunications, University of Quebec, Canada, as a Doctoral Trainee. From 1999 to 2001, he held an appointment at the Iran Telecom Research Center (ITRC), Teheran. From 2000 to 2002, he was with the Electronics Research Center at SUT and held several short-time appointments in the telecommunications industry. Since March 2002, he has been a Postdoctoral Fellow in the Department of Communication Systems, University of Duisburg-Essen, Duisburg, Germany. His research interests include array signal processing, MIMO systems, wireless communications, and radar systems.

Alex B. Gershman received his Diploma and Ph.D. degrees in radiophysics from the Nizhny Novgorod University, Russia, in 1984 and 1990, respectively. From 1984 to 1989, he was with the Radiotechnical and Radiophysical Institutes, Nizhny Novgorod. From 1989 to 1997, he was with the Institute of Applied Physics, Nizhny Novgorod. From 1997 to 1999, he was a Research Associate at the Department of Electrical Engineering,



Ruhr University, Bochum, Germany. In 1999, he joined the Department of Electrical and Computer Engineering, McMaster University, Hamilton, Ontario, Canada where he is now a Professor. He also held visiting positions at the Swiss Federal Institute of Technology, Lausanne, Ruhr University, Bochum, and Gerhard-Mercator University, Duisburg. His main research interests are in statistical and array signal processing, adaptive beamforming, MIMO systems and space-time coding, multiuser communications, and parameter estimation. He has published over 220 technical papers in these areas. Dr. Gershman was a recipient of the 1993 URSI Young Scientist Award, the 1994 Outstanding Young Scientist Presidential Fellowship (Russia), the 1994 Swiss Academy of Engineering Science Fellowship, and the 1995–1996 Alexander von Humboldt Fellowship (Germany). He received the 2000 Premiers Research Excellence Award, Ontario, Canada, and the 2001 Wolfgang Paul Award, Alexander von Humboldt Foundation, Germany. He was also a recipient of the 2002 Young Explorers Prize from the Canadian Institute for Advanced Research (CIAR), which has honored Canada's

top 20 researchers aged 40 or under. He is an Associate Editor for the IEEE Transactions on Signal Processing and EURASIP Journal on Wireless Communications and Networking, as well as a Member of the SAM Technical Committee of the IEEE SP Society.