Title

Feedback assisted transmission subspace tracking for MIMO systems

Journal

IEEE Journal on Selected Areas in Communications, 21(3)

ISSN

0733-8716

Authors

Banister, B C

Zeidler, J R

Publication Date

2003-04-01

Peer reviewed

Feedback Assisted Transmission Subspace Tracking for MIMO Systems

Brian Clarke Banister, *Member, IEEE*, and James R. Zeidler, *Fellow, IEEE*

***Abstract*—This paper describes a feedback assisted stochastic gradient algorithm for transmission tracking of the dominant channel subspaces for multiple-input–multiple-output (MIMO) communications systems. Subspace tracking is introduced as a means of tracking multiple transmission weights, being the MIMO generalization of beam steering in the familiar multiple-input–single-output case. The subspace solution approximates that of water filling (WF) in some cases, without the complete rate/power allocation required by WF. The gain of subspace tracking in low rank systems is demonstrated, particularly, in the case where the number of transmit antennas exceeds the number of receive antennas. Simulations of ergodic capacity show the utility of both subspace tracking in general and of the specific adaptation algorithm, and simulations of frame-error rates show the utility in a specific coding example.**

***Index Terms*—Adaptive arrays, gradient methods, multiple-input–multiple-output (MIMO) systems, transmitting antennas.**

I. Introduction

SPACE–TIME CODING (STC) for capacity enhancement of multiple-input–multiple-output (MIMO) channels uses the independent modes of the channel-state matrix to effectively obtain multiple spatial transmission pipes, giving an increase in the effective transmission bandwidth and allowing for greater bit rates. Most of the research into STC with multiple transmit antennas has focused on "blind" techniques, where no knowledge of the forward channel state is available to the transmitter. These attain diversity with a single receive antenna [1] or multiplexing coding gain with multiple receive antennas [2]–[4]. There is also a substantial body of literature on the subject of signal processing approaches for transmit adaptation in multiple-input–single-output (MISO) environments, including closed loop techniques utilizing digital feedback [5]–[8]. While it has been clear from the basic capacity formula [9] that utilization by the transmitter of the channel state could also be advantageous in MIMO space–time coded systems, there is only a small body of literature on specific techniques for attaining such knowledge in the MIMO case, particularly in frequency-division duplex (FDD) systems. Those works have focused on antenna selection algorithms [10]–[13].

Manuscript received May 1, 2002; revised November 1, 2002.

B. C. Banister is with the Department of Electrical and Computer Engineering, University of California, San Diego, La Jolla, CA 92093 USA and also with the Qualcomm Inc., San Diego, CA 92121-1714 USA (e-mail: banister@qualcomm.com).

J. R. Zeidler is with the Department of Electrical and Computer Engineering, University of California, San Diego, La Jolla, CA 920933 USA and also with the Space and Naval Warfare Systems Center, San Diego, CA 92152-5001 USA (e-mail: zeidler@ece.ucsd.edu).

Digital Object Identifier 10.1109/JSAC.2003.809721

This paper focuses on a signal processing approach for subspace tracking, using a multimode extension of an algorithm previously introduced for MISO systems [5]. The proposed adaptive approach is similar to subspace tracking algorithms which have been proposed for receive systems [14]–[16]. Given that the principal spatial transmission modes are found and tracked by the algorithm, coding techniques can use the spatial modes in conjunction with the more traditional time/frequency modes in a conventional manner (e.g., see [17]). A specific coding technique is evaluated as an example. In general, however, the algorithm is not specific to any coding technique.

The algorithm maintains a set of transmission weight vectors, where there are fewer weight vectors than transmit antennas. Each weight vector is applied to a corresponding space–time code stream. Gaussian perturbation probing is transmitted to the receiver, which generates feedback indicating a preferred direction for this perturbation to be applied to the current weight vector set. This provides the transmitting unit with a coarse estimate of the gradient of the power delivered to the receiver and is used to update the transmission weight vectors. Gram–Schmidt orthonormalization is performed to maintain the orthonormality of the set of weight vectors. The first-order behavior of this approach is very similar to that described in [14].

The situation which most explicitly benefits from tracking the principal modes is one where some of the available transmission channel subspaces have null or near null response and, hence, deliver no power to the receiver. This condition may occur due to an ill-conditioned channel response. Such a channel response may be due to correlation from closely spaced antennas or may arise from a poor scattering environment, as can occur even with fading independent across all antennas [18]. The results presented in this paper focus on the environment which provides the most explicit subspace tracking gains, one where the channel components are independent and identically distributed (i.i.d.) and there are fewer receive antennas than transmit antennas (e.g., due to size or cost constraints). Simulations with two receive and either four or eight transmit antennas in a fading environment show the effectiveness of this tracking algorithm in attaining the potential gain over blind MIMO transmission, and results for optimal water filling and perfect subspace tracking are generated for comparison.

The paper is organized as follows. Section II motivates the algorithm with an analysis of channel capacity with subspace tracking; Section III describes the operation of the proposed feedback assisted adaptation algorithm; Section IV provides an analysis of the algorithm convergence behavior; Section V provides numerical results for convergence behavior and the realized ergodic capacity with the algorithm; Section VI provides

simulation results for a specific coding example; Section VII provides a discussion of implications for other environments.

II. Capacity Motivation for MIMO Channels

A. Capacity Formulae

The desirability of a subspace tracking technique is demonstrated through simple capacity analysis. The MIMO system will have N_T transmission antennas and N_R receive antennas. The number of tracked transmission subspaces is N_S. It is assumed that $N_T \geq N_R$, $N_S < N_T$. The channel frequency response is assumed to be flat, so there is no temporal intersymbol interference (ISI). The transmission is represented by the $N_T \times 1$ complex vector $\mathbf{t}$, which is zero mean with autocorrelation $\mathbf{\Phi_t}$. The transmission energy per Nyquist symbol is given by E_S

$$\mathbf{\Phi_t} \equiv E(\mathbf{t}\mathbf{t}^H) \tag{1}$$

$$\text{tr}(\mathbf{\Phi_t}) = E_S. \tag{2}$$

With a $N_R \times N_T$ complex channel transfer matrix $\mathbf{H}$ and $N_R \times 1$ zero-mean complex Gaussian noise vector $\mathbf{n}$ with autocorrelation $N_0\mathbf{I}$, the received vector is

$$\mathbf{r} = \mathbf{H}\mathbf{t} + \mathbf{n}. \tag{3}$$

The Shannon capacity of the channel with this modulation is then given by [9]

$$C = \log_2\left(\left|\mathbf{I} + \frac{\mathbf{1}}{N_0}\mathbf{H}\mathbf{\Phi_t}\mathbf{H}^H\right|\right) \tag{4}$$

where $|.|$ denotes the matrix determinant.

The singular value decomposition (SVD) of $\mathbf{H}$ is defined as

$$\underset{N_R\times N_T}{\mathbf{H}} = \underset{N_R\times N_R}{\mathbf{U}}\ \underset{N_R\times N_T}{\mathbf{\Lambda}^{\frac{1}{2}}}\ \underset{N_T\times N_T}{\mathbf{V}^H} \tag{5}$$

with unitary left-singular matrix $\mathbf{U}$, right-singular matrix $\mathbf{V}$ and magnitude sorted singular values $\lambda_k^{1/2}(\mathbf{\Lambda} = \mathbf{\Lambda}^{H/2}\mathbf{\Lambda}^{1/2})$, where $\lambda_1^{1/2}$ is the largest singular value.

With Gaussian signaling (maximizing entropy), we evaluate three values for capacity: 1) water-filling (WF) capacity (the optimal allocation); 2) perfect subspace tracked (PST) transmission capacity; and 3) blind transmission (Bld) capacity (no channel-state knowledge at transmitter). For these transmission strategies, the transmission vector autocorrelations are

$$\mathbf{\Phi_{t(WF)}} = \mathbf{V}(E_0\mathbf{I} - N_0\mathbf{\Lambda^{-1}})^+\mathbf{V}^H \tag{6}$$

$$\mathbf{\Phi_{t(PST)}} = \frac{E_S}{N_S}\mathbf{V}\begin{bmatrix}\underset{N_S\times N_S}{\mathbf{I}} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\mathbf{V}^H \tag{7}$$

$$\mathbf{\Phi_{t(Bld)}} = \frac{E_S}{N_T}\mathbf{I} \tag{8}$$

where $(.)^+$ denotes that the positive entries are retained and the negative entries are zeroed (note: for $\lambda_k = 0$, the entry is $-\infty$ and becomes zero), and for WF E_0 is determined such that

$$\text{tr}(E_0\mathbf{I} - N_0\mathbf{\Lambda^{-1}})^+ = E_S. \tag{9}$$

The associated capacities then become

$$C_{(\text{WF})} = \sum_{k=1}^{\min(N_T,N_R)}\left(\log_2\left(\frac{E_0\lambda_k}{N_0}\right)\right)^+ \tag{10}$$

$$C_{(\text{PST})} = \sum_{k=1}^{\min(N_S,N_R)}\log_2\left(1+\frac{E_S\lambda_k}{N_SN_0}\right) \tag{11}$$

$$C_{(\text{Bld})} = \sum_{k=1}^{\min(N_T,N_R)}\log_2\left(1+\frac{E_S\lambda_k}{N_TN_0}\right). \tag{12}$$

One limiting condition of some interest for WF is large signal to noise ratio $(N_0 \to 0)$. Assuming the channel is of maximal rank (given by N_R, as will be attained in most realistic channel conditions with probability 1), we have

$$\lim_{N_0\to 0}(E_0\mathbf{I} - N_0\mathbf{\Lambda^{-1}})^+ = \frac{E_S}{N_R}\begin{bmatrix}\underset{N_R\times N_R}{\mathbf{I}} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}. \tag{13}$$

B. Discussion

It is immediately clear from (11) and (12) that if $N_S \geq N_R$ and $N_T > N_R$, then a power gain of precisely N_T/N_S is attained for perfect subspace tracking over blind transmission. The gain is attained because transmission of power into null channel modes, where such power cannot be received by the receiver, is avoided. Hence, the received power is increased without sacrificing the multiplexing gain available from the independent spatial modes. This is a reasonable downlink antenna topology since the number of antennas at a data terminal may be restricted by cost and/or space constraints.

The performance of perfect subspace tracking relative to WF is less clear for the general case. However, from (13), we see that the WF and subspace tracking solutions will be approximately the same in the case of large signal to noise ratio. Hence, it is expected that in many practical situations subspace tracking will provide the majority of the achievable adaptation gains. Numerical results comparing the WF, subspace tracked and blind approaches are presented in Section V, after the description of the specific adaptation algorithm.

III. Algorithm Description

A. Objective and Adaptation Cost Function

This section describes the feedback adaptive algorithm for transmission subspace tracking in a FDD system. The objective of the algorithm is to track a $N_T \times N_S$ complex weight matrix $\mathbf{W}$ which maps a $N_S \times 1$ complex vector $\mathbf{s}$ of coded data to the applied signals at the antennas, $\mathbf{t} = \mathbf{W}\mathbf{s}$. The tracking is attempting to extract the principal right-singular subspaces, giving

$$\mathbf{V}^H\mathbf{W}\mathbf{W}^H\mathbf{V} = \begin{bmatrix}\underset{N_S\times N_S}{\mathbf{I}} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}. \tag{14}$$

Define the cost function J

$$J(k) \equiv \|\mathbf{H}\mathbf{W}(k)\|_F^2 = \text{tr}\left(\mathbf{W}^H(k)\mathbf{H}^H\mathbf{H}\mathbf{W}(k)\right) \tag{15}$$

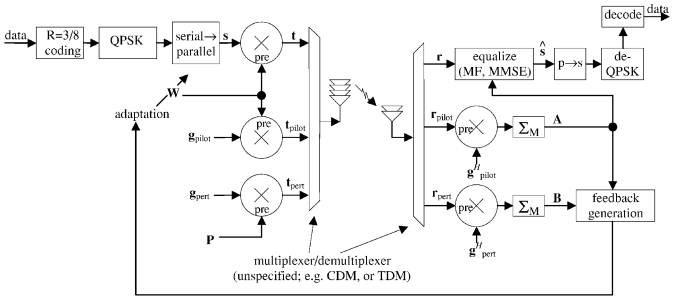


Fig. 1. Diagram of the system with the gradient algorithm.

TABLE I
ADAPTATION ALGORITHM SUMMARY

Initialize:

$\mathbf{W}(0) = \begin{bmatrix} \mathbf{I}_{N_S \times N_S} \\ \mathbf{0} \end{bmatrix}$ (transmitter)

$\mathbf{P}(0)$ = random realization of Gaussian (transmitter)

for k=1:∞

estimate: (receiver)

$\mathbf{A}(k-1) = \mathbf{H}(k-1)\mathbf{W}(k-1)$

$\mathbf{B}(k-1) = \mathbf{H}(k-1)\mathbf{P}(k-1)$

$d(k-1) = \text{sign}\left(\text{Re}\left(\sum_{m=1}^{N_R}\sum_{n=1}^{N_S} a_{m,n}^*(k-1)\cdot b_{m,n}(k-1)\right)\right)$ (receiver)

$\mathbf{W}(k) = \text{G}(\mathbf{W}(k-1) + \beta d(k-1)\mathbf{P}(k-1))$ (transmitter)

$\mathbf{P}(k)$ = new random realization of Gaussian (transmitter)

end

where $||.||_F$ is the Frobenius norm, and constrain the weight matrix so that

$$\mathbf{W}^H\mathbf{W} = \underset{N_S \times N_S}{\mathbf{I}}. \tag{16}$$

It is shown in [14] that the maximization of J subject to the constraint (16) accomplishes the desired subspace tracking.

B. Algorithm Operation

The algorithm objective will be accomplished with a feedback stochastic gradient algorithm adapted from [5] for the multimodal adaptation required for subspace tracking. A block diagram of the system is shown in Fig. 1, and an algorithm summary is provided in Table I. The system utilizes N_S parallel space–time coded transmission streams, where any STC technique can in principle be applied. Each of the transmission streams is transmitted with a different antenna weighting vector, according to the column-orthogonal $N_T \times N_S$ weight matrix $\mathbf{W}$, where this matrix is tracked by the feedback algorithm such that its columns span the principal right-singular subspaces of the channel gain matrix $\mathbf{H}$. The algorithm strategy is to transmit a probing perturbation signal, with the receiver generating feedback selecting the preferred sign to apply to the perturbation in an update to $\mathbf{W}$. It will be shown that this results in the desired subspace tracking. For ease of presentation, the description uses discrete time sampling; Nyquist pulse shaping and a temporal ISI free channel are assumed.

The transmission is comprised of three components, distinguishable at the receiver through utilization of any standard multiplexing technique (e.g., code or time-division multiplex: CDM or TDM): coded data modulation, a pilot transmission (for channel estimation at the receiver), and a perturbation probing transmission (for feedback generation). The adaptation occurs through probing and feedback every M Nyquist symbols; measurement and feedback latency is ignored in this presentation. The data modulation is given by the $N_S \times 1$ code stream vector $\mathbf{s}(i)$, which has an autocorrelation of $\mathbf{I}$. The weighting matrix translates this vector to the transmit antennas

$$\mathbf{t}(i) = \sqrt{\frac{E_S}{N_S}}\mathbf{W}\left(\left\lfloor \frac{i}{M} \right\rfloor\right)\mathbf{s}(i). \tag{17}$$

The pilot modulation is all $+1$'s multiplied by a $N_S \times 1$ vector spreading cover $\mathbf{g}_{\text{pilot}}(i)$, which is transmitted with the same weighting matrix as the data so that the channel experienced by the data transmission can be estimated at the receiver

$$\mathbf{t}_{\text{pilot}}(i) = \sqrt{\frac{E_{\text{pilot}}}{N_S}}\mathbf{W}\left(\left\lfloor \frac{i}{M} \right\rfloor\right)\mathbf{g}_{\text{pilot}}(i). \tag{18}$$

The perturbation modulation is similar to that of the pilot: all $+1$'s multiplied by a $N_S \times 1$ vector spreading cover $\mathbf{g}_{\text{pert}}(i)$, which is translated to the transmit antennas with a $N_T \times N_S$ test perturbation matrix $\mathbf{P}$. For each perturbation probing period the perturbation matrix $\mathbf{P}$ is randomly generated with i.i.d. zero-mean complex Gaussian elements with variance of two.

$$\mathbf{t}_{\text{pert}}(i) = \sqrt{\frac{E_{\text{pert}}}{2N_TN_S}}\mathbf{P}\left(\left\lfloor \frac{i}{M} \right\rfloor\right)\mathbf{g}_{\text{pert}}(i). \tag{19}$$

The spreading cover vectors are generated as uncorrelated so that the pilot and perturbation transmissions can be recovered at the receiver

$$\sum_{i=kM}^{kM+M-1}\mathbf{g}_{\text{pilot}}(i)\mathbf{g}_{\text{pilot}}^H(i) = \sum_{i-kM}^{kM+M-1}\mathbf{g}_{\text{pert}}(i)\mathbf{g}_{\text{pert}}^H(i) = M\mathbf{I}. \tag{20}$$

The data, pilot, and perturbation signals are all transmitted from the same antennas. The multiplexing technique used to distinguish these three signals at the receiver is important to the final system design, but it is not relevant to this general introduction to the algorithm and will not be considered here. In this paper, it is assumed that pilot and perturbation signals can be perfectly recovered by the receiver and their multiplexing requires an insignificant bandwidth, so that the capacity loss incurred by the multiplexing is negligible. This requires that the forward link bandwidth is much greater than the feedback rate. Together, the pilot and perturbation transmissions consume $2N_S$ orthogonal time/frequency bases out of M available,[1] where M is the ratio of frequency bandwidth to adaptation rate. Hence, the assumption is $2N_S \ll M$.

[1]The pilot and perturbation bandwidth utilization can be incorporated into the results as a bit rate loss factor $(M - 2N_S)/M$ (e.g., for a 1-MHz bandwidth and 10 kb/s feedback the bit rate loss is 0.98). This still excludes the power utilization.

The $N_R \times N_T$ channel gain matrix $\mathbf{H}$ is assumed constant over the perturbation measurement and feedback interval, so that the composite channels formed by the combination of the physical channel and the transmission weight and perturbation, as seen by the receiver, are as follows:

$$\underset{N_R \times N_S}{\mathbf{A}}(k) \equiv \mathbf{H}(k)\mathbf{W}(k) \tag{21}$$

$$\underset{N_R \times N_S}{\mathbf{B}}(k) \equiv \mathbf{H}(k)\mathbf{P}(k). \tag{22}$$

Then, the $N_R \times 1$ received data signal vector $\mathbf{r}$ is

$$\mathbf{r}(i) = \mathbf{H}\left(\left\lfloor \frac{i}{M} \right\rfloor\right)\mathbf{t}(i) + \mathbf{n}(i) = \sqrt{\frac{E_S}{N_S}}\mathbf{A}\left(\left\lfloor \frac{i}{M} \right\rfloor\right)\mathbf{s}(i) + \mathbf{n}(i). \tag{23}$$

With noise excluded, the pilot and perturbation are recovered perfectly at the receiver as

$$\frac{\mathbf{1}}{M}\sum_{i=kM}^{kM+M-1}\mathbf{H}(k)\,\mathbf{t}_{\text{pilot}}(i)\,\mathbf{g}_{\text{pilot}}^{H}(i) = \sqrt{\frac{E_{\text{pilot}}}{N_S}}\mathbf{A}(k) \tag{24}$$

$$\frac{1}{M}\sum_{i=kM}^{kM+M-1}\mathbf{H}(k)\,\mathbf{t}_{\text{pert}}(i)\,\mathbf{g}_{\text{pert}}^{H}(i) = \sqrt{\frac{E_{\text{pert}}}{N_T N_S}}\mathbf{B}(k). \tag{25}$$

Feedback is generated from the receiver using the pilot and perturbation transmissions. The feedback selects which sign-direction is preferable, in terms of maximizing receive power, as an update of $\mathbf{W}$ by $\mathbf{P}$. That is, the binary feedback decision d is determined as

$$d(k) = \text{sign}\left(\lim_{\Delta \to +0}\left(\|\mathbf{A}(k) + \Delta\mathbf{B}(k)\|_F^2 - \|\mathbf{A}(k) - \Delta\mathbf{B}(k)\|_F^2\right)\right) = \text{sign}\left(\text{Re}\left(\sum_{m=1}^{N_R}\sum_{n=1}^{N_S} a_{m,n}{}^{*}(k)\cdot b_{m,n}(k)\right)\right). \tag{26}$$

The decision defined by (26) is binary encoded and provided as feedback from the receiver to the transmitter. Since the magnitude (in expectation) of the elements of $\mathbf{P}$ are constrained (having variance two), the adaptation parameter β is introduced to capture the adaptation rate, with a larger β giving faster but noisier adaptation. Using the parameter β to define the update step size the weight matrix update at the transmitter is given by

$$\mathbf{W}'(k+1) = \mathbf{W}(k) + \beta \cdot d(k) \cdot \mathbf{P}(k) \tag{27}$$

$$\mathbf{W}(k+1) = \text{G}\left(\mathbf{W}'(k+1)\right) \tag{28}$$

where $\mathbf{W}'$ denotes an intermediate computation prior to orthonormalization, and the matrix function $\text{G}(.)$ returns the Gram–Schmidt column orthonormalization of the input matrix; e.g., $\text{G}(.)$ returns the unitary $\mathbf{Q}$ matrix from a Gram–Schmidt QR factorization. Other QR factorizations could be used, but the Gram–Schmidt is assumed throughout and will be seen to have some desirable properties. This orthonormalization ensures that the transmission streams are orthogonal at the transmit antennas at all times, so that the weight vectors (columns of $\mathbf{W}$ cannot collapse into overlapping spaces and N_S transmission modes are indeed stimulated, satisfying (16).

IV. Adaptation Analysis

A. Definitions

In this section, convergence will be considered with $\mathbf{H}$ static and nonrandom (or taken as given). A tilde will be used throughout to indicate projection into the right-singular spaces of $\mathbf{H}$. Hence, defining $\mathbf{V}$ as in (5)

$$\tilde{\mathbf{W}} = \mathbf{V}^H\mathbf{W} \tag{29}$$

$$\tilde{\mathbf{E}} = \mathbf{V}^H\mathbf{E}. \tag{30}$$

For brevity, it will be convenient to define $b(k)$ as

$$b(k) \equiv \frac{\beta\sqrt{\frac{2}{\pi}}}{\|\mathbf{H}^H\mathbf{H}\mathbf{W}(k)\|_F} = \beta\sqrt{\frac{2}{\pi \cdot \text{tr}\left(\tilde{\mathbf{W}}(k)^H\mathbf{\Lambda}^2\tilde{\mathbf{W}}(k)\right)}}. \tag{31}$$

B. Cost Function and Gradient Extraction

This algorithm is operating to maximize the performance metric J defined above, and can be considered to be a steepest ascent implementation. The gradient of J with respect to $\mathbf{W}$ is

$$\nabla_{\text{W}} J(k) = 2\mathbf{H}^H\mathbf{H}\mathbf{W}(k). \tag{32}$$

Extending a result from [5], for $\mathbf{P}$ comprised of i.i.d. random complex Gaussians with variance twice unity, the expected value of the weight change prior to orthonormalization is the scaled Frobenius normalized gradient of J with respect to $\mathbf{W}$ (see Appendix A).

$$\text{E}\left(d(k)\cdot\mathbf{P}(k)\right) = \sqrt{\frac{2}{\pi}}\frac{\mathbf{H}^H\mathbf{H}\mathbf{W}(k)}{\|\mathbf{H}^H\mathbf{H}\mathbf{W}(k)\|_F}. \tag{33}$$

Then, with negligible estimation error in the receiver and reliable feedback, the weight matrix update prior to orthonormalization is

$$\mathbf{W}'(k+1) = \mathbf{W}(k) + b(k)\mathbf{H}^H\mathbf{H}\mathbf{W}(k) + \beta\mathbf{E}(k) \tag{34}$$

where $\mathbf{E}$ is a zero-mean error matrix. The elements of $\mathbf{E}$ may be correlated due to the normalized gradient which is extracted from $\pm\mathbf{P}$ to leave $\mathbf{E}$. The right-singular projection representation in the update prior to orthonormalization is then a diagonal modification plus noise from $\mathbf{E}$

$$\tilde{\mathbf{W}}'(k+1) = \left(\mathbf{I} + b(k)\mathbf{\Lambda}\right)\tilde{\mathbf{W}}(k) + \beta\tilde{\mathbf{E}}(k). \tag{35}$$

It was observed in [14] that a gradient update with a form similar[2] to (34) and (35) with orthonormalization (28) and $\mathbf{E} = \mathbf{0}$ has the same formulation as the orthogonal iteration method of eigendecomposition [19]. Hence, a gradient update with the form of (35) will cause the weight matrix to converge to the principal right-singular subspaces of $\mathbf{H}$.

C. Adaptation Update Moments

The first and second moments of the updated right-singular projected weight matrix prior to orthonormalization are

[2]The difference is that here the gradient update is Frobenius normalized. In contrast, in [14] there is no normalization, so that the update matrix multiplying $\mathbf{W}$ is invariant.

straightforward to compute, applying (59), (61), and (62) from Appendix A

$$\mathbf{E}\left(\tilde{\mathbf{W}}'(k+1)|\tilde{\mathbf{W}}(k)\right) = (\mathbf{I}+b(k)\mathbf{\Lambda})\,\tilde{\mathbf{W}}(k) \tag{36}$$

$$\mathrm{E}\left(\tilde{\mathbf{W}}'^{H}(k+1)\tilde{\mathbf{W}}'(k+1)|\tilde{\mathbf{W}}(k)\right) = \tilde{\mathbf{W}}^{H}(k)\tilde{\mathbf{W}}(k)+2b(k)\tilde{\mathbf{W}}^{H}(k)\mathbf{\Lambda}\tilde{\mathbf{W}}(k)+2\beta^{2}N_{T}\mathbf{I} \tag{37}$$

$$\mathrm{E}\left(\tilde{\mathbf{W}}'(k+1)\tilde{\mathbf{W}}'^{H}(k+1)|\tilde{\mathbf{W}}(k)\right) = \tilde{\mathbf{W}}(k)\tilde{\mathbf{W}}^{H}(k)+b(k)\left(\mathbf{\Lambda}\tilde{\mathbf{W}}(k)\;\tilde{\mathbf{W}}^{H}(k) + \tilde{\mathbf{W}}(k)\;\tilde{\mathbf{W}}^{H}(k)\mathbf{\Lambda}\right)+2\beta^{2}N_{S}\mathbf{I}. \tag{38}$$

D. Convergence in Static Channel With Noiseless Update

In considering the convergence properties, we approximate perfect gradient estimation by setting $\mathbf{E}=\mathbf{0}$, which provides some insight to the adaptation process. With the removal of the error term, the update prior to orthonormalization (34) becomes a premultiplication of the weight matrix

$$\mathbf{W}(k+1)=\mathbf{V}\mathrm{G}\left((\mathbf{I}+b(k)\mathbf{\Lambda})\,\mathbf{V}^{H}\mathbf{W}(k)\right). \tag{39}$$

Because the Frobenius normalizing denominator in $b(k)$ is dependent on the state of the weights (35), this formula must be iteratively computed in order to completely characterize the convergence. The convergence path of the weight matrix with the assumption of zero adaptation error ($\mathbf{E}=\mathbf{0}$) is derived in Appendix B[3]

$$\mathbf{W}(k)=\mathbf{V}\cdot\mathrm{G}\left(\left[\prod_{n=0}^{k-1}\left(\mathbf{I}+\beta\sqrt{\frac{2}{\pi}}\frac{\mathbf{\Lambda}}{||\mathbf{\Lambda}\mathbf{V}^{H}\mathbf{W}(n)||_{F}}\right)\right]\mathbf{V}^{H}\mathbf{W}(0)\right). \tag{40}$$

Noting that $||\mathbf{\Lambda}\tilde{\mathbf{W}}(k)||_{F}\leq\sqrt{\mathrm{tr}(\mathbf{\Lambda}^{2})}$ then directly from (40) the convergence of $\tilde{\mathbf{w}}$ to the i^{th} singular space over lesser spaces goes as follows. This result is similar to the conclusion of [14] except that the variable gradient scaling (Frobenius normalization) requires the use of a lower bound

$$\frac{|\tilde{w}_{i,i}(k)|}{|\tilde{w}_{i,j}(k)|}\propto\prod_{n=0}^{k-1}\left(\frac{1+b(n)\lambda_{i}}{1+b(n)\lambda_{j}}\right) \geq\left(\frac{1+\beta\mathrm{tr}^{-\frac{1}{2}}(\mathbf{\Lambda}^{2})\sqrt{\frac{2}{\pi}}\lambda_{i}}{1+\beta\mathrm{tr}^{-\frac{1}{2}}(\mathbf{\Lambda}^{2})\sqrt{\frac{2}{\pi}}\lambda_{j}}\right)^{k}\qquad i<j. \tag{41}$$

This shows that the first weight column vector converges to the first singular space, and the second weight vector converges to the second singular space (since the Gram–Schmidt projects it away from the first principal space), etc. Thus, upon convergence this algorithm not only tracks the desired subspace, but extracts the sorted right-singular spaces. This has two major implications when the algorithm has converged or nearly converged: 1) there is a reduced need for a spatial equalizer at the receiver to compensate for code stream crosstalk, and 2) the elements of the space–time code stream vector $\mathbf{s}$, s_k, attain reducing degrees strength and reliability for increasing k.

[3]Matrix factorization is herein defined from the left side: $\prod_{i=1}^{k}\mathbf{A}(i)\equiv\mathbf{A}(k)\prod_{i=1}^{k-1}\mathbf{A}(i)$

V. Simulation Study of Capacity

A. Simulation Environment

In order to demonstrate the general (coding independent) properties of the applied algorithm, Monte Carlo simulation is performed and several metrics are extracted. The algorithm is simulated as described above with $N_S=N_R=2$ and both $N_T=4$ and $N_T=8$. Channel estimation at the receiver is considered to be perfect for purposes of generating the feedback and computing capacity, and pilot/perturbation multiplexing bandwidth utilization is not considered. The feedback is implemented without decision errors, and β was varied to find its best value. The channel model is independent Raleigh-flat fading with time correlation given by Jakes model, and the Doppler frequency F_D is configured relative to the feedback rate F_{FB}, so that both are captured in the ratio F_{FB}/F_D. The mean channel gain is captured in $|h|_2$

$$|h|_{2}\equiv\sqrt{E|h_{i,j}|^{2}}. \tag{42}$$

The mean gradient cost metric is evaluated as

$$\bar{J}=E\left(\frac{J(k)}{J_{\mathrm{opt}}(k)}\right) \tag{43}$$

where $J(k)$ is the time varying value of (15) and $J_{\mathrm{opt}}(k)$ is the time varying value of (15) for perfect subspace tracking. In addition, ergodic capacity values in units of bits/second/Hertz were evaluated as the mean mutual information (for Gaussian signaling at the transmitter) between the transmitted waveform and the received waveform for a transmission energy per Nyquist symbol of E_S (energy summed over all N_S code streams), according to (44). Appropriate reformulations were applied for each system example (e.g., perfect weights, blind)

$$\bar{C}=\mathbf{E}\left(\log_{2}\left(\left|\mathbf{I}+\frac{E_{S}}{N_{S}N_{0}}\mathbf{H}(k)\mathbf{W}(k)\mathbf{W}^{H}(k)\mathbf{H}^{H}(k)\right|\right)\right). \tag{44}$$

These ergodic capacities were evaluated for single-input–single-output (SISO) and single-input–multiple-output (SIMO) benchmark conditions and for several configurations with MIMO channels with $N_R=2$ and with either $N_T=4$ or $N_T=8$. The test case labels and descriptions are given in Table II.

B. Discussion

The results of the sweep of β for different frequency ratios F_{FB}/F_D are shown in Fig. 2 and Fig. 3. As expected, higher relative feedback rates allow the use of a smaller adaptation parameter β. The optimal values of β, those which maximize $\overline{J}$ from these simulations are used for all subsequent simulations.

The average convergence transient behavior is shown in Fig. 4 and Fig. 5, where the mean absolute values of the entries of the matrix $\mathbf{A}^{H}\mathbf{A}$ are shown for PST (weights are exactly the principal right-singular vectors) and for the gradient feedback (GFB) algorithm with $F_{FB}/F_D=1000$ and $\beta=0.02$. With

TABLE II
SIMULATION LABELING AND DESCRIPTIONS

label	description
SISO AWGN	SISO system, N_T=N_R=1, with no fading (AWGN channel, baseline reference)
SISO	N_T=N_R=1
SIMO	N_T=1, N_R=2
WF	N_T=4 or 8, N_R=2. MIMO with perfect water filling
PST	N_T=4 or 8, N_R=2. MIMO with perfect subspace tracking
Bld	N_T=4 or 8, N_R=2. MIMO with blind transmission into all transmit subspaces
GFB	N_T=4 or 8, N_R=2. MIMO with gradient feedback algorithm subspace tracking

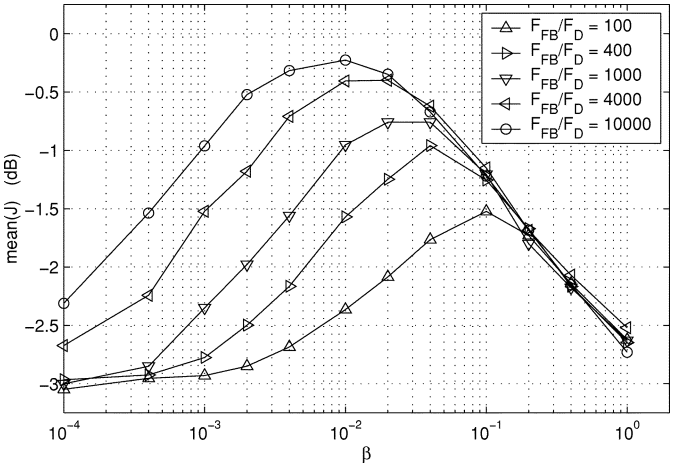


Fig. 2. Cost function, mean(J), for the gradient update, sweeping β, and the feedback frequency to Doppler frequency ratio, $N_R = 2$, $N_T = 4$.

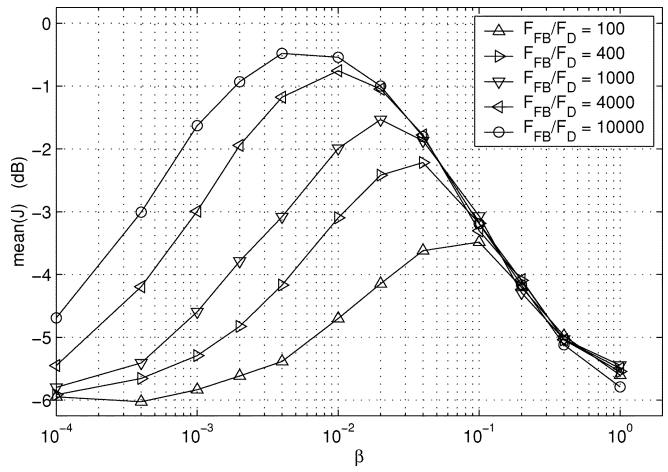


Fig. 3. Cost function, mean(J), for the gradient update, sweeping β, and the feedback frequency to Doppler frequency ratio, $N_R = 2$, $N_T = 8$.

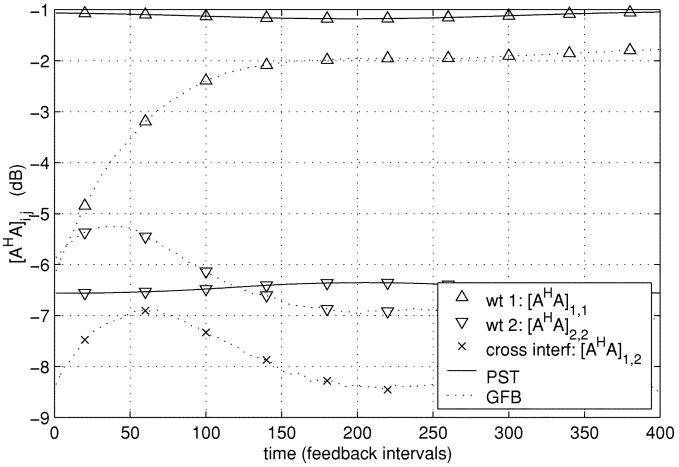


Fig. 4. Convergence transient of power and cross talk. y axis is $10 \cdot \log_{10}(\mathrm{E}|[\mathbf{A}^H\mathbf{A}]_{i,j}|/\mathrm{E}(\mathrm{tr}(\mathbf{H}^H\mathbf{H})))$. For perfect subspace weights or gradient feedback tracked weights ($F_{FB}/F_D = 1000, \beta = 0.02$). $N_R = 2$, $N_T = 4$.

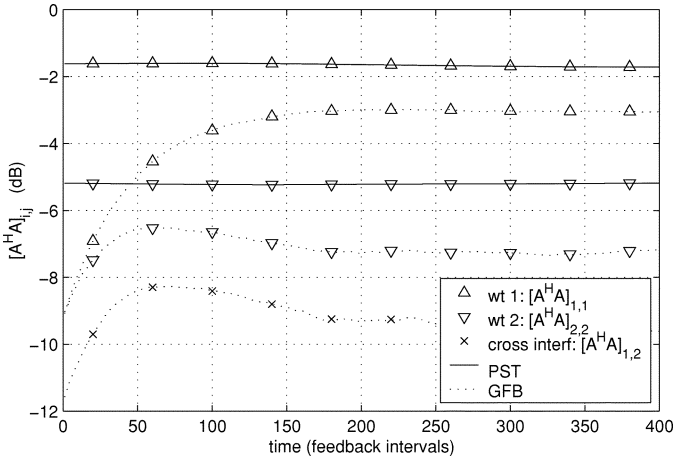


Fig. 5. Convergence transient of power and cross talk. y axis is $10 \cdot \log_{10}(\mathrm{E}|[\mathbf{A}^H\mathbf{A}]_{i,j}|/\mathrm{E}(\mathrm{tr}(\mathbf{H}^H\mathbf{H})))$. For perfect subspace weights or gradient feedback tracked weights ($F_{FB}/F_D = 1000, \beta = 0.02$). $N_R = 2$, $N_T = 8$.

PST, the mean result is constant, while GFB shows a convergence time on the order of 100 feedback intervals. It is clear that with GFB the first weight tracks toward the dominant right-singular space, while the second weight tracks toward the second right-singular space. The cross term $[\mathbf{A}^H\mathbf{A}]_{1,2}$ gives the cross-interference between the two received code streams; e.g., the signal to cross-interference power ratio experienced by received code #1 with a matched filter (no equalization) scheme at the receiver is

$$\mathrm{SIR}(s_1) = \frac{\|\mathbf{a}_1\|_2^4}{\left|\mathbf{a}_1^H\mathbf{a}_2\right|^2} = \frac{[\mathbf{A}^H\mathbf{A}]_{1,1}^2}{\left|[\mathbf{A}^H\mathbf{A}]_{1,2}\right|^2}. \tag{45}$$

Example, applying (45) to Fig. 5 ($N_T = 8$) provides a steady state $\mathrm{SIR}(s_1) \cong +13.2$ dB and $\mathrm{SIR}(s_2) \cong +4.8$ dB for GFB

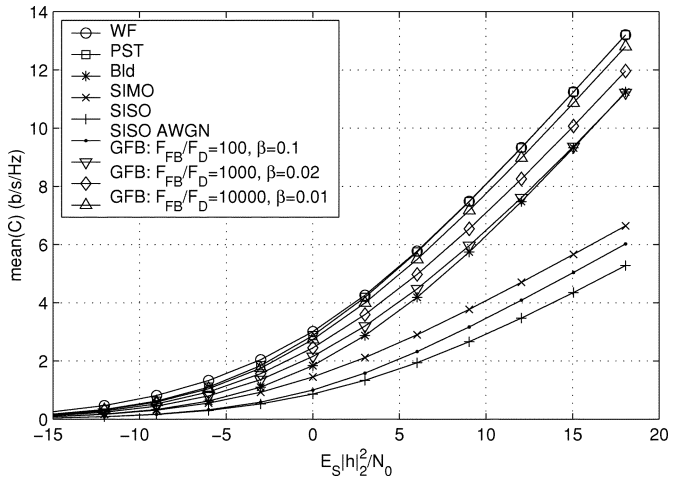


Fig. 6. Ergodic capacity (b/s/Hz) versus energy per Nyquist symbol, $N_T = 4$. See Table II.

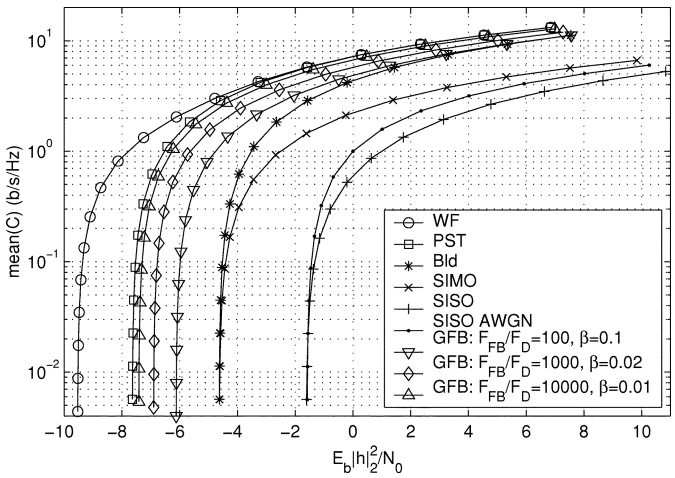


Fig. 7. Ergodic capacity (b/s/Hz) versus energy per bit, $N_T = 4$. See Table II.

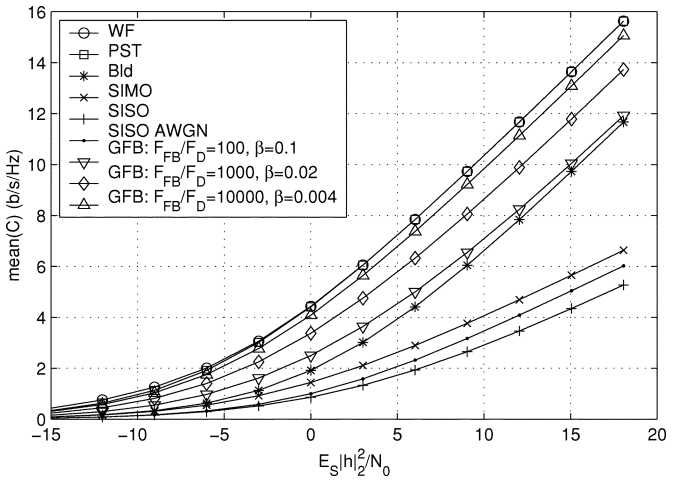


Fig. 8. Ergodic capacity (b/s/Hz) versus energy per Nyquist symbol, $N_T = 8$. See Table II.

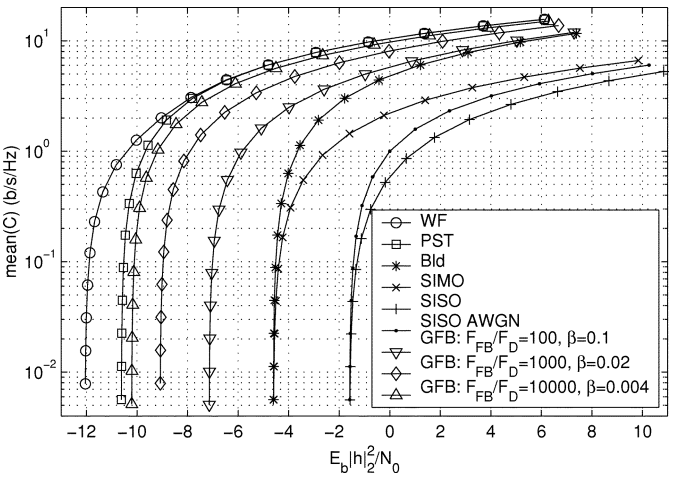


Fig. 9. Ergodic capacity (b/s/Hz) versus energy per bit, $N_T = 8$. See Table II.

(the cross interference is zero for PST). Hence, the cross interference is relatively small, and good performance could be expected with either no spatial equalizer [i.e., a simple matched-filter (MF) receiver] or with a relatively simple linear [e.g., minimum mean-square error (MMSE)] spatial equalizer. More complicated equalization approaches such as successive cancellation (e.g., "BLAST" [2]), or iterative turbo equalization (e.g., [20] and [21]) might not be expected to provide significant further gains in this case. MF and MMSE results are presented in Section VI.

The key capacity results are given in Fig. 6, Fig. 7 $(N_T = 4)$ and Fig. 8, Fig. 9 $(N_T = 8)$. These show the gains available from the use of STC with and without subspace tracking. The x axis transmitted energy per bit or energy per Nyquist symbol are shown for mean single channel gain $|h|_2$, so that the gain from directing more of the transmitted energy toward the receiver is visible. The capacity curves can be divided into two general regions: the "bandwidth limited" region highlighted by plotting capacity versus energy per Nyquist symbol (Fig. 6, Fig. 8), and the "power limited" region highlighted by showing capacity versus energy per bit (Fig. 7, Fig. 9). In the bandwidth limited region of the curve, the primary limitation to channel capacity is the number of available orthogonal bases for transmission, and the $N_S = 2$ STC (whether subspace tracked or blind) doubles the bits per power octave slope versus non space–time coded transmission. In the power limited region, the primary limitation is the received power and extra orthogonal transmission bases provide only small gain. The limiting values are summarized in Table III. With infinitesimal data rate one attains infinite time diversity for each code symbol and the fading SISO limit approaches the well known -1.59 dB $\mathrm{E_b/N_0}$ additive white Gaussian noise (AWGN) limit.

In Fig. 6–Fig. 9 the PST case performs 3.01 dB $(N_T = 4)$ or 6.02 dB $(N_T = 8)$ better than the blind transmission approach. In all cases, the gradient feedback algorithm provides an ergodic capacity with significant gain over the blind transmission approach. With $F_{FB}/F_D = 10\,000$ the GFB subspace tracking almost performs, as well as PST. Even with feedback rates as low as $F_{FB}/F_D = 100$ the GFB outperforms blind transmission, except at very high signal to noise ratios. It is worth noting in comparing these figures to Fig. 2 and Fig. 3 that the mean metric $\overline{J}$ is a very good predictor of the performance of GFB relative to PST in the power limited region, where the power transfer (captured in $\overline{J}$) becomes the most critical aspect.

TABLE III
COMPARISONS IN THE LIMITS

		SISO	SIMO	MIMO Bld	MIMO GFB F_{FB}/F_D= 1e2	MIMO GFB F_{FB}/F_D= 1e3	MIMO GFB F_{FB}/F_D= 1e4	MIMO PST	MIMO WF
N_T=4	min $E_b\cdot\|h\|_2^2/N_0$ (dB) (C→0)	-1.59	-4.60	-4.60	-6.1	-6.8	-7.4	-7.61	-9.56
	Cap. slope ((b/s/Hz)/3.01dB) (E_S→∞)	1	1	2	2	2	2	2	2
N_T=8	min $E_b\cdot\|h\|_2^2/N_0$ (dB) (C→0)	-1.59	-4.60	-4.60	-7.2	-9.2	-10.3	-10.62	-12.03
	Cap. slope ((b/s/Hz)/3.01dB) (E_S →∞)	1	1	2	2	2	2	2	2

Together, the capacity plots show the gains available by subspace tracking. Many systems will be interference limited, and additional energy per bit is extra interference to other users. This motivates operation in or near the "power limited" region of the curve. In practice, this is a motivation to avoid higher order modulations [i.e., quadrature amplitude modulation (QAM)], since the large E_b/N_0 required for reliable performance increases the system wide interference. Following this reasoning, it could be desirable to use rate 1/2 or 1/4 codes with quaternary phase-shift keying (QPSK) or 8-PSK modulation. A QPSK rate 1/2 code in a SISO AWGN channel gives 1 b/s/Hz and requires $E_b/N_0 = 0$ dB for reliable reception, which is an excess bit energy of $+1.59$ dB. Applying a rate 1/2 code with $N_S = 2$ spatial channels gives 2 b/s/Hz, and the figures show clear gains for GFB in this region versus blind transmission. It is also interesting to note that in this region the gains for blind STC are only 2.0 dB $(N_T = 4)$ or 2.3 dB $(N_T = 8)$ over standard coding with a single transmit antenna and two receive antennas (SIMO), and in the limit for low rate the extra transmit antennas provide no gain over the SIMO case. Hence, we see the motivation for adopting some form of transmission adaptation scheme, with subspace tracking forming perhaps the simplest general class of appropriate adaptation.

VI. Simulation Study With Convolutional Coding

A. Environment

In order to better illustrate the utility of the algorithm with a realistic coding scheme, simulations have been performed utilizing a simple example of convolutional coding and QPSK modulation. For comparison, a blind STC scheme is implemented with diversity coding providing the uncorrelated transmit vector assumed in (8). The code is rate 1/2, constraint length 9 with octal generator polynomials (753, 561) and free distance 12. The coding was implemented as a block convolutional code with 24 information bits and 8 tail bits (zeros), so that the true code rate is 3/8. With SISO QPSK modulation, this gives a data rate of 0.75 b/s/Hz. With $N_S = 2$ in the MIMO case two QPSK symbols are transmitted at a time and the rate is 1.5 b/s/Hz. The simulations were run with uncorrelated Gaussian entries of $\mathbf{H}$ with Jakes temporal correlation, with block static frequency flat fading constant over each frame (i.e., frame duration $\ll \mathbf{1}/\mathrm{F_D}$).

The coding for the adaptive scheme performs a serial to parallel operation on pairs of QPSK symbols to provide two symbols at a time as the code stream $\mathbf{s}$. No additional special structure on the STC is applied. This simple approach is similar to that of [17], except for the adaptive weighting of the transmitted code streams.

For comparison, blind transmission is generated using diversity STC. In the $N_T = 4$ case, the scheme of [1] is used on a first pair of QPSK symbols to generate the transmission for two antennas and two time intervals, and the scheme is applied again for the second pair of symbols to generate transmission for the remaining two antennas. Similarly, in the $N_T = 8$ case, the scheme of [22] is used on a first quartet of QPSK symbols to generate the transmission for four antennas and four time intervals, and the scheme is applied again for the second quartet of symbols to generate transmission for the remaining four antennas. This is illustrated in (46) and (47), shown at the bottom of the page, where s_n are the serial QPSK symbols directly from the encoder and QPSK modulator.

$$\begin{bmatrix} \mathbf{s}^T(i) \\ \mathbf{s}^T(i+1) \end{bmatrix}^T = \begin{bmatrix} s_{4i} & s_{4i+1} & s_{4i+2} & s_{4i+3} \\ -{s_{4i+1}}^* & {s_{4i}}^* & -{s_{4i+3}}^* & {s_{4i+2}}^* \end{bmatrix}^T \text{(blind, } N_T = 4) \tag{46}$$

$$\begin{bmatrix} \mathbf{s}^T(i) \\ \mathbf{s}^T(i+1) \\ \mathbf{s}^T(1+2) \\ \mathbf{s}^T(i+3) \end{bmatrix}^T = \begin{bmatrix} s_{8i} & s_{8i+1} & s_{8i+2} & s_{8i+3} & s_{8i+4} & s_{8i+5} & s_{8i+6} & s_{8i+7} \\ {s_{8i+1}}^* & -{s_{8i}}^* & {s_{8i+3}}^* & -{s_{8i+2}}^* & {s_{8i+5}}^* & -{s_{8i+4}}^* & {s_{8i+7}}^* & {s_{8i+6}}^* \\ s_{8i+2} & -s_{8i+3} & -s_{8i} & s_{8i+1} & s_{8i+6} & -s_{8i+7} & -s_{8i+4} & s_{8i+5} \\ {s_{8i+3}}^* & {s_{8i+2}}^* & -{s_{8i+1}}^* & -{s_{8i}}^* & {s_{8i+7}}^* & {s_{8i+6}}^* & -{s_{8i+5}}^* & -{s_{8i+4}}^* \end{bmatrix}^T \text{(blind, } N_T = 8) \tag{47}$$

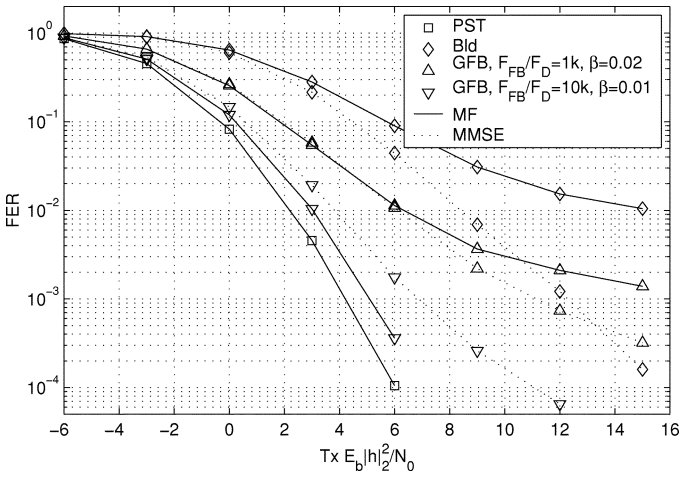


Fig. 10. Frame-error rate $N_T = 4$. See key Table II.

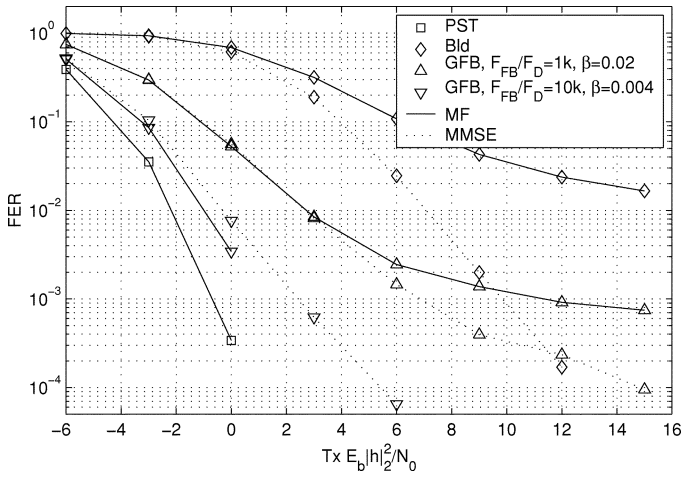


Fig. 11. Frame-error rate, $N_T = 8$. See key Table II.

In both blind transmission schemes the QPSK symbol rate is again two independent symbols per Nyquist interval. The use of four or eight transmit antennas with the diversity STC provides a diversity enhancement over simply using two transmit antennas, so that reliability is improved but the capacity enhancement is small.

The performance was evaluated with either a MF or a MMSE linear symbol estimator providing the symbols to the decoder. For the blind cases, the MF and MMSE approaches are straightforward extensions of those described in [1] and [22]; MMSE requires combining over the two $(N_T = 4)$ or four $(N_T = 8)$ received time samples of the STC. For the adaptive transmission systems (PST and GFB), the symbol estimates which are QPSK demodulated and Viterbi decoded are generated as (time index omitted for $\mathbf{A}$)

$$\hat{\mathbf{s}}(i) = \mathbf{A}^H \mathbf{r}(i) \qquad \text{(PST,GFB:MF)} \tag{48}$$

$$\hat{\mathbf{s}}(i) = \mathbf{A}^H \left(\frac{E_S}{N_S} \mathbf{A}\mathbf{A}^H + N_0 \mathbf{I} \right)^{-1} \mathbf{r}(i) \qquad \text{(PST, GFB: MMSE)}. \tag{49}$$

B. Discussion

The FER simulation results are seen in Fig. 10 $(N_T = 4)$ and Fig. 11 $(N_T = 8)$. We see that the MMSE receiver provides much better performance than the MF receiver in the blind transmission cases: this transmission scheme suffers from code stream spatial crosstalk. The use of extra antennas (more than the two required to support the coding and data rate) has attained a diversity gain, but the lack of power gain (available through subspace tracking) and the code-stream cross talk degradation with suboptimal linear equalization have hindered the performance. However, this is not true for the GFB cases; with the GFB approach the transmission adaptation tends to extract not just the dominant subspace but to separate the individual right-singular spaces, which minimizes code stream crosstalk at the receiver. Hence, in the slower GFB tracking cases $(F_{FB}/F_D = 1000)$, where the right singular spaces are not entirely separated, the MMSE performance is only slightly better than with the MF. However, with $F_{FB}/F_D = 10\,000$ the MF receiver actually outperforms the MMSE receiver. Here, the symbol weighting distortion from MMSE degrades the performance relative to MF more than the minimization of crosstalk from MMSE improves the performance.

The simulation results show the gains of the tracking algorithm and illustrate that the application of this algorithm may in fact simplify the implementation of the receiver, since the need for equalization is minimized.

VII. Closing Discussion

The numerical results presented have focused on a channel condition with explicitly limited rank. In systems with equal numbers of antennas at the receiver and the transmitter the channel rank is not explicitly limited and will generally be full rank. In such situations the gain from a transmitter subspace tracking algorithm is less clear. The tradeoff of blind transmission versus subspace tracked transmission with $N_S < N_T$ becomes a matter of optimizing the tradeoff of "bandwidth" (orthogonal bases per second) versus the delivered power. That is, by concentrating the transmitted power into the right-singular spaces with the largest gain, the subspace tracking algorithm can increase the received SNR while losing effective coding bandwidth. This provides gains for systems operating in the power-limited region or with large spatial correlation.

With the Gram–Schmidt orthonormalization step the distinct right-singular vectors of the channel are extracted. This was shown in Section VI to provide "preequalization" and reduce the need for equalization at the receiver. In an unusual case of $N_S = N_T$ this vector extraction is achieved without extracting a subspace. With equal power allocation in $\mathbf{s}$ the result would be no increase in ergodic capacity over blind STC, but a practical performance increase due to the preequalization.

VIII. Conclusion

The benefits of transmission tracking of dominant channel subspaces for communications in low rank MIMO channel conditions have been described, particularly for the specific case where the number of receive antennas is less than the number of transmit antennas. A specific feedback stochastic gradient algorithm for transmission adaptation has been introduced. Numerical results showed that the adaptive algorithm performs

well, with results approaching perfect subspace tracking for feedback rates on the order of 1000 to 10 000 times the channel Doppler frequency. In systems with fewer receive than transmit antennas, the subspace tracking approach was shown to provide significant capacity and performance power gains of N_T/N_R over blind space–time coded transmission. The performance approaches that of water filling in many cases.

Appendix A
Update Stochastic Gradient Estimate

A. Vector Gradient Extraction

Define $\mathbf{g}$ as a non-random complex vector, and $\mathbf{p}$ as a zero-mean complex Gaussian vector with autocorrelation $2\mathbf{I}$. Define the decision vector $\mathbf{z}$ and error vector $\mathbf{e}$ as follows:

$$\mathbf{z} \equiv \text{sign}(\text{Re}(\mathbf{p}^H\mathbf{g})) \cdot \mathbf{p} \tag{50}$$

$$\mathbf{e} \equiv \mathbf{z} - \text{E}(\mathbf{z}). \tag{51}$$

Then, from [5] the decision is characterized by the following first and second moments:

$$\text{E}(\mathbf{z}) = \sqrt{\frac{2}{\pi}} \cdot \frac{\mathbf{g}}{\|\mathbf{g}\|} \tag{52}$$

$$\text{E}(\mathbf{e}\mathbf{e}^H) = 2\mathbf{I} - \frac{2}{\pi} \cdot \frac{\mathbf{g}\mathbf{g}^H}{\|\mathbf{g}\|^2}. \tag{53}$$

This result is mapped to the current system as follows, with $\mathbf{A}$ and $\mathbf{B}$ from (21) and (22)

$$\mathbf{Z} \equiv \text{sign}\left(\text{Re}\left(\sum_{m=1}^{N_R}\sum_{n=1}^{N_S} {a_{m,n}}^* \cdot b_{m,n}\right)\right) \cdot \mathbf{P} \tag{54}$$

The decision summation is given by

$$\sum_{m=1}^{N_R}\sum_{n=1}^{N_S} a_{m,n}{}^*(m) \cdot b_{m,n}(m) = \sum_{m=1}^{N_R}\sum_{n=1}^{N_S}\left(\sum_{i=1}^{N_T} {h^*}_{m,i}{w^*}_{i,n}\right) \times \left(\sum_{j=1}^{N_T} h_{m,j}p_{j,n}\right) = \sum_{j=1}^{N_T}\sum_{n=1}^{N_S}[\mathbf{W}^H\mathbf{H}^H\mathbf{H}]_{n,j}p_{j,n}. \tag{55}$$

If we let the vector $\mathbf{p}$ be the result of stacking the columns of the matrix $\mathbf{P}$ and the vector $\mathbf{g}$ be the result of stacking the columns of the matrix $\mathbf{H}^H\mathbf{H}\mathbf{W}$, then (55) translates to

$$\sum_{m=1}^{N_R}\sum_{n=1}^{N_S} {a_{m,n}}^* \cdot b_{m,n} = \mathbf{g}^H\mathbf{p}. \tag{56}$$

Noting that for $\mathbf{P}$ with i.i.d. entries of variance two and arbitrary non-random matrix $\mathbf{D}$

$$\text{E}\left(\mathbf{P}\underset{N_T\times N_T}{\mathbf{D}}\mathbf{P}^H\right) = 2\text{tr}(\mathbf{D})\mathbf{I} \tag{57}$$

$$\text{E}\left(\mathbf{P}^H\underset{N_S\times N_S}{\mathbf{D}}\mathbf{P}\right) = 2\text{tr}(\mathbf{D})\mathbf{I} \tag{58}$$

then, (52) and (53) give

$$\text{E}(\mathbf{Z}) = \sqrt{\frac{2}{\pi}}\frac{\mathbf{H}^H\mathbf{H}\mathbf{W}}{\|\mathbf{H}^H\mathbf{H}\mathbf{W}\|_F} \tag{59}$$

$$\mathbf{E} \equiv \mathbf{Z} - \text{E}(\mathbf{Z}) \tag{60}$$

$$\text{E}(\mathbf{E}\mathbf{E}^H) = 2N_T \cdot \mathbf{I} - \frac{2}{\pi}\frac{\mathbf{H}^H\mathbf{H}\mathbf{W}\mathbf{W}^H\mathbf{H}^H\mathbf{H}}{\|\mathbf{H}^H\mathbf{H}\mathbf{W}\|_F^2} \tag{61}$$

$$\text{E}(\mathbf{E}^H\mathbf{E}) = 2N_S \cdot \mathbf{I} - \frac{2}{\pi}\frac{\mathbf{W}^H\mathbf{H}^H\mathbf{H}\mathbf{H}^H\mathbf{H}\mathbf{W}}{\|\mathbf{H}^H\mathbf{H}\mathbf{W}\|_F^2}. \tag{62}$$

Appendix B
Convergence in Static Channel With Noiseless Update

The iterative update of the algorithm utilizes the Gram–Schmidt procedure to maintain the orthogonality of the column vectors of $\mathbf{W}$. The expectation of the gradient based update yields a matrix premultiplication of $\mathbf{W}$ prior to the Gram–Schmidt orthogonalization, so that if this premultiplying matrix is invariant the update takes the form of the Gram–Schmidt QR iteration, which can be used perform an eigendecomposition of the premultiplying matrix [19]. With the assumption that the gradient estimation error matrix is zero, this appendix derives the convergence path of this QR adaptation, a result found in neither [19] nor [14].

Define the orthogonal projector matrix as $\mathbf{P}(.)$ and the vector normalization function as $\text{N}(.)$

$$\mathbf{P}(\mathbf{a}) \equiv \mathbf{I} - \frac{\mathbf{a}\mathbf{a}^H}{\mathbf{a}^H\mathbf{a}} \tag{63}$$

$$\text{N}(\mathbf{a}) \equiv \frac{\mathbf{a}}{\|\mathbf{a}\|}. \tag{64}$$

Then, the Gram–Schmidt column orthonormalization $\mathbf{B} = \text{G}(\mathbf{A})$ of a matrix $\mathbf{A}$ comprised of column vectors $\mathbf{a}_n$ is defined by

$$\mathbf{B} = [\mathbf{b}_1 \quad \mathbf{b}_2 \quad \ldots] = \text{G}(\mathbf{A}) \tag{65}$$

$$\mathbf{b}_n = \text{N}\left(\left(\prod_{i=1}^{n-1}\mathbf{P}(\mathbf{b}_i)\right)\mathbf{a}_n\right). \tag{66}$$

With projection into the right-singular spaces, from (35) and (31) the premultiplying matrix of the noiseless update is defined by the diagonal matrix $\mathbf{D}(k)$ as follows:

$$\mathbf{D}(k) \equiv \mathbf{I} + \beta\sqrt{\frac{2}{\pi}}\frac{\mathbf{\Lambda}}{\left\|\mathbf{\Lambda}\tilde{\mathbf{W}}(k)\right\|_F}. \tag{67}$$

The right-singular space projected matrix $\tilde{\mathbf{W}}$ (29) is composed of column vectors $\tilde{\mathbf{w}}_1, \tilde{\mathbf{w}}_2 \ldots \tilde{\mathbf{w}}_{N_s}$. We base the proof upon the induction of Lemma 1a, noting that for $n = 1$ the relationship (71) holds trivially so that the induction has a valid foundation. Note that the intermediate vector normalization is omitted, as the scaling can be applied at any time without modifying the result.

Theorem 1: If the matrix $\tilde{\mathbf{W}}(k)$ is constructed by iterating a matrix premultiplication by Hermitian symmetric $\mathbf{D}(k)$ followed by Gram–Schmidt column orthonormalization $\text{G}(.)$, where $\text{G}(.)$ returns the Q-matrix of a Gram–Schmidt QR decomposition, as follows:

$$\tilde{\mathbf{W}}(k) = \text{G}\left(\mathbf{D}(k-1)\tilde{\mathbf{W}}(k-1)\right). \tag{68}$$

Then

$$\tilde{\mathbf{W}}(k) = \text{G}\left(\left(\prod_{i=0}^{k-1}\mathbf{D}(i)\right)\cdot\tilde{\mathbf{W}}(0)\right). \tag{69}$$

$$\left(\prod_{i=1}^{n}\mathbf{P}\left(\tilde{\mathbf{w}}_i(k)\right)\right)\mathbf{D}(k-1)\left(\prod_{\substack{i=n\\ i--}}^{1}\mathbf{P}\left(\tilde{\mathbf{w}}_i(k-1)\right)\right)$$
$$=\left(\prod_{i=1}^{n-1}\mathbf{P}\left(\tilde{\mathbf{w}}_i(k)\right)\right)\cdot\left(\mathrm{I}-\frac{\left(\prod_{i=1}^{n-1}\mathbf{P}\left(\tilde{\mathbf{w}}_i(k)\right)\right)\mathbf{D}(k-1)\tilde{\mathbf{w}}_n(k-1)\tilde{\mathbf{w}}_n^H(k-1)\mathbf{D}(k-1)\left(\prod_{i=1}^{n-1}\mathbf{P}\left(\tilde{\mathbf{w}}_i(k)\right)\right)}{\mathbf{w}_n^H(k-1)\mathbf{D}(k-1)\left(\prod_{i=1}^{n-1}\mathbf{P}\left(\tilde{\mathbf{w}}_i(k)\right)\right)\mathbf{D}(k-1)\tilde{\mathbf{w}}_n(k-1)}\right)\cdot\mathbf{D}(k-1)$$
$$=\left(\prod_{i=1}^{n-1}\mathbf{P}\left(\tilde{\mathbf{w}}_i(k)\right)\right)\cdot\mathbf{D}(k-1). \tag{75}$$

Lemma 1a: If the recursion (70) defines a new column vector set $\tilde{\mathbf{w}}_n(k)$ from the prior set $\tilde{\mathbf{w}}_n(k-1)$ through premultiplication by a Hermitian symmetric matrix $\mathbf{D}(k-1)$, and the property (71) holds, relating the first $n-1$ projectors of the old and new sets, as follows:

$$\tilde{\mathbf{w}}_n(k)=\left(\prod_{i=1}^{n-1}\mathbf{P}(\tilde{\mathbf{w}}_i(k))\right)\mathbf{D}(k-1)\times\tilde{\mathbf{w}}_n(k-1), \tag{70}$$

$$\left(\prod_{i=1}^{n-1}\mathbf{P}\left(\tilde{\mathbf{w}}_i(k)\right)\right)\mathbf{D}(k-1)\left(\prod_{\substack{i=n-1\\ i--}}^{1}\mathbf{P}\left(\tilde{\mathbf{w}}_i(k-1)\right)\right)=\left(\prod_{i=1}^{n-1}\mathbf{P}\left(\tilde{\mathbf{w}}_i(k)\right)\right)\mathbf{D}(k-1) \tag{71}$$

where $i--$ indicates that the index is decremented in order to be consistent with the leftward matrix factorization, then the property (70) extends to the n^{th} vector of the new set as given in (72)

$$\left(\prod_{i=1}^{n}\mathbf{P}\left(\tilde{\mathbf{w}}_i(k)\right)\right)\mathbf{D}(k-1)\left(\prod_{i=1}^{n}\mathbf{P}\left(\tilde{\mathbf{w}}_i(k-1)\right)\right)=\left(\prod_{i=1}^{n}\mathbf{P}\left(\tilde{\mathbf{w}}_i(k)\right)\right)\mathbf{D}(k-1). \tag{72}$$

Proof of Lemma 1a: The projectors are generated from orthogonal vectors $\tilde{\mathbf{w}}_n$ from (70), so that the order of the projectors can be interchanged. Therefore

$$\prod_{\substack{i=n\\ i--}}^{1}\mathbf{P}\left(\tilde{\mathbf{w}}_i(k)\right)=\prod_{i=1}^{n}\mathbf{P}\left(\tilde{\mathbf{w}}_i(k)\right). \tag{73}$$

Inserting (71) and (73) into the left side of (72) yields

$$\left(\prod_{i=1}^{n}\mathbf{P}\left(\tilde{\mathbf{w}}_i(k)\right)\right)\mathbf{D}(k-1)\left(\prod_{\substack{i=n\\ i--}}^{1}\mathbf{P}\left(\tilde{\mathbf{w}}_i(k-1)\right)\right)=\left(\prod_{i=1}^{n-1}\mathbf{P}\left(\tilde{\mathbf{w}}_i(k)\right)\right)\mathbf{P}\left(\tilde{\mathbf{w}}_n(k)\right)\mathbf{D}(k-1)\mathbf{P}\left(\tilde{\mathbf{w}}_n(k-1)\right). \tag{74}$$

This leaves the nth projection to be proved. Replacing $\tilde{\mathbf{w}}_n(k)$ with the right-hand side of (70), utilizing the idempotency of the projectors, and performing algebraic simplification, this is shown in (75), at the top of the page. Q.E.D.

Proof of Theorem 1: Equation (68), in terms of the individual column vectors $\tilde{\mathbf{w}}_n(k)$ of $\tilde{\mathbf{W}}(k)$, is

$$\tilde{\mathbf{w}}_n(k)=\mathrm{N}\left(\left(\prod_{i=1}^{n-1}\mathbf{P}\left(\tilde{\mathbf{w}}_i(k)\right)\right)\mathbf{D}(k-1)\tilde{\mathbf{w}}_n(k-1)\right). \tag{76}$$

From (68) and the definition of the Gram–Schmidt orthonormalization the vector $\tilde{\mathbf{w}}_n(k-1)$ can be replaced, giving

$$\tilde{\mathbf{w}}_n(k)=\mathrm{N}\left(\left(\prod_{i=1}^{n-1}\mathbf{P}\left(\tilde{\mathbf{w}}_i(k)\right)\right)\mathbf{D}(k-1)\cdot\mathrm{N}\left(\left(\prod_{i=1}^{n-1}\mathbf{P}\left(\tilde{\mathbf{w}}_i(k-1)\right)\right)\mathbf{D}(k-2)\tilde{\mathbf{w}}_n(k-2)\right)\right). \tag{77}$$

From Lemma 1a, this is

$$\tilde{\mathbf{w}}_n(k)=\mathrm{N}\left(\left(\prod_{i=1}^{n-1}\mathbf{P}\left(\tilde{\mathbf{w}}_i(k)\right)\right)\mathbf{D}(k-1)\mathbf{D}(k-2)\times\tilde{\mathbf{w}}_n(k-2)\right). \tag{78}$$

Applying this inductively for $k-3,k-4,\ldots 0$ yields

$$\tilde{\mathbf{w}}_n(k)=\mathrm{N}\left(\left(\prod_{i=1}^{n-1}\mathbf{P}\left(\tilde{\mathbf{w}}_i(k)\right)\right)\left(\prod_{i=1}^{k-1}\mathbf{D}(i)\right)\tilde{\mathbf{w}}_n(0)\right)=\mathrm{G}\left(\left(\prod_{i=1}^{k-1}\mathbf{D}(i)\right)\tilde{\mathbf{w}}_n(0)\right) \tag{79}$$

Q.E.D.

REFERENCES

[1] S. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. Select. Areas Commun.*, vol. 16, no. 8, Oct. 1998.

[2] G. J. Foschini, "Layered space–time architecture for wireless communication in a fading environment when using multielement antennas," *Bell Labs Tech. J.*, pp. 41–59, Autumn 1996.

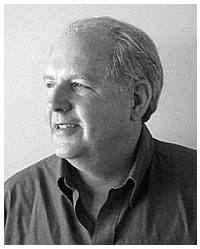
[3] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," in *Wireless Personal Communications*. New York: Kluwer, Mar. 1998, vol. 6, pp. 311–335.

[4] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block coding for wireless communications: performance results," *IEEE J. Select. Areas Commun.*, vol. 17, pp. 451–460, Mar. 1999.

[5] B. C. Banister and J. R. Zeidler, "A simple gradient sign algorithm for transmit antenna weight adaptation with feedback," *IEEE Trans. Signal Processing*, May 2003, to be published.
[6] D. Gerlach and A. Paulraj, "Adaptive transmitting antenna arrays with feedback," *IEEE Signal Processing Lett.*, vol. 1, no. 10, Oct. 1994.
[7] R. W. Heath Jr. and A. Paulraj, "A simple scheme for transmit diversity using partial channel feedback," in *Proc. 32nd Asilomar Conf. Signals, Systems and Computers*, Nov. 1998, pp. 1073–1078.
[8] J.-W.Jen-Wei Liang and A.Arogyaswami Paulraj, "Forward link antenna diversity using feedback for indoor communication systems," in *Proc. Int. Conf. Acoustics, Speech, and Signal Processing*, May 1995, pp. 1753–1755.
[9] E. Telatar, "Capacity of multiantenna gaussian channels," *Eur. Trans. Telecomm.*, vol. 10, no. 6, pp. 585–595, Nov.–Dec. 1999.
[10] R. W. Heath, S. Sandhu, and A. Paulraj, "Antenna selection for spatial multiplexing systems with linear receivers," *IEEE Commun. Lett.*, vol. 5, pp. 142–145, Apr. 2001.
[11] S. Sandhu, R. Nabar, D. A. Gore, and A. Paulraj, "Near optimal selection of transmit antennas for a MIMO channel based on Shannon capacity," in *Proc. 34th Asilomar Conf.*, Pacific Grove, CA, Nov. 1999, pp. 567–571.
[12] D. Gore and A. Paulraj, "Space-time block coding with optimal antenna selection," in *Proc. Int. Conf. Acoustics, Speech and Signal Processing*, Salt Lake City, UT, May 2001, pp. 2441–2444.
[13] A. F. Molisch, M. Z. Win, and J. H. Winters, "Capacity of MIMO systems with antenna selection," in *Proc. Int. Conf. Communications*, Helsinki, Finland, June 2001, pp. 570–574.
[14] J. F. Yang and M. Kaveh, "Adaptive Eigenspace algorithms for direction or frequency estimation and tracking," *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. 36, pp. 241–251, Feb. 1988.
[15] T. K. Sarkar and X. Yang, "Application of the conjugate gradient and steepest descent for computing the eigenvalues of an operator," *IEEE Trans. Signal Processing*, vol. 17, pp. 31–38, May 1989.
[16] B. Yang, "Projection approximation subspace tracking," *IEEE Trans. Signal Processing*, vol. 43, pp. 97–105, Jan. 1995.
[17] B. A. Bjerke and J. G. Proakis, "Multiple transmit and receive antenna diversity techniques for wireless communications," in *Proc. Adaptive Systems for Signal Processing, Communications and Control Symp.*, Lake Louise, AB, Canada, Oct. 2000, pp. 70–75.
[18] D. Gesbert, H. Bolcskei, D. Gore, and A. Paulraj, "MIMO wireless channels: capacity and performance prediction," in *IEEE GLOBAL Telecommunications Conf. Record*, Nov. 2000, pp. 1083–1088.
[19] G. H. Golub and C. F. van Loan, *Matrix Computations*, 3rd ed. Baltimore, MD: The Johns Hopkins Univ. Press, 1996.
[20] G. Bauch, H. Khorram, and J. Hagenauer, "Iterative equalization and decoding in mobile communications systems," in *Proc. 2nd European Personal Mobile Communications Conf.*, Bonn, Germany, Sept. 1997, pp. 307–312.
[21] G. Bauch, J. Hagenauer, and N. Seshadri, "Turbo processing in transmit antenna diversity systems," in *Annals of Telecommunications*: Editions Hermes, July–Aug. 2001, vol. 56.
[22] C. B. Papadias and G. J. Foschini, "A space–time coding approach for systems employing four transmit antennas," in *Proc. Int. Conf. Acoustics, Speech, and Signal Processing*, Salt Lake City, UT, May 2001, pp. 2481–2484.

Brian Clarke Banister (M'93) received the B.S., M.S., and Ph.D. degrees in electrical engineering from the University of California, San Diego, in 1993, 1999, and 2002, respectively.

From May 1996 until April 2002, he worked at the LSI Logic Wireless Design Center, San Diego, where he worked on radio and baseband modem design for cellular communications systems, as well as research into the practical implementation of multiantenna algorithms. He has been with Qualcomm, Inc., San Diego, since May 2002, where he has focused on algorithm implementation and verification for cellular products. His research has focused on the application of multiple antenna arrays to communications and signal processing problems.



James R. Zeidler (M'76–SM'84–F'94) has been a Scientist at the Space and Naval Warfare Systems Center, San Diego, CA, since 1974. He has also been an Adjunct Professor in the Electrical and Computer Engineering Department, University of California, San Diego, since 1988.

His current research interests are in adaptive signal processing, communications signal processing, and wireless communication networks.

Dr. Zeidler was an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING from 1991 to 1994. He was corecipient of the Award for Best Unclassified Paper at the IEEE Military Communications Conference in 1995 and received the Lauritsen-Bennet Award for Achievement in Science in 2000, and the Navy Meritorious Civilian Service Award in 1991.