A Simple Gradient Sign Algorithm for Transmit Antenna Weight Adaptation With Feedback

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

***Abstract*—In this paper, a simple algorithm for adaptation of the complex baseband weights of a transmit antenna array using feedback from the receiver is proposed and analyzed. The system utilizes stochastic gradient adaptation to maximize the power delivered to the receiver for a constrained transmission power, which provides both fading diversity and beam steering gain. Dual perturbed transmission weight vectors are time multiplexed onto the pilot signal, and the receiver generates feedback selecting the perturbed weight vector which delivers greater power. This feedback is used to provide weight adaptation at the transmitter, and this adaptation is shown to be an update by a coarse estimate of the gradient of the delivered power. The performance of the algorithm is analyzed in terms of convergence and tracking of an AR1 fading channel, with simulations confirming the analysis. Bit error rate (BER) simulations in a dynamic fading channel show that the algorithm outperforms previously proposed vector selection feedback, and in slower fading, the algorithm substantially outperforms diversity space time coding.**

***Index Terms*—Adaptive beamforming, antenna array, stochastic gradient algorithms, transmit beamforming.**

I. Introduction

IT is generally accepted that the downlink of next-generation cellular systems will require greater capacity than the uplink. This is largely due to the asymmetry of data traffic patterns. For example, a mobile data terminal may download large web sites while uploading only control information such as IP addresses. The use of transmit adaptive antenna arrays at the base station is a promising area for downlink capacity improvement [1]–[3]. This paper describes a gradient algorithm utilizing mobile to base feedback in order to achieve some of those possible gains.

Optimal multiple antenna transmission algorithms can be defined if the forward channel state is known [4], [5]. However, in mobile wireless applications, the channel is time varying and unknown *a priori*. In addition, many systems use frequency division duplexing (FDD), such that the downlink and uplink channels in a multipath environment are not in general the same. Thus, the uplink channel state measured at the base station cannot be used directly as a reciprocal downlink channel state. In general, transmit antenna algorithms in this environment can be classified as

a) space time codes;
b) blind adaptive beam steering algorithms;
c) adaptive steering algorithms incorporating feedback.

With multiple receive antennas, space time codes can provide diversity gain [6], [7] and "multiplexing" coding gain [8]–[10] through the application of codes across the multiple transmit antennas, but with a single receive antenna only diversity gain is possible. These space time coding techniques are "blind," where the downlink channel state is assumed unknown to the transmitter, and there is no adaptation to changing channel conditions. These algorithms do not provide beamforming or "aperture" gain. The gains from these codes are diminished when the fading channels experienced by the antennas are correlated, with no gain when the fades are fully correlated.

On the other hand, "blind" adaptive beam steering algorithms also have no direct knowledge of the downlink channel but utilize the measured uplink channel to infer characteristics of the downlink channel. These are then used to adapt transmit antenna weights. This may require accurate antenna calibration as, for example, in estimating the angle of arrival and angular dispersion of the received signal to steer the transmit beam [11], or it may assume that the long-term characteristics of the uplink and downlink channels are strongly correlated [12]. However, these algorithms do not provide fading diversity, and if the antennas experience independent fading then blind techniques will not work, as without correlation between antennas, no correlation from uplink to downlink channel can be extracted.

In order to benefit from both fading diversity and beam steering, an algorithm incorporating detailed downlink channel information must be used, which in FDD systems requires feedback from the receiving unit to the transmitting unit. This has led to several proposals for feedback, using training sequences from each antenna [13], [14] or gradient extraction of the forward complex channel vector [15]. Feedback of the complete forward channel state as in [13] and [15] is impractical in a mobile wireless channel due to the power and bandwidth required by the feedback channel, motivating vector quantization (or "codebook selection") feedback as in [14]. This requires detailed information sharing between the transmitting and receiving units as there must be an agreed codebook of possible weight vectors, and each transmission antenna must have a distinguishable pilot sequence. In addition, the vector quantization induced coarseness of the available transmission weight vectors limits the achievable performance. The algorithm presented here overcomes these limitations

Manuscript received September 14, 2000; revised Oct. 10, 2002. This work was supported in part by CoRe Research Grant core 00-10074. The first author was with LSI Logic Corporation at the time this work was performed.The associate editor coordinating the review of this paper and approving it for publication was Dr. Alexi Gorokhov.

B. C. Banister is with the Department of Electrical and Computer Engineering, University of California at San Diego, La Jolla, CA 92093 USA, and also with Qualcomm, San Diego, CA USA (e-mail: banister@ece.ucsd.edu).

J. R. Zeidler is with the Department of Electrical and Computer Engineering, University of California at San Diego, La Jolla, CA 92093 USA, and also with Space and Naval Warfare Systems Center, San Diego, CA USA (e-mail: zeidler@nosc.mil).

Digital Object Identifier 10.1109/TSP.2002.808104

with i) a simple interface requiring the receiver to be aware only of a single dedicated pilot and ii) recursive adaptation, which allows fine adjustment over time. The dedicated pilot is generally required for beam steering with coherent demodulation regardless of the specific adaptation mechanism, and in this scheme the receiver does not need to know details of the transmit algorithm such as the number of antennas in the array.

Gradient-based algorithms vary according to the reliability and precision of the gradient information, from numerical optimization which may assume accurate knowledge, sometimes including second derivatives for Newton step adjustment [16], to signal processing applications which apply a stochastic estimate such as least mean squares (LMS) [17]. The gradient algorithm of this study considers a more coarse stochastic approach with only binary (signed) gradient information, which operates upon the usable power delivered to the mobile as the metric to be maximized.[1] Weight vector perturbations and feedback are used to extract a coarse estimate of the gradient of this metric, which is then used to adjust the transmit antenna weights. This paper will provide an analysis of the convergence and tracking performance of the algorithm. The analysis includes the effect of the noise in the gradient estimate, allowing for an adaptation rate optimization that considers the tradeoff of the induced noise versus the tracking rate. In addition, bit error rate (BER) simulations show that for moderate fading rates this algorithm outperforms previously proposed algorithms.

This paper is organized as follows. Section II introduces the general multiple transmit single receive antenna system model, eigenanalysis notation for the system, the capacity motivation for selecting transmit weights to maximize the delivered power, and a brief comparison of transmit weight adaptation to blind diversity space time coding. Section III introduces the gradient sign feedback algorithm. Section IV provides the signed feedback gradient statistics. Section V analyzes the convergence of the system in a static channel. Section VI analyzes the tracking performance in an AR1 frequency flat fading channel. Detailed derivations are located in the Appendices.

II. Preliminaries

A. System Model

The system is analyzed with Nyquist pulse shaping so that a discrete time representation of the waveforms is adequate. The system contains N_T transmission antennas at the base station, and there are a maximum of L delay paths. A wideband system might provide several such paths, while a narrow band system would give $L = 1$. In order to simplify the presentation, multipath terms in (2) have delays that are a multiple of the Nyquist sample time. This leads to a clean eigenstructure by eliminating interpolation issues and provides a clearer picture of the algorithm properties.

The vector applied to weight the transmission of the multiple antennas will be constrained to have unit norm so that it does not impact the total transmitted power. The analysis addresses the weight vector norm constraint by allowing the norm of the analyzed weight vector to be unconstrained and applying a normalization to the weight vector in (1) prior to transmission. In this way, the constraint is correctly addressed in the adaptation characterization and analysis but is not explicitly a constraint on the adaptation. This unconstrained norm reflects the analysis and not a realistic implementation. Because of the normalization of (1), it is always the "direction" of the vector $\mathbf{w}$ which is of interest and never the magnitude. Divergence of the norm of $\mathbf{w}$ is inconsequential.

The transmitted signal is

$$\mathbf{t}(n) = \sqrt{P_{\text{traffic}}^{(T)}}\frac{\mathbf{w}}{||\mathbf{w}||}s_{\text{traffic}}(n) \tag{1}$$

so that the received signal is

$$r(n) = \sum_{l=0}^{L-1}\sqrt{P_{\text{traffic}}^{(T)}}\mathbf{c}_l^H\frac{\mathbf{w}}{||\mathbf{w}||}s_{\text{traffic}}(n-l) + n(n) \tag{2}$$

where $s_{\text{traffic}}(n)$ is the modulation sequence, $n(n)$ is a complex zero mean Gaussian with variance $2\sigma_n^2$ representing the received noise and interference, P_{traffic}^T is the transmit power, $\mathbf{w}$ is the $N_T \times 1$ prenormalization transmit weight vector, and $\mathbf{c}_l$ is the $N_T \times 1$ conjugated channel response vector of the l^{th} path. For the moment, this formulation includes no pilot signal and assumes perfect channel estimation for demodulation at the receiver.

Defining the mobile's gain matrix as[2]

$$\mathbf{R} \equiv \sum_{l=0}^{L-1}\mathbf{c}_l\mathbf{c}_l^H \tag{3}$$

and the total usable signal power at the receiver is given by

$$P^{(R)} = P^{(T)} \cdot \frac{\mathbf{w}^H\mathbf{R}\mathbf{w}}{\mathbf{w}^H\mathbf{w}}. \tag{4}$$

B. Performance Metric, Gradient, and Eigenanalysis

In this application, the ultimate cost function is BER or capacity loss. The ratio of the received power to transmit power is a valid surrogate for either metric, particularly in the case of either DS-CDMA with large spreading gain, as is shown below, or a narrowband system with frequency flat fading. With the weight normalization applied prior to transmission as in (1), the metric used by the algorithm, which is to be maximized, is given by the Raleigh quotient

$$J = \frac{P^{(R)}}{P^{(T)}} = \frac{\mathbf{w}^H\mathbf{R}\mathbf{w}}{\mathbf{w}^H\mathbf{w}}. \tag{5}$$

The gradient of this metric with respect to $\mathbf{w}$ is given by

$$\mathbf{g}(\mathbf{w}) \equiv \nabla_{\mathbf{w}}(J) = 2\left(\frac{\mathbf{R}\mathbf{w}}{\mathbf{w}^H\mathbf{w}} - \frac{\mathbf{w}^H\mathbf{R}\mathbf{w}}{(\mathbf{w}^H\mathbf{w})^2}\mathbf{w}\right). \tag{6}$$

The subtractive term in (6) results from projection of the gradient into the subspace orthogonal to $\mathbf{w}$. Since adjustment in the direction of $\mathbf{w}$ determines the transmission power [or would if

[1] This is counter to more the typical gradient algorithm formulation where a cost function is minimized.

[2] It is worth noting that in the case of multiple receive antennas, (2) applies to each antenna, and (3) is defined by performing an additional summation over each receive antenna. Hence, the structure of the problem and solution is the same, but the rank is increased.

not for the power normalization in (1)] and not the transmission "direction," this projection provides for superior gradient adaptation given the power constraint.

Eigenanalysis is used in the system evaluation. The gain matrix is decomposed as follows, where the indices reflect magnitude sorted eigenvalues, with λ_0 being the largest.

$$\mathbf{R} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^H = \sum_{n=0}^{N_T-1} \lambda_n \mathbf{q}_n \mathbf{q}_n^H. \tag{7}$$

Then the weight vector can be represented in terms of the eigenmodes as an eigenweight vector $\mathbf{u}$

$$\mathbf{u} = \mathbf{Q}^H \mathbf{w}. \tag{8}$$

The eigenweight representations of (5) and (6) are

$$J = \frac{\mathbf{u}^H \mathbf{\Lambda} \mathbf{u}}{\mathbf{u}^H \mathbf{u}} \tag{9}$$

$$\mathbf{g}(\mathbf{w}) = \frac{2}{\mathbf{u}^H \mathbf{u}} \cdot \mathbf{Q}\left(\mathbf{\Lambda} - J\mathbf{I}\right)\mathbf{u}. \tag{10}$$

The analysis will consider the second moments, so the vector $\mathbf{v}$, with elements v_n, is defined to be comprised of the squared magnitudes of $\mathbf{u}$. Then, with

$$\mathbf{1} \equiv [1 \quad 1 \quad \ldots \quad 1]^T \tag{11}$$

$$[\mathbf{v}]_n = v_n \equiv |u_n|^2 \tag{12}$$

we define higher order values of J as $J^{(k)}$ as in (13) and find the following:

$$J^{(k)} \equiv \frac{\mathbf{1}^T \mathbf{\Lambda}^k \mathbf{v}}{\mathbf{1}^T \mathbf{v}} \tag{13}$$

$$J \equiv J^{(1)} = \frac{\mathbf{1}^T \mathbf{\Lambda} \mathbf{v}}{\mathbf{1}^T \mathbf{v}} \tag{14}$$

$$\begin{aligned} \|\mathbf{g}(\mathbf{w})\|^2 &= \frac{4}{\left(\mathbf{1}^T \mathbf{v}\right)^2} \mathbf{1}^T \left(\mathbf{\Lambda} - J\mathbf{I}\right)^2 \mathbf{v} \\ &= \frac{4}{\mathbf{1}^T \mathbf{v}} \left(J^{(2)} - J^2\right). \end{aligned} \tag{15}$$

In contrast to the MSE problem in receive systems [17], the gradient does not consist of independent terms in each eigenmode, as the $-J\mathbf{I}$ term of (10) introduces intermodal dependence. This coupling arises from the weight normalization (5), which constrains the gradient to be orthogonal to $\mathbf{w}$.

The gradient (10) of the Raleigh quotient is zero only when the weight vector lies in one of the eigenspaces of $\mathbf{R}$ (e.g., the nullspace or the principal eigenspace) [18]. Hence, when one eigenmode dominates the weight vector $\mathbf{w}$, the resulting gradient is small. The only local maximum is the global maximum, occurring within the principal eigenspace. Within the span of other nonminimal valued eigenvectors, a saddle point gives rise to the zero gradient, where the surface rises in the direction of larger valued eigenvectors and falls in the direction of smaller valued eigenvectors. If the principal eigenspace is not unique, then the solution to the power maximization problem is not unique and lies in any of those principal eigenspaces. The only minimum is the global minimum, occurring as a bowl bottom in the minimal eigenspace (the nullspace if it exists).

C. Channel Capacity

To derive the capacity of this system, the vector frequency response of the channel is given by

$$\mathbf{C}(z) = \sum_{k=0}^{L-1} \mathbf{c}_l z^{-k}. \tag{16}$$

Then, the capacity (in bits per second per Hertz) of this discrete time channel is given by

$$C = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log_2 \left(1 + \frac{P^{(T)} \mathbf{w}^H \mathbf{C}(e^{jw}) \mathbf{C}^H(e^{jw}) \mathbf{w}}{2\sigma_n^2 \mathbf{w}^H \mathbf{w}}\right) d\omega. \tag{17}$$

In a DS-CDMA system with large spreading gain, the signal density to noise density ratio at each frequency is small so that the capacity formula is approximated using a first-order Taylor expansion as

$$\begin{aligned} C &\cong \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{P^{(T)} \mathbf{w}^H \mathbf{C}(e^{jw}) \mathbf{C}^H(e^{jw}) \mathbf{w}}{2\sigma_n^2 \ln(2) \mathbf{w}^H \mathbf{w}} d\omega \\ &= \frac{P^{(R)}}{2\sigma_n^2 \ln(2)}. \end{aligned} \tag{18}$$

Hence, the capacity is maximized in this context by selecting $\mathbf{w}$ to maximize the received power. Similarly, we can show that in a DS-CDMA system this power maximization minimizes bit and frame error rates in additive white Gaussian noise. In the rank 1 case (single time resolvable path), this conclusion is trivially extended to narrowband systems. Even in systems which may be more "bandwidth limited" than DS-CDMA and yet undergoing nonflat fading, power maximization is clearly a reasonable adaptation objective, although it is no longer necessarily optimal. Hence, maximizing the received power for a constrained transmission power is an effective strategy for weight selection. The received power is maximized by selecting $\mathbf{w}$ as $\mathbf{q}_0$, the eigenvector corresponding to the largest eigenvalue of $\mathbf{R}$ so that for arbitrary ϕ

$$\frac{\mathbf{w}_{opt}}{\|\mathbf{w}_{opt}\|} = e^{j\phi} \mathbf{q}_0. \tag{19}$$

For $L = 1$, this is the matched filter weights [3], [15], providing both fading diversity and beam steering gain.

$$\frac{\mathbf{w}_{opt}}{\|\mathbf{w}_{opt}\|} = e^{j\phi} \frac{\mathbf{c}_0}{\|\mathbf{c}_0\|}. \tag{20}$$

D. Comparison to Diversity Space Time Coding

Here, a brief comparison of optimal adaptive beam forming and diversity space time coding is provided. Optimal adaptive beam forming will outperform blind diversity space time coding because the former assumes that full channel state information is available at the transmitter, e.g., from a feedback algorithm such as that introduced in Section III. When the receiver has only a single antenna, space time coding techniques (e.g., [6], [7]) can provide diversity but cannot provide the multiplexing coding gain seen in multiple-input multiple-output systems (e.g., [8]–[10]) so that a performance comparison based only on received power can be considered adequate. In the case

of a diversity space time coding transmission, a general formulation of the transmission vector is

$$\mathbf{t}_{\mathrm{STC}}(n) = \sqrt{\frac{P_{\mathrm{traffic}}^{(T)}}{N_T}} \begin{bmatrix} s_{\mathrm{traffic},0}(n) \\ s_{\mathrm{traffic},1}(n) \\ \cdots \\ s_{\mathrm{traffic},N_T-1}(n) \end{bmatrix}. \tag{21}$$

For a good space time code design, this transmission vector is uncorrelated since the diversity gain is obtained by delivering energy to as many of the spatial modes as possible (e.g., this is true of the schemes of [6] and [7] with independent symbol inputs).

$$\mathbf{\Phi}_{\mathrm{STC}} \equiv E\left(\mathbf{t}_{\mathrm{STC}}\mathbf{t}_{\mathrm{STC}}^H\right) = \frac{P_{\mathrm{traffic}}^{(T)}}{N_T}\mathbf{I}. \tag{22}$$

This is in contrast to the adaptive weight scheme, wherein the transmission energy is concentrated in a single (optimal) spatial mode and the autocorrelation is of rank 1.

$$\mathbf{\Phi} \equiv E\left(\mathbf{t}\mathbf{t}^H\right) = P_{\mathrm{traffic}}^{(T)}\frac{\mathbf{w}\mathbf{w}^H}{\|\mathbf{w}\|^2} = P_{\mathrm{traffic}}^{(T)}\mathbf{q}_0\mathbf{q}_0^H. \tag{23}$$

The ratio of the power received in the adaptive weight scheme versus the diversity space time coding scheme is

$$\frac{P^{(R)}}{P_{\mathrm{STC}}^{(R)}} = \frac{tr\left(\mathbf{\Phi}\mathbf{R}\right)}{tr\left(\mathbf{\Phi}_{\mathrm{STC}}\mathbf{R}\right)} = \frac{N_T\lambda_0}{\sum_{k=0}^{N_T-1}\lambda_k} \tag{24}$$

$$\frac{N_T}{\min\left(L,N_T\right)} \leq \frac{P^{(R)}}{P_{\mathrm{STC}}^{(R)}} \leq N_T. \tag{25}$$

Hence, we see that optimal adaptive weighting will always outperform the diversity space time coding scheme with a gain that increases as the ratio of the number of antennas to the number of paths (i.e., rank of $\mathbf{R}$) is increased. The performance in terms of this power delivery translates directly to instantaneous capacity or bit/frame error rate for flat fading or DS-CDMA with large spreading gain, as has been discussed, so that this gives a precise performance comparison in these cases. For the single path case, the gain of optimal beam forming over diversity space time coding is $10 \cdot \log_{10}(N_T)$ dB at all time instants. Other cases would require detailed consideration of equalization techniques and coding structure for a more thorough performance comparison.

III. Algorithm Definition

The proposed algorithm uses binary sign feedback to provide a gradient estimate to the transmitting unit. This has some similarities to the "sign algorithm" (SA) applied in some receive systems attempting to minimize mean square error (MSE) [19]–[21]. The SA is an approximation to more precise gradient techniques such as LMS, motivated by the desire to minimize the complexity of the weight update. The motivation for using a gradient sign in this transmit system is very different: the minimization of the amount of information required as feedback for weight adjustment. The Raleigh quotient metric in this system, as discussed above, and the norm constrained adaptation resulting from its gradient are different from the quadratic MSE cost function applied in SA receivers.

The system can be considered to be a DS-CDMA system with a code multiplexed pilot or a time division multiple access system with a time multiplexed training sequence. The multiplexing type of the system is not specified for the analysis as it is not relevant to the basic properties of the algorithm. Time multiplexing would be obtained with appropriate zeroing of the traffic and pilot baseband signals. The base station transmits the data with a weight vector $\mathbf{w}$, whereas the pilot is transmitted using two different weight vectors, $\mathbf{w}_{\mathrm{even}}$ and $\mathbf{w}_{\mathrm{odd}}$, which are perturbed from the tracked transmission weight vector $\mathbf{w}$. All of these weight vectors are constant during a measurement interval. The transmitted waveform for the specific mobile is given by

$$\begin{aligned}\mathbf{t}(n) =& \sqrt{P_{\mathrm{traffic}}^{(T)}} \cdot \frac{\mathbf{w}}{\|\mathbf{w}\|} s_{\mathrm{traffic}}(n) \\ &+ \sqrt{P_{\mathrm{pilot}}^{(T)}} \cdot s_{\mathrm{pilot}}(n) \\ &\cdot \left\{ \begin{array}{ll} \frac{\mathbf{w}_{\mathrm{even}}}{\|\mathbf{w}_{\mathrm{even}}\|} & \text{if } \left(\lfloor \frac{n}{M} \rfloor == \text{even}\right) \\ \frac{\mathbf{w}_{\mathrm{odd}}}{\|\mathbf{w}_{\mathrm{odd}}\|} & \text{if } \left(\lfloor \frac{n}{M} \rfloor == \text{odd}\right) \end{array} \right\}. \end{aligned} \tag{26}$$

Here, n is the Nyquist sampling time index, M is the duration of each even/odd perturbation slot, $s_{\mathrm{traffic}}(n)$ is the information bearing modulated symbols, $s_{\mathrm{pilot}}(n)$ is the pilot sequence modulation, $P_{\mathrm{traffic}}^{(T)}$ is the mean traffic channel transmission power, and $P_{\mathrm{pilot}}^{(T)}$ is the mean pilot channel transmission power. The sequences $s_{\mathrm{traffic}}(n)$ and $s_{\mathrm{pilot}}(n)$ would typically be orthogonal by code or time multiplexing.

The receiver generates feedback by determining whether the even or the odd time slot provided the larger received pilot power. The feedback bit d is "+1" to select the even slot or "−1" to select the odd slot. Note that in the case of resolvable multipath [rank$(\mathbf{R}) > 1$], the receiver will be tracking several versions of the received pilot, making channel estimates for each path and combining these channel estimate powers prior to doing the decision comparison, as is implied by (5). For the purpose of this discussion, it is assumed that the base station receives the feedback bit instantaneously after the completion of the measurement period. Given this feedback, the base station generates new transmit weights as follows.

```
when feedback is received at beginning of
new test interval
 if ( feedback == +1, indicating the even
channel was better)
  w ⇐ w_even
 else
  w ⇐ w_odd
 end if
 p ⇐ new test perturbation vector
 w_even ⇐ w + ||w|| β · p
 w_odd ⇐ w − ||w|| β · p
```

The test perturbation $\mathbf{p}$ is generated as a zero mean complex random Gaussian vector with an autocorrelation matrix $2\mathbf{I}$. The

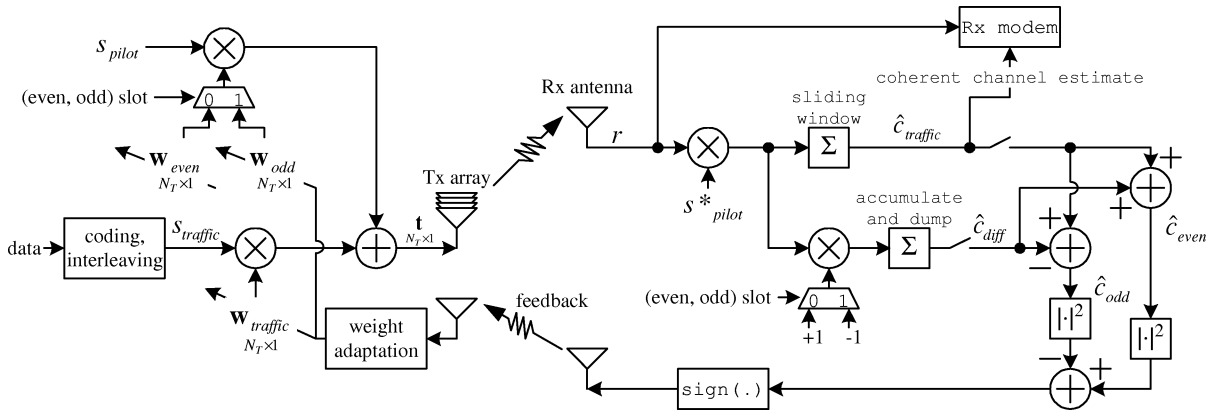

Fig. 1. Diagram of the system with example of efficient receiver channel estimation for coherent demodulation and antenna adaptation feedback.

parameter β is the adaptation rate, with a larger β giving faster but noisier tracking, as is seen in the following analysis. The normalization of the even/odd weight vectors applied in (26) results in equal power transmission in both even and odd time slots, and hence, the mobile selects the better weight vector direction rather than a larger transmission power, according to (5) and (10). Note that the norm of the weight $\mathbf{w}$ is allowed to grow for the analysis, but a practical implementation would maintain normalized vectors at every step. Note also that $\mathbf{w}$ is the mean of the even and odd pilot weight vectors, so that obtaining a channel estimate from the received pilot for the coherent demodulation of the traffic is straightforward, as discussed below.

To further illustrate the algorithm operation, let the time index i indicate the algorithm measurement/update interval, the duration of which is an (unspecified) integer multiple of $2M$ times the Nyquist sampling interval (indexed n). At time i, the weight vector $\mathbf{w}(i)$ is applied to the traffic, and the weight vectors for the even/odd pilot time slots are

$$\mathbf{w}_{\text{even}}(i) = \mathbf{w}(i) + \|\mathbf{w}(i)\|\,\beta\cdot\mathbf{p}(i) \tag{27}$$

$$\mathbf{w}_{\text{odd}}(i) = \mathbf{w}(i) - \|\mathbf{w}(i)\|\,\beta\cdot\mathbf{p}(i). \tag{28}$$

The received pilot channel power is given by inserting (27) and (28) in (4). The decision statistic $T(i)$ computed by the receiver is given by the difference between the measured even and odd received pilot power, which in the case of perfect power estimation, gives

$$T(i) = P_{\text{pilot}}^{(T)} \times\left(\frac{(\mathbf{w}(i)+\beta\|\mathbf{w}(i)\|\cdot\mathbf{p}(i))^H\mathbf{R}(\mathbf{w}(i)+\beta\|\mathbf{w}(i)\|\cdot\mathbf{p}(i))}{\|\mathbf{w}(i)+\beta\|\mathbf{w}(i)\|\cdot\mathbf{p}(i)\|^2} - \frac{(\mathbf{w}(i)-\beta\|\mathbf{w}(i)\|\cdot\mathbf{p}(i))^H\mathbf{R}(\mathbf{w}(i)-\beta\|\mathbf{w}(i)\|\cdot\mathbf{p})}{\|\mathbf{w}(i)-\beta\|\mathbf{w}(i)\|\cdot\mathbf{p}(i)\|^2}\right). \tag{29}$$

The decision feedback $d(i)$ from the receiver and the subsequent weight update at the transmitter are

$$d(i) = \text{sign}\,(T(i)) \tag{30}$$

$$\mathbf{w}\,(i+1) = \mathbf{w}(i) + d(i)\cdot\beta\,\|\mathbf{w}(i)\|\cdot\mathbf{p}(i). \tag{31}$$

New perturbed vectors (27) and (28) are then calculated for time $i+1$, and the process continues iteratively.

A block diagram for transmitter and receiver of this system is shown in Fig. 1, which in this case illustrates recovery of a single path. As shown, an efficient receiver for this mechanism can use a sliding window filter to generate an estimate of the composite complex scalar channel of the traffic transmission $\hat{c}_{\text{traffic}}$, summing the received pilot channel estimates over an equal number of both even and odd time slots. This estimate can be used directly for coherent data demodulation since the data was transmitted with the mean of the even and odd pilot weight vectors. At the same time, a difference channel estimate $\hat{c}_{\text{diff}}$ is generated via an accumulate and dump unit since it is not needed continually for demodulation. From these two estimates, the channel estimates for even and odd time slots $\hat{c}_{\text{even}}$ and $\hat{c}_{\text{odd}}$ are generated, which ultimately provide the feedback.

IV. Gradient Extraction

In the following, the feedback is assumed to be generated by the receiver without estimation error [as in (29)] and received by the transmitting unit without error. The channel $\mathbf{c}$, and hence $\mathbf{R}$, is nonrandom (or taken as given). If β is small, then the test statistic T can be considered to be generated as a weighted differential step. That is, the test statistic at time i is the gradient $\mathbf{g}(\mathbf{w}(i))$ (6) weighted by the perturbation vector $\mathbf{p}(i)$.

$$\begin{aligned} T(i) &\cong P_{\text{pilot}}^{(T)}\left(J+\beta\|\mathbf{w}(i)\|\,\text{Re}\left(\mathbf{g}\,(\mathbf{w}(i))^H\mathbf{p}(i)\right)\right) \\ &\quad - P_{\text{pilot}}^{(T)}\left(J-\beta\|\mathbf{w}(i)\|\,\text{Re}\left(\mathbf{g}\,(\mathbf{w}(i))^H\mathbf{p}(i)\right)\right) \\ &= P_{\text{pilot}}^{(T)}\beta\|\mathbf{w}(i)\|\cdot\left(\mathbf{p}(i)^H\mathbf{g}\,(\mathbf{w}(i)) + \mathbf{g}\,(\mathbf{w}(i))^H\mathbf{p}(i)\right). \end{aligned} \tag{32}$$

From (31) and (32), the update at the transmitter upon receiving the feedback is then given by

$$\mathbf{w}\,(i+1) = \mathbf{w}(i) + \beta\,\|\mathbf{w}(i)\| \cdot\text{sign}\left(\text{Re}\left(\mathbf{p}(i)^H\mathbf{g}\,(\mathbf{w}(i))\right)\right)\mathbf{p}(i). \tag{33}$$

As is shown in Appendix A, with $\mathbf{p}$ comprised of i.i.d. Gaussians and $\mathbf{g}$ given (or nonrandom), this update takes the form of

a step of fixed magnitude in the direction of the gradient $\mathbf{g}$ (in expectation) plus a zero mean error term $\mathbf{e}$.

$$\mathbf{w}(i+1) \cong \mathbf{w}(i) + \sqrt{\frac{2}{\pi}} \cdot \beta \|\mathbf{w}(i)\| \cdot \frac{\mathbf{g}(\mathbf{w}(i))}{\|\mathbf{g}(\mathbf{w}(i))\|} + \beta \|\mathbf{w}(i)\| \cdot \mathbf{e}(i). \quad (34)$$

The algorithm by its basic formulation is a "descent method" [16], and from (34), it is seen to be a stochastic gradient algorithm, performing steepest descent adaptation in expectation with a cost function $-J$, which is given by sign inversion of the maximization metric J.

From (86), the variance of the error vector cast into each of the eigenmodes is given by

$$E\left(\mathbf{Q}^H \mathbf{e}\mathbf{e}^H \mathbf{Q}\right) = 2\mathbf{I} - \frac{2}{\pi} \frac{\mathbf{Q}^H \mathbf{g}\mathbf{g}^H \mathbf{Q}}{\|\mathbf{g}\|^2}. \quad (35)$$

The variance of the update in the nth eigenmode is given by

$$\left[E\left(\mathbf{Q}^H \mathbf{e}\mathbf{e}^H \mathbf{Q}\right)\right]_{n,n} = 2 - \frac{2}{\pi} \frac{(\lambda_n - J)^2 \cdot v_n}{\mathbf{1}^T (\mathbf{\Lambda} - J\mathbf{I})^2 \mathbf{v}}. \quad (36)$$

It has been shown that the weight update defined for the algorithm provides an estimate of the gradient of the delivered power performance metric. This estimate provides the two-norm normalized direction of the gradient in expectation. The normalization of the gradient estimate is to be expected since the binary feedback does not provide any information of the gradient amplitude but only of the relative magnitudes of the gradient components. Note that a Gaussian estimation error at the receiver, in (29), would simply modify the normalization of the gradient in (34) since this would be an additional Gaussian component in the summation of (78). In addition, a feedback bit error probability of p modifies the derived update expectation derived in by a factor of $(1-p)$. Both modifications also require appropriate adjustment to the update variance. Incorporation of these effects is beyond the scope of the current work.

V. Convergence in Static Channel

A. Derivation of Learning Curve

The convergence can be visualized by inserting the gradient (6) into the update (34), giving

$$\begin{aligned}\mathbf{w}(i+1) &\cong \left(\mathbf{I} + \frac{\sqrt{\frac{2}{\pi}} \cdot \beta \|\mathbf{w}(i)\|}{\left\| \left(\frac{\mathbf{R}}{\mathbf{w}^H(i)\mathbf{w}(i)} - \frac{\mathbf{w}^H(i)\mathbf{R}\mathbf{w}(i)}{(\mathbf{w}^H(i)\mathbf{w}(i))^2} \mathbf{I} \right) \mathbf{w}(i) \right\|} \right. \\ &\quad \left. \times \left(\frac{\mathbf{R}}{\mathbf{w}^H(i)\mathbf{w}(i)} - \frac{\mathbf{w}^H(i)\mathbf{R}\mathbf{w}(i)}{(\mathbf{w}^H(i)\mathbf{w}(i))^2} \mathbf{I} \right) \right) \mathbf{w}(i) \\ &\quad + \beta \|\mathbf{w}(i)\| \cdot \mathbf{e}(i). \end{aligned} \quad (37)$$

Hence, it is seen that the expectation of the adaptation update takes the form of a matrix premultiplication of the weights by the channel gain matrix $\mathbf{R}$ plus a scaled identity. By the principle of matrix power iteration[3] [18], iterating such a modification tends to extract the eigenvector corresponding to the largest eigenvalue of the premultiplying matrix. In this case, the largest non-negative eigenvalue of the premultiplying matrix is associated with $\mathbf{q}_0$, and hence, the iteration tends to extract the desired eigenvector-space and maximize the power delivered to the receiver. While there are degenerate conditions with large negative eigenvalues, it turns out that such conditions are transient and do not yield undesirable behavior.

The behavior is best analyzed within the eigendomain. Inserting (8) and (10) into (34), the eigenweights of the weight vector adapt as follows:

$$u_n(i+1) = u_n(i) \cdot \left(1 + \sqrt{\frac{2}{\pi}} \cdot \frac{2\beta(\lambda_n - J(i))}{\|\mathbf{w}(i)\| \cdot \|\mathbf{g}(\mathbf{w}(i))\|}\right) + \beta \|\mathbf{w}(i)\| \cdot \mathbf{q}_n^H \mathbf{e}. \quad (38)$$

One can approximate this weight adjustment by assuming a noiseless update so that the new value of u_n is given simply by the expectation of the step (i.e., $\mathbf{e} \Rightarrow \mathbf{0}$). However, it is more informative to consider the expectation of the magnitude squared of the eigenweights (the vector $\mathbf{v}$ (12)), which allows the effects of the coarse gradient estimation of the update to be included. Then, from (36) and (38)

$$E(v_n(i+1)|\mathbf{v}(i)) = v_n(i) \cdot \left(1 + \sqrt{\frac{2}{\pi}} \cdot \frac{2\beta(\lambda_n - J(i))}{\|\mathbf{w}(i)\| \cdot \|\mathbf{g}(\mathbf{w}(i))\|}\right)^2 + \beta^2 \cdot \|\mathbf{w}(i)\|^2 \mathbf{q}_n^H E\left(\mathbf{e}\mathbf{e}^H\right) \mathbf{q}_n. \quad (39)$$

With the error variance from (36) and some simplification using (11)–(15), the algorithm adaptation update is parameterized only by $\mathbf{\Lambda}$, β, and the current state $\mathbf{v}(i)$.

$$\mathbf{G}_{\text{alg}}(\mathbf{v}(i)) \equiv \mathbf{I} + \sqrt{\frac{2}{\pi}} \cdot \frac{2\beta}{\sqrt{\mathbf{1}^T \left(\mathbf{\Lambda} - \mathbf{1}^T \mathbf{\Lambda} \frac{\mathbf{v}(i)}{\mathbf{1}^T \mathbf{v}(i)} \cdot \mathbf{I}\right)^2 \frac{\mathbf{v}(i)}{\mathbf{1}^T \mathbf{v}(i)}}} \cdot \left(\mathbf{\Lambda} - \mathbf{1}^T \mathbf{\Lambda} \frac{\mathbf{v}(i)}{\mathbf{1}^T \mathbf{v}(i)} \mathbf{I}\right) + 2\beta^2 \mathbf{1}\mathbf{1}^T \quad (40)$$

$$E(\mathbf{v}(i+1)|\mathbf{v}(i)) = \mathbf{G}_{\text{alg}}(\mathbf{v}(i)) \cdot \mathbf{v}(i). \quad (41)$$

Extending the expectation of (41) over $\mathbf{v}(i)$ is nontrivial. We make the approximation

$$E(\mathbf{v}(i+1)|E(\mathbf{v}(i))) \cong \mathbf{G}_{\text{alg}}(E(\mathbf{v}(i))) \cdot E(\mathbf{v}(i)). \quad (42)$$

Applying the approximation (42) iteratively gives

$$E(\mathbf{v}(i)|\mathbf{v}(0)) \cong \prod_{k=0}^{i-1} \mathbf{G}_{\text{alg}}(E(\mathbf{v}(k)|\mathbf{v}(0))) \cdot \mathbf{v}(0). \quad (43)$$

[3]Matrix power iteration yields $\lim_{n\to\infty} \left(\lambda_0^{-n} \mathbf{A}^n \mathbf{b}\right) = (\mathbf{q}_0^H \mathbf{b}) \cdot \mathbf{q}_0$, where λ_0 and $\mathbf{q}_0$ are the principal eigenvalue and eigenvector, respectively, of the Hermitian symmetric matrix $\mathbf{A}$, so that if $\mathbf{q}_0^H \mathbf{b} \neq 0$, then the principal eigenvector results.

For conciseness, the approximation implied by the iterative computation of (43) will be denoted

$$\bar{\mathbf{v}}(i) \equiv \prod_{k=0}^{i-1} \mathbf{G}_{\text{alg}}\left(\bar{\mathbf{v}}\left(k\right)\right) \cdot \mathbf{v}\left(0\right) \cong E\left(\mathbf{v}(i) \middle| \mathbf{v}\left(0\right)\right). \tag{44}$$

Finally, this result is plugged into (14) to provide an estimate of the mean performance metric J

$$E\left(J(i) \middle| \mathbf{v}\left(0\right)\right) \cong \bar{J}(i) \equiv \frac{\mathbf{1}^T \mathbf{\Lambda} \bar{\mathbf{v}}(i)}{\mathbf{1}^T \bar{\mathbf{v}}(i)}. \tag{45}$$

It is worth making note of a degenerate adaptation condition. When the weight vector lies very nearly within a single eigenspace, the gradient vector norm approaches zero. In this condition, the normalization in (39) and (40) approaches zero, and some of the diagonal components of the matrix $\mathbf{G}_{\text{alg}}$ can blow up and approach negative infinity. This would seem to violate the obvious requirement that the resultant eigenenergy vector $\mathbf{v}$ must be all non-negative. However, the update (39) is clearly all non-negative, and the update in (34) is clearly well behaved. Closer examination of the gradient norm given by (15), and the update equation shows that the diagonal entry of $\mathbf{G}$ corresponding to a particular eigenweight can only approach negative infinity as the eigenweight itself approaches zero so that the overall process remains well behaved. However, this degenerate condition does present analytic difficulties.

B. Discussion

The convergence can be visualized by the matrix power iteration formulation of (37), where the weight vector tends toward the principal eigenvector of $\mathbf{R}$, as desired. The update varies from matrix power iteration in three ways:

i) error vector $\mathbf{e}$;
ii) premultiplying matrix that is projected orthogonal to the current $\mathbf{w}$;
iii) scaling of the $\mathbf{R}$ component of the pre-multiplying matrix that is varying because of the update normalization by the norm of the gradient.

The noise introduced by $\mathbf{e}$ is averaged out through multiple iterations, establishing the tradeoff between fast convergence with a large β and a smaller residual weight error with a small β. Since we are not interested in the magnitude of the vector $\mathbf{w}$, we can note that the latter two issues modify the magnitude of the directional change of $\mathbf{w}$ but not the direction. Hence, these do not indicate any problem with a general convergence. While the orthogonal and fixed norm nature of the update indicates the convergence will never settle closer than one update step size, in practice, this effect is overwhelmed by the update noise. The maximum gradient norm is derived in Appendix B, which gives some insight into minimum adaptation rate since this divides the adaptation components of (37) and (40).

In nondegenerate conditions, the diagonal terms of $\mathbf{G}_{\text{alg}}$ are non-negative, and the diagonal element corresponding to the principal eigenvector is largest in magnitude. This causes the desired adaptation toward the principal eigenvector. However, in degenerate situations, some of the diagonal elements can be negative. This indicates a condition where the weight vector lies so close to a single eigenspace that the adaptation, which recall is perpendicular to the weight vector itself and constrained to be unit norm, will pass through the nearby eigenspace causing the weight projection into lesser eigenspaces (with smaller associated eigenvalues) to change sign. One particular example can be visualized if the weight vector is near the principal eigenspace and the update noise $\mathbf{e}$ is ignored. In this situation, the adaptation is forced to be nearly orthogonal to the desired eigenspace (i.e., orthogonal to the weight vector itself) and has a constrained norm. The adaptation would be toward the desired eigenspace but pass through it and invert the other eigenweights, like passing over the north pole and effectively inverting the east-west position. Hence, these degenerate situations would only arise when the distance of the weight vector to an eigenspace is small relative to the step size $\sqrt{2/\pi} \cdot \beta$.

As has been discussed, the Raleigh quotient performance metric contains no local maxima so that there are no inverted bowls into which the algorithm may become trapped. However, gradient algorithms can also stall when the gradient is near zero, as at some local minimum or saddle point (i.e., within a single eigenspace) [17]. For the Raleigh quotient metric with this algorithm, these conditions are degenerate, as previously mentioned. Two factors mitigate against stall problems with this algorithm. First, the normalization of the applied gradient update (which causes the analytic degeneracy) helps to avoid stalling. When the gradient approaches zero, the rate of adaptation of the algorithm will not be reduced since the norm of the expected value of the adjustment in $\mathbf{w}$ is unchanged. This is somewhat similar to normalized-LMS [17], but here, the weight update step size itself is fixed (in expectation). Note that this is the near-degeneracy (near zero gradient) working to benefit the algorithm. Second, the weighted $\mathbf{1}\mathbf{1}^T$ term in the update generator $\mathbf{G}_{\text{alg}}$ represents the introduction of weight noise through the application of the random perturbation vector so that all eigenmodes are excited. This pushes the weight vector state away from those degenerate potential stall inducing locations where the primary eigenweight of $\mathbf{w}$ is near zero. This noisy modal excitation is somewhat analogous to the stall mitigating properties of the leaky-LMS algorithm [17]. Since this induced noise has a greater root-variance than the gradient update magnitude [compare (85) with (86)], it should be expected that it is not possible for these degenerate states to persist or arise to any relevant extant.

C. Numerical Examples and Simulation

The above results have been verified by simulation. The first simulation example shows the convergence of the metric J and eigenweight energies $\mathbf{v}$ for a rank 2 (two path) case with $N_T = 4$ antennas to illustrate the modal convergence and stall avoidance characteristics. Then, examples of the metric J are presented for a rank 1 (one path) case for both $N_T = 2$ and $N_T = 4$ with various values of β to show the convergence times in a comparable environment to the tracking simulations of Section VI.

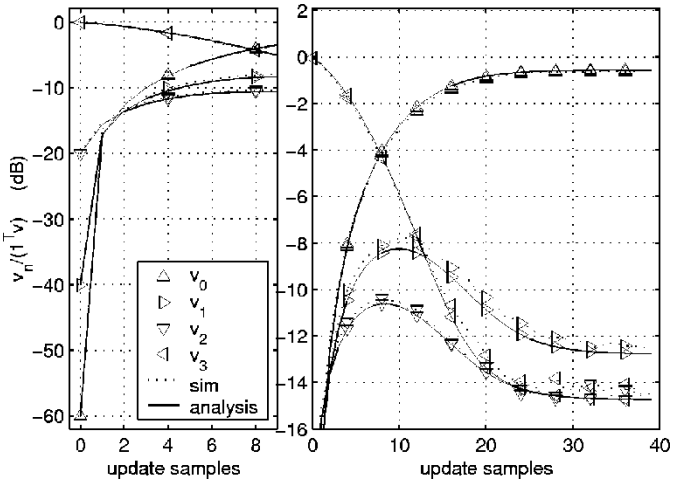

Fig. 2. Convergence of the eigenweight energies $\mathbf{v}$ for first example simulation and analytic from (43). $N_T = 4$, $\beta = 0.1$, $\lambda_0 = 0.7256$, $\lambda_1 = 0.2744$, and $\lambda_2 = \lambda_3 = 0$ (two paths).

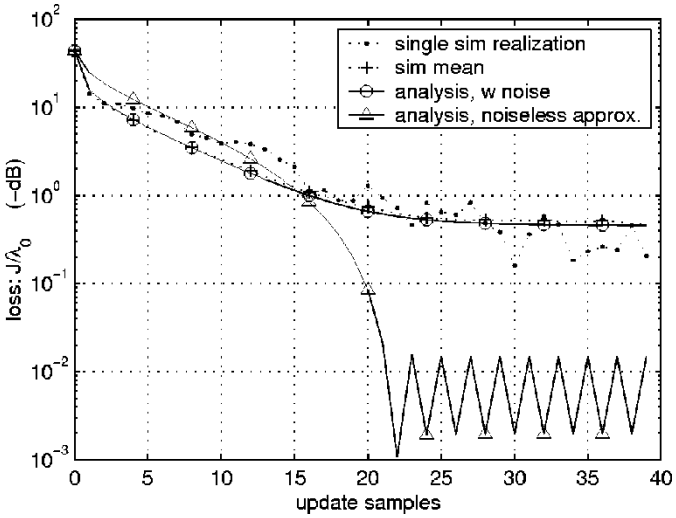

Fig. 3. Convergence of the metric J. First example, simulation, analytic from (14) and (43), analytic with noiseless approximation of (38). $N_T = 4$, $\beta = 0.1$, $\lambda_0 = 0.7256$, $\lambda_1 = 0.2744$, and $\lambda_2 = \lambda_3 = 0$ (two paths).

The first simulation example uses two paths with randomly selected coefficients generated as real Gaussians normalized by the Frobenius norm over both vectors so that $\mathrm{tr}(\mathbf{R})$ is unity.

$$\text{Example 1: } [\mathbf{c}_0 \quad \mathbf{c}_1] = \begin{bmatrix} 0.4325 & -0.2192 \\ 0.0471 & -0.2150 \\ -0.4263 & -0.5518 \\ 0.4371 & 0.1983 \end{bmatrix}. \tag{46}$$

The eigenvalues for the resultant matrix $\mathbf{R}$ are given by

$$[\lambda_0 \quad \ldots \quad \lambda_3] = [0.7256 \quad 0.2744 \quad 0 \quad 0]. \tag{47}$$

The performance metric (9) is dependent on the magnitude $\mathbf{u}$ and independent of the phase of $\mathbf{u}$, and given the post-normalization of (14), the ratios of the entries of $\mathbf{v}$ are relevant, whereas their absolute magnitudes are not. Thus the initial conditions can be captured as the ratios of the elements of $\mathbf{v}$. The initial conditions selected for the simulation represent a very poor state, where the energy of the desired mode v_0 is much smaller than the other modes. In order to demonstrate the performance near a potential stall location, this initial state is near the zero-gradient location at the bottom of the null-space bowl in J. The initial state is defined by

$$v_k(0) = 100 \cdot v_{k-1}(0). \tag{48}$$

The simulation is performed with $\beta = 0.1$. Fig. 2 shows the convergence of the eigenweight energies $\mathbf{v}$ toward dominance of the principal eigenmode, showing an average over 500 simulations against the analytic result of (43). The simulated and analytic results are nearly indistinguishable. The stall mitigation from the excitation of every eigenmode is visible in the first step at time index 1; both v_0 and v_1 jump (in expectation) from near zero to approximately $2\beta^2 \cdot \mathbf{1}^T\mathbf{v} = 0.02 \cdot \mathbf{1}^T\mathbf{v}$ (-17 dB). From there, v_0 grows to dominate the weight vector $\mathbf{w}$ by the eighth update sample and, finally, to about -0.6 dB of the weight energy ($\mathbf{1}^T\mathbf{v}$) in steady state. The lesser eigenspace energy v_1 converges toward -12.8 dB, whereas the two nullspace energies v_2 and v_3 both recede to about -14.7 dB of the weight energy. The value of J for these cases is shown in Fig. 3 as a loss from the ideal $J_{opt} = \lambda_0$. For comparison, curves are generated for a single realization of the simulated metric J, the mean over the 500 simulations, the analysis of (14) and (43) with noisy update, and the noiseless approximation computing the update of (38) with $\mathbf{e} = \mathbf{0}$. The initial conditions provide a starting point with a -44.2 dB loss from the optimal weight vector. The noiseless update curve shows a weak initial adaptation as the eigenmode excitation due to the noisy update does not occur, and the weights stay stalled near the zero-gradient at the nullspace. This noiseless curve eventually catches up and has the best steady-state performance with the weights bouncing back and forth across the principal eigenspace. Analysis of (14) and (43) with noisy update is nearly identical to the simulation mean, and the weights settle to a loss of about -0.5 dB from the optimal.

The second and third convergence examples are performed with a rank one channel (single path). In this case, the channel vector is arbitrary, and $c_0 = 1$, $c_{k\neq 0} = 0$ is selected. The weight vector is again initialized according to (48) (a pessimistic condition). The value of β is varied to provide the convergence time versus adaptation rate, and the metric J is shown for simulation and analysis in Figs. 4 and 5. As one would expect, a larger β provides a faster convergence with a noisier final result.

VI. Tracking Performance

A. Reduced-Rank Analysis

A tracking analysis for a simplified system is presented, with a single time resolvable path ($L = 1$) undergoing AR1 Raleigh fading, independent across all antennas. Since the time varying gain matrix (3) is now of rank one, the weight vector can be decomposed into two constituent eigenvectors of $\mathbf{R}$ if the eigenvectors are properly selected.

We now have some notes on notation. In Appendix C, a prime mark is used to distinguish the reduced-rank notation from the full-rank notation, but this is omitted in the following to avoid notational clutter. In this section, the vector $\mathbf{v}$ refers to the 2×1 reduced-rank formulation defined below. In addition,

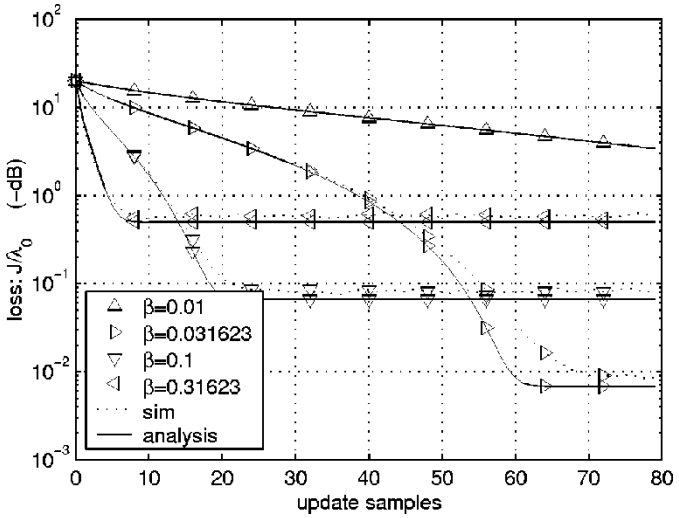

Fig. 4. Convergence of the metric J for the second example, simulation and analytic from (14) and (43). $N_T = 2$, $\lambda_0 = 1$, and $\lambda_1 = 0$ (one path).

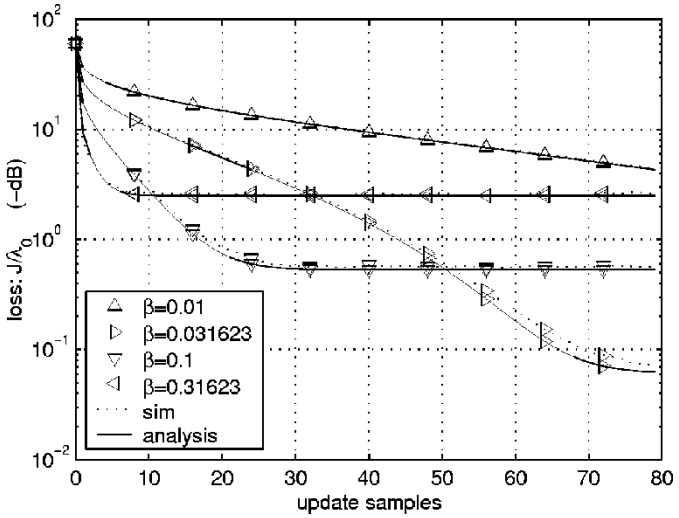

Fig. 5. Convergence of the metric J for the third example, simulation and analytic from (14) and (43). $N_T = 4$, $\lambda_0 = 1$, and $\lambda_1 = \lambda_2 = \lambda_3 = 0$ (one path).

the overbar distinguishing the update in expectation $\bar{\mathbf{v}}$ from the true eigenenergy vector $\mathbf{v}$ is omitted, and all analysis is subject to the expectation approximations applied to obtain (44) and (45).

The first eigenvector is the normalized channel vector. All other eigenvalues are zero and their eigenvectors are arbitrary within the nullspace (orthogonal to the channel vector). Hence, a specific eigenvector set can be imposed wherein only the first two eigenweights of the weight vector $\mathbf{w}$ are nonzero, and the second eigenvector is given by the normalized projection of $\mathbf{w}$ orthogonal to $\mathbf{c}$. The first eigenvector represents the desired weight vector, and the second represents the error vector. The relevant eigenvectors are

$$\mathbf{q}_0 = \frac{\mathbf{c}}{\|\mathbf{c}\|} \tag{49}$$

$$\mathbf{q}_1 = \frac{\left(\mathbf{I} - \frac{\mathbf{c}\mathbf{c}^H}{\|\mathbf{c}\|^2}\right)\mathbf{w}}{\left\|\left(\mathbf{I} - \frac{\mathbf{c}\mathbf{c}^H}{\|\mathbf{c}\|^2}\right)\mathbf{w}\right\|}. \tag{50}$$

Then, the weight vector can be represented in terms of these eigenvectors as

$$\underset{2\times 1}{\mathbf{u}} = \underset{2\times N_T}{\mathbf{Q}}{}^H \underset{N_T\times 1}{\mathbf{w}}. \tag{51}$$

Again, the vector $\mathbf{v}$ is comprised of the squared magnitudes of $\mathbf{u}$.

$$\mathbf{v} = \left[\ \frac{\left|\mathbf{c}^H\mathbf{w}\right|^2}{\|\mathbf{c}\|^2} \quad \left\|\left(\mathbf{I} - \frac{\mathbf{c}\mathbf{c}^H}{\|\mathbf{c}\|^2}\right)\mathbf{w}\right\|^2\ \right]^T. \tag{52}$$

The performance metric is

$$J = \frac{v_0}{v_0 + v_1} \cdot \|\mathbf{c}\|^2. \tag{53}$$

v_0 gives the portion of the "energy" of $\mathbf{w}$ that lies in the desired direction, which is received by the mobile, and v_1 is the portion of the "energy" of $\mathbf{w}$, which is in error and cannot be received by the mobile.

B. Dynamic System Model

The fading channel is defined as a first-order autoregressive (AR1) complex Gaussian process with a zero mean complex Gaussian stimulus $\mathbf{x}$.

$$\mathbf{c}(i+1) = a\mathbf{c}(i) + \mathbf{x}(i). \tag{54}$$

The channel is uncorrelated across the antennas so that the autocorrelation matrix of $\mathbf{x}$ is

$$\mathbf{\Phi}_{\mathbf{xx}} \equiv E\left(\mathbf{x}\mathbf{x}^H\right) = 2\sigma_x^2\mathbf{I}. \tag{55}$$

The parameter s is used in the analysis, which is defined by the expectation of the inverted scaled chi-squared distribution given as follows:

$$s \equiv E\left(\frac{2\sigma_x^2}{a^2\|\mathbf{c}(i)\|^2}\right) = \frac{1-a^2}{a^2\cdot(N_T-1)}. \tag{56}$$

C. Dynamic Performance Analysis

The tracking performance will be considered by deriving expressions for a transition matrix applied to the eigenmodal vector representation of the weight vector $\mathbf{w}$. This transition matrix incorporates the effect of both the algorithm update and the channel change for one time step. Each step is shown to be a multiplicative transformation of the vector $\mathbf{v}$ with the following approximations:

1) independence of the channel vector norm from update to update;
2) second-order Taylor approximation of the fading channel update on v_0.

The first assumption will more closely apply for large numbers of antennas as the channel norm approaches a constant in expectation. The second is reasonable because the expectation of odd order terms is zero, and the fourth-order term will be small for small values of $(1-a)$.

The step update of (34), which assumed perfect receive channel estimation and no feedback bit errors, is translated into a simplified notation for this rank one situation (only one resolvable path). As shown in Appendix C, the reduced-rank

update formula can be expressed with a simplified 2×2 generator matrix (the prime mark distinguishing reduced rank notation in Appendix C is omitted) as follows:

$$\gamma(i) \equiv \sqrt{\frac{v_0(i)}{v_1(i)}} \tag{57}$$

$$\begin{aligned}\mathbf{G}_{\text{alg}}(\gamma(i)) \equiv & \left(1 - 2\sqrt{\frac{2}{\pi}} \cdot \beta\gamma(i)\right)\mathbf{I} \\ & + 2\sqrt{\frac{2}{\pi}} \cdot \beta\left(\gamma(i) + \gamma^{-1}(i)\right)\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \\ & + 2\beta^2 \begin{bmatrix} 1 & 1 \\ N_T - 1 & N_T - 1 \end{bmatrix}. \end{aligned} \tag{58}$$

As shown in Appendix D, defining s from (56) as a function of the AR1 parameter a, the second-order Taylor approximation of the effect of an AR1 increment on the eigenenergies $\mathbf{v}$ is given by

$$\mathbf{G}_{\text{chan}} = \left(\mathbf{I} + s \cdot \begin{bmatrix} 1 - N_T & 1 \\ N_T - 1 & -1 \end{bmatrix}\right). \tag{59}$$

The measurement that leads to the algorithm update is based on the channel iterated one step from its value of the last time slot. Hence, the effect of one iteration is captured by considering the operation of first the algorithm update $\mathbf{G}_{\text{alg}}$ followed by the channel update $\mathbf{G}_{\text{chan}}$. This is described by the matrix $\mathbf{G}$ in (60), shown at the bottom of the page.

The effect of an iteration of both the algorithm and channel change is then approximated as

$$\underset{2\times 1}{\mathbf{v}}(i+1) \cong \underset{2\times 2}{\mathbf{G}}(\gamma(i)) \underset{2\times 1}{\mathbf{v}}(i). \tag{61}$$

If a steady-state solution for (61) exists, then the steady-state vector $\mathbf{v}$ must be one of the eigenvectors of $\mathbf{G}$. The eigenvectors of this 2×2 matrix are given by

$$\mathbf{r} = \begin{bmatrix} \frac{+G_{0,0} - G_{1,1} \pm \sqrt{(G_{0,0} - G_{1,1})^2 + 4G_{0,1}G_{1,0}}}{2} \\ G_{1,0} \end{bmatrix}. \tag{62}$$

Hence, in the steady state

$$\gamma^2 = \frac{+G_{0,0} - G_{1,1} \pm \sqrt{(G_{0,0} - G_{1,1})^2 + 4G_{0,1}G_{1,0}}}{2G_{1,0}}. \tag{63}$$

This gives rise to a fourth-order polynomial in γ.

$$\begin{aligned}\gamma^4\left(\left(s + 2\beta^2\right)(N_T - 1)\right) + \gamma^3 & \left(2\sqrt{\frac{2}{\pi}}\beta \cdot (sN_T - 1)\right) \\ & + \gamma^2\left((N_T - 2)\left(2\beta^2 + s\right)\right) \\ & + \gamma\left(2\sqrt{\frac{2}{\pi}}\beta \cdot (sN_T - 1)\right) \\ & - 2\beta^2 - s = 0. \end{aligned} \tag{64}$$

Two of the roots of this equation are given by $\gamma = \pm j$. These roots are not of direct interest for this problem since γ^2 is positive. The remaining two roots are given by (65), shown at the bottom of the page.

Of these two roots, one is negative and thus does not satisfy the original definition (57) [in (58) and (60), γ is non-negative]. Hence, the additive root is the only root consistent with the problem formulation and gives the solution to the steady-state value of γ. This gives the steady-state solution for the expected correlation value v_0 by using

$$\left.\frac{v_0}{v_0 + v_1}\right|_{\text{steady state}} = \frac{\gamma^2|_{\text{steady state}}}{1 + \gamma^2|_{\text{steady state}}}. \tag{66}$$

D. Discussion and Numerical Results

Equation (65) provides a mechanism for evaluating the effectiveness of an adaptation rate β for a fading rate of the AR1 process. It is of interest to note that with $\beta = 0$

$$\gamma^2|_{\text{steady state},\beta=0} = \frac{1}{N_T - 1}. \tag{67}$$

This is a confirmation of the common sense solution in this case, where $1/N_T$ of the weight vector energy lies within span of the channel vector and the remainder is in its nullspace.

To confirm the applicability of this analysis, simulations were performed for $N_T = 2$ (Figs. 6–9) and $N_T = 4$ (Figs. 10–13)

$$\begin{aligned}\mathbf{G}(\gamma(i)) \equiv & \mathbf{G}_{\text{chan}}\mathbf{G}_{\text{alg}}(\gamma(i)) \\ = & \begin{bmatrix} 1 + 2\sqrt{\frac{2}{\pi}} \cdot \beta\gamma^{-1}(i)\left(1 - s\left(N_T - 1\right)\right) + 2\beta^2 - s\left(N_T - 1\right) & 2\beta^2 + s - 2\sqrt{\frac{2}{\pi}} \cdot \beta s\gamma(i) \\ \left(s + 2s\sqrt{\frac{2}{\pi}} \cdot \beta\gamma^{-1}(i) + 2\beta^2\right)\left(N_T - 1\right) & 1 + 2 \cdot (s-1)\sqrt{\frac{2}{\pi}} \cdot \beta\gamma(i) + 2\beta^2\left(N_T - 1\right) - s \end{bmatrix} \end{aligned} \tag{60}$$

$$\gamma|_{\text{steady state}} = \frac{\sqrt{\frac{2}{\pi}}\beta \cdot (1 - sN_T) \pm \sqrt{4\left(N_T - 1\right)\beta^4 + \left(\frac{2}{\pi}sN_T\left(sN_T - 2\right) + \frac{2}{\pi} + 4s\left(N_T - 1\right)\right)\beta^2 + \left(N_T - 1\right)s^2}}{\left(s + 2\beta^2\right)\left(N_T - 1\right)} \tag{65}$$

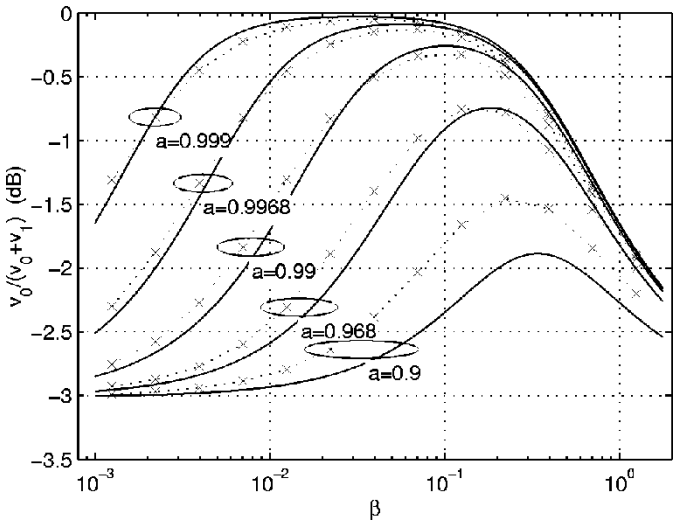

Fig. 6. Mean correlation v_0, $N_T = 2$, simulated and from (65) and (66).

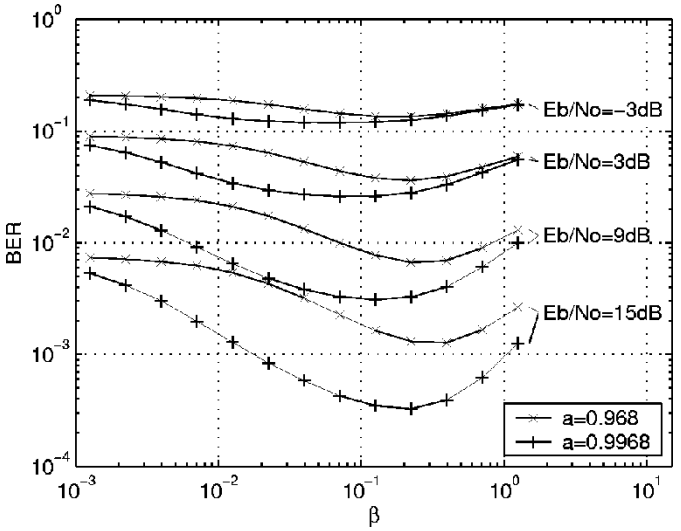

Fig. 7. BERs for $N_T = 2$, $a = [0.968, 0.9968]$.

antennas with fading uncorrelated across the antennas as per (55) over a variety of values of the AR1 parameter a and the adaptation rate β. In addition, BERs were simulated in order to compare the BER performance with the performance of the vector "cross correlation energy" between $\mathbf{w}$ and $\mathbf{c}$ given by $(v_0/(v_0+v_1))$. The simulations implemented the receiver's decision and the feedback channel with no errors.

The simulation results are compared to the analysis in Figs. 6 and 10 for two and four antennas, respectively. Both figures show that the simulated vector correlation $(v_0/(v_0+v_1))$ has a very good match with the analysis for all fading rates except $a = 0.9$, for which we note that the AR process is approaching the limit of (109). Comparing these figures with Figs. 7 and 11, we see that the analysis provides a good prediction of the best value of β for minimizing BERs. The optimal β versus fading frequency $(1-a)$ from both analysis and BER simulations is shown in Fig. 9 for $N_T = 2$ and Fig. 13 for $N_T = 4$. For low signal-to-noise ratios (SNRs), the analysis is a very good predictor of the optimal β for minimizing BER as power delivery is important in this condition. For higher SNRs, it is apparent

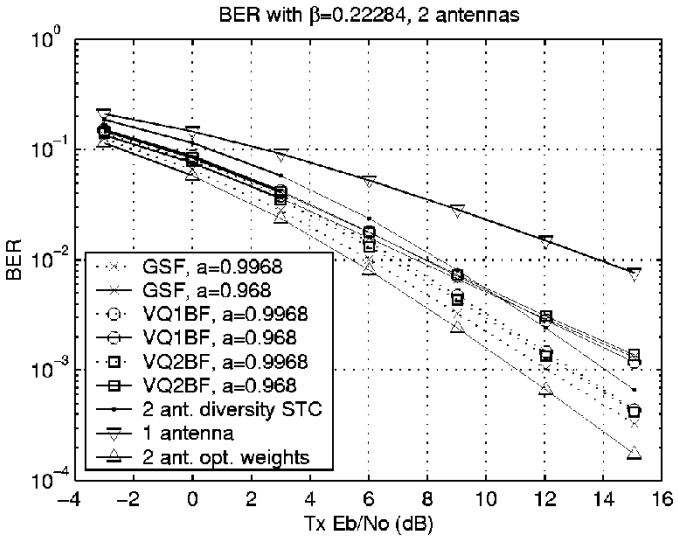

Fig. 8. Simulated BER for $N_T = 2$, $\beta = 0.1419$, and $a = [0.968, 0.9968]$, compared with optimal weights, STC, and vector quantization.

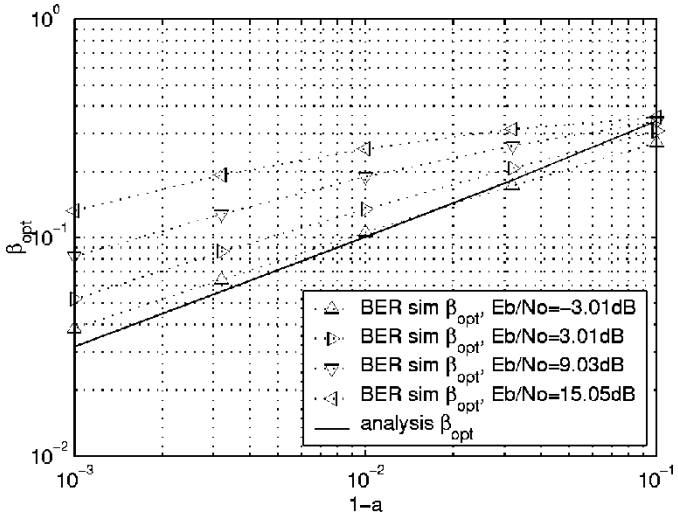

Fig. 9. Optimal values of β versus fading frequency $(1-a)$ from simulation and analysis $N_T = 2$.

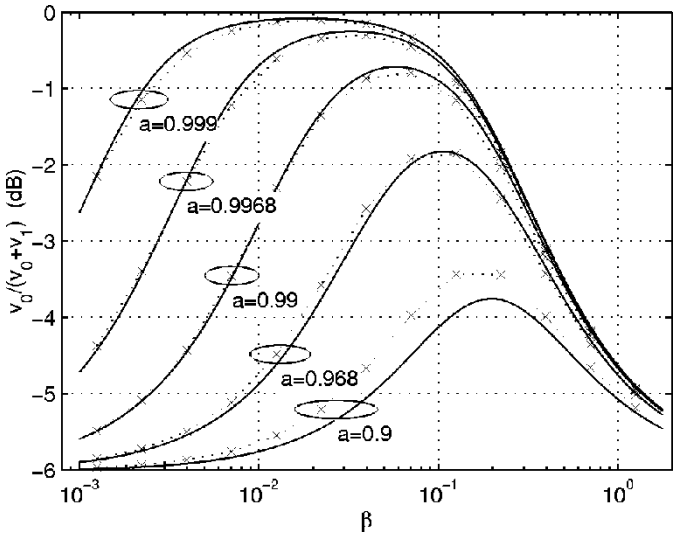

Fig. 10. Mean correlation v_0, $N_T = 4$, simulated and from (65) and (66).

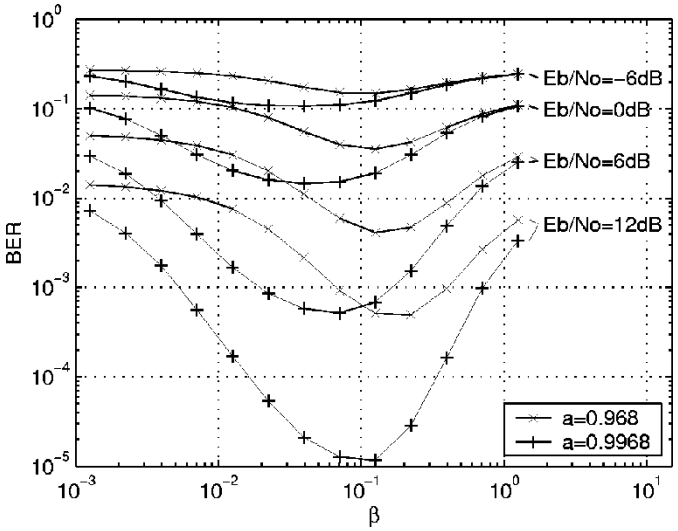

Fig. 11. BERs for $N_T = 4$, $a = [0.968, 0.9968]$.

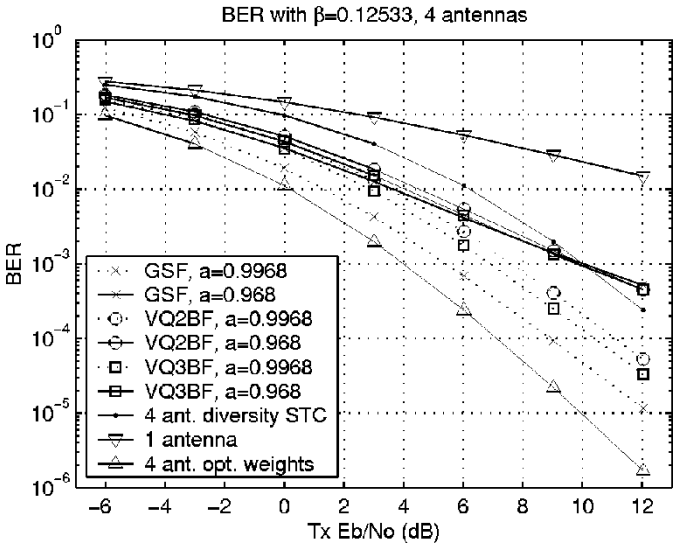

Fig. 12. Simulated BER for $N_T = 4$, $\beta = 0.079\,79$, and $a = [0.968, 0.9968]$, compared with optimal weights, STC, and vector quantization.

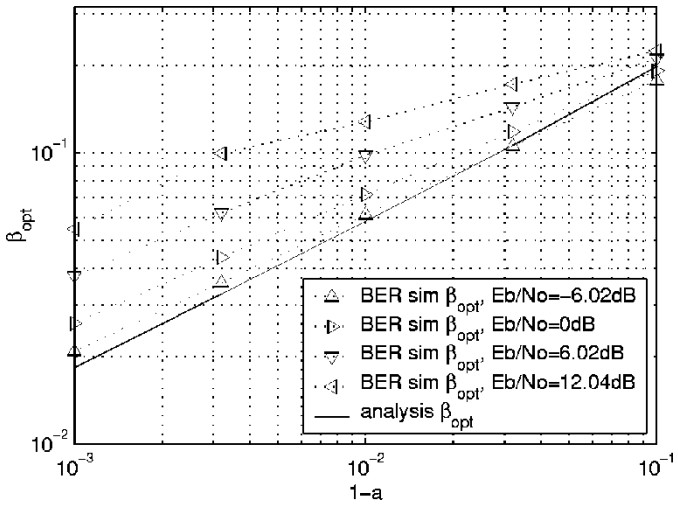

Fig. 13. Optimal values of β versus fading frequency $(1 - a)$ from simulation and analysis $N_T = 4$.

that minimizing BER requires β to be larger than is given by maximizing the analytic result for J; in this condition, fading diversity becomes more important than simple mean power delivery, and BER is best minimized by prioritizing moving the weights away from the nullspace quickly over ensuring optimal power delivery when the channel is not in a deep fade.

BERs from the simulation of the gradient sign feedback (GSF) are shown for two of the fading rates in Figs. 8 and 12. For comparison, these figures also include the analytic performance of diversity space time codes (STC) for a single Rx antenna and simulated performance of vector quantization code book selection feedback, wherein the receiver provides feedback selecting which of several weight vectors is best [14]. The space time code performance is evaluated as a loss from optimal weight adaptation for diversity codes with no coding gain, as in [6] for two Tx antennas. For four Tx antennas, such codes are a lower bounding BER abstraction as they do not exist without bandwidth expansion [22], but approximations do exist [7]. For two antennas, VQ1BF is vector quantization with one bit feedback selecting which antenna should transmit (second-order selection diversity), and VQ2BF is a 2-bit feedback selecting a phase rotation of $\langle 0, \pi/2, \pi, 3\pi/2 \rangle$ for the second weight. For four antennas, VQ2BF is a 2-bit feedback selecting which of the four antennas should transmit (fourth-order selection diversity), and VQ3BF is a 3-bit feedback selecting a phase rotation of $\langle 0, \pi \rangle$ for weights w_1, w_2, and w_3. For 2- or 3-bit feedback, the feedback decision interval is lengthened so that the feedback data rate is unchanged.

BERs are plotted versus Tx E_b/N_0 normalized to the Rx E_b/N_0 for one Tx antenna so that the 3.01- and 6.02–dB array gain from two- or four-antenna systems can be seen. For slow fading, the algorithm performs close to the theoretic limit and outperforms the vector quantization feedback approaches. The gradient adaptation provides a simple recursive update that uses the history to provide better resolution in the weights than is available from vector quantization, and the gain from beam forming provides better performance than the diversity space time codes. For faster fading, the gradient approach gives similar performance to vector quantization, but both feedback approaches are outperformed by diversity space time coding, which does not require the transmitter to adapt to the time varying channel.

VII. Conclusion

A new gradient sign algorithm for transmit antenna array adaptation has been defined, and the convergence and tracking performance of the algorithm has been analyzed. The algorithm makes use of gradient sign feedback from the receiver to generate a coarse gradient estimate used by the transmitter to recursively adjust the transmit weights. The mechanism employed by the receiver to generate the feedback is simple and can be employed with no knowledge of the specifics of the transmitter antenna algorithm, i.e., the receiver need not know how many antennas are employed or exactly how the update is performed. The convergence and tracking behavior was found to match the analysis through simulation verification, and the algorithm is found through simulations of bit error

rate to outperform previously proposed antenna weight vector selection feedback and diversity space time coding algorithms for moderate fading speeds.

APPENDIX A
SIGNED GRADIENT EXTRACTION

This appendix derives the first and second moments of the signed gradient algorithm update.

1) Theorem 1: For a nonrandom $\mathbf{g}$ and a zero mean complex Gaussian vector $\mathbf{p}$ with autocorrelation $2\mathbf{I}$, define the decision vector $\mathbf{x}$ and error vector $\mathbf{e}$ as follows:

$$\mathbf{x} \equiv \text{sign}\left(\text{Re}\left(\mathbf{p}^H\mathbf{g}\right)\right)\cdot\mathbf{p} \tag{68}$$

$$\mathbf{e} \equiv \mathbf{x} - E(\mathbf{x}). \tag{69}$$

Then, the update is characterized by the following first and second moments:

$$E(\mathbf{x}) = \sqrt{\frac{2}{\pi}}\cdot\frac{\mathbf{g}}{\|\mathbf{g}\|} \tag{70}$$

$$E\left(\mathbf{e}\mathbf{e}^H\right) = 2\mathbf{I} - \frac{2}{\pi}\cdot\frac{\mathbf{g}\mathbf{g}^H}{\|\mathbf{g}\|^2}. \tag{71}$$

Proof of Theorem 1: Consider the vectors generated as the serialized real vectors containing the real and imaginary components of $\mathbf{p}$, $\mathbf{x}$, and $\mathbf{g}$, which will be denoted by a prime

$$\mathbf{p}' \equiv \begin{bmatrix}\text{Re}(\mathbf{p})\\ \text{Im}(\mathbf{p})\end{bmatrix} \tag{72}$$

$$\mathbf{g}' \equiv \begin{bmatrix}\text{Re}(\mathbf{g})\\ \text{Im}(\mathbf{g})\end{bmatrix} \tag{73}$$

$$\mathbf{x}' \equiv \begin{bmatrix}\text{Re}(\mathbf{x})\\ \text{Im}(\mathbf{x})\end{bmatrix}. \tag{74}$$

Hence

$$\mathbf{x}' = \begin{cases}+\mathbf{p}', & \mathbf{p}'^T\mathbf{g}' \geq 0\\ -\mathbf{p}', & \mathbf{p}'^T\mathbf{g}' < 0.\end{cases} \tag{75}$$

Let

$$y \equiv \frac{\sum_{n=0,n\neq k}^{2N-1} p'_n g'_n}{|g_k|}. \tag{76}$$

Recalling that $\mathbf{p}$ is complex Gaussian with autocorrelation $2\mathbf{I}$, p'_n is real Gaussian of unit variance so that

$$E(y) = 0 \tag{77}$$

$$\sigma^2 \equiv \text{var}(y) = \frac{\sum_{n=0,n\neq k}^{2N-1} g_n'^2}{g_k'^2}. \tag{78}$$

Then, the kth element of the update vector is given by

$$x'_k(p'_k, y) = \begin{cases}+p'_k, & \text{sign}(g'_k)\cdot p'_k \geq -y\\ -p'_k, & \text{sign}(g'_k)\cdot p'_k < -y.\end{cases} \tag{79}$$

Noting that y and p'_k are statistically independent, the expectation of the kth element of the update vector is

$$E(x'_k) = \frac{1}{2\pi\sigma}\int_{p=-\infty}^{\infty}\int_{y=-\text{sign}(g'_k)\cdot p}^{\infty} pe^{-p^2/2}e^{-y^2/2\sigma^2}dydp - \frac{1}{2\pi\sigma}\int_{p=-\infty}^{\infty}\int_{y=-\infty}^{-\text{sign}(g_k)\cdot p} pe^{-p^2/2}e^{-y^2/2\sigma^2}dydp. \tag{80}$$

With some manipulation, this is found to be

$$E(x'_k) = \frac{\text{sign}(g'_k)}{\pi\sigma}\int_{p=0}^{\infty}\int_{y=-p}^{p} pe^{-p^2/2}e^{-y^2/2\sigma^2}dydp. \tag{81}$$

Substituting $\sigma z = y$ and folding the even function of y into the positive half plane

$$E(x'_k) = \frac{2\text{sign}(g'_k)}{\pi}\int_{p=0}^{\infty}\int_{z=0}^{p/\sigma} pe^{-p^2/2}e^{-z^2/2}dzdp. \tag{82}$$

Substituting $2t = p^2$ and swapping the order of integration gives

$$E(x'_k) = \frac{2\text{sign}(g'_k)}{\pi}\int_{z=0}^{\infty} e^{-z^2/2}\int_{t=\sigma^2z^2/2}^{\infty} e^{-t}dzdt = \frac{2\text{sign}(g'_k)}{\pi}\int_{z=0}^{\infty} e^{-z^2(1+\sigma^2)/2}dz. \tag{83}$$

Considering that this is now an integral over a Gaussian PDF, the result can be found simply as

$$E\left(x'_k|_{u(p)=p}\right) = \sqrt{\frac{2}{\pi}}\frac{g'_k}{\sqrt{\sum_{n=0}^{2N} g_n'^2}}. \tag{84}$$

Extending the result of (84) to all elements of $\mathbf{x}$, the final complex vector solution is

$$E(\mathbf{x}) = \sqrt{\frac{2}{\pi}}\frac{\mathbf{g}}{\|\mathbf{g}\|}. \tag{85}$$

The error vector $\mathbf{e}$(69) clearly has zero mean, and its autocorrelation is simply determined by

$$E\left(\mathbf{e}\mathbf{e}^H\right) = E\left(\left(\mathbf{x} - \sqrt{\frac{2}{\pi}}\frac{\mathbf{g}}{\|\mathbf{g}\|}\right)\left(\mathbf{x} - \sqrt{\frac{2}{\pi}}\frac{\mathbf{g}}{\|\mathbf{g}\|}\right)^H\right) = 2\mathbf{I} - \frac{2}{\pi}\frac{\mathbf{g}\mathbf{g}^H}{\|\mathbf{g}\|^2}. \tag{86}$$

Q.E.D.

APPENDIX B
MAXIMUM GRADIENT NORM

In order to bound the adaptation performance, it is useful to determine the maximum realizable norm of the gradient vector subject to the constraint $\|\mathbf{w}\| = 1$. Equivalently, we can determine the maximum of the norm squared. That is, we wish to find the maximum

$$\max_{\mathbf{w}}\left(\|\mathbf{g}(\mathbf{w})\|^2 \;\middle|\; \|\mathbf{w}\| = 1\right). \tag{87}$$

From the decomposition of (10) using the constraint of (87)

$$\|\mathbf{g}\|^2_{\|\mathbf{w}\|=1} = 4\sum_{n=0}^{N-1} |u_n|^2 \lambda_n^2 - 4\left(\sum_{n=0}^{N-1} |u_n|^2 \lambda_n\right)^2. \quad (88)$$

To maximize this quantity with the weight norm constraint, the method of Lagrange multipliers is applied. Since the gradient norm is a function only of the magnitude of the eigenweight values u_n, we consider the function parameterized by the real quantity $|u_n|$, and for simplicity, we denote this as u_n (the eigenvectors can be rotated to accomplish this). Using a Lagrange multiplier of γ, we find the partial derivatives

$$\begin{aligned}\frac{\partial}{\partial u_k}&\left(4\sum_{n=0}^{N-1} u_n^2\lambda_n^2 - 4\left(\sum_{n=0}^{N-1} u_n^2\lambda_n\right)^2 + 4\gamma\left(\sum_{n=0}^{N-1} u_n^2 - 1\right)\right)\\ &=8u_k\left(\lambda_k^2 - 2\lambda_k\left(\sum_{n=0}^{N-1} u_n^2\lambda_n\right) + \gamma\right). \quad (89)\end{aligned}$$

One way to zero all of the derivatives would be to zero each of the eigenweights that correspond to nonzero eigenvalues, except that this violates the unit norm requirement if there is no null space. Beyond this we have only 2 degrees of freedom available in selecting the values u_k: the Lagrange multiplier γ and the summation $\left(\sum_{n=0}^{N-1} u_n^2\lambda_n\right)$. Hence, any inflection points in the gradient norm are achieved when one of the following is satisfied.

a) The weight vector falls in the null space of $\mathbf{R}$.
b) All but one of the eigenweights is zero.
c) All but two of the eigenweights is zero, and the remaining two eigenweights i and k conform to the relations of (90) and (91) below.

In the case of a) or b), the gradient is zero, and a minimum of the gradient norm is attained. In the case of c), a local maximum of the gradient norm is attained. The requirement for a local maximum from c) is

$$\lambda_k^2 - 2u_k^2\lambda_k^2 - 2u_i^2\lambda_i\lambda_k + \gamma = 0 \quad (90)$$

$$\lambda_i^2 - 2u_k^2\lambda_k\lambda_i - 2u_i^2\lambda_i^2 + \gamma = 0. \quad (91)$$

The solution to these two equations is given by

$$u_k^2 = v_k = \frac{1}{2} \quad (92)$$

$$u_i^2 = v_i = \frac{1}{2}. \quad (93)$$

Hence, we see that the inflection points of the gradient norm occur when the weight vector is comprised of equal contributions from only two eigenvectors. Plugging this into (88), the gradient norm at the inflection point is given by

$$\|\mathbf{g}\|^2_{\|\mathbf{w}\|=1} = (\lambda_k - \lambda_i)^2. \quad (94)$$

The global maximum value of the gradient norm is attained when the two nonzero eigenweights are the maximal and minimal modes so that

$$\max_{\mathbf{w}}(\|\mathbf{w}\| \cdot \|\mathbf{g}(\mathbf{w})\|) = \lambda_0 - \lambda_{N-1}. \quad (95)$$

APPENDIX C
REDUCED-RANK REPRESENTATION OF THE GRADIENT ALGORITHM UPDATE

We wish to represent the update generator matrix $\mathbf{G}_{\text{alg}}$ (40) in terms of the reduced-rank eigendecomposition described in Section VI-A for the rank one (single path) channel condition. In this condition, the first eigenvalue of $\mathbf{R}$ is given by the channel vector norm.

$$\lambda_0 = \|\mathbf{c}\|^2. \quad (96)$$

All other eigenvalues of $\mathbf{R}$ are zero. Hence, with some straightforward algebraic simplification, the algorithm update matrix (40) reduces to

$$\begin{aligned}\mathbf{G}_{\text{alg}}(\mathbf{v}(i)) = &\left(1 - 2\sqrt{\frac{2}{\pi}}\cdot\beta\sqrt{\frac{\frac{v_0}{\mathbf{1}^T\mathbf{v}}}{1-\frac{v_0}{\mathbf{1}^T\mathbf{v}}}}\right)\cdot\mathbf{I}\\ &+ 2\sqrt{\frac{2}{\pi}}\cdot\beta\sqrt{\frac{1}{\frac{v_0}{\mathbf{1}^T\mathbf{v}}\left(1-\frac{v_0}{\mathbf{1}^T\mathbf{v}}\right)}}\\ &\cdot\begin{bmatrix}1 & \mathbf{0}_{1\times(N_T-1)}\\ \mathbf{0}_{(N_T-1)\times 1} & \mathbf{0}_{(N_T-1)\times(N_T-1)}\end{bmatrix}\\ &+ 2\beta^2\mathbf{1}\mathbf{1}^T. \quad (97)\end{aligned}$$

Relating the reduced-rank variable γ (57) to $\mathbf{v}$ ($\mathbf{v}$ has dimension $N\times 1$) for the rank 1 channel

$$\frac{v_0}{\mathbf{1}^T\mathbf{v}} = \frac{\gamma^2}{1+\gamma^2}. \quad (98)$$

Inserting (98) into (97) with some algebraic simplification provides

$$\begin{aligned}\mathbf{G}_{\text{alg}}(\mathbf{v}(i)) \equiv &\left(1 - 2\sqrt{\frac{2}{\pi}}\beta\cdot\gamma\right)\cdot\mathbf{I}\\ &+ 2\sqrt{\frac{2}{\pi}}\cdot\beta\left(\gamma+\gamma^{-1}\right)\cdot\begin{bmatrix}1 & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix} + 2\beta^2\mathbf{1}\mathbf{1}^T. \quad (99)\end{aligned}$$

Define the reduced-rank 2×1 vector with a prime, $\mathbf{v}'$. Then

$$\mathbf{v}' = \begin{bmatrix} v_0 \\ \sum_{n\neq 0} v_n\end{bmatrix} \quad (100)$$

and the update from (99) is (101), shown at the bottom of the next page, where $\mathbf{G}'_{\text{alg}}(\gamma)$ is of dimension 2×2, and

$$\begin{aligned}\mathbf{G}'_{\text{alg}}(\gamma(i)) \equiv &\left(1 - 2\sqrt{\frac{2}{\pi}}\beta\cdot\gamma(i)\right)\cdot\mathbf{I}\\ &+ 2\sqrt{\frac{2}{\pi}}\cdot\beta\left(\gamma(i)+\gamma^{-1}(i)\right)\\ &\cdot\begin{bmatrix}1 & 0\\ 0 & 0\end{bmatrix} + 2\beta^2\mathbf{1}\mathbf{1}^T. \quad (102)\end{aligned}$$

Note that the expression is simplified by parameterizing $\mathbf{G}'_{\text{alg}}$ by the single parameter γ rather than by $\mathbf{v}'$, although the latter would be the equivalent to the portrayal of the full rank matrix in (40).

APPENDIX D
DERIVATION OF AR1 UPDATE SECOND–ORDER TAYLOR APPROXIMATION

For the transition of $\mathbf{v}$ due to the time varying channel, a second-order Taylor approximation of the incremental step due to the new stimulus $\mathbf{x}$ is applied. The nomenclature of this appendix is 2×1 reduced-rank vector $\mathbf{v}$ (52). The approximation is given by the gradient $\mathbf{h}$ and Hessian $\mathbf{H}$ of v_0 with respect to $\mathbf{c}$ so that

$$v_0(i+1) \cong v_0(i) + \frac{\mathbf{h}(i)^H \mathbf{x}(i+1) + \mathbf{x}(i+1)^H \mathbf{h}(i)}{2a} + \frac{\mathbf{x}(i)^H \mathbf{H}(i) \mathbf{x}(i)}{2a^2}. \quad (103)$$

Ignoring the time index for the moment and considering differentiating v_0 in (52), the derivatives are

$$\mathbf{h} = 2\frac{\mathbf{w}\mathbf{w}^H\mathbf{c}}{\|\mathbf{c}\|^2} - 2\frac{\mathbf{c}^H\mathbf{w}\mathbf{w}^H\mathbf{c}}{\|\mathbf{c}\|^4}\mathbf{c} \quad (104)$$

$$\mathbf{H} = 2\frac{1}{\mathbf{c}^H\mathbf{c}}\mathbf{w}\mathbf{w}^H - 4\frac{\mathbf{c}^H\mathbf{w}}{(\mathbf{c}^H\mathbf{c})^2}\mathbf{c}\mathbf{w}^H - 4\frac{\mathbf{w}^H\mathbf{c}}{(\mathbf{c}^H\mathbf{c})^2}\mathbf{w}\mathbf{c}^H + 8\frac{\mathbf{c}^H\mathbf{w}\mathbf{w}^H\mathbf{c}}{(\mathbf{c}^H\mathbf{c})^3}\mathbf{c}\mathbf{c}^H - 2\frac{\mathbf{c}^H\mathbf{w}\mathbf{w}^H\mathbf{c}}{(\mathbf{c}^H\mathbf{c})^2}\mathbf{I}. \quad (105)$$

Incorporating the eigenweight energy nomenclature $\mathbf{v}$ where convenient, the Hessian is

$$\mathbf{H} = \frac{2}{\|\mathbf{c}\|^2}\mathbf{w}\mathbf{w}^H - 4\frac{\mathbf{c}^H\mathbf{w}}{\|\mathbf{c}\|^4}\mathbf{c}\mathbf{w}^H - 4\frac{\mathbf{w}^H\mathbf{c}}{\|\mathbf{c}\|^4}\mathbf{w}\mathbf{c}^H + 8\frac{v_0}{\|\mathbf{c}\|^4}\mathbf{c}\mathbf{c}^H - 2\frac{v_0}{\|\mathbf{c}\|^2}\mathbf{I}. \quad (106)$$

Plugging (104) and (106) into (103) and taking the expectation of (103) with respect to $\mathbf{x}$ using (55) gives

$$v_0(i+1) \cong v_0(i) + \frac{2\sigma_x^2}{a^2\|\mathbf{c}(i)\|^2}\left(\mathbf{1}^T\mathbf{v}(i) - N_T v_0(i)\right). \quad (107)$$

This is used as an approximation to the update in the desired eigenspace. Given this update in v_0, the update in v_1 is known, as the weight norm is not changed by a change of the channel. Applying the second approximation of the analysis, the realization of $\mathbf{c}$ is assumed independent of the prior realizations that gave rise to $\mathbf{v}(i)$ so that the expectation with respect to $\mathbf{c}$ can be used. Using (107) and (56) while conserving the weight vector norm gives the channel update transition.

$$\mathbf{v}(i+1) \cong \mathbf{G}_{\text{chan}}\mathbf{v}(i) = \left(\mathbf{I} + s \cdot \begin{bmatrix} 1-N_T & 1 \\ N_T - 1 & -1 \end{bmatrix}\right)\mathbf{v}(i). \quad (108)$$

Note that the update of (108) satisfies intuitive expectations. For small $(1-a)$, the update represents a transfer of an equal fraction of modal power from each eigenmode to all the other eigenmodes. Each null eigenmode receives one share from v_0, but since v_1 represents these $N_T - 1$ eigenmodes of the nullspace, it receives $N_T - 1$ shares from v_0. At the same time, each of the null eigenmodes contributes one share of its power to every other eigenmode. In the nullspace, this is a zero sum game, but these contributions go from the nullspace energy to v_0. Since v_1 represents the sum of all the powers of the null eigenmodes, this is a transfer of only one share from v_1 to v_0. Clearly, this approximation can only be valid if the resultant values of v_0 remain positive so that the AR1 rate of change must be slow enough to satisfy

$$1 - a < 1 - \frac{1}{\sqrt{2}}. \quad (109)$$

REFERENCES

[1] L. Godara, "Applications of antenna arrays to mobile communications, Part I: performance improvement, feasibility and system considerations," *Proc. IEEE*, vol. 85, pp. 1031–1060, July 1997.

[2] ——, "Applications of antenna arrays to mobile communications, Part II: Beam-forming and direction of arrival considerations," *Proc. IEEE*, vol. 85, no. 85, pp. 1195–1245, Aug. 1997.

[3] A. Naguib, A. Paulraj, and T. Kailath, "Capacity improvement with base-station antenna arrays in cellular CDMA," *IEEE Trans. Veh. Technol.*, vol. 43, pp. 691–698, Aug. 1994.

[4] F. Rashid-Farrokhi, K. J. R. Liu, and L. Tassiulas, "Transmit beamforming and power control for cellular wireless systems," *IEEE J. Select. Areas Commun.*, vol. 16, pp. 1437–1450, Oct. 1998.

[5] Y. C. Liang and F. Chin, "Transmit antenna array techniques for cellular CDMA systems," in *Proc. Ninth Int. Symp. Personal, Indoor, Mobile Radio Commun.*, Sept. 1998.

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

[7] C. B. Papadias and G. J. Foschini, "A space-time coding approach for systems employing four transmit antennas," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, May 2001.

[8] G. J. Foschini, "Layered space time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs Tech. J.*, 1996.

[9] 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.

[10] E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Trans. Telecommun.*, vol. 10, no. 6, pp. 585–595, Nov.-Dec. 1999.

[11] P. Zetterberg and B. Ottersten, "The spectrum efficiency of a base station antenna array system for spatially selective transmission," *IEEE Trans. Veh. Technol.*, vol. 44, pp. 651–660, Aug. 1995.

[12] G. G. Raleigh, S. N. Diggavi, V. K. Jones, and A. Paulraj, "A blind adaptive transmit antenna algorithm for wireless communication," *Proc. IEEE Int. Conf. Commun.*, June 1995.

[13] D. Gerlach and A. Paulraj, "Adaptive transmitting antenna arrays with feedback," *IEEE Signal Processing Lett.*, vol. 1, pp. 150–152, Oct. 1994.

$$\mathbf{v}' \Leftarrow \begin{bmatrix} \left(\left(1 - 2\sqrt{\frac{2}{\pi}}\beta\cdot\gamma\right) + 2\sqrt{\frac{2}{\pi}}\cdot\beta\left(\gamma+\gamma^{-1}\right)\right)v_0 + 2\beta^2\sum_{n=0}^{N_T-1} v_n \\ \sum_{k=1}^{N-1}\left(\left(1 - 2\sqrt{\frac{2}{\pi}}\beta\cdot\gamma\right)v_n + 2\beta^2\sum_{n=0}^{N_T-1} v_n\right) \end{bmatrix} = \mathbf{G}'_{\text{alg}}(\gamma)\,\mathbf{v}' \quad (101)$$

[14] R. W. Heath Jr and A. Paulraj, "A simple scheme for transmit diversity using partial channel feedback," *Conf. Rec. Thirty Second Asilomar Conf. Signals, Syst., Comput.*, Nov. 1998.
[15] J. W. Liang and A. Paulraj, "Forward link antenna diversity using feedback for indoor communication systems," in *Proc. Int. Conf. Acoust., Speech, Signal Processing*, May 1995.
[16] R. Fletcher, *Practical Methods of Optimization*, Second ed. New York: Wiley.
[17] S. Haykin, *Adaptive Filter Theory*, Third ed. Upper Saddle River, NJ: Prentice-Hall, 1996.
[18] G. H. Golub and C. F. van Loan, *Matrix Computations*, Third ed. Baltimore, MD: Johns Hopkins Univ. Press, 1996.
[19] T. Claasen and W. Mecklenbrauker, "Comparison of the convergence of two algorithms for adaptive FIR digital filters," *IEEE Trans. Circuits Syst.*, vol. CAS-28, pp. 510–518, June 1981.
[20] N. Verhoeckx and T. Claasen, "Some considerations on the design of adaptive digital filters equipped with the sign algorithm," *IEEE Trans. Commun.*, vol. COM-32, pp. 258–266, Mar. 1984.
[21] N. Verhoeckx, H. Van Den Elzen, F. Snijders, and P. Van Gerwen, "Digital echo cancellation for baseband data transmission," *IEEE Trans. Acoust., Speech Signal Processing*, vol. ASSP-27, pp. 768–781, Dec. 1979.
[22] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-Time block codes from orthogonal designs," *IEEE Trans. Inform. Theory*, vol. 45, pp. 1456–1467, July 1999.

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

His research has focused on the application of multiple antenna arrays to communications and signal processing problems. 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, San Diego, since May 2002, where he has focused on algorithm implementation and verification for cellular products.

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 with the Electrical and Computer Engineering Department, the University of California at San Diego, La Jolla, 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 co-recipient 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.