Fast Stochastic Power Control Algorithms for Nonlinear Multiuser Receivers

Mahesh K. Varanasi, *Senior Member, IEEE,* and Deepak Das, *Member, IEEE*

***Abstract*—Uplink communication in a cellular radio network is considered where the base station in each cell employs linear or nonlinear (decision feedback) multiuser receivers. For any such receiver, the problem of interest is that of minimizing the total transmit power under the constraint that all the users of the network achieve their quality-of-service objective in terms of signal-to-interference ratio (SIR). When the solution is feasible for the desired SIR requirements, the optimum powers are computed with a distributed iterative power control strategy suitable for implementation at each base station. While the deterministic algorithm requires both in-cell and out-of-cell user information, the stochastic algorithm proposed in this paper can be implemented at the base stations in a truly distributed manner requiring knowledge of only in-cell parameters. Such an algorithm was proposed recently for the case where base stations use linear (single user) matched filter (MF) receivers. However, the feasibility region in terms of attainable SIRs for a well-designed multiuser receiver, particularly for a nonlinear receiver that employs decision feedback, is generally much larger than it is for the linear MF receiver. The stochastic power control algorithm in this paper, for linear or nonlinear multiuser receivers, converges in the mean-square sense to the minimal powers when the target SIRs are feasible. The second major focus of this paper is to improve the convergence properties of the conventional stochastic approximation based power control strategy by using the more recent results on *averaging*. Convergence issues of both the "nonaveraged" and "averaged" algorithms are investigated, and numerical examples are presented to demonstrate the performance improvement due to averaging.**

***Index Terms*—Code-division multiple access, decision-feedback receivers, multiuser detection, power control, stochastic approximation.**

I. Introduction

RECENT WORK in cellular wireless systems has recognized power control as a flexible means of meeting different quality-of-service (QoS) constraints in an efficient way (cf. [1]). A combined approach to multiuser receivers and power control for wireless networks was introduced earlier in [2]–[4]. In [3], it was shown that nonlinear multiuser receivers have the potential for significantly higher performance than do their linear counterparts. Those results were obtained for a single-cell channel and deterministic power control. This paper investigates multicell, distributed stochastic inner-loop power control on the uplink for a cellular radio network where the base stations employ multiuser receivers, with or without decision feedback (see conference version of this paper [5]). For the case where base stations use linear single-user receivers, a stochastic power control algorithm was proposed in [6]. In this paper, we show that: 1) the use of multiuser receivers, particularly those with decision feedback, results in significantly more power and/or bandwidth efficient systems; and 2) substantial improvements in rates of convergence can be achieved over the Robbins–Monro stochastic approximation-based power control algorithm of [6] by using some of the newer developments in the general stochastic approximation literature.

A power control algorithm that converges quickly to the target signal-to-interference ratios (SIRs) is of singular importance in a cellular environment which is typically nonstationary, since power updates are conveyed to the mobiles at a low feedback rate, and a quick recalculation of the optimum user power needs to be done whenever the channel changes significantly. To improve the convergence properties of stochastic power control algorithms with very little additional complexity, the concept of "averaging" is introduced in this paper.

The issue of convergence rate of general stochastic approximation methods has, for a long time, been of great interest both from theoretical and practical points of view, and has been largely addressed in terms of adaptively optimizing the step-size sequence used. In a couple of fundamental papers on stochastic approximation, Ruppert [7] and Polyak [8] showed under certain assumptions that with a decreasing step-size sequence that decays slower than the usual Robbins–Monro *constant*/n step-size sequence (where n is the iteration index), significant improvement in convergence can be achieved if the output of the conventional stochastic approximation is averaged. In fact, not only does this achieve the optimal rate of convergence, but also the optimal asymptotic error covariance with respect to (w.r.t.) the trace. In this paper, we show the convergence of an "averaged" stochastic *power control* algorithm and demonstrate the resulting improvement in performance.

II. System Model

We consider the uplink of a cellular network [15] in which there are B base stations and K active users with K_j users assigned to base j. For simplicity of presentation, we assume

Paper approved by H. Leib, the Editor for Communication and Information Theory of the IEEE Communications Society. Manuscript received December 14, 2000; revised November 16, 2001. This work was supported in part by the National Science Foundation under Grant CCR-9814996 and Grant ANIR-9725778. This paper was presented in part at the Allerton 1999 Conference, Allerton, IL, and at the International Conference on Personal Wireless Communications (ICPWC) 2000, Hyderabad, India.

M. K. Varanasi is with the Department of Electrical and Computer Engineering, University of Colorado, Boulder, CO 80309-0425 USA (e-mail: varanasi@schof.colorado.edu).

D. Das was with the Department of Electrical and Computer Engineering, University of Colorado, Boulder, CO 80309-0425 USA. He is now with the Wireless Technology Laboratory, Lucent Technologies, Whippany, NJ 07981 USA (e-mail: ddas2@lucent.com).

Digital Object Identifier 10.1109/TCOMM.2002.805274

that each base employs a common set of N matched filters (MFs) matched to orthonormal basis functions that span the entire signal space. The discussion in this paper remains valid for the case where any base station, say j, uses a different set of $N_j \leq N$ MFs matched to the orthonormal basis of some possibly "reduced" signal space that includes the signals received from the users within its cell. While the transmissions of the mobile users of a particular cell arrive at the bases of other cells symbol-asynchronously, it is assumed, for the sake of simplicity, that they arrive at their own base symbol-synchronously. It is also assumed that while the receiver at base j has knowledge of the common timing of the received signals of its own cell, it does not have the timing information of signals of other cells.

The discrete-time model for the N MF outputs at base j can be expressed as

$$\mathbf{y}_j = \sum_{i=1}^{K_j} \sqrt{w_{ij} g_{ijj}} \mathbf{s}_{ij} x_{ij} + \sum_{l \neq j}^{B} \sum_{i=1}^{K_l} \sqrt{w_{il} g_{ilj}} \left(\mathbf{s}_{ilj}^{-} x_{il}^{-} + \mathbf{s}_{ilj}^{+} x_{il}^{+} \right) + \mathbf{n}_j \quad (1)$$

where w_{il} and x_{il} denote the transmit power and the transmitted symbol, respectively, of user i of base l. The channel gain of the ith user of base l to base j is denoted by g_{ilj}. The vectors $\mathbf{s}_{ilj}$ denote the vector representations (the "signature sequence") of the signal of user i of base j w.r.t. the orthonormal basis functions employed at base j. The vectors $\mathbf{s}_{ilj}^{-}$ and $\mathbf{s}_{ilj}^{+}$ denote the vector representations of the segments of the signals associated with the two symbols of user i of base l that overlap with the signals of base j. $\mathbf{n}_j$ is an N-dimensional zero-mean Gaussian random vector with a covariance matrix equal to $\sigma_j^2 \mathbf{I}$. Define the in-cell signal matrix $\mathbf{S}_j = \left[\mathbf{s}_{1j} \mathbf{s}_{2j} \cdots \mathbf{s}_{K_j j} \right]$ whose columns are the signature sequences of the users of cell j. Also define the signal matrices of out-of-cell users $\mathbf{S}_{lj}^{+} = \left[\mathbf{s}_{1lj}^{+} \mathbf{s}_{2lj}^{+} \cdots \mathbf{s}_{K_l lj}^{+} \right]$ and $\mathbf{S}_{lj}^{-} = \left[\mathbf{s}_{1lj}^{-} \mathbf{s}_{2lj}^{-} \cdots \mathbf{s}_{K_l lj}^{-} \right]$. The received signal can, therefore, be written as

$$\mathbf{y}_j = \mathbf{S}_j \mathbf{G}_{jj}^{1/2} \mathbf{W}_j^{1/2} \mathbf{x}_j + \sum_{l \neq j}^{B} \mathbf{S}_{lj}^{-} \mathbf{G}_{lj}^{1/2} \mathbf{W}_l^{1/2} \mathbf{x}_l^{-} + \sum_{l \neq j}^{B} \mathbf{S}_{lj}^{+} \mathbf{G}_{lj}^{1/2} \mathbf{W}_l^{1/2} \mathbf{x}_l^{+} + \mathbf{n}_j \quad (2)$$

where $\mathbf{G}_{lj} = \operatorname{diag}\{g_{1lj}, g_{2lj}, \ldots, g_{K_l lj}\}$, and $\mathbf{W}_l = \operatorname{diag}\{w_{1l}, w_{2l}, \ldots, w_{K_l l}\}$. Moreover, $\mathbf{x}_j$, $\mathbf{x}_l^{+}$, and $\mathbf{x}_l^{-}$ are vectors whose ith elements are x_{ij}, x_{il}^{+}, and x_{il}^{-}, respectively.

III. Multiuser Receivers

Each base station uses either a linear or a decision feedback receiver [16] to decode the information transmitted by users in its own cell. In the decision feedback case, assume for notational simplicity that the users in each cell are numbered according to the order in which they are decoded. The soft output of the linear or decision feedback receiver for user i of base j can be written in the common form $v_{ij} = \mathbf{f}_{ij}^T \left(\mathbf{y}_j - \sum_{k=1}^{i-1} \mathbf{b}_{kij} \sqrt{w_{kj} g_{kjj}} \hat{x}_{kj} \right)$. $\mathbf{f}_{ij}$ is the feedforward filter. The feedback filters $\mathbf{b}_{kij} = \mathbf{0}$ for linear receivers (no decision feedback), and for decision feedback receivers, we let $\mathbf{b}_{kij} = \mathbf{s}_{kj}$ for $k = 1, \ldots, i-1$, without loss of generality [16]. The symbol T denotes transposition. $\hat{x}_{kj}$ denotes the detected (in uncoded transmission) or decoded (in the coded case) symbols of "past" users $1, \ldots, i-1$ relative to user i of base j. There is an issue of decoding delay in coded systems that we will discuss later in this paper. It is clear that base j has knowledge of the codes, signals, transmit powers, and channel gains of the users within its own cell. Hence, we consider some single-cell receivers that require only in-cell information. The feed-forward filter for both the single-user MF receiver and the matched-filter decision feedback (MF-DF) receiver is given by $\mathbf{f}_{ij} = \mathbf{s}_{ij}$. The latter has nonzero feedback filters $\mathbf{b}_{kij} = \mathbf{s}_{kj}$ for $k = 1, \ldots, i-1$. A "partial" decorrelator for base j is given as

$$\mathbf{f}_{ij}^T = \operatorname{row}_i \left(\mathbf{S}_j^T \mathbf{S}_j \right)^{-1} \mathbf{S}_j^T. \quad (3)$$

This receiver tunes out the interference for each user that arises due to the other users in the same cell, but ignores interference from out-of-cell users. The decision feedback version of this decorrelator (denoted as D-DF) has feedback filter coefficients $\mathbf{b}_{kij} = \mathbf{s}_{kj}$ for $k = 1, \ldots, i-1$, and the feedforward filters are defined as follows: consider a partition of the columns of $\mathbf{S}_j$ into two matrices, one that contains the first $i-1$ columns, and the second that contains the other $K_j - i + 1$ columns. Accordingly, let $\mathbf{S}_j = \left[\bar{\mathbf{S}}_j(i) \mathbf{S}_j(i) \right]$. The feedforward filter of user i is then given as

$$\mathbf{f}_{ij}^T = \operatorname{row}_1 \left(\mathbf{S}_j^T(i) \mathbf{S}_j(i) \right)^{-1} \mathbf{S}_j^T(i). \quad (4)$$

This receiver subtracts interference contributed by the already decoded users in cell j and decorrelates interference from the as yet undecoded users in cell j [17] without consideration for out-of-cell users.

The following discussion on power control is also valid for other linear receivers that do not depend on user powers but might require out-of-cell user information, like the decorrelator that removes inteference from both in-cell and out-of-cell users.

IV. Optimum Power Control

Given that each base uses an arbitrary but linear or decision feedback receiver (that does not depend on the user powers), consider the problem of minimizing total transmitted power subject to the constraints that the SIR of user i in cell j, γ_{ij}, is not less than some target value γ_{ij}^* for all $j \in \{1, \ldots, B\}$ and $i \in \{1, \ldots, K_j\}$

$$\begin{aligned} \min \quad & \sum_{j=1}^{B} \sum_{i=1}^{K_j} w_{ij} \\ \text{s.t.} \quad & \gamma_{ij} \geq \gamma_{ij}^*. \end{aligned} \quad (5)$$

For $j \in \{1, 2, \ldots, B\}$, define the diagonal matrices $\mathbf{D}_j$ according to

$$\mathbf{D}_j = \operatorname{diag}\left\{ \gamma_{1j}^* / g_{1jj} |\mathbf{f}_{1j}^T \mathbf{s}_{1j}|^2, \ldots, \gamma_{K_j j}^* / g_{K_j jj} |\mathbf{f}_{K_j j}^T \mathbf{s}_{K_j j}|^2 \right\}.$$

Let $\mathbf{t}_j = \sigma_j^2 \left[\mathbf{f}_{1j}^T \mathbf{f}_{1j}, \ldots, \mathbf{f}_{K_j j}^T \mathbf{f}_{K_j j}\right]^T$. Define the vector of transmitted powers of users in cell j as $\mathbf{w}_j = [w_{1j}, \ldots, w_{K_j j}]^T$. Define matrices $\{\mathbf{Q}_{jl}\}$ for $j, l \in \{1, 2, \ldots, B\}$, with $Q_{jl}(i,k)$ denoting the (i,k)th element of $\mathbf{Q}_{jl}$, as follows: the matrices $\mathbf{Q}_{jj}$ are defined for $j = 1, \ldots, B$ as

$$Q_{jj}(i,k) = \begin{cases} \left|\mathbf{f}_{ij}^T(\mathbf{s}_{kj} - \mathbf{b}_{kij})\right|^2, & \text{if k < i} \\ 0, & \text{if k=i} \\ \left|\mathbf{f}_{ij}^T \mathbf{s}_{kj}\right|^2, & \text{if k>i} \end{cases} \tag{6}$$

for $i, k \in \{1, \ldots, K_j\}$, and the matrices $\mathbf{Q}_{jl}$ for $j, l \in \{1, 2, \ldots, B\}$ with $j \neq l$ are defined as

$$Q_{jl}(i,k) = \left|\mathbf{f}_{ij}^T \mathbf{s}_{klj}^-\right|^2 + \left|\mathbf{f}_{ij}^T \mathbf{s}_{klj}^+\right|^2 \tag{7}$$

for $i \in \{1, \ldots, K_j\}$ and $k \in \{1, \ldots, K_l\}$. In order to describe the multicellular power control problem succinctly, let us define the square, nonnegative, block matrix $\mathcal{X}$ with B^2 matrix elements, such that the (j,l)th matrix element is given by the product

$$\mathbf{X}_{jl} \triangleq \mathbf{D}_j \mathbf{Q}_{jl} \mathbf{G}_{lj} \tag{8}$$

for each $j, l \in \{1, 2, \ldots, B\}$. Further, let $\mathbf{D} \triangleq \mathrm{diag}\{\mathbf{D}_1, \ldots, \mathbf{D}_B\}$, $\mathbf{w} \triangleq \left[\mathbf{w}_1^T, \ldots, \mathbf{w}_B^T\right]^T$, and $\mathbf{t} \triangleq \left[\mathbf{t}_1^T, \ldots, \mathbf{t}_B^T\right]^T$. With this notation, the SIR constraints in (5) can be written as

$$\mathbf{w} \geq \mathcal{X}\mathbf{w} + \mathbf{D}\mathbf{t}. \tag{9}$$

Assuming $\gamma_{ij}^* > 0$, $\sigma_j^2 > 0$, and $|\mathbf{f}_{ij}^T \mathbf{s}_{ij}|^2 > 0$ for all $i \in \{1, \ldots, K_j\}$, $j \in \{1, \ldots, B\}$, the target SIRs are said to be *feasible* if there exist positive and finite transmit powers that satisfy the above inequalities. In fact, we can show that the component-wise optimum (minimum) powers $\mathbf{w}^* > 0$ which solve the power control problem in (5) are given by the solution (*if it exists*) of the case when (9) holds with equality

$$\mathbf{w} = \mathcal{X}\mathbf{w} + \mathbf{D}\mathbf{t} \tag{10}$$

which implies that

$$\mathbf{w}^* = (\mathbf{I} - \mathcal{X})^{-1}\mathbf{D}\mathbf{t}. \tag{11}$$

The key issue of the existence of a solution is linked to the matrix $\mathcal{X}$. Note that this matrix is not necessarily irreducible (cf. [19]). It is reducible, for instance, in the particular case of a single-cell channel when the base employs any decision feedback receiver. In either case, since $\mathcal{X}$ is a square nonnegative matrix, there exists a real nonnegative eigenvalue $\lambda_0(\mathcal{X})$ of $\mathcal{X}$ with nonnegative left and right eigenvectors, such that the magnitude of any other eigenvalue of $\mathcal{X}$ is no greater than $\lambda_0(\mathcal{X})$. This eigenvalue is called the maximal eigenvalue of $\mathcal{X}$ [19]. Our next result establishes conditions under which (10) has a unique positive solution ensuring feasibility of the target SIRs. In the rest of this discussion, unless otherwise stated, we shall denote by $\mathbf{X}_1 \geq \mathbf{X}_2$ that every element of the matrix or vector $\mathbf{X}_1 - \mathbf{X}_2$ is nonnegative.

Theorem 1: A necessary and sufficient condition for (10) to have a unique positive solution, ensuring the feasibility of the target SIRs with minimum powers, is that $\lambda_0(\mathcal{X}) < 1$.

Proof: We consider the case when $\mathcal{X}$ is **reducible** (the case of **irreducible** $\boldsymbol{\mathcal{X}}$ can be seen as a special case). Clearly, $\mathcal{X} \geq \mathbf{0}$. It can be shown that we can always find a permutation matrix $\mathbf{P}$ such that [25, p. 506]

$$\mathbf{P}^T \mathcal{X} \mathbf{P} = \tilde{\mathcal{X}} = \begin{bmatrix} \tilde{\mathcal{X}}_{11} & \tilde{\mathcal{X}}_{12} & \cdots & \\ \mathbf{0} & \tilde{\mathcal{X}}_{22} & \tilde{\mathcal{X}}_{23} & \cdots \\ \vdots & \ddots & \ddots & \\ \mathbf{0} & \cdots & \mathbf{0} & \tilde{\mathcal{X}}_{LL} \end{bmatrix} \tag{12}$$

in which each $\tilde{\mathcal{X}}_{ii} \in \mathbb{R}^{l_i \times l_i}$ is either irreducible, or is the 1-by-1 zero matrix. Denoting $\tilde{\mathbf{r}} = \mathbf{P}^T\mathbf{r} = \left[\tilde{\mathbf{r}}_1^T \tilde{\mathbf{r}}_2^T \cdots \tilde{\mathbf{r}}_L^T\right]^T$, $\tilde{\mathbf{r}}_i \in \mathbb{R}^{l_i \times 1}$, and $\tilde{\mathbf{c}} = \mathbf{P}^T\mathbf{c} == \left[\tilde{\mathbf{c}}_1^T \tilde{\mathbf{c}}_2^T \cdots \tilde{\mathbf{c}}_L^T\right]^T$, $\tilde{\mathbf{c}}_i \in \mathbb{R}^{l_i \times 1}$, we note that the solution to the equation $(s\mathbf{I} - \tilde{\mathcal{X}})\tilde{\mathbf{r}} = \tilde{\mathbf{c}}$ is a permuted version of the solution to the general equation

$$(s\mathbf{I} - \mathcal{X})\mathbf{r} = \mathbf{c} \tag{13}$$

with $\mathbf{c} > \mathbf{0}$. Two different permutations, $\mathbf{P_1}$ and $\mathbf{P_2}$ [satisfying the condition in (12)], each with unique positive solution to the permuted set of equations, will yield the same solution for $\mathbf{r}$. Now, the solution to $(s\mathbf{I} - \tilde{\mathcal{X}}_{LL})\tilde{\mathbf{x}}_L = \tilde{\mathbf{c}}_L$ is guaranteed to be uniquely positive if $\lambda_{PF}(\tilde{\mathcal{X}}_{LL}) < s$ (cf. [19]). Substituting this value of $\tilde{\mathbf{r}}_L$ in the next set of equations $(s\mathbf{I} - \tilde{\mathcal{X}}_{L-1,L-1})\tilde{\mathbf{r}}_{L-1} = \tilde{\mathcal{X}}_{L-1,L}\tilde{\mathbf{r}}_L + \tilde{\mathbf{c}}$, we are again guaranteed a unique positive solution to $\tilde{\mathbf{r}}_{L-1}$ if $\lambda_{PF}(\tilde{\mathcal{X}}_{L-1,L-1}) < s$. Proceeding in this way, and noting that $\mathrm{eigenvalues}(\tilde{\mathcal{X}}) = \bigcup_i \mathrm{eigenvalues}(\tilde{\mathcal{X}}_{ii})$, we obtain

$$\lambda_0(\tilde{\mathcal{X}}) = \max\left\{\mathrm{eigenvalues}(\tilde{\mathcal{X}}_{ii})\right\} < s. \tag{14}$$

Clearly, s, $\mathbf{r}$, and $\mathbf{c}$ in our problem are 1, $\mathbf{w}$, and $\mathbf{D}\mathbf{t}$, respectively, and therefore, we have proved our claim. ■

Using the fact that $\lambda_0(c\psi) = c\lambda_0(\psi)$, scalar $c > 0$, square matrix $\psi \geq \mathbf{0}$, we can show the following consequence of *Theorem 1* in the special case of a common target SIR.

Corollary 1: In the case of a common SIR requirement for all users, $\gamma_{ij}^* = \gamma^*$, the following upper bound holds on the achievable SIR:

$$\gamma^* < \frac{1}{\lambda_0(\tilde{\mathcal{X}})} \tag{15}$$

where $\tilde{\mathcal{X}}$ is equal to $(1/\gamma^*)\mathcal{X}$ (and independent of γ^*).

Note that in the nonlinear multiuser receiver case, there is an implicit assumption that feedback is perfect. While not valid in general, it is possible, by decoding the users in the decreasing order of their target SIRs, to mitigate the error propagation effects to a large extent [17]. In this regard, equal target SIRs may represent something of a worst-case scenario. In such cases, it may be possible to change the problem slightly and set the target SIRs to be slightly different from each other, while ensuring that the worst-case target SIR is equal to the required target SIR. The slight increase in the powers required to achieve these higher sets of target SIRs may be a small price to offset the effects of error propagation and achieve better overall performance. For instance, the feasibility region in terms of achievable SIRs for the MF-DF receiver (assuming perfect feedback) is greater than that for the MF receiver. More precisely, since $\mathbf{0} \leq \mathcal{X}_{\mathrm{MF-DF}} \leq \mathcal{X}_{\mathrm{MF}}$, the fact that $\lambda_0(\mathcal{X}_1) \leq \lambda_0(\mathcal{X}_2)$ when $\mathbf{0} \leq \mathcal{X}_1 \leq \mathcal{X}_2$ [19] implies:

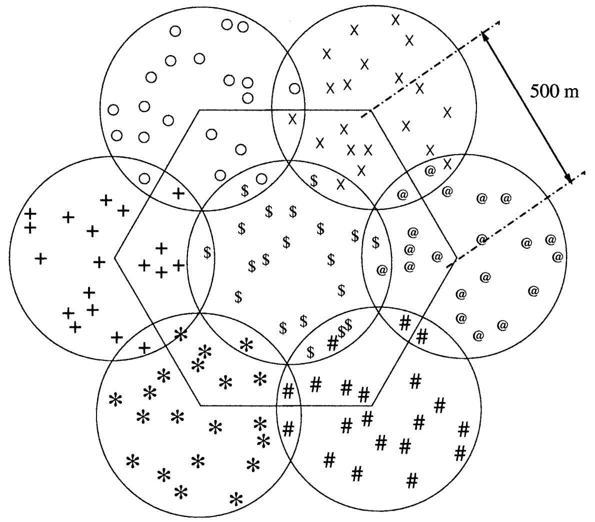


Fig. 1. Seven-cell cellular system with 20 users per cell.

Theorem 2: $\lambda_0(\mathcal{X})$ corresponding to the MF-DF receiver, with decision feedback assumed perfect, is no larger than $\lambda_0(\mathcal{X})$ corresponding to the MF receiver.

A. Distributed Power Control

There are several iterative methods for solving the system of linear equations in (10). Of these, the point Jacobi method [21] is particularly simple and relevant to the problem of distributed power control at the mobiles (see also [22]). In particular, consider the distributed algorithm

$$\mathbf{w}(n) = \mathcal{X}\mathbf{w}(n-1) + \mathbf{D}\mathbf{t} \tag{16}$$

which is easily shown to converge to the optimum powers when the target SIRs are feasible. Defining $\epsilon(n) = \mathbf{w}(n) - \mathbf{w}^*$, we see that by subtracting $\mathbf{w}^*$ from both sides of (16), we have $\epsilon(n) = \mathcal{X}^n\epsilon(0)$ so that as $n \to \infty$, $\epsilon(n) \to \mathbf{0}$, since $\mathcal{X}^n \to \mathbf{0}$ as a result of $\lambda_0(\mathcal{X}) < 1$ [19]. Moreover, the asymptotic rate of convergence is proportional to $-\log\lambda_0(\mathcal{X})$ and, therefore, it follows from *Theorem 2* that the power control algorithm for the MF-DF receiver (assumed perfect) converges no slower than that for the linear MF receiver.

Writing (16) element-wise, we can arrive at the following distributed algorithm for the optimum user powers:

$$w_{ij}(n) = \gamma_{ij}^* \left(\frac{E[|v_{ij}(n)|^2]}{g_{ijj}|\mathbf{f}_{ij}^T\mathbf{s}_{ij}|^2} - w_{ij}(n-1) \right). \tag{17}$$

It is now clear that (17) can be implemented at the (i,j)th mobile or in the jth base station, provided it has perfect knowledge of the mean-squared value or power of the "normalized" decision statistic $v_{ij}(n)/\mathbf{f}_{ij}^T\mathbf{s}_{ij}$, in addition to its target SIR value, the path gain to its own base station, and the transmit power in the previous interval.

Example 1: We consider a simple seven-cell cellular system shown in Fig. 1 with 20 users in each cell. The users are placed randomly (in a uniform manner) across each cell. All the simulations are run with fixed base station assignment (non-macro-diversity handover scenario). The gains to every base station for each user are calculated using a $1/d^\alpha$ path-loss law, where d is the distance to the base station of interest. The path loss exponent is $\alpha = 4$. The additive white noise power in the received signal is set to a fixed value for the entire simulation, ensuring on an average (over all users) a signal-to-noise ratio (SNR) of about 20 dB. For the case where all the users employed decorrelators, the maximum received SNR turned out to be 27 dB (for the user who was assigned the maximum power) and the minimum turned out to be 18 dB. The processing gain is chosen to be 35 (high bandwidth efficiency), and the signature sequences are chosen to be unit norm random vectors. Once chosen, the signature sequences remain fixed throughout the adaptation. In this example, the signature sequences of in-cell users are linearly independent.

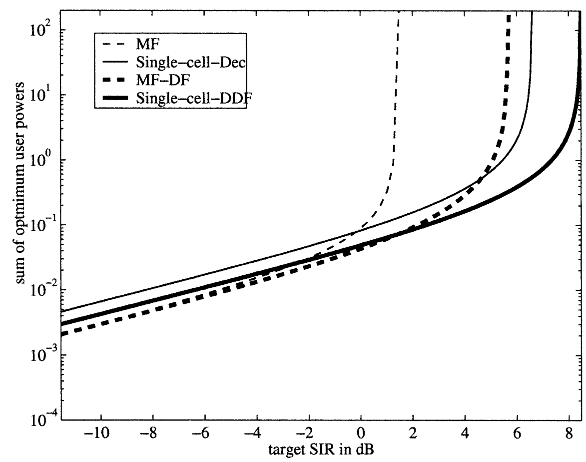


Fig. 2. Minimum total power versus common target SIR for the MF and single-cell decorrelating receivers, linear and decision feedback versions.

Fig. 2 shows the minimum total power required as a function of the common target SIR for the single-cell receivers discussed in Section III, namely, MF, decorrelator, MF-DF, and D-DF. Note that for each of the receivers, there is a feasibility interval of common achievable target SIRs. Beyond the upper limit of this interval, the target SIRs are not feasible independently of how much power one is willing to expend. The feasibility interval for the MF receiver is significantly smaller than that for the multiuser receivers. Further, the decision feedback versions have larger SIR feasibility intervals than the linear versions. Note that at a common target SIR of 6 dB, the total power required by the single-cell decorrelator is nearly ten times that required by the single-cell D-DF receiver with the power savings being even higher at larger target SIRs. Although based on one realization of signature sequences and user locations, the results for this example are typical in that they are very similar to those found for several other choices of randomly chosen signature sequences (of the same length) and user locations as well.

Let us now consider the power control algorithm in (17) in the stochastic setting. If the transmit powers are updated once every several symbols rather than once every symbol, it is possible for the base station to estimate the "normalized" mean-squared values of the decision statistics by computing the corresponding sample mean-squared values. It can then transmit these estimates via feedback channels to the respective mobiles, which can then implement (17) to update their transmit powers for the next block of symbols. However, due to the availability of better computational resources, it may be more sensible to compute the new powers via (17) at the base station and send the updated powers (or some quantized version) to the mobiles through low-rate feedback channels.

In either of these implementations, however, the trajectory of the distributed algorithm in (17), with the mean-squared values of the decision variables replaced by their sample estimates, evolves stochastically, and the convergence of the corresponding deterministic algorithm is no longer sufficient. In what follows, we obtain a stochastic distributed power control algorithm which will be shown to converge to the optimal solution in (11) in the mean-square sense.

V. Distributed Stochastic Power Control

Let the mobile transmit powers be updated once every several symbols. It is then possible for the base station to estimate the mean-squared values of the decision statistics by computing the corresponding sample mean-squared values. We will specify a synchronous power control algorithm where the transmit powers are recursively updated (either at the base stations or at the mobiles) once every M symbols. When the recursion is implemented at the base stations (mobiles), the bases communicate the newly updated powers (sample mean-squared values of the decision statistics) to their respective mobiles via low-rate feedback channels. The mobiles then transmit the next M symbols at the newly updated powers.

Form M-length blocks of the vectors of matched filter outputs at each base. Let $\mathbf{y}_j(n,m)$ denote the mth vector in the nth block. The model in (1) can be written, including the time indexes, as

$$\begin{aligned}\mathbf{y}_j(n,m) = &\sum_{i=1}^{K_j} \sqrt{w_{ij}(n-1)g_{ijj}}\mathbf{s}_{ij}x_{ij}(n,m)\\ &+ \mathbf{n}_j(n,m) + \sum_{l\neq j}^{B}\sum_{i=1}^{K_l}\sqrt{w_{il}(n-1)g_{ilj}}\\ &\times\left(\mathbf{s}_{ilj}^{-}x_{il}^{-}(n,m) + \mathbf{s}_{ilj}^{+}x_{il}^{+}(n,m)\right)\end{aligned} \tag{18}$$

where $w_{il}(n-1)$ denotes the transmit power of user i of base l during the nth block, and $x(n,m)$ with the appropriate sub- and superscripts denote the corresponding transmitted symbols in the mth duration in the nth block. The decision statistic for user i of cell j is

$$v_{ij}(n,m) = \mathbf{f}_{ij}^T\left(\mathbf{y}_j - \sum_{k=1}^{i-1}\mathbf{b}_{kij}\sqrt{w_{kj}(n-1)g_{kjj}}\hat{x}_{kj}(n,m)\right) \tag{19}$$

where $\hat{x}_{kj}(n,m)$ is the detected symbol in the uncoded case, and it is the decoded and re-encoded symbol in a coded system with decoders that are based on the soft outputs $v_{ij}(n,m)$ obtained during the nth block. The time-averaged power in the decision statistic is given by

$$z_{ij}(n) = \frac{1}{M-1}\sum_{m=1}^{M-1}|v_{ij}(n,m)|^2. \tag{20}$$

The averaging in (20) is done over $M-1$ symbol intervals (ignoring the Mth interval), because the received data in the Mth interval reflects two different powers for each out-of-cell user. If M is relatively large, this does not lead to any significant loss of efficiency. Moreover, it allows the convergence analysis to remain fairly simple. Note that in a coded system with decision feedback, after decoding and re-encoding (cf. [16]), it is necessary to choose M to be an integer multiple of the block length of the code and to synchronize the times of power updates with the beginning of a new codeword. Consider the following stochastic algorithm, based on the Robbins–Monro stochastic approximation method [23], that iteratively updates the transmit power of mobile i in cell j:

$$\begin{aligned}w_{ij}(n) = &\, w_{ij}(n-1)\\ &+a_n\left(\frac{z_{ij}(n)\gamma_{ij}^*}{g_{ijj}|\mathbf{f}_{ij}^T\mathbf{s}_{ij}|^2} - (1+\gamma_{ij}^*)w_{ij}(n-1)\right)\end{aligned} \tag{21}$$

where $\{a_n\}$ denotes the sequence of step sizes. The algorithm in (21) is applicable to arbitrary fixed linear and decision-feedback multiuser receivers. The special case of the MF receiver reduces to the one proposed in [6]. The mean-square convergence of (21) is stated next. Let $\mathbf{w}_j(n)$ be the vector of power updates of users in cell j in the nth iteration. Let $\mathbf{w}(n)$ be the power vector obtained by stacking $\{\mathbf{w}_j(n)\}_{j=1}^B$.

Theorem 3: If the target SIRs are feasible, and assuming perfect feedback (in the case of decision-feedback receivers), the stochastic algorithm in (21) converges to the optimal powers, $\mathbf{w}^*$, in the mean-square sense, i.e.,

$$\lim_{n\to\infty} E\left[||\mathbf{w}(n) - \mathbf{w}^*||^2\right] = 0 \tag{22}$$

for a positive-valued step-size sequence a_n that satisfies the Robbins–Monro conditions [23] $\sum_{n=1}^{\infty} a_n = \infty$ and $\sum_{n=1}^{\infty} a_n^2 < \infty$.

In the Appendix, where we discuss the proof of the above theorem, we also very briefly discuss the fixed step-size case. We note [see (31)], that for the fixed step-size case, there is the usual tradeoff between the asymptotic mean-squared error (MSE) and the rate of convergence. More, precisely, the larger the fixed step size chosen, the faster the rate of convergence, but more the asymptotic MSE. We also note that taking expectations of both sides of (29), we get the following equation to study the rate of convergence, in the mean, of the power control algorithm in (21):

$$E[\mathbf{w}(n)] = [\mathbf{I} - a_n\mathbf{A}]\, E[\mathbf{w}(n-1)] - a_n\Gamma\mathbf{D}\mathbf{t} \tag{23}$$

where $\mathbf{A} \triangleq (\mathbf{I} - \mathcal{X})$. For $0 < a_n = a < 1$, we note that $[\mathbf{I} - a\mathbf{A}]$ is a nonnegative matrix for both linear and (perfect) decision-feedback receivers. The recursion in (23), in terms of the expected values, has an asymptotic rate of convergence proportional to $-\log\left(\lambda_0([\mathbf{I} - a\mathbf{A}])\right)$ [21]. As in *Theorem 2*, it is easy to show, using properties of nonnegative matrices and the structure of $[\mathbf{I} - a\mathbf{A}]$ for the MF and MF-DF receivers, that for the decision-feedback case, $\lambda_0([\mathbf{I} - a\mathbf{A}])$ is no bigger than that for the linear case. For the MF-based receivers, therefore, not only does decision feedback help meet the target SIRs with smaller powers (*Theorem 2*), it also ensures faster convergence in the mean for the stochastic power control algorithm.

VI. Stochastic Power Control With Averaging

A recent fundamental development in stochastic approximation is the idea of averaging, as introduced in [7] and [8]. In the former work, a linear algorithm for the one-dimensional case was considered, and asymptotic normality of the procedure was proved. Multidimensional problems were considered in [8], and under certain assumptions, mean-square convergence for decreasing step-size adaptive algorithms followed by averaging was demonstrated. The improvement in convergence is essentially a result of averaging a stochastic approximation that uses a step-size sequence that decays more slowly (or is relatively "larger") than the a/n step size that was used in the original Robbins–Monro formulation. It was further shown in [9] that almost surely (a.s.) convergence is achieved, even for a suitable fixed step-size stochastic approximation strategy with averaging, and this was "optimal" in terms of the convergence rate and the asymptotic error covariance matrix. A lot of work has since followed in stochastic approximation using the new ideas on averaging, see, for example, [10] and [11]. Using the averaging techniques discussed in [9] to modify our recursion in (21) to include an averaging step after the "basic" recursion, we obtain the following two-step recursion:

$$w_{ij}(n) = w_{ij}(n-1) + a_n\left(\frac{z_{ij}(n)\gamma_{ij}^*}{g_{ijj}|\mathbf{f}_{ij}^T\mathbf{s}_{ijj}|^2} - (1+\gamma_{ij}^*)w_{ij}(n-1)\right) \tag{24}$$

$$\tilde{w}_{ij}(n) = \frac{1}{n}\left((n-1)\tilde{w}_{ij}(n-1) + w_{ij}(n)\right) \tag{25}$$

where a_n is a suitable decreasing or fixed step-size sequence. Note that the system requirements and the order of computational complexity for the recursion with averaging remain the same as for the nonaveraged algorithm in (21). The intuition behind averaging is that the step size in the basic recursion can be chosen to be quite large, such that (by itself) it shows rapid convergence but large fluctuations in the convergence trajectory and large asymptotic error. The averaging step essentially "smooths" the fluctutions and removes the asymptotic error while retaining the rapid convergence. Our analysis and numerical examples later in the paper demonstrate this improvement in convergence.

A key point to note, however, is that the averaging step is essentially "offline," in that the basic recursion is independent of the averaging step. This implies that until the algorithm has sufficiently converged, the mobile transmit powers have to come from the basic recursion step. This, of course, means that during this period, the mobiles cannot really benefit from the improved powers computed by the averaging step. In fact, if the mobiles were to transmit with the powers obtained from the averaging step (resulting in "coupling" between the basic recursion and the averaging step), the performance would be worse than that of just the basic recursion by itself. So the averaging step has to be kept independent of the basic recursion. This does not mean, however, that the applicability of the averaged algorithm specified in (24) is limited. In a typical modern wireless network [12], [13], once the target SIRs have been fixed by outer-loop power control for certain channel conditions, the goal is to converge, via inner-loop power control, to the optimum powers and target SIRs within a short initial convergence interval that may be fixed by the standards. So, if we care not so much about the transmit powers during this short interval, but more about the powers computed at the end of it, the averaged algorithm will definitely give us better estimates of the optimum powers than the nonaveraged version, and these can be conveyed to the mobiles at the end of the initial convergence interval. One can continue using the simpler nonaveraged recursion in (21) (or the well-known "bang-bang" or up–down power control algorithms) to counter fast fading effects until the outer loop requires a fresh computation of mobile transmit powers to achieve new target SIRs, at which point the averaged algorithm in (24) can be used again.

For our convergence analysis of the averaging algorithm (24), we shall consider a fixed step size version that averages over all past estimates. Our main convergence result for the averaged algorithm is as follows.

Theorem 4: For the averaged algorithm in (24) and (25), when the target SIRs are feasible, and assuming perfect decision feedback for the nonlinear receivers

$$\lim_{n\to\infty} E\left[\|\tilde{\mathbf{w}}(n) - \mathbf{w}^*\|^2\right] = 0 \tag{26}$$

where the fixed step size a is s.t. $0 < a < 2\left(\min_i \mathrm{Re}\lambda_i(\mathcal{H})\right)^{-1}$.

An outline of the proof is provided in the Appendix. It follows that in [9], and some of the details that have been omitted can be found in the reference. The theorem implies that the MSE error goes to zero in the limit (as opposed to the nonaveraged algorithm, where the MSE can only be guaranteed to be bounded for a fixed step-size value). Although a discussion on "optimality" of the rate of convergence due to averaging will not be included here (and can be found in [8]–[11]), we will see in the following numerical examples section that the rate of convergence does significantly improve with averaging.

Based on more recent results on stochastic approximation [14], we can modify the power control algorithm in (24) to suitably exploit the averaged iterates in the basic recursion. This can be done with simultaneous use of averaging and feedback as suggested by [14], such that the basic recursion itself yields power estimates that benefit from the averaging step. This could possibly eliminate the need to wait till the end of the initial convergence interval, as with the simple averaging scheme. More specifically, for fixed step-size parameters $a > 0$ and $\mu > 0$, the two-step improved averaging algorithm that uses an additive correction for the basic recursion based on the difference between the basic and averaged iterates, would be

$$w_{ij}(n) = w_{ij}(n-1) + a\left(\frac{z_{ij}(n)\gamma_{ij}^*}{g_{ijj}|\mathbf{f}_{ij}^T\mathbf{s}_{ijj}|^2} - (1+\gamma_{ij}^*)w_{ij}(n-1)\right) + a\mu\left(\tilde{w}_{ij}(n-1) - w_{ij}(n-1)\right) \tag{27}$$

$$\tilde{w}_{ij}(n) = \frac{1}{n}\left((n-1)\tilde{w}_{ij}(n-1) + w_{ij}(n)\right). \tag{28}$$

With suitable choice of a and μ, the basic recursion in the first step of the above algorithm (27) alone can yield better power

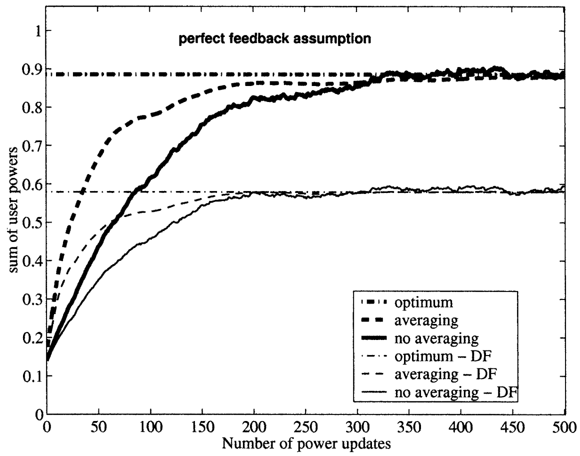


Fig. 3. Sum of powers as a function of iteration index in *Example 2* for (linear) single-cell decorrelating and (nonlinear) D-DF receivers. Common target SIR is 8. Optimum sum of powers for the linear receivers is 0.89 units, and that for the nonlinear ones is 0.58 units. Decision feedback is assumed perfect.

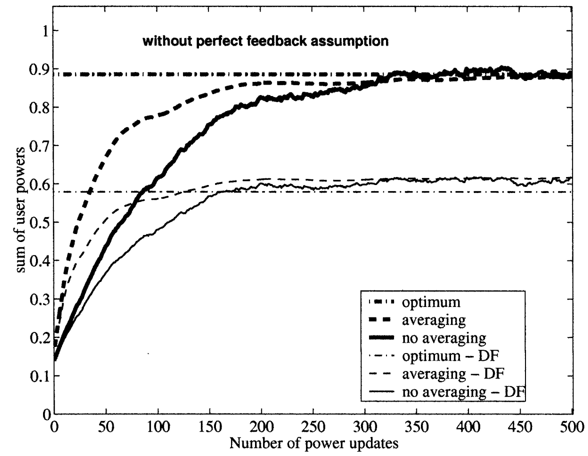


Fig. 4. Sum of powers as a function of iteration index in *Example 2* for (linear) single-cell decorrelating and (nonlinear) D-DF receivers. Common target SIR is 8. Note slight degradation in adaptation performance due to imperfect feedback.

estimates (in terms of convergence) than the nonaveraged recursions in (21) or in (24). In fact, even the averaged estimates from (28) can converge faster than those from (25). The convergence analysis in [14] provides some insight into the choice of the right step-size parameters, but a discussion of this will not be included here. We will, however, include a numerical example to show the improvement due to feedback combined with averaging over averaging alone.

VII. Numerical Examples

Example 2: We consider again the seven-cell system setup of *Example 1*. All the parameters that are not explicitly mentioned in the following description are assumed to remain the same. The processing gain is chosen to be 64, and once again, the sequences of in-cell users are linearly independent. We consider the single-cell decorrelating receiver (3) and the single-cell D-DF receiver (4) for each user for detection on the uplink. The user powers are updated every 20 symbol intervals. We consider the case when the target SIRs are all equal. We find from (15) that the maximum common target SIR that is feasible for the single-cell decorrelating and D-DF receivers is upper bounded by 10.84 and 12.86, respectively. We first choose a common target SIR of 8, which is well within the allowed limits for both the receivers. Fig. 3 shows the sum of the powers (assuming perfect feedback for the D-DF receiver) computed by the non-averaged (21) and averaged (24) power control algorithms. The optimum sum of powers [as yielded by a deterministic computation in (11)] is 0.89 units for the linear receiver and 0.58 units for the decision-feedback receiver. Nonlinear detection in this example is 52% more power efficient than linear detection. Note that for this example, a common target SIR of 8 is not feasible with MF (single-user) detection with or without decision feedback. We implement the fixed step-size versions of the power control algorithms in (21) and (24), and choose by trial and error the step-size values that give the best convergence results for each algorithm. The algorithm with averaging uses a suitably larger fixed step size (same for the linear and nonlinear receivers) than that for the nonaveraged version (again, same step size for the linear and nonlinear receivers), and as is evident from the figure, achieves a faster rate of convergence. A comparable rate of convergence for the nonaveraged algorithm with fixed step size in either case will come at the price of larger asymptotic MSE. The same example *without* the perfect feedback assumption in Fig. 4 shows that the power adaptation for the decision-feedback receiver does not suffer in a significant way when feedback is not assumed perfect.

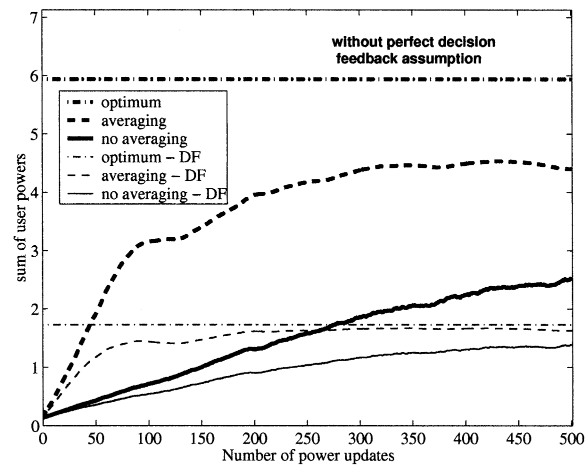


Fig. 5. Sum of powers as a function of iteration index in *Example 2* for single cell, linear decorrelating, and decision feedback decorrelating receivers. Common target SIR is 10. Optimum sum of powers for the linear receivers is 5.9 units, and that for the nonlinear ones is 1.7 units.

The advantage of using decision feedback becomes remarkably apparent as the SIR requirements are pushed higher. We next set the common target SIR for each user to 10, a value quite close to the allowed limit for the linear single-cell decorrelator. The optimum sum of powers [as yielded by a deterministic computation (11)] to achieve the target SIRs is 5.9 units for the linear decorrelating receiver, and 1.7 units for the decision-feedback decorrelating receiver—an improvement of almost 350% in power efficiency! Notice in Fig. 5 that averaging again achieves a significantly faster rate of convergence and

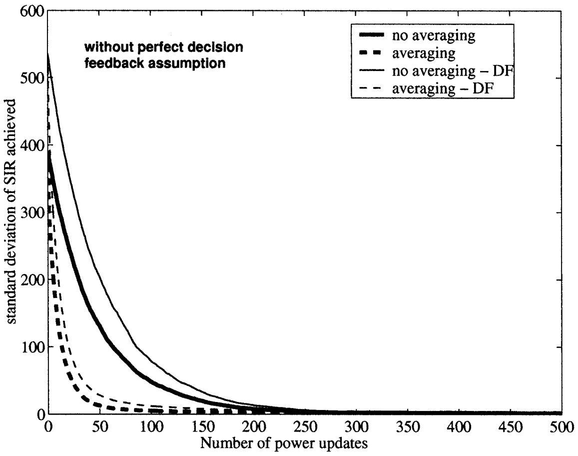


Fig. 6. Standard deviation (over users) of SIR achieved as a function of iteration index for *Example 2* when common target SIR is 10.

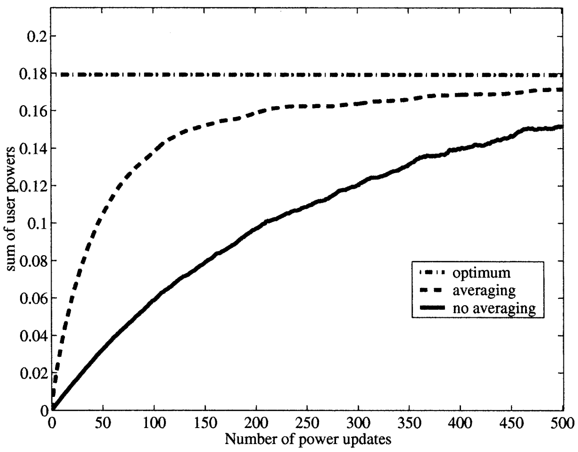


Fig. 7. Sum of powers as a function of iteration index for *Example 3*. Optimum sum of powers is 0.1792 units. The receivers employed are MF (single-user) receivers.

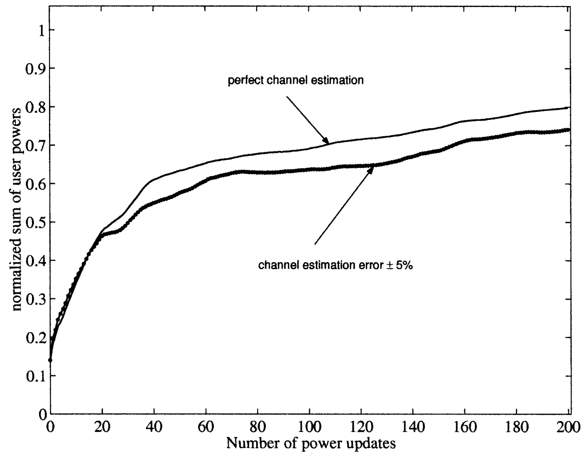


Fig. 8. Sum of user powers (normalized w.r.t. to sum of optimum powers) as a function of iteration index for *Example 4*. The receivers employed are single-cell decorrelators.

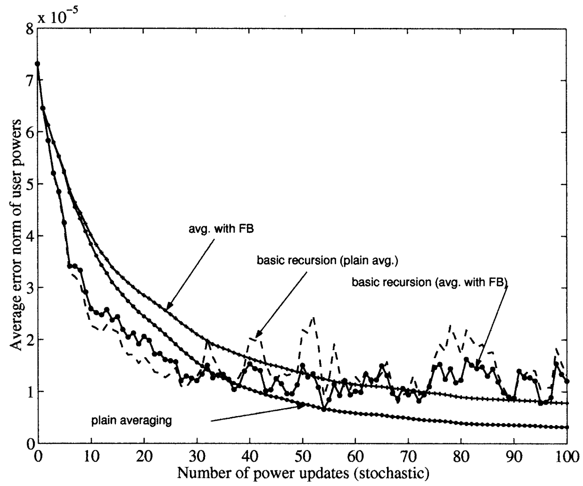


Fig. 9. Average error norm of user powers (w.r.t. to the optimum powers) as a function of iteration index for *Example 5*—comparison of averaging with feedback and plain averaging. The receivers employed are single-cell decorrelators.

smaller asymptotic MSE, although the convergence to the optimum powers in the linear single-cell decorrelation case is quite poor with and without averaging. Therefore, in this more demanding higher SIR scenario, decision feedback significantly outperforms the linear receiver case, both in terms of power efficiency, as well as convergence properties. In Fig. 6, we plot the standard deviation of the SIRs achieved over all the users, and note that the target SIRs are achieved significantly faster with averaging. In this example, if the coherence time of the frequency nonselective channel was 2000 symbol intervals (100 power updates), then without averaging, our power control algorithm would lead to an unacceptable SIR performance (with and without decision feedback), whereas, with averaging, the SIRs would be reasonably met for almost half the coherence interval. As in *Example 1*, these results were found to be typical for several randomly chosen signature-sequence assignments.

Example 3: We include, next, a smaller example (seven cells, five users in each cell) to show that averaging improves the performance of the power control algorithm for the MF receiver. In Fig. 7, we plot the performance of the averaged and nonaveraged power control algorithms (the latter is equivalent to the one in [6]) in terms of the sum of powers versus iteration index n. As in the previous example, averaging yields significant gains in terms of convergence.

Example 4: Consider another seven-cell example where each user employs the single-cell decorrelator. This time, we do not assume perfect knowledge of the channel gains in step one of the averaging algorithm in (24). Fig. 8 compares the performance of the averaging algorithm with unbiased channel gain estimates (in error up to $\pm 5\%$) with the perfect estimates case. The figure plots the sum of user powers (normalized w.r.t. the sum of the optimum powers) versus iteration index. The averaging algorithm is found to be fairly robust to small channel estimation errors. An analysis of the effect of channel gain estimation errors for the MF receiver-based, nonaveraged, power control algorithm can be found in [6], where it was shown that a large error variance could convert a feasible power control problem into an infeasible one.

Example 5: We finally investigate the performance of the averaging algorithm with feedback (27). Fig. 9 plots the average error norm of the user powers (w.r.t. the optimum powers) obtained from averaging with feedback and from averaging alone.

The receiver employed by every user in this example is the single-cell decorrelator. We plot the trajectories corresponding to both the basic recursion and the averaging step in (24) and (25), and both the basic recursion with feedback and the averaging step in (27) and (28). First, notice that the trajectory of the basic recursion in (24) fluctuates significantly (this is because the step size has been chosen aggressively so as to yield better estimates when "smoothed" by the averaging step). With feedback in algorithm (27) and the same step size a, the basic recursion shows a somewhat better convergence behavior. The feedback factor μ was adjusted by trial and error to try to get this performance. However, we do notice that there is a slight degradation in the feedback case when comparing the averaged iterates. The choice of step size and feedback factor seems to be critical to ensure performance improvement in both the basic recursion and the averaged iterate when using feedback, as suggested by the results in [14].

VIII. Conclusions

We have shown in this paper that a proper integration of multiuser receivers and power control algorithms to achieve specified SIR performance on the uplink of a cellular system can have a significant impact on the design of bandwidth- and power-efficient cellular networks. Moreover, stochastic power control algorithms that have been proposed earlier for MF as well as for multiuser receivers here can be considerably improved by using the technique of averaging for stochastic approximation. Since the averaging is achieved without any increase in the order of complexity and enables faster convergence than the corresponding nonaveraged algorithms, it makes the practical implementation of the stochastic power control algorithms a more attractive proposition.

Appendix

Proof of Theorem 3

Let Γ be a $K \times K$ diagonal matrix with the desired SIRs along its diagonal, i.e., $\mathrm{diag}\left([\gamma_{11}^*, \ldots, \gamma_{K_1 1}^*, \ldots, \gamma_{1B}^*, \ldots, \gamma_{K_B B}^*]\right)$. Similarly, let $\mathbf{z}(n) \triangleq [z_{11}(n), \ldots, z_{K_1 1}(n), \ldots, z_{1B}, \ldots, z_{K_B B}]^T$. We can rewrite the recursion in (21) as

$$\mathbf{w}(n) = \mathbf{w}(n-1) - a_n \left[(\mathbf{I} - \mathcal{X})\,\mathbf{w}(n-1) + \beta(n) - \mathbf{Dt}\right] \tag{29}$$

where the vector $\beta(n)$ is implicitly defined by the recursion as $\beta(n) = (\Gamma + \mathcal{X})\,\mathbf{w}(n-1) + \mathbf{Dt} - \mathbf{Dz}(n)$. $E\left[\beta(n)|\mathbf{w}(n-1) = \mathbf{w}\right] = \mathbf{0}$ and, therefore, $\beta(n)$ is a zero-mean vector. The matrix $\mathbf{A} \triangleq (\mathbf{I} - \mathcal{X})$ has eigenvalues given by $\lambda_i(\mathbf{A}) = 1 - \lambda_i(\mathcal{X})$. By the assumption that the power control algorithm with the chosen receivers is feasible, $|\lambda_i(\mathcal{X})| < 1$, and, therefore, $\mathrm{Re}(\lambda_i(\mathbf{A})) > 0$, or in other words, $-\mathbf{A}$ is a stable matrix. Noting that $(\mathbf{I} - \mathcal{X})\,\mathbf{w}^* = \mathbf{Dt}$, we can rewrite (29) in terms of the error, $\epsilon(n) \triangleq \mathbf{w}(n) - \mathbf{w}^*$, as

$$\epsilon(n) = (\mathbf{I} - a_n \mathbf{A})\,\epsilon(n-1) - a_n \beta(n). \tag{30}$$

Before proceeding any further, we note that a similar result (as in the *Theorem*) was proved in [6] for the case of stochastic power control using MF receivers, where the convergence analysis essentially followed the lines of [24]. Using Lyapunov's theorem for stable matrices and a "sandwiching" argument, the authors of that work showed for a fixed step size $a_n = a$, the asymptotic MSE is bounded as

$$\frac{c_0' a^2}{k_0' a - \alpha_0' a^2} \le \lim_{n\to\infty} E\left[\|\epsilon(n)\|^2\right] \le \frac{c_0 a^2}{k_0 a - \alpha_0 a^2} \tag{31}$$

where c_0, c_0', α_0, α_0', k_0, and k_0' are all suitable positive constants. We will adopt a somewhat different approach to prove our theorem, noting that this same approach will be used later to prove convergence for the averaging algorithm. Essentially, we use convergence results for the norm of a product of matrices resulting from the recursion of interest. First, we make the assumption that user powers are bounded above by a sufficiently large number, i.e., $0 < w_{ij}(n) \le \mathcal{K}_1 < \infty$ and our starting guess $w_{ij}(0) \le \mathcal{K}_1$ is nonrandom. The finite assumption for user powers holds, for instance, if whenever any user power computed by our stochastic recursion exceeds this upper bound, it has to be set equal to it. In practice, for suitably small step-size values and sufficiently small starting error, $\epsilon(0)$, our adaptive algorithm would never violate this large upper bound and it is only of interest for analytical reasons. With our assumptions, we can verify that the following are true:

$$E\left[\beta(n)|\beta(n-1), \ldots, \beta(1)\right] = \mathbf{0} \qquad \text{for } n > 1 \tag{32}$$

$$E\left[\|\beta(n)\|\right] < \infty \tag{33}$$

$$\sup_n E[\|\beta(n)\|^2|\beta(n-1), \ldots, \beta(1)] \le \mathcal{K}_2 < \infty \; a.s., \quad \text{for } n > 1. \tag{34}$$

$\beta(n)$ is, therefore, a martingale-difference process [26], where $E\left[\beta(1)\right] = \mathbf{0}$ and $E[\|\beta(1)\|^2] \le \mathcal{K}_2 < \infty$. Let $\{\mathbf{X}_j^n\}_{n\ge j}$ be the sequence of matrices $\mathbf{X}_j^n \in \mathbb{R}^{N\times N}$, determined by the following recursive relation:

$$\mathbf{X}_j^{n+1} = \mathbf{X}_j^n - a_n \mathbf{A} \mathbf{X}_j^n \tag{35}$$

where $\mathbf{X}_j^j = \mathbf{I}$. Note that the recursion implies that $\mathbf{X}_j^{n+1} = (\mathbf{I} - a_n\mathbf{A})(\mathbf{I} - a_{n-1}\mathbf{A}) \cdots (\mathbf{I} - a_j\mathbf{A})$. Further, from (30), we get that

$$\epsilon(n) = \mathbf{X}_1^{n+1}\epsilon(0) - \sum_{i=1}^{n} \mathbf{X}_{i+1}^{n+1} a_i \beta(i). \tag{36}$$

Taking norm squared and expectation, we get

$$\begin{aligned} E\left[\|\epsilon(n)\|^2\right] =& E\left[\epsilon(0)^T \left(\mathbf{X}_1^{n+1}\right)^T \left(\mathbf{X}_1^{n+1}\right)\epsilon(0)\right] \\ &- E\left[\sum_{i=1}^{n}\left[\epsilon(0)^T\left(\mathbf{X}_1^{n+1}\right)^T \mathbf{X}_{i+1}^{n+1} a_i \beta(i) \right.\right. \\ &\left.\left. + \beta(i)^T \left(\mathbf{X}_{i+1}^{n+1}\right)^T a_i \mathbf{X}_1^{n+1}\epsilon(0)\right]\right] \\ &+ E\left[\left(\sum_{i=1}^{n} \mathbf{X}_{i+1}^{n+1} a_i \beta(i)\right)^T \right. \\ &\left. \times \left(\sum_{j=1}^{n} \mathbf{X}_{j+1}^{n+1} a_j \beta(j)\right)\right]. \end{aligned} \tag{37}$$

The first term on the right-hand side is a deterministic term and can be bounded as

$$E\left[\epsilon(0)^T(\mathbf{X}_1^{n+1})^T(\mathbf{X}_1^{n+1})\epsilon(0)\right] \le \|\epsilon(0)\|^2\|\mathbf{X}_1^{n+1}\|^2. \tag{38}$$

Taking the expectation in the remaining two terms on the right-hand side of (37), first conditioned on $\{\beta(n-1) = \beta_{n-1}, \ldots, \beta(1) = \beta_1\}$ and then over $\{\beta(n-1), \ldots, \beta(1)\}$, we obtain (since, $\beta(n)$ is a martingale-difference sequence) that the second term is zero, and the third term is equal to

$$E\left[\|\beta(1)a_1\mathbf{X}_2^{n+1}\|^2\right] + \sum_{i=2}^{n} E\left[\cdots E\left[E\left[\|\beta(i)a_i\mathbf{X}_{i+1}^{n+1}\|^2|\beta(i-1) = \beta_{i-1}, \ldots, \beta(1) = \beta_1\right]\right]\cdots\right]. \quad (39)$$

Along the lines of [9, p. 845] we can show the following.

Lemma 1: With the step-size conditions in *Theorem 3*, and $\mathrm{Re}(\lambda_i(\mathbf{A})) > 0$, there are constants $c_1 > 0$ and $0 < \mathcal{K}_3 < \infty$, such that for all j and sufficiently large $n \geq j$

$$\|\mathbf{X}_j^n\| \leq \mathcal{K}_3 \exp\left(-c_1 \sum_{i=j}^{n-1} a_i\right). \quad (40)$$

Proof: By Lyapunov's theorem [25] for stable matrices, we know that there exists a symmetric, positive-definite matrix $\mathbf{G}$ that solves the equation $\mathbf{GA} + \mathbf{A}^T\mathbf{G} = \mathbf{I}$. Define $L_1 \triangleq \max \lambda_i(\mathbf{G})$, $l_1 \triangleq \min \lambda_i(\mathbf{G})$, and $\mathbf{U}_n = (\mathbf{X}_j^n)^T\mathbf{X}_j^n$. Then

$$\mathbf{U}_{n+1} = \left(\mathbf{X}_j^n\right)^T (\mathbf{I} - a_n\mathbf{A})^T\mathbf{G}(\mathbf{I} - a_n\mathbf{A})\left(\mathbf{X}_j^n\right) \quad (41)$$

$$\mathbf{U}_n - a_n\left(\left(\mathbf{X}_j^n\right)^T \mathbf{A}^T\mathbf{G} + \mathbf{GA}\right) + a_n^2(\mathbf{X}_j^n)^T\mathbf{A}^T\mathbf{GAX}_j^n. \quad (42)$$

Note that $\left(\mathbf{X}_j^n\right)^T \mathbf{X}_j^n \geq (1/L_1)\left(\mathbf{X}_j^n\right)^T \mathbf{GX}_j^n$ and $\left(\mathbf{X}_j^n\right)^T \mathbf{A}^T\mathbf{GAX}_j^n \leq c_2\left(\mathbf{X}_j^n\right)^T \mathbf{GX}_j^n$, where $c_2 = \|\mathbf{A}\|^2 L_1/l_1$. Then, for n sufficiently large and some $\lambda > 0$, we get

$$\begin{aligned}\mathbf{U}_{n+1} &\leq \mathbf{U}_n\left(1 - \frac{1}{L}a_n + c_2a_n^2\right)\\ &\leq (1-\lambda a_n)\mathbf{U}_n \leq e^{-\lambda a_n}\mathbf{U}_n.\end{aligned} \quad (43)$$

Thus, $\mathbf{U}_n \leq \mathbf{U}_j \exp(-\lambda \sum_{i=j}^{n-1} a_i)$. However,

$$\|\mathbf{U}_n\| \geq l_1\|\mathbf{X}_j^n\|^2 \text{ and } \|\mathbf{U}_n\| \leq L_1\|\mathbf{X}_j^j\|^2 = L. \quad (44)$$

So, we obtain that

$$\|\mathbf{X}_j^n\| \leq \sqrt{\frac{L_1}{l_1}} \exp\left(-\frac{\lambda}{2}\sum_{i=j}^{n-1} a_i\right). \quad (45)$$

With this result and the fact that $\sum_1^\infty a_i = \infty$, the upper bound in (38) [and, therefore, the first term on the right-hand side of (37)] can be seen to go to zero as $n \to \infty$. Using (34), we can bound (39) by $\mathcal{K}_2 \sum_{i=1}^n \|a_i\mathbf{X}_{i+1}^{n+1}\|^2$, and *Lemma 1* further implies that

$$\begin{aligned}&\lim_{n\to\infty} \mathcal{K}_2 \sum_{i=1}^{n} \|a_i\mathbf{X}_{i+1}^{n+1}\|^2\\ &\leq \mathcal{K}_2\mathcal{K}_3 \lim_{n\to\infty} \sum_{i=1}^{n} a_i^2 \exp\left(-c_1\sum_{j=i+1}^{n} a_j\right) \quad (46)\\ &= \mathcal{K}_2\mathcal{K}_3 \sum_{i=1}^{\infty} a_i^2 \lim_{n\to\infty} \exp\left(-c_1\sum_{j=i+1}^{n} a_j\right)\\ &= 0. \quad (47)\end{aligned}$$

We could interchange the limit and the summation in (47), since the sum on the right side of (46) is absolutely convergent. This completes our proof.

Proof of Theorem 4

We first note that this proof borrows some ideas and definitions from the proof of *Theorem 3*. Further, to avoid repetition, we will provide only a sketch of the proof here, encouraging the reader to peruse the exposition in [9] for details.

We consider the fixed step-size case, $a_n = a$, chosen to be $0 < a < 2\left(\min_i \mathrm{Re}\lambda_i(\mathbf{A})\right)^{-1}$. In addition to $\epsilon(n) = (\mathbf{I} - a\mathbf{A})\,\epsilon(n-1) - a\beta(n)$ and $\mathbf{X}_j^n = (\mathbf{I} - a\mathbf{A})^{n-j}$, defined as in the proof of *Theorem 3*, we define $\bar{\epsilon}(n) = \bar{\mathbf{w}}(n) - \mathbf{w}^*$, $\bar{\mathbf{X}}_j^n = a\sum_{i=j}^{n-1} \mathbf{X}_j^i$, and $\phi_j^n = \mathbf{A}^{-1} - \bar{\mathbf{X}}_j^n$. Again, the feasibility of the target SIRs implies that $\mathrm{Re}(\lambda_i(\mathbf{A})) > 0$.

Lemma 2: There is a constant $\mathcal{K}_4 < \infty$ such that for all j and sufficently large $n \geq j$

$$\|\phi_j^n\| \leq \mathcal{K}_4 \quad (48)$$

$$\lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} \phi_j^n = \mathbf{0}. \quad (49)$$

Proof:

$$\begin{aligned}\bar{\mathbf{X}}_j^n &= a\left(\mathbf{I} + (\mathbf{I} - a\mathbf{A}) + \cdots + (\mathbf{I} - a\mathbf{A})^{n-j}\right)\\ &= \mathbf{A}^{-1} - (\mathbf{I} - a\mathbf{A})^{n-j+1}\mathbf{A}^{-1}.\end{aligned} \quad (50)$$

The eigenvalues of the matrix $\mathbf{I} - a\mathbf{A}$ are $\lambda_i(\mathbf{I} - a\mathbf{A}) = \mathbf{I} - \lambda_i(\mathbf{A})$ and $|\lambda_i(\mathbf{I} - a\mathbf{A})| < 1$. So

$$\lim_{n\to\infty} (\mathbf{I} - a\mathbf{A})^n = \mathbf{0}. \quad (51)$$

This proves (48). Further

$$\begin{aligned}\frac{1}{n}\sum_{j=0}^{n-1} \phi_j^n &= \frac{1}{n}\sum_{j=0}^{n-1} (\mathbf{I} - a\mathbf{A})^{n-j+1}\mathbf{A}^{-1}\\ &= \frac{1}{n}\sum_{j=2}^{n+1} (\mathbf{I} - a\mathbf{A})^{j}\mathbf{A}^{-1} \to \mathbf{0} \text{ as } n \to \infty.\end{aligned} \quad (52)$$

Now, consider the error of the averaging algorithm in (30)

$$\bar{\epsilon}(n) = \frac{1}{n}\sum_{i=0}^{n} \epsilon(i). \quad (53)$$

Lemma 3: The following is true of the error with averaging:

$$\sqrt{n}\bar{\epsilon}(n) = \frac{1}{\sqrt{n}a}\mathcal{W}_n\epsilon(0) + \frac{1}{\sqrt{n}}\sum_{j=1}^{n-1} \mathbf{A}^{-1}\beta(j) + \frac{1}{\sqrt{n}}\sum_{j=1}^{n-1} \mathcal{Z}_j^n\beta(j) \quad (54)$$

where $\mathcal{W}_n$, $\mathcal{Z}_j^n \in \mathbb{R}^{N\times N}$, are such that $\|\mathcal{W}_n\| \leq \mathcal{K}_5$, $\|\mathcal{Z}_j^n\| \leq \mathcal{K}_5$ for some $\mathcal{K}_5 < \infty$, and

$$\frac{1}{n}\sum_{j=1}^{n-1} \|\mathcal{Z}_j^n\| \to \mathbf{0}, \qquad \text{as } n \to \infty. \quad (55)$$

The proof of the above *Lemma* can be found in [8, p. 846]. With this result, and denoting $\kappa = (1/\sqrt{n})\sum_{j=1}^{n-1} \mathbf{A}^{-1}\beta(j)$ and

$$nE\bar{\epsilon}^T(n)\bar{\epsilon}(n) = E\|\kappa\|^2 + \tau_n \tag{56}$$

we can further show [using in the process the martingale-difference property of the vector sequence $\beta(n)$] that $\tau_n \to 0$ as $n \to \infty$. Therefore

$$\begin{aligned}\lim_{n\to\infty} nE\bar{\epsilon}^T(n)\bar{\epsilon}(n) &= \lim_{n\to\infty}\frac{1}{n}\sum_{j=1}^{n-1} E\beta(j)^T(\mathbf{A}^{-1})^T \\ &\quad \times \mathbf{A}^{-1}\beta(j) \qquad (57)\\ &\le \lim_{n\to\infty}\frac{1}{n}\|\mathbf{A}^{-1}\|^2\sum_{j=1}^{n-1}E\|\beta(j)\|^2 \\ &\le \frac{1}{n}\|\mathbf{A}^{-1}\|^2 n\mathcal{K}_2 < \infty. \qquad (58)\end{aligned}$$

This implies the claim in our theorem, i.e., $\lim_{n\to\infty} E\left[\|\tilde{\mathbf{w}}(n) - \mathbf{w}^*\|^2\right] = 0$.

References

[1] S. V. Hanly and D. N. Tse, "Power control and capacity of spread spectrum wireless networks," *Automatica*, vol. 35, no. 12, pp. 1987–2012, Dec. 1999.

[2] P. S. Kumar and J. Holtzman, "Power control for a spread spectrum system with multiuser receivers," in *Proc. IEEE Personal Indoor and Mobile Radio Communications Conf.*, Sept. 1995, pp. 955–959.

[3] M. K. Varanasi, "Power control for multiuser detection," in *Proc. Conf. Information Science and Systems*, Princeton, NJ, Mar. 1996, pp. 866–873.

[4] S. Ulukus and R. Yates, "Adaptive power control with SIR maximizing multiuser detectors," in *Proc. Conf. Information Science and Systems*, Baltimore, MD, Mar. 1997.

[5] M. Varanasi, "Nonlinear multiuser receivers with distributed power control in cellular radio networks," in *Proc. 37th Annu. Allerton Conf. on Communication, Control and Computers*, Allerton, IL, Sept. 1999, pp. 820–830.

[6] S. Ulukus and R. Yates, "Stochastic power control for cellular radio systems," *IEEE Trans. Commun.*, vol. 46, pp. 784–798, June 1998.

[7] D. Ruppert, "Efficient estimators from a slowly convergent Robbins–Monro procedure," School of Oper. Res. and Ind. Eng., Cornell Univ., Ithaca, NY, Rep. 781, 1988.

[8] B. T. Polyak, "New stochastic approximation type procedures" (in Russian), *Automat I Telemekh*, vol. N7, pp. 98–107, 1990.

[9] B. T. Polyak and A. B. Juditsky, "Acceleration of stochastic approximation by averaging," *SIAM J. Contr. Optim.*, vol. 30, no. 4, pp. 838–855, 1992.

[10] H. J. Kushner and J. Yang, "Stochastic approximation with averaging of the iterates: Optimal asymptotic rate of convergence for general processes," *SIAM J. Contr. Optim.*, vol. 31, no. 4, pp. 1045–1062, 1993.

[11] L. Gyorfi and H. Walk, "On the averaged stochastic approximation for linear regression," *SIAM J. Contr. Optim.*, vol. 34, no. 1, pp. 31–61, 1996.

[12] A. J. Viterbi, *CDMA: Principles of Spread Spectrum Communication*. Englewood Cliffs, NJ: Prentice-Hall, 1995.

[13] H. Holma and A. Toskala, *WCDMA for UMTS: Radio Access for Third Generation Mobile Communications*. New York: Wiley, 2001.

[14] H. J. Kushner and J. Yang, "Stochastic approximation with averaging and feedback: Rapidly convergent on-line algorithms," *IEEE Trans. Automat. Contr.*, vol. 40, pp. 24–34, Jan. 1995.

[15] M. Varanasi, "Stochastic power control for nonlinear multiuser receivers in cellular receivers in cellular radio networks," in *Proc. IEEE Information Theory Workshop*, South Africa, June 20–25, 1999.

[16] M. K. Varanasi and T. Guess, "Optimum decision feedback multiuser equalization with successive decoding achieves the total capacity of the Gaussian multiple-access channel," in *Proc. 31st Annu. Asilomar Conf. Signals, Systems, and Computers*, Pacific Grove, CA, Nov. 1997, pp. 1405–1409.

[17] M. K. Varanasi, "Decision feedback multiuser detection: A systematic approach," *IEEE Trans. Inform. Theory*, vol. 45, pp. 219–241, Jan. 1999.

[18] U. Madhow and M. L. Honig, "MMSE interference suppression for direct-sequence spread-spectrum CDMA," *IEEE Trans. Commun.*, vol. 38, pp. 509–519, Apr. 1990.

[19] E. Seneta, *Non-Negative Matrices*. New York: Wiley, 1973.

[20] R. Yates, "A framework for uplink power control in cellular radio systems," *IEEE J. Select. Areas Commun.*, vol. 13, pp. 1341–1347, Sept. 1995.

[21] L. A. Hageman and D. M. Young, *Applied Iterative Methods*. New York: Academic, 1971.

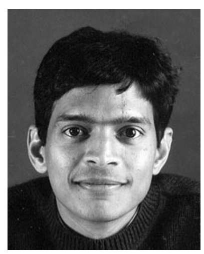
[22] G. Foschini and Z. Miljanic, "A simple distributed autonomous power control algorithm and its convergence," *IEEE Trans. Veh. Technol.*, vol. 42, pp. 641–646, Nov. 1993.

[23] H. Robbins and S. Monro, "A stochastic approximation method," *Ann. Math. Statist.*, vol. 22, pp. 400–407, 1951.

[24] D. J. Sakrison, "Stochastic approximation: A recursive method for solving regression problems," in *Advances in Communications Systems 2*, A. V. Balakrishnan, Ed. New York: Academic, 1966, pp. 51–106.

[25] R. Horn and C. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1993.

[26] J. Doob, *Stochastic Processes*. New York: Wiley, 1953.



Mahesh K. Varanasi (S'87–M'89–SM'95) received the B.E. degree in electronics and communication engineering from Osmania University, Hyderabad, India, in 1984, and the M.S. and Ph.D. degrees in electrical engineering from Rice University, Houston, TX, in 1987 and 1989, respectively.

In 1989, he joined the faculty of the College of Engineering and Applied Sciences, University of Colorado, Boulder, where he is currently Professor of Electrical and Computer Engineering. His teaching and research interests are in communication theory, information theory, and signal processing. His research has been in the areas of multiuser detection, space–time communications, equalization, signal design, and power- and bandwidth-efficient multiuser communications.

Deepak Das (S'96–M'01) received the B.E. (with honors) degree in electrical and electronics engineering from the Birla Institute of Technology and Science, Pilani, India, in 1992, and the M.S. and Ph.D. degrees in electrical and computer engineering from the University of Colorado at Boulder, in 1997 and 2000, respectively.

From 1992 to 1995, he was a Software Consultant at Mahindra-British Telecom, and worked with British Telecom, London, U.K. He is currently with the Wireless Advanced Technologies Laboratory, Lucent Technologies, Whippany, NJ. His research interests have been primarily in the areas of multiuser detection, power control, and intelligent antennas.