Distributed Transmit Beamforming Using Feedback Control

Raghuraman Mudumbai, *Member, IEEE*, Joao Hespanha, *Fellow, IEEE*, Upamanyu Madhow, *Fellow, IEEE*, and Gwen Barriac, *Member, IEEE*

***Abstract*—The concept of distributed transmit beamforming is implicit in many key results of network information theory. However, its implementation in a wireless network involves the fundamental challenge of ensuring phase coherence of the radio frequency signals from the different transmitters in the presence of unknown phase offsets between the transmitters and unknown channel gains from the transmitters to the receiver. In this paper, it is shown that such phase alignment can be achieved using distributed adaptation by the transmitters with minimal feedback from the receiver. Specifically, each transmitter independently makes a small random adjustment to its phase at each iteration, while the receiver broadcasts a single bit of feedback, indicating whether the signal-to-noise ratio (SNR) improved or worsened after the current iteration. The transmitters keep the "good" phase adjustments and discard the "bad" ones, thus implementing a distributed ascent algorithm. It is shown that, for a broad class of distributions for the random phase adjustments, this procedure leads to asymptotic phase coherence with probability one. A simple analytical model, borrowing ideas from statistical mechanics, is used to characterize the progress of the algorithm, and to provide guidance on parameter choices. This analytical model is based on a conjecture on the distribution of the received phases when the number of transmitters becomes large. Finally, the proposed system is shown to be scalable: the random phase perturbations can be chosen such that the convergence time is linear in the number of collaborating nodes.**

***Index Terms*—Distributed beamforming, sensor networks, space–time communication, synchronization.**

I. Introduction

DISTRIBUTED transmit beamforming refers to a form of cooperative communication in which transmitters agree upon a common message, and then transmit it such that their signals add up coherently at the receiver. Such constructive interference leads to a factor of N gain in power efficiency, where N is the number of collaborating transmitters. Thus, if the power of each transmitter is fixed, then distributed beamforming leads to an N^2 gain in received signal-to-noise ratio (SNR): a factor of N gain due to increase in total transmit power, and a factor of N gain in power efficiency due to increased directivity from beamforming. In essence, the transmitters organize themselves as a virtual antenna array, and cooperate with each other to focus their transmission towards the intended receiver.

Distributed beamforming using virtual arrays is at the heart of both classical and recent results in network information theory. For instance, the communication model for the Gaussian relay channel [1], [2], and some of the optimal coding schemes for the large-scale ad-hoc network [3] are all implicitly based on distributed beamforming, since they assume that the transmitted signals add up coherently at the receiver. However the technical feasibility of this assumption has received little attention until relatively recently. This has perhaps been the most important barrier to the implementation of information-theoretic schemes in wireless networks.

To achieve beamforming with distributed transmitters, it is necessary to compensate for unknown channel gains from each transmitter to the receiver as well as unknown phase offsets *between* the transmitters. The latter offsets arise from the fact that each transmitter generates its carrier signal from a separate local oscillator, and therefore has no fixed phase relationship with the others. It is possible to obtain carrier signals that are synchronized in *frequency* [4], using a *master-slave* architecture, where the *slave* transmitters use phase locked loops (PLLs) to lock to a reference carrier signal broadcast by a *master* transmitter. However this process still leaves unknown phase offsets between the carrier signals because of unknown propagation delays in the master-slave channels. Furthermore the accuracy of conventional timing synchronization techniques e.g., using GPS, is adequate only for beamforming at very low frequencies (on the order of 10 MHz), and makes such methods inapplicable to communication at the higher RF frequencies (~ 1.0 GHz) used in most wireless systems. Without first correcting for these unknown offsets, it is fundamentally impossible [5], [6] to use multiple-input multiple-output (MIMO)-like methods to measure the channel gains (e.g., using reciprocity). This is, perhaps, the most important difference between centralized and distributed beamforming.

In this paper, we investigate a simple iterative procedure, based on feedback from the receiver, for achieving phase coherence, and show that this procedure provides a powerful method

Manuscript received January 21, 2008; revised March 03, 2009. Current version published December 23, 2009. This work was supported by the National Science Foundation under Grants CCF-0431205 and CNS-0520335, by the Office of Naval Research under Grant N00014-06-1-0066, and by the Institute for Collaborative Biotechnologies through Grant DAAD19-03-D-0004 from the U.S. Army Research Office.

R. Mudumbai is with the Department of Electrical and Computer Engineering, University of Iowa, Iowa City, IA 52242 USA (e-mail: rmudumbai@engineering.uiowa.edu).

J. Hespanha and U. Madhow are with the Department of Electrical and Computer Engineering, University of California, Santa Barbara, CA 93106 USA (e-mail: hespanha@ece.ucsb.edu; madhow@ece.ucsb.edu).

G. Barriac is with Qualcomm Inc., San Diego, CA 92121 USA (e-mail: gbarriac@qualcomm.com).

Communicated by A. J. Goldsmith, Associate Editor for Communications.

Color versions of Figs. 2 and 5 in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2009.2034786

to satisfy the requirements for distributed beamforming. The basic idea behind the feedback algorithm is as follows: each transmitter adjusts its phase randomly at each iteration and the receiver broadcasts one bit of feedback per iteration indicating whether its net SNR is better or worse than before. If it is better, all transmitters keep their latest phase perturbations, otherwise they all undo the phase perturbation. This randomized ascent procedure is repeated until the transmitters converge to phase coherence. This procedure is especially appealing because it avoids the previously mentioned difficulties in channel estimation due to the unknown phase offsets; by using SNR measurements, it completely removes the need for any explicit channel estimation procedure.

The preceding algorithm was first introduced in our earlier work [7], and the present paper focuses on developing a fundamental understanding of this algorithm in an idealized setting, in which the RF carrier signals of the different transmitters are assumed to be synchronized in *frequency,* with constant (but unknown) relative phase offsets between transmitters, constant (but unknown) channel gains to the receiver and error-free estimation of SNR at the receiver. Our main results are as follows.

1) We show that for a broad class of distributions for the random phase perturbations used by the transmitters, the distributed adaptation converges to phase coherence with probability one.
2) We develop an analytical framework for characterizing the dynamics of the algorithm which provides excellent agreement with simulation results, and allows for optimization of algorithm parameters. The key steps are as follows.
   a) We use a version of the Central Limit Theorem to show that when the number of transmitters becomes large, the effect of the random phase perturbations is an additive Gaussian perturbation to the received signal amplitude.
   b) We then use the Gibbs conditioning principle of statistical mechanics to derive a probability distribution that we conjecture applies to the received phases under the feedback algorithm. While we are unable to prove the conjecture, we present theoretical arguments as well as extensive numerical simulations to show its plausibility.
   c) Using 2a) and 2b), we derive a simple expression for the expected convergence rate of the algorithm. Using this expression, we show that the convergence time of the algorithm is linear in the number of transmitters, so that the procedure scales well for large networks. For an optimized (time-varying) choice for the distribution of the phase perturbations, we also show that convergence is locally exponential, with a time constant smaller than $3N$.

While the above idealized assumptions allow us to obtain analytical insights on the convergence behavior of the algorithm, it is worth noting that the algorithm itself can be easily adapted to be robust to noise, phase jitter, quantization and estimation errors and is also capable of tracking a time-varying channel. This robustness was experimentally demonstrated in a proof-of-concept prototype [8], where transmitters using separate PLLs to obtain a local oscillator signal from a common clock signal, and implement a slightly modified version of the feedback algorithm to beamform towards the receiver which estimates the received signal strength and periodically broadcasts a 1-bit feedback signal to the transmitters. In the modified algorithm, the receiver measures the SNR averaged over a large number of symbols to minimize the effect of noise, and also checks if its SNR estimate is greater than previous estimates over a *finite* window of time, unlike the idealized algorithm where the present SNR estimate is compared against the entire past. In this case, even in the presence of noise, phase jitter, estimation error and quantization errors both in phase as well as in SNR estimation, the algorithm achieved more than 90% of the maximum possible beamforming gains. Another variant of the feedback algorithm was demonstrated in [9], where the receiver feedback was used to synchronize the carrier *frequencies* in addition to the phases.

Furthermore, simulation results show that for low levels of noise, estimation errors and channel variations, the analytical model derived using idealized assumptions accurately predicts the *initial* convergence of the algorithm. It is only when the algorithm gets close to convergence that these impairments become important: whereas the idealized algorithm asymptotically converges to full coherence, the robust version, in general, may reach a steady state where the beamforming gains fluctuate around a level less than the idealized maximum. While the analytical methods developed in this paper can be extended to model this steady-state, and preliminary results along these lines were reported in [8], we defer a complete analytical treatment to future work.

It is worth noting that beamforming with a centralized antenna array of N elements, requires $\mathcal{O}(N)$ bits or training symbols [10] to learn N unknown channel gains. Surprisingly, for our 1-bit feedback algorithm with optimized phase perturbations the average time to convergence (and consequently the number of bits) also scales as $\mathcal{O}(N)$. While such linear scaling can also be achieved using a scheme where the receiver estimates and feeds back the phase of the transmitters one at a time, the feedback algorithm offers some key advantages over this alternative approach as follows.

1) It avoids the need for coordination among the transmitters for training which can be very difficult for a large number of transmitters.
2) When the signal from individual transmitters is too weak, it is difficult for the receiver to obtain the phase estimates. Under the feedback algorithm, the receiver only needs to estimate the strength of the aggregate signal which is usually much stronger.
3) The feedback algorithm does not require a dedicated training phase; thus the transmitters can send data to the receiver during the beamforming process and the receiver can easily perform SNR estimation using the data-carrying signal. This is especially important for large networks where the training phase can be quite long.

Related work. As mentioned previously, many results [2], [3] in network information theory are implicitly based on distributed beamforming. However, it is only recently that the importance of the synchronization problem for beamforming has been recognized, and the effects of synchronization errors been systematically studied. It was shown in [11] that even partial

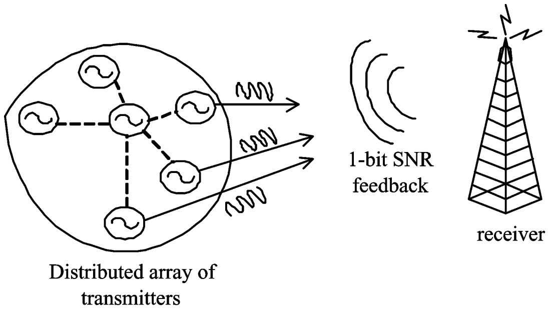

Fig. 1. Phase synchronization using receiver feedback.

phase synchronization leads to significant increase in power efficiency in wireless ad hoc networks. In our own previous work [4], we proposed a *master-slave* architecture for frequency synchronization, and analyzed the effect of phase noise on the beamforming gain. We also showed that the SNR gains are substantial even with imperfect synchronization [5]. A method of phase synchronization for beamforming with two transmitters is presented in [12]. While most work on distributed beamforming focuses on the directivity gains from beamforming, the authors in [13] adopt a different approach and examine the statistics of the side-lobes of the resulting beam patterns because of random node placement. Thus, the results in [13] provide insight not only into the SNR at the intended receiver, but also the interference at other locations.

Our feedback algorithm can be considered as a distributed version of a *stochastic approximation* algorithm such as the classical Robbins–Monro algorithm [14]. Even though we use a different analytical technique based on statistical mechanics, our model for convergence is partly motivated by the "mean ODE" method from the literature on stochastic approximation [15]. Recently, other authors have proposed some interesting variations to our feedback algorithm that are also motivated by the relationship with stochastic approximation [16]. A stochastic beamforming algorithm for the (centralized) MIMO downlink channel was proposed in [17]. Extensions of our feedback algorithm to distributed spatial multiplexing [18] and wireless relay networks [19] have also been recently proposed. In addition, other authors [20], [21] have independently derived analytical proofs for the convergence and scaling properties of the feedback algorithm. As already mentioned, the algorithm in this paper has been prototyped in [8], [9]. Finally, [6] provides a tutorial survey of the state of art and open issues for distributed transmit beamforming.

The rest of this paper is organized as follows. The feedback control algorithm for distributed beamforming is described in Section II-A and its asymptotic convergence is established in Section II-B. Section III presents an analytical model for the algorithm dynamics based on statistical ideas. This model is motivated by the predictable behavior of the algorithm when the number of transmitters is large. In Section III-A, it is shown that a version of the Central Limit Theorem applies to the variations in the received signal as a result of the phase perturbations. This leads to a simple formula for the average convergence rate derived in Section III-B. The statistical analysis depends on an "Exp-Cosine" conjecture on the phase distributions which is discussed in Section III-C. Simulation results are presented to show that the Exp-Cosine distribution closely matches the empirical histogram of the received phases, and also that the convergence rate of the algorithm is accurately predicted by the statistical model. Section IV uses this model to establish that the convergence time of the algorithm is linear on the number of transmitters, and to find an optimized distribution for the phase perturbations. Section V concludes the paper with a short discussion of open issues.

II. Feedback Control Protocol

As shown in Fig. 1, we consider a system of N transmitters transmitting a common narrowband message signal $m(t)$ to a Base Station receiver. Specifically, all the transmitters simultaneously send RF signals, each obtained by modulating a carrier with a scaled version of the message. The transmitters are organized in a master-slave architecture that assures the carrier signals are synchronized in *frequency*. The baseband signal of transmitter i can be written as $s_i(t) = Ae^{j\theta_i}m(t)$. Our goal is to adjust the complex gains $Ae^{j\theta_i}$ so as to achieve phase coherence at the receiver. We ignore distortions in the message due to small timing mismatches[1] between the transmitters, which allows us to ignore the presence of the message in what follows.

We can assume that each transmitter sends at a fixed power determined by a power constraint, which we normalize to unity i.e., $A = 1$. We note that because the transmitters obtain their RF carrier from different local oscillators, their carrier signals have unknown phase offsets between them. As discussed earlier, this is true even though carrier frequency synchronization among the transmitters is established using the master-slave architecture. The effect of this phase offset is that the phase of the base-band signal transmitted from transmitter i gets rotated by an unknown amount γ_i.

We denote the complex channel gain of transmitter i to the receiver as $h_i = a_i e^{j\psi_i}$, where $a_i \geq 0$ represents the attenuation and ψ_i the phase response of the wireless channel. The received signal due to transmitter i is given by $s_i(t)\, e^{j\gamma_i}\, h_i = a_i e^{j(\theta_i+\gamma_i+\psi_i)}\, m(t)$, with the overall received signal at the receiver resulting from the superposition of the signals received from each transmitter.

The net complex gain at the receiver is therefore given by

$$Y \triangleq \left|\sum_{i=1}^{N} a_i e^{j(\theta_i+\gamma_i+\psi_i)}\right| = \left|\sum_{i=1}^{N} a_i e^{j\Phi_i}\right|$$

where $Y \geq 0$ is the amplitude, or received signal strength (RSS), and $\Phi_i = \theta_i + \gamma_i + \psi_i$ is the phase at the receiver corresponding to the signal from transmitter i. Note that the RSS only depends on the unknowns γ_i and ψ_i through the sum $\gamma_i + \psi_i$; nevertheless we write them separately to emphasize their different physical origins.

[1]This requirement of time synchronization is unrelated to the phase synchronization required for beamforming; timing errors cause some inter-symbol interference and message signal distortion, but do not affect the beamforming gain.

Our objective is to adapt the transmitter phases $\{\theta_i\}$ so as to maximize Y. This happens if the received carrier phases Φ_i are all equal

$$Y = \left|\sum_{i=1}^{N} a_i e^{j\Phi_i}\right| \leq Y_{\text{opt}} \triangleq \left(\sum_{i=1}^{N} a_i\right)$$
$$\text{with equality if and only if } \Phi_i = \Phi_{\text{const}} \quad \forall\, i. \quad (1)$$

The purpose of the feedback control algorithm is to allow transmitter i to dynamically compute the optimal value of θ_i in (1), without requiring knowledge of either ψ_i or γ_i.

A. Description of Algorithm

The adaptation is performed in time-slotted fashion, with each transmitter adapting its phase in a timeslot in response to feedback from the receiver. At the beginning of slot n, let $\theta_i[n]$ denote the best known carrier phase at transmitter i. At each timeslot n, each transmitter i applies a random phase perturbation $\delta_i[n]$ to $\theta_i[n]$ in order to probe for a potentially better phase. The transmitted "probe" phase in slot n is then given by

$$\theta_i^{\text{probe}}[n] = \theta_i[n] + \delta_i[n].$$

The corresponding RSS is given by $Y[n] = \left|\sum_i a_i e^{j\Phi_i[n]}\right|$, where $\Phi_i[n] = \theta_i^{\text{probe}}[n] + \gamma_i + \psi_i$. The receiver measures $Y[n]$, and broadcasts one bit of feedback indicating whether $Y[n]$ is bigger or smaller than its record of the highest observed signal strength so far, which we denote by

$$Y_{\text{best}}[n] \triangleq \max_{k<n} Y[k].$$

If the feedback from the receiver indicates an improvement in RSS, then the transmitters keep their random phase perturbations, otherwise they undo their perturbations. Thus, the best known phases at the transmitters are updated as follows:

$$\theta_i[n+1] = \begin{cases} \theta_i[n] + \delta_i[n], & Y[n] > Y_{\text{best}}[n] \\ \theta_i[n], & \text{otherwise.} \end{cases} \quad (2)$$

Simultaneously, the receiver also updates its record of the highest RSS so far as follows:

$$Y_{\text{best}}[n+1] = \max\left(Y_{\text{best}}[n], Y[n]\right). \quad (3)$$

The preceding procedure is repeated over multiple timeslots. Equations (2) and (3) ensure that we retain phase perturbations that increase RSS, while discarding unfavorable ones. This distributed ascent procedure eventually converges to a set of transmit phases that satisfy (1) and achieve distributed beamforming. Fig. 2 shows the convergence to beamforming with $N = 10$ transmitters.

The random perturbations $\delta_i[n]$ are chosen independently across transmitters from a symmetric probability distribution $\delta_i[n] \sim g_n(\delta_i)$, where the density function $g_n(\delta_i)$ is a parameter of the protocol. We show in Section IV that the behavior of the algorithm is mostly characterized by the variance of the distribution $g_n(\delta_i)$ and depends only weakly on the actual distribution. In general, the distribution $g_n(\delta_i)$ can be adapted dynamically in time to optimize the speed of convergence (cf. Section IV-A).

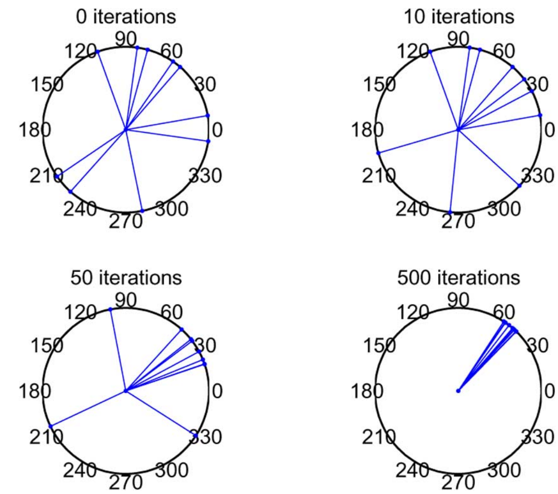

Fig. 2. Convergence of feedback control algorithm.

It follows from (2) that if the algorithm were to be terminated at timeslot n, the best achievable signal strength using the feedback information received so far, is equal to $Y_{\text{best}}[n]$, which corresponds to transmitter i transmitting with the phase $\theta_i[n]$

$$Y_{\text{best}}[n] = \left|\sum_i a_i e^{\Phi_i[n]}\right|, \text{ where } \Phi_i[n] = \theta_i[n] + \gamma_i + \psi_i \quad (4)$$

B. Asymptotic Coherence

We now show that the feedback control protocol outlined in Section II-A asymptotically achieves phase coherence for any initial values of the phases Φ_i. We define some notation first.

Let $\bar{\Phi}$ denote the N-vector of the received phase angles Φ_i. We define the function $\text{RSS}(\bar{\Phi})$ to be the received signal strength corresponding to received phase $\bar{\Phi}$:

$$\text{RSS}(\bar{\Phi}) \triangleq \left|\sum_i a_i e^{j\Phi_i}\right|. \quad (5)$$

Phase coherence means $\Phi_i = \Phi_{\text{const}}, \forall\, i$, for some arbitrary phase constant Φ_{const}. In order to remove this ambiguity, it is convenient to work with a vector $\bar{\phi}$ of rotated phase values:

$$\phi_i \triangleq \Phi_i - \Phi_0 \quad (6)$$

where Φ_0 is a constant chosen such that the phase of the total received signal is zero. This is just a convenient shift of the receiver's phase reference and as (5) shows, such a shift has no impact on the received signal strength, i.e., $\text{RSS}(\bar{\phi}) = \text{RSS}(\bar{\Phi})$, $\forall\, \bar{\Phi}$. This phase shift permits two simple expressions for the RSS which will be useful in the sequel: for every vector $\bar{\phi}$ of rotated phase values and every vector $\bar{\delta}$ of phase perturbations, we have that

$$\begin{aligned} \sum_i a_i e^{j\phi_i} &= \sum_i a_i \cos(\phi_i) \geq 0 \\ \sum_i a_i e^{j(\phi_i+\delta_i)} &= \gamma \sum_i a_i \cos(\phi_i) + \sum_i a_i e^{j\phi_i}\left(e^{j\delta_i} - \gamma\right) \end{aligned} \quad (7)$$

where γ can be any constant.

We interpret the feedback control algorithm as a discrete-time vector random process $\bar{\phi}[n]$, where $\bar{\phi}[n]$ is a N-dimensional vector of phases $\phi_i[n]$ constrained by the condition that the total phase of the received signal is zero as defined in (6). This random process is a Markov process because the phase perturbations $\bar{\delta}[n]$ are chosen independently at each timeslot n.

We now provide an argument that (under appropriate conditions on the probability density function $g_n(\delta_i)$), shows that $\{Y_{\text{best}}[n]\}$ converges *almost surely* to the constant Y_{opt} for arbitrary initial phases $\bar{\phi}[0]$. (Note that $Y_{\text{best}}[n] \to Y_{\text{opt}}$ is equivalent to $\bar{\phi}[n] \to 0$.) The following proposition will be needed to establish convergence. Roughly speaking, it states that as long as the received phases $\phi_i[n]$ are not fully coherent, there is always a finite probability of obtaining a finite increase in RSS in every timeslot.

Proposition 1: Suppose that the density $g_n(\delta_i)$ is bounded away from zero over an interval $(-\Delta_0, \Delta_0)$, where $\Delta_0 > 0$. Then, for any $\epsilon > 0$, there exist positive constants ϵ_1, ρ for which

$$\begin{aligned} \text{RSS}(\bar{\varphi}) \leq & Y_{\text{opt}} - \epsilon \\ & \Rightarrow \Pr\left(Y_{\text{best}}[n+1] - Y_{\text{best}}[n] \geq \epsilon_1 \middle| \bar{\phi}[n] = \bar{\varphi}\right) \\ & \geq \rho, \quad \forall\, \bar{\varphi}. \end{aligned}$$

□

Proof of Proposition 1: See Appendix A. ■

Theorem 1: For the class of distributions $g_n(\delta_i)$ considered in Proposition 1, starting from an arbitrary $\bar{\phi}$, the feedback algorithm converges to perfect coherence of the received signals *almost surely*, i.e., $Y_{\text{best}}[n] \to Y_{\text{opt}}$ or equivalently $\bar{\phi}[n] \to \bar{0}$ (i.e., $\phi_i[n] \to 0, \forall\, i$) with probability 1. □

Proof of Theorem 1: Pick some $\epsilon > 0$ and define the random variable $t_\epsilon \in [0, \infty)$ to be the first integer n for which the monotone nondecreasing sequence $Y_{\text{best}}[n]$ becomes strictly larger than $Y_{\text{opt}} - \epsilon$. Proposition 1 shows that there exists some $\mu(\epsilon) > 0$ such that

$$\begin{aligned} \text{RSS}(\bar{\varphi}) \leq & Y_{\text{opt}} - \epsilon \\ & \Rightarrow \text{E}\left[Y_{\text{best}}[n+1] - Y_{\text{best}}[n] \middle| \bar{\phi}[n] = \bar{\varphi}\right] \\ & \geq \mu(\epsilon), \quad \forall\, \bar{\varphi} \end{aligned}$$

since we can take $\mu(\epsilon) \triangleq \epsilon_1 \rho$. Since by the definition of t_ϵ, $Y_{\text{best}}[n] = \text{RSS}(\bar{\phi}[n]) \leq Y_{\text{opt}} - \epsilon, \forall\, n < t_\epsilon$, we conclude that

$$\text{E}\left[Y_{\text{best}}[n+1] - Y_{\text{best}}[n]\right] \geq \mu(\epsilon), \quad \forall\, n < t_\epsilon.$$

From this and the fact that $Y_{\text{best}}[n]$ is bounded above by Y_{opt}, $Y_{\text{best}}[n+1] - Y_{\text{best}}[n] \geq 0$, and $Y_{\text{best}}[0] \geq 0$ with probability one, we have that

$$\begin{aligned} Y_{\text{opt}} \geq & \text{E}\left[\lim_{n\to\infty} Y_{\text{best}}[n]\right] = \text{E}\left[Y_{\text{best}}[0]\right] \\ & + \text{E}\left[\sum_{n=0}^{\infty} \left(Y_{\text{best}}[n+1] - Y_{\text{best}}[n]\right)\right] \\ \geq & \text{E}\left[Y_{\text{best}}[0]\right] + \text{E}\left[\sum_{n=0}^{t_\epsilon - 1} \left(Y_{\text{best}}[n+1] - Y_{\text{best}}[n]\right)\right] \\ \geq & \text{E}\left[Y_{\text{best}}[0]\right] + \mu(\epsilon)\text{E}[t_\epsilon] \end{aligned}$$

from which we conclude that

$$\begin{aligned} \text{E}[t_\epsilon] &= \sum_{x=0}^{\infty} x \Pr(t_\epsilon = x) \equiv \sum_{x=1}^{\infty} \Pr(t_\epsilon \geq x) \\ &\leq \frac{Y_{\text{opt}} - \text{E}\left[Y_{\text{best}}[0]\right]}{\mu(\epsilon)} < \infty \quad \Rightarrow \quad \lim_{n\to\infty} \Pr(t_\epsilon \geq n) = 0 \end{aligned}$$

It then follows that

$$\begin{aligned} \lim_{n\to\infty} & \Pr\left(\sup_{k\geq n} |Y_{\text{opt}} - Y_{\text{best}}[k]| \geq \epsilon\right) \\ &= \lim_{n\to\infty} \Pr(Y_{\text{opt}} - Y_{\text{best}}[n] \geq \epsilon) \\ &= \lim_{n\to\infty} \Pr(t_\epsilon \geq n) = 0. \end{aligned} \tag{8}$$

where the first equality is due to the monotonicity of $\{Y_{\text{best}}[k]\}$. From (8), we finally conclude that $Y_{\text{best}}[n] \to Y_{\text{opt}}$ with probability one (cf. [22, Theorem 1, p. 253]). ■

III. Dynamics of the RSS

Although Theorem 1 guarantees that $Y_{\text{best}}[n]$ converges to the optimum RSS Y_{opt}, this result sheds little light on how long this will take. We now present an analytical model for the convergence rate that allows us to choose the distribution of the perturbations $\delta_i[n]$ for fast convergence, and to study the scalability of the algorithm with the number of transmitters N. Since the RSS at convergence, Y_{opt}, scales linearly with N, it is convenient to work with the normalized RSS, $\frac{1}{N} Y_{\text{best}}[n]$.

The key steps in our argument are as follows.

1) We use a version of the central limit theorem to characterize (for large N) the conditional distribution of $Y_{\text{best}}[n+1]$, conditioned on the value of $Y_{\text{best}}[n]$.
2) Noting that the normalized RSS is tightly clustered around its mean, we introduce a "mean ODE" style equation to define a deterministic sequence which tracks the evolution of the *average* normalized RSS, $\frac{1}{N} E\left[Y_{\text{best}}[n]\right]$.
3) We then present a statistical characterization of the phases $\{\phi_i[n]\}$; specializing to a system with equal gains, we observe that these phases are interchangeable random variables. Using a statistical argument, we then derive a simple single-parameter "Exp-Cosine" probability distribution that we conjecture applies universally to the phases $\phi_i[n]$ under the feedback algorithm.
4) Using this conjectured conditional distribution for $\phi_i[n]$, we derive an analytical expression for the average normalized RSS. This expression is the main result in this section, and will be used to obtain insights on the convergence rate of the algorithm and to optimize it.

We start by considering the variations in the RSS due to the phase perturbations $\delta_i[n]$ when N is large.

A. Central Limit Theorem Based Characterization

Consider the received signal $\sum_{i=1}^{N} a_i e^{j(\phi_i[n] + \delta_i[n])}$. We begin by using (7) with $\gamma = \chi_n$, to express the effect of the phase perturbations $\delta_i[n]$ as an increase or decrease in the RSS, combined with a rotation of the complex received signal (see Fig. 3) as shown in (9)–(11) at the bottom of the next page.

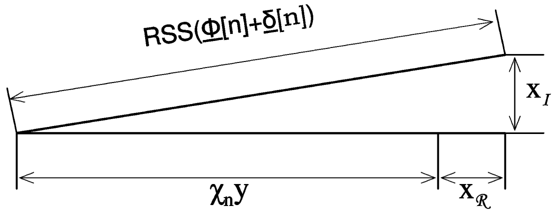

Fig. 3. Effect of phase perturbations on the total received signal.

Proposition 2: Conditioned on $\frac{1}{N}Y_{\text{best}}[n] = y$, the random variables $\sqrt{N}x_{\Re}$ and $\sqrt{N}x_{\Im}$ tend in distribution to zero mean Gaussian random variables as $N \rightarrow \infty$ with the following variances:

$$\begin{aligned}\sigma^2_{\Re}[n] &\triangleq \mathrm{Var}_{y,n}[x_{\Re}[n]]\\ &= \frac{1}{2N^2}\sum_i a_i^2\left(1-\chi_n^2-\rho_n \mathrm{E}_{y,n}[\cos(2\phi_i[n])]\right) \quad (12)\\ \sigma^2_{\Im}[n] &\triangleq \mathrm{Var}_{y,n}[x_{\Im}[n]]\\ &= \frac{1}{2N^2}\sum_i a_i^2\left(1-\chi_n^2+\rho_n \mathrm{E}_{y,n}[\cos(2\phi_i[n])]\right) \quad (13)\end{aligned}$$

where the subscript in $\mathrm{Var}_{y,n}[\cdot]$ indicates conditioning on $\frac{1}{N}Y_{\text{best}}[n] = y$. □

Proof of Proposition 2: The terms in all the summations that define $x_{\Re}[n]$ and $x_{\Im}[n]$ in (10)–(11) are of the form $x_i z_i$ where the x_i are sines or cosines of $\phi_i[n]$ (not necessarily independent) and the z_i only depend on the $\delta_i[n]$. Since the $\delta_i[n]$ are chosen i.i.d. from a symmetric distribution, all the z_i have zero mean and are independent of each other and of all the x_i. This means that the sequence $\{x_i z_i\}$ is a uniformly bounded martingale difference, i.e.,

$$\mathrm{E}[x_{i+1}z_{i+1}|x_1z_1,x_2z_2,\ldots,x_iz_i]=0, \quad |x_iz_i|\le 1, \quad \forall\, i. \quad (14)$$

We can therefore apply the CLT for sums of dependent variables in [22, Theorem 1, p. 541] to $x_{\Re}$ and $x_{\Im}$. Equations (12) and (13) then follow from straightforward trigonometric algebra (see Appendix C). ■

We now define the conditional expectation of the increment in normalized RSS as

$$\begin{aligned}h_n(y) &= \frac{1}{N}\mathrm{E}_{y,n}\left[Y_{\text{best}}[n+1]-Y_{\text{best}}[n]\right]\\ &\triangleq \frac{1}{N}\mathrm{E}\left[Y_{\text{best}}[n+1]-Y_{\text{best}}[n]\,\middle|\,\frac{1}{N}Y_{\text{best}}[n]=y\right].\end{aligned} \quad (15)$$

Here, and in the sequel, we use the subscripted notation $\mathrm{E}_{y,n}[\cdot]$ as a shorthand for the conditioning on $\frac{1}{N}Y_{\text{best}}[n] = y$. With this definition, we have

$$\begin{aligned}&\frac{1}{N}E[Y_{\text{best}}[n+1]]\\ &\quad= \frac{1}{N}E[Y_{\text{best}}[n]] + E\left[h_n\left(\frac{Y_{\text{best}}[n]}{N}\right)\right]\\ &\quad\approx \frac{1}{N}E[Y_{\text{best}}[n]] + h_n\left(\frac{1}{N}E[Y_{\text{best}}[n]]\right)\end{aligned} \quad (16)$$

where the last approximation is based on the observation that $Y_{\text{best}}[n]$ is highly concentrated around its expected value when N is large. A rigorous proof of this concentration property is left as an open problem, however we provide a heuristic justification at the end of Section III-B. Equation (16) suggests that we can model the evolution of the normalized RSS by the sequence y_n defined by the recursion:

$$y_{n+1} \triangleq y_n + h_n(y_n), \text{ with the initialization } y_0 \triangleq \frac{1}{\sqrt{N}}. \quad (17)$$

Using (16), we see that if $y_n = \frac{1}{N}E[Y_{\text{best}}[n]]$, then it follows that $y_{n+1} \approx \frac{1}{N}E[Y_{\text{best}}[n+1]]$. The initialization condition in (17) follows from the assumption that the initial phases $\phi_i[0]$ are independent and random in $(-\pi,\pi]$ which gives $E[Y_{\text{best}}[0]] = \sqrt{N}$. Equation (17) is analogous to the "mean-ODE" that is commonly used for convergence analysis in stochastic-approximation theory [15].

We emphasize that while the RSS sequence $Y_{\text{best}}[n]$ is a stochastic process, y_n is a *deterministic* sequence that models the

$$\frac{1}{N}\mathrm{RSS}(\bar{\phi}[n]+\bar{\delta}[n]) \equiv \frac{1}{N}\left|\sum_{i=1}^{N} a_i e^{j(\phi_i[n]+\delta_i[n])}\right| = \left|\frac{1}{N}\chi_n Y_{\text{best}}[n] + x_{\Re}[n] + jx_{\Im}[n]\right| \quad (9)$$

$$\begin{aligned}\text{where } x_{\Re}[n] &\triangleq \Re\left[\frac{1}{N}\sum_{i=1}^{N} a_i e^{j\phi_i[n]}(e^{j\delta_i[n]}-\chi_n)\right] = \frac{1}{N}\sum_{i=1}^{N} a_i \Re[e^{j\phi_i[n]}(e^{j\delta_i[n]}-\chi_n)]\\ &= \frac{1}{N}\left(\sum_{i=1}^{N} a_i\cos(\phi_i[n])(\cos(\delta_i[n])-\chi_n) - \sum_{i=1}^{N} a_i \sin(\phi_i[n])\sin(\delta_i[n])\right) \quad (10)\\ x_{\Im}[n] &\triangleq \Im\left[\frac{1}{N}\sum_{i=1}^{N} a_i e^{j\phi_i[n]}(e^{j\delta_i[n]}-\chi_n)\right] = \frac{1}{N}\sum_{i=1}^{N} a_i \Im[e^{j\phi_i[n]}(e^{j\delta_i[n]}-\chi_n)]\\ &= \frac{1}{N}\left(\sum_{i=1}^{N} a_i\cos(\phi_i[n])\sin(\delta_i[n]) + \sum_{i=1}^{N} a_i \sin(\phi_i[n])\cos(\delta_i[n])\right). \quad (11)\end{aligned}$$

average convergence rate of the normalized RSS as indicated by (16).

We now present the main result of the analytical model that provides an explicit expression for $h_n(y)$.

B. Computation of RSS Increment

Given $\frac{1}{N}Y_{\mathrm{best}}[n] = y$, we have seen in (9) that the normalized RSS is given by

$$\frac{1}{N}\mathrm{RSS}(\bar{\phi}[n] + \bar{\delta}[n]) = |\chi_n y + x_{\Re}[n] + jx_{\Im}[n]|. \tag{18}$$

Consider the inequality $\sqrt{1+z} \leq 1 + \frac{z}{2}, \forall\, z \geq -1$. Letting $z = \frac{2\chi_n y x_{\Re} + x_{\Re}^2 + x_{\Im}^2}{(\chi_n y)^2}$, we get

$$\begin{aligned}|\chi_n y + x_{\Re}[n] + jx_{\Im}[n]| &\equiv \chi_n y\sqrt{1+z} \leq \chi_n y \left(1 + \frac{z}{2}\right)\\ &\equiv \chi_n y + x_{\Re} + \frac{x_{\Re}^2 + x_{\Im}^2}{2\chi_n y}.\end{aligned} \tag{19}$$

Therefore we have the following bounds:

$$\chi_n y + x_{\Re} \leq \frac{1}{N}\mathrm{RSS}(\bar{\phi}[n] + \bar{\delta}[n]) \leq \chi_n y + x_{\Re} + \frac{x_{\Re}^2 + x_{\Im}^2}{2\chi_n y}. \tag{20}$$

We have observed that the variances $\sigma_{\Re}^2[n]$ and $\sigma_{\Im}^2[n]$ are $O(1/N)$. This implies that the last term in (20) can be neglected compared to the second term $x_{\Re}$. Specifically we can rewrite (20) as

$$\begin{aligned}\sqrt{N}x_{\Re} &\leq \sqrt{N}\left(\frac{1}{N}\mathrm{RSS}\left(\bar{\phi}[n] + \bar{\delta}[n]\right) - \chi_n y\right)\\ &\leq \sqrt{N}x_{\Re} + \frac{\sqrt{N}\left(x_{\Re}^2 + x_{\Im}^2\right)}{2\chi_n y}\end{aligned} \tag{21}$$

where $\sqrt{N}x_{\Re}$ is a zero-mean random variable whose variance $N\sigma_{\Re}^2 = \frac{1-\chi_n^2-\rho_n\kappa(y)}{2}$ is independent of N, and the last term $\frac{\sqrt{N}\left(x_{\Re}^2+x_{\Im}^2\right)}{2\chi_n y}$ is a nonnegative random variable whose mean vanishes for large N, and therefore converges to zero *in probability*. The preceding argument leads to the following equality in the limit of large N (conditioned on $\frac{1}{N}Y_{\mathrm{best}}[n] = y$)

$$\frac{1}{N}\mathrm{RSS}(\bar{\phi}[n] + \bar{\delta}[n]) \to_p (\chi_n y + x_{\Re}[n]) \tag{22}$$

where the limit indicates that $\sqrt{N}\left(\frac{1}{N}\mathrm{RSS}(\bar{\phi}[n] + \bar{\delta}[n]) - \chi_n y\right)$ converges *in probability* to $\sqrt{N}x_{\Re}$.

Theorem 2: In the limit of large N, the expected increment of the normalized RSS is given by

$$\begin{aligned}h_n(y) &= \sigma_{\Re}[n]\, i\left(\frac{y(1-\chi_n)}{\sigma_{\Re}[n]}\right)\\ \text{where } i(x) &\triangleq \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}} - xQ(x).\end{aligned} \tag{23}$$

In the above $Q(x) \triangleq \int_x^\infty \frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}\,dt$ denotes the complementary cumulative distribution function of a standard Gaussian random variable, and $\sigma_{\Re}[n]$ represents the mean deviation in the RSS because of the random perturbations $\delta_i[n]$ and is given by

$$\sigma_{\Re}^2[n] = \frac{1}{2N^2}\sum_{i=1}^{N} a_i^2\left(1 - \chi_n^2 - \rho_n \mathrm{E}_{y,n}\left[\cos(2\phi_i[n])\right]\right) \tag{24}$$

and the parameters $\chi_n \triangleq E\left[\cos(\delta_i[n])\right]$, and $\rho_n \triangleq \chi_n^2 - E\left[\cos(2\delta_i[n])\right]$ are functions of the distribution $g_n(\delta_i)$. □

Proof of Theorem 2: Using (15), (22), and Proposition 2, we have

$$\begin{aligned}h_n(y) &= \frac{1}{N}E_{y,n}\left[Y_{\mathrm{best}}[n+1]\right] - y\\ &= E_{y,n}\left[\max\left(y, \frac{1}{N}\mathrm{RSS}(\bar{\phi}[n] + \bar{\delta}[n])\right)\right] - y\\ &= E_{y,n}\left[\max\left(0, x_{\Re} - y(1-\chi_n)\right)\right]\\ &= \int_{y(1-\chi_n)}^{\infty}(x - y(1-\chi_n))\\ &\quad\times \frac{1}{\sqrt{2\pi\sigma_{\Re}^2[n]}}e^{-\frac{x^2}{2\sigma_{\Re}^2[n]}}\,dx.\end{aligned} \tag{25}$$

Carrying out the integration in (25) gives (23). ■

Similarly we can also show that

$$\begin{aligned}\frac{1}{N^2}\mathrm{Var}_{y,n}\left[Y_{\mathrm{best}}[n+1]\right] &= \sigma_{\Re}^2\, j\left(\frac{y(1-\chi_n)}{\sigma_{\Re}[n]}\right)\\ \text{with } j(x) &\triangleq \left((1+x^2)Q(x) - \frac{x}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\right)\\ &\quad - (i(x))^2\end{aligned} \tag{26}$$

where $i(x)$ is defined as in (23).

Remark. Since $i(x)$ and $j(x)$ are both bounded, and $\sigma_{\Re}$ decreases to zero as N becomes large, it follows from (23) and (26), that both the (normalized) mean RSS increment $h_n(y)$, and its variance decrease to zero as N becomes large. Therefore it takes a large number of timeslots ΔT_N for the normalized mean RSS to increase from y to $y + \Delta y$, where $\Delta y \ll 1$ is a small increment.

Since the perturbations $\delta_i[n]$ in each timeslot are chosen independently, the expected value and the variance of the RSS increments both add up over time. Assuming that $g_n(\delta_i)$ does not change significantly over the ΔT_N timeslots under consideration, the total expected increment in the normalized RSS is roughly $\Delta T_N \sigma_{\Re} i(x) \equiv \Delta y$ and the variance of the total increment over the same M timeslots is roughly $\Delta T_N \sigma_{\Re}^2 j(x) \equiv \Delta y \sigma_{\Re}\frac{j(x)}{i(x)}$.

The ratio $\frac{j(x)}{i(x)}$ depends on how the distribution $g_n(\delta_i)$ is chosen for each N. As we show in Section IV, when $g_n(\delta_i)$ is optimally chosen to maximize the expected RSS increment for a given value of N, the ratio $\frac{j(x)}{i(x)} = O(1)$. Since $\sigma_{\Re}$ decreases to zero for large N, the variance of the total RSS increments decreases to zero as N becomes large. Simulation results indicate that the same is true when some fixed distributions $g_n(\delta_i)$ are used over a range of values of N.

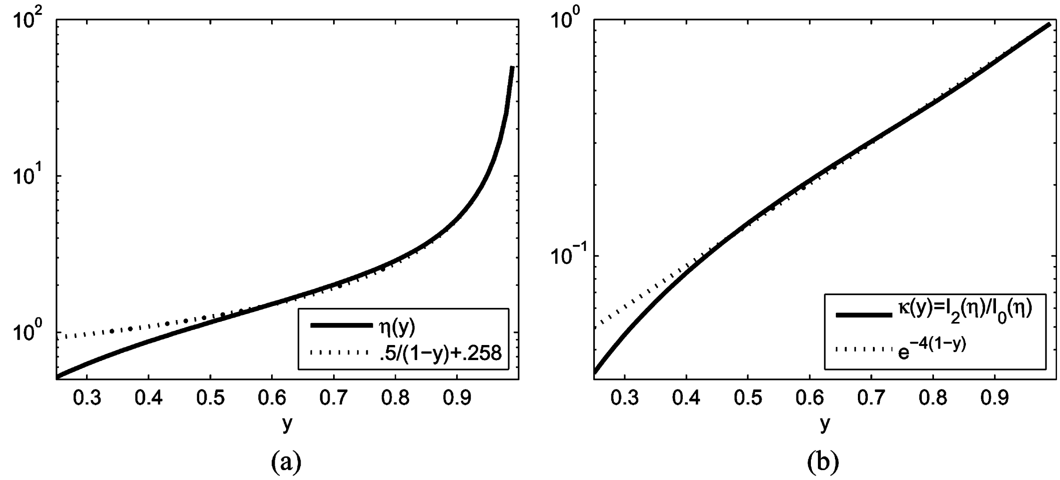

Fig. 4. Functions $\eta(y)$ such that $\frac{I_1(\eta)}{I_0(\eta)} = y$ fig:eta and $\kappa(y) = \frac{I_2(\eta)}{I_0(\eta)}$ with η such that $\frac{I_1(\eta)}{I_0(\eta)} = y$ (b). For $y \geq 0.5$, $\kappa(y) \approx e^{-4(1-y)}$. (a) $\eta(y)$. (b) $\kappa(y)$.

These observations indicate that the RSS increments, when averaged over many timeslots become almost deterministic; this justifies the approximation (16), and hence the definition (17) where $\frac{1}{N}E[Y_{\rm best}[n]]$ is modeled by a *conditional* expectation. A rigorous analysis of this concentration result is beyond our scope here. We leave this as an open problem.

The variances in (12)–(13) can be computed given the marginal distribution of the phases $\{\phi_i[n]\}$, conditioned on $\frac{1}{N}Y_{\rm best}[n] = y$. One possible approach for doing this is to keep track of the distribution of all N phases as they evolve over timeslots. However, the problem can be greatly simplified by exploiting symmetry. In particular, let us specialize to a system with equal channel gains, $a_i \equiv 1$. All transmitters begin with uniformly distributed phases, and the symmetry is preserved by the evolution of the algorithm, which depends only on the RSS seen by the receiver. This implies that, conditioned on $\frac{1}{N}Y_{\rm best}[n] = y$, the phases $\{\phi_i[n]\}$ are identically distributed, interchangeable random variables. The variances in (12)–(13) can then be rewritten as

$$\sigma^2_{\Re}[n] = \frac{1 - \chi_n^2 - \rho_n\kappa_n(y)}{2N}$$
$$\text{and } \sigma^2_{\Im}[n] = \frac{1 - \chi_n^2 + \rho_n\kappa_n(y)}{2N}. \tag{27}$$

where

$$\kappa_n(y) \triangleq \mathrm{E}_{y,n}\left[\cos(2\phi_1[n])\right]. \tag{28}$$

Note that $\kappa_n(y) \in [-1, 1]$. Intuitively, if the RSS is large, then we expect the phases to be close to zero, and κ to be close to one, whereas for small RSS, we expect the phases to exhibit a large variation, with κ being close to zero. In order to complete our description of the dynamics of the feedback algorithm, we must provide a method for computing $\kappa_n(y)$, which requires characterization of the marginal distribution of the transmitter phases conditioned on the RSS. This is addressed in Section IV.

C. The Exp-Cosine Distribution

The following Conjecture says that, for a large number of transmitters the rotated phases ϕ_i follow an "Exp-Cosine" distribution, when conditioned on the RSS.

Conjecture 1 (Exp-Cosine): For a sufficiently large number N of transmitters, the marginal distribution of $\{\phi_i[n]\}$, conditioned on $y_{\rm best}[n] = y \in (0, 1)$ is given by

$$f_n(\varphi|y) \triangleq \frac{e^{\eta(y)\cos\varphi}}{2\pi I_0(\eta(y))}, \quad \forall\, \varphi \in (-\pi, \pi] \tag{29}$$

with $\eta(y)$ chosen such that $\frac{I_1(\eta(y))}{I_0(\eta(y))} = y$, and $I_k(\cdot)$ is the modified Bessel function of the first kind and order k

$$I_k(x) \triangleq \frac{1}{2\pi}\int_{-\pi}^{\pi} \cos(k\varphi)e^{x\cos\varphi}d\varphi. \tag{30}$$

□

As we can see in Fig. 4, as $y \to 1$, the constant $\eta(y) \to +\infty$ and the distribution (29) becomes increasingly concentrated around $\varphi = 0$ and eventually converges to a Dirac-delta distribution. This is consistent with the fact that to get $Y_{\rm best}[n] = N$, all phases must be perfectly aligned.

From (29) and (30) we have the following property for the Exp-Cosine distribution

$$E\left[\cos(k\varphi)\middle|y\right] = \frac{I_k(\eta(y))}{I_0(\eta(y))}. \tag{31}$$

Using $k = 2$ in (31) and (28) gives us the desired expression for $\kappa(y)$ (based on our conjecture, $\kappa_n(y)$ depends only on y and not on n, hence we drop the dependence on n from the notation)

$$\kappa(y) \equiv \frac{I_2(\eta(y))}{I_0(\eta(y))}. \tag{32}$$

Fig. 4(b) shows the variation of $\kappa(y)$ with y.

We now present the heuristic reasoning behind the Exp-Cosine conjecture. The argument is motivated by the Gibbs conditioning principle (see [23] and the references therein) of statistical mechanics. However the version of the Gibbs principle used in this derivation requires conditional independence of the $\phi_i[n]$, which is not strictly satisfied under the feedback algorithm. A rigorous derivation of the Exp-Cosine distribution appears to require a detailed, problem-specific, large deviations analysis that is beyond the scope of this paper and that we leave as an open problem.

The Gibbs Conditioning Principle: Suppose that $\{X_1, X_2, \ldots, X_N\}$ are independent and identically distributed (i.i.d.) real-valued random variables with marginal distribution p and f and complex valued-function. Then, conditioned on

$$\left|\frac{1}{N}\sum_{i=1}^{N} f(X_i)\right| = y$$

the $\{X_1, X_2, \ldots, X_N\}$ are approximately identically distributed with the following marginal distribution:

$$q^* = \arg\min_q D(q\|p) \text{ subject to } |\mathrm{E}_q[f(X)]| = y \tag{33}$$

where $E_q[f(X)] \triangleq \int q(x)f(x)dx$ denotes the expectation with respect to the distribution q and $D(q\|p) \triangleq \int q(x)\log\frac{q(x)}{p(x)}dx$ denotes the divergence between the distributions p and q. The "approximately identically distributed" property refers to convergence in probability as $N \to \infty$. The variational problem (33) can be rewritten using Lagrange multipliers as,

$$\begin{aligned} q^* = \arg\min_q & \int q(x)\log\frac{q(x)}{p(x)}dx \\ & - \frac{\eta_1}{2}\left(\left|\int q(x)f(x)dx\right|^2 - y^2\right) \\ & - \eta_2\left(\int q(x)dx - 1\right) \end{aligned} \tag{34}$$

and its solution is given by

$$q^*(x) = ce^{\eta\Re[f(x)e^{-j\beta}]}p(x), \quad \forall\, x \tag{35}$$

where c, η, and β are normalizing constants, chosen so that

$$\int q^*(x)dx = 1, \quad \int q^*(x)f(x)dx = ye^{j\beta} \tag{36}$$

(cf. Appendix B). Suppose now that at a given iteration n of the algorithm, the rotated phases $\bar{\phi}[n] \triangleq \{\phi_i[n]\}$ are i.i.d. random variables and with conditional distribution $f_n(\varphi)$ conditioned to the current and all the past normalized RSS $\{y_{\text{best}}[0], y_{\text{best}}[1], \ldots, y_{\text{best}}[n]\}$. Consequently, the perturbed phases $\{\phi_i[n] + \delta_i[n]\}$ are also i.i.d. with marginal conditional distribution $(f_n * g_n)(\varphi)$, resulting from the circular convolution over $(-\pi, \pi]$ between $f_n(\varphi)$ and the distribution $g_n(\delta_i)$ of the $\delta_i[n]$. Assuming that at this iteration the RSS increased to some value $\alpha \triangleq y_{\text{best}}[n+1] > y_{\text{best}}[n]$, the perturbed phases must satisfy

$$\left|\frac{1}{N}\sum_{i=1}^{N} e^{j(\phi_i[n]+\delta_i[n])}\right| = \alpha. \tag{37}$$

From an application of the Gibbs conditioning principle, we conclude that, conditioned to all past RSS and (37), each perturbed phase $\phi_i[n] + \delta_i[n]$ is approximately distributed as

$$ce^{\eta\cos(\varphi-\beta)}(f_n * g_n)(\varphi), \quad \forall\, \varphi \in (-\pi, \pi] \tag{38}$$

where c, η, β are normalizing constants chosen so that (36) holds for the distribution in (38). Upon rotating the perturbed phases $\{\phi_i[n] + \delta_i[n]\}$ to obtain the new phases $\{\phi_i[n+1]\}$ corresponding to a total received signal with zero phase, the phase shift of β disappears from (38) and we conclude that each phase $\phi_i[n+1]$ is approximately distributed as

$$f_{n+1}(\varphi) \approx ce^{\eta\cos(\varphi)}(f_n * g_n)(\varphi), \quad \forall\, \varphi \in (-\pi, \pi]. \tag{39}$$

We shall see in Section IV, that it is beneficial to choose the perturbations distribution $g_n(\varphi)$ much more concentrated than the current distribution $f_n(\varphi)$. In this case, $f_n * g_n \approx f_n$ and (39) becomes

$$f_{n+1}(\varphi) \approx ce^{\eta\cos(\varphi)}f_n(\varphi), \quad \forall\, \varphi \in (-\pi, \pi].$$

Iterating this equation from $n = 0$, we conclude that

$$f_{n+1}(\varphi) \approx \bar{c}e^{\bar{\eta}\cos(\varphi)}f_0(\varphi), \quad \forall\, \varphi \in (-\pi, \pi],$$

where $\bar{c}$ and $\bar{\eta}$ are (redefined) normalization constants so that (36) holds, or equivalently, so that

$$\bar{c}\,2\pi I_0(\eta) = 1, \quad \bar{c}2\pi I_1(\eta) = \alpha \Rightarrow \frac{I_1(\eta)}{I_0(\eta)} = \alpha.$$

We thus arrive at the conclusion that, conditioned to the current and past normalized RSS $\{y_{\text{best}}[0], y_{\text{best}}[1], \ldots, y_{\text{best}}[n+1]\}$, the phases $\{\phi_i[n+1]\}$ are approximately independently distributed with marginal conditional distribution

$$f_{n+1}(\varphi) \approx \frac{1}{2\pi I_0(\eta)}e^{\eta\cos(\varphi)}, \quad \forall\, \varphi \in (-\pi, \pi],$$

with η such that $I_1(\eta)/I_0(\eta) = \alpha$. Since this distribution actually only depends on the current $y_{\text{best}}[n+1] = \alpha$, we conclude that only the conditioning to the current RSS matters.

As noted earlier, the phases $\phi_i[n]$ only approximately satisfy the conditions for the Gibbs conditioning principle because they are not completely independent. Nevertheless we conjecture that the i.i.d. perturbations $\{\delta_i[n]\}$ introduce sufficient additional independence in the perturbed phases $\{\phi_i[n] + \delta_i[n]\}$ that a version of the Gibbs principle still applies. As shown next, the Exp-Cosine distribution provides an extremely good fit for histograms obtained by simulations of the feedback algorithm.

Empirical Support for the Exp-Cosine Distribution: Conjecture 1 has been validated through an extensive set of Monte Carlo simulations, for several different values of the key parameters, which include the number of transmitters N and the distributions $g_n(\delta_i)$ for the perturbations $\delta_i[n]$. Fig. 5 shows typical results from a set of Monte Carlo simulations. We can see that the Monte Carlo results are perfectly consistent with Conjecture 1.

For comparison, Fig. 5 also shows predictions made assuming a Gaussian distribution. Clearly, the Exp-Cosine distribution provides a better fit for both the tails and the body of the distribution. This is despite the fact that we have biased the comparison in favor of the Gaussian distribution by adjusting it to match the Monte Carlo data by appropriate

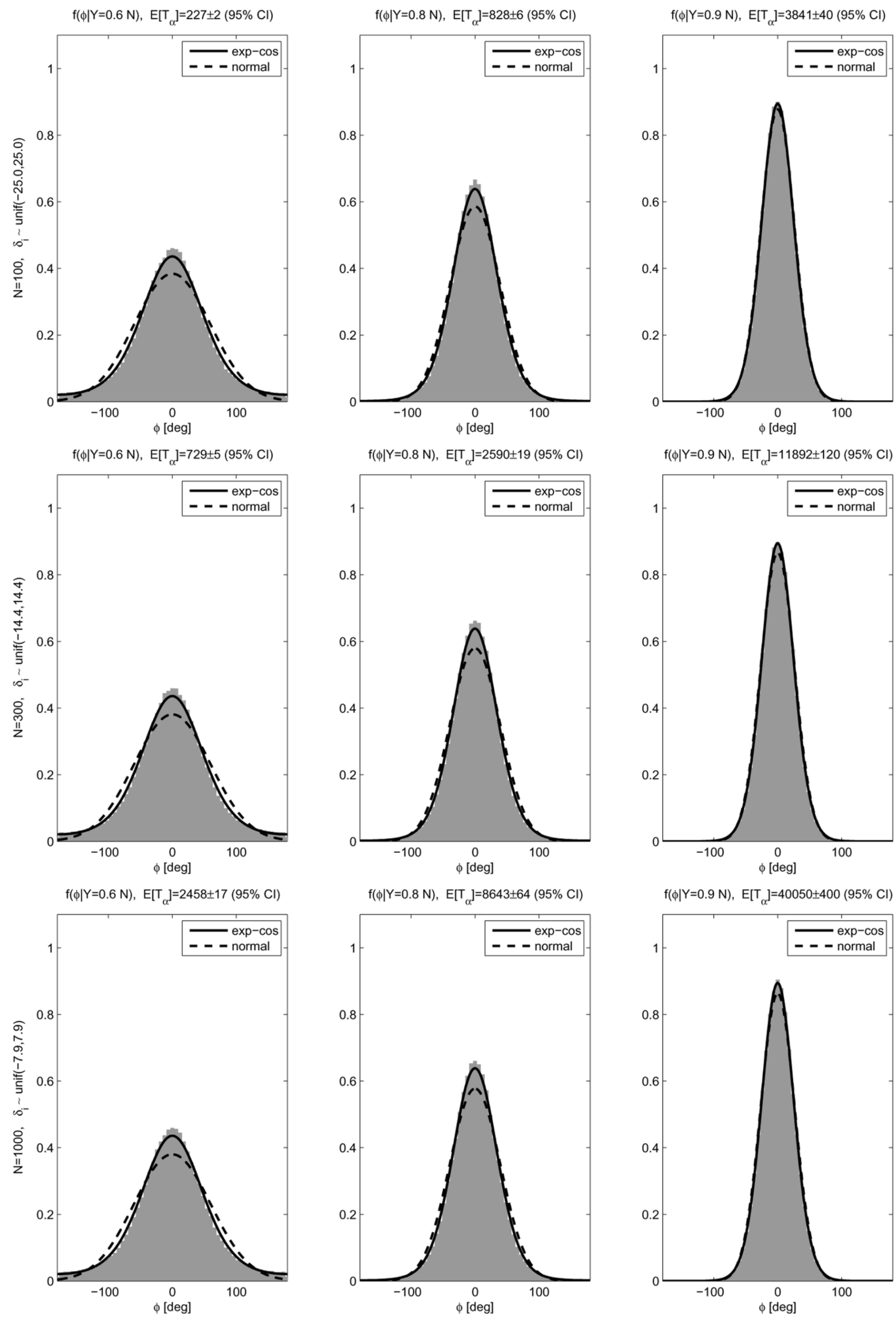

Fig. 5. Each histograms summarizes the results of a large number of Monte Carlo simulations of the distributed beamforming algorithm. The left, middle, and right columns show the distributions of the phases $\{\phi_i[n]\}$ conditioned to $\frac{1}{N}Y_{\text{best}}[n] = y$, for $y = 0.6, 0.8$, and 0.9, respectively. The top row, middle, and low rows correspond to simulations with $N = 100, 200$, and 300transmitters, respectively. Consistently with Conjecture 1, the three rows appear identical since, although obtained for different numbers of transmitters, they correspond to precisely the same values of y. Superimposed on each histogram, we see the Exp-Cosine distribution predicted by Conjecture 1 (solid line). For comparison, we also include a Normal distribution with the same mean and variance (dashed line). In all simulations shown, the $\delta_i[n]$ are uniformly distributed with the support of the distribution inversely proportional to $\sqrt{N}$, for consistency with the results in Section IV-A.

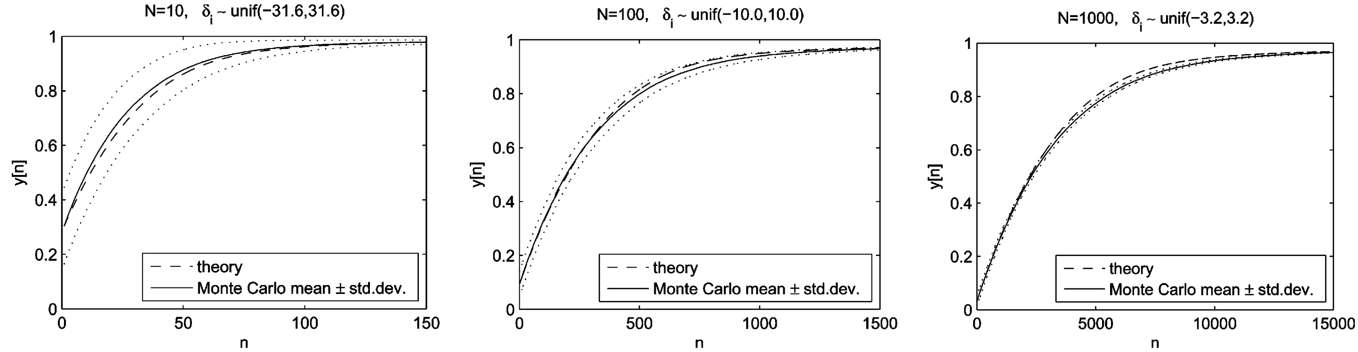

Fig. 6. Comparison between the theoretical prediction provided by (23) and Monte Carlo Simulation-based estimates of the evolution of receoved signal strength for $N \in \{10, 100, 1000\}$ transmitters. The dashed curve corresponds to (23), the solid curve is an estimate of $\frac{1}{N}E\left[Y_{\mathrm{best}}[n]\right]$ based on Monte Carlo simulations, and the dotted curves show one standard deviation of $\frac{1}{N}E\left[Y_{\mathrm{best}}[n]\right]$ around its mean, also based on Monte Carlo simulations. In all simulations shown, the $\delta_i[n]$ are uniformly distributed with the support of the distribution inversely proportional to $\sqrt{N}$, for consistency with the results in Section IV-A.

selection of its variance; in contrast, the Exp-Cosine distribution is computed using the formulas in Conjecture 1, without attempting to match the Monte Carlo data.

Fig. 6 compares the results of Monte Carlo simulations with predictions based on (17) over a wide range of values for N. We can see that even for fairly small N, (17) provides a very good match with Monte Carlo simulations. This figure also confirms that the standard deviation of $Y_{\mathrm{best}}[n]/N$ converges to zero as N increases.

IV. Performance Optimization and Scalability Analysis

In Theorem 2, we derived an analytical formula for the expected increase in RSS. We now use this result to determine the optimum distribution of the phase perturbations for the fastest rate of convergence at each timeslot. We also use the analytical model to establish the scalability of the feedback algorithm when the number of transmitters N becomes large. Finally we show that the RSS increments vary exponentially in time when the algorithm is close to convergence.

A. Performance Optimization

We consider the problem of choosing the distribution $g_n(\delta)$ for the perturbations $\delta_i[n]$ that minimizes the convergence time. Intuition suggests that it is best to choose larger perturbations initially to speed up the convergence and make the distribution narrower when the phase angles are closer to coherence; we now make this intuition precise.

We start with the observation that the expected RSS increment $h_n(y)$ depends on the distribution of $\delta_i[n]$ only through the two parameters $\chi_n \equiv E[\cos(\delta_i[n])]$, and $\rho_n \equiv \chi_n^2 - E[\cos(2\delta_i[n])]$. Therefore we can restrict ourselves to any family of distributions that allows us to freely choose these two parameters without losing optimality. Indeed, when $\delta_i[n] \ll 1$, we have

$$\begin{aligned}\chi_n &\equiv E\left[\cos(\delta_i[n])\right] \approx 1 - \frac{E\left[\delta_i^2[n]\right]}{2} \\ &= 1 - \frac{1}{2}\mathrm{Var}[\delta_i[n]] \qquad (40)\end{aligned}$$

$$\text{and } \rho_n \approx \chi_n^2 - 1 + 2E\left[\delta_i^2[n]\right] \approx \mathrm{Var}[\delta_i[n]]. \qquad (41)$$

Since we expect the phase perturbations to be relatively small, the parameters χ_n and ρ_n (and therefore the convergence rate) are largely determined by the variance of $g_n(\delta_i)$, and are largely independent of the precise form of the distribution.

Fig. 7(a) shows the convergence of $y[n]$ when $g_n(\delta_i) = \text{uniform}(-\Delta_0, \Delta_0)$ and Δ_0 is chosen numerically to maximize expected convergence rate $h_n(y)$ as given by (23) at every timeslot. The results confirm our intuition that at the initial stages of the algorithm, it is preferable to use larger perturbations (corresponding to large Δ_0), and when $Y_{\mathrm{best}}[n]$ gets closer to Y_{opt}, it is optimum to use narrower distributions (corresponding to smaller Δ_0). In general, we observe a near linear increase in RSS in the initial stage, with the convergence rate slowing with time.

We now derive a lower bound for the maximum achievable convergence rate $h_n(y)$. This derivation also yields an analytical estimate of the optimum variance of $\delta_i[n]$, which is accurate for large N. From the definition of χ_n and ρ_n, we have:

$$\begin{aligned}\rho_n &\equiv \chi_n^2 - E[\cos(2\delta_i[n])] = \chi_n^2 - E[2\cos^2(\delta_i[n]) - 1] \\ &\le 1 + \chi_n^2 - 2\left(E[\cos(\delta_i[n])]\right)^2 \equiv 1 - \chi_n^2 \qquad (42)\end{aligned}$$

where we used the Jensen's inequality in (42). We now use this inequality to rewrite (27) as:

$$\begin{aligned}\sigma_{\Re} &= \frac{\left(1 - \chi_n^2 - \rho_n\kappa(y)\right)}{2N\sigma_{\Re}} \\ &\ge \left(\frac{(1-\kappa(y))(1+\chi_n)}{2N}\right)\left(\frac{(1-\chi_n)}{\sigma_{\Re}}\right). \qquad (43)\end{aligned}$$

Using (43) in (23) gives

$$h_n(y) \ge \left(\frac{1+\chi_n}{2N}\right)\left(\frac{1-\kappa(y)}{y}\right)(x\, i(x)) \qquad (44)$$

where $x = \frac{y(1-\chi_n)}{\sigma_{\Re}}$, and $i(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}} - xQ(x)$ are defined as in (23). We can now lower bound $h_n(y)$ by maximizing the RHS of (44). We note that the RHS of (44) depends on χ_n through the term $x\, i(x)$ as well as through the term $(1 + \chi_n)$. However the latter dependence is by far weaker, therefore we focus on $x\, i(x)$. We can show by differentiation that this function achieves its maximum value of $h^* \approx 0.1012$ for $x_m \approx$

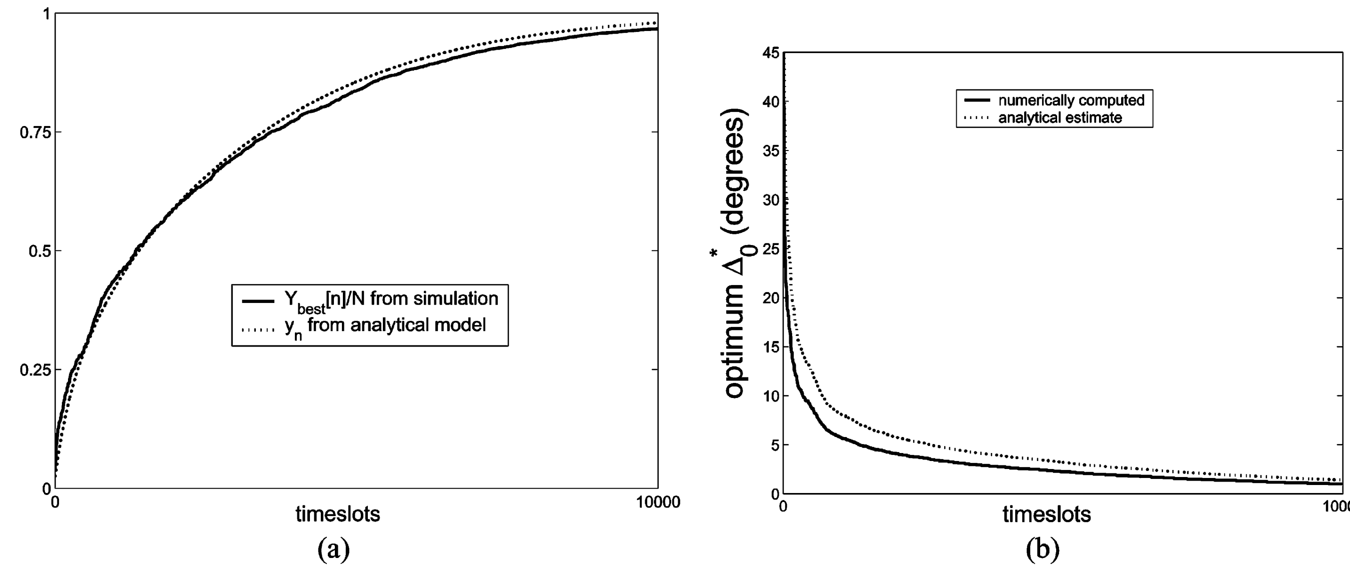

Fig. 7. Monte Carlo simulation of optimized beamforming algorithm with $N=1000$ transmitters, where the $\delta_i[n]$ are chosen from a uniform distribution whose support Δ_0 at each timeslot is chosen by numerically maximizing the RSS increment given by (23). (a) The solid curve shows $\frac{1}{N}Y_{\rm best}[n]$ from a Monte Carlo simulation using numerically optimized $\delta_i[n]$. The dotted curve shows y_n computed using (17) and (23) with $\delta_i[n]$ similarly optimized. (b) The solid curve shows the numerically optimized Δ_0. The dotted curve shows the estimate Δ_0^* of the optimum from (48).

0.6120 that satisfies $\frac{1}{\sqrt{2\pi}}e^{-\frac{x_m^2}{2}} = x_m^2 Q(x_m)$. This corresponds to a choice of χ_n given by

$$\chi_n^*(y) = \left(\frac{Ny^2 - x_m^2(1-\kappa(y))}{Ny^2 + x_m^2(1-\kappa(y))}\right). \tag{45}$$

The optimum convergence rate is at least as large as that achieved by (45), and therefore we have

$$h_n^*(y) \geq \frac{r_n(y)}{N},$$
$$\text{with } r_n(y) \triangleq \left(\frac{1+\chi_n^*(y)}{2}\right)\left(\frac{1-\kappa(y)}{y}\right)(h^*). \tag{46}$$

The RSS increments $h_n^*(y)$ varies with y roughly as $\left(\frac{1-\kappa(y)}{y}\right)$, which leads to the intuitively obvious conclusion that the convergence rate decreases as the algorithm approaches convergence. We note that the convergence rate bound in (46) is achievable, because it is possible to find a distribution that satisfies (45).

Indeed when N is large, we can use (45) and (40) to obtain an approximate formula for the optimum variance of δ_i as

$$\mathrm{Var}^*[\delta_i[n]] \approx \frac{4x_m^2}{Ny^2}(1-\kappa(y)). \tag{47}$$

One distribution that achieves this variance is the uniform distribution $g_n(\delta_i) = \frac{1}{\Delta_0^*},\ \delta_i \in (-\Delta_0^*, \Delta_0^*]$ with

$$\Delta_0^* = \left(\frac{\sqrt{12}x_m}{\sqrt{N}}\right)\left(\frac{\sqrt{(1-\kappa(y))}}{y}\right). \tag{48}$$

Remark: In deriving (47), we used the approximation in (40). In addition we made two suboptimum choices in deriving (45): one by the use of the Jensen lower bound in (42), and the other by ignoring the dependence of $h_n(y)$ on χ_n through the $(1+\chi_n)$ term in (44). However as (47) shows, the optimum $\mathrm{Var}^*[\delta_i[n]]$ approaches zero as N becomes large. In this regime (of large N), the effect of these suboptimal choices is quite small, and we find that (47) provides an excellent estimate of the best $g_n(\delta_i)$.

The numerically computed optimum Δ_0^* is plotted along with Δ_0^* from (48) in Fig. 7(b). As seen from the figure, (48) provides a good approximation for the optimal distribution, but generally overestimates the optimal Δ_0^*.

B. Scalability and Rate of Convergence

Let $T_N(y^*)$ denote the number of timeslots required for y_n to reach a given level of convergence y^* i.e., $T_N(y^*) = \arg\min_n y_n \geq y^*$. Roughly speaking, for $y^*=0.5$, $T_N(y^*)$ is the number of timeslots required for the expected RSS $E[Y_{\rm best}[n]]$ to reach 50% of $Y_{\rm opt}$.

Theorem 3: Under an optimum choice of the distribution $g_n(\delta_i)$ of the perturbations $\delta_i[n]$, the time $T_N^*(y^*)$ needed for y_n to reach y^* satisfies

$$T_N^*(y^*) < NT^*, \tag{49}$$

for any given N, y^*, where T^* is a constant that depends on y^* but is independent of N. □

Proof of Theorem 3: For this proof it is convenient to consider continuous-time function $y_c(t),\ t\in[1,\infty)$ that linearly interpolates the discrete-time function $y_n,\ n\in\{1,2,\ldots\}$:

$$y_c(t) = y_n + (y_{n+1}-y_n)(t-n) \equiv y_n + h_n(y_n)(t-n),$$
$$\forall\, t\in[n,n+1),\quad \forall\, n\in\{1,2,\ldots\}.$$

where we used (17).

Since the discrete-time function y_n is monotone strictly increasing, the continuous-time function $y_c(t)$ is also strictly increasing and therefore it has an inverse function $T_c(y)$. To find

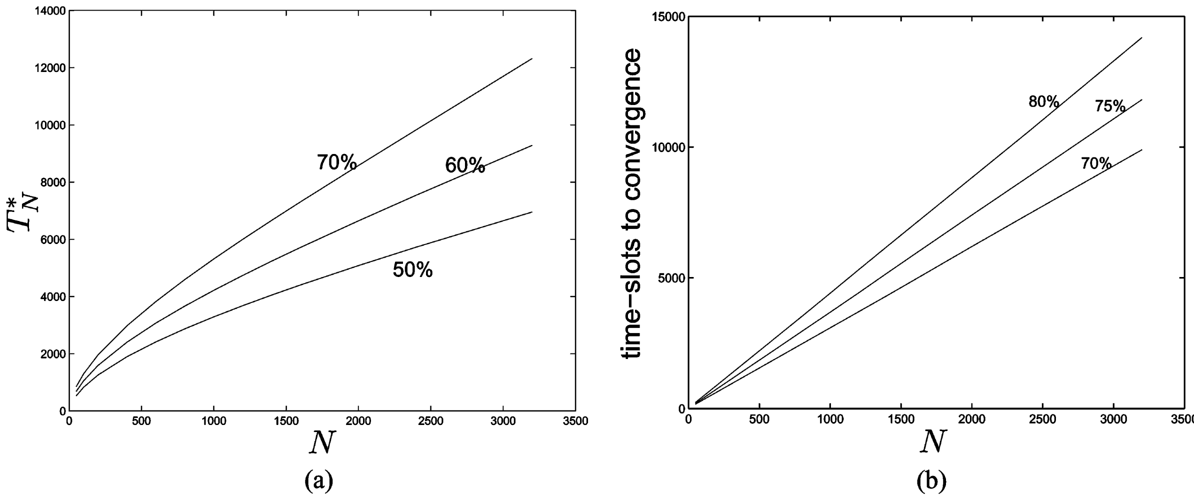

Fig. 8. Time to convergence for unoptimized and optimized phase distributions. (a) Time to convergence for $y^* = 0.5, 0.6$, and 0.7 when the phase perturbations are chosen from a fixed distribution, $\delta_i[n] \sim \text{uniform}(-2°, 2°)$. (b) Time to convergence for $y^* = 0.7, 0.75$, and 0.8 when the phase perturbations are chosen to numerically optimize the RSS increments in (23).

the time instant at which y_c reaches y^*, it suffices to compute $T_c(y^*)$, which can be done using

$$T_c(y^*) = \int_{y_0}^{y^*} \frac{\mathrm{d}T_c}{\mathrm{d}y}(y)dy. \tag{50}$$

Since $y_c(t)$ is differentiable almost everywhere, $T_c(y)$ is also differentiable almost everywhere. Further

$$\frac{\mathrm{d}y_c(t)}{\mathrm{d}t} \equiv h_n\left(y_c(\lfloor t \rfloor)\right) \geq \frac{r_n\left(y_c(\lfloor t \rfloor)\right)}{N} \geq \frac{r_n\left(y_c(t)\right)}{N} \tag{51}$$

where we used (46) and the fact that the function $r_n(y)$ is monotone decreasing. We also note from (46) that $r_n(y)$ does not depend on N. Using (51) and the inverse function theorem we have

$$\frac{\mathrm{d}T_c(y)}{\mathrm{d}y} = \left(\frac{\mathrm{d}y_c(t)}{\mathrm{d}t}\right)^{-1}_{y_c(t)=y} \leq \frac{N}{r_n(y)} \tag{52}$$

Using (52) in (50), we conclude that

$$\begin{aligned} T_N^*(y^*) &\equiv \lceil T_c(y^*) \rceil < 1 + T_c(y^*) \\ &\leq 1 + N\int_{y_0}^{y^*} \frac{1}{r_n(y)}dy \leq 1 + N\int_0^{y^*} \frac{1}{r_n(y)}dy \end{aligned}$$

from which (49) follows with $T^* \triangleq 1 + \int_0^{y^*} \frac{1}{r_n(y)}dy$. ■

Fig. 8 shows the convergence time of the algorithm with unoptimized and optimized phase distributions. The linear variation with N provides numerical confirmation of Theorem 3.

From (46), we observed earlier that the convergence becomes slower in time roughly as $\frac{1-\kappa(y)}{y}$. We now show that the rate of RSS increase is exponential in time, when the algorithm is near convergence.

Theorem 4: Suppose that the variance for the perturbations $\delta_i[n]$ satisfies (47). Then, for any choice of distribution $g_n(\delta_i)$ that satisfies (46), we have local exponential convergence of $y_n \to 1$, with a time constant equal to $\frac{N}{(4h^*)} \approx 2.5N$. □

One possible choice for $g_n(\delta_i)$ is the uniform distribution as in (48). Fig. 9 confirms that this choice does result in exponential convergence over a wide range of values for N.

Proof of Theorem 4: To prove local convergence, we need to analyze the dynamics of $r_n(y)$ around $y = 1$. Near the point $y = 1$, $\kappa(y) \approx e^{-4(1-y)}$ (cf., Fig. 4(b)), from which we can, from the comparison Lemma [24], conclude that near this point we have

$$r_n(y) = \left(\frac{1+\chi_n^*(y)}{2}\right)\left(\frac{1-\kappa(y)}{y}\right)(h^*) \approx (4h^*)\ (1-y),$$

which in view of (17) leads to

$$1 - y_{n+1} \lessapprox \left(1 - \frac{(4h^*)}{N}\right)(1 - y_n).$$

Close to the equilibrium point $y = 1$, the linear term in $(1 - y_n)$ dominates and we conclude that

$$y_n \gtrapprox 1 - e^{\frac{-(4h^*)(n-n_0)}{N}}\left(1 - y_{n_0}\right).$$

■

V. Conclusion

The results of this paper indicate that distributed transmit beamforming can be effectively realized by utilizing only one bit of feedback per iteration from the receiver. The technique is scalable, in that convergence time grows only linearly with the number of participating nodes. The basic algorithm presented here can be easily adapted for implementation in practical settings [8], and can be extended to achieve frequency as well as phase synchronization [9].

An open technical problem is a rigorous characterization of the conditional distribution of the transmitter phases, conditioned on the RSS. In particular, justification or refinement of the Exp-Cosine conjecture appears to require a deep, problem-specific large deviations analysis.

Realizing the potential gains from distributed beamforming requires the design of network protocols that support and exploit

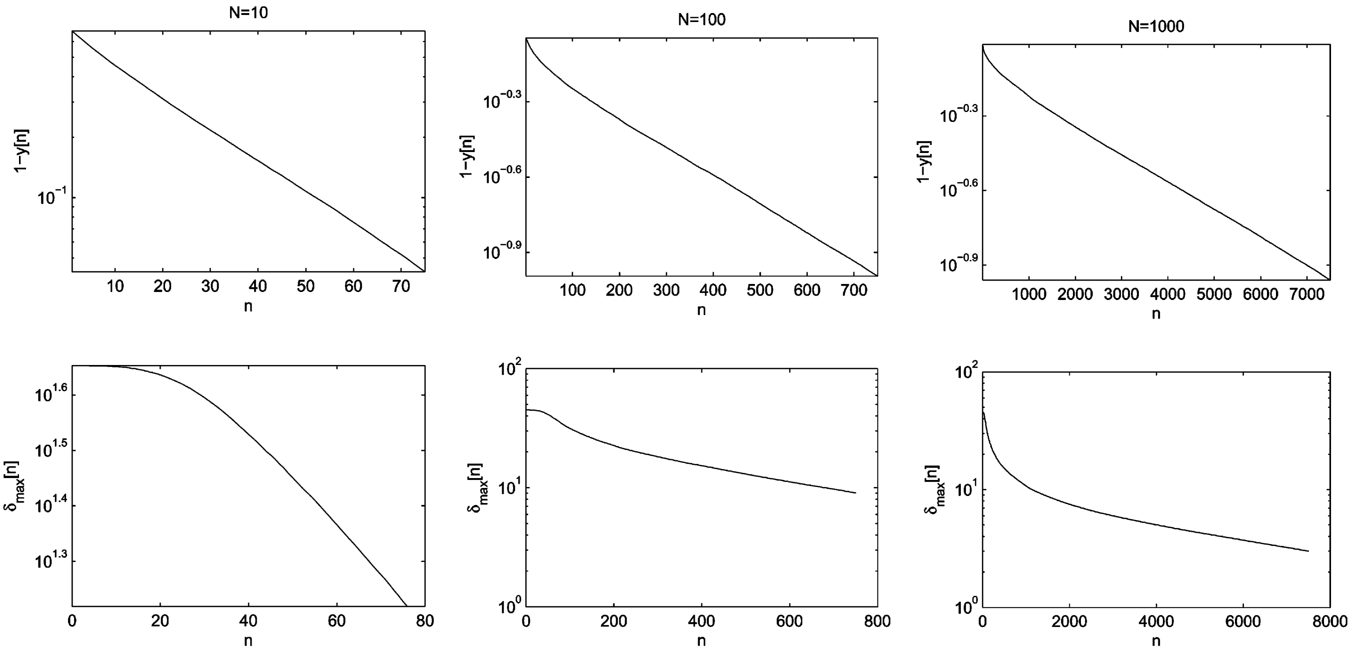

Fig. 9. Monte Carlo Simulations for $N \in \{10, 100, 1000\}$ transmitters with the optimal choice of distribution for the perturbations $\delta_i[n]$ in (48), while keeping the δ_i always below 45°. The solid curves in the top row of plots show an estimate of $1 - y_n = 1 - \frac{1}{N} E\left[Y_{\text{best}}[n]\right]$ based on Monte Carlo simulations. The straight lines confirm an exponential convergence. The bottom row of plots show the evolution of the maximum value of the uniform distribution for the $\delta_i[n]$.

it [6]. A detailed study is also required on how best to achieve and maintain the frequency synchronization across transmitters, which was assumed in our algorithm. Exploration of the effects of time variations is important for understanding the applicability of these ideas to mobile ad hoc networks. Preliminary results in [8] indicate that the analysis here can be extended to understand the tradeoffs between tracking and convergence in time-varying settings.

Appendix A
Proof of Proposition 1

Proof of Proposition 1: Let $\bar{\varphi}$ be an arbitrary N-vector of phases normalized so that the total received signal $\sum_i a_i e^{j\varphi_i}$ has zero phase. For simplicity, we assume that the elements φ_i of the N-vector $\bar{\varphi}$ are sorted so that

$$|\varphi_1| \geq |\varphi_2| \geq \cdots \geq |\varphi_N|.$$

Assuming that $\text{RSS}(\bar{\varphi}) \leq Y_{\text{opt}} - \epsilon$, we conclude that

$$\cos(\varphi_1) \sum_i a_i \leq \sum_i a_i \cos(\varphi_i) \leq Y_{\text{opt}} - \epsilon \Rightarrow |\varphi_1| \geq \phi_\epsilon \triangleq \arccos\left(\frac{Y_{\text{opt}} - \epsilon}{\sum_i a_i}\right). \quad (53)$$

Now we choose a phase perturbation δ_1 that decreases $|\varphi_1|$. This makes the most misaligned phase in $\bar{\varphi}$ closer to the received signal phase, and thus increases the magnitude of the received signal. Without loss of generality we assume $\varphi_1 > 0$, then we need to choose a $\delta_1 < 0$. Consider $\delta_1 \in \left(-\Delta_0, -\frac{\Delta_0}{2}\right)$. We have

$$\rho_1 \triangleq \Pr\left(-\Delta_0 \leq \delta_1 < -\frac{\Delta_0}{2}\right) = \int_{-\Delta_0}^{-\frac{\Delta_0}{2}} g_n(\delta) d\delta > 0. \quad (54)$$

With δ_1 chosen as above, we get

$$a_1 \cos(\varphi_1 + \delta_1) - a_1 \cos(\varphi_1) > 2\epsilon_1 \quad \text{where } \epsilon_1 \triangleq \frac{a_1 \Delta_0}{4} \sin\left(\varphi_\epsilon - \frac{\Delta_0}{2}\right). \quad (55)$$

We observe that ϵ_1 and ρ_1 depend only on ϵ and not on $\bar{\varphi}$.

Equation (55) shows that the perturbation δ_1 by itself will achieve a nonzero increase in total received signal, provided that the other phases φ_i do not get too misaligned by their respective δ_i

$$\begin{aligned}
&\text{RSS}(\bar{\varphi} + \bar{\delta}) - \text{RSS}(\bar{\varphi}) \\
&\quad \geq \sum_i a_i \left(\cos(\varphi_i + \delta_i) - \cos(\varphi_i)\right) \\
&\quad = a_1 \left(\cos(\varphi_1 + \delta_1) - \cos(\varphi_1)\right) \\
&\qquad + \sum_{i>1} a_i \left(\cos(\varphi_i + \delta_i) - \cos(\varphi_i)\right) \\
&\quad > 2\epsilon_1 + \sum_{i>1} a_i \left(\cos(\varphi_i + \delta_i) - \cos(\varphi_i)\right). \quad (56)
\end{aligned}$$

We note that since $\text{RSS}(\bar{\varphi})$ is continuous in each of the phases φ_i, we can always find a $\epsilon_i > 0$ to satisfy:

$$|a_i \left(\cos(\varphi_i + \delta_i) - \cos(\varphi_i)\right)| < \frac{\epsilon_1}{N-1}, \quad \forall\, |\delta_i| < \epsilon_i. \quad (57)$$

In particular the choice $\epsilon_i \triangleq \frac{\epsilon_1}{a_i(N-1)}$, satisfies (57), and this choice of ϵ_i is independent of $\bar{\varphi}$. With the δ_i's chosen to satisfy $|\delta_i| < \epsilon_i$, we have

$$-\epsilon_1 < \sum_{i>1} a_i \left(\cos(\varphi_i + \delta_i) - \cos(\varphi_i)\right) < \epsilon_1. \quad (58)$$

$$\begin{aligned}\mathrm{E}_{y,n}\left[x[n]^2\right] &= \sum_{i,\ell} a_i a_\ell \mathrm{E}_{y,n}\left[e^{j(\phi_i[n]+\phi_\ell[n])}\left(e^{j\delta_i[n]}-\chi_n\right)\left(e^{j\delta_\ell[n]}-\chi_n\right)\right]\\ &= \sum_i a_i^2 \mathrm{E}_{y,n}\left[e^{2j\phi_i[n]}(e^{j\delta_i[n]}-\chi_n)^2\right] = -\rho_n \sum_i a_i^2 \mathrm{E}_{y,n}[e^{2j\phi_i[n]}] \end{aligned} \tag{63}$$

Since $g_n(\delta_i)$ is bounded away from zero in each of the nonzero intervals $(-\epsilon_i, \epsilon_i)$, the probability ρ_i of choosing δ_i to satisfy (57) is nonzero, i.e., $\rho_i > 0$, which is independent of $\bar{\varphi}$. Finally, we recall that each of the δ_i are chosen independently, and therefore with probability $\rho = \prod_i \rho_i > 0$, it is possible to find δ_1 to satisfy (55) and $\delta_i, i > 1$ to satisfy (57). For $\bar{\delta}$ chosen as above, $\mathrm{RSS}(\bar{\varphi}+\bar{\delta}) - \mathrm{RSS}(\bar{\varphi}) > \epsilon_1$, and therefore Proposition 1 follows. ∎

APPENDIX B
PROOF OF (35)

The necessary optimality condition for an extremum of the functional

$$J[q] \triangleq \int q(x)\log\frac{q(x)}{p(x)}dx - \frac{\eta_1}{2}\left(\left(\int q(x)\Re[f(x)]dx\right)^2 + \left(\int q(x)\Im[f(x)]dx\right)^2 - y^2\right) - \eta_2\left(\int q(x)dx - 1\right)$$

in (34) is given by

$$1+\log\frac{q(x)}{p(x)} - 2\eta_1\left(\Re[f(x)]\int q(\bar{x})\Re[f(\bar{x})]d\bar{x} + \Im[f(x)]\int q(\bar{x})\Im[f(\bar{x})]d\bar{x}\right) - \eta_2 = 0,$$

which is equivalent to

$$q(x) = ce^{\eta_1(\Re[f(x)]\Re[\bar{f}]+\Im[f(x)]\Im[\bar{f}])}p(x) \tag{59}$$

with $c \triangleq e^{\eta_2 - 1}$ and $\bar{f} \triangleq \int q(\bar{x})f(\bar{x})d\bar{x}$. Since, one of the constraints imposes $|\bar{f}| = y$, we can express $\bar{f} = ye^{j\beta}$ and re-write (59) as

$$q(x) = ce^{\eta(\Re[f(x)e^{-j\beta}])}p(x).$$

where $\eta = \eta_1 y$, c, and β are parameters to be determined from the conditions $\int q(x)dx = 1$ and $\bar{f} = ye^{j\beta}$.

APPENDIX C
PROOF OF (12)–(13)

Defining $x[n] \triangleq x_\Re[n] + jx_\Im[n]$, we have that

$$x[n] \equiv \frac{1}{N}\sum_i a_i e^{j\phi_i[n]}(e^{j\delta_i[n]} - \chi_n).$$

Since the $\delta_i[n]$ are chosen from a symmetric distribution $g_n(\delta)$, it follows that $e^{j\delta_i[n]} - \chi_n$ has zero mean. Using this and the fact that the $\delta_i[n]$ are (conditionally) independent of the $\phi_i[n]$, we conclude that $x[n]$ has zero mean and therefore, both $x_\Re[n]$ and $x_\Im[n]$ also have zero mean.

The variances of the real and complex parts of $x[n]$ are given by

$$\begin{aligned}\mathrm{Var}[x_\Re[n]] &\triangleq \frac{\mathrm{E}_{y,n}[(x[n]+x[n]^*)^2]}{4N^2}\\ &= \frac{\mathrm{E}_{y,n}\left[x[n]x[n]^*\right] + \Re E_{y,n}\left[x[n]^2\right]}{2N^2}\end{aligned} \tag{60}$$

$$\begin{aligned}\mathrm{Var}[x_\Im[n]] &\triangleq -\frac{\mathrm{E}_{y,n}[(x[n]-x[n]^*)^2]}{4N^2}\\ &= \frac{\mathrm{E}_{y,n}\left[x[n]x[n]^*\right] - \Re \mathrm{E}_{y,n}\left[x[n]^2\right]}{2N^2}.\end{aligned} \tag{61}$$

To compute these variance, we expand

$$\mathrm{E}_{y,n}\left[x[n]x^*[n]\right] = \sum_{i,\ell} a_i a_\ell \mathrm{E}_{y,n}\left[e^{j(\phi_i[n]-\phi_\ell[n])}(e^{j\delta_i[n]}-\chi_n)(e^{-j\delta_\ell[n]}-\chi_n)\right].$$

For $i \neq \ell$ the two terms are $(e^{j\delta_i[n]} - \chi_n)$ and $(e^{-j\delta_\ell[n]} - \chi_n)$ are zero-mean and independent and therefore any term with $i \neq \ell$ disappears from the summation, which leads to

$$\begin{aligned}\mathrm{E}_{y,n}\left[x[n]x^*[n]\right] &= \sum_i a_i^2 \mathrm{E}_{y,n}\left[|e^{j\delta_i[n]} - \chi_n|^2\right]\\ &= \left(1-\chi_n^2\right)\sum_i a_i^2\end{aligned} \tag{62}$$

since $E\left[|e^{j\delta_i[n]} - \chi_n|^2\right] = 1 - \chi_n^2$. Similarly, see (63) at the top of the page, since $\rho_n = -\mathrm{E}\left[(e^{j\delta_i[n]} - \chi_n)^2\right] = \mathrm{E}[\cos(\delta_i[n])]^2 - \mathrm{E}[\cos(2\delta_i[n])]$. Replacing (62) and (63) in (60)–(61), we obtain (12)–(13).

REFERENCES

[1] T. Cover and A. Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Inf. Theory*, vol. 25, no. 5, pp. 572–584, 1979.

[2] M. Gastpar and M. Vetterli, "On the capacity of large Gaussian relay networks," *IEEE Trans. Inf. Theory*, vol. 51, no. 3, pp. 765–779, 2005.

[3] A. Ozgur, O. Leveque, and D. Tse, "How does the information capacity of ad hoc networks scale?," in *Proc. 44th Annu. Allerton Conf. Commun., Contr. Signal Processing*, Sep. 2006.

[4] G. Barriac, R. Mudumbai, and U. Madhow, "Distributed beamforming for information transfer in sensor networks," in *IPSN'04: Proc. Third Int. Symp. Inf. Processing in Sensor Netw.*, 2004, pp. 81–88.

[5] R. Mudumbai, G. Barriac, and U. Madhow, "On the feasibility of distributed beamforming in wireless networks," *IEEE Trans. Wireless Commun.*, vol. 6, no. 5, pp. 1754–1763, May 2007.

[6] R. Mudumbai, D. R. Brown, U. Madhow, and H. V. Poor, "Distributed transmit beamforming: Challenges and recent progress," *IEEE Commun. Mag.*, vol. 47, no. 2, pp. 102–110, Feb. 2009.

[7] R. Mudumbai, J. Hespanha, U. Madhow, and G. Barriac, "Scalable feedback control for distributed beamforming in sensor networks," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT'05)*, Sep. 2005.

[8] R. Mudumbai, B. Wild, U. Madhow, and K. Ramchandran, "Distributed beamforming using 1 bit feedback: From concept to realization," in *Proc. 44th Allerton Conf. Commun., Contr. Signal Processing*, Sep. 2006.
[9] M. Seo, M. Rodwell, and U. Madhow, "A feedback-based distributed phased array technique and its application to 60-GHz wireless sensor network," in *Proc. IEEE Int. Microw. Symp. (IMS)*, Jun. 2008.
[10] D. Love, J. R. W. Heath, W. Santipach, and M. Honig, "What is the value of limited feedback for MIMO channels?," *IEEE Commun. Mag.*, vol. 42, no. 10, pp. 54–59, 2004.
[11] B. Hassibi and A. Dana, "On the power efficiency of sensory and ad-hoc wireless networks," *IEEE Trans. Inf. Theory*.
[12] I. D. R. Brown, G. Prince, and J. McNeill, "A method for carrier frequency and phase synchronization of two autonomous cooperative transmitters," in *Proc. IEEE 6th Workshop Signal Processing Adv. Wireless Commun.*, 2005, pp. 260–264.
[13] H. Ochiai, P. Mitran, H. Poor, and V. Tarokh, "Collaborative beamforming for distributed wireless ad hoc sensor networks," *IEEE Trans. Signal Process. (formerly, IEEE Trans. Acoust., Speech, Signal Processing)*, vol. 53, pp. 4110–4124, 2005.
[14] H. Robbins and S. Monro, "A stochastic approximation method," *Ann. Math. Stat.*, vol. 22, pp. 400–407, 1954.
[15] H. J. Kushner and G. G. Yin, *Stochastic Approximation Algorithms and Applications*. New York: Springer-Verlag, 1997.
[16] J. A. Bucklew and W. A. Sethares, "Convergence of a class of decentralized beamforming algorithms," in *Proc. IEEE/SP 14th Workshop Stat. Signal Processing*, 2007, pp. 123–125.
[17] B. Banister and J. Zeidler, "Feedback assisted stochastic gradient adaptation of multiantenna transmission," *IEEE Trans. Wireless Commun.*, vol. 4, no. 3, pp. 1121–1135, 2005.
[18] J. Thukral and H. Bölcskei, Distributed Spatial Multiplexing With 1-bit Feedback 2007 [Online]. Available: http://arxiv.org/abs/0710.1522
[19] P. Fertl, A. Hottinen, and G. Matz, Perturbation-Based Distributed Beamforming for Wireless Relay Networks 2008 [Online]. Available: http://arxiv.org/abs/0809.5182
[20] C. Lin, V. V. Veeravalli, and S. P. Meyn, Distributed Beamforming With Feedback: Convergence Analysis 2008 [Online]. Available: http://arxiv.org/abs/0806.3023
[21] M. Johnson, M. Mitzenmacher, and K. Ramchandran, "Distributed beamforming with binary signaling," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2008, pp. 890–894.
[22] A. N. Shiriaev, *Probability*, ser. Graduate Texts in Mathematics, 2nd ed. New York: Springer Verlag, 1996, vol. 95.
[23] A. Dembo and O. Zeitouni, "Refinements of the Gibbs conditioning principle," *Prob. Theory Rel. Fields 104*, pp. 1–14, 1996.
[24] H. K. Khalil, *Nonlinear Systems*, 3rd ed. Upper Saddle River, NJ: Prentice-Hall, 2002.

Raghuraman Mudumbai (S'04–M'09) received the B.Tech. degree in electrical engineering from the Indian Institute of Technology, Madras, India, in 1998 and the M.S. degree in electrical engineering from Polytechnic University, Brooklyn, NY, in 2000. He worked as a System Engineer at Ericsson Datacom, Inc., between 2001 and 2003, and returned to graduate school at the University of California, Santa Barbara (UCSB) in 2003, where he received the Ph.D. degree in 2007.

He worked as a Postdoctoral Researcher at UCSB until June 2009. Since August 2009, he has been with the Department of Electrical and Computer Engineering at the University of Iowa, Iowa City, as Assistant Professor.

Joao Hespanha (S'95–A'98–SM'02–F'08) was born in Coimbra, Portugal, in 1968. He received the Licenciatura in electrical and computer engineering from the Instituto Superior Técnico, Lisbon, Portugal in 1991 and the Ph.D. degree in electrical engineering and applied science from Yale University, New Haven, CT, in 1998.

From 1999 to 2001, he was Assistant Professor at the University of Southern California, Los Angeles. He moved to the University of California, Santa Barbara in 2002, where he currently holds a Professor position with the Department of Electrical and Computer Engineering.

Prof. Hespanha is Associate Director for the Center for Control, Dynamical-systems, and Computation (CCDC), Vice-Chair of the Department of Electrical and Computer Engineering, and a member of the Executive Committee for the Institute for Collaborative Biotechnologies (ICB). From 2004–2007, he was an Associate Editor for the IEEE TRANSACTIONS ON AUTOMATIC CONTROL.

Upamanyu Madhow (S'96–M'90–SM'96–F'05) received the Bachelor's degree in electrical engineering from the Indian Institute of Technology, Kanpur, in 1985. He received the M.S. and Ph.D. degrees in electrical engineering from the University of Illinois, Urbana-Champaign, in 1987 and 1990, respectively.

From 1990 to 1991, he was a Visiting Assistant Professor at the University of Illinois. From 1991 to 1994, he was a research scientist at Bell Communications Research, Morristown, NJ. From 1994 to 1999, he was on the faculty of the Department of Electrical and Computer Engineering at the University of Illinois, Urbana-Champaign. Since December 1999, he has been with the Department of Electrical and Computer Engineering at the University of California, Santa Barbara, where he is currently a Professor. His research interests are in communication systems and networking, with current emphasis on wireless communication, sensor networks and multimedia security.

Dr. Madhow is a recipient of the NSF CAREER award. He has served as Associate Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS, IEEE TRANSACTIONS ON INFORMATION THEORY, and IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY. He is the author of the graduate textbook, *Fundamentals of Digital Communication* (Cambridge University Press).

Gwen Barriac (S'03–M'04) received the B.S. degree in electrical engineering from Princeton University, Princeton, NJ, in 1999, and the M.S. and Ph.D. degrees in electrical engineering from the University of California at Santa Barbara in 2004.

She is currently with Corporate R&D group, Qualcomm, San Diego, CA.