

# Distortion Outage Minimization in Rayleigh Fading Using Limited Feedback

Chih-Hong Wang *Member, IEEE*, and Subhrakanti Dey *Senior Member, IEEE*

**Abstract**—In this paper we investigate the problem of distortion outage minimization in a clustered sensor network where sensors within each cluster send their noisy measurements of a random Gaussian source to their respective clusterheads (CH) using analog forwarding and a non-orthogonal multi-access scheme under the assumption of perfect distributed beamforming. The CHs then amplify and forward their measurements to a remote fusion center over orthogonal Rayleigh distributed block-fading channels. Due to fading, the distortion between the true value of the random source and its reconstructed estimate at the fusion center becomes a random process. Motivated by delay-limited applications, we seek to minimize the probability that the distortion exceeds a certain threshold (called the “distortion outage” probability) by optimally allocating transmit powers to the CHs. In general, the outage minimizing optimal power allocation for the CH transmitters requires full instantaneous channel state information (CSI) at the transmitters, which is difficult to obtain in practice. The novelty of this paper lies in designing locally optimal and sub-optimal power allocation algorithms which are simple to implement, using *limited channel feedback* where the fusion center broadcasts only a few bits of feedback to the CHs. Numerical results illustrate that a few bits of feedback provide significant improvement over no CSI and only 6-8 bits of feedback result in outages that are reasonably close to the full CSI performance for a 6-cluster sensor network. We also present results using a simultaneous perturbation stochastic approximation (SPSA) based optimization algorithm that provides further improvements in outage performance but at the cost of a much greater computational complexity.

## I. INTRODUCTION

WIRELESS sensor networks have many useful applications such as in environmental and wildlife habitat monitoring, in tracking targets for defense applications, in aged healthcare and many other areas of human life. Wireless sensor networks usually involve large numbers of sensor nodes that are distributed geographically to monitor certain physical phenomena and collect measurements which are then sent to a central processing (often called a Fusion Center (FC)) unit via wireless channels. The FC computes estimates of the samples of the physical phenomenon from the noisy measurements collected by the sensors. Energy consumption is a unique and important issue in wireless sensor networks performing such distributed estimation tasks because sensors are expected to be deployed once only and their batteries are often irreplaceable due to high cost. Due to random fading in wireless transmission, the quality of the estimate at the FC, measured by a distortion measure becomes a random variable. In delay-limited settings, instead of minimizing a long term average distortion (or expected distortion for ergodic fading channels), it is more appropriate to minimize the probability that the distortion for each estimate exceeds a certain threshold, the so-called distortion outage probability. This is similar to the idea of minimizing the information outage probability in block-fading wireless communication channels in the information theoretic context [1]. Optimal power allocation at the sensor transmitters for such outage minimization under various power constraints is an important problem

The authors are with the ARC Special Research Centre for Ultra Broadband Information Networks (CUBIN) within the Department of Electrical & Electronic Engineering at the University of Melbourne, Parkville, Victoria 3010 (e-mail: [chwang,sdey]@ee.unimelb.edu.au). CUBIN is an affiliated program of National ICT Australia (NICTA).

from the point of view of reducing energy consumption in sensor networks. In this paper, we look at an optimal power allocation problem for distortion outage minimization in a clustered sensor network using finite rate channel feedback (or quantized channel state information (CSI)) broadcast by the FC to the transmitters at the clusterheads for the various clusters.

Energy efficiency in wireless sensor networks has been studied in the context of power optimization problems and estimation diversity [2]. In [1], optimal power allocation schemes were obtained for minimizing the information outage probability over block-fading channels subject to peak and/or long-term average power constraints. These techniques were used to obtain optimal power allocation schemes for distortion outage minimization in [3]. These power allocation schemes assume perfect CSI at both transmitter and receiver. In practice, perfect CSI at the transmitter (CSIT) relies on instantaneous channel feedback from the FC, which is difficult to implement due to the limited bandwidth, delay and error in the feedback channel. Motivated by these constraints, many work in the literature have looked at power control in the field of MIMO beamforming systems with partial CSIT using limited feedback [4], [5]. The optimal power allocation scheme for systems employing limited feedback is in general complex and hence difficult to obtain. [6] studied average reliable throughput minimization over slow fading channels. They found properties of optimal power allocation policy that aid in the design of power allocation algorithms. For outage minimization, [7], [8], [9] showed that the optimal index mapping for the power codebook has a circular structure, where the same index is assigned to both the best and the worst channel realizations. A suboptimal power allocation scheme is proposed in [10] for a single user system with multiple transmit antennas and single receive antenna with finite rate feedback power control. Practical methods with low complexity for information outage minimization are also proposed in [4] for fading relay channels. These suboptimal power allocation schemes, although not optimal, can provide significant gains over no-CSIT even for small number of feedback bits. In addition, [11] studied outage behavior of slow fading channels with power control where the feedback channel is corrupted by noise.

In this paper we study a wireless sensor network where the sensors are grouped into clusters. Each cluster has an elected clusterhead (CH). The sensors observe a Gaussian random point source and send their observed (noisy) information to the CH by uncoded analog transmission using distributed beamforming. The clusterheads then transmit the combined received signal to the FC using amplify-and-forward and orthogonal FDMA where the channel is subject to random fading. It is assumed that the channel from each clusterhead to the FC is subject to independent and ergodic block fading where each fading block is long enough for all transmissions within the clusters and between the CHs and the FC to be completed and an estimate of the random Gaussian source to be computed at the FC. The optimal power allocation that minimizes the distortion outage subject to a long-term average sum (across the CHs) power constraint for Rayleigh fading channels has been obtained in [3] where it is assumed that

full instantaneous CSIT is available via feedback from the FC. However full instantaneous CSIT is very expensive to implement in practice, as this requires an error-free, delayless, infinite-bandwidth feedback channel and extra communication overheads. To address this issue, we modify the assumption of the feedback channel to be error-free, delayless but rate-limited. In this setting, it is assumed that an optimal power codebook (to be designed) is pre-computed at the FC based on the fading channel statistics and the average power constraint, and stored at the FC as well as the CH transmitters. In real time, under the assumption of perfect channel state information at the FC, an index (corresponding to a region in a multi-dimensional space that the channel vector belongs to) is computed and broadcast to all CHs so that they can use the corresponding transmit power from their pre-computed power codebooks. This index can only take a limited number of values due to the finite rate constraint on the feedback channel. In general, solving for the globally optimal power codebook is difficult due to the non-convexity of the associated optimization problems and the difficulty of exactly computing the probability of the channel vector belonging to a specific region defined by the index of the power codebook.

We propose power allocation schemes with low computational complexity by using various useful approximations for computing the probability of the channel vectors belonging to the multi-dimensional quantized regions and distributions of average sum power across the various regions. In the case of a single cluster, we obtain locally optimum power codebooks (due to non-convexity) while in the case of a multi-clustered network, we obtain strictly sub-optimal but low-complexity algorithms for designing the power codebooks. We also show that by approximating the boundaries between the adjacent quantized channel vector space by straight lines, it is asymptotically optimal (as average power tends to infinity) to allocate equal transmit power to all CHs. Numerical results illustrate that a few bits of feedback results in significantly lower outage than no CSIT and 6-8 bits of feedback can take the outage performance of a 6 cluster network to within 3 dB Watts of the average power required for the same outage performance with full CSIT. For comparison purposes, we also derive a simultaneous perturbation stochastic approximation (SPSA) based optimization method that results in improved outage performance compared (requires approximately 0.8 dBW lower average power at an outage of  $10^{-3}$  for 4 bits of feedback) to the above sub-optimal algorithms. However, the computational complexity and convergence time of the SPSA based algorithm are exceedingly high for it to be practically useful.

## II. SENSOR NETWORK MODEL AND PROBLEM FORMULATION

A schematic diagram of the wireless sensor network studied in this paper is shown in Fig. 1. The network is composed of  $N$  clusters where the  $n$ -th cluster contains  $M_n$  sensors and each cluster has a CH. The sensors are deployed around a single point source  $\theta[k]$ , where  $k = 0, 1, 2, \dots$  denotes discrete time instants, and take measurements of the source and transmit them to the corresponding CHs. The CHs then send the data to the FC which computes an estimate of the source  $\hat{\theta}[k]$  from all the measurements gathered. We make the following assumptions for our sensor network model: (1)  $\theta[k]$  is independent and identically distributed (i.i.d.) Gaussian (band-limited) random process of mean zero and variance  $\sigma_\theta^2$ , (2) the measurement noise for the  $m$ -th sensor within the  $n$ -th cluster  $N_m^n[k]$  is i.i.d. Gaussian-distributed of zero mean and variance  $(\sigma_m^n)^2$ . We assume that the noise variance is proportional to the square of the distance between the source and the sensors, (3) CHs are capable of transmitting with larger power than sensors, (4) the sensors within a cluster

amplify-and-forward observations to CH via a non-orthogonal multi-access scheme (distributed beamforming) with power gains of  $(\alpha_m^n)^2$  (for the  $m$ -th sensor in the  $n$ -th cluster), (5) The channel noise  $N_{C1}^n[k]$  (for transmissions from sensors to CH) is AWGN (additive white Gaussian noise) of zero mean and variance  $(\sigma_{C1}^n)^2$ , (6) signals received at each CH,  $y_n[k]$ , are not interfered by any signals from other clusters, (7) channels between sensors and CHs are static and the channel power gain  $g_m^n$  is proportional to the inverse of the square of the transmission distance, (8) CHs amplify-and-forward the received sum of sensor measurements to the FC with power gains of  $\beta_n^2$  using orthogonal multi-access (e.g. FDMA (frequency division multiple access)), (9) full CSI is available at the receiver (FC) and the CH transmitters have either full CSI (in the ideal case) or quantized CSI (which can be obtained by delayless and error-free but finite rate feedback broadcast channels from the FC), (10) channel noise  $N_{C2}^n[k]$  (transmissions from CHs to FC) is AWGN of zero mean and variance  $(\sigma_{C2}^n)^2$ , (11) the channels between CHs and FC are subject to Rayleigh block-fading. The channel power gain  $h_n$  (for the  $n$ -th CH) is i.i.d. exponentially-distributed of mean equal to the inverse of the transmission distance squared, given by  $h_n = \frac{1}{\lambda_n} f_n$  where  $\frac{1}{\lambda_n}$  is the mean and  $f_n$  is i.i.d. exponentially distributed with mean unity.

With the above assumptions, the signals of this network are described as follows. The observed sample of the  $m$ th sensor in the  $n$ th cluster at time  $k$  is given as

$$x_m^n[k] = \theta[k] + N_m^n[k]$$

The signal received at the  $n$ th CH at time  $k$  is given as

$$y_n[k] = \sum_{m=1}^{M_n} [\alpha_m^n \sqrt{g_m^n} (\theta[k] + N_m^n[k])] + N_{C1}^n[k] \quad (1)$$

The signal received at the FC from the  $n$ th CH is given as

$$z_n[k] = \beta_n \sqrt{h_n} y_n[k] + N_{C2}^n[k] \quad (2)$$

We can write the received signal in vector form as  $\mathbf{z} = \boldsymbol{\theta} + \mathbf{v}$  where

$$\begin{aligned} \mathbf{z} &= [z_1[k], \dots, z_N[k]]^T \\ \mathbf{s} &= \left[ \beta_1 \sqrt{h_1} \sum_{m=1}^{M_1} \alpha_m^1 \sqrt{g_m^1}, \dots, \beta_N \sqrt{h_N} \sum_{m=1}^{M_N} \alpha_m^N \sqrt{g_m^N} \right]^T \\ \mathbf{v} &= \left[ \beta_1 \sqrt{h_1} \left( \sum_{m=1}^{M_1} \alpha_m^1 \sqrt{g_m^1} N_m^1[k] + N_{C1}^1[k] \right) + N_{C2}^1[k], \right. \\ &\quad \left. \dots, \beta_N \sqrt{h_N} \left( \sum_{m=1}^{M_N} \alpha_m^N \sqrt{g_m^N} N_m^N[k] + N_{C1}^N[k] \right) + N_{C2}^N[k] \right]^T \end{aligned}$$

where  $^T$  denotes transposition.

In what follows we suppress the time index  $k$  for simplicity. The FC uses MMSE estimator to reconstruct the source  $\theta$ , since we have the prior pdf of  $\theta$ . The MMSE estimator is given as  $\hat{\theta} = \frac{\mathbf{s}^T \mathbf{C}^{-1} \mathbf{z}}{\frac{1}{\sigma_\theta^2} + \mathbf{s}^T \mathbf{C}^{-1} \mathbf{s}}$  where  $\mathbf{C}$  is a diagonal matrix with its  $n$ th diagonal element given as  $C_{nn} = \beta_n^2 h_n \left( \sum_{m=1}^{M_n} (\alpha_m^n)^2 g_m^n (\sigma_m^n)^2 + (\sigma_{C1}^n)^2 \right) + (\sigma_{C2}^n)^2$ . The variance of  $\hat{\theta}$  (also called distortion in this context) is given by  $\text{var}(\hat{\theta}) = \left[ \frac{1}{\sigma_\theta^2} + \mathbf{s}^T \mathbf{C}^{-1} \mathbf{s} \right]^{-1}$ .

Let  $q_n$  be the total power of sensors in the  $n$ th cluster and  $P_n$  the transmit power of the  $n$ th CH. Following the assumption made in [3] that all sensors within a cluster transmit with equal power ( $q_n/M_n$ ), we can obtain the following

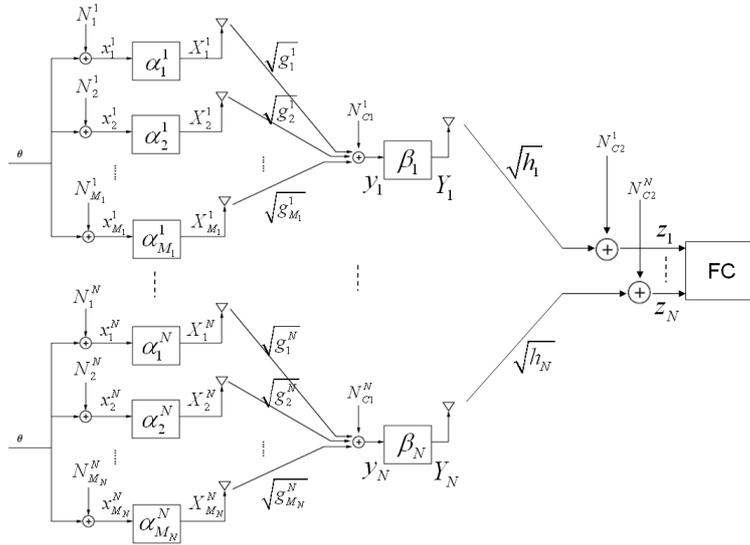


Fig. 1. Schematic diagram of a wireless sensor network for distributed estimation

expressions of sensor power gain, CH power and distortion respectively.

$$\alpha_m^n = \sqrt{\frac{q_n}{M_n [\sigma_\theta^2 + (\sigma_m^n)^2]}}, P_n = \beta_n^2 C_n \quad (3)$$

$$\text{var}[\hat{\theta}] = \sigma_\theta^2 \left( 1 + \sum_{n=1}^N \frac{\beta_n^2 h_n U_n}{\beta_n^2 h_n V_n + (\sigma_{Cn}^n)^2} \right)^{-1} \quad (4)$$

where  $C_n = (q_n/M_n) \sum_{m=1}^{M_n} g_m^n + (\sigma_{Cn}^n)^2$ ,  $U_n = (q_n/M_n) \left( \sum_{m=1}^{M_n} \sqrt{g_m^n} / (1 + (\gamma_m^n)^{-1}) \right)$ ,  $V_n = (q_n/M_n) \sum_{m=1}^{M_n} (g_m^n (\gamma_m^n)^{-1}) / (1 + (\gamma_m^n)^{-1}) + (\sigma_{Cn}^n)^2$  and  $\gamma_m^n = \sigma_\theta^2 / (\sigma_m^n)^2$ . Note that  $U_n, V_n, C_n$  are parameters available at the CH and contain information that captures the topology of each cluster.

We are now ready to formulate our optimization problems. We first present the problem of having full CSIT (channel state information at the transmitter) in Section II-A and then formulate the problem of quantized channel feedback in Section II-B.

### A. Power allocation with Full CSI

In this section we re-state the full CSI power allocation problem studied in [3] for Rayleigh-faded channels. The optimization problem is to obtain the optimal power allocation scheme that minimizes distortion outage subject to a long-term average power constraint  $P_{av}$ . Distortion outage probability is defined as the probability that distortion exceeds some maximum allowable distortion  $D_{max}$ . The problem is given as

$$\begin{aligned} \min \quad & \Pr(D(\mathbf{P}(\mathbf{h}), \mathbf{h}) > D_{max}) \\ \text{s.t.} \quad & E[\langle \mathbf{P}(\mathbf{h}) \rangle] \leq P_{av} \\ & \mathbf{P}(\mathbf{h}) \geq 0. \end{aligned} \quad (5)$$

where  $\mathbf{P}(\mathbf{h}) \triangleq (P_1(\mathbf{h}), \dots, P_N(\mathbf{h}))$  and  $\mathbf{h} \triangleq (h_1, \dots, h_N)$ .  $\Pr(A)$  denotes probability of the event  $A$  and  $\langle \mathbf{x} \rangle$  denotes the arithmetic mean of the vector  $\mathbf{x}$  of length  $M$  defined by  $\langle \mathbf{x} \rangle \triangleq \frac{1}{M} \sum_{i=1}^M x_i$ .  $D(\mathbf{P}(\mathbf{h}), \mathbf{h})$  is the distortion achieved at the FC as a function of channel gains and CH transmission power, which is also a function of the channel gains. For further details on Problem (5), see [3].

### B. Power allocation with quantized CSI

In rate-limited feedback channels, only a finite set of power values can be fed back to the transmitters. Denote the number of feedback bits as  $R$  and the number of power levels as  $L = 2^R$ . For an  $R$ -bit broadcast feedback channel and  $N$

clusters, we quantize the vector channel space  $\mathfrak{R}_+^N$  into  $L$  regions. We begin by studying the single-cluster network ( $N = 1$ ). Define the channel thresholds that divide the channel space into  $L$  regions as  $s_1, \dots, s_L$ . Without loss of generality, assume that  $s_1 < \dots < s_L$ . Following a similar result in [6], for  $N = 1$ , given some arbitrary power allocation shown in Fig. 2, it is always possible to reduce the outage probability by re-distributing the power so that outage occurs only in  $[0, s_1)$ . This essentially means that the optimal way of allocating power involves pouring enough power at each channel quantization region so that maximum distortion constraint is always met except in  $[0, s_1)$ . The remaining task is to find the optimal quantization levels such that  $s_1$  is minimized (which will in turn minimize the outage probability).

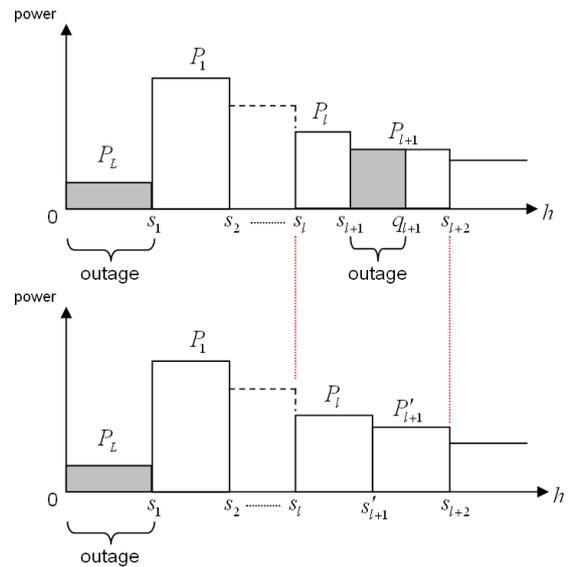


Fig. 2. Top: An arbitrary CH power allocation. Bottom: Adjusted power allocation while still satisfying the average power constraint.

We now extend the above idea to the multi-cluster case. To illustrate the problem, we give an example of a 2-cluster ( $N = 2$ ) network with an  $R$ -bit feedback channel in Fig. 3 which shows the vector channel quantization regions over the channel space  $\mathbf{h} \in \mathfrak{R}_+^2$ . Denote the power codebook as  $\mathcal{P} = \{P_1, \dots, P_L\}$ , and without loss of generality, assume  $P_1 > P_2 > \dots > P_L$ , where  $P_k > P_l$  if  $D(\mathbf{h}, P_k) < D(\mathbf{h}, P_l)$ . Note that because distortion is a non-increasing

function of  $P_{i,j}$ , the inequality  $\mathcal{P}_k > \mathcal{P}_l$  also holds in an element by element sense for the vector  $\mathcal{P}_j$ . Each power codeword  $\mathcal{P}_j = \{P_{i,j}\}$  contains a set of  $N$  power values that specify the transmitting power of the CHs, where  $i = 1, \dots, N$  is the index of the cluster. We assume the CHs and the FC know this (pre-computed) power codebook, since this power codebook can be computed purely based on the channel statistics and the available average power. Depending on the CSI at the receiver, the FC then sends the index of the power codebook to all the CHs in a broadcast manner.

The  $L$  quantization regions are formed by a series of boundaries  $B_j$ ,  $j = 1, \dots, L$ . Motivated by the above observation (that the outage region is contained within the first interval) (see also [4]), it is easily seen that outage only occurs in the region bounded by  $B_1$  and the axes. This implies that we must allocate power in all regions (except for the outage region) such that distortion constraint is always met. The boundaries  $B_j$  hence are distortion curves where the distortion is equal to  $D_{max}$  at any point on the boundaries. The boundaries are functions of the channel thresholds (or power). Denote the channel gain thresholds as  $s_{i,j}$  where  $i = 1, \dots, N$  is the index of the cluster and  $j = 1, \dots, L$  is the index of the boundary. For convenience, let  $\mathbf{s}_j = \{s_{1,j}, \dots, s_{N,j}\}$  denote the set of thresholds that defines boundary  $B_j$ . The equation of the distortion curve  $B_j$  as a function of the transmit power ( $P_{i,j} = C_i \beta_i^2$ ) is given as

$$D_{max} = \sigma_\theta^2 \left( 1 + \sum_{i=1}^N \frac{P_{i,j} h_i U_i}{P_{i,j} h_i V_i + C_i (\sigma_{C2}^i)^2} \right)^{-1} \quad (6)$$

for  $j = 1, \dots, L$ . The region  $\mathcal{R}_j$  is defined as  $\mathcal{R}_j = \{\mathbf{h} : D(\mathbf{h}, \mathcal{P}_j) \leq D_{max} \cap D(\mathbf{h}, \mathcal{P}_{j+1}) > D_{max}\}$  for  $j = 1, \dots, L-1$  and  $\mathcal{R}_L = \{\mathbf{h} : \mathcal{R}_{out} \cup D(\mathbf{h}, \mathcal{P}_L) \leq D_{max}\}$ . Following [8], [9] where it was shown that the optimal index mapping of the power codebook for outage minimization has a circular structure, the outage region  $\mathcal{R}_{out} = \{\mathbf{h} : D(\mathbf{h}, \mathcal{P}_1) > D_{max}\}$ , is a subset of  $\mathcal{R}_L$ . The circular index mapping structure dictates that we assign the power codeword  $\mathcal{P}_L$  to the  $L$ -th quantized region which includes the outage region (since in the outage region, it only makes sense to allocate the power codeword with the smallest elements).

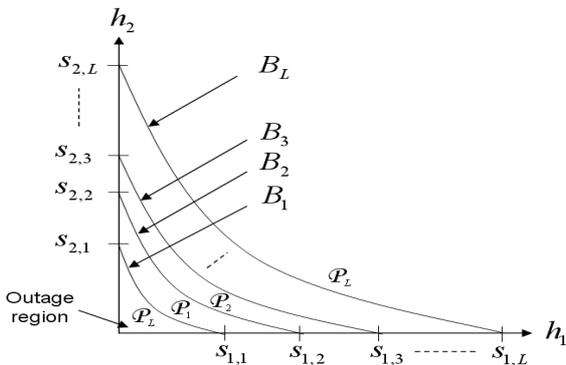


Fig. 3. Vector channel quantization regions formed by a series of distortion curves for a 2-cluster network.

We are interested in minimizing the distortion outage probability subject to a long-term average power constraint for vector channel quantization. Let  $F_N(\mathbf{s}_j)$  be defined as  $F_N(\mathbf{s}_j) = \Pr(\mathbf{h} < B_j)$  where the set  $\{\mathbf{h} < B_j\}$  is defined as  $\{\mathbf{h} : D(\mathbf{h}, \mathcal{P}_j) \geq D_{max}\}$ . An  $N$ -cluster,  $R$ -bit feedback

optimization problem is then given as

$$\begin{aligned} \min & F_N(\mathbf{s}_1) \\ \text{s.t.} & \sum_{j=1}^{L-1} \mathbf{P}_j [F_N(\mathbf{s}_{j+1}) - F_N(\mathbf{s}_j)] \\ & + \mathbf{P}_L (1 - F_N(\mathbf{s}_L) + F_N(\mathbf{s}_1)) \leq NP_{av} \\ & 0 \leq s_{j,i} \leq s_{j,i+1} \quad \forall j, i. \end{aligned} \quad (7)$$

where  $\mathbf{P}_j = \sum_{i=1}^N P_{i,j}$  denotes the sum (across the CHs) transmit power for the  $j$ -th quantized region.

### III. POWER ALLOCATION SCHEMES AND SOLUTIONS

#### A. Full CSI

The solution to problem (5) is given in [3] and will not be repeated here due to space limitations.

#### B. Quantized CSI

Problem (7) is non-convex in general, but we can still solve it using the standard Lagrange multiplier based optimization technique by finding the Karush-Kuhn-Tucker (KKT) necessary conditions. However the solutions obtained can only be locally optimal as global optimality cannot be guaranteed. It can be easily shown that the second constraint in (7) must be satisfied with a strict inequality. We will discard the second constraint in what follows as it will not affect the result. Note that the second constraint also implies that  $\mathbf{P}_1 > \dots > \mathbf{P}_L$ . The Lagrangian is

$$L = F_N(\mathbf{s}_1) + \mu \left[ \sum_{j=1}^{L-1} \mathbf{P}_j (F_N(\mathbf{s}_{j+1}) - F_N(\mathbf{s}_j)) + \mathbf{P}_L (1 - F_N(\mathbf{s}_L) + F_N(\mathbf{s}_1)) - NP_{av} \right]$$

where  $\mu$  is the Lagrange multiplier. For ease of viewing, we write the partial derivatives of the cdf  $F_N(\mathbf{s}_j)$  and the sum power function  $\mathbf{P}_j$  with respect to any of its variables in  $\mathbf{s}_j$  as  $\partial F_N(\mathbf{s}_j)/\partial s_j$  and  $\partial \mathbf{P}_j/\partial s_j$ .

*Single-cluster network ( $N = 1$ ):* The KKT conditions of (7) for  $N = 1$  are given as

$$\begin{aligned} \frac{\lambda e^{-\lambda s_{i+1}}}{s_i} - \frac{e^{-\lambda s_{i+1}} - e^{-\lambda s_{i+2}}}{s_{i+1}^2} - \frac{\lambda e^{-\lambda s_{i+1}}}{s_{i+1}} &= 0, \\ i \in \{1, \dots, L_2\} \\ \frac{\lambda e^{-\lambda s_L}}{s_{L-1}} - \frac{1 - e^{-\lambda s_1} + e^{-\lambda s_L}}{s_L^2} - \frac{\lambda e^{-\lambda s_L}}{s_L} &= 0 \quad (8) \\ \sum_{i=1}^{L-1} \frac{e^{-\lambda s_i} - e^{-\lambda s_{i+1}}}{s_i} + \frac{1 - e^{-\lambda s_1} + e^{-\lambda s_L}}{s_L} &= \frac{P_{av}}{\phi} \end{aligned}$$

where  $\phi = C \sigma_{C2}^2 \gamma_{th} / (U - V \gamma_{th})$  and  $\gamma_{th} = \sigma_\theta^2 / D_{max} - 1$ . Note that the last KKT condition (the long-term average power constraint) must be met with equality since this is implied by the optimality condition. Problem (8) can be solved by fixed point iterative methods for solving nonlinear equations.

*Multi-cluster network ( $N \geq 2$ ):* The KKT conditions of (7) for  $N \geq 2$  are given as

$$\begin{aligned} \frac{\partial \mathbf{P}_1}{\partial s_{i,1}} \Big/ \frac{\partial F_N(\mathbf{s}_1)}{\partial s_{i,1}} &= \frac{\partial \mathbf{P}_1}{\partial s_{j,1}} \Big/ \frac{\partial F_N(\mathbf{s}_1)}{\partial s_{j,1}} \quad \forall i, j \\ &\vdots \\ \frac{\partial \mathbf{P}_L}{\partial s_{i,L}} \Big/ \frac{\partial F_N(\mathbf{s}_L)}{\partial s_{i,L}} &= \frac{\partial \mathbf{P}_L}{\partial s_{j,L}} \Big/ \frac{\partial F_N(\mathbf{s}_L)}{\partial s_{j,L}} \quad \forall i, j \quad (9) \\ \sum_{j=1}^{L-1} \mathbf{P}_j (F_N(\mathbf{s}_{j+1}) - F_N(\mathbf{s}_j)) &+ \mathbf{P}_L (1 - F_N(\mathbf{s}_L) + F_N(\mathbf{s}_1)) = NP_{av}. \end{aligned}$$

In general, computing the probability of the channel vector belonging to the  $j$ -th quantized region, namely  $F_N(\mathbf{s}_j)$ , for  $N > 1$  involves multi-dimensional integrals as a function of the distortion curves and are difficult to be expressed in closed forms. Instead, we propose that the distortion curves be approximated by planes formed by joining the intersections of the distortion curves and the axes shown by the upper straight line in Fig. 4. We call this approximation the outer-straight-line approximation and denote the  $i$ th plane as  $\bar{B}_i$ . We can also construct another plane that is parallel to this plane and is also the tangent of the distortion curve, shown by the lower straight line in Fig. 4. We call this the inner-straight-line approximation and denote the  $i$ th plane as  $\underline{B}_i$ . Simulation results show that these two approximations give very comparable outage performances, hence the rest of the paper will be based on the outer-straight-line approximation (referred in this paper simply as the straight-line approximation (SLA)). An example of SLA for  $N = 2$  is shown in Fig. 5.

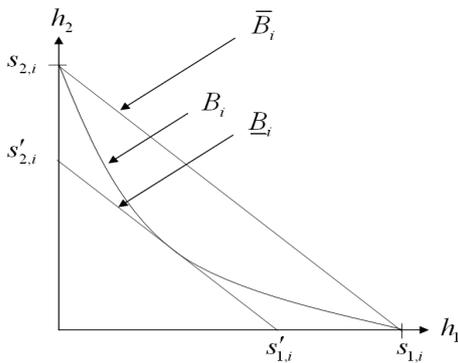


Fig. 4. Inner and outer straight-line approximations.

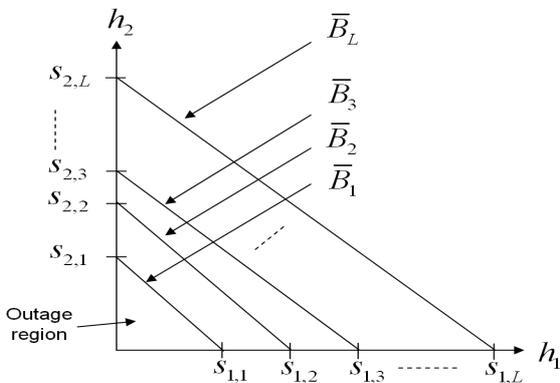


Fig. 5. Quantization regions formed by the straight-line approximation.

The KKT conditions shown in (9) constitute a set of nonlinear equations, where the number of equations grows exponentially as the number of feedback bits increases. For large number of feedback bits we use an approximation called equal average power per region (EPPR) as in [10], [4], where the authors used the mean-value theorem to obtain this approximation. However before we can write down the problem formulation with EPPR approximation, we must deal with the issue of whether we should allocate power in the outage region or not. If we put nonzero power in the outage region (NZPOR), the channel space is quantized into  $L$  regions with  $L - 1$  non-outage regions and the  $L$ -th region containing a non-outage region as well as an outage region due to the circular nature mentioned earlier. In the case of zero-power-in-outage-region (ZPOR), there would be  $L$  regions with  $L - 1$  non-outage regions and an outage region. Note

that in the general problem definition in (7), ZPOR is in fact a special case of NZPOR (NZPOR is hence a more general model than ZPOR). This case arises when the outer-most channel thresholds  $s_{i,L} \rightarrow \infty \forall i$ . The condition when ZPOR is optimal may depend on the network topology and the long-term average power. We observe through simulation that using EPPR approximation, ZPOR performs nearly optimally when the average power is very low (the actual threshold below which ZPOR performs near-optimally depends on  $N$  and  $R$ ). Similar observations for near-optimality of ZPOR can also be found in [7]. We now provide the problem formulations using the EPPR approximation for NZPOR and ZPOR, by (10) and (11) respectively.

$$\begin{aligned} \min \quad & \bar{F}_N(\mathbf{s}_1) \\ \text{s.t.} \quad & \mathbf{P}_1(\bar{F}_N(\mathbf{s}_2) - \bar{F}_N(\mathbf{s}_1)) = \frac{NP_{av}}{L} \\ & \vdots \\ & \mathbf{P}_{L-1}(\bar{F}_N(\mathbf{s}_L) - \bar{F}_N(\mathbf{s}_{L-1})) = \frac{NP_{av}}{L} \\ & \mathbf{P}_L(1 - \bar{F}_N(\mathbf{s}_L) + \bar{F}_N(\mathbf{s}_1)) = \frac{NP_{av}}{L}. \end{aligned} \quad (10)$$

$$\begin{aligned} \min \quad & \bar{F}_N(\mathbf{s}_1) \\ \text{s.t.} \quad & \mathbf{P}_1(\bar{F}_N(\mathbf{s}_2) - \bar{F}_N(\mathbf{s}_1)) = \frac{NP_{av}}{L-1} \\ & \vdots \\ & \mathbf{P}_{L-1}(\bar{F}_N(\mathbf{s}_L) - \bar{F}_N(\mathbf{s}_{L-1})) = \frac{NP_{av}}{L-1} \end{aligned} \quad (11)$$

where  $\bar{F}_N(\mathbf{s}_j)$  denotes the probability that the channel vector belongs to the  $j$ -th quantization region as obtained by the SLA.

The following result shows that at high average power, using the SLA, one can further approximate the optimal power allocation scheme. The proof is excluded due to space limitations.

*Lemma 3.1:* Based on SLA, as the long-term average power tends to infinity, it is asymptotically optimal to transmit with equal transmit power per CH for every quantization region, i.e.,  $P_{i,j} = P_{k,j} \forall i, k, i \neq k$ .

Hence problem (10) and (11) can be further simplified at high average power by letting all the CHs transmit with equal power. We will abbreviate equal power per CH as EPPC. Each boundary is now a function of a single scalar variable, denoted as  $P_j$ . The channel thresholds  $s_{i,j}$  can be found by  $s_{i,j} = \phi_i/P_j$ . We can also express channel thresholds belonging to the same boundary as a function of  $s_{1,j}$  given as  $s_{i,j} = R_i s_{1,j}$  where  $R_i = \phi_i/\phi_1 \forall i$ . Using SLA, EPPR and EPPC, problem (10) becomes

$$\begin{aligned} \min \quad & \bar{F}_N(s_{1,1}) \\ \text{s.t.} \quad & P_1(\bar{F}_N(s_{1,2}) - \bar{F}_N(s_{1,1})) = \frac{P_{av}}{L} \\ & \vdots \\ & P_{L-1}(\bar{F}_N(s_{1,L}) - \bar{F}_N(s_{1,L-1})) = \frac{P_{av}}{L} \\ & P_L(1 - \bar{F}_N(s_{1,L}) + \bar{F}_N(s_{1,1})) = \frac{P_{av}}{L} \end{aligned} \quad (12)$$

Problem (10),(11) and (12) can be solved directly from the corresponding KKT conditions.

1) *Power Allocation for Quantized CSI Using a Simultaneous Perturbation Stochastic Approximation (SPSA) Algorithm:* The vector channel quantization problem can be formulated as the classical vector quantization problem with a modified distortion measure, and the solution can be found by using the iterative Lloyd's algorithm incorporating SPSA [12]. The Lloyd's algorithm can only find locally optimal points. Nonetheless it can provide a performance benchmark (since it does not use any approximations for the distortion curves or the power allocation for various quantization regions or CHs) with which we can compare with the results obtained by solving the KKT conditions for the various approximations using SLA, EPPR and EPPC.

We are interested in finding the optimal power codebook that minimizes the outage probability for quantized CSI feedback. The Lloyd iteration for codebook improvement has two steps. In the first step, given the power codebook  $\mathcal{P}$  one finds the optimal partition for the quantization cells using the nearest neighbor condition. In our context the quantization cells are formed by solving the following optimization problem

$$\arg \max_{\mathcal{P}_j} D(\mathbf{h}, \mathcal{P}_j) \quad \text{s.t.} \quad D(\mathbf{h}, \mathcal{P}_j) \leq D_{max} \quad (13)$$

The solution of problem (13) yields the regions  $\mathcal{R}_j$ ,  $j = 1, \dots, L$ , formed by generating a large number of random channel vectors  $\mathbf{h}$  according to the joint probability distribution  $\{h_1, h_2, \dots, h_N\}$  and grouping the set of  $\mathbf{h}$  that belong to  $\mathcal{P}_j$ ,  $j = 1, \dots, L$ . In other words, we associate a channel vector to the power codeword that achieves the largest distortion from the set of power codewords that meet the distortion constraint. There are, of course, channel vectors for which none of the power codewords from the given power codebook can meet the distortion constraint. This corresponds to outage. The circular structure of the index mapping implies that we should assign this set of channel vectors to the power codeword that achieves the largest distortion ( $\mathcal{P}_L$ ). Note that  $\mathcal{R}_{out} \subseteq \mathcal{R}_L$ .

In the second step we find the improved power codebook. This involves solving the optimization problem given as

$$\begin{aligned} \min \quad & E[1(D(\mathbf{h}, \mathcal{P}_1) > D_{max}) | \mathbf{h} \in \mathcal{R}_{out}] \Pr(\mathbf{h} \in \mathcal{R}_{out}) \\ \text{s.t.} \quad & \sum_{j=1}^L (\mathbf{P}_j \Pr(\mathbf{h} \in \mathcal{R}_j)) \leq NP_{av} \end{aligned} \quad (14)$$

Because we do not have an explicit expression for the outage probability (the objective function above), we apply the SPSA algorithm to search for the new power codebook. SPSA is a stochastic optimization method which can be applied to problems where perfect information of the loss function (the objective function) is not available or corrupted by noise, e.g. the derivatives where these information are required to determine the search direction [13]. Stochastic optimization randomly chooses the search direction and iterates toward a (locally optimal) solution. SPSA has the advantage that it only requires 2 loss measurements, which is independent of the dimension of the problem, and uses these two measurements to form a gradient approximation [13], [14]. An enhanced version of the algorithm called adaptive stochastic approximation (ASP) that approximates the Hessian has also been developed in [15] but it requires 4 loss measurements. In this paper we use the SPSA due to its simplicity. The loss function  $L$  is given as  $L = \Pr(\mathbf{h} \in \mathcal{R}_{out}) + \bar{\lambda} \sum_{j=1}^L (\mathbf{P}_j \Pr(\mathbf{h} \in \mathcal{R}_j))$  where  $\bar{\lambda}$  is the Lagrangian multiplier. Since the loss function can be viewed as an unconstrained optimization problem, we will have to obtain  $P_{av}$  numerically as a function of  $\lambda$ . Once the new power codebook is found, we repeat step 1 and step 2 until the stopping criterion is met. Denote  $\tilde{\mathcal{P}}$  as the  $NL$  by 1 column vector given as  $\tilde{\mathcal{P}} = [\mathcal{P}_1, \mathcal{P}_2, \dots, \mathcal{P}_L]^T$  where  $^T$  denotes transposition. The 2-sided SPSA algorithm can be summarized by the following steps [14]:

- 1) *Initialization and Coefficient Selection*: Set counter index  $k = 0$ . Guess initial power codebook  $\tilde{\mathcal{P}}_0$  and set non-negative coefficients  $a$ ,  $c$ ,  $A$ ,  $\alpha$  and  $\gamma$  in the SPSA gain sequences  $a_k = a/(A + k + 1)^\alpha$  and  $c_k = c/(k + 1)^\gamma$ . For guideline on choosing these coefficients see [14].
- 2) *Generation of Simultaneous Perturbation*: Generate a  $NM$ -dimensional random perturbation column vector  $\Delta_k$ . Each component of  $\Delta_k$  are i.i.d Bernoulli  $\pm 1$  distributed with probability of 0.5 for each  $\pm 1$  outcome.

- 3) *Loss Function Evaluations*: Obtain two measurements of the loss function  $L$  based on the simultaneous perturbation around the current power codebook  $\tilde{\mathcal{P}}_k$ :  $L(\tilde{\mathcal{P}}_k + c_k \Delta_k)$  and  $L(\tilde{\mathcal{P}}_k - c_k \Delta_k)$  with  $c_k$  and  $\Delta_k$  from Steps 1 and 2.
- 4) *Gradient Approximation*: Generate the simultaneous perturbation approximation to the unknown gradient  $\hat{g}_k(\tilde{\mathcal{P}}_k)$  given as

$$\hat{g}_k(\tilde{\mathcal{P}}_k) = \frac{L(\tilde{\mathcal{P}}_k + c_k \Delta_k) - L(\tilde{\mathcal{P}}_k - c_k \Delta_k)}{2c_k} \begin{bmatrix} \Delta_{k,1}^{-1} \\ \Delta_{k,2}^{-1} \\ \vdots \\ \Delta_{k,NL}^{-1} \end{bmatrix}$$

where  $\Delta_{k,i}$  is the  $i$ th component of the  $\Delta_k$  vector.

- 5) *Updating power codebook*: Use the standard SA form

$$\tilde{\mathcal{P}}_{k+1} = \tilde{\mathcal{P}}_k - a_k \hat{g}_k(\tilde{\mathcal{P}}_k)$$

- 6) *Iteration or Termination*: Return to Step 2 with  $k + 1$  replacing  $k$ . Terminate the algorithm if there is little change in several successive iterations or the maximum allowable number of iterations has been reached.

*Remark 1*: SPSA is computationally intensive and requires tuning  $\bar{\lambda}$  and all the coefficients whenever network parameter changes, such as the average power constraint or feedback resolution. Convergence can be slow and may settle to different local minima depending on the initial points chosen. Hence in the next section, we will only provide limited SPSA results (up to 4 bits of feedback) for providing some performance comparison.

#### IV. SIMULATION RESULTS

We simulate a clustered wireless sensor networks as shown in Fig. 6. The clusters are deployed in an equi-distant fashion from the source in a uniformly-spaced manner. For  $N = 6$  the network topology is simply given in Fig. 6. For  $N = 2$  we take the top left and bottom right clusters and discard the rest. For  $N = 1$  we take the top left cluster only. The sensors in each cluster are organized in four equally-spaced concentric circles and the number of sensors in each circle are 6, 12, 18 and 24 from the smallest to the biggest circle respectively. All clusters have a radius of 40m. All sensors transmit with a power of  $q_n/M_n = 1mW$ . The CHs are located at the center of each cluster for simplicity. CHs are 100m apart from the next closest CH (for the 6-cluster network). The FC is located 500m away from the source. The channel noise variance are set to  $(\sigma_{C1}^n)^2 = 10^{-12}$  Watts and  $(\sigma_{C2}^n)^2 = 10^{-10}$  Watts  $\forall n$  in the first and second stage of transmission respectively. Source variance is set to  $\sigma_\theta^2 = 1$  Watt. The maximum distortion threshold  $D_{max}$  is set to 0.0043 (10% of the minimum achievable distortion of the 6-cluster network). Recall that the expressions of the outage probability for  $N \geq 2$  do not have closed form expressions. The outage performance shown in this section are obtained via Monte Carlo simulation over 1,000,000 channel realizations using the locally optimum power allocation (for  $N = 1$  and the SPSA algorithm) and the strictly sub-optimal power allocation obtained via SLA and the various other approximations such as EPPC and EPPR etc. For very low average power (the exact threshold depends on the values of  $N$  and  $R$ ), the outage performance is obtained using the ZPOR algorithm as the NZPOR algorithm can become degenerate for such low  $P_{av}$  values. The outage performance with full CSIT is simulated using the optimal power allocation results from [3]. We now present the simulation results for  $N = 1, 2, 6$ .

Fig. 7 shows the outage performance of a single-cluster network with 1 and 2-bit feedback. We compare EPPR with the optimal outage performance obtained by solving (8). It can be seen from graph that the EPPR approximation is a

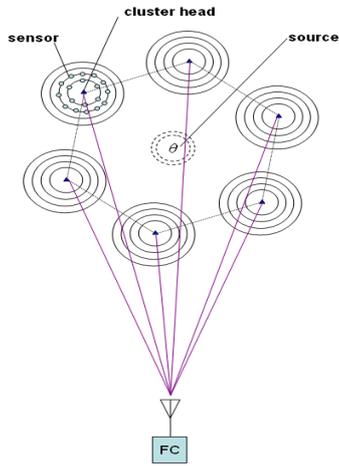


Fig. 6. Wireless sensor network topology.

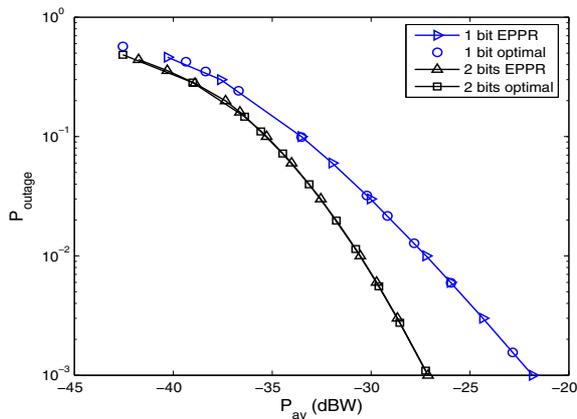


Fig. 7. Comparing the outage probability of a single-cluster network employing 1 and 2 feedback bits with and without EPPR.

very good approximation, giving results that almost coincide with the optimal outage performance.

Fig. 8 shows the outage performance for  $N = 1$  and  $R = 1, 2, 3, 4, 6$ . The outage performances of no-CSI at transmitter and the optimal power allocation with full CSI are also plotted for comparison. In the case of no-CSI, all the CHs simply transmit with equal average power. Utilizing the feedback channel for providing partial CSIT significantly improves the outage performance. As we increase the number of feedback bits to  $R = 6$ , the outage performance approaches full CSI.

We provide some performance comparison for using EPPR and EPPC approximations and exhaustive search (ES) for  $N = 2$  in Fig. 9. The results show that EPPR+EPPC approximations give results that are very close to the ones using SLA where CHs are allowed to transmit with different power and where the average power in each region may not necessarily be the same. In order to comment on how good SLA is, we obtain the outage performance of 1-bit feedback based on distortion curves using exhaustive search method (ES). Results of ES matches closely with SLA, showing that SLA is a good approximation, at least for 1-bit feedback, 2-cluster case. In obtaining the results for SLA for  $R = 2$ , the ZPOR algorithm was used to find the optimal power allocation for  $P_{av} \leq -40.44$  dBW and the NZPOR algorithm was used to find the optimal power allocation for all other values of  $P_{av}$ .

Fig. 10 shows the outage performance obtained by SPSA algorithm and EPPC+EPPR approximation for  $R = 1, 2, 4$  and  $N = 2$ . Since SPSA is sensitive to initial conditions

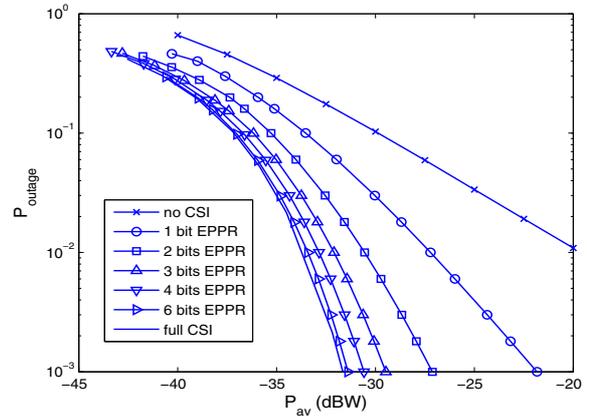


Fig. 8. Outage performance of a single-cluster network employing 1,2,3,4 and 6 feedback bits and optimal full-CSI power allocation.

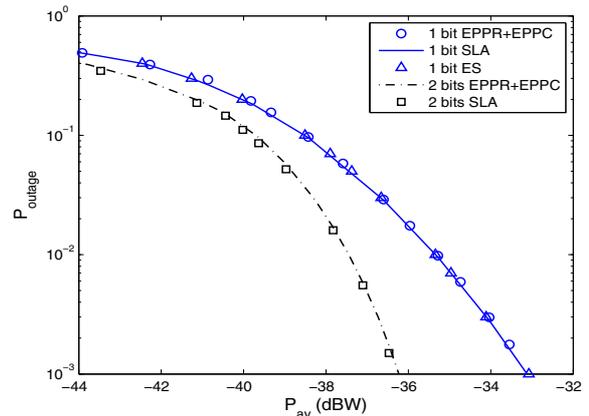


Fig. 9. Outage performance comparison: EPPC+EPPR versus SLA and ES,  $R = 1, 2$ ,  $N = 2$

and only works well with initial points close to the optimal point, we first run the EPPR+EPPC algorithm and use these solutions as our initial points for SPSA. Each point on the graph is obtained by running SPSA with different initializations for a given  $\lambda$ , and then selecting the one that provides the lowest outage performance. In step 2 of the Lloyd's algorithm outlined in section III-B.1 the probabilities are approximated by Monte Carlo simulation over 100,000 vector channel realizations. The coefficients used in SPSA algorithm are:  $c = 10^{-5}$ ,  $A = 80$ ,  $\alpha = 0.602$ ,  $\gamma = 0.101$  and  $a = 10^{-6} \cdot (A + 1)^\alpha / (\text{mean magnitude of } \hat{g}_0)$ . Simulation results show that SPSA gives similar outage performance as to EPPC+EPPR approximation.

Fig. 11 shows the outage performance of  $N = 2$  and  $R = 1, 2, 4, 6$ . The SLA curves are obtained using an optimization software called '1stOpt'<sup>1</sup>. SLA performs slightly better than EPPC because SLA allows different CH powers, while EPPR is a special case of the unequal CH power allocation. The advantage of allocating different powers to different CHs is reflected by the number of CHs  $N$ . The difference is little here because  $N$  is small. We will see that this gap widens in  $N = 6$ . Also shown in the graph is the significant improvement in performance achieved by using limited feedback compared with no CSI based in [3].

We now present simulation results for  $N = 6$  shown in Fig. 12. The outage performance of EPPR and EPPC for  $R = 1, 2, 3, 4, 6, 8$  are shown. There is significant power

<sup>1</sup>1stOpt is a mathematical program developed by 7D-Soft High Technology Inc. Levenberg-Marquardt and Universal Global Optimization methods were used to solve our problem. For more information please go to [www.7d-soft.com](http://www.7d-soft.com).

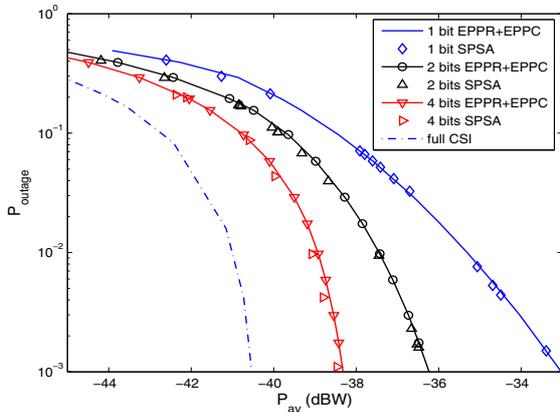


Fig. 10. Outage performance comparison of a 1- and 2-bit feedback, 2-cluster network obtained by SPSA and EPPC+EPPR approximation.

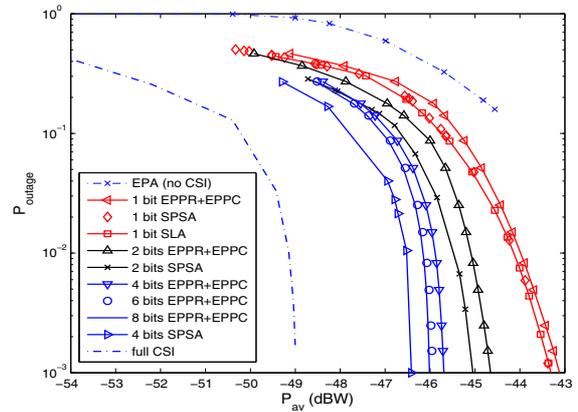


Fig. 12. Outage performance of a 6-cluster network with various power allocation schemes.

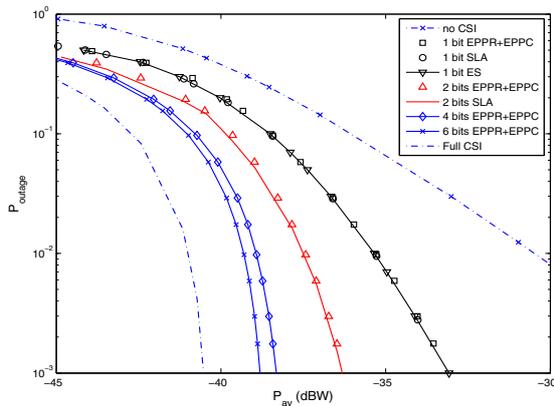


Fig. 11. Outage performance of a 2-cluster network with various power allocation schemes.

gain in employing feedback over no CSI where equal power allocation is used. As the number of feedback bit increases, less power is required to achieve the same outage probability. However there is still a gap between 8-bit feedback and full CSI (2.66dBW at  $P_{outage} = 0.01$ ). Also note that 1-bit SLA curve now has larger power gains over 1-bit EPPR+EPPC (0.17dBW at  $P_{outage} = 0.01$ ). This is reflected by a larger number of clusters that contribute to diversity when they are allowed to transmit with different power.

Fig. 12 also shows SPSA results for  $R = 1, 2$  and 4. The coefficients used in SPSA are the same as for  $N = 2$ , and we again use 100,000 vector channel realizations for Monte Carlo simulation in step 2 of the Lloyd's algorithm. The initial points are obtained by running EPPR+EPPC. For  $R = 1$  the result matches closely with 1-bit SLA. For  $R = 2$  and 4, SPSA has a power gain of 0.5dBW and 0.66dBW at  $P_{outage} = 0.01$  over EPPR+EPPC respectively. However we do not give further results using SPSA due to the computational complexity involved in the algorithm as explained in Remark 1.

## V. CONCLUSIONS

In this paper we present a number of low-complexity algorithms for minimizing distortion outage in a multi-clustered sensor network where the sensors measure a random Gaussian source and forward their noisy measurements to their respective CH by analog forwarding using a non-orthogonal multi-access scheme (under the assumption of distributed beamforming). The CHs then also use analog forwarding to transmit their received signals over orthogonal Rayleigh distributed block-fading channels to a fusion center which reconstructs an estimate of the source. The distortion (which

is a function of the random fading channels between the CHs and the fusion center) is required to be less than a certain maximum threshold and can violate this condition only with a certain small probability, called the distortion outage probability. Based on the restriction of limited channel feedback (the fusion center can only broadcast a fixed number of bits to the CHs), we propose a number of low-complexity outage minimization algorithms with various levels of useful approximations. An extensive set of numerical results are presented to demonstrate the performance of these algorithms.

## REFERENCES

- [1] G. Caire, G. Taricco, and E. Biglieri, "Optimum power control over fading channels," *IEEE Transactions on Information Theory*, vol. 5, pp. 1468–1489, July 1999.
- [2] S. Cui, J. J. Xiao, A. J. Goldsmith, Z. Q. Luo, and H. V. Poor, "Estimation diversity and energy efficiency in distributed sensing," *IEEE Transactions on Signal Processing*, vol. 9, pp. 4683–4695, Sept. 2007.
- [3] C. Wang and S. Dey, "Power allocation for distortion outage minimization in clustered wireless sensor networks," in *Proc. IEEE International Wireless Communications and Mobile Computing Conference (IWCMC'08)*, Crete, Greece, Aug. 2008, pp. 395–400.
- [4] N. Ahmed, M. A. Khojastepour, A. Sabharwal, and B. Aazhang, "Outage minimization with limited feedback for the fading relay channel," *IEEE Transactions on Communications*, vol. 54, pp. 659–669, Apr. 2006.
- [5] K. K. Mukkavilli, A. Sabharwal, E. Erkip, and B. Aazhang, "On beamforming with finite rate feedback in multiple-antenna systems," *IEEE Transactions on Information Theory*, vol. 49, pp. 2562–2579, Oct. 2003.
- [6] L. Lin, R. D. Yates, and P. Spasojevic, "Adaptive transmission with discrete code rates and power levels," *IEEE Transactions on Communications*, vol. 51, pp. 2115–2125, Dec. 2003.
- [7] T. T. Kim and M. Skoglund, "Partial power control for slowly fading MIMO channels," in *Proc. IEEE International Conference on Communications (ICC'06)*, vol. 3, June 2006, pp. 1362–1367.
- [8] —, "Diversity-multiplexing tradeoff in MIMO channels with partial CSIT," *IEEE Transactions on Information Theory*, vol. 53, pp. 2743–2759, Aug. 2007.
- [9] M. A. Khojastepour, G. Yue, X. Wang, and M. Madhian, "Optimal power control in MIMO systems with quantized feedback," *IEEE Transactions on Wireless Communications*, vol. 7, pp. 4859–4866, Dec. 2008.
- [10] A. Khoshnevis and A. Sabharwal, "Performance of quantized power control in multiple antenna systems," in *Proc. IEEE International Conference on Communications (ICC'04)*, June 2004, pp. 803–807 Vol.2.
- [11] S. Ekbatani, F. Etemadi, and H. Jafarkhani, "Outage behaviour of slow fading channels with power control using noisy quantized CSIT," presented in part at the IEEE Global Communications Conference, Washington DC, November, 2007.
- [12] A. Gersho and R. M. Gray, *Vector quantization and signal compression*. Massachusetts, USA: Kluwer Academic Publishers, 1992.
- [13] J. Gentle, W. Härdle, and Y. Mori, (Editors), *Handbook of computational statistics*. Germany: Springer, 2004.
- [14] J. C. Spall, "Implementation of the simultaneous perturbation algorithm for stochastic optimization," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 34, pp. 817–823, July 1998.
- [15] —, "Adaptive stochastic approximation by the simultaneous perturbation method," *IEEE Transactions on Automatic Control*, vol. 45, pp. 1839–1853, Oct. 2000.