

Optimal Power Scheduling for Correlated Data Fusion in Wireless Sensor Networks Via Constrained PSO

Thakshila Wimalajeewa, *Student Member, IEEE*, and Sudharman K. Jayaweera, *Member, IEEE*

Abstract—Optimal power scheduling for distributed detection in a Gaussian sensor network is addressed for both independent and correlated observations. We assume amplify-and-forward local processing at each node. The wireless link between sensors and the fusion center is assumed to undergo fading and coefficients are assumed to be available at the transmitting sensors. The objective is to minimize the total network power to achieve a desired fusion error probability at the fusion center. For i.i.d. observations, the optimal power allocation is derived analytically in closed form. When observations are correlated, first, an easy to optimize upper bound is derived for sufficiently small correlations and the power allocation scheme is derived accordingly. Next, an evolutionary computation technique based on Particle Swarm Optimization is developed to find the optimal power allocation for arbitrary correlations. The optimal power scheduling scheme suggests that the sensors with poor observation quality and bad channels should be inactive to save the total power expenditure of the system. It is shown that the probability of fusion error performance based on the optimal power allocation scheme outperforms the uniform power allocation scheme especially when either the number of sensors is large or the local observation quality is good.

Index terms: Decentralized detection, correlated observations, data fusion, optimal power scheduling, particle swarm optimization.

I. INTRODUCTION

Wireless sensor networks (WSNs) are ideal for a wide variety of applications such as environmental monitoring, smart factory instrumentation, intelligent transportation and remote surveillance [1]–[3] due to their low cost and ease of operation. Decentralized detection is becoming more attractive in many WSN applications over the centralized approach since it drastically reduces communication resource requirements. In decentralized detection, each node in the network sends a summary of its observation to the fusion center in contrast to that in centralized detection [4], [5]. The local processing at distributed nodes can be a form of lossy compression or simple relaying. The fusion center makes use of partially processed data from local nodes to make the final decision. Since only a summary of observations is transmitted, decentralized detection has the potential to extend the lifetime of the sensor network, at the expense of some performance reduction.

The fusion performance of a decentralized detection system in a low power WSN is limited by resource constraints, namely power and bandwidth. In a typical WSN, communication and

computing capabilities of sensor nodes can be limited due to various design considerations such as small battery and available bandwidth. For example, it may be impractical to replace or recharge the batteries due to cost and operating environment considerations. Therefore, power management is considered to be a core issue in designing a WSN.

A. Related work

The problem of distributed detection and fusion under resource constraints has been considered by many authors [6]–[15]. These have studied the fusion performance under given power or bandwidth constraints on the network. For example, in [7], it was shown that when the network is subjected to a joint power constraint, having identical sensor nodes (i.e. all nodes using the same transmission scheme) is asymptotically optimal for binary decentralized detection. When the whole system is subjected to a total average power constraint, [10] showed that it is better to combine as many not-so-good local decisions as possible rather than relying on a few very good local decisions in the case of deterministic signal detection. Efficient node power allocation to achieve a required fusion performance has been considered by [1], [16]–[18]. The optimal power scheduling for distributed detection in a WSN also has recently been considered in [16], where an optimal power allocation scheme was developed with respect to the so-called J -divergence performance index. They have shown that the optimal power allocation is determined by the qualities of the local decisions of the sensors and the communication channels where the channels are assumed to be imperfect. The optimal power scheduling scheme for the related problem of decentralized estimation subject to a required target mean squared error at the fusion center (with independent observations) was considered in [1] assuming quantized decisions at local nodes. It was shown that the optimal power scheduling scheme decreases the quantization resolutions of the nodes correspond to bad channels or poor observation qualities. In [18], the same problem was addressed with amplify-and-forward processing at local nodes. It was shown in [18] that such an analog forwarding scheme is optimal in the single sensor case by Shannon’s separation principle. For the case of multiple sensors the optimal power scheduling was derived in [18] via convex optimization. It was also shown that optimal power scheduling scheme improves the mean squared error performance by a large margin compared to that achieved by an uniform power allocation scheme. The minimum energy

¹Department of Electrical and Computer Engineering, University of New Mexico, MSC01 1100, Albuquerque NM 87131.

decentralized estimation with correlated data was addressed in [17]. They exploited knowledge of the noise covariance matrix to select quantization levels at nodes that minimized the power, while meeting a target mean-squared error.

B. Summary of Results

We address the problem of power allocation for detection of a constant signal in a sensor network with independent as well as correlated observations while keeping the fusion error probability under a required threshold, with amplify-and-forward local processing at sensor nodes. We consider a WSN consisting of a fusion center and an n number of spatially separated sensors. The distributed nodes collect observations corrupted by Gaussian noise and perform amplify-and-forward local processing to compute a local message that is transmitted to the fusion center. The wireless channel between the nodes and the fusion center is assumed to undergo fading.

First we consider the case where the local observations are independent and derive the optimal power allocation scheme analytically. For the correlated observations, we derive the exact, as well as an upper bound to the fusion error probability that is easy to optimize when the local observation correlations are sufficiently small and find analytically the optimal power allocation scheme. Next, we use Particle Swarm Optimization (PSO), which is a computation technique based on the movement and intelligence of particles of a swarm, to numerically find the optimal power allocation scheme for arbitrarily correlated Gaussian observations.

As we will show, according to the optimal power allocation scheme that conserves total power spent by the whole WSN, the nodes with poor observation quality and/or bad channels are turned off while the other nodes transmit locally processed data to the fusion center. We will show that when local signal-to-noise ratio (SNR) is large, only a small number of nodes needs to be active to achieve the required fusion error performance while a relatively large number of nodes should be active when local SNR is small. We also observe that the optimal power allocation scheme has considerably better performance over the uniform power allocation scheme specifically when the number of nodes in the network is large. It is also verified that the results obtained via PSO-based numerical method closely match with analytical results under the same network conditions when the observations are i.i.d.. We also investigate the performance of the analytical power allocation scheme derived under the conditionally independent assumption in a network with correlated observations. It can be seen that for large correlations, the conditionally independent assumption degrades the energy performance significantly compared to the performance of PSO-based method for correlated observations.

The remainder of this paper is organized as follows: Section II formulates the fusion problem. In Section III the optimal fusion performance is analyzed. The proposed optimal power allocation schemes are discussed in Section IV. Section V gives the performance results and finally concluding remarks are given in Section VI.

II. DATA FUSION PROBLEM FORMULATION

We consider a binary hypothesis testing problem in an n -node distributed wireless sensor network. The k -th sensor observation under the two hypotheses is given by,

$$\begin{aligned} H_0 : z_k &= v_k; \quad k = 1, 2, \dots, n \\ H_1 : z_k &= x_k + v_k; \quad k = 1, 2, \dots, n \end{aligned} \quad (1)$$

where v_k is zero-mean Gaussian observation noise with variance σ_v^2 and x_k is the signal to be detected. In vector notation, (1) becomes $\mathbf{z} = \mathbf{x} + \mathbf{v}$ where \mathbf{v} is a zero mean Gaussian n -vector of noise samples with covariance matrix Σ_v . In general we consider spatially correlated observations, so that Σ_v is not necessarily diagonal. We consider the detection of a constant signal so that $x_k = m$ for all k (the results hold straightforwardly for any deterministic signal). Let us define the local signal-to-noise ratio $\gamma_0 = \frac{m^2}{\sigma_v^2}$. The prior probabilities of the two hypotheses H_1 and H_0 are denoted by $P(H_1) = \pi_1$ and $P(H_0) = \pi_0$, respectively.

In this paper we assume that amplify-and-forward local processing is used, according to which each node retransmits an amplified version of its own observation to the fusion center. Hence the local decisions sent to the fusion center are, $u_k = g_k z_k$; $k = 1, 2, \dots, n$, where g_k is the amplifier gain at node k . The received signal r_k at the fusion center under each hypothesis is given by; $H_0 : r_k = n_k$; and $H_1 : r_k = h_k g_k x_k + n_k$; for $k = 1, 2, \dots, n$, where $n_k = h_k g_k v_k + w_k$, h_k is the channel fading coefficient and w_k is the receiver noise that is assumed to be i.i.d. with mean zero and variance σ_w^2 . Defining $\mathbf{r} = [r_1, \dots, r_n]^T$, we have, $\mathbf{r} = \mathbf{A}\mathbf{x} + \mathbf{n}$ where $\mathbf{A} = \text{diag}(h_1 g_1, h_2 g_2, \dots, h_n g_n)$. The detection problem at the fusion center can then be formulated as,

$$\begin{aligned} H_0 &: \mathbf{r} \sim p_0(\mathbf{r}) = \mathcal{N}(0, \Sigma_n) \\ H_1 &: \mathbf{r} \sim p_1(\mathbf{r}) = \mathcal{N}(\mathbf{A}\mathbf{m}, \Sigma_n) \end{aligned} \quad (2)$$

where $\Sigma_n = \mathbf{A}\Sigma_v\mathbf{A} + \sigma_w^2\mathbf{I}$, $\mathbf{m} = m\mathbf{e}$, \mathbf{e} is the n -vector of all ones and \mathbf{I} is the $n \times n$ identity matrix. The log-likelihood ratio (LLR) for the detection problem (2) can be written as, $T(\mathbf{r}) = m\mathbf{e}^T\mathbf{A}\Sigma_n^{-1}\mathbf{r} - \frac{1}{2}m^2\mathbf{e}^T\mathbf{A}\Sigma_n^{-1}\mathbf{A}\mathbf{e}$. It is well known that optimal fusion tests should be threshold tests on the above LLR. Thus the optimal Bayesian decision rule at the fusion center is given by,

$$\delta(\mathbf{r}) = \begin{cases} 1 & \text{if } T(\mathbf{r}) \geq \ln \tau \\ 0 & \text{if } T(\mathbf{r}) < \ln \tau \end{cases} \quad (3)$$

where τ is the threshold given by $\tau = \frac{\pi_1}{\pi_0}$ (assuming minimum probability of error Bayesian fusion).

III. ANALYSIS OF OPTIMAL FUSION PERFORMANCE

Note that,

$$\begin{aligned} H_0 : T(\mathbf{r}) &\sim \mathcal{N}\left(-\frac{1}{2}m^2\mathbf{e}^T\mathbf{A}\Sigma_n^{-1}\mathbf{A}\mathbf{e}, m^2\mathbf{e}^T\mathbf{A}\Sigma_n^{-1}\mathbf{A}\mathbf{e}\right) \\ H_1 : T(\mathbf{r}) &\sim \mathcal{N}\left(\frac{1}{2}m^2\mathbf{e}^T\mathbf{A}\Sigma_n^{-1}\mathbf{A}\mathbf{e}, m^2\mathbf{e}^T\mathbf{A}\Sigma_n^{-1}\mathbf{A}\mathbf{e}\right). \end{aligned} \quad (4)$$

The false alarm probability of the optimal detector at the fusion center is $P_f = P(T(\mathbf{r}) > \ln \tau | H_0) =$

$Q\left(\frac{\ln\tau + \frac{1}{2}m^2\mathbf{e}^T\mathbf{A}\Sigma_n^{-1}\mathbf{A}\mathbf{e}}{m\sqrt{\mathbf{e}^T\mathbf{A}\Sigma_n^{-1}\mathbf{A}\mathbf{e}}}\right)$ where Q -function is defined by $Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-\frac{\zeta^2}{2}} d\zeta$. Similarly, the probability of detection is given by $P_D = P(T(\mathbf{r}) > \ln\tau | H_1) = Q\left(\frac{\ln\tau - \frac{1}{2}m^2\mathbf{e}^T\mathbf{A}\Sigma_n^{-1}\mathbf{A}\mathbf{e}}{m\sqrt{\mathbf{e}^T\mathbf{A}\Sigma_n^{-1}\mathbf{A}\mathbf{e}}}\right)$. Hence the probability of error at the fusion center for a Bayesian optimal detector is given by $P(E) = P_f\pi_0 + (1 - P_D)\pi_1 = Q\left(\frac{1}{2}\sqrt{m^2\mathbf{e}^T\mathbf{A}\Sigma_n^{-1}\mathbf{A}\mathbf{e}}\right)$ (5)

where the prior probabilities are assumed to be equal so that $\tau = 1$.

A. Independent Local Observations

When the node observations are uncorrelated the noise covariance matrix Σ_v is simply $\Sigma_v = \sigma_v^2\mathbf{I}$. Then the probability of fusion error in (5) is simplified to,

$$P(E) = Q\left(\frac{1}{2}m\sqrt{\sum_{k=1}^n \frac{h_k^2 g_k^2}{h_k^2 g_k^2 \sigma_v^2 + \sigma_w^2}}\right). \quad (6)$$

It is interesting to note that $\lim_{g_k^2 \rightarrow \infty, k=1, \dots, n} \sum_{k=1}^n \frac{h_k^2 g_k^2}{h_k^2 g_k^2 \sigma_v^2 + \sigma_w^2} = \frac{n}{\sigma_v^2}$ so that the probability of fusion error has a performance floor:

$$\lim_{g_k^2 \rightarrow \infty, k=1, \dots, n} P(E) \rightarrow Q\left(\frac{\sqrt{n}\gamma_0}{2}\right). \quad (7)$$

Therefore, for a fixed n the probability of fusion error is ultimately limited by the observation quality at local sensor nodes regardless of the quality of the wireless channel.

B. Correlated Observation Noise

It is not straightforward to evaluate Σ_n^{-1} in (5) analytically in closed form for a general Σ_v when the observations are correlated. In the following we consider a specific sensor network model and obtain an upper bound for $P(E)$ in (5) that is valid for small correlations. To that end let us assume a 1-D sensor network in which adjacent nodes are separated by an equal distance d and correlation between nodes i and j is proportional to $\rho_0^{|i-j|}$ where $|\rho_0| \leq 1$. Letting $\rho_0^d = \rho$, Σ_v can be written as

$$\Sigma_v = \sigma_v^2 \begin{bmatrix} 1 & \rho & \cdot & \cdot & \rho^{n-2} & \rho^{n-1} \\ \rho & 1 & \cdot & \cdot & \rho^{n-3} & \rho^{n-2} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \rho^{n-1} & \rho^{n-2} & \cdot & \cdot & \rho & 1 \end{bmatrix}. \quad (8)$$

Note that, when ρ is sufficiently small, we may approximate (8) by its tri-diagonal version by dropping second and higher order terms of ρ . Recall, from Bergstrom's inequality [19] that, for any two positive definite matrices \mathbf{P} and \mathbf{Q}

$$\mathbf{e}^T \mathbf{P}^{-1} \mathbf{e} \geq \frac{(\mathbf{e}^T (\mathbf{P} + \mathbf{Q})^{-1} \mathbf{e}) (\mathbf{e}^T \mathbf{Q}^{-1} \mathbf{e})}{\mathbf{e}^T \mathbf{Q}^{-1} \mathbf{e} - \mathbf{e}^T (\mathbf{P} + \mathbf{Q})^{-1} \mathbf{e}}. \quad (9)$$

Since $m^2 \mathbf{e}^T \mathbf{A} \Sigma_n^{-1} \mathbf{A} \mathbf{e} = m^2 \mathbf{e}^T (\Sigma_v + \sigma_w^2 \mathbf{A}^{-2})^{-1} \mathbf{e}$, let $\mathbf{P} = (\Sigma_v + \sigma_w^2 \mathbf{A}^{-2})$ and define the matrix \mathbf{Q} such that

$$\mathbf{Q} = \sigma_v^2 \begin{bmatrix} 1 & -\rho & \cdot & \cdot & -\rho^{n-2} & -\rho^{n-1} \\ -\rho & 1 & \cdot & \cdot & -\rho^{n-3} & -\rho^{n-2} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ -\rho^{n-1} & -\rho^{n-2} & \cdot & \cdot & -\rho & 1 \end{bmatrix}.$$

For small enough ρ it can be shown that $\mathbf{e}^T \mathbf{Q} \mathbf{e} > 0$. In fact, when Σ_v has the tri-diagonal structure (implying only the adjacent node observations are correlated), it can be shown that for any $|\rho| < \frac{n}{2(n-1)}$, we will have $\mathbf{e}^T \mathbf{Q} \mathbf{e} > 0$. In general, if Σ_v is as in (8), this will be true for small enough ρ . Note that while noise covariance matrix (8) is an idealization, it can be used in many applications, such as traffic monitoring or in industrial monitoring, where the sensors are approximately equally spaced. The tri-diagonal version of (8) is a reasonable approximation when the correlation coefficient ρ is small, since then the second and higher order terms of ρ in (8) are negligible. From (9) it can be shown that,

$$\mathbf{e}^T (\Sigma_v + \sigma_w^2 \mathbf{A}^{-2})^{-1} \mathbf{e} \geq \left(\frac{1}{\sum_{k=1}^n \frac{h_k^2 g_k^2}{2h_k^2 g_k^2 \sigma_v^2 + \sigma_w^2}} - \frac{1}{D} \right)^{-1}, \quad (10)$$

where $D = \mathbf{e}^T \mathbf{Q}^{-1} \mathbf{e}$. From (5) and (10), we then have the following upper bound for the fusion error probability when the observations are correlated and ρ is sufficiently small:

$$P(E) \leq Q\left(\frac{m}{2} \left(\frac{1}{\sum_{k=1}^n \frac{h_k^2 g_k^2}{2h_k^2 g_k^2 \sigma_v^2 + \sigma_w^2}} - \frac{1}{D} \right)^{-\frac{1}{2}}\right). \quad (11)$$

When $\rho = 0$ we have $D = n/\sigma_v^2$. Then

$$\lim_{g_k^2 \rightarrow \infty, k=1, \dots, n} \left(\frac{1}{\sum_{k=1}^n \frac{h_k^2 g_k^2}{2h_k^2 g_k^2 \sigma_v^2 + \sigma_w^2}} - \frac{1}{D} \right) = \frac{n}{\sigma_v^2}.$$

That is, the fusion error probability bound (11) also has a performance floor of $Q\left(\frac{\sqrt{n}\gamma_0}{2}\right)$ as in (7), when local amplifier gains are large. Thus both the exact fusion error probability and the proposed bound exhibit the same performance in the case of i.i.d. observations at least when the channel SNR quality is good.

IV. OPTIMAL POWER ALLOCATION

In the following, we first derive the optimal power allocation scheme that minimizes the total power spent by the whole sensor network subjected to a threshold on the fusion error probability when local observations are i.i.d.. Next, we propose a numerical method based on PSO to find the optimal power allocation when local observations are arbitrarily correlated. In this case, we also obtain an analytical optimal power allocation scheme that minimizes the fusion error probability bound in (11) subjected to a required threshold for sufficiently small ρ values. We show that according to these optimal schemes the nodes with poor observation quality and/or bad channels are inactivated to save the total power of the system.

In general, the power allocation problem can be formulated as,

$$\begin{aligned} & \min_{g_k \geq 0, k=1, \dots, n} \sum_{k=1}^n g_k^2 \text{ such that} \\ & P(E) = Q\left(\frac{1}{2}\sqrt{m^2\mathbf{e}^T\mathbf{A}\Sigma_n^{-1}\mathbf{A}\mathbf{e}}\right) \leq \epsilon \text{ and} \\ & g_k \geq 0; \quad k = 1, 2, \dots, n, \end{aligned} \quad (12)$$

where ϵ is the required fusion error probability at the fusion center.

A. Optimal Power Allocation when Observations are i.i.d.

When local observations are i.i.d., the fusion error probability is given by (6). Hence, the first inequality in (12) becomes $\beta \leq \sqrt{\sum_{k=1}^n \frac{h_k^2 g_k^2}{h_k^2 g_k^2 \sigma_v^2 + \sigma_w^2}}$ where we have defined $\beta = \frac{2}{m} Q^{-1}(\epsilon)$. Since β is positive, the optimal power allocation problem can thus be rewritten as,

$$\begin{aligned} & \min_{g_k \geq 0, k=1, \dots, n} \sum_{k=1}^n g_k^2, \text{ such that} \\ & \beta^2 - \sum_{k=1}^n \frac{h_k^2 g_k^2}{h_k^2 g_k^2 \sigma_v^2 + \sigma_w^2} \leq 0, \text{ and} \\ & g_k \geq 0 \text{ for } k = 1, 2, \dots, n, \end{aligned} \quad (13)$$

The Lagrangian for the above problem is

$$\begin{aligned} G(L, \lambda_0, \mu_k) = & \sum_{k=1}^n g_k^2 + \lambda_0 \left[\beta^2 - \sum_{k=1}^n \frac{h_k^2 g_k^2}{h_k^2 g_k^2 \sigma_v^2 + \sigma_w^2} \right] \\ & + \sum_{k=1}^n \mu_k (-g_k) \end{aligned} \quad (14)$$

where $\lambda_0 \geq 0$ and $\mu_k \geq 0$ for $k = 1, 2, \dots, n$. Verifying KKT conditions, it can be shown that the optimal solution for (13) is given by,

$$g_k^2 = \begin{cases} \frac{\sigma_w^2}{h_k^2 \sigma_v^2} \left[\frac{h_k \sum_{j=1}^{K_1} \frac{1}{h_j}}{(K_1 - \beta^2 \sigma_v^2)} - 1 \right] & ; \text{ if } k < K_1 \text{ and } n > \beta^2 \sigma_v^2 \\ 0 & ; \text{ if } k > K_1 \text{ and } n > \beta^2 \sigma_v^2 \\ \text{infeasible} & ; \text{ if } n < \beta^2 \sigma_v^2 \end{cases} \quad (15)$$

where K_1 is found such that $f(K_1) < 1$ and $f(K_1+1) \geq 1$ for $1 \leq K_1 \leq n$ assuming, without loss of generality, $h_1 \geq h_2 \geq \dots \geq h_n$ where $f(k) = \frac{(k - \beta^2 \sigma_v^2)}{h_k \sum_{j=1}^k \frac{1}{h_j}}$, $1 \leq k \leq n$. The proof of the uniqueness of such a K_1 and the global optimality of the solution (15) for the optimization problem (13) are shown in the Appendix.

Since there is a feasible optimal solution only when $n > \beta^2 \sigma_v^2$, i.e. $\gamma_0 > \frac{4}{n} (Q^{-1}(P_e))^2$, this implies that we can not achieve probability of errors below $Q\left(\frac{\sqrt{n\gamma_0}}{2}\right)$. Note that this is consistent with (7). The optimal solution for g_k^2 when $f(k) - 1 < 0$ and $n > \beta^2 \sigma_v^2$ can be rewritten as $g_k^2 = \frac{\sigma_w^2}{h_k^2 \sigma_v^2} \left(\frac{\sqrt{\lambda_0} h_k}{\sigma_w} - 1 \right)$, where $\sqrt{\lambda_0} = \frac{\sigma_w \sum_{k=1}^{K_1} \frac{1}{h_k}}{K_1 - \beta^2 \sigma_v^2}$. Hence, once the fusion center calculates λ_0 and broadcasts it, each node can determine its power distributively using λ_0 as side information.

B. Optimal Power Allocation when Observations are Correlated via Particle Swarm Optimization

Since it is not possible to find a closed form optimal solution for g_k 's in (12) when observations are correlated, in the following we solve it numerically. For that, we develop a stochastic evolutionary computation technique based on PSO [20]–[22]. Since PSO is not directly applicable for constrained optimization problems, we first transform our constrained optimization problem in (12) into an unconstrained optimization problem using the exterior penalty function approach [23], [24].

1) *Penalty function approach for constrained optimization:* Suppose that the optimization problem of interest is

$$\min f(\mathbf{X}) \text{ such that } h_j(\mathbf{X}) \leq 0; \quad j = 1, \dots, m. \quad (16)$$

Then the exterior penalty function for the above minimization problem can be formulated as [23], [24],

$$\phi(\mathbf{X}, r_k) = f(\mathbf{X}) + r_k \sum_{j=1}^m (\max[0, h_j(\mathbf{X})])^q, \quad (17)$$

where r_k is a positive penalty parameter and q is a non-negative constant. Usually, the value of q is chosen to be 2 in practice [23]. The exterior penalty function algorithm that finds the optimal solution for the problem (16) can be stated as below: (Note that subscript of \mathbf{X} denotes the index corresponding to penalty parameter while the superscript of \mathbf{X} denotes the iteration number of the minimization algorithm for a particular penalty parameter).

- **step 1:** Set $k = 1$. Start from any initial solution \mathbf{X}_k^1 and a suitable value of $r_k = r_1$.
- **step 2:** Find the vector \mathbf{X}_k^* that minimizes the function given in (17).
- **step 3:** Test whether the point \mathbf{X}_k^* satisfies all the constraints. If \mathbf{X}_k^* is feasible, it is the desired optimum and hence terminate the procedure. Otherwise go to next (15) step.
- **step 4:** Choose the next value of the penalty parameter according to the relation $\frac{r_{k+1}}{r_k} = c$ where c is a constant greater than one and set $\mathbf{X}_{k+1}^1 = \mathbf{X}_k^*$ and $k = k + 1$. Go to step 2.

Assuming that $f(\mathbf{X})$ and $h_j(\mathbf{X})$, $j = 1, 2, \dots, m$ are continuous and that an optimal solution exists for (16), the unconstrained minima \mathbf{X}_k^* of (17) converge to the optimal solution of the original problem $f(\mathbf{X})$ as $k \rightarrow \infty$ and $r_k \rightarrow \infty$ [23]. In order to ensure the existence of a global minimum of $\phi(\mathbf{X}, r_k)$ in (17) for every positive value r_k , $\phi(\cdot)$ has to be a strictly convex function of \mathbf{X} . The following theorem, the proof of which can be found in [23], gives the sufficient conditions for $\phi(\mathbf{X}, r_k)$ to be strictly convex:

Theorem 1

If $f(\mathbf{X})$ and $h_j(\mathbf{X})$, for $j = 1, 2, \dots, m$ are convex and at least $f(\mathbf{X})$ or either one of $\{h_j(\mathbf{X})\}_{j=1}^m$ is strictly convex, then the function $\phi(\mathbf{X}, r_k)$ defined by (17) will be a strictly convex function of \mathbf{X} .

2) *Particle swarm optimization:* To evaluate optimal \mathbf{X}_k^* for each penalty parameter r_k as required in the step 2 above, we use the particle swarm optimization technique. A brief overview of the particle swarm language is given in Table I and more details can be found in [22].

In the following we give the algorithmic steps needed to implement the PSO for a given problem:

- (I). Define the solution space and the fitness function: Pick the parameters that need to be optimized and give them a reasonable range in which to search for the optimal solution. The fitness function should exhibit a functional dependence that is relative to the importance of each characteristic being optimized.

TABLE I
PSO TERMINOLOGY

Particle/Agent	A Single individual in the swarm
Location/Position	An agent's n -dimensional coordinates which represent a solution to the problem
Swarm	The entire collection of agents
Fitness	A single number representing the goodness of a given solution
pbest	The location in parameter space of the best fitness returned for a specific agent
gbest	The location in parameter space of the best fitness returned in the entire swarm
V_{max}	The maximum allowed velocity in a given direction

We denote the swarm size by M . For each k in (17), we perform a PSO optimization algorithm to find \mathbf{X}_k^* . For each k , let us define, $\mathbf{X}_{k,m}$ as the position vector of the m -th particle; $\mathbf{P}_{k,m}$ as the *pbest* of the m -th particle; $\mathbf{P}_{k,gbest}$ as the *gbest* of the swarm; $\phi(\mathbf{X}_{k,m}, r_k)$ as the fitness value corresponding to the location $\mathbf{X}_{k,m}$ of the m -th particle; $\phi(\mathbf{P}_{k,m}, r_k)$ as the fitness value corresponding to the *pbest* $\mathbf{P}_{k,m}$ of the m -th particle; $\phi(\mathbf{P}_{k,gbest}, r_k)$ as the fitness value corresponding to the *gbest* of the swarm and $\mathbf{V}_{k,m}$ as velocity of the m -th particle. The maximum number of iterations of PSO for each k is set to S .

(II). If $k = 1$ (i.e. the penalty parameter is r_1) initialize the swarm locations randomly. Otherwise set the initial positions of each particle to be the best *pbest* values for $k = k - 1$.

- Initializing position: For $k = 1$ and for each particle m , $m = 1, \dots, M$, $\mathbf{X}_{k,m}^1$ is chosen randomly. If $k > 1$, then $\mathbf{X}_{k,m}^1 = \mathbf{P}_{k-1,m}^S$ where $\mathbf{P}_{k-1,m}^S$ is the *pbest* of the m -th particle for $k = k - 1$ at the S -th iteration of PSO.
- Initializing *pbest*: Since its initial position is the only location encountered by each particle at the run's start, this position becomes each particle's initial *pbest*. i.e. $\mathbf{P}_{k,m}^1 = \mathbf{X}_{k,m}^1$.
- Initializing *gbest*: The first *gbest* is selected as the initial *pbest* which gives the best fitness value: $\mathbf{P}_{k,gbest}^1 = \mathbf{P}_{k,m_1}^1$ where $m_1 = \arg \min_{1 \leq m \leq M} \{\phi(\mathbf{P}_{k,m}^1, r_k)\}$.
- Initializing velocities: Initialize $\mathbf{V}_{k,m}^1$ as zeros for each particle m .

(III). Fly the particles through the solution space:

Each particle is then moved through the solution space. The following steps are performed on each particle individually.

- Evaluate the particle's fitness value and compare it with that of *pbest* and *gbest*. For each particle, if its fitness value is better than that of the respective *pbest* for that particle or the global *gbest*, then the appropriate locations are replaced with the current location. i.e., in the s -th iteration of the PSO, for each particle m , for $m = 1, \dots, M$, if $\phi(\mathbf{X}_{k,m}^s, r_k) < \phi(\mathbf{P}_{k,m}^s, r_k)$ then set $\mathbf{P}_{k,m}^s = \mathbf{X}_{k,m}^s$. Set $\mathbf{P}_{k,gbest}^s = \mathbf{P}_{k,m_s}^s$ where $m_s = \arg \min_{1 \leq m \leq M} \{\phi(\mathbf{P}_{k,m}^s, r_k)\}$.
- Update the particle's velocity: The velocity of the particle is changed according to the relative loca-

tions of *pbest* and *gbest*. The particles are "accelerated" in the directions of the locations of best fitness value according to the following equation [22], [25]:

$$\mathbf{V}_{k,m}^{s+1} = \mathcal{X}\{(w\mathbf{V}_{k,m}^s + c_1\text{rand}())(\mathbf{P}_{k,m}^s - \mathbf{X}_{k,m}^s) + c_2\text{rand}())(\mathbf{P}_{k,gbest}^s - \mathbf{X}_{k,m}^s)\}, \quad (18)$$

where \mathcal{X} is the constriction factor that is used to control and constrict velocities; w is the inertia weight that determines to what extent the particle remains along its original course unaffected by the pull of *pbest* and *gbest*, c_1 and c_2 are positive constants that determine the relative "pull" of *pbest* and *gbest* (in fact c_1 determines how much the particle is influenced by the memory of its best location and c_2 determines how much the particle is influenced by the rest of the swarm) and the random number function $\text{rand}()$ returns a number between 0 and 1.

- Move the particle: Once the velocity has been determined as in (18), move the particle to its next location as $\mathbf{X}_{k,m}^{s+1} = \mathbf{X}_{k,m}^s + \Delta t \mathbf{V}_{k,m}^{s+1}$. The velocity is applied for a given time step Δt .

(IV). Repetition: After the velocity and the position are updated the process is repeated starting at step (III) until the termination criteria are met. The termination criteria can be a user-defined maximum iteration number or a target fitness termination condition. In the latter case, the PSO is run for the user-defined number of iterations, but at any time if a solution is found that is greater than or equal to the target fitness value, then PSO is stopped at that point. In our work we set the maximum iteration number (S) for PSO as defined before. Once the termination criteria are met, the optimal solution \mathbf{X}_k^* for the unconstrained minimization problem (17) for given k is $\mathbf{P}_{k,gbest}^S$.

To solve the optimization problem in (12) when the observations are correlated we define the exterior penalty function as,

$$\phi(\mathbf{g}, r_k) = f(\mathbf{g}) + r_k \{(\max[h_1(\mathbf{g}), 0])^2 + \sum_{j=2}^m (\max[h_j(\mathbf{g}), 0])^2\}, \quad (19)$$

where $f(\mathbf{g}) = \sum_{i=1}^n g_i^2$, $h_1(\mathbf{g}) = \beta^2 - \mathbf{e}^T \mathbf{A} \Sigma_n^{-1} \mathbf{A} \mathbf{e}$ and $h_{i+1}(\mathbf{g}) = -g_i$ for $i = 1, 2, \dots, n$ and $\mathbf{g} = [g_1, \dots, g_n]^T$. Here we have $m = n + 1$. When the observation noise is i.i.d, it can be shown that $\phi(\mathbf{g}, r_k)$ is a strictly convex function

for $g_i \geq \frac{\sigma_w^2}{3h_i^2\sigma_v^2}$ for $i = 1, 2, \dots, n$ and also it can be seen that when h_i 's are small enough the convexity of $\phi(\mathbf{g}, r_k)$ holds for $g_i \geq 0$, ensuring a global minimum for $\phi(\mathbf{g}, r_k)$. We will assume that $\phi(\mathbf{g}, r_k)$ has a global minimum for each r_k even when the observation noise is correlated under above conditions. Assuming that an optimal solution for (12) exists and since $f(\mathbf{g})$ and $h_j(\mathbf{g})$ for $j = 1, 2, \dots, m$, are continuous, as $k \rightarrow \infty$ and $r_k \rightarrow \infty$ the unconstrained minima \mathbf{g}_k^* of $\phi(\mathbf{g}, r_k)$ converge to the optimal solution of the original problem (12).

3) *Selection of parameter values for PSO*: The parameter set to be optimized is $\mathbf{g} = [g_1, \dots, g_n]^T$ and we define the solution space as $[0, \infty)$ for each parameter. To run the PSO the population size was selected as 30 which has been shown to be sufficient for many engineering problems [26]. Various values for inertia weight w have been suggested in the literature. Since larger weights tend to encourage global exploration and conversely smaller initial weights encourage local exploitations, [27] has suggested to vary w linearly from 0.9 to 0.4 over the course of the run. On the other hand, [25] suggested to gradually decrease w from 1.2 towards 0.1 over the run of a PSO. We allowed w to vary between 0.9 to 0.4 linearly since it gave a fast convergence over 100 iterations. c_1 and c_2 were both set to 2.0 [22], [25]. The constriction factor \mathcal{X} was set to 0.73 [25].

One of the main advantage of the PSO based method is that once the algorithm parameters are chosen as above, the algorithm seems to work over a large range of variations in problem parameters such as fading coefficients, n, ρ and ϵ . On the other hand, the choice of step size and the initial values for a conventional method such as Newton's was observed to depend heavily on the problem parameters. The designer has to change the step sizes and the initial values every time when the system parameters change. This becomes especially problematic since fading coefficients are random. Hence, although once proper choices have been made, the Newton's and the proposed PSO-based methods show almost similar convergence properties, the PSO based method seems much easier to use.

C. Power Allocation based on the Fusion Error Probability Bound

When observations are correlated we may use the bound (11) to obtain an approximate analytical solution to the power allocation problem via

$$\begin{aligned} & \min_{g_k \geq 0, k=1, \dots, n} \sum_{k=1}^n g_k^2 \text{ such that} \\ & q - \sum_{k=1}^n \frac{h_k^2 g_k^2}{2h_k^2 g_k^2 \sigma_v^2 + \sigma_w^2} \leq 0 \quad \text{and} \\ & g_k \geq 0; \quad k = 1, 2, \dots, n, \end{aligned} \quad (20)$$

where $q = (\frac{1}{\beta^2} + \frac{1}{D})^{-1}$ and, as before, $\beta = \frac{2Q^{-1}(\epsilon)}{m}$ (Note that, $q > 0$ since $D > 0$). We can use the same method as in Section IV-A to find the optimal solution for (20). Defining a function $\tilde{f}(k) = \frac{(k-2\sigma_v^2 q)}{h_k \sum_{j=1}^k \frac{1}{h_j}}$ and assuming again, $h_1 \geq h_2 \geq \dots \geq h_n$ it can be shown that (due to space limitations we avoid details but the steps are similar to that in Section IV-A), we can find

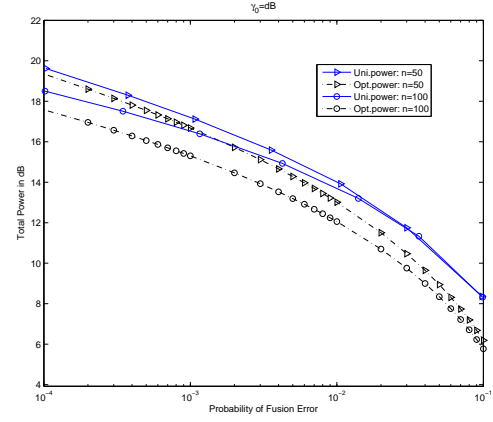


Fig. 1. Total power Vs. probability of fusion error for independent observations

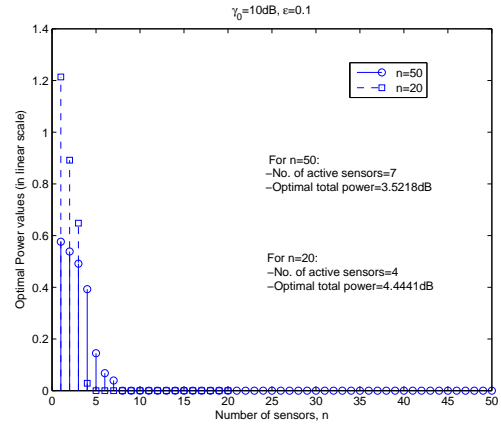


Fig. 2. Optimal power values of sensor nodes Vs. number of sensors for $n = 20$ and $n = 50$ when $\epsilon = 0.1$ and $\gamma_0 = 10dB$

a unique L_1 such that $\tilde{f}(L_1) < 1$ and $\tilde{f}(L_1 + 1) \geq 1$ for $1 \leq L_1 \leq n$. Then the solution to the problem (20) is given by,

$$g_k^2 = \begin{cases} \frac{\sigma_w^2}{2h_k^2\sigma_v^2} \left[\frac{h_k \sum_{j=1}^{L_1} \frac{1}{h_j}}{(L_1 - 2\sigma_v^2 q)} - 1 \right]; & \text{if } k < L_1 \text{ \& } n > 2\sigma_v^2 q \\ 0; & \text{if } k > L_1 \text{ \& } n > 2\sigma_v^2 q \\ \text{infeasible;} & \text{if } n < 2\sigma_v^2 q \end{cases} \quad (21)$$

Note from (21) that to achieve the required fusion error probability at the fusion center the total number of active sensors should be greater than $2\sigma_v^2 q$ in the optimal solution.

V. PERFORMANCE RESULTS

In this section we illustrate performance gains possible with the derived optimal power allocation scheme. We assume that fading coefficients h_k 's of the channel between sensors and the fusion center are Rayleigh distributed with a unit mean.

The results on Figs. 1 to 4 correspond to the optimal power allocation for i.i.d. observations. When observations are i.i.d. the optimal total power is given by $P_{Opt.} = \sum_{k=1}^{K_1} g_k^2$ where g_k^2 's are given in (15). The performance of the optimal scheme is compared with that of the uniform power allocation scheme.

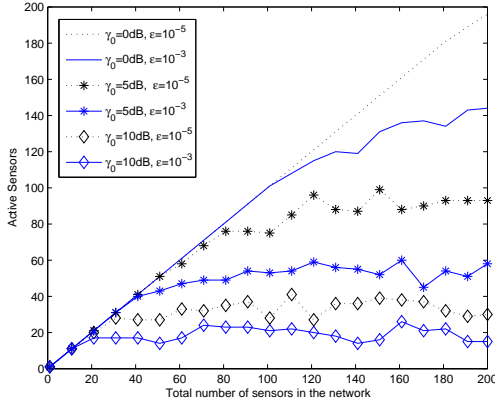


Fig. 3. Number of active sensors for independent observations.

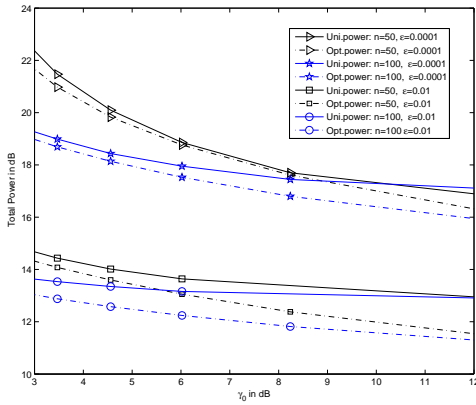


Fig. 4. Total power Vs. local SNR for independent observations

Figure 1 shows the total network power versus fusion error probability for different values of n . It can be seen that when the number of sensors is increased then the energy saving due to proposed optimal scheme is more significant compared to uniform power allocation. This is because it is more likely that there will be more channels with good channel fading coefficients. By using those channels the network can spend a smaller total power, while still ensuring the required performance at the fusion center. The power allocation to meet the same performance level with different n is shown in Fig. 2. From Fig. 1 it can also be seen that when the required fusion error probability is not significantly low, the gain of the optimal power allocation scheme over the uniform power allocation scheme is high.

The number of active sensors versus total sensors in the network for $\epsilon = 10^{-3}$ and $\epsilon = 10^{-5}$ with different γ_0 values is shown in Fig. 3. To achieve a given fusion error probability, it can be seen that only a small number of active sensors is needed when the local SNR is high. Fig. 3 also shows that a relatively large number of active sensors are needed to achieve lower fusion error probabilities compared to that of higher fusion error probabilities. This explains the high performance gain achieved at relatively higher fusion error probabilities as shown in Fig. 1.

In Figure 4 the total power versus the observation SNR γ_0 is shown for $n = 50$ and $n = 100$ parameterized by different fusion error probabilities. It can be seen that when the local SNR is high it is enough to turn on a relatively smaller number of nodes to achieve the same performance, thus decreasing the total system power. Also it is observed that when γ_0 is fixed, the fusion error performance can be improved by having a large number of nodes in the network.

In Fig. 5 we have considered the fusion performance with correlated observations based on the fusion error probability bound (11). The results are obtained assuming the observation noise covariance matrix has the tri-diagonal structure of (8). It can be seen that the optimal power allocation scheme under the fusion error probability bound performs significantly better than the uniform power allocation scheme based on either the bound (11) or the exact fusion error probability (5).

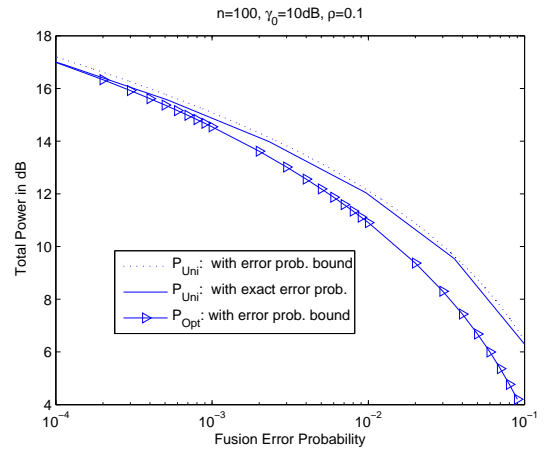
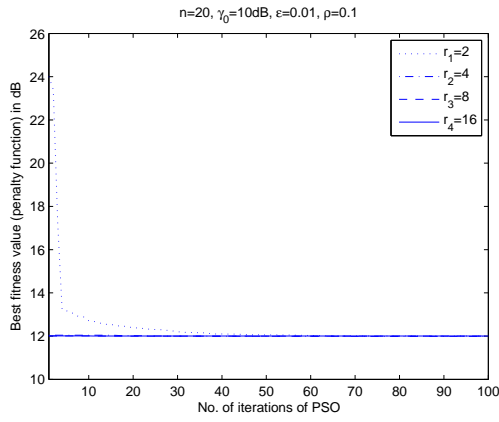


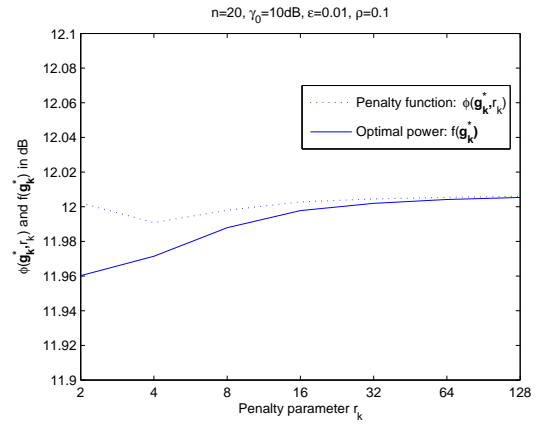
Fig. 5. Total power and the fusion error probability bound for correlated observations; $\gamma_0 = 5dB$, $n = 100$ and $\rho = 0.1$

Next we consider the performance results based on the constrained-PSO algorithm. Note that we employed the PSO-based method for each penalty parameter r_k of the unconstrained optimization problem (19) until $\phi(\mathbf{g}_k^*, r_k) \rightarrow f(\mathbf{g}_k^*)$ where $\mathbf{g}_k^* = \text{argmin}_{\mathbf{g}_k} \phi(\mathbf{g}_k, r_k)$. For a given r_k the convergence of PSO algorithm is shown in Fig. 6(a). The starting penalty parameter r_1 was set to 2, and was increased in such a way that $\frac{r_{k+1}}{r_k} = 2$. It was observed that for each r_k the PSO algorithm converges rapidly. The convergence of unconstrained minimum of $\phi(\mathbf{g}, r_k)$ to the constrained minimum of $f(\mathbf{g})$ is shown in Fig. 6(b) in which the error between the penalty function and the objective function at the convergent point is 0.0023 after 7 iterations of r_k . That is, with a relatively smaller number of iterations, the unconstrained minimum of the penalty function $\phi(\mathbf{g}, r_k)$ approaches to that of the objective function $f(\mathbf{g})$.

The comparison of \mathbf{g}^* obtained numerically (via PSO) and analytically under the same network conditions are shown in first two rows of the Table II for 10 nodes when the observations are i.i.d.. It can be seen that the numerical results closely match with the analytical solution. The third row of Table II shows the optimal \mathbf{g}^* obtained numerically when $\rho = 0.1$, $n = 10$, $\gamma_0 = 10dB$ and $\epsilon = 0.01$. It shows that

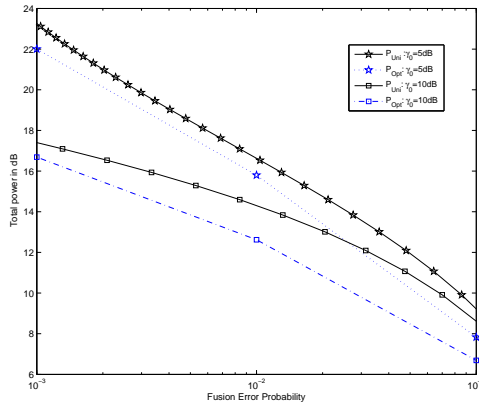


(a)

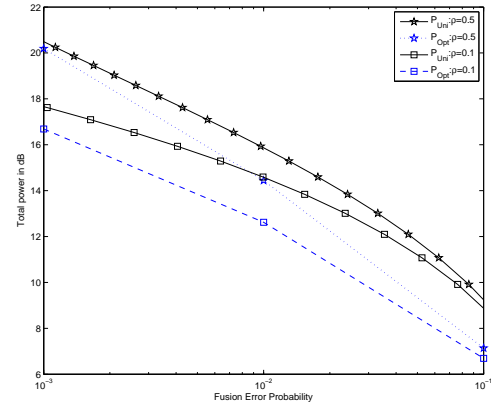


(b)

Fig. 6. The Convergence of exterior penalty function based PSO: Fusion error probability = 0.01 (a). Best fitness returned for PSO iterations for a given penalty parameter. (b). Convergence of penalty function to the original optimization problem.



(a)



(b)

Fig. 7. PSO: Total power Vs. fusion error probability when observations are correlated. (a). $n = 20, \rho = 0.1$ (b). $n = 20, \gamma_0 = 10dB$

TABLE II

COMPARISON OF ANALYTICAL AND NUMERICAL RESULTS WHEN $\rho = 0$, $\gamma_0=10dB$, $\epsilon=0.01$, $N=10$

\mathbf{g}^* : Analytical ($\rho = 0$)	[1.6172, 1.5888, 1.5555, 1.4666, 1.4616, 1.4107, 1.1231, 0, 0, 0]
\mathbf{g}^* : Numerical ($\rho = 0$)	[1.6163, 1.5696, 1.5548, 1.5014, 1.4501, 1.4099, 1.1212, 0.0013, 0.0066, 0.0008]
\mathbf{g}^* : Numerical ($\rho = 0.1$)	[1.6717, 1.5867, 1.6112, 1.5034, 1.5285, 1.4758, 1.3381, 0.3366, 0.0062, 0.0005]

when the observations are correlated the optimal solution for (12) should turn off the sensors with poor channels similar to the analytical solution for i.i.d observations. But it is seen that then the sensors need more power when the observations are correlated for the same n , γ_0 and ϵ .

The dependance of the total network power (obtained via constrained-PSO) on the required fusion error probability when local observations are correlated is shown in Fig. 7 parameterized by ρ and γ_0 . Note that, the constrained-PSO method is applicable for any arbitrary observation noise correlation model. The results in Fig. 7 are based on the noise covariance matrix in (8). It can be seen that the fusion

performance characteristics with respect to n and γ_0 for the correlated observations are similar to that with the i.i.d. observations. Figure 7(b) shows that the network needs to spend more power when the correlation coefficient of the observations is high since then the new information added by each additional sensor decreases resulting in degraded fusion performance.

Figure 8 shows the results obtained from the constrained PSO algorithm for different noise covariance models. In the noise covariance matrix in model 1, the off-diagonal elements above the main diagonal (or below the main diagonal) are generated according to a uniform distribution on $[0,1]$. Model 2 refers to the noise covariance matrix Σ_v such that $(\Sigma_v)_{i,j} = \sigma_v^2 \rho$ for $i \neq j$ and $(\Sigma_v)_{i,j} = \sigma_v^2$ for $i = j$. Model 3 refers to (8) and Model 4 is its tri-diagonal version. $\rho = 0.1$ for models 2, 3 and 4. As observed earlier, for small ρ we may approximate model 3 by model 4. As in model 2, if the observation correlation is the same among all the sensors then the system needs more power to achieve the same performance compared to models 3 and 4 in which the correlations decrease as separation between sensors increases.

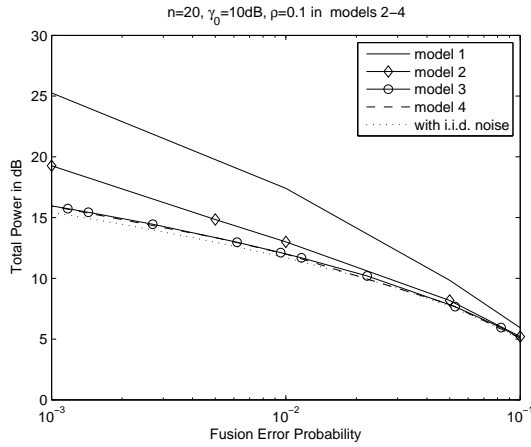


Fig. 8. PSO: Total power Vs. fusion error probability for different noise covariance models; in models 2 – 4, $\rho=0.1$

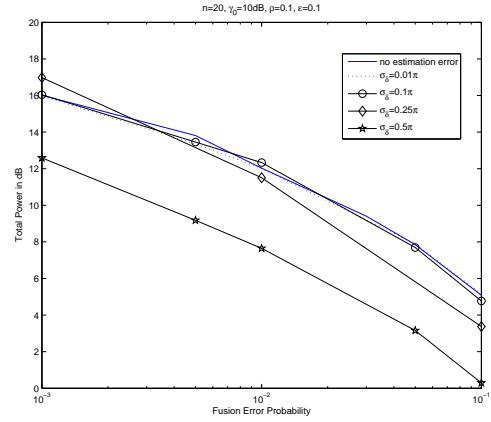


Fig. 10. PSO: Total power Vs. fusion error probability for with the estimation error of the fading coefficients at the fusion center. $n = 20$, $\gamma_0 = 10dB$, $\rho = 0.1$, $\epsilon = 0.1$

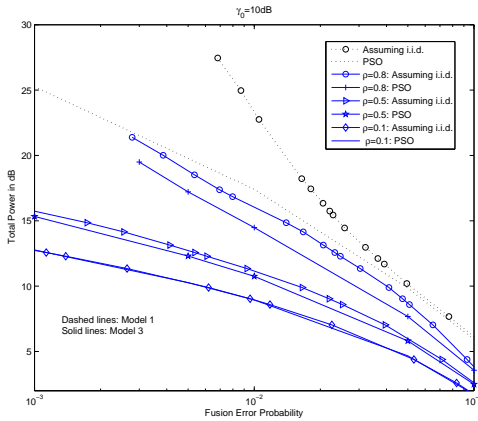


Fig. 9. PSO based results Vs. results obtained assuming i.i.d. noise for noise models 1 and 3, $n = 20$ and $\gamma_0 = 10dB$

It can also be seen that when the correlation coefficients are randomly selected between 0 and 1 as in model 1, the required power is significantly higher than that of other noise covariance models considered with small ρ values.

In Fig. 9 the results based on PSO method were compared to the results obtained assuming i.i.d. observations for different correlation profiles. Dashed line plots are corresponding to the model 1 and the solid line plots are corresponding to the noise model 3 with different ρ values as described above. With model 1, it can be seen from Fig. 9 that the assumption of conditional independence degrades the energy performance significantly. With the noise Model 3, which may be more realistic in practice, it can be seen from Fig. 9 that for large ρ values the PSO-based method has a better performance over the power allocation assuming i.i.d. observations (as given in Section IV.A). On the other hand, when ρ is small, the assumption of conditional independence might not lead to severe performance penalties, although actual observations are correlated. However, as observation correlations increase, the energy penalty becomes more significant.

So far we have assumed that transmitting nodes and the fusion center have the knowledge of exact channel fading

coefficients. In practice, the fusion center has only estimates \hat{h}_k 's of channel coefficients. Let us assume that $\hat{h}_k = h_k + \delta_k$ where estimation error $\delta_k \sim \mathcal{N}(0, \sigma_\delta^2)$ and σ_δ^2 is the estimation error variance. The affect of the estimation error on the optimal power allocation is shown in Fig. 10 with different σ_δ values. It can be seen that for small estimation errors the performance results do not change significantly.

VI. CONCLUSION

In this paper we addressed the problem of optimal power scheduling while meeting a target fusion error probability, for data fusion in a wireless sensor network with i.i.d. as well as correlated observations. When observations are i.i.d., we derived the optimal power allocation scheme analytically. For correlated observations, we derived an easy to optimize upper bound for the fusion error probability that is valid for sufficiently small data correlations. When the observations are arbitrary correlated, we also proposed an evolutionary computation technique based on PSO to evaluate the optimal power levels in the system. We showed that according to the optimal power allocation strategy the sensors with poor observation and/or channel quality must be turned off to save the total power spent by the system. Moreover, when the local observation quality is very good it is sufficient to collect data from only a small number of sensors out of the total available nodes in the network (keeping others turned off). We also noted that in the case of i.i.d. observations the derived optimal power scheduling scheme can be implemented distributively with only a small feedback from the fusion center. From numerical results based on constrained-PSO, we observed that the optimal power allocation scheme provides significant total energy savings over that of the uniform power allocation scheme especially when the number of nodes in the system is large or when the local observation quality is good. Also the PSO based method has significantly better performance compared to power allocation assuming observations are independent, especially for relatively high correlations.

APPENDIX

Uniqueness of K_1 : In the following, we show the existence of a unique K_1 , where $1 \leq K_1 \leq n$ such that $f(K_1) < 1$ and $f(K_1 + 1) \geq 1$ where $f(k) = \frac{(k - \beta^2 \sigma_v^2)}{h_k \sum_{j=1}^k \frac{1}{h_j}}$, $1 \leq k \leq n$ and we have assumed $h_1 \geq h_2 \geq \dots \geq h_n$. When $k = 1$, $f(1) = \frac{(1 - \beta^2 \sigma_v^2)}{h_1 \frac{1}{h_1}} < 1$ So, $f(k) > 1$ is not possible for all $k = 1, 2, \dots, n$. Therefore there are two possibilities: (I). $f(k) < 1$ for all $1 \leq k \leq n$: In this case we set $K_1 = n$. (II). There exists a unique K_1 such that $f(K_1) < 1$ and $f(K_1 + 1) \geq 1$, where $1 \leq K_1 \leq n$.

The uniqueness of K_1 implies that for any $k \geq K_1 + 1$, we should have that $f(k) \geq 1$. This can be proved by showing that if $f(k) \geq 1$, then $f(k + 1) \geq 1$. When $f(k) \geq 1$, it implies that

$$f(k + 1) = \frac{(k - \beta^2 \sigma_v^2) + 1}{(h_k \sum_{j=1}^k \frac{1}{h_j} + 1) + (h_{k+1} - h_k) \sum_{j=1}^k \frac{1}{h_j}} \quad (22)$$

The second term of the denominator of (22) is negative or equal to zero since we have assumed that $h_{k+1} \leq h_k$. Hence $f(k + 1) \geq \frac{(k - \beta^2 \sigma_v^2) + 1}{h_k \sum_{j=1}^k \frac{1}{h_j} + 1} > 1$ as required.

Uniqueness of the minimum of (13) : The uniqueness follows from the fact that, (15) is the only solution that satisfies the KKT conditions of the problem (13). Remaining is to show that the optimal solution (15) corresponds to a global minimum. To prove that, we will show that the Hessian matrix of the Lagrangian (14) is positive definite at the optimal solution. It can be seen that the Hessian matrix (H) of (14) is diagonal with $H_{k,k} = 2 + 2\lambda_0 h_k^2 \sigma_w^2 \frac{(3g_k^2 h_k^2 \sigma_v^2 - \sigma_w^2)}{(g_k^2 h_k^2 \sigma_v^2 + \sigma_w^2)^3}$ for $k = 1, 2, \dots, n$. As in (15), when $n > \beta^2 \sigma_v^2$ and $f(k) - 1 < 0$, optimal $g_k^2 = \frac{\sigma_w^2}{h_k^2 \sigma_v^2} \left[\frac{h_k \sum_{j=1}^{K_1} \frac{1}{h_j}}{(K_1 - \beta^2 \sigma_v^2)} - 1 \right]$. Then $H_{k,k} = 2 \left[1 - \left(4 \frac{h_{K_1}}{h_k} f(K_1) - 3 \right) \right] > 0$, since $f(K_1) < 1$ and $h_{K_1} \leq h_k$. When $n > \beta^2 \sigma_v^2$ and $f(k) - 1 > 0$, optimal $g_k^2 = 0$ and then $H_{k,k} = 2 \left[1 - \left(\frac{h_k}{h_{K_1} f(K_1)} \right)^2 \right] > 0$, since then $f(K_1) > 1$ (that is $k \geq K_1$) and therefore $h_k \leq h_{K_1}$. That is $H_{k,k} > 0$ for $k = 1, 2, \dots, n$ implies H a positive definite matrix.

REFERENCES

- [1] J.-J. Xiao, S. Cui, Z.-Q. Luo, and A. J. Goldsmith, "Joint estimation in sensor networks under energy constraints," in *Proc. IEEE first conf. Sensor and Ad Hoc commun. and Networks*, Santa Clara, CA, Oct. 2004.
- [2] L. Snidaro, R. Niu, P. Varshney, and G. L. Foresti, "Sensor fusion for video surveillance," in *Proc. 7th Int. Conf. Information Fusion*, Stockholm, June 2004.
- [3] A. Tiwari, F. L. Lewis, and S. S. Ge, "Wireless sensor network for machine condition based maintenance," in *Proc. 8th Control, Automation, Robotics and Vision Conference*, vol. 1, Dec. 2004, pp. 461–467.
- [4] J. N. Tsitsiklis, "Decentralized detection," *Advances in Statistical Signal Processing*, vol. 2, pp. 297–344, 1993.
- [5] Z. Chair and P. K. Varshney, "Optimal data fusion in multiple sensor detection systems," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-22, no. 1, pp. 98–101, Jan. 1986.
- [6] S. Appadwedula, V. V. Veeravalli, and D. L. Jones, "Energy efficient detection in sensor networks," *IEEE J. Select. Areas Commun.*, vol. 23, no. 4, pp. 693–702, Apr 2005.
- [7] J. F. Chamberland and V. V. Veeravalli, "Asymptotic results for decentralized detection in power constrained wireless sensor networks," *IEEE J. Select. Areas Commun.*, vol. 22, no. 6, pp. 1007–1015, Aug. 2004.

- [8] —, "Decentralized detection in wireless sensor systems with dependent observations," in *Proc. Int. Conf. Comput., Commun. and Contr. Technol. (CCCT)*, Austin, TX, Aug. 2004.
- [9] —, "The impact of fading on decentralized detection in power constrained wireless sensor networks," in *Proc. Acoust., Speech, Signal Processing 2004. (ICASSP '04)*, vol. 3, 2004, pp. 837–840.
- [10] S. K. Jayaweera, "Large sensor system performance of decentralized detection in noisy, bandlimited channels," in *IEEE 61st Vehicular Technology Conference (VTC) 2005 Spring*, Stockholm, Sweden, May 2005.
- [11] —, "Bayesian fusion performance and system optimization in distributed stochastic gaussian signal detection under communication constraints," *IEEE Trans. Signal Processing.*, vol. 55, no. 4, pp. 1238–1250, April 2007.
- [12] —, "Large system decentralized detection performance under communication constraints," *IEEE Commun. Lett.*, vol. 9, pp. 769–771, Sep. 2005.
- [13] R. Negi and A. Rajeswaran, "Capacity of power constrained ad-hoc networks," in *Proc. IEEE Infocom*, May 2004, pp. 443–453.
- [14] K. Altarazi, "Asymptotic fusion performance in a power constrained, distributed wireless sensor network," Master's thesis, Wichita State University, Wichita, KS, April 2006.
- [15] K. Altarazi, S. K. Jayaweera, and V. Aravinthan, "Performance of decentralized detection in a resource-constrained sensor network," in *39th Annual Asilomar Conf. on Signals, Systems and Computers*, Pacific Grove, CA, Nov. 2005.
- [16] X. Zhang, H. V. Poor, and M. Chiang, "Optimal power allocation for distributed detection in wireless sensor networks," *IEEE Trans. Signal Processing.*, Jan. 2007, Submitted.
- [17] A. Krasnopeev, J.-J. Xiao, and Z.-Q. Luo, "Minimum energy decentralized estimation in a wireless sensor network with correlated sensor noises," *EURASIP Journal on Wireless Commun. and Networking*, vol. 2005, no. 4, pp. 473–482, 2005.
- [18] J.-J. Xiao, S. Cui, and A. J. Goldsmith, "Power-efficient analog forwarding transmission in an inhomogeneous gaussian sensor network," in *6th Workshop on IEEE Signal Processing Advances in Wireless Commun.*, June 2005, pp. 121–125.
- [19] K. M. Abadir and J. R. Magnus, *Matrix Algebra*. New York, USA: Cambridge University Press, 2005.
- [20] J. Kennedy and W. M. Spears, "Matching algorithms to problems: An experimental test of the particle swarm and some genetic algorithms on multi modal problem generator," in *Proc. IEEE Int. Conf. Evolutionary Computation*, 1998.
- [21] J. Kennedy and R. C. Eberhart, "Particle swarm optimization," in *Proc. IEEE Conf. Neural Networks IV*, Piscataway, NJ, 1995.
- [22] J. Robinson and Y. R. Samii, "Particle swarm optimization in electromagnetism," *IEEE Trans. Antennas Propagat.*, vol. 52, no. 2, pp. 397–407, Feb. 2004.
- [23] S. S. Rao, *Optimization: Theory and Applications*. New Delhi, India: Wiley Eastern Limited, 1995.
- [24] J. M. Yang, Y. P. Chen, J. T. Horng, and C. Y. Kao, *Applying Family Competition to Evolution Strategies for Constrained Optimization. Lecture Notes in Computer Science*. Berlin Heidelberg, New York: Springer-Verlag, 1997.
- [25] K. E. Parsopoulos and M. N. Vrahatis, *Intelligent Technologies: New Trends in Intelligent Technologies, Book Chapter: Particle Swarm Optimization Method for Constrained Optimization Problems: pages 214-220*. IOS press, 2002.
- [26] A. Carlisle and G. Dozier, "An off-the-shelf PSO," in *Proc. 2001 Workshop on Particle Swarm Optimization*, Indianapolis, IN, 2001.
- [27] R. C. Eberhart and Y. Shi, "Evolving artificial neural networks," in *Proc. 1998 Int. Conf. Neural Networks and Brain*, Beijing, P.R.C, 1998.