Power Scheduling for Distributed Estimation in Cluster-Based Wireless Sensor Networks

Muhammad Hafeez Chaudhary and Luc Vandendorpe

***Abstract*—This work deals with distributed estimation problem in hierarchical wireless sensor networks, where the network is divided into spatially disjoint groups called clusters. The sensors in each cluster observe a separate random source which is correlated with the sources being observed by the other clusters. Each cluster has its designated cluster head (CH). The sensors in the clusters forward their observations to the CHs, which in turn communicate with a fusion center (FC). The estimation at the CHs and the FC is done based on the minimum mean square error estimation rule. To minimize the overall estimation distortion, we propose a power scheduling scheme that allocates power to the individual sensors and the CHs subject to constraints on the transmit power of individual clusters and the overall network. The correlation among the underlying sources leads to coupling of the optimization variables and the power allocation solution requires centralized computation, which may be computationally expensive. To this end, we propose an alternative formulation based on an upper-bound on the distortion function, which leads to a solution that exhibits favorable characteristics for distributed implementation. Simulation examples corroborate the effectiveness of the proposed power scheduling scheme.**

***Index Terms*—Hierarchical wireless sensor networks, parameter estimation, power scheduling, resource management, spatial correlation.**

I. Introduction

In recent years for wireless sensor networks (WSNs) several power-aware and energy-efficient estimation algorithms have been proposed under a wide variety of network models. The work of [1] considers estimation based on quantized sensor observations. In [2], the focus is on designing a power allocation scheme where sensors amplify and transmit their analog observations. The estimation schemes in these works target to estimate an unknown deterministic parameter. The works of [3] and [4] studied power allocation in WSNs with spatially correlated data. In all these aforementioned works, individual sensors send their observations to a central unit, called fusion center (FC), which forms estimate of the underlying source. In the realm of energy-efficient estimation, this centralized WSN topology, where all sensors directly transmit their observations to a central station, may not be an optimum choice. To this end, [5] investigated the minimal energy progressive estimation in sensor networks and [6] studied estimation under different network topological settings. The authors in [7] and [8] proposed power allocation schemes for estimation in cluster based WSNs. All these works propose power allocation schemes for estimation of a homogeneous unknown deterministic parameter and do not consider the effect of data correlation.

While the existing energy-aware or power-constrained estimation algorithms for WSNs ignore the effect of data correlation, in this paper we consider a system where the network is divided into clusters, where each cluster observes a separate source albeit correlated with the sources being observed by the other clusters. The estimation of underlying sources is performed in two time slots: In the first, the sensors in each cluster amplify and forward their noisy measurements to their respective CH that forms a preliminary estimate of the underlying source; and in the second, the CHs send a scaled version of their partial estimates to a FC that forms the final estimate of the sources. To this purpose, the CHs and the FC employ minimum mean square error (MMSE) estimation rule.

In [9], we proposed a power scheduling scheme in the cluster-based WSNs where we target to minimize the estimation distortion. However, under the proposed scheme, power allocations need to be computed numerically in a centralized fashion. For large networks, the computational cost and the implementation overhead associated with this centralized scheme may become prohibitive. To this end, in this paper, first we develop an upper-bound on the distortion function and subsequently we use that bound as a surrogate for the distortion function in the optimization problem for power allocation. The resulting power scheduling scheme has favorable structure for distributed implementation and gives distortion performance that matches quite well with the scheme in [9], which is based on the exact distortion function. Moreover, compared to a uniform power allocation scheme, the proposed design gives better distortion performance.

The remainder of the paper is organized as follows: Section II presents the system model, Section III formulates the power allocation problem and outlines its solution, Section IV presents some simulation examples, and finally Section V gives concluding remarks.

II. System Model and Preliminaries

Consider the cluster-based hierarchical sensor network shown in Fig. 1 in which N_0' spatially distributed sensor nodes are divided into $N_{\rm c}$ disjoint and non-overlapping clusters, indexed by $\mathcal{J} = \{1,\ldots,N_{\rm c}\}$, such that $N_0' = \sum_{j\in\mathcal{J}} N_j$, N_j being the number of sensors in cluster j indexed by $\tilde{\mathcal{I}}_j = \{1,\ldots,N_j\}$. The clusters observe zero-mean random Gaussian sources $s_j \sim \mathcal{N}(0,\sigma_{s_j}^2)$ for $j\in\mathcal{J}$ which are correlated such that $\mathrm{Cov}\{s_j,s_k\} = \rho_{s_j,s_k}\sigma_{s_j}\sigma_{s_k}$, where ρ_{s_j,s_k} specifies the correlation between s_j and s_k for all j and k in $\mathcal{J}$. The noisy observation at sensor i in cluster j is given by

$$x_{i,j} = s_j + n_{i,j}, \quad \forall i\in\tilde{\mathcal{I}}_j,\ \forall j\in\mathcal{J}, \tag{1}$$

where $n_{i,j}\sim\mathcal{N}(0,\sigma_{n_{i,j}}^2)$ denotes the observation noise, which is independent of s_j's and the observation noise across sensors.

As an example application of the sensor network, we can view that the network is deployed to observe a Gaussian spatial random field. We assume that inter-sensor distances within each cluster are small compared to the inter-cluster distances. The sensors in each cluster being close to each other have strong internal correlation and therefore we can model the field within each cluster as homogeneous. Whereas long inter-cluster distances suggest heterogeneous field values in different clusters.

The estimation of the sources is done in two phases. In the first phase, the sensors in each cluster amplify their observations and then transmit to their respective CH over orthogonal channels such that the received observations are

$$y_{i,j} = \sqrt{\phi_{i,j}\tilde{c}_{i,j}}(s_j + n_{i,j}) + w_{i,j}, \quad \forall i\in\mathcal{I}_j,\ \forall j\in\mathcal{J}, \tag{2}$$

where $\mathcal{I}_j = \tilde{\mathcal{I}}_j - \{N_j\}$. In (2) $\phi_{i,j}\in[0,\infty)$ is a scaling or an amplifying factor, $\tilde{c}_{i,j}$ is channel gain between the sensor and the

The authors are with the ICTEAM Institute, Université catholique de Louvain, 1348 Louvain-La-Neuve, Belgium (email: {Muhammad.Chaudhary, Luc.Vandendorpe}@uclouvain.be).

The authors would like to thank the ARC SCOOP, the Walloon region framework program COSMOS/TSARINE, and the EU project FP7 NEWCOM++ for the financial support and the scientific inspiration.

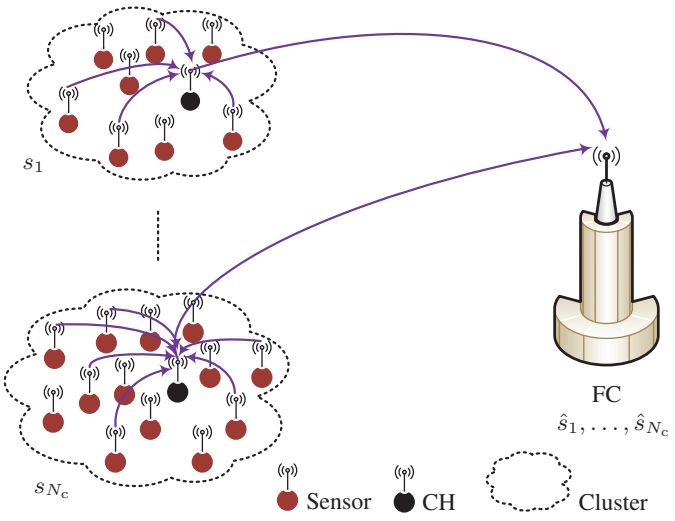

Fig. 1. A cluster-based hierarchical wireless sensor network.

CH, and $w_{i,j} \sim \mathcal{N}(0, \sigma^2_{w_{i,j}})$ is receiver noise which is assumed to be independent of s_j and $n_{i,j}$ for all i and j. Moreover, the receiver noise is assumed to be independent (of the receiver noises) across sensors in all clusters. Note that without any loss of generality we have assumed that the sensor N_j is designated as CH of cluster j. The designated CH for each cluster can be a fixed sensor or it can be dynamically selected from among the sensors in that cluster; for example, it could be a sensor with maximum remaining energy or better channel gain to the FC.

Employing the MMSE estimation rule [10], the CH j forms an estimate $\hat{s}_j$ of the source s_j based on the received observations from the sensors in the cluster. We can write the estimate $\hat{s}_j$ as

$$\hat{s}_j = D_j \left(\frac{x_{N_j}}{\sigma^2_{n_{N_j}}} + \sum_{i \in \mathcal{I}_j} \frac{\sqrt{\phi_{i,j}\tilde{c}_{i,j}}\, y_{i,j}}{\phi_{i,j}\tilde{c}_{i,j}\sigma^2_{n_{i,j}} + \sigma^2_{w_{i,j}}} \right), \tag{3}$$

where D_j is estimation MSE that can be given by

$$D_j = \left(\frac{1}{\sigma^2_{s_j}} + \frac{1}{\sigma^2_{N_j}} + \sum_{i \in \mathcal{I}_j} \frac{\phi_{i,j} c_{i,j}}{\phi_{i,j} c_{i,j} \sigma^2_{i,j} + 1} \right)^{-1}, \tag{4}$$

where $\sigma^2_{i,j} = \sigma^2_{n_{i,j}}$ for $i \in \mathcal{I}_j$ and $\sigma^2_{N_j} = \sigma^2_{n_{N_j},j}$ for $j \in \mathcal{J}$. Moreover $c_{i,j} = \tilde{c}_{i,j}/\sigma^2_{w_{i,j}}$ for all i and j.

By defining $\tilde{\sigma}^2_j$ as

$$\tilde{\sigma}^2_j = \left(\frac{1}{\sigma^2_{N_j}} + \sum_{i \in \mathcal{I}_j} \frac{\phi_{i,j} c_{i,j}}{\phi_{i,j} c_{i,j} \sigma^2_{i,j} + 1} \right)^{-1} \tag{5}$$

we can write $v_j = D_j \tilde{\sigma}^2_j \hat{s}_j$, a scaled version of $\hat{s}_j$, as follows:

$$v_j = s_j + \vartheta_j, \quad \forall j \in \mathcal{J}, \tag{6}$$

where

$$\vartheta_j = \tilde{\sigma}^2_j \left(\frac{n_{N_j,j}}{\sigma^2_{N_j}} + \sum_{i \in \mathcal{I}_j} \frac{\phi_{i,j}\tilde{c}_{i,j} n_{i,j} + \sqrt{\phi_{i,j}\tilde{c}_{i,j}}\, w_{i,j}}{\phi_{i,j}\tilde{c}_{i,j}\sigma^2_{n_{i,j}} + \sigma^2_{w_{i,j}}} \right). \tag{7}$$

We can show that $\vartheta_j \sim \mathcal{N}(0, \tilde{\sigma}^2_j)$, which is independent of s_j and ϑ_k for all $j \neq k$. We can view v_j as an equivalent observation at the CH with ϑ_j representing the equivalent observation noise.

In the second phase of an estimation cycle, the CHs amplify and transmit the observations (6) to the FC over orthogonal channels such that the received observations are

$$z_j = \sqrt{\psi_j \tilde{g}_j}(s_j + \vartheta_j) + w_j, \quad \forall j \in \mathcal{J}, \tag{8}$$

where $\psi_j \in [0, \infty)$ is an amplifying factor, $\tilde{g}_j$ is gain of the channel between CH j and the FC, and $w_j \sim \mathcal{N}(0, \sigma^2_{w_j})$ is receiver noise at the FC which is independent of s_j, ϑ_j, and w_k for all j and k with $k \neq j$. Using matrix–vector notation, (8) can be written in a compact form as follows:

$$\mathbf{z} = \tilde{\mathbf{H}}\mathbf{s} + \mathbf{r}, \tag{9}$$

where

$$\begin{aligned} &\mathbf{z} = [z_1, \ldots, z_{N_c}]^T, \quad \mathbf{s} = [s_1, \ldots, s_{N_c}]^T, \\ &\tilde{\mathbf{H}} = \mathrm{diag}\left(\sqrt{\psi_1 \tilde{g}_1}, \ldots, \sqrt{\psi_{N_c}\tilde{g}_{N_c}}\right), \\ &\mathbf{r} = \left[\sqrt{\psi_1\tilde{g}_1}\vartheta_1 + w_1, \ldots, \sqrt{\psi_{N_c}\tilde{g}_{N_c}}\vartheta_{N_c} + w_{N_c}\right]^T. \end{aligned}$$

Now based on (9) and employing MMSE estimation rule, the FC forms an estimate of the underlying source vector $\mathbf{s}$ that can be given by

$$\hat{\mathbf{s}} = \mathbf{R}_\mathbf{s}\tilde{\mathbf{H}}^T\left(\tilde{\mathbf{H}}\mathbf{R}_\mathbf{s}\tilde{\mathbf{H}}^T + \tilde{\mathbf{R}}\right)^{-1}\mathbf{z}, \tag{10}$$

where $\mathbf{R}_\mathbf{s} := \mathbb{E}\left[\mathbf{s}\mathbf{s}^T\right]$ and $\tilde{\mathbf{R}} := \mathbb{E}\left[\mathbf{r}\mathbf{r}^T\right]$. By defining $\epsilon := \mathbf{s} - \hat{\mathbf{s}}$ as the estimation error vector, we can write the covariance of the error vector, denoted by $\tilde{\mathbf{R}}_\epsilon$, as follows [9]:

$$\begin{aligned} \tilde{\mathbf{R}}_\epsilon &= \mathbf{R}_\mathbf{s} - \mathbf{R}_\mathbf{s}\tilde{\mathbf{H}}^T\left(\tilde{\mathbf{H}}\mathbf{R}_\mathbf{s}\tilde{\mathbf{H}}^T + \tilde{\mathbf{R}}\right)^{-1}\tilde{\mathbf{H}}\mathbf{R}_\mathbf{s}^T \\ &= \left(\tilde{\mathbf{H}}\tilde{\mathbf{R}}^{-1}\tilde{\mathbf{H}}^T + \mathbf{R}_\mathbf{s}^{-1}\right)^{-1}. \end{aligned} \tag{11}$$

III. Iterative Power Scheduling Scheme

We base our power scheduling scheme on the following optimization problem:

$$\begin{aligned} \underset{\psi_j \geq 0,\, \phi_{i,j} \geq 0,\, \forall i,j}{\text{minimize}} \quad & \mathrm{tr}\left(\tilde{\mathbf{R}}_\epsilon\right) \\ \text{subject to} \quad & \sum_{j \in \mathcal{J}} \left(\psi_j\left(\sigma^2_{s_j} + \tilde{\sigma}^2_j\right) + \sum_{i \in \mathcal{I}_j} \phi_{i,j}\left(\sigma^2_{s_j} + \sigma^2_{i,j}\right)\right) \leq P_\mathrm{t}, \\ & \sum_{j \in \mathcal{J}} \psi_j\left(\sigma^2_{s_j} + \tilde{\sigma}^2_j\right) \leq \psi_{\max}, \\ & \sum_{i \in \mathcal{I}_j} \phi_{i,j}\left(\sigma^2_{s_j} + \sigma^2_{i,j}\right) \leq \phi^{[j]}_{\max}, \quad \forall j \in \mathcal{J}. \end{aligned} \tag{12}$$

Where the first constraint limits the total network transmit power, which makes sense from the view point of global energy efficiency and to realize green ICT [11]. Moreover, this constraint enables a fair comparison between networks of different sizes. The second and third constraints limit inter-cluster interference and interference with any neighboring network, which is important from the perspective of spatial reuse of spectrum resources. Depending on the application of the WSN, some clusters may be located in critical areas and it may be required to keep those clusters alive for sufficiently long time; this observation gives another motivation for putting cap on total transmit power of individual clusters.

Note that the optimization variables ψ_j and $\phi_{i,j}$ are coupled in the constraints. This coupling of the variables, and the reason that the optimization problem (12) is jointly non-convex over the optimization variables make the problem difficult to solve. To this end, in Prop. 1 (due to space constraints, the proof of which is given in [12]) we reformulate the problem in an equivalent form that bears favorable characteristics as shall be seen in the ensuing development.

Proposition 1: Let $\alpha \in [0, 1)$ such that αP_t power is expended in all clusters on forwarding observations from the sensors to the CHs and $(1 - \alpha)P_\mathrm{t}$ power is expended on forwarding observations from the CHs to the FC. Moreover, assuming $\xi_j \geq 0$, $\gamma_j \geq 0$, and $\beta_{i,j} \geq 0$

such that $\sum_{j\in\mathcal{J}}\xi_j \le 1$, $\sum_{j\in\mathcal{J}}\gamma_j \le 1$, and $\sum_{i\in\mathcal{I}_j}\beta_{i,j} \le 1$, we can write for all $i \in \mathcal{I}_j$ and $j \in \mathcal{J}$ as follows:

$$\psi_j = \frac{(1-\alpha)P_{\mathrm{t}}\xi_j}{\sigma_{s_j}^2 + \tilde{\sigma}_j^2}, \quad \phi_{i,j} = \frac{\alpha P_{\mathrm{t}}\gamma_j\beta_{i,j}}{\sigma_{s_j}^2 + \sigma_{i,j}^2}. \tag{13}$$

With this we can write the problem (12) in the following equivalent form:

$$\begin{aligned} \underset{\alpha,\,\xi_j\ge0,\,\gamma_j\ge0,\,\beta_{i,j}\ge0,\,\forall i,j}{\text{minimize}} \quad & \mathrm{tr}\left(\mathbf{R}_\epsilon\right) \\ \text{subject to} \quad & \alpha \in \mathcal{T},\ \sum_{j\in\mathcal{J}}\xi_j \le 1,\ \sum_{j\in\mathcal{J}}\gamma_j \le 1, \\ & \sum_{i\in\mathcal{I}_j}\beta_{i,j} \le 1,\ \gamma_j \le \gamma_{\max}^{[j]}, \quad \forall j \in \mathcal{J}. \end{aligned} \tag{14}$$

Where $\mathcal{T} = [\alpha_0, 1)$, $\alpha_0 = \max\{0, 1 - \psi_{\max}/P_{\mathrm{t}}\}$, $\gamma_{\max}^{[j]} = \min\{1, \phi_{\max}^{[j]}/P_{\mathrm{t}}\}$, and $\mathbf{R}_\epsilon$ is given as

$$\begin{aligned} \mathbf{R}_\epsilon &= \mathbf{R}_{\mathbf{s}} - \mathbf{R}_{\mathbf{s}}\mathbf{H}^T\left(\mathbf{H}\mathbf{R}_{\mathbf{s}}\mathbf{H}^T + \mathbf{R}\right)^{-1}\mathbf{H}\mathbf{R}_{\mathbf{s}}^T \\ &= \left(\mathbf{H}\mathbf{R}^{-1}\mathbf{H}^T + \mathbf{R}_{\mathbf{s}}^{-1}\right)^{-1}, \end{aligned} \tag{15}$$

with

$$\begin{aligned} \mathbf{H} &= \mathrm{diag}\left(\sqrt{(1-\alpha)P_{\mathrm{t}}\xi_1 g_1}, \ldots, \sqrt{(1-\alpha)P_{\mathrm{t}}\xi_{N_{\mathrm{c}}} g_{N_{\mathrm{c}}}}\right), \\ \mathbf{R} &= \mathrm{diag}\left((1-\alpha)P_{\mathrm{t}}\xi_1 g_1\sigma_1^2 + \sigma_1^2 + \sigma_{s_1}^2, \ldots,\right. \\ &\qquad \left.(1-\alpha)P_{\mathrm{t}}\xi_{N_{\mathrm{c}}} g_{N_{\mathrm{c}}}\sigma_{N_{\mathrm{c}}}^2 + \sigma_{N_{\mathrm{c}}}^2 + \sigma_{s_{N_{\mathrm{c}}}}^2\right). \end{aligned} \tag{16}$$

In (16), for all j, $g_j = \tilde{g}_j/\sigma_{w_j}^2$ and

$$\sigma_j^2 = \left(\frac{1}{\sigma_{N_j}^2} + \sum_{i\in\mathcal{I}_j}\frac{\alpha P_{\mathrm{t}}\gamma_j\beta_{i,j}c_{i,j}}{\alpha P_{\mathrm{t}}\gamma_j\beta_{i,j}c_{i,j}\sigma_{i,j}^2 + \sigma_{i,j}^2 + \sigma_{s_j}^2}\right)^{-1}. \tag{17}$$

Proof: For proof see [12]. □

The equivalent formulation in (14) has linear constraints and the constraints are independent in the sense that each constraint function depends on a separate set of optimization variables (i.e., α, ξ_j's, γ_j's, or $\beta_{i,j}$'s). This independence of the constraints is a nice property that allows us to solve the problem using the block-coordinate descent method (BCoDM) [13]: Whereby we divide the problem into subproblems and in each subproblem we can optimize over a separate set of optimization variables. This approach is used in [9] where the solution of each subproblem relies on centralized numerical methods, because the optimization variables are nonlinearly coupled in the objective function due to the correlation of the underlying sources. The centralized numerical solution may be computationally expensive and may incur high overhead in implementation. To this end, in the ensuing development, to solve the problem (14), first we develop an upper-bound for the objective function and subsequently we use this upper-bound as a surrogate for the objective function and solve the optimization problem. The resulting solution is amenable for distributed implementation as we shall see in the ensuing development.

***Proposition** 2:* The trace of $\mathbf{R}_\epsilon$ can be upper bounded as

$$\mathrm{tr}\left(\mathbf{R}_\epsilon\right) \le \Big(\sum_{j\in\mathcal{J}}\sigma_{s_j}^2\Big) - \Upsilon, \tag{18}$$

where

$$\Upsilon = \frac{\left(\sum_{j\in\mathcal{J}}\xi_j g_j\Psi_j\right)^2}{\sum_{j\in\mathcal{J}}\xi_j g_j\left(\sum_{k\in\mathcal{J}}\xi_k g_k\tilde{\mathbf{Q}}_{[j,k]} + \xi_j g_j\Psi_j\sigma_j^2 + \frac{\Psi_j\left(\sigma_j^2+\sigma_{s_j}^2\right)}{(1-\alpha)P_{\mathrm{t}}}\right)},$$

$$\Psi_j = \sum_{k\in\mathcal{J}}\mathbf{R}_{\mathbf{s}[j,k]}^2 = \sum_{k\in\mathcal{J}}\mathrm{Cov}\left\{s_j, s_k\right\}^2, \quad \tilde{\mathbf{Q}} = \mathbf{R}_{\mathbf{s}} \circ (\mathbf{R}_{\mathbf{s}}^T\mathbf{R}_{\mathbf{s}})$$

with $\mathbf{R}_{\mathbf{s}} \circ (\mathbf{R}_{\mathbf{s}}^T\mathbf{R}_{\mathbf{s}})$ denoting the Hadamard or Schur product of matrices $\mathbf{R}_{\mathbf{s}}$ and $(\mathbf{R}_{\mathbf{s}}^T\mathbf{R}_{\mathbf{s}})$.

Proof: The proof is given in [12]. □

Now we consider the optimization problem (14) where, instead of the actual distortion function, we target to minimize the upper-bound on the distortion. For this purpose, it is sufficient to consider the following optimization problem:

$$\begin{aligned} \underset{\alpha,\,\xi_j\ge0,\,\gamma_j\ge0,\,\beta_{i,j}\ge0,\,\forall i,j}{\text{minimize}} \quad & -\Upsilon \\ \text{subject to} \quad & \alpha \in \mathcal{T},\ \sum_{j\in\mathcal{J}}\xi_j \le 1,\ \sum_{j\in\mathcal{J}}\gamma_j \le 1, \\ & \sum_{i\in\mathcal{I}_j}\beta_{i,j} \le 1,\ \gamma_j \le \gamma_{\max}^{[j]}, \quad \forall j \in \mathcal{J}. \end{aligned} \tag{19}$$

The power allocation problem (19) can be solved using the BCoDM, which cyclically/iteratively minimizes the cost function with respect to each set of optimization variables subject to the associated constraints while the other optimization variables are held fixed. Specifically to solve the problem do the following.

Initialize α, ξ_j's, γ_j's, and $\beta_{i,j}$'s in their respective feasible region, and repeat step 1 to step 4 until there is no appreciable decrease in the objective function.

1: For given ξ_j's, γ_j's, and $\beta_{i,j}$'s, find α by solving

$$\underset{\alpha\in\mathcal{T}}{\text{minimize}} \ -\Upsilon. \tag{20}$$

2: For given α, γ_j's, and $\beta_{i,j}$'s, find ξ_j's by solving

$$\underset{\xi_j\ge0,\,\forall j}{\text{minimize}} \ -\Upsilon \quad \text{subject to} \sum_{j\in\mathcal{J}}\xi_j \le 1. \tag{21}$$

3: For given α, ξ_j's, and $\beta_{i,j}$'s, find γ_j's by solving

$$\underset{\gamma_j\ge0,\forall j}{\text{minimize}} \ -\Upsilon \quad \text{subject to} \sum_{j\in\mathcal{J}}\gamma_j \le 1,\ \gamma_j \le \gamma_{\max}^{[j]},\ \forall j. \tag{22}$$

4: For given α, ξ_j's, and γ_j's, find $\beta_{i,j}$'s for each $j \in \mathcal{J}$ by solving

$$\underset{\beta_{i,j}\ge0,\,\forall i}{\text{minimize}} \ -\Upsilon \quad \text{subject to} \sum_{i\in\mathcal{I}_j}\beta_{i,j} \le 1. \tag{23}$$

Where the solution of problem (20) to problem (23) is outlined in Section III-A to Section III-D, respectively.

A. Optimization of α

Let $f_j(\alpha) = \xi_j g_j\Psi_j\left(\xi_j g_j\sigma_j^2 + \frac{(\sigma_j^2+\sigma_{s_j}^2)}{(1-\alpha)P_{\mathrm{t}}}\right)$ such that $f(\alpha) = \sum_{j\in\mathcal{J}} f_j(\alpha)$. Now to solve the optimization problem (20) for α, it is sufficient to consider the following problem:

$$\underset{\alpha\in\mathcal{T}}{\text{minimize}} \ f(\alpha), \tag{24}$$

where an explicit solution for α is intractable. However, to find α, we may resort to numerical methods such as line search methods for one-dimensional minimization (e.g., the Golden Section method [13]). Note that $f(\alpha) = \sum_{j\in\mathcal{J}} f_j(\alpha)$, where $f_j(\alpha)$ depends on the parameters concerning the cluster j, has separable structure along the clusters. From the implementation perspective, the FC broadcasts an initial value of $\alpha \in \mathcal{T}$ to all CHs. Then each CH computes $f_j(\alpha)$ and/or $\partial f_j(\alpha)/\partial\alpha$ (as required by the numerical method) and sends to the FC. The FC then updates the value of α and broadcasts it to the CHs. This procedure is repeated until a stopping criterion for the numerical method is satisfied.

B. Optimization of ξ_j's

For optimization of ξ_j's, we proceed as follows. By defining

$$\begin{aligned}
&\xi = [\xi_1, \ldots, \xi_{N_\mathrm{c}}]^T, \\
&\breve{\mathbf{Q}} = \mathrm{diag}\left(g_1^2 \Psi_1 \sigma_1^2, \ldots, g_{N_\mathrm{c}}^2 \Psi_{N_\mathrm{c}} \sigma_{N_\mathrm{c}}^2\right), \\
&\mathbf{q} = \left[\frac{g_1 \Psi_1 (\sigma_1^2 + \sigma_{s_1}^2)}{(1-\alpha)P_\mathrm{t}}, \ldots, \frac{g_{N_\mathrm{c}} \Psi_{N_\mathrm{c}} (\sigma_{N_\mathrm{c}}^2 + \sigma_{s_{N_\mathrm{c}}}^2)}{(1-\alpha)P_\mathrm{t}}\right]^T, \\
&\mathbf{u} = [g_1 \Psi_1, \ldots, g_{N_\mathrm{c}} \Psi_{N_\mathrm{c}}]^T, \quad \mathbf{g} = [g_1, \ldots, g_{N_\mathrm{c}}]^T, \\
&\mathbf{U} = \mathbf{u}\mathbf{u}^T, \quad \mathbf{G} = \mathbf{g}\mathbf{g}^T, \\
&\mathbf{Q} = \mathbf{R_s} \circ (\mathbf{R_s}^T \mathbf{R_s}) \circ \mathbf{G} + \breve{\mathbf{Q}},
\end{aligned}$$

we can write Υ as follows:

$$\Upsilon = \frac{\xi^T \mathbf{U} \xi}{\xi^T \mathbf{Q} \xi + \mathbf{q}^T \xi}. \tag{25}$$

Now by defining $\mathbf{1} := [1, \ldots, 1]^T$, the optimization problem (21) can be written as

$$\underset{\xi \geq 0}{\text{maximize}}\ \Upsilon \quad \text{subject to } \mathbf{1}^T \xi = 1, \tag{26}$$

where we have replaced the inequality constraint with equality to exclude the case where the denominator of Υ is zero over the feasible region.

Let $\mathcal{F} = \left\{\xi \in \mathbb{R}^{N_\mathrm{c}} | \xi \geq \mathbf{0},\ \mathbf{1}^T \xi = 1\right\}$ denote the feasible region of the problem (26). Note that $\mathcal{F}$ is a compact convex set in $\mathbb{R}^{N_\mathrm{c}}$ (the set of N_c-dimensional real numbers). It is easy to show that $\mathbf{U}$ and $\mathbf{Q}$ are positive semidefinite matrices, which means that the numerator and denominator of Υ are convex functions of ξ. Thus, the problem (26) is a convex–convex type quadratic fractional programming problem. To solve (26), in what follows, we develop an algorithm based on the parametric programming approach, which is a powerful scheme for solving fractional programs.

For $\theta \geq 0$, let

$$f(\xi; \theta) = \xi^T \mathbf{U} \xi - \theta \left(\xi^T \mathbf{Q} \xi + \mathbf{q}^T \xi\right) \tag{27}$$

be a parametrized function associated with the problem (26). We have the following proposition, which is based on the well-known result by Dinkelbach [15].

***Proposition** 3:* For given θ, define

$$\varphi(\theta) = \underset{\xi \in \mathcal{F}}{\text{maximize}}\ f(\xi; \theta) \tag{28}$$

with the corresponding optimal ξ vector as

$$\xi(\theta) = \arg\max_{\xi \in \mathcal{F}} f(\xi; \theta). \tag{29}$$

If there exists some $\theta^* \geq 0$ such that $\varphi(\theta^*) = 0$ then $\xi^* = \xi(\theta^*)$ is an optimal solution of the problem (26) and the corresponding optimal value is $\theta^* = \Upsilon(\xi^*)$.

Proof: For a detailed proof see [12]. □

Based on Prop. 3, the optimization problem (26) can be solved using the following iterative procedure.

i: Set $\iota = 0$ and initialize $\xi^{(\iota)} \in \mathcal{F}$.
ii: Compute $\theta^{(\iota+1)} = \Upsilon\left(\xi^{(\iota)}\right)$.
iii: Solve the following optimization problem to obtain the global optimal solution for $\xi^{(\iota+1)}$:

$$\underset{\xi \in \mathcal{F}}{\text{maximize}}\ f(\xi; \theta^{(\iota+1)}). \tag{30}$$

iv: If $|f(\xi^{(\iota+1)}; \theta^{(\iota+1)})| \leq \delta$ for some $\delta > 0$ then terminate; else set $\iota = \iota + 1$ and go to step ii.

The preceding algorithm is guaranteed to converge to the optimal solution of the problem (26) provided that the problem (30) can be solved [15]. To this end, note that although the feasible region $\mathcal{F}$ is a convex set, the function $f(\xi; \theta^{(\iota)})$ is not concave. Therefore the problem (30) is a non-concave maximization problem wherein many different local maxima may exist, which are different from the globally optimal solution. To this end, by introducing a slack variable $\tau = \xi^T \mathbf{U} \xi$ we can reformulate the problem (30) in the following equivalent form:

$$\begin{aligned}
&\underset{\tau_{\min} \leq \tau \leq \tau_{\max};\, \xi \geq 0}{\text{minimize}}\ \theta^{(\iota)} \left(\xi^T \mathbf{Q} \xi + \mathbf{q}^T \xi + \epsilon\right) - \tau \\
&\text{subject to} \quad \mathbf{1}^T \xi = 1,\ \xi^T \mathbf{U} \xi - \tau \leq 0,
\end{aligned} \tag{31}$$

which is a convex quadratically constrained quadratic programming (QCQP) problem that can be efficiently solved by numerical methods, for example, the interior point method [14]. In (31), the $\tau_{\min}$ and $\tau_{\max}$ are given as follows:

$$\tau_{\min} = \underset{\xi \in \mathcal{F}}{\text{minimize}}\ \xi^T \mathbf{U} \xi = \underset{\xi \in \mathcal{F}}{\text{minimize}}\ \left(\mathbf{u}^T \xi\right)^2, \tag{32}$$

$$\tau_{\max} = \underset{\xi \in \mathcal{F}}{\text{maximize}}\ \xi^T \mathbf{U} \xi = \underset{\xi \in \mathcal{F}}{\text{maximize}}\ \left(\mathbf{u}^T \xi\right)^2, \tag{33}$$

where we can show that $\tau_{\min} = (\min\{u_j, \ldots, u_{N_\mathrm{c}}\})^2$ and $\tau_{\max} = (\max\{u_j, \ldots, u_{N_\mathrm{c}}\})^2$.

C. Optimization of γ_j's

For optimization over γ_j's we have problem (22) to solve wherein it is sufficient to consider the following problem:

$$\begin{aligned}
&\underset{\gamma_j \geq 0,\, \forall j}{\text{minimize}}\ \sum_{j \in \mathcal{J}} \sigma_j^2 \xi_j g_j \Psi_j \left(1 + (1-\alpha) P_\mathrm{t} \xi_j g_j\right) \\
&\text{subject to} \sum_{j \in \mathcal{J}} \gamma_j \leq 1,\ \gamma_j \leq \gamma_{\max}^{[j]}, \quad \forall j \in \mathcal{J}.
\end{aligned} \tag{34}$$

We can prove that the objective function is decreasing with respect to γ_j's and the problem is jointly convex over γ_j's. Note that the objective function and the constraints are separable along clusters, consequently the Lagrange dual-decomposition method can be used to solve the problem [13].

In the optimization problem (34), as we are minimizing a decreasing function, therefore the optimum is always at the boundary of the constraints set. In the optimization, one of the following three scenarios may arise. First, if $\sum_{j \in \mathcal{J}} \gamma_{\max}^{[j]} < 1$ then the sum-constraint (i.e., $\sum_{j \in \mathcal{J}} \gamma_j \leq 1$) is inactive and all individual-constraints (i.e., $\gamma_j \leq \gamma_{\max}^{[j]}$ for all j) are active. In this particular case, the optimization problem is trivial and all clusters simply transmit with $\gamma_j = \gamma_{\max}^{[j]}$ for all j. Second, if $\sum_{j \in \mathcal{J}} \gamma_{\max}^{[j]} = 1$ then the sum- and all individual-constraints are active, and we simply have $\gamma_j = \gamma_{\max}^{[j]}$ for all j. Finally, if $\sum_{j \in \mathcal{J}} \gamma_{\max}^{[j]} > 1$ then the sum-constraint is always active and some of the individual-constraints may be active while others remain inactive. To solve the optimization problem in this later case, we proceed as follows. First we ignore the individual constraints and solve the problem with only the sum-constraint; then later on in this section we shall show how to incorporate the individual-constraints into the solution.

For solution to the problem (34) without considering the constraints on individual γ_j's, that is, to solve the problem

$$\begin{aligned}
&\underset{\gamma_j \geq 0,\, \forall j}{\text{minimize}}\ \sum_{j \in \mathcal{J}} \sigma_j^2 \xi_j g_j \Psi_j \left(1 + (1-\alpha) P_\mathrm{t} \xi_j g_j\right) \\
&\text{subject to} \sum_{j \in \mathcal{J}} \gamma_j \leq 1,
\end{aligned} \tag{35}$$

we propose a primal–dual algorithm based on the Lagrange dual-decomposition approach. For this purpose, we can write the Lagrange

function associated with the problem as follows:

$$\Lambda(\gamma_1,\ldots,\gamma_{N_\mathrm{c}};\mu)=\sum_{j\in\mathcal{J}}\Lambda_j(\gamma_j,\mu)-\mu, \tag{36}$$

where μ is a Lagrange multiplier (also called dual variable or price value) associated with the constraint $\sum_{j\in\mathcal{J}}\gamma_j\leq 1$ and

$$\Lambda_j(\gamma_j,\mu)=\sigma_j^2\xi_j g_j\Psi_j\big(1+(1-\alpha)P_\mathrm{t}\xi_j g_j\big)+\mu\gamma_j. \tag{37}$$

The corresponding dual objective function can be given by

$$\begin{aligned}\Omega(\mu)&=\underset{\gamma_j\geq 0,\,\forall j}{\text{minimize}}\ \Lambda(\gamma_1,\ldots,\gamma_{N_\mathrm{c}};\mu)\\&=\sum_{j\in\mathcal{J}}\underset{\gamma_j\geq 0}{\text{minimize}}\ \Lambda_j(\gamma_j,\mu)-\mu,\end{aligned} \tag{38}$$

and the dual optimization problem can be written as

$$\underset{\mu\geq 0}{\text{maximize}}\ \Omega(\mu). \tag{39}$$

For the dual objective $\Omega(\mu)$, we need to find γ_j's that minimize $\Lambda(\gamma_1,\ldots,\gamma_{N_\mathrm{c}};\mu)$. To this end, for given μ, $\Omega(\mu)$ can be obtained by solving N_c separate problems as follows:

$$\gamma_j(\mu)=\arg\min_{\gamma_j\geq 0}\ \Lambda_j(\gamma_j,\mu),\quad j\in\mathcal{J}. \tag{40}$$

Note that, (40) corresponds to cluster j that can be solved by the corresponding CH using simple line search algorithm for one-dimensional minimization.

The optimal dual variable μ can be obtained by finding μ such that $\sum_{j\in\mathcal{J}}\gamma_j(\mu)=1$. This can be done by a one-dimensional numerical search, for example, using bi-sectional search method or can be done using gradient-ascent method that leads to the following updation rule [13]:

$$\mu^{(\kappa+1)}=\left[\mu^{(\kappa)}+\delta^{(\kappa)}\bigg(\sum_{j\in\mathcal{J}}\gamma_j\big(\mu^{(\kappa)}\big)-1\bigg)\right]^+, \tag{41}$$

where κ is an iteration-index, δ is a positive step-size parameter, and $\gamma_j\big(\mu^{(\kappa)}\big)$ is solution of (40) for given $\mu^{(\kappa)}$. Because the underlying primal optimization problem is convex and satisfies the Slater constraint qualification conditions, therefore the primal variables $\gamma_j\big(\mu^{(\kappa)}\big)$'s and the dual variable $\mu^{(\kappa)}$ converge to their optimal values as $\kappa\to\infty$ and at convergence there is zero-duality gap [13].

To solve the problem (34) including the constraints on individual γ_j's, we adopt the following iterative procedure.

i: Assume $\gamma_\mathrm{t}^{(0)}=1$ such that the sum-constraint can be written as $\sum_{j\in\mathcal{J}}\gamma_j\leq\gamma_\mathrm{t}^{(0)}$.
ii: Solve the optimization problem as outlined in (36)–(41) ignoring the individual constraints.
iii: Construct the index-set $\mathcal{L}=\big\{j\in\mathcal{J}|\gamma_j\geq\gamma_{\max}^{[j]}\big\}$, and for all $j\in\mathcal{L}$ set $\gamma_j=\gamma_{\max}^{[j]}$.
iv: Recalculate the sum-constraint as $\gamma_\mathrm{t}^{[\iota]}=\gamma_\mathrm{t}^{[\iota-1]}-\sum_{j\in\mathcal{L}}\gamma_{\max}^{[j]}$, where ι is an iteration-index.
v: Recalculate γ_r for all $r\in\mathcal{R}^{[\iota]}$, where $\mathcal{R}^{[\iota]}=\mathcal{R}^{[\iota-1]}\setminus\mathcal{J}$ with $\mathcal{R}^{[0]}=\mathcal{J}$, as in step ii with the sum-constraint $\sum_{r\in\mathcal{R}^{[\iota]}}\gamma_r\leq\gamma_\mathrm{t}^{[\iota]}$. Note that $\mathcal{R}^{[\iota-1]}\setminus\mathcal{L}$ means all elements of $\mathcal{R}^{[\iota-1]}$ that are not in $\mathcal{L}$.
vi: Repeat step ii to step v until all constraints are satisfied.

As the given optimization problem is jointly convex over γ_j's and the objective function is a decreasing function of γ_j's, thus the solution given in (36)–(41) with the sum-constraint and the solution obtained by the preceding iterative procedure incorporating both the sum- and the individual-constraints are optimal.

From the viewpoint of implementation, the solution to given optimization problem based on the primal–dual approach as outlined in this section can be obtained in a distributed fashion with the assistance of the CHs. Specifically, the FC first broadcasts an initial price value, that is, the value of μ. This value is used by the CHs to calculate γ_j's by solving (40). Note that for CH j, the problem (40) entirely depends on the local information concerning that cluster. The new γ_j's are then sent to the FC so that to update the price μ. This updated value is then broadcasted to the CHs. This procedure is repeated until γ_j's and μ converge to their optimal value.

D. Optimization of $\beta_{i,j}$'s

For optimization of $\beta_{i,j}$'s, it is sufficient to consider the following optimization problem for $j\in\mathcal{J}$:

$$\underset{\beta_{i,j}\geq 0,\,\forall i}{\text{minimize}}\ \sigma_j^2\quad\text{subject to}\ \sum_{i\in\mathcal{I}_j}\beta_{i,j}\leq 1, \tag{42}$$

which is equivalent to

$$\begin{aligned}&\underset{\beta_{i,j}\geq 0,\,\forall i}{\text{minimize}}\ \sum_{i\in\mathcal{I}_j}\frac{-\alpha P_\mathrm{t}\gamma_j\beta_{i,j}c_{i,j}}{\alpha P_\mathrm{t}\gamma_j\beta_{i,j}c_{i,j}\sigma_{i,j}^2+\sigma_{i,j}^2+\sigma_{s_j}^2}\\&\text{subject to}\ \sum_{i\in\mathcal{I}_j}\beta_{i,j}\leq 1,\end{aligned} \tag{43}$$

where we can show that the objective function is a decreasing function of $\beta_{i,j}$ and the problem is jointly convex over $\beta_{i,j}$'s. The optimal solution for $\beta_{i,j}$'s is outlined in the following, which is obtained by solving the Karush–Kuhn–Tucker optimality conditions [14] associated with the problem (43).

$$\beta_{i,j}=\frac{\sigma_{s_j}^2+\sigma_{i,j}^2}{\alpha P_\mathrm{t}\gamma_j c_{i,j}\sigma_{i,j}^2}\left(\sqrt{\frac{\alpha P_\mathrm{t}\gamma_j c_{i,j}}{(\sigma_{s_j}^2+\sigma_{i,j}^2)\eta_j}}-1\right)^+,\ \forall i\in\mathcal{I}_j, \tag{44}$$

$$\eta_j=\left(\frac{\sum_{\kappa\in\mathcal{A}_j}\frac{1}{\sigma_{\kappa,j}^2}\sqrt{\frac{\sigma_{s_j}^2+\sigma_{\kappa,j}^2}{\alpha P_\mathrm{t}\gamma_j c_{\kappa,j}}}}{1+\sum_{\kappa\in\mathcal{A}_j}\frac{\sigma_{s_j}^2+\sigma_{\kappa,j}^2}{\alpha P_\mathrm{t}\gamma_j c_{\kappa,j}\sigma_{\kappa,j}^2}}\right)^2, \tag{45}$$

$$\mathcal{A}_j=\Big\{i\in\mathcal{I}_j\Big|\frac{\alpha P_\mathrm{t}\gamma_j c_{i,j}}{(\sigma_{s_j}^2+\sigma_{i,j}^2)\eta_j}>1\Big\}. \tag{46}$$

For $P_\mathrm{t}\to\infty$, the solution for $\beta_{i,j}$'s given in (44)–(46) converges to

$$\lim_{P_\mathrm{t}\to\infty}\beta_{i,j}=\sqrt{\frac{\sigma_s^2+\sigma_{i,j}^2}{c_{i,j}\sigma_{i,j}^4}}\left(\sum_{l\in\mathcal{I}_j}\sqrt{\frac{\sigma_s^2+\sigma_{l,j}^2}{c_{l,j}\sigma_{l,j}^4}}\right)^{-1},\ \forall i\in\mathcal{I}_j. \tag{47}$$

From the implementation point of view, the CH j determines the Lagrange multiplier η_j and broadcasts its value to all sensors in the cluster. After knowing η_j, the sensors in the cluster j can calculate $\beta_{i,j}$'s by (44).

IV. Performance Evaluation

In this section, we corroborate the performance of the proposed power scheduling scheme with some simulation examples. For this purpose, we consider a WSN comprising $N_\mathrm{c}=16$ clusters with sizes $N_j=N_{j-1}+4$ for all $j\in\mathcal{J}$ and $N_0=0$. In each cluster, the sensors are randomly and uniformly distributed as in [9]. The correlation between the underlying sources of clusters j and k is modeled as $\rho_{s_j,s_k}=e^{-d_{j,k}^\mathrm{c}/\theta}$ for all j and k, where $d_{j,k}^\mathrm{c}$ denotes the CH-to-CH distance between cluster j and k; and $\theta>0$ is a scale parameter that controls how fast the correlation decays with distance. We assume $\sigma_{s_j}^2=1$ for all j. The CH selection criterion and the values of other system parameters (namely $\sigma_{i,j}^2$'s, $c_{i,j}$'s, and g_j's) are same as in [9].

We compare the distortion performance of the proposed adaptive power allocation (APA) design with a uniform power allocation (UPA) scheme. In the UPA scheme we have $\alpha=\alpha_\mathrm{u}=0.5$, $\xi_j=\xi_\mathrm{u}=1/N_\mathrm{c}$, $\gamma_j=\gamma_\mathrm{u}=1/N_\mathrm{c}$, and $\beta_{i,j}=\beta_{\mathrm{u}_j}=1/(N_j-1)$

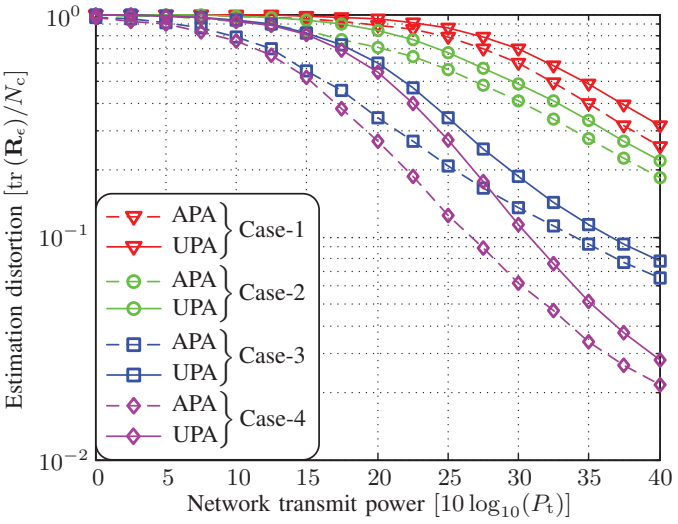

Fig. 2. Estimation Distortion comparison for correlated sources: (Case-1) $\theta = 5e1$, (Case-2) $\theta = 5e2$, (Case-3) $\theta = 5e3$, and (Case-4) $\theta = 5e4$.

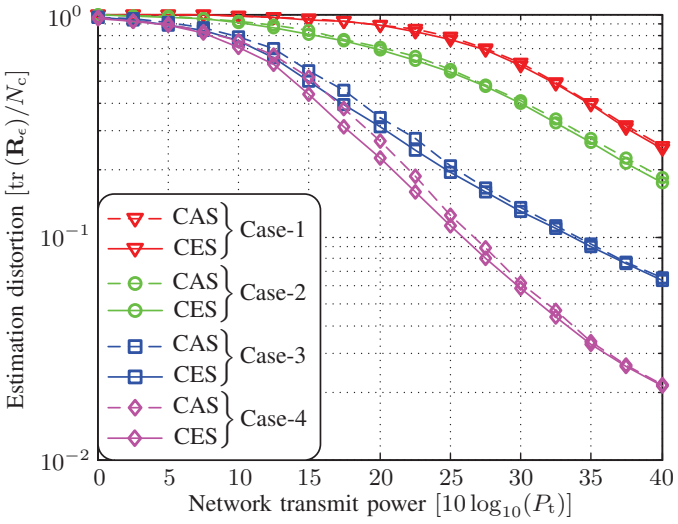

Fig. 3. Estimation Distortion comparison for correlated sources: (Case-1) $\theta = 5e1$, (Case-2) $\theta = 5e2$, (Case-3) $\theta = 5e3$, and (Case-4) $\theta = 5e4$.

for all $i \in \mathcal{I}_j$ and $j \in \mathcal{J}$. We also compare the performance of the proposed power scheduling scheme with the the scheme given in [9]. The results are averaged over 10^3 random deployment of the sensors in each cluster. Here in all simulations, we assume $\alpha_0 = 0$ and $\gamma_{\max}^{[j]} = 1$ for all j in $\mathcal{J}$.

Fig. 2 plots and compares the estimation distortion of the APA and UPA schemes as a function of P_t for different level of correlation between the underlying sources. The correlation increases with increasing value of θ. The figure shows that the distortion performance of the proposed APA scheme is considerably better than that of the UPA scheme and difference in performance of the two schemes increases as the level of correlation increases. Moreover, we can see that the performance of the APA scheme monotonically converges to the UPA scheme as P_t increases, which is typical of the power-constrained estimation schemes.

Next we compare the distortion achieved by the APA scheme proposed in this paper, which is based on an approximation of the distortion function by an upper bound, with the APA scheme we proposed in [9], which is based on the exact distortion function; here, the two schemes are denoted as CAS and CES, respectively. The results are plotted in Fig. 3 for different level of correlation between the sources. The figure shows that the CAS achieves distortion which is quite close to that achieved by the CES for a wide range of network transmit power and the correlation values. This observation illustrates the effectiveness of the APA scheme based on the distortion approximation vis-à-vis the scheme based on exact distortion function.

V. Concluding Remarks

In this work we derived an adaptive power scheduling strategy for distributed estimation of underlying correlated sources in cluster-based hierarchical WSNs. The proposed power scheduling scheme is based on the optimization problem where we target to minimize the estimation distortion subject to constraints on transmit power of the clusters as well as the network as a whole. Due to the correlation of the sources, the solution based on exact distortion function requires centralized scheduler. To this end, based on an upper-bound on the estimation distortion, we proposed an alternative solution, which bears favorable characteristics for distributed implementation and gives distortion performance that matches quite closely with the solution that relies on the exact formulation of estimation distortion. We also showed that compared to a uniform power allocation scheme, the proposed adaptive power allocation design gives better distortion performance.

References

[1] J.-J. Xiao, S. Cui, Z.-Q. Luo, and A. J. Goldsmith, "Power scheduling of universal decentralized estimation in sensor Networks," *IEEE Trans. on Signal Processing*, vol. 54, no. 2, pp. 413–422, Feb. 2006.

[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 Trans. on Signal Processing*, vol. 55, no. 9, pp. 4683–4695, Sept. 2007.

[3] M. H. Chaudhary and L. Vandendorpe, "Power constrained linear estimation in wireless sensor networks with correlated data and digital modulation," *IEEE Trans. on Signal Processing*, vol. 60, no. 2, pp. 570–584, Feb. 2012.

[4] M. H. Chaudhary and L. Vandendorpe "Adaptive power allocation in wireless sensor networks with spatially correlated data and analog modulation: perfect and imperfect CSI," *EURASIP Jr. on Wireless Commun. and Net.*, vol. 2010, Article ID 817961, 14 pages, 2010.

[5] Y. Huang and Y. Hua, "Energy planning for progressive estimation in multihop sensor networks," *IEEE Trans on Signal Processing*, vol. 57, no. 10, pp. 4052–4065, Oct. 2009.

[6] G. Thatte and U. Mitra, "Sensor selection and power allocation for distributed estimation in sensor networks: Beyond the star topology," *IEEE Trans. on Signal Processing*, vol. 56, no. 7, pp. 2649–2661, July 2008.

[7] J. Matamoros and C. Antón-Haro, "Hierarchical organizations of sensors for decentralized parameter estimation," in *Proc. of IEEE Int. Symp. on Signal Process. and Inf. Techn. (ISSPIT'07)*, Cairo, pp. 513–518, Dec. 2007.

[8] J. Li and G. AlRegib, "Energy-efficient cluster-based distributed estimation in wireless sensor networks," in *Proc. of the Conf. on Military Commun. (MILCOM'06)*, Washington, D.C., pp. 1–7, Oct. 2006.

[9] M. H. Chaudhary and L. Vandendorpe "Power constrained linear estimation of correlated sources in hierarchical wireless sensor networks," in *Proc. of IEEE 12th Int. Workshop on Signal Process. Advances in Wireless Commun. (SPAWC'11)*, San-Francisco, pp. 146–150, June 2011.

[10] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, Prentice-Hall, 1993.

[11] J. Hoydis, M. Kobayashi, and M. Debbah, "Green small-cell networks," *IEEE Vehicular Tech. Magazine*, vol. 6, no. 1, pp. 37–43, Mar. 2011.

[12] M. H. Chaudhary and L. Vandendorpe "Performance of power-constrained estimation in hierarchical wireless sensor networks," *submitted for publication in IEEE Trans. on Signal Processing*.

[13] D. P. Bertsekas, *Nonlinear Programming*, Athena Scientific, Belmont, Mass, USA, second edition, 1999.

[14] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2008.

[15] W. Dinkelbach, "On nonlinear fractional programming," *Management Science*, vol. 13, no. 7, pp. 492–498, 1967.