Performance of Power-Constrained Estimation in Hierarchical Wireless Sensor Networks

Muhammad Hafeez Chaudhary, *Member, IEEE,* and Luc Vandendorpe, *Fellow, IEEE*

***Abstract*—In this paper we study estimation in a power-constrained wireless sensor network, where the network is divided into disjoint groups called clusters. The sensors in each cluster observe a random source that is correlated with the sources being observed by the other clusters. Each cluster has a designated cluster head (CH). Estimation of the sources is performed in two time slots: In the first slot, the sensors in each cluster amplify and forward their noisy measurements to the CH that forms a preliminary estimate of the underlying source; and in the second slot, the CHs send a scaled version of their partial estimates to a remote fusion center (FC) that forms the final estimate of the sources. The CHs and the FC use minimum mean square error estimation rule. To minimize the overall estimation distortion, we propose a power scheduling scheme which allocates power to the sensors and the CHs subject to constraints on the transmit powers of the individual clusters and the overall network. We show that when the sources are fully uncorrelated or fully correlated then the solution to the power allocation problem has a computationally favorable structure and is amenable for distributed implementation. However, the partial correlation between the 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 shares characteristics of the fully uncorrelated and correlated cases. Simulation examples illustrate the effectiveness of the proposed power scheduling scheme.**

***Index Terms*—Cluster-based WSNs, correlation, parameter estimation, power scheduling, resource management.**

I. Introduction

SPURRED by the ease of deployment provided by the wireless communication paradigm, wireless sensor networking is an emerging technology which finds application in many fields [1]. A wireless sensor network (WSN) consists of spatially distributed sensors that cooperatively monitor physical or environmental conditions. The sensors are usually battery powered that can provide limited sensing, communication, and computational functionalities. A significant research has focused on developing distributed data processing and cooperative communication strategies in the context of realizing energy-efficient sensor networks with acceptable sensing capabilities. In this work, we study the problem of power-constrained estimation in cluster-based WSNs where our objective is to minimize a distortion measure subject to constraints on the transmit powers.

In recent years, for WSNs, several energy-efficient estimation algorithms have been proposed under a wide variety of network models. For instance, [2]–[4] consider estimation based on quantized sensor observations. In [5], the focus is on designing a power allocation scheme where sensors amplify and transmit their analog observations. The estimation schemes in these works target the estimation of an unknown deterministic parameter. The work in [6] proposed a power allocation scheme for estimation that also takes into account the power consumed in estimating the channels from the sensors to the fusion center (FC). The works of [7]–[10] studied power allocation in sensor networks with spatially correlated data. In all these aforementioned works, individual sensors send their observations, via single hop, to a centralized unit which forms the estimate of the underlying source. This so-called centralized network topology is not favorable from the perspective of energy-efficient estimation. Such a topology may also pose a challenge in medium access scheduling—specifically for networks with large number of sensors—because it would not allow spatial reuse of spectrum resources among sensors. In this regard, [11] investigated minimal energy progressive estimation in sensor networks and [12] studied estimation under different network topological settings. In [13] and [14] power allocation schemes are proposed for estimation in cluster-based wireless sensor networks. The power scheduling schemes in all these works target estimation of a homogeneous unknown deterministic parameter and neglect the effect of data correlation.

Recently [15], [16] proposed a power scheduling scheme that minimizes outage probability of estimation distortion in cluster-based WSN observing spatially homogeneous source. However, therein, optimized power allocation to cluster-heads (CHs) is studied whereas power among the sensors in each cluster is distributed uniformly. The authors of [17] proposed an estimation scheme in cluster-based WSN in which they target dimensionality reduction of the observations at each CH and optimize power to transmit the compressed observations to the FC. The authors, however, assumed ideal communication channels from sensors to their respective CHs. This assumption of ideal communication channels, though simplifies the design, is not a reasonable assumption. In any real world application, the communication channel from each sensor to its CH will experience path-loss and receiver noise, and may

The authors are with the Institute of Information and Communication Technologies, Electronics and Applied Mathematics (ICTEAM), 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 project, the Walloon region framework program COSMOS/TSARINE, and EU project FP7 NEWCOM++ for the financial support and the scientific inspiration.

The material in this paper is partly presented in SPAWC 2011 [31] and ICC 2012 [32].

also be subject to multi-path fading and shadowing. In the realm of energy-efficient estimation in sensor networks, it is imperative to account for the channel imperfections when designing a power allocation scheme. In recent past, some studies emerged about clustering and route optimization for correlated data gathering in WSNs, for example, see [18]–[20] and references therein. However, these studies do not consider power scheduling for energy-efficient estimation.

The existing energy-aware or power-constrained estimation algorithms ignore the effect of data correlation in sensor networks, while in this paper we consider a network that is partitioned into clusters, where each cluster observes a separate source albeit correlated with the sources being observed by the other clusters. The estimation of the underlying sources is performed in two steps. In the first step, the sensors in each cluster amplify and forward their noisy observations to their respective cluster head (CH) that forms a preliminary estimate of the underlying source. Subsequently, in the second step, the CHs amplify and forward their partial estimates to a remote FC that forms the final estimate of the sources. To form the estimates, both the CHs and the FC employ the minimum mean square error (MMSE) estimation rule. Communication between the sensors in each cluster and their respective CH, and similarly between the CHs and the FC take place over orthogonal multiple access channels.

From the perspective of energy-efficient estimation, uniform power allocation (UPA) is not an optimal strategy due to the variability of the quality of observations at the sensors, the channel gains between the sensors and the CHs, the channel gains between the CHs and the FC, the correlation structure, and the cluster sizes. Towards this end, in this work we propose an adaptive power allocation (APA) design that takes into account all these factors in allocating power to the sensors and the CHs, and gives distortion performance better than the UPA scheme. Furthermore, compared to a centralized WSN where all sensors send their observations directly to the FC, the proposed cluster-based WSN performs better in terms of estimation distortion. The power allocation design is based on an optimization problem where we target to minimize the overall estimation distortion subject to constraints on transmit power of the individual clusters and the network as a whole. We formulate the power allocation problem as a convex optimization problem and outline its solution using a block coordinate descent method (BCoDM) based approach.

We show that for the case of fully uncorrelated (and likewise for the case of fully correlated) sources the solution to the power allocation problem embodies favorable structure from the point of view of computational cost and is amendable for distributed implementation. Specifically, we show that the underlying optimization problem can be decomposed into simpler problems, which can be conveniently solved either analytically or numerically. The resulting solution shows separable structure along the clusters as well as the sensors. On the other hand, for the case when sources are neither fully uncorrelated nor fully correlated the solution to the power allocation problem needs to be computed numerically in a centralized way. The computational cost and the control overhead associated with the centralized solution may become prohibitive, in particular, for a large size network. In an effort to cut corners, we develop an upper-bound for the distortion function and then solve the optimization problem with that bound as a surrogate for the distortion. The resulting solution shares favorable properties as exhibited by the solution of the uncorrelated case (and likewise the correlated case). The proposed power allocation design, in all cases, shows significant performance gain compared to a power allocation design based on the UPA scheme. Moreover, energy efficiency comparison with the centralized network topology shows that the proposed cluster-based network scheme gives better performance.

Remainder of the article is organized as follows. Section II presents the system model, formulates the optimization problem, and introduces the adopted approach to solve it. Section III outlines the solution for the case of uncorrelated sources. Section IV outlines the solution for correlated sources: Section IV-A and Section IV-B present solutions based on the exact distortion function and the upper-bound of the distortion, respectively. Section V outlines the solution for the case of fully correlated sources. Section VI presents simulation examples. Finally, Section VII gives some concluding remarks.

II. System Model and Preliminaries

We consider a hierarchical sensor network shown in Fig. 1, where N_0' spatially distributed sensor nodes are divided into N_c disjoint and non-overlapping clusters, indexed by $\mathcal{J} = \{1, \ldots, N_\text{c}\}$ such that $N_0' = \sum_{j\in\mathcal{J}} N_j$. Where N_j is the number of sensors in cluster j, indexed by $\tilde{\mathcal{I}}_j = \{1, \ldots, N_j\}$. We assume number of clusters and distribution of sensors in the clusters as given, and we study the problem of power allocation for energy-efficient estimation. The clusters observe random Gaussian sources, $s_j \sim \mathcal{N}(0, \sigma_{s_j}^2)$ for $j \in \mathcal{J}$, that are correlated such that $\text{Cov}\{s_j, s_k\} = \sigma_{s_j}\sigma_{s_k}\rho_{s_j,s_k}$, where ρ_{s_j,s_k} specifies the correlation between s_j and s_k for all j and k in $\mathcal{J}$. Specifically, we assume that the sources are jointly Gaussian distributed. The observation at each sensor is corrupted by observation noise, which is independent of the underlying sources s_j's and the observation noises across sensors. The noisy observation at sensor i in cluster j is $x_{i,j} = s_j + n_{i,j}$ for all $j \in \mathcal{J}$ and $i \in \tilde{\mathcal{I}}_j$. Where $n_{i,j} \sim \mathcal{N}(0, \sigma_{n_{i,j}}^2)$ denotes the observation noise. By allowing the variance of the observation noise to vary across sensors, we can model a scenario where observation channels from the sources to the sensors have different quality across sensors. In a special case of $N_j = 1$ for all j—that is, each cluster comprises only one sensor—the system model converges to the case discussed in [7]. For the same single sensor per cluster setup, if $s_j = s$ for all j (i.e., a spatially homogeneous source), the system model converges to the case discussed in [5], [6]. In this work, however, we investigate a general case, in which $N_j \geq 1$ for all clusters, that encompasses [5]–[7] as special cases.

The estimation problem we study here essentially corresponds to estimation of a spatial random field, where we are interested in the field values in each cluster. As an example application, we can view the sensor network as deployed to observe a Gaussian spatial random field. We assume that the inter-sensor distances within each cluster are small compared

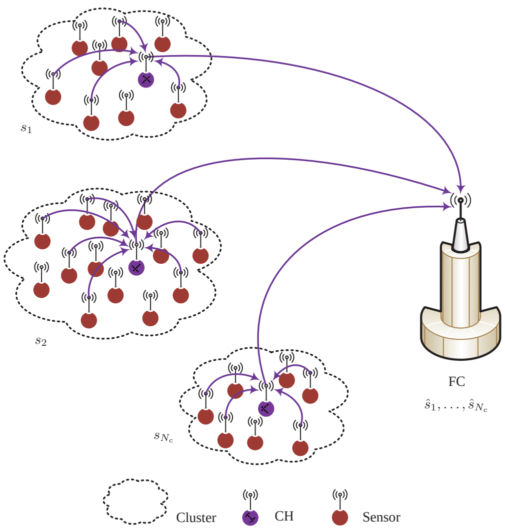

Fig. 1. Architecture of the hierarchical sensor network.

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 the long inter-cluster distances suggest heterogeneous correlation among the field values in different clusters. Note that by allowing the number of sensors to vary across clusters we model the general setup, which gives flexibility to observe the field in certain areas with high fidelity.

We consider estimation on a *sample-by-sample* basis, that is, we do not study temporal dependencies of the observations. In our future work, however, we will investigate such dependencies. The estimation of the sources is performed in two phases. In the first phase of an estimation cycle, the sensors in each cluster amplify their observations and then transmit to their respective CH such that the received observations are $y_{i,j} = \sqrt{\phi_{i,j}\tilde{c}_{i,j}}(s_j + n_{i,j}) + w_{i,j}$ for all $j \in \mathcal{J}$ and $i \in \mathcal{I}_j$, with $\mathcal{I}_j = \tilde{\mathcal{I}}_j - \{N_j\}$. Where $\phi_{i,j} \in [0,\infty)$ is a scaling or an amplifying factor, $\tilde{c}_{i,j}$ is gain of the channel between the sensor and the CH, and $w_{i,j} \sim \mathcal{N}(0,\sigma^2_{w_{i,j}})$ is the receiver noise. The noise is assumed to be independent of s_j and $n_{i,j}$ for all i and j. Moreover, the noise is assumed to be independent (of the noises) across sensors in all clusters. Here without any loss of generality we have assumed the sensor N_j as the designated 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 [21]. Using matrix–vector notation, we can write the received signals at CH j in a compact form as $\mathbf{y}_j = \mathbf{b}_j s_j + \mathbf{f}_j$, where $\mathbf{y}_j = [y_{1,j}\ldots,y_{N_j,j}]^T$, $\mathbf{b}_j = \left[\sqrt{\phi_{1,j}\tilde{c}_{1,j}},\ldots,\sqrt{\phi_{N_j-1,j}\tilde{c}_{N_j-1,j}},1\right]^T$, and $\mathbf{f}_j = \left[\sqrt{\phi_{1,j}\tilde{c}_{1,j}}n_{1,j}+w_{1,j},\ldots,\sqrt{\phi_{N_j-1,j}\tilde{c}_{N_j-1,j}}n_{N_j-1,j}+w_{N_j-1,j},n_{N_j,j}\right]^T$.

Employing the MMSE estimation rule [22], the CH j forms an estimate $\hat{s}_j$ of the source s_j based on the received observations. By defining $\mathbf{R}_{\mathbf{f}_j} := \mathbb{E}[\mathbf{f}_j\mathbf{f}_j^T]$, we can write $\hat{s}_j$ as

$$\begin{aligned}\hat{s}_j &= \mathbb{E}\left[s_j\mathbf{y}_j^T\right]\left(\mathbb{E}\left[\mathbf{y}_j\mathbf{y}_j^T\right]\right)^{-1}\mathbf{y}_j\\ &= \sigma^2_{s_j}\mathbf{b}_j^T\left(\sigma^2_{s_j}\mathbf{b}_j\mathbf{b}_j^T+\mathbf{R}_{\mathbf{f}_j}\right)^{-1}\mathbf{y}_j \overset{(a)}{=} \frac{\mathbf{b}_j^T\mathbf{R}_{\mathbf{f}_j}^{-1}\mathbf{y}_j}{1/\sigma^2_{s_j}+\mathbf{b}_j^T\mathbf{R}_{\mathbf{f}_j}^{-1}\mathbf{b}_j}\\ &= \left(\frac{1}{\sigma^2_{s_j}}+\frac{1}{\sigma^2_{n_{N_j}}}+\sum_{i\in\mathcal{I}_j}\frac{\phi_{i,j}\tilde{c}_{i,j}}{\phi_{i,j}\tilde{c}_{i,j}\sigma^2_{n_{i,j}}+\sigma^2_{w_{i,j}}}\right)^{-1}\\ &\qquad\left(\frac{x_{N_j}}{\sigma^2_{n_{N_j}}}+\sum_{i\in\mathcal{I}_j}\frac{\sqrt{\phi_{i,j}\tilde{c}_{i,j}}y_{i,j}}{\phi_{i,j}\tilde{c}_{i,j}\sigma^2_{n_{i,j}}+\sigma^2_{w_{i,j}}}\right),\end{aligned} \tag{1}$$

where equality (a) follows from the Woodbury identity or the matrix-inversion lemma [23]. The associated mean square estimation error $D_j := \mathbb{E}[(\hat{s}_j - s_j)^2]$ is given by

$$\begin{aligned}D_j &= \mathbb{E}\left[s_j^2\right]-\mathbb{E}\left[s_j\mathbf{y}_j^T\right]\left(\mathbb{E}\left[\mathbf{y}_j\mathbf{y}_j^T\right]\right)^{-1}\mathbb{E}\left[s_j\mathbf{y}_j\right]\\ &= \sigma^2_{s_j}-\sigma^4_{s_j}\mathbf{b}_j^T\left(\sigma^2_{s_j}\mathbf{b}_j\mathbf{b}_j^T+\mathbf{R}_{\mathbf{f}_j}\right)^{-1}\mathbf{b}_j\\ &= \left(\frac{1}{\sigma^2_{s_j}}+\frac{1}{\sigma^2_{N_j}}+\sum_{i\in\mathcal{I}_j}\frac{\phi_{i,j}c_{i,j}}{\phi_{i,j}c_{i,j}\sigma^2_{i,j}+1}\right)^{-1},\end{aligned} \tag{2}$$

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.

Let $\tilde{\sigma}^2_j$ be defined as

$$\tilde{\sigma}^2_j = \left(\frac{1}{\sigma^2_{N_j}}+\sum_{i\in\mathcal{I}_j}\frac{\phi_{i,j}c_{i,j}}{\phi_{i,j}c_{i,j}\sigma^2_{i,j}+1}\right)^{-1}. \tag{3}$$

Now, let v_j be a scaled version of $\hat{s}_j$ as $v_j = D_j\tilde{\sigma}^2_j\hat{s}_j$. The v_j can be written as follows: $v_j = s_j + \vartheta_j$ for all $j \in \mathcal{J}$, where $\vartheta_j = \tilde{\sigma}^2_j\left(\frac{n_{N_j,j}}{\sigma^2_{N_j}}+\sum_{i\in\mathcal{I}_j}\frac{\phi_{i,j}\tilde{c}_{i,j}n_{i,j}+\sqrt{\phi_{i,j}\tilde{c}_{i,j}}w_{i,j}}{\phi_{i,j}\tilde{c}_{i,j}\sigma^2_{n_{i,j}}+\sigma^2_{w_{i,j}}}\right)$. Here, it is fairly straightforward to show that $\vartheta_j \sim \mathcal{N}(0,\tilde{\sigma}^2_j)$. Moreover, ϑ_j is independent of s_j for all j and is independent of ϑ_k for all $k \neq j$. We can view v_j as an equivalent observation at the CH with ϑ_j as the equivalent observation noise. Expressing $\hat{s}_j$ by v_j simplifies ensuing formulation of the final estimate at the FC and thereby helps in solving the optimization problem in later development.

In the second phase of the estimation cycle, the CHs amplify and transmit the equivalent observations v_j's to the FC such that the received observations are $z_j = \sqrt{\psi_j\tilde{g}_j}(s_j+\vartheta_j)+w_j$ for all $\mathcal{J}$. Where $\psi_j \in [0,\infty)$ is an amplifying factor, $\tilde{g}_j$ is gain of the channel between the CH j and the FC, and $w_j \sim \mathcal{N}(0,\sigma^2_{w_j})$ is the receiver noise at the FC, which is independent of s_j, ϑ_j, and w_k for all j and k with $k \neq j$. By defining $\mathbf{z} = [z_1,\ldots,z_{N_c}]^T$, $\mathbf{s} = [s_1,\ldots,s_{N_c}]^T$, $\tilde{\mathbf{H}} = \operatorname{diag}\left(\sqrt{\psi_1\tilde{g}_1},\ldots,\sqrt{\psi_{N_c}\tilde{g}_{N_c}}\right)$, and $\mathbf{r} = \left[\sqrt{\psi_1\tilde{g}_1}\vartheta_1+w_1,\ldots,\sqrt{\psi_{N_c}\tilde{g}_{N_c}}\vartheta_{N_c}+w_{N_c}\right]^T$, the received observations at the FC can be written in a compact form as $\mathbf{z} = \tilde{\mathbf{H}}\mathbf{s}+\mathbf{r}$ where $\mathbf{s}$ is the underlying source vector to be estimated. Now based on the received signals and employing the MMSE estimation rule, the FC forms an estimate of the source vector which is given by

$$\hat{\mathbf{s}} = \mathbf{R}_{\mathbf{sz}}\mathbf{R}_{\mathbf{z}}^{-1}\mathbf{z} = \mathbf{R}_{\mathbf{s}}\tilde{\mathbf{H}}^T\left(\tilde{\mathbf{H}}\mathbf{R}_{\mathbf{s}}\tilde{\mathbf{H}}^T+\tilde{\mathbf{R}}\right)^{-1}\mathbf{z}, \tag{4}$$

where $\mathbf{R}_{\mathbf{sz}} := \mathbb{E}\left[\mathbf{sz}^T\right]$, $\mathbf{R}_{\mathbf{z}} := \mathbb{E}\left[\mathbf{zz}^T\right]$, $\mathbf{R}_{\mathbf{s}} := \mathbb{E}\left[\mathbf{ss}^T\right]$, and $\tilde{\mathbf{R}} := \mathbb{E}\left[\mathbf{rr}^T\right]$. By defining $\epsilon := \mathbf{s} - \hat{\mathbf{s}}$ as the estimation error vector, we can write its covariance in the following form:

$$\begin{aligned}\tilde{\mathbf{R}}_\epsilon &= \mathbb{E}\left[\epsilon\epsilon^T\right] = \mathbf{R}_{\mathbf{s}} - \mathbf{R}_{\mathbf{sz}}\mathbf{R}_{\mathbf{z}}^{-1}\mathbf{R}_{\mathbf{sz}}^T \\ &= \mathbf{R}_{\mathbf{s}} - \mathbf{R}_{\mathbf{s}}\tilde{\mathbf{H}}^T\left(\tilde{\mathbf{H}}\mathbf{R}_{\mathbf{s}}\tilde{\mathbf{H}}^T + \tilde{\mathbf{R}}\right)^{-1}\tilde{\mathbf{H}}\mathbf{R}_{\mathbf{s}}^T \\ &= \left(\tilde{\mathbf{H}}\tilde{\mathbf{R}}^{-1}\tilde{\mathbf{H}}^T + \mathbf{R}_{\mathbf{s}}^{-1}\right)^{-1}, \end{aligned} \tag{5}$$

where the last equality follows from the matrix-inversion lemma.

A. Formulation of the Optimization Problem

In order to allocate power to the sensors and the CHs we consider the following optimization problem:

$$\begin{aligned}\underset{\psi_j \geq 0,\, \phi_{i,j} \geq 0,\, \forall i,j}{\text{minimize}} \quad & \operatorname{tr}\left(\tilde{\mathbf{R}}_\epsilon\right) \\ \text{subject to} \quad & \sum_{j\in\mathcal{J}}\Big(\psi_j\left(\sigma_{s_j}^2 + \tilde{\sigma}_j^2\right) + \sum_{i\in\mathcal{I}_j}\phi_{i,j}\left(\sigma_{s_j}^2 + \sigma_{i,j}^2\right)\Big) \leq P_{\mathrm{t}}, \\ & \sum_{j\in\mathcal{J}}\psi_j\left(\sigma_{s_j}^2 + \tilde{\sigma}_j^2\right) \leq \psi_{\max}, \\ & \sum_{i\in\mathcal{I}_j}\phi_{i,j}\left(\sigma_{s_j}^2 + \sigma_{i,j}^2\right) \leq \phi_{\max}^{(j)}, \quad \forall j \in \mathcal{J}, \end{aligned} \tag{6}$$

where we target to minimize the sum of mean-square estimation errors of the underlying sources. In (6), the constraint on the total network power (first constraint) enables a fair comparison between the networks of different sizes. Moreover, putting a cap on the total network power consumption conserves energy, which makes sense from the viewpoint of global energy efficiency and to realize green information and communication technology [24], [25]. The second and third constraints of the optimization problem limit inter-cluster interference and interference with any other network in the neighborhood, which is important from the perspective of spatial reuse of the 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 the individual clusters and the CHs.

Note that the optimization variables ψ_j and $\phi_{i,j}$ are coupled in the constraints. This coupling of the variables and the fact that the optimization problem (6) is jointly non-convex over the optimization variables make the problem difficult to solve. To this end, in Prop. 1 we reformulate the problem in an alternative form that bears favorable characteristics, as we shall see in the subsequent 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 $\sum_{j\in\mathcal{J}}\xi_j \leq 1$, $\sum_{j\in\mathcal{J}}\gamma_j \leq 1$, and $\sum_{i\in\mathcal{I}_j}\beta_{i,j} \leq 1$, we can write

$$\begin{aligned}\psi_j &= \frac{(1-\alpha)P_{\mathrm{t}}\xi_j}{\sigma_{s_j}^2 + \tilde{\sigma}_j^2}, \quad \forall j \in \mathcal{J}, \\ \phi_{i,j} &= \frac{\alpha P_{\mathrm{t}}\gamma_j\beta_{i,j}}{\sigma_{s_j}^2 + \sigma_{i,j}^2}, \quad \forall j \in \mathcal{J},\, i \in \mathcal{I}_j. \end{aligned}$$

Here ξ_j's, γ_j's, and $\beta_{i,j}$'s respectively decide the power split among the CHs, the clusters, and the sensors in each cluster. With this we can write the problem (6) in the following form:

$$\begin{aligned}\underset{\alpha,\, \xi_j \geq 0,\, \gamma_j \geq 0,\, \beta_{i,j} \geq 0,\, \forall i,j}{\text{minimize}} \quad & \operatorname{tr}\left(\mathbf{R}_\epsilon\right) \\ \text{subject to } \alpha \in \mathcal{T},\ & \sum_{j\in\mathcal{J}}\xi_j \leq 1,\ \sum_{j\in\mathcal{J}}\gamma_j \leq 1, \\ & \sum_{i\in\mathcal{I}_j}\beta_{i,j} \leq 1,\ \gamma_j \leq \gamma_{\max}^{(j)}, \quad \forall j \in \mathcal{J}, \end{aligned} \tag{7}$$

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

$$\begin{aligned}\mathbf{R}_\epsilon &= \mathbf{R}_{\mathbf{s}} - \mathbf{R}_{\mathbf{s}}\mathbf{H}^T\left(\mathbf{H}\mathbf{R}_{\mathbf{s}}\mathbf{H}^T + \mathbf{R}\right)^{-1}\mathbf{H}\mathbf{R}_{\mathbf{s}}^T \\ &= \left(\mathbf{H}\mathbf{R}^{-1}\mathbf{H}^T + \mathbf{R}_{\mathbf{s}}^{-1}\right)^{-1}, \end{aligned} \tag{8}$$

with

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

Where, for all $j \in \mathcal{J}$, $g_j = \tilde{g}_j/\sigma_{w_j}^2$ and

$$\sigma_j^2 = \left(\frac{1}{\sigma_{N_j}^2} + \sum_{i\in\mathcal{I}_j}\frac{\alpha P_{\mathrm{t}}\gamma_j\beta_{i,j}c_{i,j}}{\alpha P_{\mathrm{t}}\gamma_j\beta_{i,j}c_{i,j}\sigma_{i,j}^2 + \sigma_{i,j}^2 + \sigma_{s_j}^2}\right)^{-1}. \tag{9}$$

Proof: See Appendix A. □

The alternative formulation in (7) has linear constraints and the constraints are independent in the sense that each constraint function depends on a separate set of optimization variables (namely α, ξ_j's, γ_j's, or $\beta_{i,j}$'s). The independence of the constraints is a nice property that allows us to divide the problem into subproblems, where in each subproblem we can optimize over a separate set of optimization variables, as outlined in the ensuing development. There we shall see that each of these subproblems is jointly convex over the given set of optimization variables.

B. BCoDM based Algorithm for Power Allocation

The solution to the power allocation problem (7) will be obtained using a BCoDM based approach, which cyclically/iteratively minimizes the cost function $\operatorname{tr}\left(\mathbf{R}_\epsilon\right)$ with respect to (w.r.t.) each set of optimization variables subject to the associated constraints while the other optimization variables are held fixed. Specifically to solve (7) do the following:

1: Initialize α, ξ_j's, γ_j's, and $\beta_{i,j}$'s in their respective feasible region.
2: For given ξ_j's, γ_j's, and $\beta_{i,j}$'s, find α by solving

$$\underset{\alpha\in\mathcal{T}}{\text{minimize}}\ \operatorname{tr}\left(\mathbf{R}_\epsilon\right). \tag{10}$$

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

$$\underset{\xi_j \geq 0, \forall j}{\text{minimize}} \ \text{tr}\left(\mathbf{R}_\epsilon\right) \ \text{subject to} \sum_{j \in \mathcal{J}} \xi_j \leq 1. \tag{11}$$

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

$$\begin{aligned} &\underset{\gamma_j \geq 0, \forall j}{\text{minimize}} \ \text{tr}\left(\mathbf{R}_\epsilon\right) \\ &\text{subject to} \sum_{j \in \mathcal{J}} \gamma_j \leq 1, \, \gamma_j \leq \gamma_{\max}^{[j]}, \quad \forall j \in \mathcal{J}. \end{aligned} \tag{12}$$

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

$$\underset{\beta_{i,j} \geq 0, \forall i}{\text{minimize}} \ \text{tr}\left(\mathbf{R}_\epsilon\right) \ \text{subject to} \sum_{i \in \mathcal{I}_j} \beta_{i,j} \leq 1. \tag{13}$$

6: Repeat step 2 to step 5 until there is no appreciable decrease in the objective function.

The given algorithm is guaranteed to converge to a minimum point of the optimization problem (7) as stated next.

***Proposition** 2:* Let κ be the iteration index of the BCoDM algorithm. For any feasible initialization point $\alpha^{(0)}$, $\xi_j^{(0)}$'s, $\gamma_j^{(0)}$'s, and $\beta_{i,j}^{(0)}$'s, the iterates $\alpha^{(\kappa)}$, $\xi_j^{(\kappa)}$'s, $\gamma_j^{(\kappa)}$'s, and $\beta_{i,j}^{(\kappa)}$'s generated by the BCoDM algorithm converge monotonically to a minimum point of the problem (7).

Proof: The proof is based on the result of Prop. 2.7 in [28], which proves the convergence of BCoDM provided the minimum in each of the optimization problems (10) through (13) is uniquely attained. Towards this end, in the ensuing sections, we shall prove that each of these subproblems is jointly convex over the respective set of optimization variables and thus has a unique minimum. The minimum can be obtained by using tools from the convex optimization theory. With that we conclude the convergence proof of the BCoDM algorithm. □

Given that each step of the algorithm minimizes a convex function by solving a convex optimization problem. Therefore, the distortion function $\text{tr}\left(\mathbf{R}_\epsilon\right)$ decreases monotonically from one iteration to the other of the algorithm. In the sequel, we outline solution of the optimization problem using the partitioning approach (of diving the problem into subproblems via BCoDM) for three distinct cases: Where the underlying sources s_j's are uncorrelated, partially correlated, and fully correlated.

III. Uncorrelated Sources

For uncorrelated sources s_j's, the covariance matrix $\mathbf{R_s}$ is diagonal, that is, $\mathbf{R_s} = \text{diag}\left(\sigma_{s_1}^2, \ldots, \sigma_{s_{N_c}}^2\right)$. In this particular case, we can write

$$\text{tr}\left(\mathbf{R}_\epsilon\right) = \sum_{j \in \mathcal{J}} \left(\sigma_{s_j}^2 - \frac{\sigma_{s_j}^4}{\sigma_{s_j}^2 + \sigma_j^2} \frac{(1-\alpha) P_{\text{t}} \xi_j g_j}{(1-\alpha) P_{\text{t}} \xi_j g_j + 1} \right).$$

The uncorrelated case is of particular interest because in this case the solution of the problems (10) to (13) can be implemented in a distributed way in a certain sense. We use the solution proposed for this case as a baseline and later on show that even in case of partially and fully correlated sources, the optimization problems could be solved by reverting to the techniques outlined under the uncorrelated case.

A. Optimization of α

For optimization over α we need to solve the optimization problem (10). To this end, let $f(\alpha) = \text{tr}\left(\mathbf{R}_\epsilon\right)$. The function $f(\alpha)$ is strictly convex over $\mathcal{T}$ and consequently has a unique global minimizer in $\mathcal{T}$, see Appendix B. The convexity of $f(\alpha)$ over $\mathcal{T}$ means the following condition is both necessary and sufficient for $\alpha^* \in \mathcal{T}$ to minimize $f(\alpha)$ over $\mathcal{T}$ (cf. Prop. 2.1.2 in [28]):

$$\frac{\partial f(\alpha^*)}{\partial \alpha}(\alpha - \alpha^*) \geq 0, \quad \forall \alpha \in \mathcal{T}. \tag{14}$$

For $\alpha^* \neq \alpha_0$, the condition (14) reduces to $\partial f(\alpha^*)/\partial \alpha = 0$. 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, for example, the Golden Section method [28]. Thanks to the convexity of $f(\alpha)$ over $\mathcal{T}$, the convergence of these numerical methods to α^* is guaranteed.

Remark 1: Note that $f(\alpha) = \sum_{j \in \mathcal{J}} f_j(\alpha)$, with $f_j(\alpha) = \sigma_{s_j}^2 - \frac{\sigma_{s_j}^4}{\sigma_{s_j}^2 + \sigma_j^2} \frac{(1-\alpha) P_{\text{t}} \xi_j g_j}{(1-\alpha) P_{\text{t}} \xi_j g_j + 1}$, has a separable structure along the clusters where factor $f_j(\alpha)$ depends on parameters concerning the cluster j. As a perspective on implementing the solution, 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 the stopping criterion is satisfied.

B. Optimization of ξ_j's

The optimization of ξ_j's is based on the problem (11). For this purpose, as shown in Appendix C, the function $\text{tr}\left(\mathbf{R}_\epsilon\right)$ is decreasing w.r.t. ξ_j's and the given optimization problem is jointly convex over ξ_j's. The convexity of the problem means the Karush-Kuhn-Tucker (KKT) conditions are sufficient for optimality of the solution for ξ_j's [26]. Solving the KKT conditions we get

$$\xi_j = \frac{1}{(1-\alpha) P_{\text{t}} g_j} \left(\sigma_{s_j}^2 \sqrt{\frac{(1-\alpha) P_{\text{t}} g_j}{(\sigma_{s_j}^2 + \sigma_j^2)\lambda}} - 1 \right)^+, \quad \forall j, \tag{15}$$

where $(x)^+ = \max\{0, x\}$ and λ is a Lagrange multiplier associated with the sum-constraint. Because the objective function is a convex decreasing function, therefore the optimum solution is at the constraint boundary, that is, the sum-constraint is always active. Consequently the multiplier λ should be determined so that it satisfies the constraint with equality, that is, $\sum_{j \in \mathcal{K}} \xi_j = 1$ which gives

$$\lambda = \left(\frac{\sum_{k \in \mathcal{K}} \frac{\sigma_{s_k}^2}{\sqrt{(1-\alpha) P_{\text{t}} g_k (\sigma_{s_k}^2 + \sigma_k^2)}}}{1 + \sum_{k \in \mathcal{K}} \frac{1}{(1-\alpha) P_{\text{t}} g_k}} \right)^2, \tag{16}$$

where $\mathcal{K} = \left\{ j \in \mathcal{J} \middle| \frac{(1-\alpha) P_{\text{t}} g_j \sigma_{s_j}^4}{(\sigma_{s_j}^2 + \sigma_j^2)\lambda} > 1 \right\}$. From (15) and (16), it is fairly simple to show that the solution for ξ_j's converges to

$$\lim_{P_{\rm t}\to\infty} \xi_j = \frac{\sigma_{s_j}^2}{\sqrt{g_j(\sigma_{s_j}^2+\sigma_j^2)}} \left(\sum_{k\in\mathcal{J}} \frac{\sigma_{s_k}^2}{\sqrt{g_k(\sigma_{s_k}^2+\sigma_k^2)}} \right)^{-1}. \quad (17)$$

Remark 2: The structure of the solution for ξ_j's is same as the power allocation solution in [5], [6] for estimation in networks comprising single sensor per cluster and all observing the same source. From an implementation point of view, the computation of ξ_j's can be done via a coordination mechanism where the FC determines the value of λ and broadcasts it to the CHs, which then calculate ξ_j's by (15). Note that for given λ and $(1-\alpha)P_{\rm t}$, the expression (15) depends on the local information available at each CH.

C. Optimization of γ_j's

In order to optimize γ_j's, we have the optimization problem given in (12). On the same lines as in Appendix C, we can show that the objective function is decreasing w.r.t. γ_j's and the problem is jointly convex over γ_j's. In the optimization problem as we are minimizing a decreasing function, therefore, the optimum solution is always at the boundary of the constraints set. In the optimization, one of the following three scenarios may arise. Firstly, 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. Secondly, 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 last case, we proceed as follows: Initially, we ignore the individual-constraints and solve the problem with only the sum-constraint. Afterwards. later in this section, we shall show how to incorporate the individual-constraints into the solution obtained with only the sum-constraint.

For solution to the problem (12) without considering the constraints on individual γ_j's, we propose a primal–dual type algorithm based on the Lagrangian dual-decomposition approach [27]. For this purpose, we can write the Lagrange function associated with the problem (called the primal problem) as follows:

$$\begin{aligned} \Lambda(\gamma_1,\ldots,\gamma_{N_{\rm c}};\mu) &= \operatorname{tr}\left(\mathbf{R}_\epsilon\right) + \mu\Bigg(\sum_{j\in\mathcal{J}} \gamma_j - 1\Bigg) \\ &= \sum_{j\in\mathcal{J}} \left[\sigma_{s_j}^2 - \frac{\sigma_{s_j}^4}{\sigma_{s_j}^2+\sigma_j^2} \frac{(1-\alpha)P_{\rm t}\xi_j g_j}{(1-\alpha)P_{\rm t}\xi_j g_j + 1} + \mu\gamma_j\right] - \mu, \\ &= \sum_{j\in\mathcal{J}} \Lambda_j(\gamma_j,\mu) - \mu. \end{aligned} \quad (18)$$

where μ is a Lagrange multiplier (also known as the dual variable or the price value) associated with the constraint $\sum_{j\in\mathcal{J}} \gamma_j \leq 1$. The corresponding Lagrangian-dual function

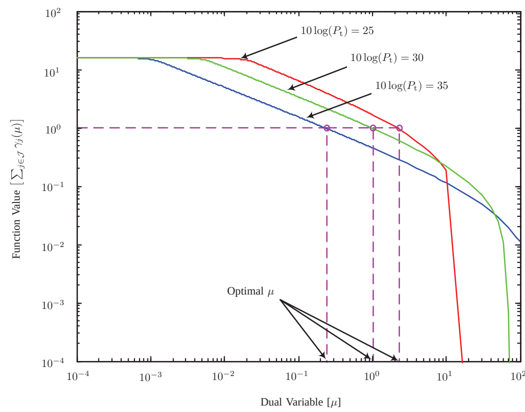

Fig. 2. Variation of $\sum_{j\in\mathcal{J}} \gamma_j(\mu)$ as a function of μ. $N_{\rm c} = 16$, $N_j = N_{j-1} + 4$, $\sigma_{i,j}^2 = \bar{\sigma}_j^2$, $\bar{\sigma}_j^2 = \bar{\sigma}_{j-1}^2 + 0.1$, $c_{i,j} = 1$, and $g_j = 1$ for $j \in \mathcal{J}$ and $i \in \mathcal{I}_j$ with $N_0 = 0$ and $\bar{\sigma}_0^2 = 0$. $\alpha = 0.5$, $\xi_j = 1/N_{\rm c}$, and $\beta_{i,j} = 1/(N_j - 1)$, $\forall i,j$. $\log(P_{\rm t}) = \log_{10}(P_{\rm t})$.

can be given by

$$\begin{aligned} \Omega(\mu) &= \underset{\gamma_j\geq 0,\,\forall j}{\operatorname{minimize}}\ \Lambda(\gamma_1,\ldots,\gamma_{N_{\rm c}};\mu) \\ &= \sum_{j\in\mathcal{J}} \underset{\gamma_j\geq 0}{\operatorname{minimize}}\ \Lambda_j(\gamma_j,\mu) - \mu \end{aligned} \quad (19)$$

and the dual optimization problem can be written as follows:

$$\underset{\mu\geq 0}{\operatorname{maximize}}\ \Omega(\mu). \quad (20)$$

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

$$\gamma_j(\mu) = \arg\min_{\gamma_j\geq 0} \Lambda_j(\gamma_j,\mu) \quad (21)$$

for $j \in \mathcal{J}$. Note that, (21) corresponds to the cluster j which can be solved by the corresponding CH using some line search algorithm for one-dimensional minimization.

The optimal dual variable μ can be obtained by finding μ such that $\sum_{j\in\mathcal{J}} \gamma_j(\mu) = 1$ as illustrated in Fig. 2. This can be done by a one-dimensional numerical search, for instance, bisectional search method, or can be done using gradient-ascent method that leads to the following updation rule [28]:

$$\mu^{(\tau+1)} = \left[\mu^{(\tau)} + \delta^{(\tau)}\Bigg(\sum_{j\in\mathcal{J}} \gamma_j\big(\mu^{(\tau)}\big) - 1\Bigg)\right]^{+}, \quad (22)$$

where τ is an iteration index, δ is a positive step-size parameter, and $\gamma_j(\mu^{(\tau)})$ is the solution of (21) for given $\mu^{(\tau)}$. Because the primal problem is convex having linear constraints and is feasible over the domain of the problem thus the problem satisfies the weak Slater's condition for constraint qualifications. Given that, the primal variables $\gamma_j(\mu^{(\tau)})$'s and the dual variable $\mu^{(\tau)}$ converge to their optimal values as $\tau \to \infty$, and at convergence the duality gap is zero [26], [28].

Now, to solve (12) including the constraints on individual γ_j's, we adopt the following procedure:

i: Set $\iota = 0$ and assume $\gamma_{\rm t}^{(\iota)} = 1$ such that $\sum_{j\in\mathcal{J}} \gamma_j \leq \gamma_{\rm t}^{(\iota)}$, where ι is an iteration index.

ii: Solve the optimization problem as outlined in (18) to (22) ignoring the individual constraints.

iii: Construct the set $\mathcal{L} = \{j \in \mathcal{J} | \gamma_j \geq \gamma_{\max}^{(j)}\}$, and for all $l \in \mathcal{L}$ set $\gamma_l = \gamma_{\max}^{(l)}$.

iv: Recalculate the sum-constraint as $\gamma_{\mathrm{t}}^{(\iota+1)} = \gamma_{\mathrm{t}}^{(\iota)} - \sum_{l\in\mathcal{L}} \gamma_{\max}^{(l)}$.

v: Recalculate γ_r for all $r \in \mathcal{R}^{(\iota+1)}$ as in step ii with sum-constraint $\sum_{r\in\mathcal{R}^{(\iota+1)}} \gamma_r \leq \gamma_{\mathrm{t}}^{(\iota+1)}$, where $\mathcal{R}^{(\iota+1)} = \mathcal{R}^{(\iota)} \setminus \mathcal{L}$ with $\mathcal{R}^{(0)} = \mathcal{J}$. Note that $\mathcal{R}^{(\iota)} \setminus \mathcal{L}$ means all elements of $\mathcal{R}^{(\iota)}$ that are not in $\mathcal{L}$.

vi: Set $\iota = \iota + 1$ and repeat step iii to step v until all constraints are satisfied.

The solution given in (18) to (22) with only the sum-constraint, and the solution obtained by the preceding procedure incorporating both the sum- and the individual-constraints are optimal. The optimality can be justified by the convexity of the underlying optimization problem and the decreasing property of its objective function w.r.t. γ_j's.

Remark 3: From the viewpoint of implementation, the solution based on the Lagrangian dual-decomposition approach, as outlined here, can be computed 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 (21). Note that for CH j, the problem (21) 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 the γ_j's and μ converge to their optimal values.

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

For optimization of $\beta_{i,j}$'s, we need to solve the problem (13). In this case, the cost function $\mathrm{tr}\left(\mathbf{R}_\epsilon\right)$ and the constraints decouple along the clusters. Consequently, the problem (13) decouples into N_{c} independent problems, one for each cluster. Here it is sufficient to consider the following problem for each $j \in \mathcal{J}$.

$$\underset{\beta_{i,j}\geq 0,\forall i}{\mathrm{minimize}}\ \sigma_j^2 \ \text{ subject to } \sum_{i\in\mathcal{I}_j} \beta_{i,j} \leq 1, \tag{23}$$

which is equivalent to

$$\begin{aligned}
&\underset{\beta_{i,j}\geq 0,\forall i}{\mathrm{minimize}}\ -\sum_{i\in\mathcal{I}_j} \frac{\alpha P_{\mathrm{t}}\gamma_j\beta_{i,j}c_{i,j}}{\alpha P_{\mathrm{t}}\gamma_j\beta_{i,j}c_{i,j}\sigma_{i,j}^2 + \sigma_{i,j}^2 + \sigma_{s_j}^2} \\
&\text{subject to} \sum_{i\in\mathcal{I}_j} \beta_{i,j} \leq 1,
\end{aligned} \tag{24}$$

where, on the same lines as Appendix C, we can show that the objective function is a decreasing function 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 associated KKT conditions in the same way as in Section III-B for ξ_j's.

$$\beta_{i,j} = \frac{\sigma_{s_j}^2 + \sigma_{i,j}^2}{\alpha P_{\mathrm{t}}\gamma_j c_{i,j}\sigma_{i,j}^2}\left(\sqrt{\frac{\alpha P_{\mathrm{t}}\gamma_j c_{i,j}}{(\sigma_{s_j}^2 + \sigma_{i,j}^2)\eta_j}} - 1\right)^+, \ \forall i, \tag{25}$$

$$\eta_j = \left(\frac{\sum_{\kappa\in\mathcal{A}_j} \frac{1}{\sigma_{\kappa,j}^2}\sqrt{\frac{\sigma_{s_j}^2+\sigma_{\kappa,j}^2}{\alpha P_{\mathrm{t}}\gamma_j c_{\kappa,j}}}}{1 + \sum_{\kappa\in\mathcal{A}_j} \frac{\sigma_{s_j}^2+\sigma_{\kappa,j}^2}{\alpha P_{\mathrm{t}}\gamma_j c_{\kappa,j}\sigma_{\kappa,j}^2}}\right)^2, \tag{26}$$

where $\mathcal{A}_j = \left\{i \in \mathcal{I}_j \middle| \frac{\alpha P_{\mathrm{t}}\gamma_j c_{i,j}}{(\sigma_{s_j}^2+\sigma_{i,j}^2)\eta_j} > 1\right\}$. For $P_{\mathrm{t}} \to \infty$, the $\beta_{i,j}$'s converges to

$$\lim_{P_{\mathrm{t}}\to\infty} \beta_{i,j} = \sqrt{\frac{\sigma_s^2 + \sigma_{i,j}^2}{c_{i,j}\sigma_{i,j}^4}}\left(\sum_{l\in\mathcal{I}_j}\sqrt{\frac{\sigma_s^2 + \sigma_{l,j}^2}{c_{l,j}\sigma_{l,j}^4}}\right)^{-1}. \tag{27}$$

Remark 4: To implement the solution, the CH j determines the Lagrange multiplier η_j and broadcasts its value to all sensors in that cluster. After knowing η_j (as well as $\alpha P_{\mathrm{t}}\gamma_j$), the sensors can calculate $\beta_{i,j}$'s by (25).

IV. Correlated Sources

Herein we outline solution to the optimization problem (7) for the case of correlated sources s_j's. This section consists of two subsections. In the first, we solve the optimization problem with the exact formulation of the cost function $\mathrm{tr}\left(\mathbf{R}_\epsilon\right)$ given in (8). The function $\mathrm{tr}\left(\mathbf{R}_\epsilon\right)$ is not separable, because the covariance matrix $\mathbf{R}_{\mathbf{s}}$ is not diagonal in this case. That is why, the resulting solution for α, ξ_j's, γ_j's, and $\beta_{i,j}$'s (given in Section IV-A) does not admit a distributed implementation, and may computationally be expensive as the solution has to be computed numerically in an iterative fashion that involves matrix computations. In order to address these concerns, in the second subsection, we develop an upper-bound for $\mathrm{tr}\left(\mathbf{R}_\epsilon\right)$ and use this as a cost function to solve the problem eqrefOptimizationProblem-II. The resulting solution is amenable for distributed implementation and carries much of the favorable properties as exhibited by the case of uncorrelated sources. In what follows, we ignore the constraints $\gamma_j \leq \gamma_{\max}^{(j)}$ for all j. Nevertheless, these constraints can similarly be incorporated as we have discussed in Section III-C.

A. Exact Solution

Here, we solve the problem (7) with the exact cost function given in (8) by employing the BCoDM based approach outlined in Section II-B. Unlike the uncorrelated case, the solution outlined here has to be computed in a centralized way because the objective function does not support a separable structure.

1) Optimization of α: The optimization problem (10) has a unique global minimization point, as shown in Appendix D. The corresponding optimal α can be found by using the numerical methods for one-dimensional search as discussed in Section III-A.

2) Optimization of ξ_j's: The objective function of the problem (11) is a decreasing function of ξ_j for all j. Besides, the problem is jointly convex over ξ_j's, as shown in Appendix E. As the objective function is a convex decreasing function, therefore, the optimum is at the constraint boundary, that is, the sum-constraint $\sum_{j\in\mathcal{J}} \xi_j \le 1$ is always active. Moreover, the convexity of the problem means the following conditions are sufficient for optimality of the solution for ξ_j's [28].

$$\xi_j, \forall j = \arg\min_{\xi_j \ge 0, \forall j} \operatorname{tr}(\mathbf{R}_\epsilon) + \lambda\Bigg(\sum_{j\in\mathcal{J}} \xi_j - 1\Bigg), \tag{28}$$

$$\lambda > 0, \quad \sum_{j\in\mathcal{J}} \xi_j - 1 = 0, \tag{29}$$

where λ is a Lagrange multiplier. The optimal ξ_j's can be obtained by numerically solving (28) for given λ (e.g., using the gradient projection methods [28]) and finding $\lambda > 0$ such that $\sum_{j\in\mathcal{J}} \xi_j(\lambda) = 1$. To find the optimal λ, we can use the bi-sectional search or the dual-ascent method described in Section III-C.

3) Optimization of γ_j's: For γ_j's, the optimization problem to solve is (11), where (as shown in Appendix F) the objective function is decreasing w.r.t. γ_j and the problem is jointly convex over γ_j's. The optimal γ_j's can be computed numerically by a similar procedure as proposed for ξ_j's in Section IV-A2.

4) Optimization of $\beta_{i,j}$'s: As shown in Appendix G, for each j, the optimization problem (13) is jointly convex and the cost function is decreasing over $\beta_{i,j}$'s. Once again, to solve the problem we can use the procedure of Section IV-A2.

B. Approximate Solution

In the ensuing development, to solve the problem (7) for the case of correlated sources, 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.

***Proposition** 3:* The trace of $\mathbf{R}_\epsilon$ can be upper bounded as follows:

$$\operatorname{tr}(\mathbf{R}_\epsilon) \le \Bigg(\sum_{j\in\mathcal{J}} \sigma_{s_j}^2\Bigg) - \Upsilon, \tag{30}$$

where

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

$$\tilde{\mathbf{Q}} = \mathbf{R_s} \circ (\mathbf{R_s}^T \mathbf{R_s}), \quad \Psi_j = \sum_{k\in\mathcal{J}} \mathbf{R}^2_{\mathbf{s}[j,k]} = \sum_{k\in\mathcal{J}} \operatorname{Cov}\{S_j, S_k\}^2$$

in which the operator '$\circ$' denotes the Hadamard or Schur product of the matrices.

Proof: See Appendix H. $\square$

Now we consider the optimization problem (7) where we target to minimize the given upper-bound on the distortion. For this purpose, it is sufficient to consider the following optimization problem:

$$\begin{aligned} &\underset{\alpha\in\mathcal{T},\, \xi_j\ge 0,\, \gamma_j \ge 0,\, \beta_{i,j}\ge 0,\, \forall i,j}{\text{minimize}} \; -\Upsilon \\ &\text{subject to } \sum_{j\in\mathcal{J}} \xi_j \le 1,\ \sum_{j\in\mathcal{J}} \gamma_j \le 1,\ \sum_{i\in\mathcal{I}_j} \beta_{i,j} \le 1,\ \forall j, \end{aligned} \tag{31}$$

where we have ignored the constraints on individual γ_j's. Nevertheless, these constraints can be incorporated in the solution by the procedure outlined in Section III-C. Note that in the special case of $N_j = 1$ for all j—that is, where all clusters comprise one sensor each—we have $\alpha = 0$, $\gamma_j = 0$, and $\beta_{i,j} = 0$ for all i and j. In this particular case, what remains to be optimized is ξ_j's and the optimization problem (31) converges to the case studied in [8]. For the general case, in the sequel, we outline a solution to the problem (31) employing the BCoDM approach of Section II-B.

1) Optimization of α: To minimize '$-\Upsilon$' w.r.t. α it is sufficient to consider minimization of the denominator of Υ—the numerator does not depend on α. Assuming $f_j(\alpha) = \xi_j g_j \Psi_j \Big(\xi_j g_j \sigma_j^2 + \frac{\sigma_j^2+\sigma_{s_j}^2}{(1-\alpha)P_{\mathrm{t}}}\Big)$ and $f(\alpha) = \sum_{j\in\mathcal{J}} f_j(\alpha)$, the optimization problem for α can be written as follows:

$$\underset{\alpha\in\mathcal{T}}{\text{minimize}}\ f(\alpha). \tag{32}$$

We can solve (32) using one-dimensional numerical search methods. As the function $f(\alpha)$ decouples along the CHs indicating that the numerical algorithm to find α can be implemented in a distributed manner similar to the uncorrelated case.

2) Optimization of ξ_j's: For optimization of ξ_j's, we proceed as follows. By defining

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

we can write Υ as follows

$$\Upsilon = \frac{\xi^T \mathbf{U} \xi}{\xi^T \mathbf{Q} \xi + \mathbf{q}^T \xi}. \tag{33}$$

Now assuming $\mathbf{1} := [1, \ldots, 1]^T$, the optimization problem for ξ_j's can be written as

$$\underset{\xi \ge 0}{\text{maximize}}\ \Upsilon \quad \text{subject to } \mathbf{1}^T\xi = 1, \tag{34}$$

where we have replaced the inequality constraint with equality to exclude the case in which the objective function is unbounded over the feasible region of the problem.

Let $\mathcal{F} = \left\{\xi \in \mathbb{R}^{N_{\mathrm{c}}} | \xi \ge \mathbf{0},\ \mathbf{1}^T\xi = 1\right\}$ denote the feasible region of the problem (34). 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 PSD matrices, which means the numerator and denominator of Υ are convex functions of

ξ. Thus, the problem (34) is a *convex–convex* type quadratic fractional programming problem—for a detailed classification of the types of fractional programs and their solution methods see [29] and references therein. To solve (34), in what follows, we develop an algorithm based on the parametric programming approach, which is a powerful scheme for solving the fractional programs.

For $\theta \geq 0$, let

$$f(\xi;\theta) = \xi^T \mathbf{U} \xi - \theta \left(\xi^T \mathbf{Q} \xi + \mathbf{q}^T \xi\right) \tag{35}$$

be a parametrized function associated with the problem (34). Next we have the following proposition, which is based on the well-known result by Dinkelbach [30].

***Proposition** 4:* For given θ, define

$$\varphi(\theta) = \underset{\xi\in\mathcal{F}}{\operatorname{maximize}}\ f(\xi;\theta) \tag{36}$$

with the corresponding optimal ξ vector as

$$\xi(\theta) = \arg\max_{\xi\in\mathcal{F}} f(\xi;\theta). \tag{37}$$

If there exists some $\theta^* \geq 0$ such that $\varphi(\theta^*) = 0$ then $\xi^* = \xi(\theta^*)$ is an optimal solution of the problem (34) and the corresponding optimal value is $\theta^* = \Upsilon(\xi^*)$.

Proof: See Appendix I. □

***Proposition** 5:* The function $\varphi(\theta)$ defined in (36) is a strictly decreasing convex function of θ. Moreover, for θ^* as defined in Prop. 4, following holds

$$\varphi(\theta) > 0,\ \forall\theta < \theta^*, \quad \text{and} \quad \varphi(\theta) < 0,\ \forall\theta > \theta^*. \tag{38}$$

Proof: See Appendix J. □

Based on Prop. 4, the optimization problem (34) 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 $\xi^{(\iota+1)}$:

$$\underset{\xi\in\mathcal{F}}{\operatorname{maximize}}\ f(\xi;\theta^{(\iota+1)}). \tag{39}$$

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.

This procedure is guaranteed to converge to the optimal solution of the problem (34) provided the problem (39) can be solved [30]. In this regard, although the feasible region $\mathcal{F}$ is a convex set, the function $f(\xi;\theta^{(\iota)})$ is not concave. Thus, the problem (39) is a non-concave maximization problem wherein many different local maxima may exist, which are different from the globally optimal solution. In what follows, we show that the problem (39) can be reformulated as a convex *quadratically constrained quadratic programming* (QCQP) problem, which can be solved by numerical methods, for example, the interior point method [26].

By introducing a slack variable $\tau = \xi^T \mathbf{U} \xi$, we can write (39) in the following equivalent form:

$$\begin{aligned} &\underset{\tau_{\min}\leq\tau\leq\tau_{\max};\,\xi\geq 0}{\operatorname{minimize}}\ \theta^{(\iota)}\left(\xi^T\mathbf{Q}\xi + \mathbf{q}^T\xi + \epsilon\right) - \tau \\ &\text{subject to}\ \mathbf{1}^T\xi = 1, \quad \xi^T\mathbf{U}\xi - \tau \leq 0, \end{aligned} \tag{40}$$

where the objective and the constraint functions are convex, and the problem (40) is a convex QCQP problem. The $\tau_{\min}$ and $\tau_{\max}$ in (40) are solution to the following problems:

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

where we can show that $\tau_{\min} = (\min\{u_j,\ldots,u_{N_\mathrm{c}}\})^2$ and $\tau_{\max} = (\max\{u_j,\ldots,u_{N_\mathrm{c}}\})^2$. That is, the corresponding solution for ξ_j's in both cases is like the *winner-take-all* policy—meaning only one of the ξ_j's is equal to 1.

3) Optimization of γ_j's: For optimization over γ_j's, it is sufficient to consider the following problem:

$$\begin{aligned} &\underset{\gamma_j,\,\forall j}{\operatorname{minimize}}\ \sum_{j\in\mathcal{J}} \sigma_j^2 \xi_j g_j \Psi_j \big(1 + (1-\alpha)P_\mathrm{t}\xi_j g_j\big) \\ &\text{subject to}\ \sum_{j\in\mathcal{J}} \gamma_j \leq 1, \end{aligned} \tag{41}$$

where we can prove that the objective function is decreasing w.r.t. γ_j's and the problem is jointly convex over γ_j's. Moreover, note that the objective as well as the constraint functions are separable along clusters. Consequently, the problem (41) can be solved using the Lagrangian dual-decomposition approach outlined in Section III-C.

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

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

whose solution is same as given in (25) to (27) under Section III-D.

V. Fully Correlated Sources

In the case of fully correlated sources s_j's, that is, $s_j = s$ for all j, the sensors in all clusters essentially observe a spatially homogeneous source. In this case, the mean squared estimation distortion $D_\epsilon := \mathbf{E}[(\hat{s} - s)^2]$ at the FC based on the MMSE estimate $\hat{s}$ of s can be written as

$$D_\epsilon = \left(\frac{1}{\sigma_s^2} + \sum_{j\in\mathcal{J}} \frac{(1-\alpha)P_\mathrm{t}\xi_j g_j}{(1-\alpha)P_\mathrm{t}\xi_j g_j \sigma_j^2 + \sigma_j^2 + \sigma_s^2}\right)^{-1},$$

where σ_j^2 is given in (9) with $\sigma_{s_j}^2 = \sigma_s^2$ for all j. Now we consider the optimization problem (7) with the cost function $\operatorname{tr}(\mathbf{R}_\epsilon)$ replaced by D_ϵ. The optimization problem can be solved by the BCoDM algorithm outlined in Section II-B. Similar to the uncorrelated case, the algorithm to find optimal α can be implemented in a distributed manner. To optimize over α, we can prove the existence and uniqueness of the optimal α; and we can find the optimal value using the approach given in Section III-A. For optimal ξ_j's, we can show that the objective function is decreasing and is jointly convex over ξ_j's. We can find the optimal ξ_j's by solving the KKT conditions. The resulting solution has same structure as in the uncorrelated case in Section III-B. To optimize γ_j's, we can show that the Lagrangian dual-decomposition approach

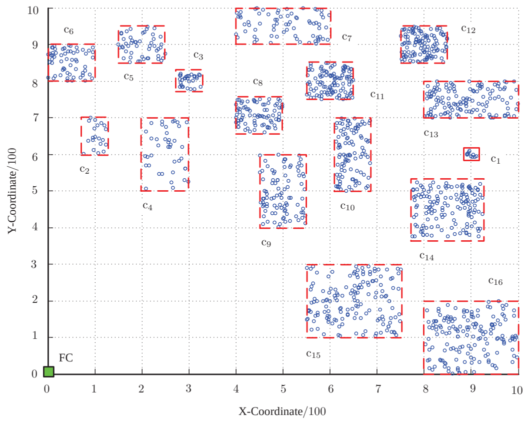

Fig. 3. Network setup for simulation.

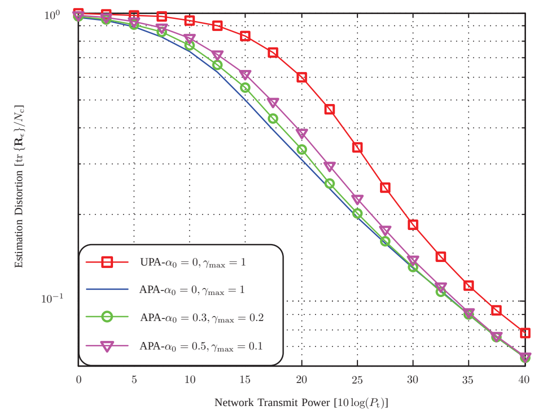

Fig. 4. Estimation distortion increases with increasing α_0 and decreasing $\gamma_{\max}$ for $\theta = 5000$.

outlined in Section III-C can be employed. The solution for optimal $\beta_{i,j}$'s comes out to be the same as outlined in (25) to (27) under Section III-D with $\sigma_{s_j}^2 = \sigma_s^2$ for all j.

If each cluster only contains one sensor, that is, $N_j = 1$ for all j then $\alpha = 0$, $\gamma_j = 0$, and $\beta_{i,j} = 0$ for all i and j; and what remains to be optimized is ξ_j's. In this scenario, the optimization problem considered in this section converges to the case studied in [5], [6]. Besides, for $N_j = N$, $\gamma_j = 1/N_c$, and $\sigma_{i,j}^2 = \sigma_n^2$ for all i and j the problem converges to the case discussed in [13], with the exception of how to find α, as we have outlined herein—[13] does not explicitly state how to find α and does not delineate on the associated convergence issue. Thus, as a conclusion, our proposed framework models the general case which includes related works as special cases.

VI. Performance Evaluation

In order to evaluate the performance of the proposed power allocation scheme we consider a WSN comprising $N_c = 16$ clusters with sizes $N_j = N_{j-1} + 4$ for all $j \in \mathcal{J}$ with $N_0 = 0$. In each cluster, the sensors are randomly and uniformly distributed as shown in Fig. 3, where c_1 through c_{16} denotes the clusters. The correlation between the underlying sources of clusters j and k is modeled as

$$\rho_{s_j,s_k} = e^{-d_{j,k}^{c}/\theta}, \quad \forall j,k \in \mathcal{J},$$

where $d_{j,k}^{c}$ denotes the CH-to-CH distance between clusters 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. Moreover, we assume that the observation noise variances (i.e., $\sigma_{i,j}^2$'s) are uniformly distributed between 0.1 and 10. The channel SNR from sensor i in cluster j to the FC is modeled as $g_{i,j} = |\tilde{h}_{i,j}|^2/\tilde{d}_{i,j}^2\sigma_{w_j}^2$, where $\tilde{h}_{i,j} \sim \mathcal{CN}(0,1)$ and $\tilde{d}_{i,j}$ denotes the distance between the sensor i in cluster j and the FC. For cluster j, the CH is selected by the following rule:

$$i^* = \arg\max_i \{g_{i,j}, i \in \mathcal{I}_j\}$$

such that $g_j = g_{i^*,j}$. For CH selection, other criteria are also possible, for instance, in each cluster select the sensor with largest remaining energy [21]. We also have $c_{i,j} = |h_{i,j}|^2/d_{i,j}^2\sigma_{w_{i,j}}^2$; where $h_{i,j} \sim \mathcal{CN}(0,1)$, and $d_{i,j}$ denotes the distance between the sensor i and CH j. We assume $\sigma_{w_j}^2 = \sigma_{w_{i,j}}^2 = 10^{-4}$ for all i and j.

We compare the distortion performance of the proposed APA design with a UPA scheme. In the UPA scheme we have $\alpha = \alpha_u = 0.5$, $\xi_j = \xi_u = 1/N_c$, $\gamma_j = \gamma_u = 1/N_c$, and $\beta_{i,j} = \beta_{u_j} = 1/(N_j - 1)$ for all i and j. The results are averaged over 10^3 random deployments of the sensors in each cluster. Unless stated otherwise, in the figures the estimation distortion $\mathrm{tr}\,(\mathbf{R}_\epsilon)$ is normalized by the number of underlying sources being estimated. The purpose of the simulations is to observe how the distortion performance of the APA scheme varies with P_t for different values of α_0 and $\gamma_{\max}^{(j)}$, and compares with the performance of the UPA scheme over the degree of correlation among the sources. Furthermore, the simulations show how the distortion achieved by the approximate solution matches with the exact solution for different levels of correlation among the sources. The simulations also compare the distortion performance of the cluster-based WSN with a centralized WSN. In the figures, $\log(P_t) = \log_{10}(P_t)$.

Let $\gamma_{\max}^{(j)} = \gamma_{\max}$ for all j. Increasing the value of α_0 and decreasing the value of $\gamma_{\max}$ reduce the feasibility region of the underlying optimization problem and consequently the estimation distortion is expected to increase. This is illustrated in Fig. 4. When the value of α_0 is increased, it means there would be less power available to transmit the partial estimates from the CHs to the FC. In this case even though CHs may have very good estimate of the sources, however they will not have enough power assigned to reliably transmit those estimates to the FC. On the other hand, when $\gamma_{\max}$ is reduced, there would be less power available to the sensors to deliver their observations to the CHs, which would deteriorate the quality of the partial estimates at the CHs. The deterioration of the estimates contribute to the increase in the overall estimation distortion at the FC. Unless stated otherwise, in the subsequent simulation examples, we assume $\alpha_0 = 0$ and $\gamma_{\max} = 1$, which effectively means we only consider constraint on the transmit power of the overall network specified by P_t.

Fig. 5 plots and compares the estimation distortion of the APA and the UPA schemes as a function of P_t for different levels of correlation among the underlying sources.

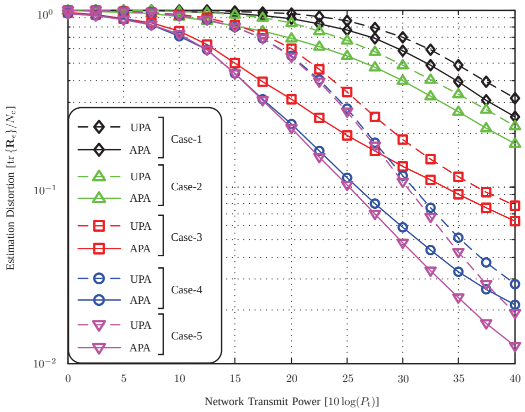

Fig. 5. Estimation distortion comparison of the APA and the UPA schemes: Case-1) uncorrelated sources, Case-2) correlated sources with $\theta = 500$, Case-3) correlated sources with $\theta = 5000$, Case-4) correlated sources with $\theta = 50000$, and Case-5) fully correlated sources.

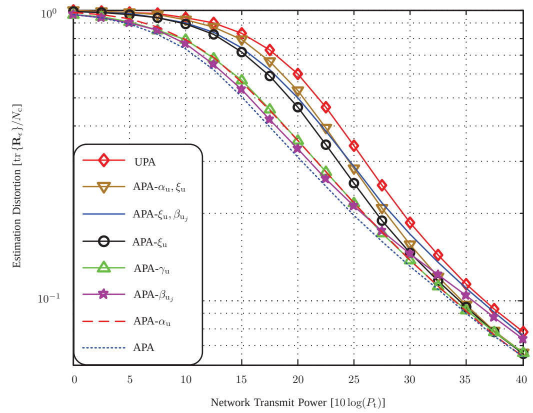

Fig. 6. Impact of different optimization variables on estimation distortion for $\theta = 5000$.

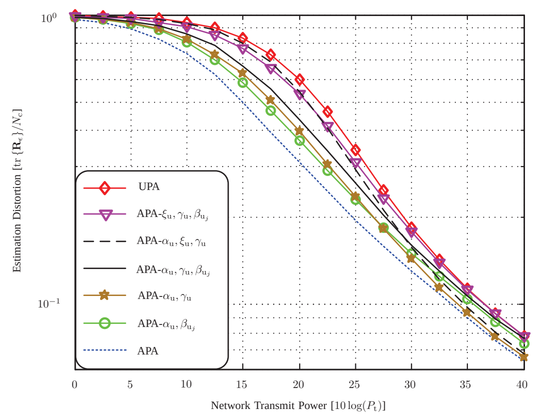

Fig. 7. Impact of different optimization variables on estimation distortion for $\theta = 5000$.

According to the assumed model, the correlation among the sources increases as the value of θ increases. The figure shows that the proposed APA scheme gives distortion performance which is better than the UPA scheme and the performance difference increases as the level of correlation increases. In the perspective of efficient utilization of energy, this observation illustrates the advantage of the proposed power allocation scheme. Besides, 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.

In the proposed APA scheme, even optimizing over some or anyone of the variables (α, ξ_j's, γ_j's, and $\beta_{i,j}$'s) may give significant performance gain. This is illustrated in Fig. 6 and Fig. 7, where the APA scheme means all variables are optimized and the UPA scheme refers to the case of all uniform variables. The other schemes refer to mixed situations with all variables optimized but the ones indicated, which are uniform. The figures underline the relative importance of different optimization variables and their impact on the distortion. For instance, Fig. 6 shows that roughly before $\log_{10}(P_{\mathrm{t}}) = 30$, among the four types of optimization variables, selecting $\beta_{i,j} = \beta_{\mathrm{u}_j}$ causes the least increase in the distortion (compared to the APA case) whereas in the same P_{t} range selecting $\xi_j = \xi_{\mathrm{u}}$ causes the highest increase in the distortion. However, at P_{t} values higher than the given value, the converse behavior can be observed. Effectively this means, for the given setup, if the network transmit power falls below the given value then splitting power uniformly among the sensors in each cluster, to transmit observations to their respective CHs, does not cost much in terms of distortion performance while implementation of the power allocation protocol will significantly be simplified. Similarly, when the network transmit power is greater than the given value, then splitting power in a uniform way among the CHs, to transmit their equivalent observations to the FC, does not cause significant increase in distortion. One important conclusion from these simulation examples, we could choose to optimize over a subset of the variables as a trade-off between the computational cost and implementation overhead versus performance gain with respect to the UPA scheme.

Next we compare the estimation distortion achieved by the APA schemes proposed under the exact solution and the approximate solution in Section IV-A and Section IV-B, respectively. The results are plotted in Fig. 8 for different level of correlation among the sources, where CES and CAS denote the exact solution and the approximate solution, respectively. 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 P_{t} and the correlation values. Specifically, when the sources are not highly correlated then the difference in the distortion performance is essentially negligible over almost the entire range of P_{t}. This observation illustrates the effectiveness of the APA scheme based on the distortion approximation vis-à-vis the scheme based on the exact distortion function.

Finally, we compare the distortion performance of the APA scheme for the cluster-based hierarchical WSN with that of the APA scheme for a centralized WSN in which all sensors directly send their observations to the FC. Under the centralized scheme, we can show that the estimation error covariance $\mathbf{R}_\epsilon$ can be given by (8) with $\mathbf{H}\mathbf{R}^{-1}\mathbf{H}^T$ replaced

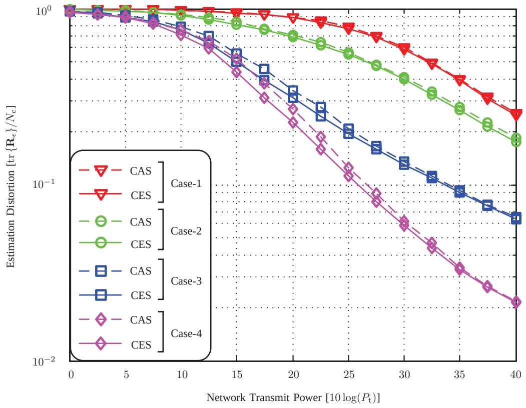

Fig. 8. Distortion comparison of the exact solution (CES) and the approximate solution (CAS) for correlated sources: Case-1) $\theta = 50$, Case-2) $\theta = 500$, Case-3) $\theta = 5000$, and Case-4) $\theta = 50000$.

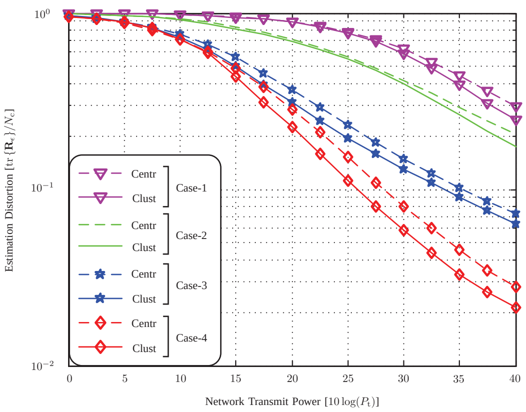

Fig. 9. Cluster-based (Clust) and centralized (Centr) WSNs distortion comparison for correlated sources: Case-1) $\theta = 50$, Case-2) $\theta = 500$, Case-3) $\theta = 5000$, and Case-4) $\theta = 50000$.

by a diagonal matrix $\mathbf{C}$ defined as

$$\mathbf{C}_{[j,j]} = \sum_{i \in \tilde{\mathcal{I}}_j} \frac{P_{\rm t}\gamma_j \beta_{i,j} g_{i,j}}{P_{\rm t}\gamma_j \beta_{i,j} g_{i,j} \sigma_{i,j}^2 + \sigma_{i,j}^2 + \sigma_{s_j}^2}, \quad \forall j \in \mathcal{J}.$$

In this particular case, the optimization problem (7) reduces to

$$\begin{aligned} &\underset{\gamma_j \geq 0,\, \beta_{i,j} \geq 0,\, \forall i,j}{\text{minimize}} \quad \mathrm{tr}\left(\mathbf{R}_\epsilon\right) \\ &\text{subject to} \ \sum_{j \in \mathcal{J}} \gamma_j \leq 1, \sum_{i \in \tilde{\mathcal{I}}_j} \beta_{i,j} \leq 1, \ \gamma_j \leq \gamma_{\max}^{[j]}, \ \forall j, \end{aligned} \qquad (43)$$

which can be solved on the same lines as the problem (7). In Fig. 9 the estimation distortions achieved by the APA designs for the cluster-based and the centralized WSNs are plotted. The figure shows that the cluster-based scheme outperforms the centralized scheme in terms of achieved distortion. The gap in distortion performance is significant when either the network transmit power is reasonable large or the correlation among the sources is high, or both. In the realm of energy-efficient estimation, this illustrates the potential advantage of the cluster-based hierarchical WSN topology.

VII. Concluding Remarks

In this work we addressed the problem of power-constrained estimation in cluster-based WSNs, where clusters observe underlying correlated sources. The proposed power scheduling design minimizes the estimation distortion with constraints on the transmit power of the clusters as well as the network as a whole. The estimation in the network is performed in two stages. In the first stage, the CH of each cluster forms a preliminary estimate of the source based on the observations received from the sensors in the cluster. In the second phase, the CHs transmit their estimates to the FC where the final estimate is formed. In this work, we formulated the power allocation problem as a convex optimization problem and outlined its solution based on a block coordinate descent method using partitioning principle by exploiting the independence property of the constraints. The proposed power allocation design for fully correlated and fully uncorrelated cases can be implemented in a distributed fashion. However, for partially correlated case the design requires a centralized scheduler to optimize the power allocations. Subsequently, we proposed an approximate solution based on an upper-bound of the distortion function. Thus obtained solution bears favorable characteristics for distributed implementation (very much like the fully correlated and fully uncorrelated cases) and gives distortion performance that matches quite closely to that of the exact solution. We showed that the proposed power allocation design gives distortion performance better than a uniform power allocation scheme. We also showed the advantage of the cluster-based hierarchical WSN compared to the centralized WSN in the perspective of realizing energy-efficient estimation.

In this paper we have investigated power-constrained estimation of spatially correlated sources in WSNs that are organized in two tiers. It would be interesting to extend the work to networks comprising more than two tiers, and analyze other optimization criteria like minimization of network power consumption and outage probability, among others. We assumed perfect knowledge of the instantaneous channel gains, it would be for instance intriguing to study more realistic scenarios where there are errors in the channel gains or to adapt power allocations according to channel statistics instead of the instantaneous channel gains. While in our analysis we focused on the energy consumed in the transmitting operations, it would be interesting to take the analysis one step further and also incorporate the energy consumed in the receiving operations. We assumed fixed wireless medium access scheme for transmission of the observations from the sensors to the CHs and from the CHs to the FC. It would be for example of practical interest to study the estimation problem with joint optimization of the transmit powers and the medium access scheduling assuming random channel access mechanism like slotted ALOHA or use schedule based mechanism like TDMA but with spatial reuse of the time slots.

Appendix A
Proof of Prop. 1

By defining $g_j = \tilde{g}_j/\sigma^2_{w_j}$ for $j \in \mathcal{J}$, it is straight forward to obtain $\mathbf{R}_\epsilon$ and σ^2_j from $\tilde{\mathbf{R}}_\epsilon$ and $\tilde{\sigma}^2_j$, respectively, by substituting $\psi_j = (1-\alpha)P_{\rm t}\xi_j/(\sigma^2_{s_j}+\tilde{\sigma}^2_j)$ and $\phi_{i,j} = \alpha P_{\rm t}\gamma_j\beta_{i,j}/(\sigma^2_{s_j}+\sigma^2_{i,j})$ for $i \in \mathcal{I}_j$ and $j = \mathcal{J}$ in (5) and (3). Now to prove Prop. 1, we need to show that the constraints of the problem (6) and the problem (7) are equivalent. To this end, note that according to the definition of α we can split the constraint $\sum_{j\in\mathcal{J}}\big(\psi_j\big(\sigma^2_{s_j}+\tilde{\sigma}^2_j\big)+\sum_{i\in\mathcal{I}_j}\phi_{i,j}\big(\sigma^2_{s_j}+\sigma^2_{i,j}\big)\big) \le P_{\rm t}$ into two as follows:

$$\sum_{j\in\mathcal{J}}\psi_j\big(\sigma^2_{s_j}+\tilde{\sigma}^2_j\big) \le (1-\alpha)P_{\rm t}, \tag{44}$$

$$\sum_{j\in\mathcal{J}}\sum_{i\in\mathcal{I}_j}\phi_{i,j}\big(\sigma^2_{s_j}+\sigma^2_{i,j}\big) \le \alpha P_{\rm t}. \tag{45}$$

Now substituting $\psi_j = (1-\alpha)P_{\rm t}\xi_j/(\sigma^2_{s_j}+\tilde{\sigma}^2_j)$ in (44) we get $\sum_{j\in\mathcal{J}}\xi_j \le 1$. Similarly substituting $\phi_{i,j} = \alpha P_{\rm t}\gamma_j\beta_{i,j}/(\sigma^2_{s_j}+\sigma^2_{i,j})$ in (45) we get $\sum_{j\in\mathcal{J}}\gamma_j\big(\sum_{i\in\mathcal{I}_j}\beta_{i,j}\big) \le 1$, where the inequality holds for $\sum_{i\in\mathcal{I}_j}\beta_{i,j} \le 1$ and $\sum_{j=1}^{N_{\rm c}}\gamma_j \le 1$.

Next, (44) in conjunction with $\sum_{j\in\mathcal{J}}\psi_j\big(\sigma^2_{s_j}+\tilde{\sigma}^2_j\big) \le \psi_{\max}$ imply $(1-\alpha)P_{\rm t} \le \psi_{\max}$ that gives $\alpha \ge \max\{0, 1-\psi_{\max}/P_{\rm t}\} := \alpha_0$. Finally substituting $\phi_{i,j} = \alpha P_{\rm t}\gamma_j\beta_{i,j}/(\sigma^2_{s_j}+\sigma^2_{i,j})$ in $\sum_{i\in\mathcal{I}_j}\phi_{i,j}\big(\sigma^2_{s_j}+\sigma^2_{i,j}\big) \le \phi^{(j)}_{\max}$ and noting that $\sum_{i\in\mathcal{I}_j}\beta_{i,j} \le 1$ and $\alpha < 1$ give $\gamma_j \le \phi^{(j)}_{\max}/P_{\rm t} := \gamma^{(j)}_{\max}$. Because $\gamma_j \ge 0$ and $\sum_{j\in\mathcal{J}}\gamma_j \le 1$ which imply that $\gamma_j \le 1$; thus we have $\gamma^{(j)}_{\max} = \min\{1, \phi^{(j)}_{\max}/P_{\rm t}\}$.

Appendix B
Convexity of $f(\alpha)$ over $\mathcal{T}$ for Uncorrelated Sources

Assuming $q_{i,j} := \frac{(\sigma^2_{s_j}+\sigma^2_{i,j})P_{\rm t}\gamma_j\beta_{i,j}c_{i,j}}{(\alpha P_{\rm t}\gamma_j\beta_{i,j}c_{i,j}\sigma^2_{i,j}+\sigma^2_{i,j}+\sigma^2_{s_j})^2}$ and $\bar{q}_{i,j} := \frac{(\sigma^2_{s_j}+\sigma^2_{i,j})P^2_t\gamma^2_j\beta^2_jc^2_{i,j}\sigma^2_{i,j}}{(\alpha P_{\rm t}\gamma_j\beta_{i,j}c_{i,j}\sigma^2_{i,j}+\sigma^2_{i,j}+\sigma^2_{s_j})^3}$, the second-order derivative of $f(\alpha)$ w.r.t. α can be written as

$$\frac{\partial^2 f(\alpha)}{\partial\alpha^2} = \sum_{j\in\mathcal{J}}\frac{2P_{\rm t}\xi_jg_j\sigma^4_{s_j}}{\big((1-\alpha)P_{\rm t}\xi_jg_j+1\big)\big(\sigma^2_{s_j}+\sigma^2_j\big)}$$
$$\Bigg[\frac{P_{\rm t}\xi_jg_j}{\big((1-\alpha)P_{\rm t}\xi_jg_j+1\big)^2}+\frac{\sigma^4_j\sum_{i\in\mathcal{I}_j}q_{i,j}}{\big((1-\alpha)P_{\rm t}\xi_jg_j+1\big)\big(\sigma^2_{s_j}+\sigma^2_j\big)}$$
$$+\frac{\sigma^2_{s_j}\sigma^6_j\big(\sum_{i\in\mathcal{I}_j}q_{i,j}\big)^2}{\big(\sigma^2_{s_j}+\sigma^2_j\big)^2}+\frac{\sigma^4_j\sum_{i\in\mathcal{I}_j}\bar{q}_{i,j}}{\sigma^2_{s_j}+\sigma^2_j}\Bigg]. \tag{46}$$

Clearly $\partial^2 f(\alpha)/\partial\alpha^2 > 0$ for any $\alpha \in \mathcal{T}$, which establishes that $f(\alpha)$ is strictly convex over $\mathcal{T}$. This proves the existence and uniqueness of the global minimizer of $f(\alpha)$ over $\mathcal{T}$ (cf., Prop. B.10 [28]).

Appendix C
Convexity of $\operatorname{tr}\big(\mathbf{R}_\epsilon\big)$ over ξ_j's for Uncorrelated Sources

The first-order derivative of $\operatorname{tr}\big(\mathbf{R}_\epsilon\big)$ w.r.t. ξ_j is negative for any valid ξ_j for all j, that is,

$$\frac{\partial\operatorname{tr}\big(\mathbf{R}_\epsilon\big)}{\partial\xi_j} = \frac{-\sigma^4_{s_j}}{\sigma^2_{s_j}+\sigma^2_j}\frac{(1-\alpha)P_{\rm t}g_j}{((1-\alpha)P_{\rm t}\xi_jg_j+1)^2} < 0, \forall j, \tag{47}$$

which shows that $\operatorname{tr}\big(\mathbf{R}_\epsilon\big)$ is a decreasing function of ξ_j. As the constraints of the problem (11) are linear, therefore, to prove the convexity of the problem it is sufficient to show that the objective function is convex. For this purpose, we can show that the second-order derivatives of $\operatorname{tr}\big(\mathbf{R}_\epsilon\big)$ w.r.t. ξ_j's are

$$\frac{\partial^2\operatorname{tr}\big(\mathbf{R}_\epsilon\big)}{\partial\xi^2_j} = \frac{\sigma^4_{s_j}}{\sigma^2_{s_j}+\sigma^2_j}\frac{2(1-\alpha)^2P^2_tg^2_j}{((1-\alpha)P_{\rm t}\xi_jg_j+1)^3} \ge 0, \forall j,$$
$$\frac{\partial^2\operatorname{tr}\big(\mathbf{R}_\epsilon\big)}{\partial\xi_l\partial\xi_j} = 0, \forall l \ne j, \tag{48}$$

which tell us that the Hessian of the objective function is positive-semidefinite (PSD) that in turn proves that the function is jointly convex over ξ_j's.

Appendix D
Existence and Uniqueness of Global Minimizer α for Correlated Sources

Let $\mathbf{F}$ and $\mathbf{G}$ be diagonal matrices with, for $j \in \mathcal{J}$,

$$\frac{\mathbf{F}_{[j,j]}}{P_{\rm t}\xi_jg_j} = \frac{\sigma^2_j+\sigma^2_{s_j}+(1-\alpha)\big((1-\alpha)P_{\rm t}\xi_jg_j+1\big)\sigma^4_j\sum_{i\in\mathcal{I}_j}q_{i,j}}{\big((1-\alpha)P_{\rm t}\xi_jg_j\sigma^2_j+\sigma^2_j+\sigma^2_{s_j}\big)^2}$$
$$\frac{\Pi^3_j\mathbf{G}_{[j,j]}}{2P_{\rm t}\xi_jg_j} = (\sigma^2_{s_j}+\sigma^2_j)P_{\rm t}\xi_jg_j+(1-\alpha)(3-2\alpha)\Lambda_j\sigma^2_j\sum_{i\in\mathcal{I}_j}\bar{q}_{i,j}+$$
$$\big(\sigma^2_{s_j}+\sigma^2_j+(2\sigma^2_{s_j}+\sigma^2_j)(1-\alpha)P_{\rm t}\xi_jg_j\big)\sum_{i\in\mathcal{I}_j}q_{i,j}+$$
$$(1-\alpha)\sigma^2_{s_j}\sigma^4_j\big(1+(1-\alpha)P_{\rm t}\xi_jg_j\big)\Big(\sum_{i\in\mathcal{I}_j}q_{i,j}\Big)^2,$$

where $\Pi_j = (1-\alpha)P_{\rm t}\xi_jg_j\sigma^2_j+\sigma^2_j+\sigma^2_{s_j}$. Note that, for all j, $\mathbf{F}_{[j,j]} \ge 0$ and $\mathbf{G}_{[j,j]} \ge 0$ for any $\alpha \in \mathcal{T}$.

The second-order derivative of $\operatorname{tr}\big(\mathbf{R}_\epsilon\big)$ can be written as

$$\frac{\partial^2\operatorname{tr}\big(\mathbf{R}_\epsilon\big)}{\partial\alpha^2} = 2\operatorname{tr}\big(\mathbf{F}\mathbf{R}_\epsilon\mathbf{F}\mathbf{R}_\epsilon\mathbf{R}_\epsilon\big)+\operatorname{tr}\big(\mathbf{R}_\epsilon\mathbf{G}\mathbf{R}_\epsilon\big). \tag{49}$$

For any $\alpha \in \mathcal{T}$, note that $\operatorname{tr}\big(\mathbf{F}\mathbf{R}_\epsilon\mathbf{F}\mathbf{R}_\epsilon\mathbf{R}_\epsilon\big) \ge 0$ because $\mathbf{F}\mathbf{R}_\epsilon\mathbf{F} = \big(\mathbf{R}^{0.5}_\epsilon\mathbf{F}\big)^T\mathbf{R}^{0.5}_\epsilon\mathbf{F}$ and $\mathbf{R}_\epsilon\mathbf{R}_\epsilon = \mathbf{R}^T_\epsilon\mathbf{R}_\epsilon$ are PSD[1], and for any real PSD matrices $\mathbf{A}$ and $\mathbf{B}$ following holds $\operatorname{tr}\big(\mathbf{A}\mathbf{A}\big) \ge 0$. Moreover note that $\operatorname{tr}\big(\mathbf{R}_\epsilon\mathbf{G}\mathbf{R}_\epsilon\big) = \sum_{j\in\mathcal{J}}\sum_{k\in\mathcal{J}}\mathbf{G}_{[k,k]}\mathbf{R}^2_{\epsilon[j,k]} > 0$. Therefore, $\partial^2\operatorname{tr}\big(\mathbf{R}_\epsilon\big)/\partial\alpha^2 > 0$ for any $\alpha \in \mathcal{T}$ and thus shows that $\operatorname{tr}\big(\mathbf{R}_\epsilon\big)$ is strictly convex over $\mathcal{T}$. This strict convexity proves the existence and uniqueness of the minimizer of $\operatorname{tr}\big(\mathbf{R}_\epsilon\big)$ over $\mathcal{T}$.

[1]Note that $\mathbf{R}_\epsilon$ is a covariance matrix and thus is PSD; for any PSD matrix $\mathbf{A}$ its square root (i.e., $\mathbf{A}^{0.5}$) exists such that $\mathbf{A} = (\mathbf{A}^{0.5})^T\mathbf{A}^{0.5}$. Also, $\mathbf{A}^T\mathbf{A}$ is PSD for any real matrix $\mathbf{A}$.

Appendix E
Convexity of $\mathrm{tr}(\mathbf{R}_\epsilon)$ over ξ_j's for Correlated Sources

The first-order derivative of $\mathrm{tr}(\mathbf{R}_\epsilon)$ w.r.t. ξ_j, for all j, is given by

$$\frac{\partial\,\mathrm{tr}(\mathbf{R}_\epsilon)}{\partial \xi_j} = \frac{(\alpha-1)P_{\mathrm{t}}(\sigma_j^2+\sigma_{s_j}^2)g_j r_j}{((1-\alpha)P_{\mathrm{t}}\xi_j g_j \sigma_j^2 + \sigma_j^2 + \sigma_{s_j}^2)^2}, \tag{50}$$

where $r_j = \sum_{k\in\mathcal{J}} \mathbf{R}^2_{\epsilon[j,k]}$. Note that $\partial\,\mathrm{tr}(\mathbf{R}_\epsilon)/\partial\xi_j < 0$ for any valid ξ_j, which shows that the objective function is decreasing over ξ_j. To prove the convexity of the problem, it is sufficient to show that the objective function is convex. For this purpose, we can write the Hessian of $\mathrm{tr}(\mathbf{R}_\epsilon)$ w.r.t. ξ_j's as

$$\nabla_\xi^2\,\mathrm{tr}(\mathbf{R}_\epsilon) = 2\mathbf{B} + 2(\mathbf{A}\mathbf{R}_\epsilon\mathbf{A})\circ(\mathbf{R}_\epsilon\mathbf{R}_\epsilon), \tag{51}$$

where $\mathbf{A}$ and $\mathbf{B}$ are diagonal matrices with

$$\mathbf{A}_{[j,j]} = \frac{\sigma_j^2+\sigma_{s_j}^2}{((1-\alpha)P_{\mathrm{t}}\xi_j g_j\sigma_j^2+\sigma_j^2+\sigma_{s_j}^2)^2},$$
$$\mathbf{B}_{[j,j]} = \frac{(1-\alpha)^2P_t^2(\sigma_j^2+\sigma_{s_j}^2)g_j^2\sigma_j^2 r_j}{((1-\alpha)P_{\mathrm{t}}\xi_j g_j\sigma_j^2+\sigma_j^2+\sigma_{s_j}^2)^3},$$

for all j. In (51) $\mathbf{A}\circ\mathbf{B}$ denotes the Hadamard product of $\mathbf{A}$ and $\mathbf{B}$. Note that $\mathbf{R}_\epsilon\mathbf{R}_\epsilon = \mathbf{R}_\epsilon^T\mathbf{R}_\epsilon$ and $\mathbf{A}\mathbf{R}_\epsilon\mathbf{A} = (\mathbf{R}_\epsilon^{0.5}\mathbf{A})^T\mathbf{R}_\epsilon^{0.5}\mathbf{A}$ are always PSD. Besides, $\mathbf{B}$ is PSD and the Hadamard product of two PSD matrices is always a PSD matrix by the *Schur product theorem*. We conclude that the Hessian $\nabla_\xi^2\,\mathrm{tr}(\mathbf{R}_\epsilon)$ is PSD, which in turn proves that the function is jointly convex over ξ_j's.

Appendix F
Convexity of $\mathrm{tr}(\mathbf{R}_\epsilon)$ over γ_j's for Correlated Sources

Let $\tilde{b}_j := \sum_{j\in\mathcal{J}} \alpha q_{i,j}/\gamma_j$ and $\bar{b}_j := \sum_{i\in\mathcal{I}_j}\alpha^2\bar{q}_{i,j}/\gamma_j^2$ with $q_{i,j}$ and $\bar{q}_{i,j}$ defined in Appendix B. For any valid γ_j, we can show that

$$\frac{\partial\,\mathrm{tr}(\mathbf{R}_\epsilon)}{\partial\gamma_j} = \frac{(\alpha-1)P_{\mathrm{t}}\xi_j g_j((1-\alpha)P_{\mathrm{t}}\xi_j g_j+1)\sigma_j^4\tilde{b}_j r_j}{((1-\alpha)P_{\mathrm{t}}\xi_j g_j\sigma_j^2+\sigma_j^2+\sigma_{s_j}^2)^2} \tag{52}$$

is negative for all j, which tells us that $\mathrm{tr}(\mathbf{R}_\epsilon)$ is a decreasing function of γ_j's. Moreover, the Hessian of $\mathrm{tr}(\mathbf{R}_\epsilon)$ w.r.t. γ_j's can be given by

$$\nabla_\gamma^2\,\mathrm{tr}(\mathbf{R}_\epsilon) = 2\bar{\mathbf{B}} + 2(\bar{\mathbf{A}}\mathbf{R}_\epsilon\bar{\mathbf{A}})\circ(\mathbf{R}_\epsilon\mathbf{R}_\epsilon), \tag{53}$$

which is PSD by the same arguments as employed in Appendix E. In (53) $\bar{\mathbf{A}}$ and $\bar{\mathbf{B}}$ are diagonal matrices with

$$\bar{\mathbf{A}}_{[j,j]} = \frac{(1-\alpha)P_{\mathrm{t}}\xi_j g_j((1-\alpha)P_{\mathrm{t}}\xi_j g_j+1)\sigma_j^4\tilde{b}_j}{((1-\alpha)P_{\mathrm{t}}\xi_j g_j\sigma_j^2+\sigma_j^2+\sigma_{s_j}^2)^2},$$
$$\bar{\mathbf{B}}_{[j,j]} = \frac{\bar{\mathbf{A}}_{[j,j]}}{\tilde{b}_j}\left(\frac{\sigma_j^2\sigma_{s_j}^2\tilde{b}_j^2}{(1-\alpha)P_{\mathrm{t}}\xi_j g_j\sigma_j^2+\sigma_j^2+\sigma_{s_j}^2}+\bar{b}_j\right)r_j,$$

for all j. The preceding discussion proves that the given optimization problem is jointly convex over γ_j's.

Appendix G
Convexity of $\mathrm{tr}(\mathbf{R}_\epsilon)$ over $\beta_{i,j}$'s for Correlated Sources

For any non-zero vector $\mathbf{v}\in\mathcal{R}^{N_j-1}$, we can show that

$$\mathbf{v}^T\nabla_{\beta_j}^2\,\mathrm{tr}(\mathbf{R}_\epsilon)\mathbf{v} = a_j\sum_{i\in\mathcal{I}_j}v_i^2\check{q}_{i,j} + \bar{a}_j\Big(\sum_{i\in\mathcal{I}_j}v_i\tilde{q}_{i,j}\Big)^2, \tag{54}$$

where $\nabla_{\beta_j}^2\,\mathrm{tr}(\mathbf{R}_\epsilon)$ is Hessian of $\mathrm{tr}(\mathbf{R}_\epsilon)$ w.r.t. $\beta_{i,j}$'s for each j, $\check{q}_{i,j} = \alpha^2\bar{q}_{i,j}/\beta_{i,j}^2$, $\tilde{q}_{i,j} = \alpha q_{i,j}/\beta_{i,j}$, $a_j = 2\bar{\mathbf{A}}_{[j,j]}r_j/\tilde{b}_j$, and

$$\bar{a}_j = \frac{\sigma_{s_j}^2\sigma_j^2 a_j}{(1-\alpha)P_{\mathrm{t}}\xi_j g_j\sigma_j^2+\sigma_j^2+\sigma_{s_j}^2} + \frac{\bar{\mathbf{A}}_{[j,j]}}{\tilde{b}_j}a_j\mathbf{R}_{\epsilon[j,j]}.$$

As each of a_j, $\bar{a}_j$, and $\check{q}_{i,j}$ is positive, therefore $\mathbf{v}^T\nabla_{\beta_j}^2\,\mathrm{tr}(\mathbf{R}_\epsilon)\,\mathbf{v} \geq 0$, which shows that the Hessian is PSD and thus proves that, for each j, the given optimization problem is jointly convex w.r.t. $\beta_{i,j}$'s. We also have $\partial\,\mathrm{tr}(\mathbf{R}_\epsilon)/\partial\beta_{i,j} = -\tilde{q}_{i,j}a_j/2 < 0$ for any valid $\beta_{i,j}$, which means $\mathrm{tr}(\mathbf{R}_\epsilon)$ is a decreasing function of $\beta_{i,j}$.

Appendix H
Proof of Prop. 3

To prove Prop. 3, we use the following lemma from [8].

Lemma 1: Let $\mathbf{G}$ and $\mathbf{Q}$ are two positive definite matrices, then

$$\mathrm{tr}\left(\mathbf{G}^T\mathbf{Q}^{-1}\mathbf{G}\right) \geq \frac{\left(\mathrm{tr}\left(\mathbf{G}^T\mathbf{G}\right)\right)^2}{\mathrm{tr}\left(\mathbf{G}^T\mathbf{Q}\mathbf{G}\right)} \tag{55}$$

Proof: By defining $\mathbf{A} = \mathbf{G}^T\mathbf{Q}^{-0.5}$ $\mathbf{B} = \mathbf{G}^T\mathbf{Q}^{0.5}$, (55) follows from the following Cauchy–Swarz inequality $\mathrm{tr}\left(\mathbf{A}\mathbf{A}^T\right)\mathrm{tr}\left(\mathbf{B}\mathbf{B}^T\right) \geq \left(\mathrm{tr}\left(\mathbf{A}\mathbf{B}^T\right)\right)^2$, which concludes the proof of the lemma. □

Now, from (8) we can write

$$\mathrm{tr}(\mathbf{R}_\epsilon) = \mathrm{tr}(\mathbf{R}_\mathbf{s}) - \mathrm{tr}\left(\mathbf{R}_\mathbf{s}\mathbf{H}^T\left(\mathbf{H}\mathbf{R}_\mathbf{s}\mathbf{H}^T+\mathbf{R}\right)^{-1}\mathbf{H}\mathbf{R}_\mathbf{s}^T\right),$$
$$\leq \mathrm{tr}(\mathbf{R}_\mathbf{s}) - \frac{\left(\mathrm{tr}\left(\mathbf{R}_\mathbf{s}\mathbf{H}^T\mathbf{H}\mathbf{R}_\mathbf{s}^T\right)\right)^2}{\mathrm{tr}\left(\mathbf{R}_\mathbf{s}\mathbf{H}^T\left(\mathbf{H}\mathbf{R}_\mathbf{s}\mathbf{H}^T+\mathbf{R}\right)\mathbf{H}\mathbf{R}_\mathbf{s}^T\right)}, \tag{56}$$

which follows from (55) with $\mathbf{G} = \mathbf{H}\mathbf{R}_\mathbf{s}^T$ and $\mathbf{Q} = \mathbf{H}\mathbf{R}_\mathbf{s}\mathbf{H}^T + \mathbf{R}$. We can show that

$$\mathrm{tr}\left(\mathbf{R}_\mathbf{s}\mathbf{H}^T\mathbf{H}\mathbf{R}_\mathbf{s}^T\right) = \mathrm{tr}\left(\mathbf{H}^T\mathbf{H}\mathbf{R}_\mathbf{s}^T\mathbf{R}_\mathbf{s}\right)$$
$$= (1-\alpha)P_{\mathrm{t}}\sum_{j\in\mathcal{J}}\xi_j g_j\Psi_j, \tag{57}$$

where $\Psi_j = \sum_{k\in\mathcal{J}}\mathbf{R}^2_{\mathbf{s}[j,k]} = \sum_{k\in\mathcal{J}}\mathrm{Cov}\left\{S_j,S_k\right\}^2$. Moreover, we can show that

$$\mathrm{tr}\left(\mathbf{R}_\mathbf{s}\mathbf{H}^T\left(\mathbf{H}\mathbf{R}_\mathbf{s}\mathbf{H}^T+\mathbf{R}\right)\mathbf{H}\mathbf{R}_\mathbf{s}^T\right)$$
$$= \mathrm{tr}\left(\mathbf{R}_\mathbf{s}\mathbf{H}^T\mathbf{H}\mathbf{R}_\mathbf{s}^T\mathbf{R}_\mathbf{s}\mathbf{H}^T\mathbf{H}\right) + \mathrm{tr}\left(\mathbf{H}^T\mathbf{R}\mathbf{H}\mathbf{R}_\mathbf{s}^T\mathbf{R}_\mathbf{s}\right), \tag{58}$$

Next, we can write

$$\begin{aligned}\operatorname{tr}\left(\mathbf{H}^T\mathbf{R}\mathbf{H}\mathbf{R}_{\mathbf{s}}^T\mathbf{R}_{\mathbf{s}}\right) &= (1-\alpha)^2P_t^2\sum_{j\in\mathcal{J}}\xi_j^2g_j^2\Psi_j\sigma_j^2\\ &\quad+(1-\alpha)P_{\mathrm{t}}\sum_{j\in\mathcal{J}}\xi_jg_j\Psi_j(\sigma_j^2+\sigma_{s_j}^2),\\ \operatorname{tr}\left(\mathbf{R}_{\mathbf{s}}\mathbf{H}^T\mathbf{H}\mathbf{R}_{\mathbf{s}}^T\mathbf{R}_{\mathbf{s}}\mathbf{H}^T\mathbf{H}\right) &= \operatorname{diag}(\mathbf{H}^T\mathbf{H})^T\tilde{\mathbf{Q}}\operatorname{diag}(\mathbf{H}^T\mathbf{H})\\ &= (1-\alpha)^2P_t^2\sum_{j\in\mathcal{J}}\sum_{k\in\mathcal{J}}\xi_jg_j\xi_kg_k\tilde{\mathbf{Q}}_{[j,k]}, \quad (59)\end{aligned}$$

where $\tilde{\mathbf{Q}} = (\mathbf{R_s})\circ(\mathbf{R_s}^T\mathbf{R_s})$. By substituting (57) to (59) in (56) we get (30).

Appendix I
Proof of Prop. 4

In order to prove Prop. 4, we need to show that the following statements are equivalent, that is, (60)⇔(61).

$$\theta^* = \max_{\xi\in\mathcal{F}}\frac{\xi^T\mathbf{U}\xi}{\xi^T\mathbf{Q}\xi+\mathbf{q}^T\xi}, \quad (60)$$

$$\max_{\xi\in\mathcal{F}}\xi^T\mathbf{U}\xi-\theta^*\left(\xi^T\mathbf{Q}\xi+\mathbf{q}^T\xi\right)=0. \quad (61)$$

Let the solution of (60) be ξ^*. We can write for any $\xi\in\mathcal{F}$

$$\theta^* = \frac{\xi^{*T}\mathbf{U}\xi^*}{\xi^{*T}\mathbf{Q}\xi^*+\mathbf{q}^T\xi^*}\geq\frac{\xi^T\mathbf{U}\xi}{\xi^T\mathbf{Q}\xi+\mathbf{q}^T\xi}\geq 0, \quad (62)$$

which gives

$$\begin{aligned}&\xi^{*T}\mathbf{U}\xi^*-\theta^*\left(\xi^{*T}\mathbf{Q}\xi^*+\mathbf{q}^T\xi^*\right)=0,\\ &\xi^T\mathbf{U}\xi-\theta^*\left(\xi^T\mathbf{Q}\xi+\mathbf{q}^T\xi\right)\leq 0, \qquad \forall\xi\in\mathcal{F}. \quad (63)\end{aligned}$$

From (63), we have

$$\max_{\xi\in\mathcal{F}}\left\{\xi^T\mathbf{U}\xi-\theta^*\left(\xi^T\mathbf{Q}\xi+\mathbf{q}^T\xi\right)\right\}=0 \quad (64)$$

and the corresponding maximizer is ξ^*. This proves that (60) leads to (61) that is, (60)⇒(61). Similarly we can also show that (61)⇒(60). For this purpose, assume ξ^* be the solution of (61), which implies that for any $\xi\in\mathcal{F}$

$$\begin{aligned}&\xi^T\mathbf{U}\xi-\theta^*\left(\xi^T\mathbf{Q}\xi+\mathbf{q}^T\xi\right)\leq\\ &\xi^{*T}\mathbf{U}\xi^*-\theta^*\left(\xi^{*T}\mathbf{Q}\xi^*+\mathbf{q}^T\xi^*\right)=0. \quad (65)\end{aligned}$$

Dividing (65) by $\xi^{*T}\mathbf{Q}\xi^*+\mathbf{q}^T\xi^*$ gives

$$\theta^* = \frac{\xi^{*T}\mathbf{U}\xi^*}{\xi^{*T}\mathbf{Q}\xi^*+\mathbf{q}^T\xi^*} \quad (66)$$

and similarly dividing (65) by $\xi^T\mathbf{Q}\xi+\mathbf{q}^T\xi$ gives

$$\theta^*\geq\frac{\xi^T\mathbf{U}\xi}{\xi^T\mathbf{Q}\xi+\mathbf{q}^T\xi}. \quad (67)$$

From (66) and (67), we have

$$\max_{\xi\in\mathcal{F}}\frac{\xi^T\mathbf{U}\xi}{\xi^T\mathbf{Q}\xi+\mathbf{q}^T\xi}=\theta^*, \quad (68)$$

which proves (61)⇒(60). With this, we have proved (60)⇔(61).

J Proof of Prop. 5

For some θ_ι let

$$\begin{aligned}\xi(\theta_\iota) &= \arg\max_{\xi\in\mathcal{F}} f(\xi;\theta_\iota)\\ \varphi(\theta_\iota) &= f(\xi(\theta_\iota);\theta_\iota), \quad (69)\end{aligned}$$

then for $\theta_2>\theta_1\geq 0$ it is easy to show that

$$\varphi(\theta_2) = f(\xi(\theta_2);\theta_2)\overset{(a)}{<}f(\xi(\theta_2);\theta_1), \quad (70)$$

$$\varphi(\theta_1) = f(\xi(\theta_1);\theta_1)\overset{(b)}{\geq}f(\xi(\theta_2);\theta_1), \quad (71)$$

where the inequality (a) follows from the definition of $f(\xi;\theta)$ and the inequality (b) results from the sub-optimality of $\xi(\theta_2)$ for θ_1. Combining (70) and (71) we get

$$\varphi(\theta_2)<\varphi(\theta_1), \quad (72)$$

which proves that $\varphi(\theta)$ is a strictly decreasing function of θ. As defined in Prop. 4, for some θ^* such that $\varphi(\theta^*)=0$, the decreasing nature of $\varphi(\theta)$ means that $\varphi(\theta)>0$ for $\theta<\theta^*$ and $\varphi(\theta)<0$ for $\theta>\theta^*$.

Moreover, for some $\omega\in[0,1]$ we can write

$$\begin{aligned}\omega\varphi(\theta_1)+(1-\omega)\varphi(\theta_2) &= \omega f(\xi(\theta_1);\theta_1)+(1-\omega)f(\xi(\theta_2);\theta_2)\\ &\overset{(c)}{\geq}\omega f(\xi(\theta_1);\theta_1)+(1-\omega)f(\xi(\theta_1);\theta_2)\\ &=\omega\left\{\xi(\theta_1)^T\mathbf{U}\xi(\theta_1)-\theta_1\left(\xi(\theta_1)^T\mathbf{Q}\xi(\theta_1)+\mathbf{q}^T\xi(\theta_1)\right)\right\}+\\ &\quad(1-\omega)\left\{\xi(\theta_1)^T\mathbf{U}\xi(\theta_1)-\theta_2\left(\xi(\theta_1)^T\mathbf{Q}\xi(\theta_1)+\mathbf{q}^T\xi(\theta_1)\right)\right\}\\ &=\xi(\theta_1)^T\mathbf{U}\xi(\theta_1)-(\omega\theta_1+(1-\omega)\theta_2)\Big(\xi(\theta_1)^T\mathbf{Q}\xi(\theta_1)\\ &\qquad+\mathbf{q}^T\xi(\theta_1)\Big)\\ &=f(\xi(\theta_1);\omega\theta_1+(1-\omega)\theta_2)\overset{(d)}{\geq}\varphi(\omega\theta_1+(1-\omega)\theta_2)), \quad (73)\end{aligned}$$

where the inequality (c) is due to the sub-optimality of $\xi(\theta_1)$ for $\varphi(\theta_2)$ and the inequality (d) is due to the decreasing nature of $\varphi(\theta)$. Thus (72) together with (73) prove that $\varphi(\theta)$ is a strictly decreasing convex function.

References

[1] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensor networks: a survey," *Elsevier Jr. on Computer Nets.*, vol. 38, no. 4, pp. 393–422, Mar. 2002.

[2] 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 Proces.*, vol. 54, no. 2, pp. 413–422, Feb. 2006.

[3] J. Y. Wu, Q. Z. Huang, and T. S. Lee, "Minimal energy decentralized estimation via exploiting the statistical knowledge of sensor noise variance," *IEEE Trans. on Signal Proces.*, vol. 56, no. 5, pp. 2171–2176, May 2008.

[4] J. Li and G. AlRegib, "Distributed estimation in energy-constrained wireless sensor networks," *IEEE Trans. on Signal Proces.*, vol. 57, no. 10, pp. 3746–3758, Oct. 2009.

[5] 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 Proces.*, vol. 55, no. 9, pp. 4683–4695, Sep. 2007.

[6] H. Senol and C. Tepedelenlioglu "Performance of distributed estimation over unknown parallel fading channels," *IEEE Trans. on Signal Proces.*, vol. 56, no. 12, pp. 6057–6068, Dec. 2008.

[7] I. bahceci, and A. K. Khandani, "Linear estimation of correlated data in wireless sensor networks with optimum power allocation and analog modulation," *IEEE Trans. on Comm.*, vol. 56, no. 7, pp. 1146–1156, July 2008.

[8] J. Fang and H. Li "Power constrained distributed estimation with correlated sensor data," *IEEE Trans. on Signal Proces.*, vol. 57, no. 10, pp. 4052–4065, Oct. 2009.

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

[10] 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 Comm. and Net.*, vol. 2010, pp. 1–14, Article ID 817961, 2010.

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

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

[13] J. Matamoros and C. Antón-Haro, "Hierarchical organizations of sensors for decentralized parameter estimation," *IEEE Int. Symp. on Signal Proces. and Inform. Techn. (ISSPIT'07)*, Cairo, Dec. 2007.

[14] J. Li and G. AlRegib, "Energy-efficient cluster-based distributed estimation in wireless sensor networks," *IEEE Conf. on Milt. Comm. (MILCOM'06)*, Washington D.C., Oct. 2006.

[15] C.-H. Wang and S. Dey, "Distortion outage minimization in Nakagami fading using limited feedback," *EURASIP Jr. on Advances in Signal Proces.*, pp. 1–16, doi:10.1186/1687-6180-2011-92, 2011.

[16] C.-H. Wang and S. Dey, "Distortion outage minimization in Rayleigh fading using limited feedback," *IEEE Global Comm. Conf. (GLOBECOM'09)*, Honolulu, Nov.–Dec. 2009.

[17] J. Fang and H. Li, "Power constrained distributed estimation with cluster-based sensor collaboration," *IEEE Trans. on Signal Proces.*, vol. 8, no. 7, pp. 3822–3832, July 2009.

[18] A. Dabirmoghaddam, M. Ghaderi, and C. Williamson, "Cluster-based correlated data gathering in wireless sensor networks," *IEEE Int. Symp. on Modeling, Analysis & Simulation of Computer and Telecomm. Sys. (MASCOTS'10)*, Miami, Aug. 2010.

[19] R. Rajesh and V. Sharma, "Amplify and forward for correlated data gathering over hierarchical sensor networks," *IEEE Conf. on Wireless Comm. and Net. (WCNC'09)*, Budapest, Apr. 2009.

[20] K. Yuen, B. Liang, and B. Li, "A distributed framework for correlated data gathering in sensor networks," *IEEE Trans. on Veh. Tech.*, vol. 57, no. 1, pp. 578–593, Jan. 2008.

[21] A. A. Abbasi and M. Younis, "A survey on clustering algorithms for wireless sensor networks," *Elsevier Jr. on Computer Comm.*, vol. 30, no. 14–15, pp. 2826–2841, Oct. 2007.

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

[23] K. B. Petersen and M. S. Pedersen, *The Matrix Cookbook*, 2008.

[24] S. Murugesan, "Harnessing green IT: principles and practices," *IEEE IT Professional*, vol. 10, no. 1, pp. 24–33, Jan.–Feb. 2008.

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

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

[27] D. P. Palomar and M. Chiang, "A tutorial on decomposition methods for network utility maximization," *IEEE Jr. on Sel. Areas in Comm.*, vol. 24, no. 8, pp. 1439–1451, Aug. 2006.

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

[29] H. Frenk, and S. Schaible *Fractional Programming*. In: C. A. Floudas, and P. M. Pardolas (Eds), *Encyclopedia of Optimization*, 2*nd* Edition, Springer Press, NK, USA, 2009.

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

[31] M. H. Chaudhary and L. Vandendorpe "Power constrained linear estimation of correlated sources in hierarchical wireless sensor networks," *IEEE 12th Int. Works. on Signal Proces. Adv. in Wireless Comm. (SPAWC'11)*, San Francisco, June 2011.

[32] M. H. Chaudhary and L. Vandendorpe "Power scheduling for distributed estimation in cluster-based wireless sensor networks," *IEEE Int. Conf. on Comm. (ICC'12)*, Ottawa, pp. 344–349, June 2012.

Muhammad Hafeez Chaudhary was born in the Punjab region of Pakistan. He attended Govt. College University Lahore, Pakistan, for his pre-engineering studies. He received his B.Sc. Electrical Engineering degree with *honours* from the University of Engineering and Technology Lahore in 2001. He worked at the National Engineering and Scientific Commission Islamabad, Pakistan, from 2001 to 2005 where he was involved in the research and development of sensor systems.

In 2007, he did European Master of Science in Research on Information and Communication Technologies (MERIT), a double-degree M.S. in Electrical Engineering jointly organized by the Université catholique de Louvain (UCL), Louvain-La-Neuve, Belgium and the Karlsruhe Institute of Technology, Karlsruhe, Germany with the overall grade *sehr gut* (very good: an *outstanding achievement*). He received his Ph.D. degree in Electrical Engineering from UCL in 2012. From October 2007 to November 2011, he was with the Institute of Information and Communication Technologies, Electronics and Applied Mathematics (ICTEAM) of UCL as a research assistant. Since November 2011, he is affiliated with the Royal Military Academy Belgium as a researcher. His research interests include algorithm and protocol design for information communication and processing in next generation wireless networks.

Luc Vandendorpe (M'93–SM'99–F'06) was born in Mouscron, Belgium, in 1962. He received the Electrical Engineering degree (summa cum laude) and the Ph.D. degree from the Université catholique de Louvain (UCL), Louvain-la-Neuve, Belgium, in 1985 and 1991, respectively.

Since 1985, he has been with the Communications and Remote Sensing Laboratory of UCL, where he first worked in the field of bit rate reduction techniques for video coding. In 1992, he was a Visiting Scientist and Research Fellow at the Telecommunications and Traffic Control Systems Group of the Delft Technical University, The Netherlands, where he worked on spread spectrum techniques for personal communications systems. From October 1992 to August 1997, he was Senior Research Associate of the Belgian NSF at UCL, and invited Assistant Professor. He is currently a Professor and head of the Institute for Information and Communication Technologies, Electronics and Applied Mathematics. His current interest is in digital communication systems and more precisely resource allocation for OFDM(A)-based multicell systems, MIMO and distributed MIMO, sensor networks, turbo-based communications systems, physical layer security and UWB based positioning.

Dr. Vandendorpe was corecipient of the 1990 Biennal Alcatel–Bell Award from the Belgian NSF for a contribution in the field of image coding. In 2000, he was corecipient (with J. Louveaux and F. Deryck) of the Biennal Siemens Award from the Belgian NSF for a contribution about filter-bank-based multicarrier transmission. In 2004, he was co-winner (with J. Czyz) of the Face Authentication Competition, FAC 2004. He is or has been TPC member for numerous IEEE conferences (VTC Fall, Globecom Communications Theory Symposium, SPAWC, ICC) and for the Turbo Symposium. He was Co-Technical Chair (with P. Duhamel) for the IEEE ICASSP 2006. He was an Editor for Synchronization and Equalization of the IEEE TRANSACTIONS ON COMMUNICATIONS between 2000 and 2002, Associate Editor of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS between 2003 and 2005, and Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING between 2004 and 2006. He was Chair of the IEEE Benelux joint chapter on Communications and Vehicular Technology between 1999 and 2003. He was an elected member of the Signal Processing for Communications committee between 2000 and 2005, and an elected member of the Sensor Array and Multichannel Signal Processing committee of the Signal Processing Society between 2006 and 2008. Currently, he is an elected member of the Signal Processing for Communications committee. He is the Editor-in-Chief for the EURASIP Journal on Wireless Communications and Networking.