Wireless Power Transfer for Distributed Estimation in Sensor Networks

Vien V. Mai, Won-Yong Shin, *Senior Member, IEEE*, and Koji Ishibashi, *Member, IEEE*

***Abstract*—This paper studies power allocation for distributed estimation of an unknown scalar random source in sensor networks with a multiple-antenna fusion center (FC), where wireless sensors are equipped with radio-frequency based energy harvesting technology. The sensors' observation is locally processed by using an *uncoded amplify-and-forward* scheme. The processed signals are then sent to the FC, and are coherently combined at the FC, at which the best linear unbiased estimator (BLUE) is adopted for reliable estimation. We aim to solve the following two power allocation problems: 1) minimizing distortion under various power constraints; and 2) minimizing total transmit power under distortion constraints, where the distortion is measured in terms of mean-squared error of the BLUE. Two iterative algorithms are developed to solve the nonconvex problems, which converge at least to a local optimum. In particular, the above algorithms are designed to jointly optimize the *amplification coefficients*, *energy beamforming*, and *receive filtering*. For each problem, a suboptimal design, a single-antenna FC scenario, and a *common harvester* deployment for colocated sensors, are also studied. Using the powerful *semidefinite relaxation* framework, our result is shown to be valid for any number of sensors, each with different noise power, and for an arbitrarily number of antennas at the FC.**

***Index Terms*—Amplify-and-forwarding, best linear unbiased estimator (BLUE), distributed estimation, mean-squared error (MSE), wireless power transfer (WPT).**

I. Introduction

Distributed inference in wireless sensor networks (WSNs) has been extensively studied for applications such as environmental monitoring, weather forecasts, health care, and home automation (see, e.g., [1]–[7] and references therein). Sensors in WSNs are powered typically by batteries, and hence the network lifetime is highly limited. In practice, periodically replacing or recharging batteries may be hard or even impossible (due to the fact that sensors are located inside toxic environments, building structures, or human bodies [8]). Therefore, although there have been many efforts in power management policies, the network lifetime remains a performance bottleneck and limits the wide-range deployment of WSNs.

This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2014R1A1A2054577) and by the Ministry of Science, ICT & Future Planning (MSIP) (2015R1A2A1A15054248).

V. V. Mai was with Dankook University, Yongin 448-701, Republic of Korea. He is now with the Department of Automatic Control, KTH Royal Institute of Technology, SE-100 44 Stockholm, Sweden. (e-mail: maivv@kth.se).

W.-Y. Shin (corresponding author) is with the Department of Computer Science and Engineering, Dankook University, Yongin 448-701, Republic of Korea (e-mail: wyshin@dankook.ac.kr)

K. Ishibashi is with the Advanced Wireless & Communication Research Center (AWCC), The University of Electro-Communications, Tokyo 182-8585, Japan (e-mail: koji@ieee.org)

A. Previous Work

The optimal power allocation strategies for distributed estimation in WSNs have received a great research interest both from analog and digital encoding perspectives [3]–[5], [9]–[13]. Among encoding schemes, the *uncoded amplify-and-forward* scheme has been extensively studied due to its simplicity and information-theoretic-optimality properties under certain conditions [14]. In particular, the authors in [9] studied power allocation for orthogonal multiple access channels (MACs), when the best linear unbiased estimator (BLUE) is adopted. The same problem was considered in [3] for a coherent MAC. The effects of channel estimation error were reported in [10] for orthogonal MACs adopting a linear minimum mean-squared error estimator, while in [13], the sensing noise uncertainty was investigated by adopting the BLUE. Recently, the optimal transmit strategy for cooperative linear estimation was studied in [4].

The tremendous performance gains achieved by multiple-antenna techniques highly motivate us to integrate this technology into future wireless systems including WSNs. The benefits of such technology in the context of WSNs have been recently studied for distributed inference [5], [7], [15]–[18]. For a large-scale fusion center (FC) over a Rayleigh fading channel, it has been shown in [17] that the detection/estimation performance remains asymptotically constant if the transmit power at each sensor decreases proportionally with increasing number of antennas at the FC. The benefits of the multiple-antenna FC in distributed detection were analyzed in terms of asymptotic error exponents in [7]. Power allocation strategies for distributed estimation were studied for the correlated source case [18] and for the correlated noise case [5].

Although the network life span can be prolonged by applying the aforementioned strategies, one needs a disruptive design that fundamentally changes the limitation of a WSN. One of the promising solution is the so-called *energy harvesting* (EH), in which sensors scavenge energy from the ambient environment (e.g., solar, wind, and vibration) that can guarantee an infinite life span in theory [19]. However, due to the unpredictable nature of energy sources, EH is typically uncontrolled, and thus can be critical for some reliable-sensitive applications. In addition to commonly used energy sources such as solar and wind, ambient radio-frequency (RF) signals can be a viable new source for energy scavenging. Most of the researches on *wireless power transfer* (WPT) have been focused on *cellular* networks, where user terminals replenish energy from the received signals sent by the base station via the far-field RF-based WPT [8], [20]–[25]. For example, the fundamental trade-off between the achievable rate and the transferred power was characterized in [20]. Several practical receiver architectures for simultaneous information and power transfer were investigated in [8], [21]. Exploiting multiple antenna technologies in WPT has been widely studied: multiple-input-multiple-output broadcast channels [8], beamforming designs for multiuser multiple-input-single-output (MISO) [23], physical-layer security problems for multiuser MISO [24], and multiple-antenna interference channels [25]. On the other hand, there are a relatively limited number of studies on WPT for *WSNs*; different WPT technologies for addressing energy/lifetime problems in WSNs were reviewed in [26], [27]; in [28], the authors studied a distributed estimation system in which some of the multiple-antenna sensors, named super sensors, are capable of WPT to its neighboring sensors via beamforming; and in [29], several multiple-antenna RF-based chargers were used to replenish the wireless sensors and then to switch to the information transmission phase, where each sensor sent a quantized version of its measurement to the FC for estimation.

B. Main Contributions

For distributed estimation in WSNs, an important question is how to intelligently exploit multiple-antenna technologies and WPT to improve both the inference performance and network lifetime. In this paper, we devote to studying the optimality of WPT and the optimal allocation of harvested energy for distributed estimation of an unknown random source in WSNs with a multiple-antenna FC. Our main contributions are summarized as follows:

- When the BLUE is adopted at the FC for estimation, we jointly optimize the amplification coefficients, energy beamforming, and receiver filtering by adopting alternative minimization methods (see Algorithms 1 and 2). To that end, we first solve the mean-squared error (MSE) minimization problem under the total power constraint at the FC as well as the causal power constraint at each sensor. Then, we solve a converse problem where the total transmit power at the FC is minimized subject to an MSE requirement.
- A key ingredient of our algorithms is the so-called *semidefinite relaxation*. We show that such a relaxation does not sacrifice the optimality of the relaxed problems. We derive the properties of the optimal solutions (see Theorems 1 and 2).
- A special deployment of WPT in WSNs is also discussed, where a common energy harvester is used to collect energy from the FC. We show that the optimization problems are significantly simplified in this case. The optimal power–distortion trade-off is also characterized (see Theorem 3).

C. Organization

The rest of the paper is organized as follows. The system model and problem formulation are described in Section II.Section III studies the problem of minimizing the MSE subject to power constraints. In Section IV, the converse problem in Section III is studied. The numerical results are shown in Section V. Finally, we conclude the paper in Section VI.

D. Notations

The operators $(\cdot)^\top$, $(\cdot)^*$, $(\cdot)^\dagger$ are the transpose, complex conjugate, and transpose conjugate, respectively. The notation $\mathbf{I}_n$ denotes the $n \times n$ identity matrix; $\mathrm{tr}\,(\mathbf{A})$ denotes the trace of a matrix $\mathbf{A}$; $\mathsf{rank}\,(\mathbf{A})$ denotes the rank of a matrix $\mathbf{A}$; $\mathrm{diag}\,(\mathbf{a})$ denotes a diagonal matrix with vector $\mathbf{a}$ being its diagonal, $\mathbf{A} \succeq \mathbf{0}$ denotes the positive semidefinite $\mathbf{A}$; $\mathbb{E}\{\cdot\}$ denotes the expectation operator; $\dim\,(\mathbf{A})$ denotes a dimension of the subspace $\mathbf{A}$. We use the Bachmann–Landau notation: $f(x) = O(g(x))$ if $\lim_{x\to x_0} \frac{f(x)}{g(x)} = c < \infty$. Finally, we use the notation $[n]$ to denote the set of positive natural numbers up to n, i.e., $[n] = \{i : i = 1, 2, \ldots, n\}$.

II. System Model and Problem Formulation

A. System Model

As illustrated in Fig. 1, we consider a distributed estimation system where an n_{r}-antenna FC collects data from n_{s} spatially distributed sensors. Let θ be an unknown scalar random parameter (source) with variance of δ_θ^2 to be estimated.[1] Examples of such a parameter include physical phenomena such as pressure, temperature, sound intensity, radiation level, pollution concentration, seismic activity, etc.. We assume that all sensors do not have conventional energy supplies and hence need to harvest energy from the RF signal transferred by the FC for future use. We also assume that there is no cooperation among the sensors since they are spatially distributed. In this paper, we adopt a time-switching *harvest-then-forward* protocol [30] in which for each τT amount of time, where T is the length of one time slot, the FC transmits its energy signal to the sensors, and for the remaining $(1-\tau)T$ amount of time, the sensors observe and forward their observations to the FC for estimation while using the harvested energy from the RF signal. For analytical convenience, we set $\tau = 1/2$ in the sequel unless otherwise specified.[2]

In the first phase (i.e., the energy harvesting phase) of a time slot, the FC broadcasts its energy signal to the sensors through energy beamforming. More precisely, $n_{\mathrm{b}} \le n_{\mathrm{r}}$ energy beams are assigned to n_{s} sensors without loss of generality. The energy signal received at the kth sensor is then given by

$$r_k = \boldsymbol{g}_k^\dagger \boldsymbol{x}_{\mathrm{e}} + m_k = \boldsymbol{g}_k^\dagger \sum_{i=1}^{n_{\mathrm{b}}} \boldsymbol{w}_i s_i + m_k, \tag{1}$$

where $\boldsymbol{x}_{\mathrm{e}} = \sum_{i=1}^{n_{\mathrm{b}}} \boldsymbol{w}_i s_i$ is the energy signal transmitted from the FC; s_i is the energy-carrying signal for the ith energy beam

[1] Our WPT framework is not designed only for a scalar source. Note that there is no restriction to apply it for the vector case even if finding a theoretical optimal solution with no approximation for estimating vector-valued sources remains an open problem.

[2] For a Gaussian sensor network using the source-channel encoding strategy [14], the rate-distortion theorems (e.g., [31, Theorem 10.3.3] for orthogonal MACs and [14, Section IV] for a coherent MAC) enable us to characterize the effect of τ (via the rate expressions) on the distortion performance (see, e.g., [32] and references therein). In this work, we adopt an *analog uncoded amplify-and-forward* scheme without bandwidth expansion, in which the nature of information is in an analog form, but not in a bitwise form [14]. As a result, the power-distortion tradeoff (e.g., [33, Theorem 1]) is independent of τ, and hence in our work, the value of τ is assumed to be a constant. In practice, the value $T_0 = (1-\tau)T$ corresponds to the amount of time that each sensor needs for observing, amplifying, and forwarding its observation to the FC.

satisfying $\mathbb{E}\{|s_i|^2\} = 1$ and $\mathbb{E}\{s_i s_j\} = 0$ for $i \neq j$, which can be any arbitrary random signal provided that its power spectral density satisfies certain regulations on microwave radiation [23]; $\boldsymbol{w}_i \in \mathbb{C}^{n_\mathrm{r}\times 1}$ is the ith energy bemforming vector; $\boldsymbol{g}_k \in \mathbb{C}^{n_\mathrm{r}\times 1}$ is the channel between the FC and kth sensor; and m_k is the additive noise at the kth sensor. By ignoring the background noise for the sake of simplicity, the harvested energy at the kth sensor in each slot is given by [8]

$$E_k = \frac{\zeta_k T}{2}\sum_{i=1}^{n_\mathrm{b}} \left|\boldsymbol{w}_i^\dagger \boldsymbol{g}_k\right|^2, \tag{2}$$

where $0 \leq \zeta_k \leq 1$ is the energy harvesting efficiency at the kth sensor. Then, the average power P_k available for the information transmission phase at the kth sensor can be expressed as

$$P_k = \frac{2\left(E_k - E_k^{\mathrm{cir}}\right)}{T} = \zeta_k \sum_{i=1}^{n_\mathrm{b}} \left|\boldsymbol{w}_i^\dagger \boldsymbol{g}_k\right|^2 - \frac{2E_k^{\mathrm{cir}}}{T}, \tag{3}$$

where $E_k^{\mathrm{cir}} \geq 0$ is the circuit energy consumption at the kth sensor, which is assumed to be constant over time slots. Similarly as in [25], [30], we simply assume $\zeta_k = 1$ and unit slot duration in the rest of this work (note that using an arbitrary ζ_k does not fundamentally change our power allocation problems). Similarly as in [30], [34], for easy of presentation, we also assume that $\{E_k^{\mathrm{cir}}\}_{k=1}^{n_\mathrm{s}} = 0$ to focus on the transmit power of the sensors.[3] The FC has a total transmit power constraint P; we thus have

$$\mathbb{E}\left\{\boldsymbol{x}_\mathrm{e}^\dagger \boldsymbol{x}_\mathrm{e}\right\} = \sum_{i=1}^{n_\mathrm{b}} \|\boldsymbol{w}_i\|^2 \leq P. \tag{4}$$

Now, let us turn to describing the second phase (i.e., the information transmission phase) of a time slot. The observation at the kth sensor can be expressed as

$$x_k = \theta + u_k, \quad k = 1, \ldots, n_\mathrm{s}, \tag{5}$$

where u_k is the additive noise at the kth sensor with variance $\sigma_{\mathrm{u},k}^2$. The noise at each sensor is assumed to be independent of each other. In this paper, we adopt an analog uncoded amplify-and-forward scheme, i.e., the kth sensor just simply amplifies its observation by a factor α_k. Therefore, by stacking the transmit signals from all sensors into a single vector $\boldsymbol{t}$, it can be expressed as

$$\boldsymbol{t} = \mathbf{A}\mathbf{1}\theta + \mathbf{A}\boldsymbol{u}, \tag{6}$$

where $\mathbf{A} = \mathrm{diag}\left(\alpha_1, \ldots, \alpha_{n_\mathrm{s}}\right) \in \mathbb{C}^{n_\mathrm{s}\times n_\mathrm{s}}$ is the amplification matrix; $\boldsymbol{u} = [u_1\, u_2 \cdots u_{n_\mathrm{s}}]^\top \in \mathbb{C}^{n_\mathrm{s}\times 1}$ is the noise vector at the sensors with zero mean and covariance matrix $\boldsymbol{R}_\mathrm{s} = \mathrm{diag}\left(\sigma_{\mathrm{u},1}^2, \sigma_{\mathrm{u},2}^2, \ldots, \sigma_{\mathrm{u},n_\mathrm{s}}^2\right)$; and $\mathbf{1}$ is the all one vector. Then, the received signal $\boldsymbol{z} \in \mathbb{C}^{n_\mathrm{r}\times 1}$ at the FC can be written as

$$\boldsymbol{z} = \boldsymbol{H}\mathbf{A}\mathbf{1}\theta + \boldsymbol{H}\mathbf{A}\boldsymbol{u} + \boldsymbol{n}, \tag{7}$$

where $\boldsymbol{H} \in \mathbb{C}^{n_\mathrm{r}\times n_\mathrm{s}}$ is the channel between the sensors and FC; and $\boldsymbol{n} \in \mathbb{C}^{n_\mathrm{r}\times 1}$ is the noise vector at the FC with zero mean and covariance matrix $\boldsymbol{R}_\mathrm{n} = \mathrm{diag}\left(\sigma_{\mathrm{n},1}^2, \sigma_{\mathrm{n},2}^2, \ldots, \sigma_{\mathrm{n},n_\mathrm{r}}^2\right)$.

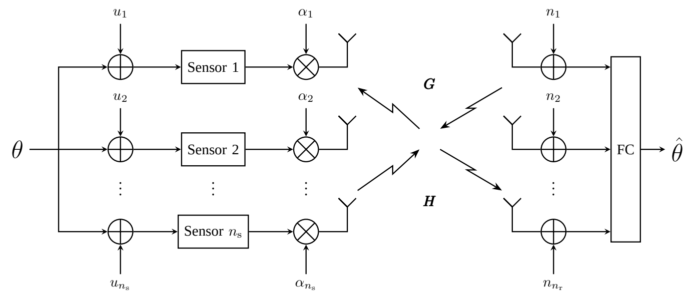

Fig. 1. The distributed estimation system with an n_r-antenna FC and n_s spatially distributed sensors where $\boldsymbol{G} = \left[\boldsymbol{g}_1\ \boldsymbol{g}_2 \cdots\ \boldsymbol{g}_{n_\mathrm{s}}\right]^\dagger$.

Here, the random quantities θ, $\boldsymbol{u}$, and $\boldsymbol{n}$ are statistically independent.

Since we consider a coherent MAC, we assume that there is perfect synchronization between the sensors and the FC. All wireless channels are assumed to be quasi-static flat fading, i.e., once each channel is realized, it remains fixed during each time slot and changes independently between slots. We further assume that full channel state information (CSI) is available at the FC. In practice, the sensors-to-FC channel $\boldsymbol{H}$ can be estimated at the FC via periodic pilot transmissions from the sensors, while the FC-to-sensors channels $\{\boldsymbol{g}_i\}_{i=1}^{n_\mathrm{s}}$ can be acquired owing to channel reciprocity between the sensors-to-FC and FC-to-sensors channels when the system operates in time-division-duplex mode.

B. Problem Formulation

The received signal $\boldsymbol{z}$ is constructively combined at the FC by a filtering vector $\boldsymbol{v} \in \mathbb{C}^{n_\mathrm{r}\times 1}$. Then, by adopting the well-known BLUE [35, Theorem 6.1], the FC estimates the parameter θ based on the minimal sufficient statistic $y = \boldsymbol{v}^\dagger \boldsymbol{z}$ as follows:[4].

$$\hat{\theta} = \left[\boldsymbol{a}^\dagger \boldsymbol{H}^\dagger \boldsymbol{v}\sigma_{\mathrm{tot}}^{-2}\boldsymbol{v}^\dagger \boldsymbol{H}\boldsymbol{a}\right]^{-1} \boldsymbol{a}^\dagger \boldsymbol{H}^\dagger \boldsymbol{v}\sigma_{\mathrm{tot}}^{-2} y, \tag{8}$$

where $\sigma_{\mathrm{tot}}^2 = \boldsymbol{v}^\dagger\left[\boldsymbol{H}\mathbf{A}\boldsymbol{R}_\mathrm{s}\mathbf{A}^\dagger\boldsymbol{H}^\dagger + \boldsymbol{R}_\mathrm{n}\right]\boldsymbol{v}$ is the total noise power after post-processing at the FC; and $\boldsymbol{a} = [\alpha_1\, \alpha_2\, \ldots\, \alpha_{n_\mathrm{s}}]^\top$. The MSE of the BLUE can be written as

$$\begin{aligned} \mathsf{mse} = \mathbb{E}\left\{|\theta - \hat{\theta}|^2\right\} &= \left[\boldsymbol{a}^\dagger \boldsymbol{H}^\dagger \boldsymbol{v}\sigma_{\mathrm{tot}}^{-2}\boldsymbol{v}^\dagger \boldsymbol{H}\boldsymbol{a}\right]^{-1} \\ &= \left[\frac{\left|\boldsymbol{v}^\dagger\boldsymbol{H}\boldsymbol{a}\right|^2}{\boldsymbol{v}^\dagger\left(\boldsymbol{H}\mathbf{A}\boldsymbol{R}_\mathrm{s}\mathbf{A}^\dagger\boldsymbol{H}^\dagger + \boldsymbol{R}_\mathrm{n}\right)\boldsymbol{v}}\right]^{-1}. \end{aligned} \tag{9}$$

Since the three quantities $\boldsymbol{a}$, $\{\boldsymbol{w}_k\}_{k=1}^{n_\mathrm{b}}$, and $\boldsymbol{v}$ critically affect both the power requirement and estimation performance of the entire system, we jointly design the optimum sensor amplification coefficients $\boldsymbol{a}$, receive filtering vector $\boldsymbol{v}$, and energy beamforming $\{\boldsymbol{w}_k\}_{k=1}^{n_\mathrm{b}}$ under practical constraints. To that end, we solve two types of minimization problems: 1)

[3]Otherwise, we can rewrite our problem along with a power offset, i.e., $E_k^{\mathrm{cir}} > 0$, as a problem without any power offset for a smaller ζ_k.

[4]The *minimal sufficient statistic* is defined in the sense that we no longer need the individual sample since all the information has been captured by the sufficient statistic [35]

minimizing the MSE of the BLUE under causal individual power constraints at the sensors and a total power constraint at the FC; and 2) minimizing the total power consumed at the FC given a minimum requirement of the MSE. In particular, we aim to find the solution to the first problem, named (P1), by solving the following optimization problem.

(P1) :

$$\begin{aligned} \underset{\boldsymbol{v},\boldsymbol{a},\{\boldsymbol{w}_i\}_{i=1}^{n_\mathrm{b}}}{\text{maximize}} \quad & \frac{\left|\boldsymbol{v}^\dagger \boldsymbol{H}\boldsymbol{a}\right|^2}{\boldsymbol{v}^\dagger\left(\boldsymbol{H}\mathbf{A}\boldsymbol{R}_\mathrm{s}\mathbf{A}^\dagger\boldsymbol{H}^\dagger+\boldsymbol{R}_\mathrm{n}\right)\boldsymbol{v}} \\ \text{subject to} \quad & |\alpha_k|^2\left(\delta_\theta^2+\sigma_{\mathrm{u},k}^2\right)\le\sum_{i=1}^{n_\mathrm{b}}\left|\boldsymbol{w}_i^\dagger\boldsymbol{g}_k\right|^2, \quad \forall k\in[n_\mathrm{s}] \\ & \sum_{i=1}^{n_\mathrm{b}}\|\boldsymbol{w}_i\|^2\le P. \end{aligned}$$

As a counterpart of (P1), for a given MSE threshold $\mathsf{mse} = 1/\gamma$, the second optimization problem is stated as follows.

(P2) :

$$\begin{aligned} \underset{\boldsymbol{v},\boldsymbol{a},\{\boldsymbol{w}_i\}_{i=1}^{n_\mathrm{b}}}{\text{minimize}} \quad & \sum_{i=1}^{n_\mathrm{b}}\|\boldsymbol{w}_i\|^2 \\ \text{subject to} \quad & \frac{\left|\boldsymbol{v}^\dagger \boldsymbol{H}\boldsymbol{a}\right|^2}{\boldsymbol{v}^\dagger\left(\boldsymbol{H}\mathbf{A}\boldsymbol{R}_\mathrm{s}\mathbf{A}^\dagger\boldsymbol{H}^\dagger+\boldsymbol{R}_\mathrm{n}\right)\boldsymbol{v}}\ge\gamma, \\ & |\alpha_k|^2\left(\delta_\theta^2+\sigma_{\mathrm{u},k}^2\right)\le\sum_{i=1}^{n_\mathrm{b}}\left|\boldsymbol{w}_i^\dagger\boldsymbol{g}_k\right|^2, \quad \forall k\in[n_\mathrm{s}]. \end{aligned}$$

Note that closed-form solutions to the global optimization of these two problems are generally unknown. Indeed, both problems are non-convex due to the coupled amplification vector $\boldsymbol{a}$ and receive filtering $\boldsymbol{v}$. Therefore, we turn to a simple approach—alternative minimization—which guarantees convergence, at least to a local optimum.

III. Minimizing MSE under Power Constraints

In this section, we propose an alternative minimization algorithm to obtain the minimum solution to problem (P1). We also study the MSE performance for a large-scale antenna FC as well as a single-antenna FC.

A. Proposed Solution to Problem (P1)

Since problem (P1) is non-convex due to a non-concave objective function, we solve (P1) by using the alternative minimization method. Our goal is to progressively increase the objective function in (P1) by iteratively optimizing (P1) over $\boldsymbol{a}$ and $\{\boldsymbol{w}_i\}_{i=1}^{n_\mathrm{b}}$ for given $\boldsymbol{v}$, and then over $\boldsymbol{v}$ for given $\boldsymbol{a}$. In order to find $\boldsymbol{v}$, we first fix $\boldsymbol{a}$ and solve the following unconstrained optimization problem:

$$\underset{\boldsymbol{v}}{\text{maximize}} \quad \frac{\left|\boldsymbol{v}^\dagger \boldsymbol{H}\boldsymbol{a}\right|^2}{\boldsymbol{v}^\dagger\left(\boldsymbol{H}\mathbf{A}\boldsymbol{R}_\mathrm{s}\mathbf{A}^\dagger\boldsymbol{H}^\dagger+\boldsymbol{R}_\mathrm{n}\right)\boldsymbol{v}}, \tag{10}$$

which is a Rayleigh quotient and hence can be recasted as

$$\begin{aligned} \underset{\boldsymbol{v}}{\text{minimize}} \quad & \boldsymbol{v}^\dagger\left(\boldsymbol{H}\mathbf{A}\boldsymbol{R}_\mathrm{s}\mathbf{A}^\dagger\boldsymbol{H}^\dagger+\boldsymbol{R}_\mathrm{n}\right)\boldsymbol{v} \\ \text{subject to} \quad & \boldsymbol{v}^\dagger\boldsymbol{H}\boldsymbol{a}=1. \end{aligned} \tag{11}$$

Solving the above problem, we obtain

$$\boldsymbol{v}^+=\kappa\left(\boldsymbol{H}\mathbf{A}\boldsymbol{R}_\mathrm{s}\mathbf{A}^\dagger\boldsymbol{H}^\dagger+\boldsymbol{R}_\mathrm{n}\right)^{-1}\boldsymbol{H}\boldsymbol{a}. \tag{12}$$

Note that the value of κ is chosen to guarantee the equality constraint in (11). However, any selected value of κ will not affect the objective function in (P1), and thus we simply choose $\kappa = 1$ without loss of optimality. For a given $\boldsymbol{v}$ in (12), we are now ready to find an update of $\boldsymbol{a}$ and $\{\boldsymbol{w}_i\}_{i=1}^{n_\mathrm{b}}$ in (P1). To facilitate the calculations, we define $\boldsymbol{f} = \left[\boldsymbol{v}^\dagger\boldsymbol{h}_1\,\boldsymbol{v}^\dagger\boldsymbol{h}_2\,\ldots\,\boldsymbol{v}^\dagger\boldsymbol{h}_{n_\mathrm{s}}\right]^\top$ and $\boldsymbol{F} = \mathrm{diag}\,(\boldsymbol{f})$, where $\boldsymbol{h}_i$ is the ith column of the matrix $\boldsymbol{H}$. Then, for a fixed receive filtering $\boldsymbol{v}$, problem (P1) can be expressed as

$$\begin{aligned} \underset{\boldsymbol{a},\{\boldsymbol{w}_i\}_{i=1}^{n_\mathrm{b}}}{\text{maximize}} \quad & \frac{\left|\boldsymbol{a}^\top\boldsymbol{f}\right|^2}{\boldsymbol{a}^\top\boldsymbol{F}\boldsymbol{R}_\mathrm{s}\boldsymbol{F}^\dagger\boldsymbol{a}^*+\boldsymbol{v}^\dagger\boldsymbol{R}_\mathrm{n}\boldsymbol{v}} \\ \text{subject to} \quad & |\alpha_k|^2\left(\delta_\theta^2+\sigma_{\mathrm{u},k}^2\right)\le\sum_{i=1}^{n_\mathrm{b}}\left|\boldsymbol{w}_i^\dagger\boldsymbol{g}_k\right|^2, \forall k\in[n_\mathrm{s}] \\ & \sum_{i=1}^{n_\mathrm{b}}\|\boldsymbol{w}_i\|^2\le P. \end{aligned} \tag{13}$$

We remark that even with a fixed receive filtering $\boldsymbol{v}$, problem (P1) is still non-convex, and thus needs to be transformed to a simple form. We further introduce $\boldsymbol{Q} = \boldsymbol{a}^*\boldsymbol{a}^\top$, $\boldsymbol{W} = \sum_{i=1}^{n_\mathrm{b}}\boldsymbol{w}_i\boldsymbol{w}_i^\dagger$, $\boldsymbol{\Sigma} = \boldsymbol{f}\boldsymbol{f}^\dagger$, $\boldsymbol{\Psi} = \boldsymbol{F}\boldsymbol{R}_\mathrm{s}\boldsymbol{F}^\dagger$, $\boldsymbol{G}_k = \boldsymbol{g}_k\boldsymbol{g}_k^\dagger$, and $\mathbf{D}_k = \mathrm{diag}(0,\ldots,\delta_\theta^2+\sigma_{\mathrm{u},k}^2,\ldots,0)$. Then, we can rewrite the optimization problem (13) as

$$\begin{aligned} \underset{\boldsymbol{Q},\boldsymbol{W}}{\text{maximize}} \quad & \frac{\mathrm{tr}\,(\boldsymbol{Q}\boldsymbol{\Sigma})}{\mathrm{tr}\,(\boldsymbol{Q}\boldsymbol{\Psi})+\boldsymbol{v}^\dagger\boldsymbol{R}_\mathrm{n}\boldsymbol{v}} \\ \text{subject to} \quad & \mathrm{tr}\,(\mathbf{D}_k\boldsymbol{Q})-\mathrm{tr}\,(\boldsymbol{G}_k\boldsymbol{W})\le 0, \quad \forall k\in[n_\mathrm{s}] \\ & \mathrm{tr}\,(\boldsymbol{W})\le P \\ & \boldsymbol{W}\succeq\mathbf{0},\boldsymbol{Q}\succeq\mathbf{0} \\ & \mathsf{rank}\,(\boldsymbol{Q})=1. \end{aligned} \tag{14}$$

In (14), if there exist a rank one solution of the optimal $\boldsymbol{Q} = \boldsymbol{Q}^\star$ and a rank n_b solution of the optimal $\boldsymbol{W} = \boldsymbol{W}^\star$, then one can recover the optimal $\boldsymbol{a}^\star$ and $\{\boldsymbol{w}_i^\star\}_{i=1}^{n_\mathrm{b}}$ by taking the eigenvalue decomposition of the matrices $\boldsymbol{Q}^\star$ and $\boldsymbol{W}^\star$, respectively. Note that problem (14) is non-convex due to the linear fractional structure of its objective function. However, we can use the Charnes-Cooper transformation [36] to reformulate the quasi-convex objective function in (14) into a simpler form as follows:[5]

$$\begin{aligned} \underset{\bar{\boldsymbol{Q}},\bar{\boldsymbol{W}},\eta}{\text{maximize}} \quad & \mathrm{tr}\left(\bar{\boldsymbol{Q}}\boldsymbol{\Sigma}\right) \\ \text{subject to} \quad & \mathrm{tr}\left(\bar{\boldsymbol{Q}}\boldsymbol{\Psi}\right)+\eta\boldsymbol{v}^\dagger\boldsymbol{R}_\mathrm{n}\boldsymbol{v}=1 \\ & \mathrm{tr}\left(\mathbf{D}_k\bar{\boldsymbol{Q}}\right)-\mathrm{tr}\left(\boldsymbol{G}_k\bar{\boldsymbol{W}}\right)\le 0, \quad \forall k\in[n_\mathrm{s}] \\ & \mathrm{tr}\left(\bar{\boldsymbol{W}}\right)\le\eta P \\ & \bar{\boldsymbol{W}}\succeq\mathbf{0},\bar{\boldsymbol{Q}}\succeq\mathbf{0},\eta>0 \\ & \mathsf{rank}\left(\bar{\boldsymbol{Q}}\right)=1. \end{aligned} \tag{15}$$

Note that $\eta = 0$ is not feasible because from the third constraint, we must have $\bar{\boldsymbol{W}} = \mathbf{0}$ if $\eta = 0$. Thus, from the

[5]Here, we use the transformations $\eta^{-1} = \mathrm{tr}\left((\boldsymbol{Q}\boldsymbol{\Psi})+\boldsymbol{v}^\dagger\boldsymbol{R}_\mathrm{n}\boldsymbol{v}\right)$, $\bar{\boldsymbol{Q}} = \eta\boldsymbol{Q}$, and $\bar{\boldsymbol{W}} = \eta\boldsymbol{W}$.

second constraint for any k, it follows that $\bar{\boldsymbol{Q}} = \mathbf{0}$, which however violates the first constraint in (15).

Remark 1 (The Equivalence of Problems (14) *and* (15)*):* If $(\bar{\boldsymbol{Q}}^\star, \bar{\boldsymbol{W}}^\star, \eta^\star)$ is the optimal solution to problem (15), then $(\bar{\boldsymbol{Q}}^\star/\eta^\star, \bar{\boldsymbol{W}}^\star/\eta^\star)$ is feasible to problem (14) and achieves the same objective value as that of problem (15). On the other hand, let $t^\star = \mathrm{tr}\left(\boldsymbol{Q}^\star \boldsymbol{\Psi}\right) + \boldsymbol{v}^\dagger \boldsymbol{R}_{\mathrm{n}} \boldsymbol{v}$. Then, if $(\boldsymbol{Q}^\star, \boldsymbol{W}^\star)$ is the optimal solution to problem (14), then $(\boldsymbol{Q}^\star/t^\star, \boldsymbol{W}^\star/t^\star, 1/t^\star)$ is feasible to problem (15) and achieves the same objective value as that of problem (14). This implies that the Charnes-Cooper transform is a one-to-one mapping between the feasible sets of problems (14) and (15). We can thus obtain the optimal solution to problem (14) by solving problem (15), which has a simpler form in the sense that the non-convexity of the objective function in problem (14) is eliminated.

Note that problem (15) is still non-convex due to the rank constraint, which makes problem (15) intractable in general. Hence, we will solve a relaxed version of (15) by ignoring the rank constraint on $\boldsymbol{Q}$, which leads to the semidefinite relaxation (SDR) of problem (15).

$$
\begin{aligned}
&(\mathsf{SDR1}): \\
&\underset{\bar{\boldsymbol{Q}}, \bar{\boldsymbol{W}}, \eta}{\text{maximize}} \quad \mathrm{tr}\left(\bar{\boldsymbol{Q}} \boldsymbol{\Sigma}\right) \\
&\text{subject to} \quad \mathrm{tr}\left(\bar{\boldsymbol{Q}} \boldsymbol{\Psi}\right) + \eta \boldsymbol{v}^\dagger \boldsymbol{R}_{\mathrm{n}} \boldsymbol{v} = 1 \\
&\qquad\qquad\quad \mathrm{tr}\left(\mathbf{D}_k \bar{\boldsymbol{Q}}\right) - \mathrm{tr}\left(\boldsymbol{G}_k \bar{\boldsymbol{W}}\right) \le 0, \ \ \forall k \in [n_{\mathrm{s}}] \\
&\qquad\qquad\quad \mathrm{tr}\left(\bar{\boldsymbol{W}}\right) \le \eta P \\
&\qquad\qquad\quad \bar{\boldsymbol{W}} \succeq \mathbf{0}, \bar{\boldsymbol{Q}} \succeq \mathbf{0}, \eta > 0.
\end{aligned}
$$

The relaxed problem (SDR1) is now convex—indeed semidefinite program (SDP)—whose optimal solution can be found, for example, by using the interior-point method (e.g., CVX [37]). The following theorem characterizes the properties of the optimal solution to problem (SDR1).

Theorem 1 (Properties of Optimal Solution): Let $\nu^\star$ and $\beta^\star$ be the optimal dual solutions associated with the first and third constraint in (SDR1), respectively. We also let $\bar{\boldsymbol{Q}}^\star$ and $\bar{\boldsymbol{W}}^\star$ be the optimal primal solutions to problem (SDR1). Then, the following three properties are fulfilled:

1) $\nu^\star > 0$, $\beta^\star > 0$;
2) $\mathsf{rank}\left(\bar{\boldsymbol{W}}^\star\right) \le \min\left(n_{\mathrm{s}}, n_{\mathrm{r}}\right)$;
3) $\mathsf{rank}\left(\bar{\boldsymbol{Q}}^\star\right) = 1$.

Proof: See Appendix A. □

Remark 2: The condition $\beta^\star > 0$ implies that the total power constraint at the FC must be satisfied with equality, while property 2) implies that at most $n_{\mathrm{b}} = \min\left(n_{\mathrm{s}}, n_{\mathrm{r}}\right)$ energy beams are required for the optimal solution of problem (SDR1). It is worth noting that for fixed $\boldsymbol{v}$, at the optimal solution $(\bar{\boldsymbol{Q}}^\star, \bar{\boldsymbol{W}}^\star, \eta^\star)$, the individual power constraints in (SDR1) are not necessarily all tight, i.e., there may exist some k such that $\mathrm{tr}\left(\mathbf{D}_k \bar{\boldsymbol{Q}}\right) - \mathrm{tr}\left(\boldsymbol{G}_k \bar{\boldsymbol{W}}\right) < 0$. This fact reveals that the sensors do not always transmit all the power budget harvested from the energy harvesting phase, but power control is required to guarantee the MSE optimality. A similar observation was made in throughput optimization for multiple-antenna multiuser cellular systems in [30].

Remark 3 (The Equivalence of Problems (15) *and* (SDR1)*):* We remark that since problem (SDR1) is a relaxed version of problem (15), in general, the solution to problem (SDR1) provides an upper bound on the optimal solution to problem (15), or equivalently, an upper bound on problem (P1) for a given $\boldsymbol{v}$. Fortunately, we can show that the optimal solution to (SDR1) is also optimal to (15). To do that, let $\Phi_\eta\left(\bar{\boldsymbol{Q}}, \bar{\boldsymbol{W}}\right)$ be the objective function of problem (15) or (SDR1) for a given feasible η, and $(\bar{\boldsymbol{Q}}^\star, \bar{\boldsymbol{W}}^\star)$ and $(\bar{\boldsymbol{Q}}_\star, \bar{\boldsymbol{W}}_\star)$ be the optimal solutions to problems (SDR1) and (15), respectively. Since the optimization problem (SDR1) is a relaxation of problem (15), we must have

$$\Phi_\eta\left(\bar{\boldsymbol{Q}}^\star, \bar{\boldsymbol{W}}^\star\right) \ge \Phi_\eta\left(\bar{\boldsymbol{Q}}_\star, \bar{\boldsymbol{W}}_\star\right). \tag{16}$$

On the other hand, since $\mathsf{rank}\left(\bar{\boldsymbol{Q}}^\star\right) = 1$, the solution $(\bar{\boldsymbol{Q}}^\star, \bar{\boldsymbol{W}}^\star, \eta)$ is also a feasible solution to problem (15). Therefore, we have

$$\Phi_\eta\left(\bar{\boldsymbol{Q}}^\star, \bar{\boldsymbol{W}}^\star\right) \le \Phi_\eta\left(\bar{\boldsymbol{Q}}_\star, \bar{\boldsymbol{W}}_\star\right). \tag{17}$$

From (16) and (17), it follows that $\Phi_\eta\left(\bar{\boldsymbol{Q}}^\star, \bar{\boldsymbol{W}}^\star\right) = \Phi_\eta\left(\bar{\boldsymbol{Q}}_\star, \bar{\boldsymbol{W}}_\star\right)$. In other words, $(\bar{\boldsymbol{Q}}^\star, \bar{\boldsymbol{W}}^\star)$ is also optimal solution to problem (15). Note that the above equivalence holds for any feasible η, and hence it holds for the optimal η.

Remark 3 suggests that we can solve the original problem (P1) for a given $\boldsymbol{v}$ by equivalently solving the relaxed problem (SDR1) without loss of optimality. Finally, we summarize the overall procedure for solving problem (P1) in Algorithm 1 below. In this algorithm, the FC iteratively updates $\boldsymbol{v}$, $\boldsymbol{a}$ and $\{\boldsymbol{w}_i\}_{i=1}^{n_{\mathrm{b}}}$ in Step 3 and 4, respectively. The convergence and complexity of Algorithm 1 are analyzed in the following remark.

Remark 4 (Convergence and Complexity): Note that the objective function in (P1) is increased in each step of Algorithm 1. Moreover, the objective function is upper-bounded by a certain value due to the finite total power at the FC, which implies that the algorithm must converge. However, the algorithm may converge to a local optimum due to the non-convex nature of the optimization problem. We now provide the complexity analysis of the proposed algorithm. Specifically, in each iteration of Algorithm 1, the worst-case computational complexity for solving the generic convex problem in (SDR1), corresponding to Step 4 in Algorithm 1, using the interior point method is given by $O\left(\left(2n_{\mathrm{s}} + n_{\mathrm{r}}\right)^{1/2}\left(n_{\mathrm{s}}^4 + n_{\mathrm{r}}^3 n_{\mathrm{s}} + n_{\mathrm{r}}^2 n_{\mathrm{s}}^2\right)\log\left(\frac{1}{\xi}\right)\right)$ for an ξ-optimal solution [38, Chapter 6.6.3]. For the receive filtering update, based on the elementary vector matrix calculation [39], one can show that the computational complexity of Step 3 in Algorithm 1 is $O\left(n_{\mathrm{r}}^3 + n_{\mathrm{r}}^2 n_{\mathrm{s}}\right)$. In Step 6, the amplification vector is constructed by using eigenvalue decomposition of rank one matrix $\bar{\boldsymbol{Q}}$, and hence the complexity is $O\left(n_{\mathrm{s}}^2\right)$. Thus, the overall complexity per iteration of Algorithm 1 is at most $O\left(\left(2n_{\mathrm{s}} + n_{\mathrm{r}}\right)^{1/2}\left(n_{\mathrm{s}}^4 + n_{\mathrm{r}}^3 n_{\mathrm{s}} + n_{\mathrm{r}}^2 n_{\mathrm{s}}^2\right)\log\left(\frac{1}{\xi}\right) + n_{\mathrm{r}}^3 + n_{\mathrm{r}}^2 n_{\mathrm{s}}\right)$. We remark that although the complexity of the alternative minimization algorithms are typically unknown [40], [41], it is observed via simulations that they converge within 10 to 20 iterations in general.

B. Large-Scale Antenna FC

The following proposition shows the property of the asymptotic MSE of the BLUE.

Proposition 1 (Asymptotic MSE): Consider the distributed estimation system in Section II, where the channel matrix $\boldsymbol{H}$ is a random matrix with independent and identical elements, each of which has zero mean and unit variance. As the number of antennas at the FC tends to infinity, the MSE defined in (9) converges to that of centralized estimation systems.[6] That is, as $n_{\rm r} \to \infty$, we have

$$\mathsf{mse} \overset{\rm a.s}{\to} \left[\mathbf{1}^\top \boldsymbol{R}_{\rm s}^{-1}\mathbf{1}\right]^{-1}, \tag{18}$$

where $\overset{\rm a.s}{\to}$ denotes the almost sure convergence.

Proof: Given the receive filtering in (12), the MSE can be written as

$$\mathsf{mse} = \left[\boldsymbol{a}^\dagger \boldsymbol{H}^\dagger \left(\boldsymbol{H}\mathbf{A}\boldsymbol{R}_{\rm s}\mathbf{A}^\dagger\boldsymbol{H}^\dagger + \boldsymbol{R}_{\rm n}\right)^{-1}\boldsymbol{H}\boldsymbol{a}\right]^{-1}. \tag{19}$$

Therefore, it suffices to prove that

$$\boldsymbol{a}^\dagger \boldsymbol{H}^\dagger \left(\boldsymbol{H}\mathbf{A}\boldsymbol{R}_{\rm s}\mathbf{A}^\dagger\boldsymbol{H}^\dagger + \boldsymbol{R}_{\rm n}\right)^{-1}\boldsymbol{H}\boldsymbol{a} \overset{\rm a.s}{\to} \mathbf{1}^\top \boldsymbol{R}_{\rm s}^{-1}\mathbf{1} \tag{20}$$

as $n_{\rm r}$ tends to infinity. Using the matrix inversion lemma, we can show that

$$\begin{aligned}&\left(\boldsymbol{H}\mathbf{A}\boldsymbol{R}_{\rm s}\mathbf{A}^\dagger\boldsymbol{H}^\dagger + \boldsymbol{R}_{\rm n}\right)^{-1}\\ &= \boldsymbol{R}_{\rm n}^{-1} - \boldsymbol{R}_{\rm n}^{-1}\boldsymbol{H}\left(\boldsymbol{K}^{-1} + \boldsymbol{H}^\dagger\boldsymbol{R}_{\rm n}^{-1}\boldsymbol{H}\right)^{-1}\boldsymbol{H}^\dagger\boldsymbol{R}_{\rm n}^{-1},\end{aligned} \tag{21}$$

where $\boldsymbol{K} = \mathbf{A}\boldsymbol{R}_{\rm s}\mathbf{A}^\dagger$. Substituting (21) into (19), we obtain

$$\begin{aligned}\mathsf{mse}^{-1} &= \boldsymbol{a}^\dagger\boldsymbol{H}^\dagger\boldsymbol{R}_{\rm n}^{-1}\boldsymbol{H}\boldsymbol{a}\\ &- \boldsymbol{a}^\dagger\boldsymbol{H}^\dagger\boldsymbol{R}_{\rm n}^{-1}\boldsymbol{H}\left(\boldsymbol{K}^{-1} + \boldsymbol{H}^\dagger\boldsymbol{R}_{\rm n}^{-1}\boldsymbol{H}\right)^{-1}\boldsymbol{H}^\dagger\boldsymbol{R}_{\rm n}^{-1}\boldsymbol{H}\boldsymbol{a}.\end{aligned} \tag{22}$$

Note that as $n_{\rm r} \to \infty$, we have [42]

$$\frac{1}{n_{\rm r}}\boldsymbol{H}^\dagger\boldsymbol{R}_{\rm n}^{-1}\boldsymbol{H} \overset{\rm a.s}{\to} \boldsymbol{R}_{\rm n}^{-1}. \tag{23}$$

Using this identity, we obtain

$$\begin{aligned}\frac{\mathsf{mse}^{-1}}{n_{\rm r}} &\overset{\rm a.s}{\to} \boldsymbol{a}^\dagger\boldsymbol{R}_{\rm n}^{-1}\boldsymbol{a} - \boldsymbol{a}^\dagger\boldsymbol{R}_{\rm n}^{-1}\left(\frac{\boldsymbol{K}^{-1}}{n_{\rm r}} + \boldsymbol{R}_{\rm n}^{-1}\right)^{-1}\boldsymbol{R}_{\rm n}^{-1}\boldsymbol{a}\\ &= \boldsymbol{a}^\dagger\boldsymbol{R}_{\rm n}^{-1}\left(\mathbf{I} - \left(\frac{\boldsymbol{R}_{\rm n}\boldsymbol{K}^{-1}}{n_{\rm r}} + \mathbf{I}\right)^{-1}\right)\boldsymbol{a}\\ &= \frac{1}{n_{\rm r}}\boldsymbol{a}^\dagger\left(\boldsymbol{K} + \frac{\boldsymbol{R}_{\rm n}}{n_{\rm r}}\right)^{-1}\boldsymbol{a}.\end{aligned} \tag{24}$$

where the second equality follows from the matrix inversion lemma. From (24) and the definitions of the matrices $\boldsymbol{K}$ and $\mathbf{A}$, as $n_{\rm r} \to \infty$, we finally have

$$\mathsf{mse} \overset{\rm a.s}{\to} \left[\mathbf{1}^\top \boldsymbol{R}_{\rm s}^{-1}\mathbf{1}\right]^{-1}, \tag{25}$$

which concludes the proof of the proposition. □

Proposition 1 implies that as the number of antennas grows large, the effects of fading and noise at the FC disappear, and hence the performance benchmark is determined by the sensing quality. From (18), if the sensing noise at the sensors is equal to $\boldsymbol{R}_{\rm s} = \sigma_{\rm n}^2\mathbf{I}_{n_{\rm s}}$, then it follows that $\left[\mathbf{1}^\top \boldsymbol{R}_{\rm s}^{-1}\mathbf{1}\right]^{-1} = \frac{\sigma_{\rm n}^2}{n_{\rm s}}$. This means that the MSE linearly decreases according to $n_{\rm s}$.

[6]We use the term centralized estimation to refer to the case for which the sensors' data are perfectly available at the FC, which serves as a performance benchmark.

Algorithm 1 proposed algorithm to solve (P1)

1: **Initialization**: set $n := 0$, and generate $\boldsymbol{a}^{(0)}$ and $\mathbf{A}^{(0)}$.
2: **repeat**
3: $\quad \boldsymbol{v}^{(n)} = \left(\boldsymbol{H}\mathbf{A}^{(n)}\boldsymbol{R}_{\rm s}\mathbf{A}^{(n)\dagger}\boldsymbol{H}^\dagger + \boldsymbol{R}_{\rm n}\right)^{-1}\boldsymbol{H}\boldsymbol{a}^{(n)}$
4: $\quad$ Solve problem (SDR1) with $\boldsymbol{v} = \boldsymbol{v}^{(n)}$ to obtain the optimal solution $(\bar{\boldsymbol{Q}}^\star, \bar{\boldsymbol{W}}^\star, \eta^\star)$.
5: $\quad$ Set $(\bar{\boldsymbol{Q}}^{(n+1)}, \bar{\boldsymbol{W}}^{(n+1)}, \eta^{(n+1)}) := (\bar{\boldsymbol{Q}}^\star, \bar{\boldsymbol{W}}^\star, \eta^\star)$.
6: $\quad$ Construct $\{\boldsymbol{a}^{(n+1)}, \mathbf{A}^{(n+1)}\}$ from $\bar{\boldsymbol{Q}}^{(n+1)}/\eta^{(n+1)}$.
7: $\quad$ Update $n := n+1$.
8: **until convergence**
9: **Output:** $(\boldsymbol{Q} = \bar{\boldsymbol{Q}}^{(n)}/\eta^{(n)}, \boldsymbol{W} = \bar{\boldsymbol{W}}^{(n)}/\eta^{(n)}, \boldsymbol{v} = \boldsymbol{v}^{(n)})$

C. Single-Antenna FC

It is of importance to study the single-antenna FC scenario separately not only for comparison but also because the problem is remarkably simplified. Specifically, for a single-antenna FC, the design of energy beamforming and receive filtering is neglected, and thus we aim to simply find the optimal amplification coefficients $\boldsymbol{a}$ that minimize the MSE of the BLUE. During the energy transmission phase, we assume that the FC transmits an energy signal s such that $\mathbb{E}\{|s|^2\} = P$. In this case, the harvested energy at the kth sensor is given by

$$E_k = \frac{P\,|g_k|^2}{2}. \tag{26}$$

The MSE of the BLUE is boiled down to

$$\mathsf{mse} = \left[\frac{\boldsymbol{a}^\top\boldsymbol{h}\boldsymbol{h}^\dagger\boldsymbol{a}^*}{\boldsymbol{a}^\top\boldsymbol{F}\boldsymbol{R}_{\rm s}\boldsymbol{F}^\dagger\boldsymbol{a}^* + \sigma_{\rm n}^2}\right]^{-1}, \tag{27}$$

where $\boldsymbol{h}^\dagger \in \mathbb{C}^{1\times n_{\rm s}}$ is the channel between the sensors and the FC; $\boldsymbol{F} = {\rm diag}(\boldsymbol{h})$; and $n \in \mathbb{C}$ is the additive noise at the FC. Given the MSE in (27), we aim to solve the following problem

$$\begin{aligned}\underset{\boldsymbol{a}}{\text{maximize}} \quad & \frac{\boldsymbol{a}^\top\boldsymbol{h}\boldsymbol{h}^\dagger\boldsymbol{a}^*}{\boldsymbol{a}^\top\boldsymbol{F}\boldsymbol{R}_{\rm s}\boldsymbol{F}^\dagger\boldsymbol{a}^* + \sigma_{\rm n}^2}\\ \text{subject to} \quad & |\alpha_k|^2\left(\delta_\theta^2 + \sigma_{{\rm u},k}^2\right) \le P\,|g_k|^2, \quad \forall k \in [n_{\rm s}].\end{aligned} \tag{28}$$

The above problem—quadratically constrained ratio of quadratic functions (QCRQ)—has been studied for parameter tracking using the Kalman filter at the FC [6], where the optimal solution is given by

$$\boldsymbol{a}^\star = \frac{1}{\sqrt{(\boldsymbol{P}^\star)_{n_{\rm s}+1,n_{\rm s}+1}}}\bar{\boldsymbol{a}}^*. \tag{29}$$

Here, $(\boldsymbol{P}^\star)_{i,j}$ is the (i,j)th element of the matrix $\boldsymbol{P}^\star$; $\bar{\boldsymbol{a}}$ is the vector satisfying $\bar{\boldsymbol{a}}^*\bar{\boldsymbol{a}}^\top = \boldsymbol{P}^\star_{n_\mathrm{s}}$; $\boldsymbol{P}^\star_{n_\mathrm{s}}$ is the n_sth order leading principal submatrix of $\boldsymbol{P}^\star$ obtained by excluding the $(n_\mathrm{s}+1)$th row and column; and $\boldsymbol{P}^\star \in \mathbb{C}^{(n_\mathrm{s}+1)\times(n_\mathrm{s}+1)}$ is the optimal solution to the following problem

$$\begin{aligned} \underset{\boldsymbol{P}\succeq 0}{\text{maximize}} \quad & \mathrm{tr}\left(\boldsymbol{P}\,\boldsymbol{\Xi}\right) \\ \text{subject to} \quad & \mathrm{tr}\left(\boldsymbol{P}\,\boldsymbol{C}\right) = 1 \\ & \mathrm{tr}\left(\boldsymbol{P}\,\bar{\mathbf{D}}_k\right) \le P\left|g_k\right|^2, \quad \forall k \in [n_\mathrm{s}], \end{aligned} \tag{30}$$

where $\boldsymbol{P} = \left[t\,\boldsymbol{a}^\top\ t\right]^\dagger\left[t\,\boldsymbol{a}^\top\ t\right]$; $\boldsymbol{\Xi} = \begin{pmatrix} \boldsymbol{h}\boldsymbol{h}^\dagger & \mathbf{0} \\ \mathbf{0} & 0 \end{pmatrix}$; $\boldsymbol{C} = \begin{pmatrix} \boldsymbol{F}\boldsymbol{R}_\mathrm{s}\boldsymbol{F}^\dagger & \mathbf{0} \\ \mathbf{0} & t\,\sigma_\mathrm{n}^2 \end{pmatrix}$; $\bar{\mathbf{D}}_k = \begin{pmatrix} \mathbf{D}_k & \mathbf{0} \\ \mathbf{0} & -P\left|g_k\right|^2 \end{pmatrix}$; and t is an auxiliary variable.

Remark 5: Note that the solution in (29) is indeed global optimum. This is different from the multiple-antenna FC case in which we may only achieve a local optimum.

D. A Common Energy Harvester

We now consider a special deployment case in WPT-enabled sensor networks, where a common energy harvester is used to collect energy from the FC.[7] Assume that the common energy harvester is equipped with a single antenna, then the optimization problem can be stated as

$$\begin{aligned} (\mathsf{P1{-}Sum}): \quad \underset{\boldsymbol{v},\boldsymbol{a},\boldsymbol{w}}{\text{maximize}} \quad & \frac{\left|\boldsymbol{v}^\dagger\boldsymbol{H}\boldsymbol{a}\right|^2}{\boldsymbol{v}^\dagger\left(\boldsymbol{H}\mathbf{A}\boldsymbol{R}_\mathrm{s}\mathbf{A}^\dagger\boldsymbol{H}^\dagger + \boldsymbol{R}_\mathrm{n}\right)\boldsymbol{v}} \\ \text{subject to} \quad & \boldsymbol{a}^\dagger\mathbf{D}\boldsymbol{a} \le \left|\boldsymbol{w}^\dagger\boldsymbol{h}_\mathrm{e}\right|^2 \\ & \|\boldsymbol{w}\|^2 \le P, \end{aligned}$$

where $\boldsymbol{h}_\mathrm{e}$ is the channel between the FC and the common harvester and $\mathbf{D} = \mathrm{diag}\left(\delta_\theta^2 + \sigma_{\mathrm{u},1}^2, \cdots, \delta_\theta^2 + \sigma_{\mathrm{u},n_\mathrm{s}}^2\right)$. Since the objective function in problem (P1–Sum) is monotonically increasing with the norm of $\boldsymbol{a}$, the sum power constraint (i.e., the first constraint) should be satisfied with equality and the right-hand side of this constraint should be as large as possible to be the optimal solution. This implies that the optimal energy beamforming is $\boldsymbol{w}^\star = \sqrt{P}\frac{\boldsymbol{h}_\mathrm{e}}{\|\boldsymbol{h}_\mathrm{e}\|}$. Similarly as in problem (P1), we iteratively solve the above problem for $\boldsymbol{v}$ and $\boldsymbol{a}$, where $\boldsymbol{v}$ is given in (12). Let $E = \frac{1}{2}|\boldsymbol{h}_\mathrm{e}^\dagger\boldsymbol{w}^\star|^2 = \frac{1}{2}P\left\|\boldsymbol{h}_\mathrm{e}\right\|^2$ be the harvested energy at the harvester. For a given $\boldsymbol{v}$, the optimal $\boldsymbol{a}$ is then the solution of the following problem:

$$\begin{aligned} \underset{\boldsymbol{a}}{\text{maximize}} \quad & \frac{\left|\boldsymbol{a}^\top\boldsymbol{f}\right|^2}{\boldsymbol{a}^\top\boldsymbol{F}\boldsymbol{R}_\mathrm{s}\boldsymbol{F}^\dagger\boldsymbol{a}^* + \boldsymbol{v}^\dagger\boldsymbol{R}_\mathrm{n}\boldsymbol{v}} \\ \text{subject to} \quad & \boldsymbol{a}^\dagger\mathbf{D}\boldsymbol{a} = P\left\|\boldsymbol{h}_\mathrm{e}\right\|^2, \end{aligned} \tag{31}$$

where $\boldsymbol{f} = \left[\boldsymbol{v}^\dagger\boldsymbol{h}_1\,\boldsymbol{v}^\dagger\boldsymbol{h}_2\,\cdots\,\boldsymbol{v}^\dagger\boldsymbol{h}_{n_\mathrm{s}}\right]^\top$ and $\boldsymbol{F} = \mathrm{diag}\left(\boldsymbol{f}\right)$. The problem is equivalent to

$$\underset{\boldsymbol{a}}{\text{maximize}} \quad \frac{\boldsymbol{a}^\top\boldsymbol{X}\boldsymbol{a}^*}{\boldsymbol{a}^\top\boldsymbol{Y}\boldsymbol{a}^*}, \tag{32}$$

[7]This differs from what we have considered so far, where each sensor has its own energy harvester. This approach reduces the hardware complexity of sensors. However, it is feasible only if the sensors are closely, or even colocated.

where $\boldsymbol{X} = \boldsymbol{f}\boldsymbol{f}^\dagger$ and $\boldsymbol{Y} = \boldsymbol{F}\boldsymbol{R}_\mathrm{s}\boldsymbol{F}^\dagger + \frac{\boldsymbol{v}^\dagger\boldsymbol{R}_\mathrm{n}\boldsymbol{v}}{P\|\boldsymbol{h}_\mathrm{e}\|^2}\mathbf{D}$. Note that $\boldsymbol{X} \succeq \mathbf{0}$ and $\boldsymbol{Y} \succ \mathbf{0}$, and problem (32) is indeed Rayleigh quotient, thus the optimal solution can be expressed as

$$\boldsymbol{a}^\star = \sqrt{\frac{P\left\|\boldsymbol{h}_\mathrm{e}\right\|^2}{\boldsymbol{f}^\dagger\boldsymbol{Y}^{-1}\mathbf{D}\boldsymbol{Y}^{-1}\boldsymbol{f}}}\boldsymbol{Y}^{-1}\boldsymbol{f}^*. \tag{33}$$

Then, the optimal value in (31) is given by

$$\max_{\boldsymbol{a}} \frac{\left|\boldsymbol{a}^\top\boldsymbol{f}\right|^2}{\boldsymbol{a}^\top\boldsymbol{F}\boldsymbol{R}_\mathrm{s}\boldsymbol{F}^\dagger\boldsymbol{a}^* + \boldsymbol{v}^\dagger\boldsymbol{R}_\mathrm{n}\boldsymbol{v}} = \lambda_{\max}\left(\boldsymbol{Y}^{-1}\boldsymbol{X}\right) = \boldsymbol{f}^\dagger\boldsymbol{Y}^{-1}\boldsymbol{f}, \tag{34}$$

where $\lambda_{\max}\left(\cdot\right)$ denotes the maximum eigenvalue of a matrix. It can be seen that the sum power constraint enables to significantly reduce the complexity of the optimization problem.

IV. Minimizing Power under an MSE Constraint

In this section, we study the power minimization for distributed estimation with an MSE constraint.

A. Proposed Solution to Problem (P2)

Similarly as in problem (P1), we adopt an alternative minimization method to iteratively solve problem (P2). Specifically, we first solve problem (P2) over $\boldsymbol{v}$ for given $\boldsymbol{a}$ by finding a solution to the following feasibility problem:

$$\begin{aligned} \underset{\boldsymbol{v}}{\text{minimize}} \quad & 0 \\ \text{subject to} \quad & \frac{\left|\boldsymbol{v}^\dagger\boldsymbol{H}\boldsymbol{a}\right|^2}{\boldsymbol{v}^\dagger\left(\boldsymbol{H}\mathbf{A}\boldsymbol{R}_\mathrm{s}\mathbf{A}^\dagger\boldsymbol{H}^\dagger + \boldsymbol{R}_\mathrm{n}\right)\boldsymbol{v}} \ge \gamma. \end{aligned} \tag{35}$$

Since the left-hand side (LHS) of the constraint in (35) is increasing with the norm of $\boldsymbol{a}$, one should choose $\boldsymbol{v}$ such that the LHS term is as large as possible. Hence, problem (35) can be rewritten as an unconstrained optimization problem as follows:

$$\underset{\boldsymbol{v}}{\text{maximize}} \quad \frac{\left|\boldsymbol{v}^\dagger\boldsymbol{H}\boldsymbol{a}\right|^2}{\boldsymbol{v}^\dagger\left(\boldsymbol{H}\mathbf{A}\boldsymbol{R}_\mathrm{s}\mathbf{A}^\dagger\boldsymbol{H}^\dagger + \boldsymbol{R}_\mathrm{n}\right)\boldsymbol{v}}. \tag{36}$$

Solving the above problem, we obtain

$$\boldsymbol{v}^+ = \left(\boldsymbol{H}\mathbf{A}\boldsymbol{R}_\mathrm{s}\mathbf{A}^\dagger\boldsymbol{H}^\dagger + \boldsymbol{R}_\mathrm{n}\right)^{-1}\boldsymbol{H}\boldsymbol{a}. \tag{37}$$

For fixed $\boldsymbol{v}$ given in (37), we now solve problem (P2) over $\boldsymbol{a}$ and $\{\boldsymbol{w}_i\}_{i=1}^{n_\mathrm{b}}$ as in the following:

$$\begin{aligned} \underset{\boldsymbol{a},\{\boldsymbol{w}_i\}_{i=1}^{n_\mathrm{b}}}{\text{minimize}} \quad & \sum_{i=1}^{n_\mathrm{b}}\|\boldsymbol{w}_i\|^2 \\ \text{subject to} \quad & \frac{\left|\boldsymbol{a}^\top\boldsymbol{f}\right|^2}{\boldsymbol{a}^\top\boldsymbol{F}\boldsymbol{R}_\mathrm{s}\boldsymbol{F}^\dagger\boldsymbol{a}^* + \boldsymbol{v}^\dagger\boldsymbol{R}_\mathrm{n}\boldsymbol{v}} \ge \gamma, \\ & \left|\alpha_k\right|^2\left(\delta_\theta^2 + \sigma_{\mathrm{u},k}^2\right) \le \sum_{i=1}^{n_\mathrm{b}}\left|\boldsymbol{w}_i^\dagger\boldsymbol{g}_k\right|^2, \ \forall k \in [n_\mathrm{s}], \end{aligned} \tag{38}$$

where $\boldsymbol{f} = \left[\boldsymbol{v}^\dagger\boldsymbol{h}_1\,\boldsymbol{v}^\dagger\boldsymbol{h}_2\,\cdots\,\boldsymbol{v}^\dagger\boldsymbol{h}_{n_\mathrm{s}}\right]^\top$; $\boldsymbol{h}_i$ is the ith column of the matrix $\boldsymbol{H}$; and $\boldsymbol{F} = \mathrm{diag}\left(\boldsymbol{f}\right)$. We remark that for a fixed $\boldsymbol{v}$, the MSE constraint (i.e., the first constraint) at the optimal solution to problem (38) must be fulfilled with equality. We

Algorithm 2 proposed algorithm to solve (P2)

1: **Initialization**: set $n := 0$, and generate $\boldsymbol{a}^{(0)}$ and $\mathbf{A}^{(0)}$.
2: **repeat**
3: $\boldsymbol{v}^{(n)} = \left(\boldsymbol{H}\mathbf{A}^{(n)}\boldsymbol{R}_{\mathrm{s}}\mathbf{A}^{(n)\dagger}\boldsymbol{H}^{\dagger} + \boldsymbol{R}_{\mathrm{n}}\right)^{-1}\boldsymbol{H}\boldsymbol{a}^{(n)}$
4: Solve problem (SDR2) with $\boldsymbol{v} = \boldsymbol{v}^{(n)}$ to obtain the optimal value $(\boldsymbol{Q}^{\star}, \boldsymbol{W}^{\star})$.
5: Set $(\boldsymbol{Q}^{(n+1)}, \boldsymbol{W}^{(n+1)}) := (\boldsymbol{Q}^{\star}, \boldsymbol{W}^{\star})$.
6: Construct $\{\boldsymbol{a}^{(n+1)}, \mathbf{A}^{(n+1)}\}$ from $\boldsymbol{Q}^{(n+1)}$.
7: Update $n := n + 1$.
8: **until convergence**
9: **Output:** $(\boldsymbol{Q} = \boldsymbol{Q}^{(n)}, \boldsymbol{W} = \boldsymbol{W}^{(n)}, \boldsymbol{v} = \boldsymbol{v}^{(n)})$

prove it by contradiction. Assume that the MSE constraint is satisfied with a strict inequality at the optimal solution $(\boldsymbol{a}^{\star}, \{\boldsymbol{w}_i^{\star}\}_{i=1}^{n_{\mathrm{b}}})$. By letting $\bar{\boldsymbol{a}} = t\boldsymbol{a}^{\star}$ for $0 < t < 1$, we can choose a sufficient large t such that

$$\frac{|\boldsymbol{a}^{\star}\boldsymbol{f}|^2}{\boldsymbol{a}^{\star\top}\boldsymbol{F}\boldsymbol{R}_{\mathrm{s}}\boldsymbol{F}^{\dagger}\boldsymbol{a}^{\star *} + \boldsymbol{v}^{\dagger}\boldsymbol{R}_{\mathrm{n}}\boldsymbol{v}} > \frac{\left|\bar{\boldsymbol{a}}^{\top}\boldsymbol{f}\right|^2}{\bar{\boldsymbol{a}}^{\top}\boldsymbol{F}\boldsymbol{R}_{\mathrm{s}}\boldsymbol{F}^{\dagger}\bar{\boldsymbol{a}}^{*} + \boldsymbol{v}^{\dagger}\boldsymbol{R}_{\mathrm{n}}\boldsymbol{v}} \geq \gamma. \tag{39}$$

When $\bar{\boldsymbol{w}}_i = t\boldsymbol{w}_i^{\star}$, $(\bar{\boldsymbol{a}}, \{\bar{\boldsymbol{w}}_i\}_{i=1}^{n_{\mathrm{s}}})$ can also be a feasible solution to problem (38) with the new objective value $t^2\sum_{i=1}^{n_{\mathrm{b}}}\|\boldsymbol{w}_i^{\star}\|^2$, which is definitely smaller than the optimal value when the optimal solution is $(\boldsymbol{a}^{\star}, \{\boldsymbol{w}_i^{\star}\}_{i=1}^{n_{\mathrm{s}}})$. This contradicts to the assumption that $(\boldsymbol{a}^{\star}, \{\boldsymbol{w}_i^{\star}\}_{i=1}^{n_{\mathrm{s}}})$ is optimal. Therefore, the MSE constraint must hold with equality. Let $\boldsymbol{Q} = \boldsymbol{a}^{*}\boldsymbol{a}^{\top}$, $\boldsymbol{W} = \sum_{i=1}^{n_{\mathrm{b}}}\boldsymbol{w}_i\boldsymbol{w}_i^{\dagger}$, $\boldsymbol{\Sigma} = \boldsymbol{f}\boldsymbol{f}^{\top}$, $\boldsymbol{\Psi} = \boldsymbol{F}\boldsymbol{R}_{\mathrm{s}}\boldsymbol{F}^{\dagger}$, $\boldsymbol{G}_k = \boldsymbol{g}_k\boldsymbol{g}_k^{\dagger}$, and $\mathbf{D}_k = \mathrm{diag}(0, \ldots, \delta_\theta^2 + \sigma_{\mathrm{u},k}^2, \ldots, 0)$. Then, as in problem (15), we will omit the rank constraint on $\boldsymbol{Q}$ and solve a relaxed version of (38), which leads to

$$\begin{aligned}
&(\mathsf{SDR2}):\\
&\underset{\boldsymbol{Q},\boldsymbol{W}}{\text{minimize}} \quad \mathrm{tr}\,(\boldsymbol{W})\\
&\text{subject to} \quad \mathrm{tr}\,(\boldsymbol{Q}\boldsymbol{\Sigma}) = \gamma\,\mathrm{tr}\,(\boldsymbol{Q}\boldsymbol{\Psi}) + \gamma\boldsymbol{v}^{\dagger}\boldsymbol{R}_{\mathrm{n}}\boldsymbol{v}\\
&\qquad\qquad\quad \mathrm{tr}\,(\mathbf{D}_k\boldsymbol{Q}) - \mathrm{tr}\,(\boldsymbol{G}_k\boldsymbol{W}) \leq 0, \quad \forall k \in [n_{\mathrm{s}}]\\
&\qquad\qquad\quad \boldsymbol{W} \succeq \mathbf{0}, \boldsymbol{Q} \succeq \mathbf{0}.
\end{aligned}$$

The following theorem characterizes the properties of the optimal solution to problem (SDR2).

Theorem 2 (Properties of Optimal Solution): Let $\beta^{\star}$ be the dual optimal solutions associated with the equality constraint in (SDR2). We also let $\boldsymbol{Q}^{\star}$ and $\boldsymbol{W}^{\star}$ be the primal optimal solutions to (SDR2). Then the following three properties are fulfilled:

1) $\beta^{\star} > 0$;
2) $\mathsf{rank}\,(\boldsymbol{W}^{\star}) \leq \min(n_{\mathrm{s}}, n_{\mathrm{r}})$;
3) $\mathsf{rank}\,(\boldsymbol{Q}^{\star}) = 1$.

Proof: The proof can be found using the similar steps to the proof for Theorem 1. □

Similarly as in Section III, we summarize the overall procedure for solving problem (P2) in Algorithm 2. In this algorithm, the objective value is monotonically reduced in each step, and for a given feasible threshold γ, it is lower-bounded by a certain value. As a result, the algorithm converges at least to a local optimum. Finally, it can be verified that the computational complexity of Algorithm 2 is same as that of Algorithm 1.

B. Single-Antenna FC

It would also be of interest to study problem (P2) for the single-antenna FC scenario. In this case, problem (P2) can be rewritten as

$$\begin{aligned}
&\underset{\boldsymbol{a},P}{\text{minimize}} \quad P\\
&\text{subject to} \quad \frac{\boldsymbol{a}^{\top}\boldsymbol{h}\boldsymbol{h}^{\dagger}\boldsymbol{a}^{*}}{\boldsymbol{a}^{\top}\boldsymbol{F}\boldsymbol{R}_{\mathrm{s}}\boldsymbol{F}^{\dagger}\boldsymbol{a}^{*} + \sigma_{\mathrm{n}}^2} = \gamma\\
&\qquad\qquad\quad |\alpha_k|^2\left(\delta_\theta^2 + \sigma_{\mathrm{u},k}^2\right) \leq P\,|g_k|^2, \ \forall k \in [n_{\mathrm{s}}].
\end{aligned} \tag{40}$$

Define the matrices $\boldsymbol{\Omega} = \begin{pmatrix} \boldsymbol{a}^{*}\boldsymbol{a}^{\top} & \mathbf{0}\\ \mathbf{0} & P\end{pmatrix}$; $\boldsymbol{P} = \begin{pmatrix} \mathbf{0} & \mathbf{0}\\ \mathbf{0} & 1\end{pmatrix}$; $\bar{\mathbf{D}}_k = \begin{pmatrix} \frac{1}{|g_k|^2}\mathbf{D}_k & \mathbf{0}\\ \mathbf{0} & -1\end{pmatrix}$; and $\boldsymbol{E} = \boldsymbol{h}\boldsymbol{h}^{\dagger} - \gamma\boldsymbol{F}\boldsymbol{R}_{\mathrm{s}}\boldsymbol{F}^{\dagger}$. Then, problem (40) can be recast as

$$\begin{aligned}
&\underset{\boldsymbol{\Omega}}{\text{minimize}} \quad \mathrm{tr}\,(\boldsymbol{\Omega}\boldsymbol{P})\\
&\text{subject to} \quad \mathrm{tr}\,(\boldsymbol{\Omega}\boldsymbol{E}) = \gamma\sigma_{\mathrm{n}}^2\\
&\qquad\qquad\quad \mathrm{tr}\left(\boldsymbol{\Omega}\bar{\mathbf{D}}_k\right) \leq 0, \quad \forall k \in [n_{\mathrm{s}}]\\
&\qquad\qquad\quad \mathsf{rank}\,(\boldsymbol{\Omega}) = 1.
\end{aligned} \tag{41}$$

By dropping the rank constraint on $\boldsymbol{\Omega}$, problem (41) is a SDP and thus can be solved efficiently. If we denote $\boldsymbol{\Omega}^{\star}$ by the optimal solution to the relaxed problem of (41), then $\mathsf{rank}\,(\boldsymbol{\Omega}^{\star}) = 1$ and the optimal $\boldsymbol{a}$ and P can be found from $\boldsymbol{\Omega}^{\star}$. Particularly,

$$P^{\star} = (\boldsymbol{\Omega}^{\star})_{n_{\mathrm{s}}+1, n_{\mathrm{s}}+1} \tag{42}$$

$$\boldsymbol{a}^{\star} = \sqrt{\mathrm{tr}(\boldsymbol{\Omega}_{n_{\mathrm{s}}}^{\star})}\,\boldsymbol{u}_1^{*}, \tag{43}$$

where $\boldsymbol{\Omega}_{n_{\mathrm{s}}}^{\star}$ is the n_{s}th order leading principal submatrix of $\boldsymbol{\Omega}^{\star}$ obtained by excluding the $(n_{\mathrm{s}}+1)$th row and column and $\boldsymbol{u}_1$ is the eigenvector associated with the largest eigenvalue of $\boldsymbol{\Omega}_{n_{\mathrm{s}}}^{\star}$. Similarly as in problem (P1), in this case, the optimal solution $(P^{\star}, \boldsymbol{a}^{\star})$ is indeed a global optimum.

C. A Common Energy Harvester

Now, we consider the converse problem of (P1–Sum), in which we aim to minimize the transmit power at the FC subject to a minimum requirement of the MSE performance,

$$\begin{aligned}
(\mathsf{P2\text{–}Sum}):\ &\underset{\boldsymbol{v},\boldsymbol{a},\boldsymbol{w}}{\text{minimize}} \quad \|\boldsymbol{w}\|^2\\
&\text{subject to} \quad \frac{\left|\boldsymbol{v}^{\dagger}\boldsymbol{H}\boldsymbol{a}\right|^2}{\boldsymbol{v}^{\dagger}\left(\boldsymbol{H}\mathbf{A}\boldsymbol{R}_{\mathrm{s}}\mathbf{A}^{\dagger}\boldsymbol{H}^{\dagger} + \boldsymbol{R}_{\mathrm{n}}\right)\boldsymbol{v}} \geq \gamma\\
&\qquad\qquad\quad \boldsymbol{a}^{\dagger}\mathbf{D}\boldsymbol{a} \leq \left|\boldsymbol{w}^{\dagger}\boldsymbol{h}_{\mathrm{e}}\right|^2.
\end{aligned}$$

If we multiply $\boldsymbol{w}$ and $\boldsymbol{a}$ by a scalar $\alpha > 1$ and $\beta < 1$, respectively, then the left-hand side of the MSE constraint (i.e., the first constraint) is strictly increased while the right-hand side of the sum power constraint (i.e., the second constraint) as well as the objective function are strictly decreased. Thus, the optimality for (P2–Sum) is achieved when all the above constraints are satisfied with equality. Problem (P2–Sum) can be formulated as a SDP, and hence solved efficiently by CVX. In the following, we establish a fundamental relationship between two problems (P1–Sum) and (P2–Sum).

Theorem 3 (Power–Distortion Trade-off): For a distributed estimation system using the BLUE with a common energy harvester, if we assume that the alternative algorithms solving (P1–Sum) and (P2–Sum) are initialized with $\boldsymbol{a}^{(0)}$, then the optimal *power–distortion trade-off* is given by

$$\frac{1}{\mathsf{mse}} = \boldsymbol{f}^\dagger \left(\boldsymbol{F} \boldsymbol{R}_\mathrm{s} \boldsymbol{F}^\dagger + \frac{\boldsymbol{v}^\dagger \boldsymbol{R}_\mathrm{n} \boldsymbol{v}}{P \left\| \boldsymbol{h}_\mathrm{e} \right\|^2} \mathbf{D} \right)^{-1} \boldsymbol{f}. \tag{44}$$

Proof: See Appendix B. □

Theorem 3 is important since it enables to (numerically) find the *power–distortion trade-off* for distributed estimation in the cumulative power constraint case.

V. Numerical Results

In this section, we provide numerical examples by evaluating our proposed algorithms in Sections III and IV. In the simulations, we consider the widely used 915 MHz frequency band in WSNs [45] for both energy and information transmissions. For energy transmission, we consider the use of both the commercially available power transmitter (Powercast TX91501) with transmit power $P = 1$W (30 dBm) and the RF power harvester (Powercast P2110). The detailed system parameters are summarized in Table I. To model a small-scale fading, we assume that the elements of the channel matrices are drawn independently from the Gaussian distribution with zero mean and unit variance. To further evaluate the effectiveness of the proposed algorithms, we also perform comparisons to low-complexity baseline schemes specified below.

A. Baseline Schemes

1) Suboptimal Design for (P1): We divide the optimization procedure into two phases. In the first phase, the energy beamforming vectors $\{\boldsymbol{w}_i\}_{i=1}^{n_\mathrm{b}}$ are designed such that the total harvested energy is maximized, which leads to the following maximization problem:

$$\begin{aligned} \underset{\{\boldsymbol{w}_i\}_{i=1}^{n_\mathrm{b}}}{\text{maximize}} \quad & \sum_{k=1}^{n_\mathrm{s}} \beta_k \left(\sum_{i=1}^{n_\mathrm{b}} \left| \boldsymbol{w}_i^\dagger \boldsymbol{g}_k \right|^2 \right) \\ \text{subject to} \quad & \sum_{i=1}^{n_\mathrm{b}} \left\| \boldsymbol{w}_i \right\|^2 \leq P. \end{aligned} \tag{45}$$

Here, $\{\beta_k\}_{k=1}^{n_\mathrm{s}}$ denote the energy weights indicating the priority (e.g., sensors with weaker channels can be assigned to a higher weight to guarantee fairness) of the corresponding sensors. It has been shown in [8] that the optimal strategy is to allocate all the power budget to the direction of $\boldsymbol{\eta}$—the eigenvector associated with the largest eigenvalue of the matrix $\sum_{k=1}^{n_\mathrm{s}} \beta_k \boldsymbol{g}_k \boldsymbol{g}_k^\dagger$. The optimal value in problem (45) is achieved when $\boldsymbol{w}_i^\star = \sqrt{p_i} \boldsymbol{\eta}$ with $p_i \geq 0$ such that $\sum_{i=1}^{n_\mathrm{b}} p_i = P$.

In the second phase, we find the amplification vector $\boldsymbol{a}$ and the receive filtering $\boldsymbol{v}$ in terms of minimizing the MSE subject to the energy harvested in the first phase. In particular, we solve the following problem:

$$\begin{aligned} \underset{\boldsymbol{a}, \boldsymbol{v}}{\text{maximize}} \quad & \frac{\left| \boldsymbol{a}^\top \boldsymbol{f} \right|^2}{\boldsymbol{a}^\top \boldsymbol{F} \boldsymbol{R}_\mathrm{s} \boldsymbol{F}^\dagger \boldsymbol{a}^* + \boldsymbol{v}^\dagger \boldsymbol{R}_\mathrm{n} \boldsymbol{v}} \\ \text{subject to} \quad & \left| \alpha_k \right|^2 \left(\delta_\theta^2 + \sigma_{\mathrm{u},k}^2 \right) \leq P_k, \quad \forall k \in [n_\mathrm{s}], \end{aligned} \tag{46}$$

where $\boldsymbol{f} = \left[\boldsymbol{v}^\dagger \boldsymbol{h}_1 \, \boldsymbol{v}^\dagger \boldsymbol{h}_2 \, \cdots \, \boldsymbol{v}^\dagger \boldsymbol{h}_{n_\mathrm{s}} \right]^\top$; $\boldsymbol{F} = \mathrm{diag}\left(\boldsymbol{f} \right)$; and $P_k = \left| \boldsymbol{g}_k^\dagger \boldsymbol{w}_i^\star \right|^2$. Problem (46) corresponds to problem (P1) without the total power constraint and can be solved by iteratively updating $\boldsymbol{v}$ and $\boldsymbol{a}$.

2) Suboptimal Design for (P2): To reduce the computational burden of the joint optimization for (P2), we propose a suboptimal design, in which the optimization procedure is divided into two phases. In the first phase, we aim to solve the following problem:

$$\begin{aligned} \underset{\boldsymbol{a}, \boldsymbol{v}}{\text{minimize}} \quad & \boldsymbol{a}^\dagger \mathbf{D} \, \boldsymbol{a} \\ \text{subject to} \quad & \frac{\left| \boldsymbol{v}^\dagger \boldsymbol{H} \boldsymbol{a} \right|^2}{\boldsymbol{v}^\dagger \left(\boldsymbol{H} \mathbf{A} \boldsymbol{R}_\mathrm{s} \mathbf{A}^\dagger \boldsymbol{H}^\dagger + \boldsymbol{R}_\mathrm{n} \right) \boldsymbol{v}} \geq \gamma. \end{aligned} \tag{47}$$

We note that the objective function in problem (47) is the total transmit power of the sensors. Since the receive filtering $\boldsymbol{v}$ appears only in the constraint, we can iteratively solve problem (47) for $\boldsymbol{a}$ and $\boldsymbol{v}$. Since the left-hand side of the constraint is nondecreasing with the norm of $\boldsymbol{a}$, the constraint must be satisfied with equality. For a fixed $\boldsymbol{v}$, the above problem can be expressed as follows:

$$\begin{aligned} \underset{\boldsymbol{a}, \boldsymbol{v}}{\text{minimize}} \quad & \boldsymbol{a}^\top \mathbf{D} \, \boldsymbol{a}^* \\ \text{subject to} \quad & \boldsymbol{a}^\top \boldsymbol{E} \, \boldsymbol{a}^* = \gamma \, \boldsymbol{v}^\dagger \boldsymbol{R}_\mathrm{n} \boldsymbol{v}, \end{aligned} \tag{48}$$

where $\boldsymbol{E} = \boldsymbol{f}\boldsymbol{f}^\dagger - \gamma \boldsymbol{F} \boldsymbol{R}_\mathrm{s} \boldsymbol{F}^\dagger$. To guarantee the feasibility of problem (48), the value of γ must be chosen such that $\left| \boldsymbol{a}^\dagger \boldsymbol{f} \right|^2 \geq \gamma \boldsymbol{a}^\top \boldsymbol{F} \boldsymbol{R}_\mathrm{s} \boldsymbol{F}^\dagger \boldsymbol{a}^*$. Since the quantities $\boldsymbol{a}^\top \mathbf{D} \, \boldsymbol{a}^* \geq 0$ and $\boldsymbol{a}^\top \boldsymbol{E} \, \boldsymbol{a}^* \geq 0$ are positive, problem (48) can be rewritten as

$$\begin{aligned} \underset{\boldsymbol{a}, \boldsymbol{v}}{\text{maximize}} \quad & \frac{\boldsymbol{a}^\top \boldsymbol{E} \, \boldsymbol{a}^*}{\boldsymbol{a}^\top \mathbf{D} \, \boldsymbol{a}^*} \\ \text{subject to} \quad & \boldsymbol{a}^\top \boldsymbol{E} \, \boldsymbol{a}^* = \gamma \, \boldsymbol{v}^\dagger \boldsymbol{R}_\mathrm{n} \boldsymbol{v} \end{aligned}, \tag{49}$$

which is a Rayleigh quotient. Thus, the optimal solution to problem (49) is given by

$$\boldsymbol{a}^\star = \sqrt{\frac{\gamma \, \boldsymbol{v}^\dagger \boldsymbol{R}_\mathrm{n} \boldsymbol{v}}{\boldsymbol{u}_1^\dagger \mathbf{D}^{-1/2} \boldsymbol{E} \mathbf{D}^{-1/2} \boldsymbol{u}_1}} \mathbf{D}^{-1/2} \boldsymbol{u}_1^*, \tag{50}$$

where $\boldsymbol{u}_1$ denotes the unit-norm eigenvector associated with the largest eigenvalue of the matrix $\mathbf{D}^{-1/2} \boldsymbol{E} \mathbf{D}^{-1/2}$, $\lambda_{\max} \left(\mathbf{D}^{-1/2} \boldsymbol{E} \mathbf{D}^{-1/2} \right)$. It follows that the minimum total transmit power of the sensors, $P_\mathsf{s}^\star$, required to achieve the MSE of $1/\gamma$ is given by

$$P_\mathsf{s}^\star = \frac{\gamma \, \boldsymbol{v}^\dagger \boldsymbol{R}_\mathrm{n} \boldsymbol{v}}{\lambda_{\max} \left(\mathbf{D}^{-1/2} \boldsymbol{E} \mathbf{D}^{-1/2} \right)}. \tag{51}$$

In the second phase, we aim to minimize the total transmit power at the FC with the amplification coefficients $\{\alpha_k\}_{k=1}^{n_\mathrm{s}}$

TABLE I
SYSTEM PARAMETERS

Parameter	Value
Network Topology	10m $\times$ 10m square box
Fusion Center	Located at the origin $(0, 0)$
	Transmission power 30 dBm
	Receiver noise power -103.16 dBm (effective noise bandwidth 2 MHz and noise figure 7 dB)
Sensors	Placed uniformly over $\{x, y \mid x, y \in [-10, 10]\}$
	Energy harvesting efficiency 51%
Path loss [44]	$\mathsf{PL}(d) = 31.7 + 27.6 \log_{10}(d_{[\text{meters}]})$ [dB]

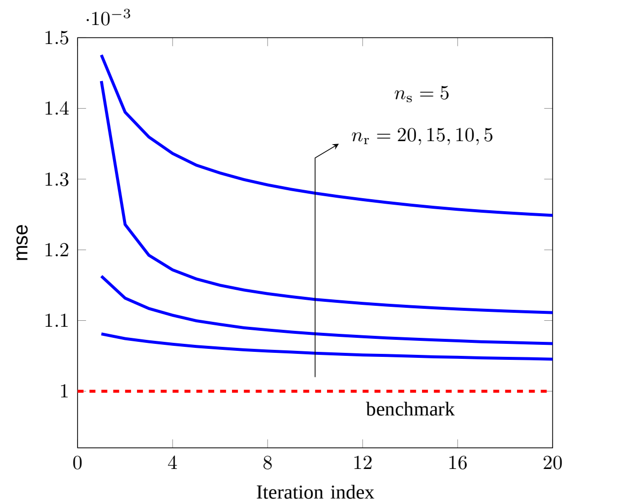

Fig. 2. The average MSE for distributed estimation versus iteration index when $n_s = 5$, $\boldsymbol{R}_s = 0.1\mathbf{I}_{n_s}$, $P = 30$ dBm, $\delta_\theta = 1$, and $n_r = 5, 10, 15, 20$.

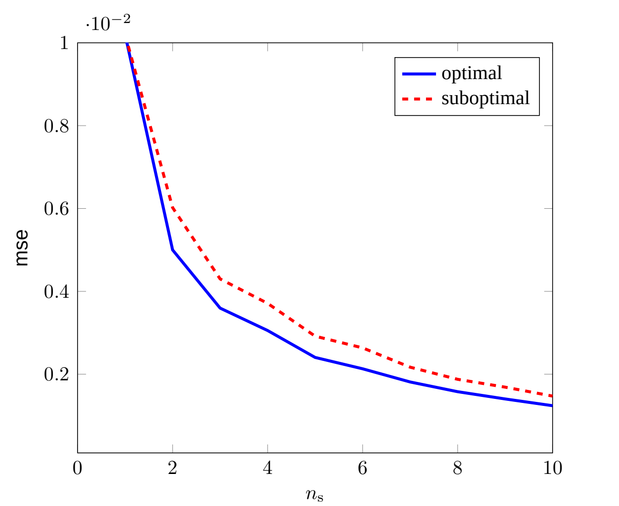

Fig. 3. The average MSE for distributed estimation as a function of n_s when $n_r = 5$, $P = 30$ dBm, $\boldsymbol{R}_s = 0.1\mathbf{I}_{n_s}$, and $\delta_\theta = 1$.

that are the solutions to problem (47). In other words, we find the optimal solution to the following minimization problem:

$$\begin{aligned} &\underset{\{\boldsymbol{w}_i\}_{i=1}^{n_b}}{\text{minimize}} \quad \sum_{i=1}^{n_b} \|\boldsymbol{w}_i\|^2 \\ &\text{subject to} \quad \sum_{i=1}^{n_b} \left|\boldsymbol{w}_i^\dagger \boldsymbol{g}_k\right|^2 \geq |\alpha_k|^2 \left(\delta_\theta^2 + \sigma_{u,k}^2\right), \forall k \in [n_s]. \end{aligned} \tag{52}$$

Problem (52) can be effectively solved by CVX. In the following subsections, we use these suboptimal designs as the baseline schemes to assess the effectiveness of our proposed algorithms.

B. MSE Minimization

Figure 2 shows the average MSE for distributed estimation versus iteration index when $n_s = 5$, $\boldsymbol{R}_s = 10^{-2}\mathbf{I}_{n_s}$, $P = 30$ dBm, $\delta_\theta = 1$, and $n_r = 5, 10, 15, 20$. One can see that the average MSE monotonically decreases while the algorithm converges within a few iterations. It can be obviously seen that the MSE performance is improved with the increasing number of antennas at the FC, n_r. In this figure, we also plot a benchmark ideal case for distributed estimation, where all the observations at the sensors are assumed to be directly available at the FC, which will give a lower bound on the MSE performance. One can see that the average MSE evaluated via our simulation tends to approach the benchmark value, $\left[\mathbf{1}^\dagger \boldsymbol{R}_s^{-1} \mathbf{1}\right]^{-1}$, as n_r increases. In Fig. 3, the average MSE for distributed estimation is shown as a function of n_s for the optimal and suboptimal solutions when $P = 30$ dBm, $n_r = 5$, $\boldsymbol{R}_s = 0.1\mathbf{I}_{n_s}$, $\boldsymbol{R}_n = 0.5\mathbf{I}_{n_r}$, and $\delta_\theta = 1$. As expected, the MSE performance is improved as n_s increases. In this example, we can see that the suboptimal solution shows a reasonable performance compared to the optimal one.

In Table II, in order to elaborate on the attributes of the optimal solution to the MSE minimization problem, we present the values of the harvested and transmit power of each sensor at the optimal solution to problem (P1). In this example, we set $P = 30$ dBm, $\boldsymbol{R}_s = 0.1\mathbf{I}_{n_s}$, $n_s = 7$, and $n_r = 2$. One can see that some of the sensors do not use their maximum power harvested from the FC, which implies that power control is needed to guarantee the optimal solution. In other words, some of individual power constraints (i.e., the first constraint) in problem (P1) may not be fully utilized, or equivalently, the corresponding dual variables may be zero. This is attributed from the fact that transmission with the full power may increase the interference level at the FC, which in turn reduces the estimation reliability. In this example, sensors 2, 5, 6, and 7 use only a fraction of their harvested power.

C. Total Power Minimization

Figure 4 illustrates the average minimum transmit power at the FC for distributed estimation versus iteration index at the distortion target of $\gamma^{-1} = 0.015$ when $n_s = 10$, $\boldsymbol{R}_s = 0.1\mathbf{I}_{n_s}$, $\delta_\theta = 1$, and $n_r = 5, 10, 15, 20$. As shown in this figure, the proposed algorithm converges quickly, and the transmit power is reduced as the number of antennas, n_r, increases. In Fig. 5, the average minimum transmit power at the FC for distributed estimation is shown as a function of the distortion target γ^{-1} when $n_s = 10$, $\boldsymbol{R}_s = 0.1\mathbf{I}_{n_s}$, $\delta_\theta = 1$, and $n_r = 5$. It is clear that the more strict distortion requirement is, the more power

TABLE II
POWER CONTROL

Sensor index	Harvested power [dBm]	Transmit power [dBm]
1	-31.449	-31.449
2	-27.687	-34.347
3	-30.737	-30.737
4	-32.865	-32.865
5	-13.847	-48.067
6	-29.886	-31.999
7	-28.307	-32.964

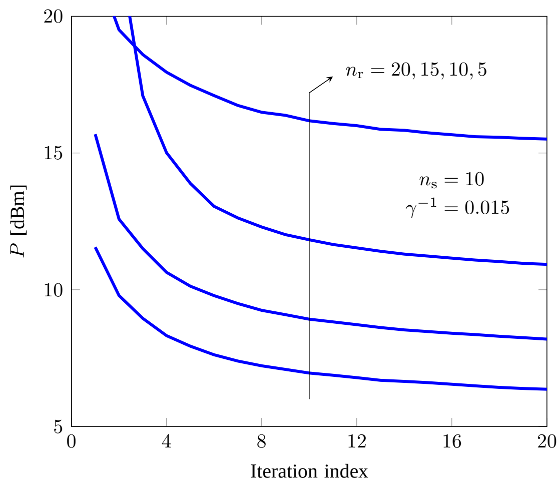

Fig. 4. The average minimum transmit power at the FC for distributed estimation versus iteration index when $n_s = 10$, $\boldsymbol{R}_s = 0.1\mathbf{I}_{n_s}$, $\gamma^{-1} = 0.015$, $\delta_\theta = 1$, and $n_r = 5, 10, 15, 20$.

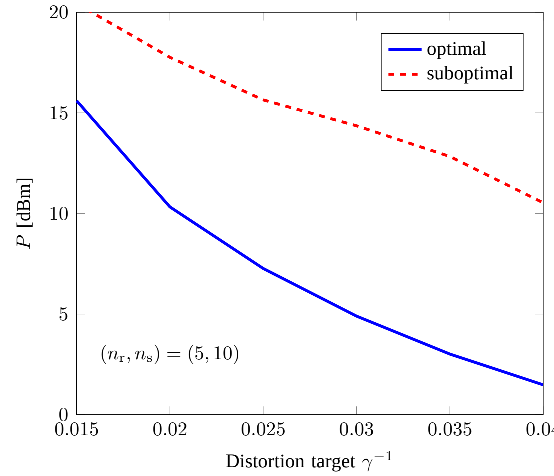

Fig. 5. The average minimum transmit power at the FC for distributed estimation as a function of the distortion target γ^{-1} when $n_s = 10$, $\boldsymbol{R}_s = 0.1\mathbf{I}_{n_s}$, $\delta_\theta = 1$, and $n_r = 5$.

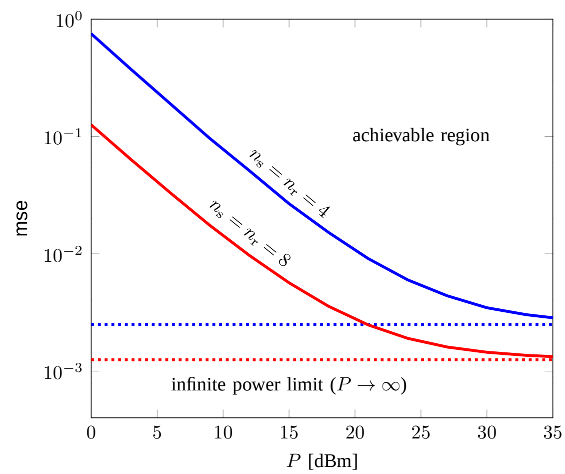

Fig. 6. The distortion–power trade-off for distributed estimation with a common energy harvester when $\boldsymbol{R}_s = 10^{-2}\mathbf{I}_{n_s}$, $\delta_\theta = 1$, $n_s = n_r = 4$, and $n_s = n_r = 8$.

is needed. Note that the distortion target must be no smaller than the benchmark MSE value such that the optimization problem is feasible. One can also see that the optimal scheme should be used for power saving. In this example, we can save the amount of transmit power of 7.44, 9.46, and 9.05 dBm at $\gamma^{-1} = 0.02, 0.03, 0.04$, respectively, compared to the suboptimal case.

D. A Common Energy Harvester

Finally, we validate the performance of the distributed estimation system with a common energy harvester. Specifically, the power–distortion trade-off is ascertained by referring to Fig. 6, where the optimal MSE is depicted as a function of the minimum transmit power P for distributed estimation when $\boldsymbol{R}_s = 10^{-2}\mathbf{I}_{n_s}$, $\delta_\theta = 1$, $n_s = n_r = 4$, and $n_s = n_r = 8$. In this figure, the region above each trade-off curve is achievable. As P tends to infinity, the MSE converges to that of centralized estimations, i.e., $\left[\mathbf{1}^\dagger \boldsymbol{R}_s^{-1}\mathbf{1}\right]^{-1}$, plotted with the dotted curve. Moreover, as expected, the achievable region gets broader for a larger (n_s, n_r) pair.

VI. Concluding Remarks

Using the SDR, we developed a new framework for solving the network lifetime problem of a WSN. To that end, we adopted the notion of RF-based WPT as well as the multiple-antenna technology so that both the life span and the estimation performance are substantially improved. In this paper, two optimization problems were formulated and iteratively solved by two proposed algorithms, which turned out to guarantee the convergence at least to a local optimum. We showed that power control is indeed required at the optimal solution. It was also shown that having multiple antennas at the FC provides a significant improvement in the estimation performance. Especially, it was shown that as the number of antennas grows large, the MSE of the distributed estimation with the BLUE approaches that of centralized estimations.

Appendix A
Proof of Theorem 1

Proof: We start by proving the first property of Theorem 1. We exploit the strong duality and then examine the Karush-Kuhn-Tucker condition of (SDR1). Let ν, $\{\lambda_k\}_{k=1}^{n_s}$, and β be the dual variables of problem (SDR1). The Lagrangian of problem (SDR1) is defined as

$$\mathcal{L}\left(\bar{\boldsymbol{Q}},\bar{\boldsymbol{W}},\eta,\nu,\lambda_k,\beta\right)=-\operatorname{tr}\left(\bar{\boldsymbol{Q}}\boldsymbol{\Sigma}\right)+\beta\left(\operatorname{tr}\left(\bar{\boldsymbol{W}}\right)-\eta P\right)$$
$$+\sum_{k=1}^{n_{\mathrm{s}}}\lambda_k\operatorname{tr}\left(\mathbf{D}_k\bar{\boldsymbol{Q}}-\boldsymbol{G}_k\bar{\boldsymbol{W}}\right)+\nu\left(\operatorname{tr}\left(\bar{\boldsymbol{Q}}\boldsymbol{\Psi}\right)+\eta\boldsymbol{v}^\dagger\boldsymbol{R}_{\mathrm{n}}\boldsymbol{v}-1\right).$$

Then, the dual function of problem (SDR1) is given by

$$\min_{\bar{\boldsymbol{Q}}\succeq\mathbf{0},\bar{\boldsymbol{W}}\succeq\mathbf{0},\eta>0}\mathcal{L}\left(\bar{\boldsymbol{Q}},\bar{\boldsymbol{W}},\eta,\nu,\lambda_k,\beta\right),$$

which can be equivalently expressed as

$$\min_{\bar{\boldsymbol{Q}}\succeq\mathbf{0},\bar{\boldsymbol{W}}\succeq\mathbf{0},\eta>0}\operatorname{tr}\left(\bar{\boldsymbol{Q}}\boldsymbol{Y}\right)+\operatorname{tr}\left(\bar{\boldsymbol{W}}\boldsymbol{Z}\right)+\eta\xi-\nu, \tag{53}$$

where

$$\begin{aligned}\xi&=\nu\boldsymbol{v}^\dagger\boldsymbol{R}_{\mathrm{n}}\boldsymbol{v}-\beta P\\ \boldsymbol{Y}&=-\boldsymbol{\Sigma}+\sum_{k=1}^{n_{\mathrm{s}}}\lambda_k\mathbf{D}_k+\nu\boldsymbol{\Psi}\\ \boldsymbol{Z}&=-\sum_{k=1}^{n_{\mathrm{s}}}\lambda_k\boldsymbol{G}_k+\beta\mathbf{I}.\end{aligned}$$

When we let $\nu^\star$, $\{\lambda_k^\star\}_{k=1}^{n_{\mathrm{s}}}$, and $\beta^\star$ be the optimal dual solutions to problem (SDR1), we define

$$\boldsymbol{Y}^\star=-\boldsymbol{\Sigma}+\sum_{k=1}^{n_{\mathrm{s}}}\lambda_k^\star\mathbf{D}_k+\nu^\star\boldsymbol{\Psi} \tag{54}$$

$$\boldsymbol{Z}^\star=-\sum_{k=1}^{n_{\mathrm{s}}}\lambda_k^\star\boldsymbol{G}_k+\beta^\star\mathbf{I}. \tag{55}$$

Then, the optimal $\bar{\boldsymbol{Q}}^\star$ must be the solution to the following problem:

$$\underset{\bar{\boldsymbol{Q}}\succeq\mathbf{0}}{\text{minimize}}\quad\operatorname{tr}\left(\bar{\boldsymbol{Q}}\boldsymbol{Y}^\star\right). \tag{56}$$

To guarantee a bounded optimal value, we must have $\boldsymbol{Y}^\star\succeq\mathbf{0}$, and hence we obtain the optimal value $\operatorname{tr}(\bar{\boldsymbol{Q}}^\star\boldsymbol{Y}^\star)=0$, which implies that

$$\bar{\boldsymbol{Q}}^\star\boldsymbol{Y}^\star=\mathbf{0}. \tag{57}$$

In the same manner, it follows that $\bar{\boldsymbol{W}}^\star\boldsymbol{Z}^\star=\mathbf{0}$ and $\eta^\star\xi^\star=0$, or equivalently $\xi^\star=0$ since $\eta^\star>0$, where $\xi^\star=\nu^\star\boldsymbol{v}^\dagger\boldsymbol{R}_{\mathrm{n}}\boldsymbol{v}-\beta^\star P$. From (53), the dual problem to problem (SDR1) can be rewritten as follows:

$$\begin{aligned}\underset{\nu,\{\lambda_k\}_{k=1}^{n_{\mathrm{s}}},\beta}{\text{minimize}}\quad&\nu\\ \text{subject to}\quad&\boldsymbol{Y}\succeq\mathbf{0},\boldsymbol{Z}\succeq\mathbf{0},\xi\geq 0\\ &\beta\geq 0,\lambda_k\geq 0,\quad\forall k\in[n_{\mathrm{s}}].\end{aligned} \tag{58}$$

Since the duality gap between problem (SDR1) and (58) is zero, $\nu^\star$ is equal to the optimal value of problem (SDR1), which is positive. Thus, we conclude that $\nu^\star>0$. Next, we will show that $\beta^\star>0$. First, if there exits a $\lambda_k>0$, then from the condition $\boldsymbol{Z}\succeq\mathbf{0}$, it follows that $\beta^\star>0$. From the condition $\xi^\star=0$ and the facts that $\boldsymbol{R}_{\mathrm{n}}\succ\mathbf{0}$ and $\nu^\star>0$, we also conclude that $\beta^\star>0$.

To prove the second property of Theorem 1, we use the fact that for any two matrices of the same size $\mathbf{A}$ and $\mathbf{B}$, $\operatorname{rank}(\mathbf{A}-\mathbf{B})\geq|\operatorname{rank}(\mathbf{A})-\operatorname{rank}(\mathbf{B})|$ [43]. Since $\beta^\star>0$ and $\operatorname{rank}\left(\sum_{k=1}^{n_{\mathrm{s}}}\lambda_k^\star\boldsymbol{G}_k\right)\leq n_{\mathrm{s}}$, it follows from (55) that $\operatorname{rank}(\boldsymbol{Z}^\star)\geq|n_{\mathrm{r}}-n_{\mathrm{s}}|\geq n_{\mathrm{r}}-n_{\mathrm{s}}$. Let $\operatorname{Null}(\boldsymbol{Z}^\star)$ be the null space of $\boldsymbol{Z}^\star$. Then from the condition $\bar{\boldsymbol{W}}^\star\boldsymbol{Z}^\star=\mathbf{0}$, we must have $\bar{\boldsymbol{W}}^\star\in\operatorname{Null}(\boldsymbol{Z}^\star)$. Since $\operatorname{rank}(\boldsymbol{Z}^\star)\geq n_{\mathrm{r}}-n_{\mathrm{s}}$ and $\operatorname{rank}\left(\bar{\boldsymbol{W}}^\star\right)\leq\dim(\operatorname{Null}(\boldsymbol{Z}^\star))$, it follows that $\operatorname{rank}\left(\bar{\boldsymbol{W}}^\star\right)\leq n_{\mathrm{s}}$. Using the fact that $\bar{\boldsymbol{W}}^\star$ is an $n_{\mathrm{r}}\times n_{\mathrm{r}}$ matrix, we conclude that $\operatorname{rank}\left(\bar{\boldsymbol{W}}^\star\right)\leq n_{\mathrm{r}}$.

Finally, we prove the property of the optimal solution $\bar{\boldsymbol{Q}}^\star$. Since $\boldsymbol{\Psi}\succ\mathbf{0}$, $\sum_{k=1}^{n_{\mathrm{s}}}\lambda_k^\star\mathbf{D}_k\succeq\mathbf{0}$, and $\nu^\star>0$, we obtain

$$\operatorname{rank}\left(\nu^\star\boldsymbol{\Psi}+\sum_{k=1}^{n_{\mathrm{s}}}\lambda_k^\star\mathbf{D}_k\right)=n_{\mathrm{s}}. \tag{59}$$

Hence, from the definition of $\boldsymbol{Y}^\star$ in (54), it follows that

$$\operatorname{rank}(\boldsymbol{Y}^\star)\geq n_{\mathrm{s}}-\operatorname{rank}(\boldsymbol{\Sigma})=n_{\mathrm{s}}-1. \tag{60}$$

From the condition (57), $\bar{\boldsymbol{Q}}^\star$ must lie in the null space of $\boldsymbol{Y}^\star$. Therefore, $\operatorname{rank}\left(\bar{\boldsymbol{Q}}^\star\right)\leq\dim(\operatorname{Null}(\boldsymbol{Y}^\star))$, which is upper-bounded by one due to (60). Now, assume that $\operatorname{rank}(\boldsymbol{Y}^\star)=n_{\mathrm{s}}$. Then from (57), it follows that $\bar{\boldsymbol{Q}}^\star=\mathbf{0}$, which cannot be the optimal solution to problem (SDR1). In consequence, we must have $\operatorname{rank}(\boldsymbol{Y}^\star)=n_{\mathrm{s}}-1$ and thus $\operatorname{rank}\left(\bar{\boldsymbol{Q}}^\star\right)=1$, which completes the proof of Theorem 1. □

Appendix B
Proof of Theorem 3

Proof: We derive the result in (44) by showing that the optimal MSE in (P1–Sum) and the optimal power in (P2–Sum) are the inverse of each other. We start the proof by introducing the following lemma.

Lemma 1: For a given $\boldsymbol{v}$ in (12), let $(\boldsymbol{a}_1^\star,\boldsymbol{w}_1^\star)$ and $(\boldsymbol{a}_2^\star,\boldsymbol{w}_2^\star)$ be the optimal solutions to problems (P1–Sum) and (P2–Sum), respectively. We also let $f_1(\cdot,\cdot)$ and $f_2(\cdot,\cdot)$ be the objective functions in (P1–Sum) and (P2–Sum), respectively. Then, they obey the property: if $\gamma=f_1(\boldsymbol{a}_1^\star,\boldsymbol{w}_1^\star)$, then $(\boldsymbol{a}_2^\star,\boldsymbol{w}_2^\star):=(\boldsymbol{a}_1^\star,\boldsymbol{w}_1^\star)$; and if $P=f_2(\boldsymbol{a}_2^\star,\boldsymbol{w}_2^\star)$, then $(\boldsymbol{a}_1^\star,\boldsymbol{w}_1^\star):=(\boldsymbol{a}_2^\star,\boldsymbol{w}_2^\star)$.

Proof: First, it is worth noting that all the inequality constraints in (P1–Sum) and (P2–Sum) are satisfied with equality at the optimal solutions. For a given $\boldsymbol{v}$, we have $\|\boldsymbol{w}_1^\star\|^2=P$. We will prove that $(\boldsymbol{a}_1^\star,\boldsymbol{w}_1^\star)$ is also a solution to (P2–Sum), i.e., $(\boldsymbol{a}_2^\star,\boldsymbol{w}_2^\star):=(\boldsymbol{a}_1^\star,\boldsymbol{w}_1^\star)$. We prove it by contradiction. Assume that $(\boldsymbol{a}_1^\star,\boldsymbol{w}_1^\star)$ is not a solution to (P2–Sum), that is, there exists a feasible solution $(\boldsymbol{a}_1',\boldsymbol{w}_1')$ to (P2–Sum) such that $\|\boldsymbol{w}_1'\|<\|\boldsymbol{w}_1^\star\|$. In other words, we can find a constant $c>1$ such that

$$\|\boldsymbol{w}_1'\|<c\|\boldsymbol{w}_1'\|\leq\|\boldsymbol{w}_1^\star\|. \tag{61}$$

Since the objective function in (P1–Sum) is monotonically increasing with the norm of $\boldsymbol{a}$, it follows that $f_1(c\boldsymbol{a}_1',c\boldsymbol{w}_1')>f_1(\boldsymbol{a}_1',\boldsymbol{w}_1')$. Since $(\boldsymbol{a}_1',\boldsymbol{w}_1')$ is feasible to (P1–Sum), we also have that $f_1(\boldsymbol{a}_1',\boldsymbol{w}_1')\geq f_1(\boldsymbol{a}_1^\star,\boldsymbol{w}_1^\star)$. Thus, we obtain

$$f_1(c\boldsymbol{a}_1',c\boldsymbol{w}_1')>f_1(\boldsymbol{a}_1',\boldsymbol{w}_1')\geq f_1(\boldsymbol{a}_1^\star,\boldsymbol{w}_1^\star). \tag{62}$$

From (61) and the second constraint in (P1–Sum), the solution $(c\boldsymbol{a}_1',c\boldsymbol{w}_1')$ is also feasible to (P1–Sum), and yields a higher objective value than the optimal $(\boldsymbol{a}_1^\star, \boldsymbol{w}_1^\star)$ does. This contradicts to the assumption that $(\boldsymbol{a}_1^\star, \boldsymbol{w}_1^\star)$ is optimal to (P1–Sum), and thus $(\boldsymbol{a}_1^\star, \boldsymbol{w}_1^\star)$ must be a solution to (P2).

The proof for the second claim can be found using the similar steps to the proof for the first one, and hence is omitted here. □

From (34), the optimal mse satisfies (44). Since the optimal P and mse are the inverse of each other, it follows that optimal P also satisfies (34). Due to the fact that the above property holds for any $\boldsymbol{v}$, it holds for the optimal $\boldsymbol{v}$ as well, which completes the proof of Theorem 3. □

REFERENCES

[1] J. Yick, B. Mukherjee, and D. Ghosal, "Wireless sensor network survey," *Comput. Networks*, vol. 52, no. 12, pp. 2292–2330, Aug. 2008.

[2] R. Viswanathan and P. K. Varshney, "Distributed detection with multiple sensors: Part I–fundamentals," *Proc. IEEE*, vol. 85, no. 1, pp. 54–63, Jan. 1997.

[3] J. J. Xiao, S. Cui, Z. Q. Luo, and A. J. Goldsmith, "Linear coherent decentralized estimation," *IEEE Trans. Signal Process.*, vol. 56, no. 2, pp. 757–770, Feb. 2008.

[4] S. Kar and P. K. Varshney, "Linear coherent estimation with spatial collaboration," *IEEE Trans. Inf. Theory*, vol. 59, no. 6, pp. 3532–3553, Jun. 2013.

[5] A. S. Behbahani, A. M. Eltawil, and H. Jafarkhani, "Decentralized estimation under correlated noise," *IEEE Trans. Signal Process.*, vol. 62, no. 21, pp. 5603–5614, Nov. 2014.

[6] F. Jiang, J. Chen, and A. L. Swindlehurst, "Optimal power allocation for parameter tracking in a distributed amplify-and-forward sensor network," *IEEE Trans. Signal Process.*, vol. 62, no. 9, pp. 2200–2211, May 2014.

[7] V. V. Mai, Y. Jeong, and H. Shin, "Error exponents for distributed detection," *IEEE Commun. Lett.*, vol. 20, no. 1, pp. 121–124, Jan. 2016.

[8] R. Zhang and C. K. Ho, "MIMO broadcasting for simultaneous wireless information and power transfer," *IEEE Trans. Wireless Commun.*, vol. 12, no. 5, pp. 1989–2001, May 2013.

[9] 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. Signal Process.*, vol. 55, no. 9, pp. 4683–4695, Sep. 2007.

[10] H. Şenol and Tepedelenlioğlu, "Performance of distributed estimation over unknown parallel fading channels," *IEEE Trans. Signal Process.*, vol. 56, no. 12, pp. 6057–6068, Dec. 2008.

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

[12] J. Fang and H. Li, "Power constrained distributed estimation with correlated sensor data," *IEEE Trans. Signal Process.*, vol. 57, no. 8, pp. 3292–3297, Aug. 2009.

[13] J. Y. Wu and T. Y. Wang, "Power allocation for robust distributed best-linear-unbiased estimation against sensing noise variance uncertainty," *IEEE Trans. Wireless Commun.*, vol. 12, no. 6, pp. 2853–2869, Jun. 2013.

[14] M. Gastpar, "Uncoded transmission is exactly optimal for a simple Gaussian sensor networks," *IEEE Trans. Inf. Theory*, vol. 54, no. 11, pp. 5247–5251, Nov. 2008.

[15] M. K. Banavar, A. D. Smith, C. Tepedelenlioğlu, and A. Spanias, "On the effectiveness of multiple antennas in distributed detection over fading MACs," *IEEE Trans. Wireless Commun.*, vol. 11, no. 5, pp. 1744–1752, May 2012.

[16] I. Nevat, G. W. Peters, and I. B. Collings, "Distributed detection in sensor networks over fading channels with multiple antennas at the fusion centre," *IEEE Trans. Signal Process.*, vol. 62, no. 3, pp. 671–683, Feb. 2014.

[17] F. Jiang, J. Chen, A. L. Swindlehurst, and J. A. López-Salcedo, "Massive MIMO for wireless sensing with a coherent multiple access channel," *IEEE Trans. Signal Process.*, vol. 63, no. 12, pp. 3005–3017, Jun. 2015.

[18] A. Shirazinia, S. Dey, D. Ciuonzo, and P. Salvo Rossi, "Massive MIMO for decentralized estimation of a correlated source," *IEEE Trans. Signal Process.*, vol. 64, no. 10, pp. 2499–2512, May 2016.

[19] M. Nourian, S. Dey, and A. Ahlén, "Distortion minimization in multi-sensor estimation with energy harvesting," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 3, pp. 524–539, Mar. 2015.

[20] P. Grover and A. Sahai, "Shannon meets Tesla: Wireless information and power transfer," in *Proc. ISIT'10*, Austin, TX, Jun 2010, pp. 2363–2367.

[21] X. Zhou, R. Zhang, and C. K. Ho, "Wireless information and power transfer: Architecture design and rate-energy tradeoff," *IEEE Trans. Commun.*, vol. 61, no. 11, pp. 4754–4767, Nov. 2013.

[22] Q. Shi, L. Liu, W. Xu, and R. Zhang, "Joint transmit beamforming and receive power splitting for MISO SWIPT systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 6, pp. 3269–3280, Jun. 2014.

[23] J. Xu, L. Liu, and R. Zhang, "Multiuser MISO beamforming for simultaneous wireless information and power transfer," *IEEE Trans. Signal Process.*, vol. 62, no. 18, pp. 4798–4810, Sep. 2014.

[24] L. Liu, R. Zhang, and K. C. Chua, "Secrecy wireless information and power transfer with MISO beamforming," *IEEE Trans. Signal Process.*, vol. 62, no. 7, pp. 1850–1863, Apr. 2014.

[25] C. Shen, W. C. Li, and T. H. Chang, "Wireless information and energy transfer in multi-antenna interference channel," *IEEE Trans. Signal Process.*, vol. 62, no. 23, pp. 6249–6264, Dec. 2014.

[26] L. Xie, Y. Shi, Y. T. Hou, and A. Lou, "Wireless power transfer and applications to sensor networks," *IEEE Wireless Commun. Mag.*, vol. 20, no. 4, pp. 140–145, Aug. 2013.

[27] S. Kim, R. Vyas, J. Bito, K. Niotaki, A. Collado, A. Georgiadis, and M. M. Tentzeris, "Ambient RF energy-harvesting technologies for self-sustainable standalone wireless sensor platforms," *IEEE Trans. Signal Process.*, vol. 102, no. 11, pp. 1649–1666, Nov. 2014.

[28] G. Yang, W. P. Tay, and Y. L. Guan, "Optimal wireless power transfer and harvested power allocation for diffusion LMS in wireless sensor networks," in *Proc. GlobalSIP'15*, Orlando, FL, Dec 2015, pp. 1–5.

[29] T. C. Hsu and Y. W. P. Hong, "Wireless power transfer for distributed estimation in wireless passive sensor network," in *Proc. GlobalSIP'15*, Orlando, FL, Dec 2015, pp. 48–52.

[30] L. Liu, R. Zhang, and K. C. Chua, "Multi-antenna wireless powered communication with energy beamforming," *IEEE Trans. Commun.*, vol. 62, no. 12, pp. 4349–4361, Dec. 2014.

[31] T. M. Cover and J. Thomas, *Elements of Information Theory*, 2nd ed. New York, NY, USA: Wiley.

[32] O. Orhan, D. Gündüz, and E. Erkip, "Source-channel coding under energy, delay, and buffer constraints," *IEEE Trans. Wireless Commun.*, vol. 14, no. 7, pp. 3836–3849, Jul. 2015.

[33] M. Gastpar and M. Vetterli, *Source-channel communication in sensor networks*, ser. Lecture Notes in Computer Science. New York, NY, USA: Springer, 2006, vol. 2634.

[34] H. Ju and R. Zhang, "Throughput maximization in wireless powered communication networks," *IEEE Trans. Signal Process.*, vol. 13, no. 1, pp. 418–428, Jan. 2014.

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

[36] A. Charnes and W. W. Cooper, "Programming with linear fractional functionals," *Naval Res. Logist. Quarter.*, vol. 9, pp. 181–186, 1962.

[37] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1," http://cvxr.com/cvx, Mar. 2014.

[38] B.-T. Aharon and A. Nemirovski, *Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications*, ser. MOS-SIAM Series on Optimization. Philadelphia, PA: SIAM, 2001.

[39] R. Hunger, "Floating point operations in matrix-vector calculus," Technische Universitat Munchen, Munchen, Germany, Tech. Rep., 2007.

[40] K. Gomadam, V. R. Cadambe, and S. A. Jafar, "A distributed numerical approach to interference alignment and applications to wireless interference networks," *IEEE Trans. Inf. Theory*, vol. 57, no. 6, pp. 3309–3322, Jun. 2011.

[41] Q. Shi, M. Razaviyayn, Z. Q. Luo, and C. He, "An iteratively weighted MMSE approach to distributed sum-utility maximization for a MIMO interfering broadcast channel," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4331–4340, Sep. 2011.

[42] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

[43] R. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1985.

[44] S. Y. Seidel and T. S. Rappaport, "914 MHz path loss prediction models for indoor wireless communications in multifloored buildings," *IEEE Trans. Antennas and Propagation.*, vol. 40, no. 2, pp. 207–217, Feb. 1992.

[45] J. A. Gutierrez, M. Naeve, E. Callaway, M. Bourgeois, V. Mitter, and B. Heile, "IEEE 802.15.4: A Developing Standard for Low-Power, Low-Cost Wireless Personal Area Networks," *IEEE Network*, vol. 15, no. 5, Sept. 2001, pp. 12–19.