Linear Precoding to Optimize Throughput, Power Consumption and Energy Efficiency in MIMO Wireless Sensor Networks

Yang Liu and Jing Li

***Abstract*— This paper considers joint precoder design to optimize throughput, power consumption and energy efficiency (EE) in the context of multi-antenna wireless sensor networks with coherent multiple access channels. To maximize throughput, both centralized and decentralized algorithms are developed. Our centralized algorithm obtains a new second order cone programming formulation of the problem, which is different from related works and can apply to more generic system setup compared to existing literature. In addition, noting the fact that all existing solutions in literature are centralized based, we propose a novel decentralized solution and analyses its convergence. Besides the throughput maximization, the power consumption and EE problems are also attacked. To optimize these two metrics, a decentralized algorithm based on dual-decomposition and block successive upper-bound method has been developed, which runs in parallel with semi-analytical solutions and has provable strong convergence. A sufficient condition for the validity of the decentralized method is obtained. Extensive numerical results are presented to consolidate our findings.**

***Index Terms*— Wireless sensor network, throughput, energy efficiency (EE), power consumption, decentralized algorithm.**

I. Introduction

CONSIDER a multi-input multi-output (MIMO) wireless sensor network (WSN) deployed for surveillance or tracking purpose (as shown in Fig.1). When an event occurs, multiple sensors may have sensed it with each sensor's observation being distorted (due to, for example, thermal noise or environmental disturbance). There exists a fusion center (FC) in the network collecting the observations from sensors and performing data fusion. Each sensor may be equipped with a different number of antennas and provisioned with a different power supply. As introduced in [1] and [2], this flat single-tier network structure is a typical deployment scenario for wireless sensor networks and can also serve as a basic component from which a multi-tier network architecture can be constructed. Many challenging issues arise along with the pervasive applications of wireless sensor networks. One particular concern is how to jointly design the transmitting scheme to perform reliable, high-throughput and power efficient communication.

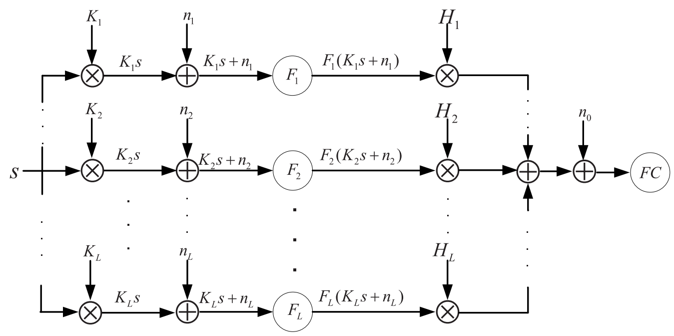

Fig. 1. A wireless sensor network system.

A. Related Works

For the model mentioned above, a good number of literature considers network design from different viewpoints.

One important line of research focuses on precoder design to perform effective signal estimation, which aims at minimizing mean square error (MSE) of the estimate at FC [3]–[9]. The early papers [3] study compressive precoding scheme under simplified channel/power assumptions (e.g. noise-free sensor-FC channels with no power constraints). A generic model of the MIMO WSN is introduced by [4], which considers additive noise, separate power constraints and fading orthogonal/coherent multiple access channels (MAC). Based on the general model in [4] and [4]–[6] have developed effective algorithms to optimize precoders for sensing with uncorrelated and correlated observation noise. Very recently, [7] and [8] have made progress in researching the optimal linear precoding schemes with FC equipped with very large arrays of antennas (i.e. massive MIMO).

Another critical performance metric is the throughput of the network. A most pervasive deployment scenario of wireless sensor networks is sensing task of image/video information, e.g. wireless multimedia sensor networks (MWSN) [1] (also known as visual sensing networks (VSN) [2]). One remarkable characteristic of VSN/WMSN is the great amount of information that is collected and a arising challenge is how

Manuscript received August 27, 2017; revised December 22, 2017; accepted January 5, 2018. Date of publication January 18, 2018; date of current version May 15, 2018. This work was supported by National Science Foundation under Grant 0928092, Grant 1133027, and Grant 1343372. The associate editor coordinating the review of this paper and approving it for publication was Y. Wu. *(Corresponding author: Yang Liu.)*

The authors are with the Electrical and Computer Engineering Department, Lehigh University, Bethlehem, PA 18015 USA (e-mail: liuocean613@gmail.com; jingli@ece.lehigh.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCOMM.2018.2794363

to effectively transport the information to the fusion center. One solution is to perform data compression. For instance the compressive sensing technique [12] has been widely adopted in image processing. In some specific applications, extracted feature instead of the pixels are transmitted to decrease the demanding bandwidth [13]. Another practical solution is to improve the throughput of the network. As recommended in [1], the multi-antenna technology is a promising method to satisfy the high data rate requirement of WMSN. In recent years, linear precoders/beamformers have been widely used in multi-antenna communication systems to boost sum-rate [19]–[23]. One recent note-worthy work [10] considers the sum-rate maximization of the WSN with orthogonal MAC. Compared to the orthogonal MAC, coherent MAC is a more appealing access scheme due to its high time/frequency efficiency, especially for large networks with huge number of sensors. Hence the throughput maximization for wireless sensor network with coherent MAC is a meaningful issue worth study, which will be one focus of this article.

Besides the network throughput, power consumption is also a critical consideration for network design. Reducing the power consumption can effectively increase the battery life. Furthermore, combining with the emerging technology of energy harvesting [11], it is possible to deploy rechargeable wireless sensor networks, where battery change will become unnecessary. The power consumption minimization problem is considered in [6]–[8] with MSE constraint. Besides another closed related metric—the energy efficiency (EE) of the system has attracted substantial attention recently [15]–[18]. To relieve energy shortage and greenhouse effects, the EE has become a key performance metric for the next generation cellular networks [14]. Therefore to optimize the power consumption as well as EE while maintaining the throughput above a predefined level will also be considered in this article.

B. Contributions

This article focuses on the MIMO wireless sensor networks equipped with linear precoders. Our goal is to design the linear precoders to optimize i) throughput, ii) power consumption and iii) EE. Here we adopt the Gaussian signaling assumptions [10], [20]–[24], i.e. the signal and noise all follow Gaussian distribution. The contributions of this article is specified as follows:

(i) For throughput maximization, we design a centralized algorithm based on second order cone programming (SOCP) technique. Compared to the few existing works considering the batch-mode method [4] and [19], our SOCP formulation has no special restrictions on the dimensions/ranks of the system setup and can be used in wider range of applications. Besides we also design a decentralized algorithm, which has parallel and semi-analytical update and provable strong convergence. To the best of our knowledge, decentralized algorithm has never been developed in the context of wireless sensor networks [3]–[8], [10], [20]–[23].

(ii) For power consumption minimization, an SOCP-based centralized algorithm is developed. Furthermore inspired by the seminal dual-decomposition method [33] and block successive upper-bound method (BSUM) [25], a decentralized algorithm is designed, which has parallel semi-analytically solvable subproblems. A sufficient condition for validity of the decentralized algorithm is obtained. It can be proved that both the centralized and decentralized algorithms guarantee the limit points of their solutions are stationary points.

(iii) Noticing the fact that decentralized solution is still missing in existing works on EE-maximization beamforming [15]–[18], we also design a decentralized implementation for EE-maximization. The distinctions between the EE-maximization and power-minimization beamformers are studied by numerical experiments.

The rest of the paper is organized as follows. Section II introduces the system model. Section III, IV and V discuss the beamforming design to optimize throughput, power consumption and EE respectively. Section VI compares the centralized and decentralized algorithms and Section VIII concludes the article.

II. System Model

A MIMO wireless sensor network with L sensors is illustrated in Fig.1. It is assumed that each sensor independently obtains noisy observation of a common target source vector $\mathbf{s}$ and reports to the fusion center. Following the previous discussion we assume that source signal $\mathbf{s}$ is a circularly symmetric complex Gaussian random vector of dimension K, i.e. $\mathbf{s} \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{\Sigma}_\mathbf{s})$ with the covariance matrix $\boldsymbol{\Sigma}_\mathbf{s}$ being positive definite.

The observation obtained at the i-th sensor is modeled as $\mathbf{K}_i\mathbf{s}+\mathbf{n}_i$, where $\mathbf{K}_i \in \mathbb{C}^{J_i \times K}$ is the observation filter, which is a known constant matrix. In reality, the observation filter can be the compressive sampling matrix in WMSN. $\mathbf{n}_i \in \mathbb{C}^{J_i \times 1}$ is the observation noise. Since sensors are usually geometrically distributed, it is reasonable to assume that $\mathbf{n}_i$'s are mutually uncorrelated.

It is assumed that the i-th sensor and the FC has N_i and M antennas respectively. Each sensor employs a linear precoder $\mathbf{F}_i \in \mathbb{C}^{N_i \times J_i}$ to encode its observation before sending it into the channel. Denote $\mathbf{H}_i \in \mathbb{C}^{M \times N_i}$ as the fading channel between the i-th sensor and the fusion center. In this article we assume that $\{\mathbf{H}_i\}_{i=1}^L$ is known (e.g. via standard channel estimation techniques). The additive circularly symmetric complex Gaussian noise at the fusion center is denoted as $\mathbf{n}_0 \in \mathbb{C}^{M\times 1}$. Without loss of generality, we assume $\mathbf{n}_0 \sim \mathcal{CN}(\mathbf{0}, \sigma_0^2\mathbf{I}_M)$. The received signal at the fusion center is therefore given by:

$$\mathbf{r} = \Big(\sum_{i=1}^{L}\mathbf{H}_i\mathbf{F}_i\mathbf{K}_i\Big)\mathbf{s} + \mathbf{n} \tag{1}$$

where the overall noise $\mathbf{n} \triangleq \sum_{i=1}^{L}\mathbf{H}_i\mathbf{F}_i\mathbf{n}_i + \mathbf{n}_0$ is still Gaussian: $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{\Sigma}_\mathbf{n})$ whose covariance matrix $\boldsymbol{\Sigma}_\mathbf{n}$ is

$$\boldsymbol{\Sigma}_\mathbf{n} = \sigma_0^2\mathbf{I}_M + \sum_{i=1}^{L}\mathbf{H}_i\mathbf{F}_i\boldsymbol{\Sigma}_i\mathbf{F}_i^H\mathbf{H}_i^H. \tag{2}$$

A. Problem Formulation

For the WSN introduced above, several critical metrics will be considered to depict the system's performance from different aspects.

The throughput of the above WSN is given by the sum-rate function $\mathsf{SR}(\cdot)$ defined in (3) shown in the next page [24]. The total power consumption of the above WSN system can be modelled as [18]

$$\mathsf{Pwr}\big(\{\mathbf{F}_i\}_{i=1}^L\big) = \mathsf{P_T}\big(\{\mathbf{F}_i\}_{i=1}^L\big) + P_0 \tag{4}$$

$$= \kappa \sum_{i=1}^{L} \mathsf{Tr}\big\{\mathbf{F}_i\big(\mathbf{K}_i\boldsymbol{\Sigma}_\mathbf{s}\mathbf{K}_i^H + \boldsymbol{\Sigma}_i\big)\mathbf{F}_i^H\big\} + P_0. \tag{5}$$

where κ is the inverse of power amplifier efficiency ($\kappa^{-1} \in (0,1)$), $\mathsf{P_T}(\cdot)$ is the average total transmit power and P_0 is a constant representing the circuit power.

Besides the throughput and power consumption, the energy efficiency will also be studied, which is defined as follows

$$\mathsf{EE}\big(\{\mathbf{F}_i\}_{i=1}^L\big) = \frac{\mathsf{SR}\big(\{\mathbf{F}_i\}_{i=1}^L\big)}{\mathsf{Pwr}\big(\{\mathbf{F}_i\}_{i=1}^L\big)}. \tag{6}$$

In this article we consider beamforming design to optimize the above metrics. One problem is the throughput maximization under limited power supplies, which is formulated as the following problem:

$$(\mathsf{P0}): \ \max_{\{\mathbf{F}_i\}_{i=1}^L}. \ \mathsf{SR}\big(\{\mathbf{F}_i\}_{i=1}^L\big), \tag{7a}$$

$$\text{s.t. } \mathsf{Tr}\big\{\mathbf{F}_i\big(\mathbf{K}_i\boldsymbol{\Sigma}_\mathbf{s}\mathbf{K}_i^H + \boldsymbol{\Sigma}_i\big)\mathbf{F}_i^H\big\} \le P_i, \ i \in \{1, \cdots, L\}. \tag{7b}$$

where P_i is the power supply limit for the i-th sensor.

Another meaningful problem is to minimize the power consumption while maintaining the network throughput above some predefined level r_0, as presented in ($\mathsf{P1}$).

$$(\mathsf{P1}): \ \min_{\{\mathbf{F}_i\}_{i=1}^L}. \ \mathsf{Pwr}\big(\{\mathbf{F}_i\}_{i=1}^L\big), \tag{8a}$$

$$\text{s.t. } \mathsf{SR}\big(\{\mathbf{F}_i\}_{i=1}^L\big) \ge r_0, \tag{8b}$$

$$\mathsf{Tr}\big\{\mathbf{F}_i\big(\mathbf{K}_i\boldsymbol{\Sigma}_\mathbf{s}\mathbf{K}_i^H + \boldsymbol{\Sigma}_i\big)\mathbf{F}_i^H\big\} \le P_i, \ i \in \{1, \cdots, L\}. \tag{8c}$$

In fact the optimizations of throughput and power minimization deviate from each other (especially in low channel SNR scenarios). In contrast, the EE maximizing beamformers can often obtain a reasonable balance between the power and throughput, which can be obtained by solving the following problem

$$(\mathsf{P2}): \ \min_{\{\mathbf{F}_i\}_{i=1}^L}. \ \mathsf{EE}\big(\{\mathbf{F}_i\}_{i=1}^L\big), \tag{9a}$$

$$\text{s.t. } \mathsf{SR}\big(\{\mathbf{F}_i\}_{i=1}^L\big) \ge r_0, \tag{9b}$$

$$\mathsf{Tr}\big\{\mathbf{F}_i\big(\mathbf{K}_i\boldsymbol{\Sigma}_\mathbf{s}\mathbf{K}_i^H + \boldsymbol{\Sigma}_i\big)\mathbf{F}_i^H\big\} \le P_i, \ i \in \{1, \cdots, L\}. \tag{9c}$$

All the above problems will be tackled in the subsequent sections.

III. Throughput Maximization

A. Centralized Method

For problem ($\mathsf{P0}$) we adopt block coordinate ascent (BCA) method. Inspired by the weighted mean square error (WMMSE) method in [22] and [23], we introduce auxiliary variables to convert the $\mathsf{SR}(\cdot)$ function into a BCA-friendly form.

According to [22] and [23], the following facts stand:

(R1) For any $n \times n$ positive definite matrix $\mathbf{E}$, we have $-\log\det(\mathbf{E}) = \max_{\mathbf{W}\succcurlyeq 0}.\{\ \log\det\big(\mathbf{W}\big) - \mathsf{Tr}\{\mathbf{W}\mathbf{E}\} + n\}$, with the optimal $\mathbf{W}^\star$ given as $\mathbf{W}^\star = \mathbf{E}^{-1}$.

(R2) Assume that $\boldsymbol{\Sigma}_\mathbf{s}$, $\boldsymbol{\Sigma}_\mathbf{n}$ and $\mathbf{W}$ are given positive definite matrices and the function $\mathbf{E}\big(\mathbf{G}\big) \triangleq \big(\mathbf{I} - \mathbf{G}^H\mathbf{H}\big)\boldsymbol{\Sigma}_\mathbf{s}\big(\mathbf{I} - \mathbf{G}^H\mathbf{H}\big)^H + \mathbf{G}^H\boldsymbol{\Sigma}_\mathbf{n}\mathbf{G}$. Then the optimization problem $\min_{.\mathbf{G}} \mathsf{Tr}\big\{\mathbf{W}\mathbf{E}(\mathbf{G})\big\}$ is solved by the optimal solution $\mathbf{G}^\star = \big(\mathbf{H}\boldsymbol{\Sigma}_\mathbf{s}\mathbf{H}^H + \boldsymbol{\Sigma}_\mathbf{n}\big)^{-1}\mathbf{H}\boldsymbol{\Sigma}_\mathbf{s}$. Its optimal value is $\mathbf{E}(\mathbf{G}^\star) = \big(\mathbf{H}^H\boldsymbol{\Sigma}_\mathbf{n}^{-1}\mathbf{H} + \boldsymbol{\Sigma}_\mathbf{s}^{-1}\big)^{-1}$.

The detailed proof for the above results are referred to [22, eq. (21)] or [23, Th. 1].

By utilizing the notation in (2) and defining $\widetilde{\mathbf{H}} \triangleq \sum_{i=1}^L \mathbf{H}_i\mathbf{F}_i\mathbf{K}_i$, the $\mathsf{SR}(\cdot)$ function can be transformed as:

$$\mathsf{SR}\Big(\{\mathbf{F}_i\}_{i=1}^L\Big)$$

$$= \log\det\Big(\mathbf{I}_M + \widetilde{\mathbf{H}}\boldsymbol{\Sigma}_\mathbf{s}\widetilde{\mathbf{H}}^H\boldsymbol{\Sigma}_\mathbf{n}^{-1}\Big) \tag{10}$$

$$= -\log\det\Big(\widetilde{\mathbf{H}}^H\boldsymbol{\Sigma}_\mathbf{n}^{-1}\widetilde{\mathbf{H}} + \boldsymbol{\Sigma}_\mathbf{s}^{-1}\Big)^{-1} + \log\det\big(\boldsymbol{\Sigma}_\mathbf{s}\big) \tag{11}$$

$$= \max_{\mathbf{W}\succcurlyeq 0}.\Big\{\log\det\big(\mathbf{W}\big) - \mathsf{Tr}\big\{\mathbf{W}\big(\widetilde{\mathbf{H}}^H\boldsymbol{\Sigma}_\mathbf{n}^{-1}\widetilde{\mathbf{H}} + \boldsymbol{\Sigma}_\mathbf{s}^{-1}\big)^{-1}\big\} + K\Big\} + \log\det\big(\boldsymbol{\Sigma}_\mathbf{s}\big)$$

$$= \max_{\substack{\mathbf{W}\succcurlyeq 0,\\ \mathbf{G}}}.\Big\{\log\det\big(\mathbf{W}\big) - \mathsf{Tr} \times \Big\{\mathbf{W}\big[\big(\mathbf{I} - \mathbf{G}^H\widetilde{\mathbf{H}}\big)\boldsymbol{\Sigma}_\mathbf{s}\big(\mathbf{I} - \mathbf{G}^H\widetilde{\mathbf{H}}\big)^H + \mathbf{G}^H\boldsymbol{\Sigma}_\mathbf{n}\mathbf{G}\big]\Big\}\Big\} + K + \log\det\big(\boldsymbol{\Sigma}_\mathbf{s}\big), \tag{12}$$

where the last two steps follow (R1) and (R2) stated above.

We denote the augmented sum-rate function after introducing additional variables inside the maximization operation in (12) as $\widetilde{\mathsf{SR}}$. Thus the original sum-rate maximization problem ($\mathsf{P0}$) is transformed into ($\mathsf{P3}$) in (13) shown on the bottom of the next page.

$$(\mathsf{P3}) \ \max_{\mathbf{W}\succcurlyeq 0, \{\mathbf{F}_i\}_{i=1}^L, \mathbf{G}}. \ \widetilde{\mathsf{SR}}\Big(\{\mathbf{F}_i\}_{i=1}^L, \mathbf{W}, \mathbf{G}\Big) \tag{13a}$$

$$\text{s.t. } \mathsf{Tr}\big\{\mathbf{F}_i\big(\mathbf{K}_i\boldsymbol{\Sigma}_\mathbf{s}\mathbf{K}_i^H + \boldsymbol{\Sigma}_\mathbf{n}\big)\mathbf{F}_i^H\big\} \le P_i, \ i \in \{1, \cdots, L\}. \tag{13b}$$

To solve ($\mathsf{P3}$) we alternatively update $\{\mathbf{F}_i\}_i^L$, $\mathbf{W}$ and $\mathbf{G}$. The results in (R1) and (R2) readily provide the analytical update of the two latter variables.

(i) When $\{\mathbf{F}_i\}_{i=1}^L$ and $\mathbf{G}$ are given, the optimal $\mathbf{W}^\star$ maximizing $\widetilde{\mathsf{SR}}$ is given by

$$\begin{aligned}\mathbf{W}^\star &= \arg \max_{\mathbf{W}\in\mathcal{H}_{++}^K}. \widetilde{\mathsf{SR}}\Big(\mathbf{W}\Big|\{\mathbf{F}_i\}_{i=1}^L, \mathbf{G}\Big)\\ &= \Big[\mathbf{G}^H\boldsymbol{\Sigma}_\mathbf{n}\mathbf{G} + \Big(\mathbf{I} - \mathbf{G}^H\big(\sum_{i=1}^L \mathbf{H}_i\mathbf{F}_i\mathbf{K}_i\big)\Big)\\ &\qquad \times \boldsymbol{\Sigma}_\mathbf{s}\Big(\mathbf{I} - \mathbf{G}^H\big(\sum_{i=1}^L \mathbf{H}_i\mathbf{F}_i\mathbf{K}_i\big)\Big)^H\Big]^{-1}. \end{aligned} \tag{14}$$

(ii) When $\{\mathbf{F}_i\}_{i=1}^L$ and $\mathbf{W}$ are given, the optimal $\mathbf{G}^\star$ that achieves the maximum of $\widetilde{\mathsf{SR}}$ in (13) is given by

$$\begin{aligned}\mathbf{G}^\star &= \arg\max_{\mathbf{G}}. \widetilde{\mathsf{SR}}\Big(\mathbf{G}\Big|\{\mathbf{F}_i\}_{i=1}^L, \mathbf{W}\Big)\\ &= \Big[\big(\sum_{i=1}^L \mathbf{H}_i\mathbf{F}_i\mathbf{K}_i\big)\boldsymbol{\Sigma}_\mathbf{s}\big(\sum_{i=1}^L \mathbf{H}_i\mathbf{F}_i\mathbf{K}_i\big)^H + \boldsymbol{\Sigma}_\mathbf{n}\Big]^{-1}\\ &\qquad \times \Big(\sum_{i=1}^L \mathbf{H}_i\mathbf{F}_i\mathbf{K}_i\Big)\boldsymbol{\Sigma}_\mathbf{s}, \end{aligned} \tag{15}$$

where $\boldsymbol{\Sigma}_\mathbf{n}$ is given in (2).

We now proceed to optimize $\{\mathbf{F}_i\}_{i=1}^L$ when $\mathbf{W}$ and $\mathbf{G}$ are given. First we introduce the notations $\mathbf{f}_i \triangleq vec\big(\mathbf{F}_i\big)$, $\mathbf{g} \triangleq vec\big(\mathbf{G}\big)$, $\mathbf{A}_{ij} \triangleq \big(\mathbf{K}_i^*\boldsymbol{\Sigma}_\mathbf{s}^*\mathbf{K}_j^T\big) \otimes \Big(\mathbf{H}_i^H\mathbf{G}\mathbf{W}\mathbf{G}^H\mathbf{H}_j\Big)$, $\mathbf{B}_i \triangleq \big(\mathbf{W}^*\boldsymbol{\Sigma}_\mathbf{s}^*\mathbf{K}_i^T\big) \otimes \mathbf{H}_i$, $\mathbf{C}_i \triangleq \boldsymbol{\Sigma}_i^* \otimes \Big(\mathbf{H}_i^H\mathbf{G}\mathbf{W}\mathbf{G}^H\mathbf{H}_i\Big)$, $\mathbf{f} \triangleq \big[\mathbf{f}_1^T, \cdots, \mathbf{f}_L^T\big]^T$, $\mathbf{A} \triangleq \big[\mathbf{A}_{ij}\big]_{i,j=1}^L$, $\mathbf{B} \triangleq \big[\mathbf{B}_1, \cdots, \mathbf{B}_L\big]$, $\mathbf{C} \triangleq \mathsf{Diag}\big\{\mathbf{C}_1, \cdots, \mathbf{C}_L\big\}$, $\mathbf{E}_i \triangleq \big(\mathbf{K}_i\boldsymbol{\Sigma}_\mathbf{s}\mathbf{K}_i^H + \boldsymbol{\Sigma}_i\big)^T \otimes \mathbf{I}_{N_i}$, $c \triangleq \mathsf{Tr}\big\{\mathbf{W}\boldsymbol{\Sigma}_\mathbf{s}\big\} + \sigma_0^2\mathsf{Tr}\big\{\mathbf{G}\mathbf{W}\mathbf{G}^H\big\}$ and $\mathbf{D}_i \triangleq \mathsf{Diag}\Big\{\mathbf{O}_{\sum_{j=1}^{i-1}K_jN_j}, \mathbf{E}_i, \mathbf{O}_{\sum_{j=i+1}^{L}K_jN_j}\Big\}$. By the identity $\mathsf{Tr}\{\mathbf{ABCD}\} = vec^T\big(\mathbf{D}^T\big)\big[\mathbf{C}^T \otimes \mathbf{A}\big]vec\big(\mathbf{B}\big)$, we rewrite $(\mathsf{P}4)$ with respect to $\{\mathbf{F}_i\}_{i=1}^L$ into a compact form as:

$$(\mathsf{P}4): \ \min_{\mathbf{f}} \ \mathbf{f}^H\big(\mathbf{A} + \mathbf{C}\big)\mathbf{f} - 2\mathsf{Re}\big\{\mathbf{g}^H\mathbf{B}\mathbf{f}\big\} + c, \tag{16a}$$

$$s.t. \ \mathbf{f}^H\mathbf{D}_i\mathbf{f} \le P_i, \quad i \in \{1, \cdots, L\}. \tag{16b}$$

By writing in an epigraph form and introducing a slack variable s, the problem $(\mathsf{P}4)$ can be formulated into a standard SOCP problem as:

$$(\mathsf{P}4_{SOCP}): \ \min_{\mathbf{f},t,s}. \ t, \tag{17a}$$

$$\text{s.t. } s - 2\mathsf{Re}\{\mathbf{g}^H\mathbf{B}\mathbf{f}\} + c \le t; \tag{17b}$$

$$\left\| \begin{matrix} (\mathbf{A}+\mathbf{C})^{\frac{1}{2}}\mathbf{f} \\ \frac{s-1}{2} \end{matrix} \right\|_2 \le \frac{s+1}{2}; \tag{17c}$$

$$\left\| \begin{matrix} \mathbf{D}_i^{\frac{1}{2}}\mathbf{f} \\ \frac{P_i-1}{2} \end{matrix} \right\|_2 \le \frac{P_i+1}{2}, \ i \in \{1, \cdots, L\}. \tag{17d}$$

SOCP problem can then be conveniently solved by standard numerical tools like CVX [27]. The entire approach is summarized in Algorithm 1.

Algorithm 1 SOCP-Based Centralized Algorithm

Initialization: randomly generate feasible $\{\mathbf{F}_i^{(0)}\}_{i=1}^L$; obtain $\mathbf{G}^{(0)}$ by (15); obtain $\mathbf{W}^{(0)}$ by (14);
repeat
  fix $\mathbf{G}^{(j-1)}$ and $\mathbf{W}^{(j-1)}$, solve $(\mathsf{P}4)$, obtain $\{\mathbf{F}_i^{(j)}\}_{i=1}^L$;
  fix $\{\mathbf{F}_i^{(j)}\}_{i=1}^L$ and $\mathbf{W}^{(j-1)}$, obtain $\mathbf{G}^{(j)}$ by (15);
  fix $\{\mathbf{F}_i^{(j)}\}_{i=1}^L$ and $\mathbf{G}^{(j)}$, obtain $\mathbf{W}^{(j)}$ by (14);
until *convergence*;

Remark 1 [19]: obtains an SOCP formulation which requires $\big(\mathbf{A}+\mathbf{C}\big)$ and $\mathbf{D}_i$'s to be positive definite (invertible). However, the $\mathbf{D}_i$'s are in fact always non-invertible in our system setup and therefore the SOCP formulation in [19] is inapplicable to our problem. Comparatively our formulation only requires $\mathbf{D}_i$'s and $\big(\mathbf{A}+\mathbf{C}\big)$ to be positive semidefinite and can tackle a larger scope of scenarios.

Following similar arguments in [23], it can be proved that any limit point of the solution sequence generated by Alg.1 is a stationary point of the original problem $(\mathsf{P}0)$.

B. Decentralized Method

In order to deploy large scale wireless sensor networks having great number of sensors, decentralized and, hopefully, closed-form solutions are highly desirable. Note that the objective in $(\mathsf{P}4)$ has all the precoders $\mathbf{F}_i$'s coupled and thus is hard to perform parallel computing directly. To enable distributive processing, inspired by the classical block successive upper-bound method (BSUM) proposed in [25], we introduce a modified objective function $h(\mathbf{f};\hat{\mathbf{f}}|\mathbf{G},\mathbf{W})$ defined as follows,

$$\begin{aligned} &h(\mathbf{f};\hat{\mathbf{f}}|\mathbf{G},\mathbf{W})\\ &\quad \triangleq \mathbf{f}^H\mathbf{C}\mathbf{f} - 2\mathsf{Re}\{\mathbf{g}^H\mathbf{B}\mathbf{f}\} + c\\ &\qquad + \lambda_{\max}(\mathbf{A})\big\|\mathbf{f}-\hat{\mathbf{f}}\big\|_2^2 + 2\mathsf{Re}\{\hat{\mathbf{f}}^H\mathbf{A}\big(\mathbf{f}-\hat{\mathbf{f}}\big)\} + \hat{\mathbf{f}}^H\mathbf{A}\hat{\mathbf{f}} \end{aligned} \tag{18}$$

with $\hat{\mathbf{f}}$ being some given feasible solution and $\lambda_{\max}(\mathbf{A})$ denoting the maximal eigenvalue of matrix $\mathbf{A}$. Compared to the original objective in $(\mathsf{P}4)$, the modified objective $h(\mathbf{f};\hat{\mathbf{f}}|\mathbf{G},\mathbf{W})$ has the following the properties

(T1) $h(\mathbf{f};\hat{\mathbf{f}}|\mathbf{G},\mathbf{W}) \ge -\widetilde{\mathsf{SR}}(\{\mathbf{F}_i\}_{i=1}^L|\mathbf{G},\mathbf{W})$, for any $\mathbf{f}$ and $\hat{\mathbf{f}}$. This can be seen by noting that $h(\mathbf{f};\hat{\mathbf{f}}|\mathbf{G},\mathbf{W})$ is obtained from $-\widetilde{\mathsf{SR}}(\{\mathbf{F}_i\}_{i=1}^L|\mathbf{G},\mathbf{W})$ via replacing the term $\mathbf{f}^H\mathbf{A}\mathbf{f}$ by its second-order Taylor expansion $(\mathbf{f}-\hat{\mathbf{f}})^H\mathbf{A}(\mathbf{f}-\hat{\mathbf{f}}) + 2\mathsf{Re}\{\hat{\mathbf{f}}^H\mathbf{A}(\mathbf{f}-\hat{\mathbf{f}})\} + \hat{\mathbf{f}}^H\mathbf{A}\hat{\mathbf{f}}$ and then upper-bounding $(\mathbf{f}-\hat{\mathbf{f}})^H\mathbf{A}(\mathbf{f}-\hat{\mathbf{f}})$ by $\lambda_{\max}(\mathbf{A})\big\|\mathbf{f}-\hat{\mathbf{f}}\big\|_2^2$;

(T2) $h(\hat{\mathbf{f}};\hat{\mathbf{f}}|\mathbf{G},\mathbf{W}) = -\widetilde{\mathsf{SR}}(\{\hat{\mathbf{F}}_i\}_{i=1}^L|\mathbf{G},\mathbf{W})$. This is readily verified by substituting $\mathbf{f}$ with $\hat{\mathbf{f}}$;

$$\mathsf{SR}\Big(\{\mathbf{F}_i\}_{i=1}^L\Big) = \log\det\bigg\{\mathbf{I}_M + \Big(\sum_{i=1}^L \mathbf{H}_i\mathbf{F}_i\mathbf{K}_i\Big)\boldsymbol{\Sigma}_\mathbf{s}\Big(\sum_{i=1}^L \mathbf{H}_i\mathbf{F}_i\mathbf{K}_i\Big)^H\Big(\sigma_0^2\mathbf{I} + \sum_{i=1}^L \mathbf{H}_i\mathbf{F}_i\boldsymbol{\Sigma}_i\mathbf{F}_i^H\mathbf{H}_i^H\Big)^{-1}\bigg\} \tag{3}$$

(T3) $h(\mathbf{f};\hat{\mathbf{f}}|\mathbf{G},\mathbf{W})$ has decoupled variables. In fact, by defining $\mathbf{p}_i \triangleq [\mathbf{A}\hat{\mathbf{f}}]_{\sum_{j=1}^{i-1}(J_jN_j)+1:\sum_{j=1}^{i}(J_jN_j)}$ and $\mathbf{t}_i \triangleq \big[\mathbf{B}_i^H\mathbf{g} + \lambda_{\max}(\mathbf{A})\hat{\mathbf{f}}_i - \mathbf{p}_i\big]$, the modified objective $h(\mathbf{f};\hat{\mathbf{f}}|\mathbf{G},\mathbf{W})$ can be decomposed as $\sum_{i=1}^{L}\big(\mathbf{f}_i^H\mathbf{C}_i\mathbf{f}_i + \lambda_{\max}(\mathbf{A})\|\mathbf{f}_i\|_2^2 - 2\mathsf{Re}\big\{\mathbf{t}_i^H\mathbf{f}_i\big\}\big) + c_2$ with c_2 being the constant terms independent of $\mathbf{f}$.

Instead of solving the problem ($\mathsf{P4}$), we optimize its modified objective $h(\mathbf{f};\hat{\mathbf{f}}|\mathbf{G},\mathbf{W})$. Due to the property (T3) above, optimizing the modified objective $h(\mathbf{f};\hat{\mathbf{f}}|\mathbf{G},\mathbf{W})$ is equivalent to solving L separate smaller problems $\{(\mathsf{P4}_i)\}_{i=1}^{L}$, formulated as below

$$(\mathsf{P4}_i): \ \min_{\mathbf{f}_i}.\ \mathbf{f}_i^H\big(\lambda_{\max}(\mathbf{A})\mathbf{I}_{KN_i} + \mathbf{C}_i\big)\mathbf{f}_i - 2\mathsf{Re}\big\{\mathbf{t}_i^H\mathbf{f}_i\big\}, \tag{19a}$$

$$\text{s.t. } \mathbf{f}_i^H\mathbf{E}_i\mathbf{f}_i \le P_i, \tag{19b}$$

which can be implemented in an distributive and parallel manner among L processing units.

The above problem ($\mathsf{P4}_i$) is actually equivalent to a convex trust-region subproblem. In fact for a convex trust-region subproblem we have the following semi-analytical solution:

Theorem 1 (Convex Trust Region Problem): Consider the following convex trust region subproblem:

$$(\mathsf{P}_{Cvx-TRSP}) \ \min_{\boldsymbol{x}_i\in\mathbb{C}^n}.\ \boldsymbol{x}^H\boldsymbol{Q}\boldsymbol{x} - 2\mathsf{Re}\big\{\boldsymbol{q}^H\boldsymbol{x}\big\}, \tag{20a}$$

$$\text{s.t. } \|\boldsymbol{x}\|_2^2 \le \rho^2, \tag{20b}$$

where $\boldsymbol{Q} \succcurlyeq 0$. Suppose that $\boldsymbol{Q}$'s dimension and rank are n and r respectively and its eigenvalue decomposition is $\boldsymbol{Q} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^H$ with $\boldsymbol{\Lambda}$ having its diagonal elements $\{\lambda_i\}_{i=1}^{n}$ arranged in a descending order. Further denote $\boldsymbol{p} \triangleq \boldsymbol{U}^H\boldsymbol{q}$, p_i as the i-th element of $\boldsymbol{p}$ and $\mathcal{N}(\boldsymbol{Q})$ as the null space of $\boldsymbol{Q}$. Then the optimal solution to ($\mathsf{P}_{Cvx-TRSP}$) is given as follows.

- *If $\boldsymbol{q} \perp \mathcal{N}(\boldsymbol{Q})$ and $\sum_{k=1}^{r}|p_k|^2\lambda_k^{-2} \le \rho^2$, one optimal solution to ($\mathsf{P}_{Cvx-TRSP}$) is given by $\boldsymbol{x}^\star = \boldsymbol{U}\boldsymbol{\Lambda}^\dagger\boldsymbol{U}^H\boldsymbol{q}$, which has the minimum l_2-norm among all possible optimal solutions.*
- *Otherwise, the solution to ($\mathsf{P}_{Cvx-TRSP}$) is unique and given by $\boldsymbol{x}^\star = \boldsymbol{U}\big(\boldsymbol{\Lambda} + \mu^\star\boldsymbol{I}_n\big)^{-1}\boldsymbol{U}^H\boldsymbol{q}$, where $\mu^\star$ is the unique positive solution to the equation $\sum_{k=1}^{n}|p_k|^2(\lambda_k + \mu)^{-2} = \rho^2$.*

Proof: Refer to Appendix A. □

Remark 2: Theorem 1 presents an enhancement of the standard result for the general trust region problem (e.g. [30, Sec. 4.3]) when the problem is convex. Compared to conventional result [30], Theorem 1 fully exploits the convexity and has advantages in two ways. First, Theorem 1 has a much lower complexity, which only requires a one-time eigenvalue decomposition. Comparatively, standard methods (e.g. [30, Algorithm 4.3]) would run for multiple iterations with each iteration performing Cholesky decomposition and two matrix inversions. Second, the new result clearly specifies the conditions for unique and multiple solutions, and in the presence of multiple solutions, it gives out the one with the minimum l_2-norm, which corresponds to the power preserving optimal solution.

By transforming $\widetilde{\mathbf{f}}_i \triangleq \mathbf{E}_i^{\frac{1}{2}}\mathbf{f}_i$, we can formulate ($\mathsf{P4}_i$) into a standard form of trust region problem as in ($\mathsf{P}_{Cvx-TRSP}$) and efficiently solve it via Theorem 1. Detailed manipulations are omitted for brevity.

In practice, we implement the decentralized algorithm in the BCA framework. Whenever the subproblem ($\mathsf{P4}$) is to be solved, we turn to optimize the modified objective $h(\mathbf{f};\mathbf{f}^{(t-1)}|\mathbf{G}^{(t-1)},\mathbf{W}^{(t-1)})$, with $\hat{\mathbf{f}}$ set as the entire precoders obtained in last iteration. This leads to simultaneously updating all precoders in parallel, with each precoder obtained by solving a convex trust-region subproblem via Theorem 1. The overall decentralized algorithm to solve ($\mathsf{P0}$) is summarized in Alg.2.

Algorithm 2 Decentralized Algorithm to Solve ($\mathsf{P0}$)

Initialization: Randomly generate feasible $\{\mathbf{F}_i^{(0)}\}_{i=1}^{L}$; Compute $\mathbf{G}^{(0)}$ using (15) and $\mathbf{W}^{(0)}$ by (14); $t = 0$;
repeat
  for $i = 1 : L$ **do** `// In Parallel`
    Update $\mathbf{f}_i^{(t)}$ by Thm.1 ;
  end
  Update $\mathbf{G}^{(t)}$ and $\mathbf{W}^{(t)}$ by (15) and (14); $t++$;
until *convergence*;

To introduce the convergence analysis result, we first present the following useful lemma, which is proved in Appendix VIII-B.

Lemma 1: Assume that the specific set of variables $\big\{\{\bar{\mathbf{F}}_i\}_{i=1}^{L},\bar{\mathbf{W}},\bar{\mathbf{G}}\big\}$ satisfy the relations in (14) and (15). Then the following identities hold true for $i \in \{1,\cdots,L\}$:

$$\frac{\partial\widetilde{\mathsf{SR}}}{\partial\mathbf{F}_i^T}\bigg|_{\{\bar{\mathbf{F}}_i\}_{i=1}^{L},\bar{\mathbf{G}},\bar{\mathbf{W}}} = \frac{\partial\mathsf{SR}}{\partial\mathbf{F}_i^T}\bigg|_{\{\bar{\mathbf{F}}_i\}_{i=1}^{L}}; \tag{21a}$$

$$\frac{\partial\widetilde{\mathsf{SR}}}{\partial\mathbf{F}_i^H}\bigg|_{\{\bar{\mathbf{F}}_i\}_{i=1}^{L},\bar{\mathbf{G}},\bar{\mathbf{W}}} = \frac{\partial\mathsf{SR}}{\partial\mathbf{F}_i^H}\bigg|_{\{\bar{\mathbf{F}}_i\}_{i=1}^{L}}. \tag{21b}$$

For the decentralized algorithm proposed above, we have the following statement

Theorem 2: Algorithm 2 generates a monotonically increasing objective sequence for (P0). The solution sequence has limit points. Each limit point of the solution sequence is a stationary point of (P0).

Proof: Refer to Appendix C. □

One comparable decentralized method is the Jacobi algorithm [31], which is a widely used naive decentralization method. Jacobi method simply updates each variable simultaneously with the other variables being fixed as the values of last iteration. One major drawback of Jacobi algorithm is that the convergence cannot be guaranteed. In fact, the objective sequence obtained by Jacobi's method is usually not monotonic and its iterates can easily get stuck by bad solutions (as shown in Section VII).

IV. Power Consumption Minimization

A. Centralized Method

In this section, we study the power consumption minimization problem ($\mathsf{P1}$), which aims at minimizing the total power consumption while maintaining the network throughput above a predefined target level r_0 and respecting each sensor's power constraint.

$(\mathsf{P}1)$ is intractable due to the nonconvex sum-rate constraint. To tackle the problem, we substitute $\mathsf{SR}\big(\{\mathbf{F}\}_{i=1}^L\big)$ in the constraint (8b) with $\widetilde{\mathsf{SR}}\big(\{\mathbf{F}\}_{i=1}^L, \mathbf{G}, \mathbf{W}\big)$ and turn to solve the modified problem $(\mathsf{P}5)$ as follows

$$(\mathsf{P}5):\ \min_{\{\mathbf{F}_i^L\}_{i=1}^L, \mathbf{G}, \mathbf{W}} . \sum_{i=1}^{L} \mathsf{Tr}\big\{\mathbf{F}_i\big(\mathbf{K}_i\boldsymbol{\Sigma}_\mathbf{s}\mathbf{K}_i^H + \boldsymbol{\Sigma}_i\big)\mathbf{F}_i^H\big\} \tag{22a}$$

$$\text{s.t. } \widetilde{\mathsf{SR}}\big(\{\mathbf{F}_i\}_{i=1}^L, \mathbf{G}, \mathbf{W}\big) \geq r_0, \tag{22b}$$

$$\mathsf{Tr}\big\{\mathbf{F}_i\big(\mathbf{K}_i\boldsymbol{\Sigma}_\mathbf{s}\mathbf{K}_i^H + \boldsymbol{\Sigma}_i\big)\mathbf{F}_i^H\big\} \leq P_i,\ i \in \{1, \cdots, L\}. \tag{22c}$$

Note that the feasibility region of $\{\mathbf{F}_i\}_{i=1}^L$ in $(\mathsf{P}1)$ and $(\mathsf{P}5)$ are identical. To see this, suppose that $\{\widetilde{\mathbf{F}}_i\}_{i=1}^L$ is feasible to (P5). Then according to (12), $\mathsf{SR}\big(\{\widetilde{\mathbf{F}}_i\}_{i=1}^L\big) \geq \widetilde{\mathsf{SR}}\big(\{\widetilde{\mathbf{F}}_i\}_{i=1}^L\big), \mathbf{G}, \mathbf{W}) \geq r_0$, i.e. $\{\widetilde{\mathbf{F}}_i\}_{i=1}^L$ is also feasible to (P1). Conversely, if $\{\widehat{\mathbf{F}}_i\}_{i=1}^L$ is feasible to $(\mathsf{P}1)$, then by (R1) and (R2), $\widetilde{\mathsf{SR}}\big(\{\widehat{\mathbf{F}}_i\}_{i=1}^L\big), \widehat{\mathbf{G}}, \widehat{\mathbf{W}}) = \mathsf{SR}\big(\{\widehat{\mathbf{F}}_i\}_{i=1}^L\big) \geq r_0$ with $\widehat{\mathbf{G}}$ and $\widehat{\mathbf{W}}$ satisfying (15) and (14) with $\{\widehat{\mathbf{F}}_i\}_{i=1}^L$ given. Therefore the feasible region of $\{\mathbf{F}_i\}_{i=1}^L$ associated with $(\mathsf{P}1)$ and $(\mathsf{P}5)$ are the same.

We solve $(\mathsf{P}5)$ by alternatively updating $\{\mathbf{F}_i\}_{i=1}^L$, $\mathbf{G}$ and $\mathbf{W}$. When $\{\mathbf{F}_i\}_{i=1}^L$ are fixed, we update $\mathbf{G}$ and $\mathbf{W}$ by (15) and (14) respectively, which can guarantee a monotonic decreasing power consumption (as will be shown in Theorem 3).

When $\mathbf{G}$ and $\mathbf{W}$ are given, following the notations in Sec.III, $(\mathsf{P}5)$ with respect to $\{\mathbf{F}_i\}_{i=1}^L$ can be rewritten as

$$(\mathsf{P}5):\ \min_{\mathbf{f}} .\ \mathbf{f}^H\mathbf{E}\mathbf{f}, \tag{23a}$$

$$\text{s.t. } \mathbf{f}^H\big(\mathbf{A}+\mathbf{C}\big)\mathbf{f} - 2\mathsf{Re}\{\mathbf{g}^H\mathbf{B}\mathbf{f}\} - r_1 \leq 0. \tag{23b}$$

$$\mathbf{f}^H\mathbf{D}_i\mathbf{f} \leq P_i, \quad i \in \{1, \cdots, L\}. \tag{23c}$$

where $\mathbf{E} \triangleq \mathsf{Diag}\{\mathbf{E}_1, \cdots, \mathbf{E}_L\}$ and $r_1 \triangleq \log\det(\mathbf{W}) + \log\det(\boldsymbol{\Sigma}_\mathbf{s}) + K - r_0 - \sigma_0^2\mathsf{Tr}\{\mathbf{G}\mathbf{W}\mathbf{G}^H\} - \mathsf{Tr}\{\mathbf{W}\boldsymbol{\Sigma}_\mathbf{s}\}$. Still $(\mathsf{P}5)$ can be transformed into an SOCP form and numerically solved, which results in an SOCP-based centralized algorithm.

B. Decentralized Method

In the sequel we discuss a decentralized way to solve $(\mathsf{P}5)$. The dual problem of $(\mathsf{P}5)$ is given as

$$\max_{\nu\geq 0} . \Big\{f(\nu) \triangleq \min_{\{\mathbf{f}_i\in\mathcal{X}_i\}_{i=1}^L} .\ \mathbf{f}^H\big[\mathbf{E}+\nu\big(\mathbf{A}+\mathbf{C}\big)\big]\mathbf{f} - 2\nu\mathsf{Re}\{\mathbf{g}^H\mathbf{B}\mathbf{f}\} - \nu r_1\Big\} \tag{24}$$

where ν is the Lagrangian multiplier associated with the sum-rate constraint (23b) and $\mathcal{X}_i \triangleq \{\mathbf{f}_i : \mathbf{f}_i^H\mathbf{E}_i\mathbf{f}_i \leq P_i\}$.

When strong duality holds, $(\mathsf{P}5)$ can be solved through dual decomposition method [33] by solving its dual problem $\max_{\nu\geq 0} f(\nu)$ as shown above. For each fixed ν, the inner minimization problem in (24) can still be solved in a decentralized manner. Specifically to parallelize computation, we introduce the tight block-decoupling upper bound of the term $\mathbf{f}^H\mathbf{A}\mathbf{f}$ as in (18). Following similar lines as in Sec.III-B, the inner minimization in (24) can be performed by iteratively solving L smaller problems in parallel with each small problem given as

$$(\mathsf{P}5_i):\ \min_{\mathbf{f}_i\in\mathcal{X}_i} .\ \mathbf{f}_i^H\big[\nu\big(\lambda_{\max}(\mathbf{A})\mathbf{I}_{KN_i} + \mathbf{C}_i\big) + \mathbf{E}_i\big]\mathbf{f}_i - 2\nu\mathsf{Re}\big\{\mathbf{t}_i^H\mathbf{f}_i\big\},$$

where $\mathbf{t}_i$ is identically defined as in (T3) of Sec.III-B, i.e.

$$\mathbf{t}_i \triangleq \big[\mathbf{B}_i^H\mathbf{g} + \lambda_{\max}(\mathbf{A})\hat{\mathbf{f}}_i - \mathbf{p}_i\big] \tag{26}$$

with $\mathbf{p}_i \triangleq [\mathbf{A}\hat{\mathbf{f}}]_{\sum_{j=1}^{i-1}(J_jN_j)+1:\sum_{j=1}^{i}(J_jN_j)}$ and $\hat{\mathbf{f}}$ being $\mathbf{f}$ obtained in the latest iteration. The distributive method to solve the inner minimization in (24) is summarized in Alg.4.

Algorithm 3 Solve the Inner Min. in (24) Distributively

Initialization: given a feasible $\mathbf{f}^{(0)}$; $t = 0$;
repeat
  update $\mathbf{t}_i$ via (24) using $\mathbf{f}^{(t)}$;
  for $i = 1 : L$ **do** `// In Parallel`
    solve $(\mathsf{P}5_i)$ to obtain $\mathbf{f}_i^{(t+1)}$ via Thm.1 ;
  end
  $t++$;
until *convergence*;

Notice that the inner minimization problem in (24) is strictly convex and thus has unique minimizer. According to Theorem 6.3.3 in [32], $f(\nu)$ is differentiable and it can be readily seen that $\nabla_\nu f(\nu) = \mathbf{f}^{\star H}(\mathbf{A}+\mathbf{C})\mathbf{f}^\star - 2\mathsf{Re}\{\mathbf{g}^H\mathbf{B}\mathbf{f}^\star\} - r_1$ with $\mathbf{f}^\star$ being the optimal solution to the inner minimization in (24) with ν given. Therefore according to [33], to obtain $\nu^\star = \arg\max_{\nu\geq 0} f(\nu)$, the multiplier ν can be updated as

$$\nu := \big[\nu + \alpha\nabla_\nu f(\nu)\big]_+ \tag{27}$$

where $[\cdot]_+ \triangleq \max\{\cdot, 0\}$.

Thus the dual decomposition method to solve $(\mathsf{P}5)$ which enables parallel computation is summarized as in Alg.4.

Algorithm 4 Solve $(\mathsf{P}5)$ Distributively

Initialization: $\nu^{(0)}$ is arbitrary positive value; feasible $\mathbf{f}^{(0)}$ is given; $s = 0$;
repeat
  invoke Alg.3 to obtain $\mathbf{f}^{(s+1)}$;
  $\nu^{(s+1)} := \Big[\nu^{(s)} + \alpha\nabla_\nu f(\nu)\big|_{\mathbf{f}^{(s+1)}}\Big]_+$; $s++$;
until $\nu^{(s)}$ *converges to positive value or* $\nu^{(s)} = 0$;
invoke Alg.3 to obtain $\mathbf{f}^{(s+1)}$;

Therefore the power minimizing precoders can be obtained by Alg.5, which can be implemented in both centralized and decentralized manner.

For Alg.5, we have the following convergence result, which is proved in Appendix.D.

Theorem 3: The sequence $\big(\{\mathbf{F}_i^{(t)}\}_{i=1}^L, \mathbf{W}^{(t)}, \mathbf{G}^{(t)}\big)$ *generated by Algorithm 5 yields a monotonic decreasing objective sequence and has limit points, with each limit point being a stationary point to (P1).*

Algorithm 5 Power Consumption Minimization

1 **Initialization**: generate a feasible $\{\mathbf{F}_i^{(0)}\}_{i=1}^L$; obtain $\mathbf{G}^{(0)}$ by (15); obtain $\mathbf{W}^{(0)}$ by (14); t=1;
2 **repeat**
3 fix $\mathbf{G}^{(t-1)}$ and $\mathbf{W}^{(t-1)}$, update $\{\mathbf{F}_i^{(t)}\}_{i=1}^L$ by solving (P5) via SOCP solver or invoking Alg.4 ;
4 fix $\{\mathbf{F}_i^{(t)}\}_{i=1}^L$, update $\mathbf{G}^{(t)}$ and $\mathbf{W}^{(t)}$ in order by (15) and (14) respectively; j++ ;
5 **until** *convergence*;

C. A Sufficient Condition for Strong Duality

Recall that the premise to exploit dual decomposition method [33] is the strong duality holds for (P5). Then a natural question rises—when strong duality is valid? To answer this question, we first introduce the notation $(\mathsf{P5}_{t,\epsilon})$, which is a modified version of the problem (P5) defined as follows

$$(\mathsf{P5}_{t,\epsilon}): \ \min_{\mathbf{f}}. \ \mathbf{f}^H\mathbf{E}\mathbf{f} \tag{28a}$$

$$\text{s.t. } \widetilde{\mathsf{SR}}\big(\mathbf{f}, \mathbf{G}^{(t-1)}, \mathbf{W}^{(t-1)}\big) \ge r_0 + \frac{\epsilon}{2^t}, \tag{28b}$$

$$\mathbf{f}^H\mathbf{D}_i\mathbf{f} \le P_i, \quad i \in \{1, \cdots, L\}. \tag{28c}$$

with t being the iteration index of Alg.5 and $\epsilon > 0$. Then a sufficient condition implying strong duality of $(\mathsf{P5}_{t,\epsilon})$ is given as follows

Lemma 2: If there exists a feasible $\{\boldsymbol{F}\}_{i=1}^L$ yielding $\mathsf{SR}\big(\{\boldsymbol{F}\}_{i=1}^L\big) > r_0$, then we can find sufficiently small ϵ such that $(P5_{t,\epsilon})$ has strong duality for $t = 1, 2, \cdots$.

Proof: Suppose $\{\widehat{\mathbf{F}}\}_{i=1}^L$ is feasible and $\mathsf{SR}\big(\{\widehat{\mathbf{F}}\}_{i=1}^L\big) = r_0 + \epsilon_0$ with $\epsilon_0 > 0$. Then we set $\{\mathbf{F}^{(0)}\}_{i=1}^L = \{\widehat{\mathbf{F}}\}_{i=1}^L$ and choose ϵ satisfying $0 < \epsilon < \epsilon_0$.

Notice that $\widetilde{\mathsf{SR}}\big(\{\mathbf{F}_i^{(0)}\}_{i=1}^L, \mathbf{G}^{(0)}, \mathbf{W}^{(0)}\big) = \mathsf{SR}\big(\mathbf{F}_i^{(0)}\big) = r_0 + \epsilon_0 > r_0 + \frac{\epsilon}{2}$. We claim that strong duality holds for $(\mathsf{P5}_{1,\epsilon})$. In fact, choose $\widetilde{\mathbf{F}}_i = (1-\gamma)\mathbf{F}_i^{(0)}$ $\forall i$ such that $0 < \gamma < 1$. Then all the power constraints (28c) $\forall i$ will become strictly feasible (inactive). Since $\widetilde{\mathsf{SR}}$ is continuous function in $\{\mathbf{F}\}_{i=1}^L$, $\widetilde{\mathsf{SR}}\big(\{\widetilde{\mathbf{F}}_i\}_{i=1}^L, \mathbf{G}^{(0)}, \mathbf{W}^{(0)}\big) > r_0 + \frac{\epsilon}{2}$ stands as long as γ is sufficiently small. Therefore Slater's condition is satisfied for $(\mathsf{P5}_{1,\epsilon})$ and strong duality holds.

At the same time the following facts stand:

$$\begin{aligned}&\mathsf{SR}\big(\{\mathbf{F}_i^{(1)}\}_{i=1}^L\big)\\ &= \widetilde{\mathsf{SR}}\big(\{\mathbf{F}_i^{(1)}\}_{i=1}^L, \mathbf{G}^{(1)}, \mathbf{W}^{(1)}\big)\\ &\ge \widetilde{\mathsf{SR}}\big(\{\mathbf{F}_i^{(1)}\}_{i=1}^L, \mathbf{G}^{(1)}, \mathbf{W}^{(0)}\big) \ge \widetilde{\mathsf{SR}}\big(\{\mathbf{F}_i^{(1)}\}_{i=1}^L, \mathbf{G}^{(0)}, \mathbf{W}^{(0)}\big)\\ &\ge r_0 + \frac{\epsilon}{2} > r_0 + \frac{\epsilon}{2^2}.\end{aligned} \tag{29}$$

Starting from (29) again and following the literally identical arguments as above, we can readily prove that $(\mathsf{P5}_{2,\epsilon})$ has strong duality and $\mathsf{SR}\big(\{\mathbf{F}_i^{(2)}\}_{i=1}^L\big) \ge r_0 + \frac{\epsilon}{2^3}$. The assert is indeed proved by repeating the above procedure. □

According to Lemma 2, as long as an initial point yielding a sum-rate strictly larger than r_0 can be found, we can adjust Alg.5 by replacing the (P5) therein with $(\mathsf{P5}_{t,\epsilon})$ to ensure that the decentralized algorithm is always valid to invoke.

D. Initialization

Another issue worth attention is the initialization of Alg.5. To this end, we first identify an upper bound of *admissible* r_0 (a r_0 is called admissible if it makes (P1) feasible).

The throughput limit can be obtained by considering the scenario of centralized sensing, where all sensors convey their noisy observations to the FC through an ideal channel, as if all the observations are obtained locally at the FC. In this case, the signal collected at the FC is given as:

$$\mathbf{r}_0 = \widetilde{\mathbf{K}}\mathbf{s} + \widetilde{\mathbf{n}} \tag{30}$$

where $\widetilde{\mathbf{K}} \triangleq [\mathbf{K}_1^T, \cdots, \mathbf{K}_L^T]^T$ and the noise $\widetilde{\mathbf{n}} \triangleq [\mathbf{K}_1^T, \cdots, \mathbf{K}_L^T]^T$ has covariance $\widetilde{\boldsymbol{\Sigma}} \triangleq \mathsf{Diag}\{\boldsymbol{\Sigma}_1, \cdots, \boldsymbol{\Sigma}_L\}$. In this case the information rate of $\mathbf{s}$ is

$$R_{bd} = \log\det\big(\mathbf{I}_{\sum_{i=1}^L K_i} + \widetilde{\mathbf{K}}\boldsymbol{\Sigma}_{\mathbf{s}}\widetilde{\mathbf{K}}^H\widetilde{\boldsymbol{\Sigma}}^{-1}\big). \tag{31}$$

To see that R_{bd} is an upper bound of admissible r_0, we notice the fact that $\mathbf{s} \to \mathbf{r}_0 \to \mathbf{r}$ actually forms a two-stage Markov chain [34]. Therefore according to the fundamental Data-Processing Inequality ([34, Th. 2.8.1]), R_{bd} is actually an upper bound of the information rate from $\mathbf{s}$ to $\mathbf{r}$.

As verified by numerical results in Fig.6, when the power supplies are sufficiently large, the throughput bound in (31) can be arbitrarily approached.

When $r_0 < R_{bd}$, to find a feasible solution to (P1) for given finite power supplies $\{P_i\}_{i=1}^L$ is to solve the following problem

$$\min_{\mathbf{f}}. \ 0 \tag{32a}$$

$$\text{s.t. } \mathsf{SR}\big(\mathbf{f}\big) \ge r_0, \tag{32b}$$

$$\mathbf{f}^H\mathbf{D}_i\mathbf{f} \le P_i, \quad i \in \{1, \cdots, L\}. \tag{32c}$$

Obviously this is equivalent to check the optimal value of (P0) (with constraints (32c)) is larger than r_0 or not. Therefore Alg.1/Alg.2 can be invoked to initialize Alg.5. If the solution provided by Alg.1/Alg.2 yields a sum-rate no smaller than r_0, a feasible solution of Alg.5 is obtained. Otherwise we have to claim that (P1) is infeasible.

V. Energy Efficiency

In this section, we consider the energy efficiency(EE) optimization problem (P2). We still replace the original sum-rate function SR with its lower bound $\widetilde{\mathsf{SR}}$, which is also adopted by [15], [17], and [18]. The modified EE problem is formulated as follows

$$(\mathsf{P6}): \ \min_{\{\mathbf{F}_i\}_{i=1}^L, \mathbf{G}, \mathbf{W}}. \ \frac{\widetilde{\mathsf{SR}}\big(\{\mathbf{F}_i\}_{i=1}^L, \mathbf{W}, \mathbf{G}\big)}{\mathsf{Pwr}\big(\{\mathbf{F}_i\}_{i=1}^L\big)} \tag{33a}$$

$$\text{s.t. } \widetilde{\mathsf{SR}}\big(\{\mathbf{F}_i\}_{i=1}^L\big) \ge r_0. \tag{33b}$$

$$\mathsf{Tr}\big\{\mathbf{F}_i\big(\mathbf{K}_i\boldsymbol{\Sigma}_{\mathbf{s}}\mathbf{K}_i^H + \boldsymbol{\Sigma}_i\big)\mathbf{F}_i^H\big\} \le P_i, \ i \in \{1, \cdots, L\} \tag{33c}$$

Problem (P6) can also be solved via BCD methodology. $\mathbf{G}$ and $\mathbf{W}$ are still updated via (15) and (14) respectively. When $\mathbf{G}$ and $\mathbf{W}$ are given, the optimization of (P6) with respect to $\{\mathbf{F}_i\}_{i=1}^L$ is a nonlinear fractional maximization problem and can be solved through the seminal method developed by [29]. Define $g(\{\mathbf{F}_i\}_{i=1}^L|\eta) \triangleq \widetilde{\mathsf{SR}}(\{\mathbf{F}_i\}_{i=1}^L) - \eta\mathsf{Pwr}(\{\mathbf{F}_i\}_{i=1}^L)$ and $\mathcal{Z} \triangleq \{\{\mathbf{F}_i\}_{i=1}^L \text{s.t.(33b-33c) holds}\}$, (P6) can be solved via the iterative procedure presented in Alg.6 (details are referred to [15], [17], and [29])

Algorithm 6 Solving (P6)

1 **Initialization**: given feasible $\big(\{\mathbf{F}_i^{(0)}\}_{i=1}^L, \mathbf{W}^{(0)}, \mathbf{G}^{(0)}\big)$; $s=0$;
2 **repeat**
3 $\quad \{\mathbf{F}_i^{(s,0)}\}_{i=1}^L = \{\mathbf{F}_i^{(s)}\}_{i=1}^L$;
4 $\quad \eta^{(s,0)} = \widetilde{\mathsf{SR}}\big(\{\mathbf{F}_i^{(s)}\}_{i=1}^L\big)/\mathsf{Pwr}(\{\mathbf{F}_i^{(s)}\}_{i=1}^L)$; $t=0$;
5 $\quad$ **repeat**
6 $\qquad \{\mathbf{F}_i^{(s,t+1)}\}_{i=1}^L = \arg\max_{\mathcal{Z}} \; g(\{\mathbf{F}_i\}_{i=1}^L|\eta^{(s,t)})$;
7 $\qquad \eta^{(s,t+1)} = \dfrac{\widetilde{\mathsf{SR}}\big(\{\mathbf{F}_i^{(s,t+1)}\}_{i=1}^L\big)}{\mathsf{Pwr}(\{\mathbf{F}_i^{(s,t+1)}\}_{i=1}^L)}$; $t++$;
8 $\quad$ **until** *convergence*;
9 $\quad \{\mathbf{F}_i^{(s+1)}\}_{i=1}^L = \arg\max_{\mathcal{Z}} \; g(\{\mathbf{F}_i\}_{i=1}^L|\eta^{(s,t)})$;
10 $\quad$ obtain $\mathbf{G}^{(s+1)}$ and $\mathbf{W}^{(s+1)}$ by (15) and (14); $s++$;
11 **until** *convergence*;

Notice that the major complexity of Alg.6 comes from the step maximizing $g(\cdot)$. Following the previous notations, the problem $\max_{\mathcal{Z}} \; g(\{\mathbf{F}_i\}_{i=1}^L|\eta)$ is equivalently written as

$$(\mathsf{P7}): \min_{\mathbf{f}_i\in\mathcal{X}_i}. \; \mathbf{f}^H\big(\mathbf{A}+\mathbf{C}+\eta\kappa\mathbf{E}\big)\mathbf{f} - 2\mathsf{Re}\{\mathbf{g}^H\mathbf{B}\mathbf{f}\} \tag{34a}$$

$$\text{s.t. } \mathbf{f}^H\big(\mathbf{A}+\mathbf{C}\big)\mathbf{f} - 2\mathsf{Re}\{\mathbf{g}^H\mathbf{B}\mathbf{f}\} - r_1 \leq 0. \tag{34b}$$

where irrelevant constant terms are omitted in (34a). Observing the similarity between the problem (P7) and (P5), it is readily realized that the dual-decomposition based decentralized algorithm in Sec.IV-B can also be used to solve (P7). Specifically as performed in Sec.IV, we attack (P7) via solving its dual problem as follows

$$\max_{\xi\geq 0}. \Big\{ f_{EE}(\xi) \triangleq \min_{\{\mathbf{f}_i\in\mathcal{X}_i\}_{i=1}^L}. \; \mathbf{f}^H\big[(1+\xi)\big(\mathbf{A}+\mathbf{C}\big)+\eta\kappa\mathbf{E}\big]\mathbf{f} - 2(1+\xi)\mathsf{Re}\{\mathbf{g}^H\mathbf{B}\mathbf{f}\} - \xi r_1 \Big\} \tag{35}$$

with ξ being the Lagrangian multiplier associated with the sum-rate constraint. For each fixed ξ, we again adopt the BSUM methodology to modify the inner minimization problem in (35) as Sec.III and Sec.IV and decompose it into L small problems given as

$$(\mathsf{P7}_i): \min_{\mathbf{f}_i\in\mathcal{X}_i}. \; \mathbf{f}_i^H\big[(1+\xi)\big(\lambda_{\max}(\mathbf{A})\mathbf{I}_{KN_i}+\mathbf{C}_i\big)+\eta\kappa\mathbf{E}_i\big]\mathbf{f}_i - 2(1+\xi)\mathsf{Re}\big\{\mathbf{t}_i^H\mathbf{f}_i\big\}. \tag{36a}$$

The Lagrangian multiplier ξ can be determined via an iterative gradient update

$$\xi := \big[\xi + \alpha\nabla_\xi f_{EE}(\xi)\big]_+ \tag{37}$$

Therefore a decentralized solution to update $\{\mathbf{F}_i\}_{i=1}^L$ for EE maximization can also be obtained via implementing the step-6 and step-9 of Alg.6 by invoking Alg.4, with the update of $(\mathsf{P5}_i)$ and ν therein being literally replaced by $(\mathsf{P7}_i)$ and update of ξ respectively. Besides the arguments on strong duality in Sec.IV-C and the discussion on initialization in Sec.IV-D are still valid for Alg.6.

TABLE I
MATLAB RUN-TIME PER OUTER-LOOP (IN SEC.)

Dim. \ L	Algorithms	$L=5$	$L=10$	$L=20$
$K=1$	our SOCP cntr	0.2655	0.3956	0.7675
$J=1$	our decntr	0.0025	0.0044	0.0088
$M=2$	WMMSE-SDR [20]	0.7037	1.475	3.244
$N_i=2$	TSTINR [21]	0.7295	1.479	3.269
$K=4$	our SOCP cntr	0.5137	1.914	13.88
$J=4$	our decntr	0.0066	0.0106	0.0191
$M=4$	WMMSE-SDR [20]	1.139	2.293	4.883
$N_i=4$	TSTINR [21]	1.178	2.463	5.325
$K=8$	our SOCP cntr	10.01	80.98	668.0
$J=8$	our decntr	0.0099	0.0143	0.0238
$M=8$	WMMSE-SDR [20]	5.217	10.31	20.73
$N_i=8$	TSTINR [21]	4.951	10.17	20.71

Notes: SDPT3 solver of CVX is chosen to implement the algorithms in [20], [21] and our centralized algorithm.

TABLE II
MATLAB RUNTIME(IN SEC.) TO UPDATE $\{\mathbf{F}_i\}_{i=1}^L$ IN ALG.6

mthd. \ L	20	50	100	150	200	250	300
socp-cntr.	5.91	90.1	407.7	2836	7015	15464	—
dual-dcntr.	40.2	40.9	42.3	47.4	55.6	56.0	65.7

Notes: SDPT3 solver of CVX is chosen to solve SOCP
'—': the problem is too large to be solved by solver

VI. DISCUSSION: CENTRALIZED V.S. DECENTRALIZED SOLUTIONS

In previous sections both the centralized and decentralized algorithms are developed for different metric optimizations. Here in this section we briefly discuss the advantages and shortcomings of these two alternatives.

First we consider the complexity. The dominating part of the complexity of all the above algorithms comes from updating $\{\mathbf{F}_i\}_{i=1}^L$. Take the centralized solution Alg.1 as an example, solving the SOCP (P4) can be shown to have a complexity of $\mathcal{O}\big(\sqrt{L}(\sum_{i=1}^L J_i N_i)^3\big)$ [28]. Comparatively Alg.2 solves L small problems in parallel with each small problem having a complexity of $\mathcal{O}(J_i^3 N_i^3)$. Obviously the decentralized algorithm has highly advantageous time complexity over its centralized counterpart. As shown by Table I and Table II, the complexity of centralized algorithm grows aggressively when the number of sensors increases. For large-scale WSN having several hundreds of sensors, the processing time becomes unrealistically tremendous even though the problem was still solvable. Besides, the centralized algorithm imposes a high computing/storage demand for the processing units, which may become bottleneck of WSN deployment.

Another aspect to be concerned is the communication cost. The centralized solution has a low communication cost. Usually one node (e.g. the FC or some computation center in the network) performs the entire optimization process and then acknowledges each sensor its associated precoder.

In contrast, the communication cost of the decentralized algorithm seems much higher since in each iteration the information of different precoders should be exchanged. When the parallel processing units are the sensors themselves, the decentralized algorithm may seem impractical since it is highly demanding in communication.

However it is worth pointing out that communication cost is indeed not always an issue depending to the deployment scenarios. As suggested by [1] and [2], many wireless multimedia sensor networks have processing hubs that are equipped with multi-core processors. The processing cores are typically connected with wired ethernets and thus the intra-core information exchange is not issue at all. Our proposed decentralized algorithms can be well accommodated by these multi-core processing platforms and impose low computation requirement (since semi-closed form solutions are provided) on each processing core.

To implement the proposed decentralized algorithms on a multi-core platform, the processing cores can be divided into two groups. One group of cores acts as workhorses, which solves one/multiple decomposed small problems (i.e. ($\mathsf{P}4_i$), ($\mathsf{P}5_i$) and ($\mathsf{P}7_i$)). The other group of cores acts as "switches' to perform information exchanges, which collects the latest precoder information from computing cores, possibly performs a linear transformation (e.g. $\mathbf{t}_i$ and $\mathbf{p}_i$ in the proposals) and then distributes the updated to computing cores.

VII. Numerical Results

In this section we provide numerical results to verify our proposals. In the following numerical experiments two different network settings are tested: (i) heterogeneous network and (ii) homogeneous network. The settings of these two networks are elaborated as follows:

- heterogeneous network: the source has dimension $K = 3$. There are $L = 3$ sensors, which use $N_1 = 3$, $N_2 = 4$ and $N_3 = 5$ antennas respectively. All the observation matrices are identity matrices. The FC uses $M = 4$ antennas. Observation SNR and transmission power constraints at sensors are: $\mathsf{SNR}_1 = 6dB$, $\mathsf{SNR}_2 = 7dB$, $\mathsf{SNR}_3 = 8dB$, $P_1 = 2$, $P_2 = 2$ and $P_3 = 3$.
- homogeneous network: the source has dimension $K = 3$. There are $L = 5$ sensors. Each sensor has $N_i = 4$ antennas and identity observation matrices, transmission power $P_i = 2$ and observation level $\mathsf{SNR}_i = 7dB$. The FC uses $M = 4$ antennas.

We set the covariance matrices of the source signal and of all the observation noise as $\boldsymbol{\Sigma}_{\mathbf{s}} = \sigma_s^2 \boldsymbol{\Sigma}_0$ and $\boldsymbol{\Sigma}_i = \sigma_i^2 \boldsymbol{\Sigma}_0$ for $i \in \{1, \ldots, L\}$, where $\boldsymbol{\Sigma}_0$ is a Toeplitz matrix with appropriate dimension given as $\left[\boldsymbol{\Sigma}_0\right]_{j,k} = \gamma^{|j-k|}$. The parameter γ here is used to adjust the correlation level between different components of the signal or noise. In our experiments γ is set as 0.5. We define the sensing signal-to-noise-ratio (SNR) at the i-th sensor as $\mathsf{SNR}_i \triangleq \sigma_s^2/\sigma_i^2$ and the channel SNR as $\mathsf{SNR} \triangleq \sigma_s^2/\sigma_0^2$. To solve the SOCP problem or SDP problems, SDPT3 solver of CVX is used.

Fig. 2 demonstrates the performances of our centralized and decentralized algorithms. For comparison, the WMMSE-SDR

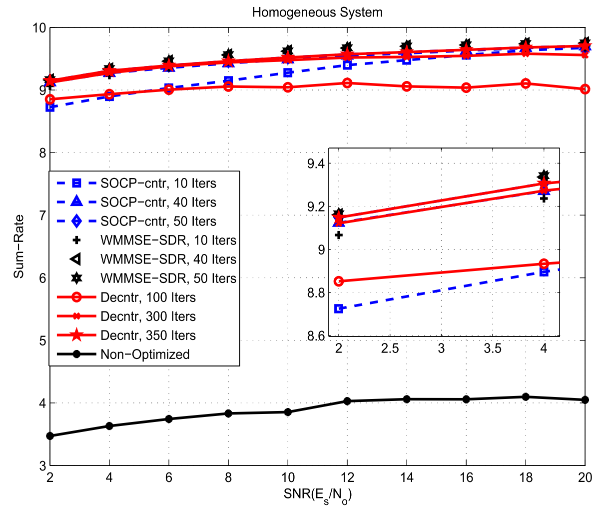

Fig. 2. Test for homogeneous network.

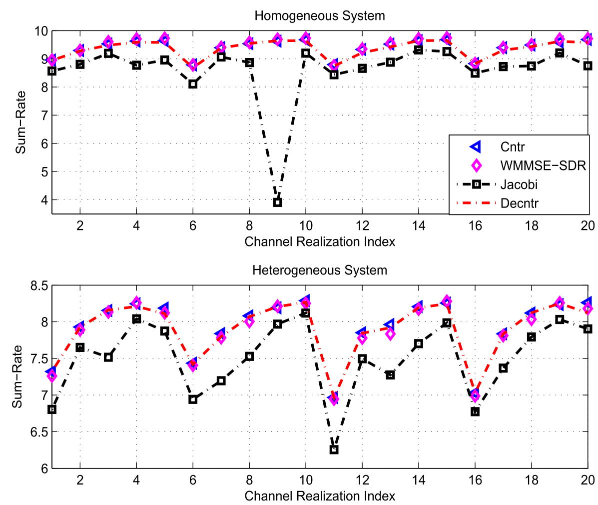

Fig. 3. Impact of different channel settings to our algorithms. Both the heter. and homo. systems are tested.

algorithm in [20] are also tested. For each specific channel noise level, 100 different channel realizations and tested. Different algorithms are tested starting from an identical initial feasible point. The average sum-rate among all channel realizations is plotted in figures. For the two centralized algorithms, i.e. our Alg.1 and the WMMSE-SDR in [20], it usually takes 50 iterations to achieve convergence. Our decentralized algorithm typically takes several hundreds of iterations to reach convergence. It should be highlighted that this result could be somewhat misleading since it appears that our decentralized algorithm is slower than the centralized ones. In fact, each iteration of the decentralized algorithm requires very little time since semi-analytical solution is available by Thm.1. This can be clearly seen in Table I.

Fig.3 compares our centralized and decentralized solutions. A group of channels matrices are randomly generated with various channel noise levels. As a benchmark, WMMSE-SDR algorithm in [20] and the Jacobi method [31] are also presented. The figure suggests that our proposed algorithms

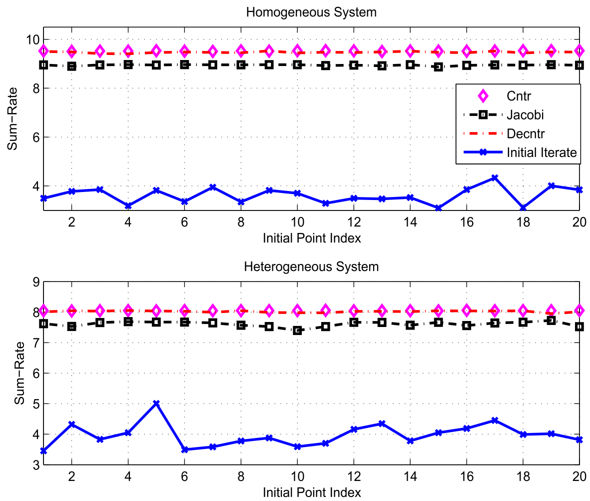

Fig. 4. Impact of different initial points to our algorithms. Heter. and homo. systems are tested. $\mathsf{SNR} = 8dB$.

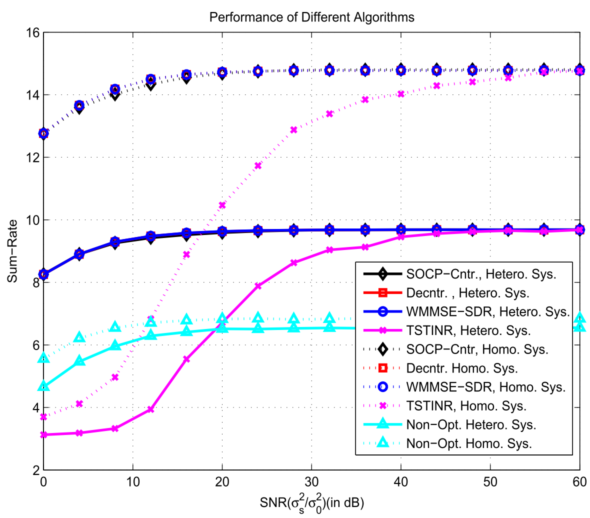

Fig. 5. Performance: Our Algorithms vs. the WMMSE-SDR in [20] v.s. the TSTINR Algorithm in [21].

work well for variant channel settings. In contrast, the Jacobi algorithm can be easily stuck by bad solutions.

Fig.4 shows the impact of variant initials to our algorithms. For a given (randomly generated) channel realization, we randomly generate a group of feasible points as the starting points of our algorithms. We see that both of our algorithms are rather robust, exhibiting insensitivity to the initial points, and capable of converging to identical performance.

We further compare our proposed algorithms with the closely relevant solutions developed in [20] and [21]. For the TSTINR algorithm developed in [21], although it was mentioned in [21] that it could be advantageous at high SNR, we have not observed it in our system model. As shown in Figure 5, in both homogeneous and heterogeneous system setups, our proposed methods deliver consistently good performance for the entire channel SNR range.

As an indication to complexity, we evaluate in Table I the average MATLAB runtime of our algorithms and the solutions

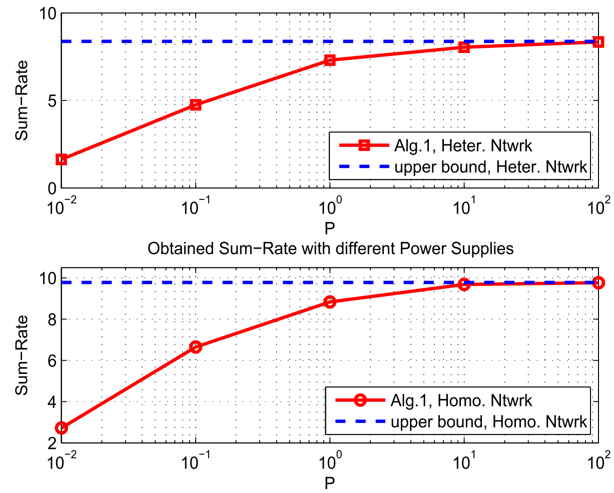

Fig. 6. Approachable throughput upper bound.

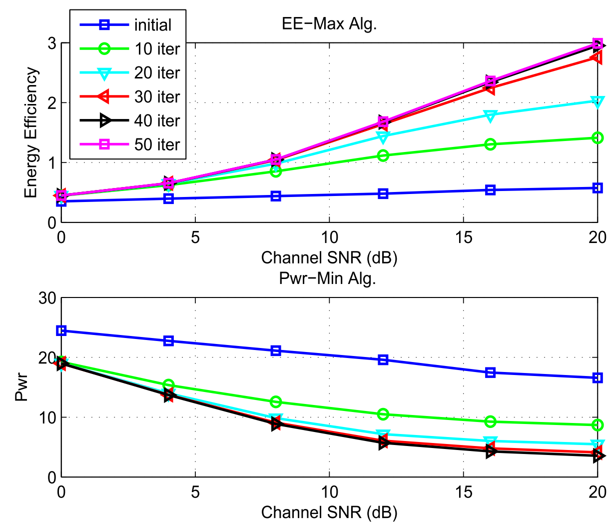

Fig. 7. Pwr-Min. and EE-Max.

in [20] and [21]. A homogeneous system is tested with different parameter settings as shown in the table. As expected, the semi-closed-form solution in Thm.1 enables our decentralized algorithm to run extremely fast.

We also compare our algorithms' performance with the theoretical upper limit given by (31). In our experiment, all sensors have identical power supply P. We vary P from 0.01 to 100 and run Alg.1. The obtained results are plotted in Fig.6 with the upper limit R_{bd} in (31) presented as well. As demonstrated in the figure, when supplied with sufficient power, our algorithms yield limit-approaching throughput. This result also suggests that R_{bd} is an upper bound for admissible r_0 for power-minimization and EE-maximization problems.

In Fig.7 we present Alg.5 and Alg.6, which optimize power consumption and energy efficiency respectively. Each point in the plot is obtained by averaging the results associated with 50 different channel realizations in heterogeneous network. Similar results can be observed for homogeneous network settings. Alg.5 and Alg.6 converge within 30 and 40 outer

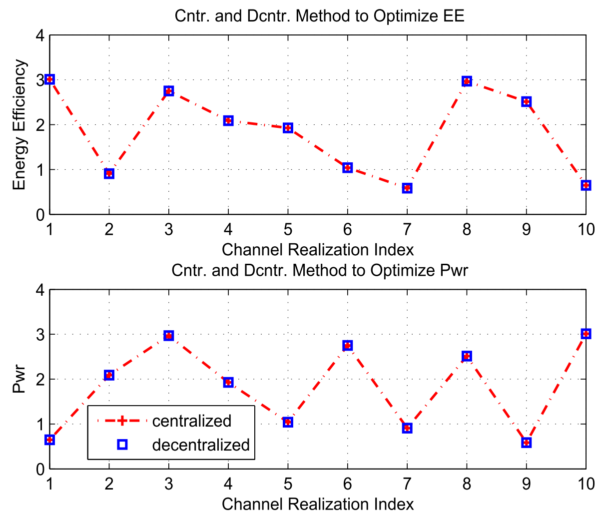

Fig. 8. Centralized and decentralized method to optimize Pwr and EE.

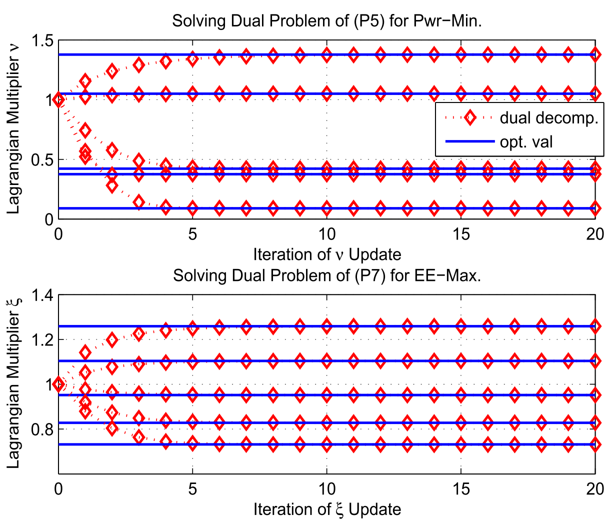

Fig. 9. Convergence of lagrangian multipliers for decentralized Pwr-Min. and EE-Max.

layer iterations respectively. Both the optimized power consumption and energy efficiency improves when the channel SNR increases.

Fig.8 presents the performance of decentralized implementation of Alg.5 and Alg.4, with each point in the figure obtained from a randomly generated channel realization and channel noise level. Usually the values within $[0.5, 2]$ is a good choice for the step size α to update the Lagrangian multipliers. In our experiment, we take the step size $\alpha = 1.2$ and set the initial values for Lagrangian multiplier ν and ξ as 1.0. As shown by Fig.9, the one-dimension Lagrangian multipliers usually converge to the optima with a satisfying precision (10^{-3}) within several tens of iterations.

Fig. 10 clearly presents the distinction between the power-minimizing and EE-maximizing beamformers from different aspects. As can be seen, the power-minimizing beamformers always provide a tightly sufficient sum-rate which is required by the sum-rate constraint so as to reduce the power to the most extent. While the EE-maximizing beamformers seem to

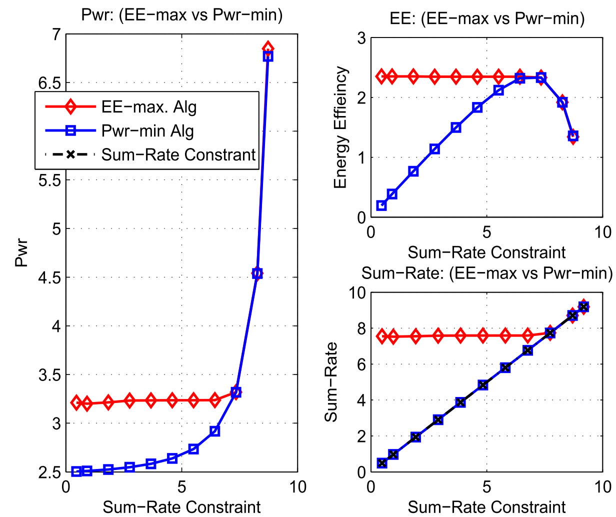

Fig. 10. EE-Max. v.s. Pwr-Min.

achieve a better rate-and-power balance for low and medium rate requirement. In the case of high sum-rate constraint, where the sum-rate constraint is active for both solutions, the two beamforming design criteria coincide with each other.

Table II evaluates the MATLAB runtime of centralized and decentralized methods to update $\{\mathbf{F}_i\}_{i=1}^L$ for EE maximization. For the decentralized implementation, the iterations numbers for Lagrangian multiplier ξ update and $\{\mathbf{F}_i\}_{i=1}^L$ update for (P7) is 40 and 500 respectively, which ensures sufficient convergence of the dual and primal variables. As can be seen from the figure, the centralized algorithm has an aggressively growing complexity with L and finally fails due the to prohibitively high demand for processing capability. In contrast, the complexity of decentralized method, which has all its updates in closed/semi-closed form, is obviously more efficient and less computation/memory-demanding.

VIII. Conclusion

This paper considers the linear precoder design to optimize throughput, power consumption and EE in multi-antenna wireless sensor networks. For the throughput maximization problem, a centralized algorithm with new SOCP formulation is developed. A decentralized algorithm with high efficiency and provable convergence is also designed. For the power consumption and EE optimization problem, a dual-decomposition based decentralized algorithm is designed, with its convergence, initialization and sufficient condition for strong duality being carefully treated. Extensive numerical results are presented to demonstrate the effectiveness of our new approaches.

Appendix

A. Proof of Theorem 1

Proof: Since ($\mathsf{P}_{Cvx-TRSP}$) is convex and Slater's condition is obviously satisfied, it suffices to show the solution satisfies the KKT conditions [26]. Associate the Lagrangian multiplier μ with the power constraint, the KKT conditions are given as

$$(\mathbf{Q}+\mu\mathbf{I}_n)\mathbf{x}=\mathbf{q};\ \|\mathbf{x}\|_2^2\le\rho^2;\ \mu\big(\|\mathbf{x}\|_2^2-\rho^2\big)=0;\ \mu\ge 0. \quad (38)$$

First we consider the case $\mu > 0$. Define $f(\mu) \triangleq \sum_{k=1}^{n} |p_k|^2(\lambda_k+\mu)^{-2}$. By the KKT conditions (38) we have

$$\|\mathbf{x}\|_2^2 = \left\|\left(\mathbf{Q}+\mu\mathbf{I}_n\right)^{-1}\mathbf{q}\right\|_2^2 = \mathbf{q}^H\mathbf{U}\left(\mathbf{\Lambda}+\mu\mathbf{I}_n\right)^{-2}\mathbf{U}^H\mathbf{q} \quad (39)$$

$$= \sum_{k=1}^{n} |p_k|^2(\lambda_k+\mu)^{-2} = f(\mu) = \rho^2. \quad (40)$$

Note $f(\mu)$ is strictly decreasing with respect to μ. If (40) has a positive solution, the supremum of $f(\mu)$, i.e. $\lim_{\mu\to 0} f(\mu)$, must be larger than ρ^2. Only two specific cases will make this condition stand: i) $\mathbf{q} \not\perp \mathcal{N}(\mathbf{Q})$, which means $\exists i \in \{r+1,\cdots,n\}$ such that $p_i \neq 0$; ii) $\mathbf{q} \perp \mathcal{N}(\mathbf{Q})$ and $f(0) = \sum_{k=1}^{r}|p_k|^2\lambda_k^{-2} > \rho^2$. If either i) or ii) occurs, there exists a positive $\mu^\star$ satisfying (40) (which can be efficiently found via bisection search), and the unique solution to ($\mathsf{P}_{Cvx-TRSP}$) can therefore be derived as $\mathbf{x}^\star = \left(\mathbf{Q}+\mu\mathbf{I}_n\right)^{-1}\mathbf{q} = \mathbf{U}\left(\mathbf{\Lambda}^{-1}+\mu^{-1}\mathbf{I}_n\right)\mathbf{U}^H\mathbf{q}$.

Next we consider the case $\mu = 0$. From the above discussion, neither case i) or case ii) occurs at this time, that means $\mathbf{q} \perp \mathcal{N}(\mathbf{Q})$ and $\sum_{k=1}^{r}|p_k|^2\lambda_k^{-2} \le \rho^2$. From (38), we get

$$\mathbf{Q}\mathbf{x} = \mathbf{q}. \quad (41)$$

Since $\mathbf{q} \perp \mathcal{N}(\mathbf{Q})$, $\mathbf{q}$ belongs to the range space $\mathcal{R}(\mathbf{Q})$. So the equation (41) is consistent. Left multiplying $\mathbf{U}^H$ to both sides of (41), we obtain

$$\mathbf{\Lambda}\mathbf{U}^H\mathbf{x} = \mathbf{p}. \quad (42)$$

Let $\bar{\mathbf{\Lambda}}$ be the top-left $r\times r$ sub-matrix of $\mathbf{\Lambda}$, $\bar{\mathbf{U}}$ be the left-most r columns of $\mathbf{U}$, $\hat{\mathbf{U}}$ be the right-most $n-r$ columns of $\mathbf{U}$, and $\bar{\mathbf{p}} \triangleq [p_1,\cdots,p_r]^T$. We can then simplify (42) to

$$\bar{\mathbf{\Lambda}}\bar{\mathbf{U}}^H\mathbf{x} = \bar{\mathbf{p}}. \quad (43)$$

Since $\mathbf{U}$ is unitary, we can represent $\mathbf{x}$ as a linear combination of its columns, i.e. $\mathbf{x} \triangleq \sum_{k=1}^{n}\alpha_k\mathbf{u}_k$. Recall the fact that $\mathcal{R}(\hat{\mathbf{U}}) = \mathcal{N}(\mathbf{Q}) \perp \mathcal{R}(\mathbf{Q}) = \mathcal{R}(\bar{\mathbf{U}})$ and $\mathbf{q} \in \mathcal{R}(\mathbf{Q})$. This implies that the values $\{\alpha_k\}_{k=r+1}^{n}$ have no impact on the objective (20a) of the problem ($\mathsf{P}_{Cvx-TRSP}$) in the case $\mu = 0$. To obtain the solution with the minimal l_2-norm, $\{\alpha_k\}_{k=r+1}^{n}$ should therefore be set to zero. Now from (43), we get

$$\mathbf{x}^\star = \bar{\mathbf{U}}\bar{\mathbf{\Lambda}}^{-1}\bar{\mathbf{p}} = \bar{\mathbf{U}}\bar{\mathbf{\Lambda}}^{-1}\bar{\mathbf{U}}^H\mathbf{q} = \mathbf{U}\mathbf{\Lambda}^\dagger\mathbf{U}^H\mathbf{q} = \mathbf{Q}^\dagger\mathbf{q}, \quad (44)$$

To verify that (44) is indeed the solution, we need to test the KKT conditions. Obviously, $\mu_i = 0$ naturally satisfies all the KKT conditions in (38) except the power constraint. To check power constraint, we utilize (44) to obtain

$$\|\mathbf{x}^\star\|_2^2 = \bar{\mathbf{p}}^H\bar{\mathbf{\Lambda}}^{-1}\bar{\mathbf{U}}^H\bar{\mathbf{U}}\bar{\mathbf{\Lambda}}^{-1}\bar{\mathbf{p}} = \sum_{k=1}^{r}|p_k|^2\lambda_k^{-2} \le \rho^2, \quad (45)$$

where the above inequality follows from the fact that $\sum_{k=1}^{r}|p_i|^2\lambda_k^{-2} \le \rho^2$ when $\mu = 0$. Hence, all the KKT conditions are satisfied and (44) is the optimal solution when $\mu = 0$ with minimum l_2-norm. The proof is complete. $\square$

B. Proof of Lemma 1

Proof: To simplify the following exposition, we introduce notations $\bar{\mathbf{H}} \triangleq \sum_{i=1}^{L}\mathbf{H}_i\bar{\mathbf{F}}_i\mathbf{K}_i$ and $\bar{\mathbf{\Sigma}}_\mathbf{n} \triangleq \sigma_0^2\mathbf{I} + \sum_{i=1}^{L}\mathbf{H}_i\bar{\mathbf{F}}_i\mathbf{\Sigma}_i\bar{\mathbf{F}}_i^H\mathbf{H}_i^H$ Then the relations connecting $\bar{\mathbf{W}}$, $\bar{\mathbf{G}}$ and $\left\{\bar{\mathbf{F}}_i\right\}_{i=1}^{L}$ in (14) and (15) can be written in a compact form as:

$$\bar{\mathbf{G}} = \left[\bar{\mathbf{H}}\mathbf{\Sigma}_\mathbf{s}\bar{\mathbf{H}}^H + \bar{\mathbf{\Sigma}}_\mathbf{n}\right]^{-1}\bar{\mathbf{H}}\mathbf{\Sigma}_\mathbf{s}, \bar{\mathbf{W}} = \bar{\mathbf{H}}^H\bar{\mathbf{\Sigma}}_\mathbf{n}^{-1}\bar{\mathbf{H}} + \mathbf{\Sigma}_\mathbf{s}^{-1}. \quad (46)$$

By (46) we obtain the following identities, whose derivations are shown in (47) and (48) respectively in next page:

$$\bar{\mathbf{G}}\bar{\mathbf{W}}\left(\mathbf{I}-\bar{\mathbf{G}}^H\bar{\mathbf{H}}\right) = \left(\bar{\mathbf{H}}\mathbf{\Sigma}_\mathbf{s}\bar{\mathbf{H}}^H+\bar{\mathbf{\Sigma}}_\mathbf{n}\right)^{-1}\bar{\mathbf{H}}; \quad (49a)$$

$$\bar{\mathbf{G}}\bar{\mathbf{W}}\bar{\mathbf{G}}^H = \left(\bar{\mathbf{H}}\mathbf{\Sigma}_\mathbf{s}\bar{\mathbf{H}}^H+\bar{\mathbf{\Sigma}}_\mathbf{n}\right)^{-1}\bar{\mathbf{H}}\mathbf{\Sigma}_\mathbf{s}\bar{\mathbf{H}}^H\bar{\mathbf{\Sigma}}_\mathbf{n}^{-1}. \quad (49b)$$

Now we calculate $\frac{\partial\widetilde{\mathsf{SR}}}{\partial\mathbf{F}_i^*}\Big|_{\{\bar{\mathbf{F}}_i\}_{i=1}^{L},\bar{\mathbf{G}},\bar{\mathbf{W}}}$ as follows

$$\frac{\partial\widetilde{\mathsf{SR}}}{\partial\mathbf{F}_i^*}\Big|_{\{\bar{\mathbf{F}}_i\}_{i=1}^{L},\bar{\mathbf{G}},\bar{\mathbf{W}}} = \mathbf{H}_i^H\bar{\mathbf{G}}\bar{\mathbf{W}}\bar{\mathbf{G}}^H\mathbf{H}_i\bar{\mathbf{F}}_i\mathbf{\Sigma}_i - \mathbf{H}_i^H\bar{\mathbf{G}}\bar{\mathbf{W}}\Big(\mathbf{I}-\bar{\mathbf{G}}^H\big(\sum_{i=1}^{L}\mathbf{H}_i\bar{\mathbf{F}}_i\mathbf{K}_i\big)\Big)\mathbf{\Sigma}_\mathbf{s}. \quad (50)$$

Substituting the identities (49) into (50), we can rewrite the $\frac{\partial\widetilde{\mathsf{SR}}}{\partial\mathbf{F}_i^*}\Big|_{\{\bar{\mathbf{F}}_i\}_{i=1}^{L},\bar{\mathbf{G}},\bar{\mathbf{W}}}$ associated with only $\{\bar{\mathbf{F}}_i\}_{i=1}^{L}$ as shown in (51) in next page.

Next we check the gradient $\frac{\partial\mathsf{SR}}{\partial\mathbf{F}_i^*}\Big|_{\{\bar{\mathbf{F}}_i\}_{i=1}^{L}}$. By defining $\mathbf{H} \triangleq \sum_{i=1}^{L}\mathbf{H}_i\mathbf{F}_i\mathbf{K}_i$ the derivative of SR with respective to $\mathbf{F}_i$ is shown in (52) in next page, with $C(\mathrm{d}\mathbf{F}_i)$ being terms independent of $\mathrm{d}(\mathbf{F}_i^*)$. Comparing (52b) with (51) we obtain (21b). Similarly (21a) can be proved. $\square$

C. Proof of Theorem 2

Proof: Part of this proof is inspired by [25]. First we prove that the obtained objective sequence for both problems is monotonic. The update of $\mathbf{G}$ and $\mathbf{W}$ will not decrease $\widetilde{\mathsf{SR}}$ since they are obtained by maximizing $\widetilde{\mathsf{SR}}$. We have

$$\widetilde{\mathsf{SR}}(\{\mathbf{F}_i^{(t+1)}\}_{i=1}^{L},\mathbf{G}^{(t+1)},\mathbf{W}^{(t+1)})$$
$$\ge \widetilde{\mathsf{SR}}(\{\mathbf{F}_i^{(t+1)}\}_{i=1}^{L}|\mathbf{G}^{(t)},\mathbf{W}^{(t)}) \overset{(a)}{\ge} -h(\mathbf{f}^{(t+1)};\mathbf{f}^{(t)}|\mathbf{G}^{(t)},\mathbf{W}^{(t)})$$
$$\overset{(b)}{\ge} -h(\mathbf{f}^{(t)};\mathbf{f}^{(t)}|\mathbf{G}^{(t)},\mathbf{W}^{(t)}) \overset{(c)}{=} \widetilde{\mathsf{SR}}(\{\mathbf{F}_i^{(t)}\}_{i=1}^{L},\mathbf{G}^{(t)},\mathbf{W}^{(t)}). \quad (53)$$

where (a) and (c) follows the properties (P1) and (P2) of $h(\cdot)$ respectively, as discussed in subsection III-B, and (b) follows the fact that $\mathbf{f}^{(t+1)}$ is the optimal solution to optimizing $h(\cdot,\mathbf{f}^{(t)}|\mathbf{G}^{(t)},\mathbf{W}^{(t)})$. Thus the $\widetilde{\mathsf{SR}}$ sequence is non-decreasing. Notice that $\mathsf{SR}(\{\mathbf{F}_i^{(t+1)}\}_{i=1}^{L}) = \widetilde{\mathsf{SR}}(\{\mathbf{F}_i^{(t+1)}\}_{i=1}^{L},\mathbf{G}^{(t+1)},\mathbf{W}^{(t+1)})$ since $\mathbf{G}^{(t+1)}$ and $\mathbf{W}^{(t+1)}$ satisfy (15) and (14) with $\{\mathbf{F}_i^{(t)}\}_{i=1}^{L}$ given. Thus SR is non-decreasing. Obviously the sum-rate is bounded, so the objective SR (or equivalently $\widetilde{\mathsf{SR}}$) converges. Denote its limit as $\bar{s}$.

It is not hard to see that the feasible set for the entire beamformers is compact, so there exists at least one subsequence of $\mathbf{f}$ that converges. Since $\mathbf{G}$ and $\mathbf{W}$ are updated by continuous functions of $\mathbf{f}$ in (15) and (14), there exits a convergent subsequence of $\left(\mathbf{f},\mathbf{G},\mathbf{W}\right)$. Assume that $\left(\mathbf{f}^{(t_k)},\mathbf{G}^{(t_k)},\mathbf{W}^{(t_k)}\right)$ is a subsequence converging to $(\bar{\mathbf{f}},\bar{\mathbf{G}},\bar{\mathbf{W}})$. Obviously $\bar{\mathbf{f}}$ is feasible.

$$\begin{aligned}\bar{\mathbf{G}}\bar{\mathbf{W}}\Big(\mathbf{I}-\bar{\mathbf{G}}^H\bar{\mathbf{H}}\Big) &= \big(\bar{\mathbf{H}}\boldsymbol{\Sigma}_{\mathbf{s}}\bar{\mathbf{H}}^H+\bar{\boldsymbol{\Sigma}}_{\mathbf{n}}\big)^{-1}\bar{\mathbf{H}}\boldsymbol{\Sigma}_{\mathbf{s}}\big(\bar{\mathbf{H}}^H\bar{\boldsymbol{\Sigma}}_{\mathbf{n}}^{-1}\bar{\mathbf{H}}+\boldsymbol{\Sigma}_{\mathbf{s}}^{-1}\big)\Big(\mathbf{I}-\boldsymbol{\Sigma}_{\mathbf{s}}\bar{\mathbf{H}}^H\big(\bar{\mathbf{H}}\boldsymbol{\Sigma}_{\mathbf{s}}\bar{\mathbf{H}}^H+\bar{\boldsymbol{\Sigma}}_{\mathbf{n}}\big)^{-1}\bar{\mathbf{H}}\Big)\\ &= \big(\bar{\mathbf{H}}\boldsymbol{\Sigma}_{\mathbf{s}}\bar{\mathbf{H}}^H+\bar{\boldsymbol{\Sigma}}_{\mathbf{n}}\big)^{-1}\bar{\mathbf{H}}\boldsymbol{\Sigma}_{\mathbf{s}}\Big[\underbrace{\bar{\mathbf{H}}^H\bar{\boldsymbol{\Sigma}}_{\mathbf{n}}^{-1}\big(\bar{\mathbf{H}}\boldsymbol{\Sigma}_{\mathbf{s}}\bar{\mathbf{H}}^H+\bar{\boldsymbol{\Sigma}}_{\mathbf{n}}\big)-\big(\bar{\mathbf{H}}^H\bar{\boldsymbol{\Sigma}}_{\mathbf{n}}^{-1}\bar{\mathbf{H}}+\boldsymbol{\Sigma}_{\mathbf{s}}^{-1}\big)\boldsymbol{\Sigma}_{\mathbf{s}}\bar{\mathbf{H}}^H}_{=\mathbf{O}}\Big]\big(\bar{\mathbf{H}}\boldsymbol{\Sigma}_{\mathbf{s}}\bar{\mathbf{H}}^H+\bar{\boldsymbol{\Sigma}}_{\mathbf{n}}\big)^{-1}\bar{\mathbf{H}}\\ &\quad+\big(\bar{\mathbf{H}}\boldsymbol{\Sigma}_{\mathbf{s}}\bar{\mathbf{H}}^H+\bar{\boldsymbol{\Sigma}}_{\mathbf{n}}\big)^{-1}\bar{\mathbf{H}} \end{aligned} \tag{47}$$

$$\begin{aligned}\bar{\mathbf{G}}\bar{\mathbf{W}}\bar{\mathbf{G}}^H &= \big(\bar{\mathbf{H}}\boldsymbol{\Sigma}_{\mathbf{s}}\bar{\mathbf{H}}^H+\bar{\boldsymbol{\Sigma}}_{\mathbf{n}}\big)^{-1}\bar{\mathbf{H}}\boldsymbol{\Sigma}_{\mathbf{s}}\big(\bar{\mathbf{H}}^H\bar{\boldsymbol{\Sigma}}_{\mathbf{n}}^{-1}\bar{\mathbf{H}}+\boldsymbol{\Sigma}_{\mathbf{s}}^{-1}\big)\boldsymbol{\Sigma}_{\mathbf{s}}\bar{\mathbf{H}}^H\big(\bar{\mathbf{H}}\boldsymbol{\Sigma}_{\mathbf{s}}\bar{\mathbf{H}}^H+\bar{\boldsymbol{\Sigma}}_{\mathbf{n}}\big)^{-1}\\ &= \big(\bar{\mathbf{H}}\boldsymbol{\Sigma}_{\mathbf{s}}\bar{\mathbf{H}}^H+\bar{\boldsymbol{\Sigma}}_{\mathbf{n}}\big)^{-1}\Big[\bar{\mathbf{H}}\boldsymbol{\Sigma}_{\mathbf{s}}\bar{\mathbf{H}}^H\bar{\boldsymbol{\Sigma}}_{\mathbf{n}}^{-1}\big(\bar{\mathbf{H}}\boldsymbol{\Sigma}_{\mathbf{s}}\bar{\mathbf{H}}^H+\bar{\boldsymbol{\Sigma}}_{\mathbf{n}}\big)\Big]\big(\bar{\mathbf{H}}\boldsymbol{\Sigma}_{\mathbf{s}}\bar{\mathbf{H}}^H+\bar{\boldsymbol{\Sigma}}_{\mathbf{n}}\big)^{-1} = \big(\bar{\mathbf{H}}\boldsymbol{\Sigma}_{\mathbf{s}}\bar{\mathbf{H}}^H+\bar{\boldsymbol{\Sigma}}_{\mathbf{n}}\big)^{-1}\bar{\mathbf{H}}\boldsymbol{\Sigma}_{\mathbf{s}}\bar{\mathbf{H}}^H\bar{\boldsymbol{\Sigma}}_{\mathbf{n}}^{-1} \end{aligned} \tag{48}$$

$$\begin{aligned}\frac{\partial\widetilde{\mathsf{SR}}}{\partial\mathbf{F}_i^*}\bigg|_{\{\bar{\mathbf{F}}_i\}_{i=1}^L,\bar{\mathbf{G}},\bar{\mathbf{W}}} &= \mathbf{H}_i^H\Big[\big(\sigma_0^2\mathbf{I}+\sum_{i=1}^L\mathbf{H}_i\bar{\mathbf{F}}_i\boldsymbol{\Sigma}_i\bar{\mathbf{F}}_i^H\mathbf{H}_i^H\big)+\big(\sum_{i=1}^L\mathbf{H}_i\bar{\mathbf{F}}_i\mathbf{K}_i\big)\boldsymbol{\Sigma}_{\mathbf{s}}\big(\sum_{i=1}^L\mathbf{H}_i\bar{\mathbf{F}}_i\mathbf{K}_i\big)^H\Big]^{-1}\big(\sum_{i=1}^L\mathbf{H}_i\bar{\mathbf{F}}_i\mathbf{K}_i\big)\boldsymbol{\Sigma}_{\mathbf{s}}\Big[\mathbf{I}\\ &\quad-\big(\sum_{i=1}^L\mathbf{H}_i\bar{\mathbf{F}}_i\mathbf{K}_i\big)^H\big(\sigma_0^2\mathbf{I}+\sum_{i=1}^L\mathbf{H}_i\bar{\mathbf{F}}_i\boldsymbol{\Sigma}_i\bar{\mathbf{F}}_i^H\mathbf{H}_i^H\big)^{-1}\mathbf{H}_i\bar{\mathbf{F}}_i\boldsymbol{\Sigma}_i\Big] \end{aligned} \tag{51}$$

$$\mathrm{d}(\mathsf{SR}) = \mathsf{Tr}\Big\{\mathbf{H}_i^H\big(\boldsymbol{\Sigma}_{\mathbf{n}}+\mathbf{H}\boldsymbol{\Sigma}_{\mathbf{s}}\mathbf{H}^H\big)^{-1}\mathbf{H}\boldsymbol{\Sigma}_{\mathbf{s}}\Big[\mathbf{I}-\mathbf{H}^H\boldsymbol{\Sigma}_{\mathbf{n}}^{-1}\mathbf{H}_i\mathbf{F}_i\boldsymbol{\Sigma}_i\Big]\mathrm{d}(\mathbf{F}_i)^H\Big\}+C(\mathrm{d}\mathbf{F}_i), \tag{52a}$$

$$\begin{aligned}\Rightarrow\frac{\partial\mathsf{SR}}{\partial\mathbf{F}_i^*} &= \mathbf{H}_i^H\Big[\big(\sigma_0^2\mathbf{I}+\sum_{i=1}^L\mathbf{H}_i\mathbf{F}_i\boldsymbol{\Sigma}_i\mathbf{F}_i^H\mathbf{H}_i^H\big)+\big(\sum_{i=1}^L\mathbf{H}_i\mathbf{F}_i\mathbf{K}_i\big)\boldsymbol{\Sigma}_{\mathbf{s}}\big(\sum_{i=1}^L\mathbf{H}_i\mathbf{F}_i\mathbf{K}_i\big)^H\Big]^{-1}\big(\sum_{i=1}^L\mathbf{H}_i\mathbf{F}_i\mathbf{K}_i\big)\boldsymbol{\Sigma}_{\mathbf{s}}\Big[\mathbf{I}\\ &\quad-\big(\sum_{i=1}^L\mathbf{H}_i\mathbf{F}_i\mathbf{K}_i\big)^H\big(\sigma_0^2\mathbf{I}+\sum_{i=1}^L\mathbf{H}_i\mathbf{F}_i\boldsymbol{\Sigma}_i\mathbf{F}_i^H\mathbf{H}_i^H\big)^{-1}\mathbf{H}_i\mathbf{F}_i\boldsymbol{\Sigma}_i\Big],\ i\in\{1,\cdots,L\}. \end{aligned} \tag{52b}$$

By (53), for arbitrary $\mathbf{f}$ we have

$$\begin{aligned} h(\bar{\mathbf{f}};\bar{\mathbf{f}}|\bar{\mathbf{G}},\bar{\mathbf{W}}) &= -\bar{s},\\ h(\mathbf{f};\mathbf{f}^{(t_k)}|\mathbf{G}^{(t_k)},\mathbf{W}^{(t_k)}) &\le h(\mathbf{f}^{(t_k+1)};\mathbf{f}^{(t_k)}|\mathbf{G}^{(t_k)},\mathbf{W}^{(t_k)}). \end{aligned} \tag{54}$$

Taking $k\to\infty$ to both sides of the inequality in (54) we obtain

$$h(\mathbf{f};\bar{\mathbf{f}}|\bar{\mathbf{G}},\bar{\mathbf{W}}) \le -\bar{s},\quad \forall\mathbf{f}. \tag{55}$$

Combining (54) and (55), it is known that $\bar{\mathbf{f}}$ is the optimal solution of ($\mathsf{P}2$) with the objective $h(\mathbf{f};\bar{\mathbf{f}},\bar{\mathbf{G}},\bar{\mathbf{W}})$. As optimality condition, $\bar{\mathbf{f}}$ is stationary, which implies

$$\frac{\partial h}{\partial\mathbf{f}^T}\bigg|_{\bar{\mathbf{f}},\bar{\mathbf{G}},\bar{\mathbf{W}}}\big(\mathbf{f}-\bar{\mathbf{f}}\big)+\frac{\partial h}{\partial\mathbf{f}^H}\bigg|_{\bar{\mathbf{f}},\bar{\mathbf{G}},\bar{\mathbf{W}}}\big(\mathbf{f}^*-\bar{\mathbf{f}}^*\big)\ge 0,\quad\forall\text{feasible }\mathbf{f}. \tag{56}$$

Recall that $h(\cdot)$ is an approximation of $-\widetilde{\mathsf{SR}}(\cdot)$ by introducing second order Taylor expansion at the point of $\bar{\mathbf{f}}$. So $h(\cdot)$ and $-\widetilde{\mathsf{SR}}(\cdot)$ have identical first order derivative at $\bar{\mathbf{f}}$. Thus the (56) can be equivalently written as

$$\begin{aligned}\sum_{i=1}^L\Bigg(\mathsf{Tr}\bigg\{\frac{\partial\widetilde{\mathsf{SR}}}{\partial\mathbf{F}_i^T}\bigg|_{\{\bar{\mathbf{F}}_i\}_{i=1}^L,\bar{\mathbf{G}},\bar{\mathbf{W}}}\Big(\mathbf{F}_i-\bar{\mathbf{F}}_i\Big)\\ +\frac{\partial\widetilde{\mathsf{SR}}}{\partial\mathbf{F}_i^H}\bigg|_{\{\bar{\mathbf{F}}_i\}_{i=1}^L,\bar{\mathbf{G}},\bar{\mathbf{W}}}\Big(\mathbf{F}_i^*-\bar{\mathbf{F}}_i^*\Big)\bigg\}\Bigg)\le 0,\ \forall\text{feasible }\{\mathbf{F}_i\}_{i=1}^L. \end{aligned} \tag{57}$$

Notice that $\mathbf{G}^{(j)}$ and $\mathbf{W}^{(j)}$ are obtained by (15) and (14) with $\{\mathbf{F}_i^{(j)}\}_{i=1}^L$ given. Therefore, by taking $j\to\infty$, $\big(\{\bar{\mathbf{F}}_i^{(j)}\}_{i=1}^L,\bar{\mathbf{G}},\bar{\mathbf{W}}\big)$ satisfies (15) and (14). Combining the results in Lemma 1 and (57), we obtain

$$\begin{aligned}\sum_{i=1}^L\Bigg(\mathsf{Tr}\bigg\{\frac{\partial\mathsf{SR}}{\partial\mathbf{F}_i^T}\bigg|_{\{\bar{\mathbf{F}}_i\}_{i=1}^L}\Big(\mathbf{F}_i-\bar{\mathbf{F}}_i\Big)\\ +\frac{\partial\mathsf{SR}}{\partial\mathbf{F}_i^H}\bigg|_{\{\bar{\mathbf{F}}_i\}_{i=1}^L}\Big(\mathbf{F}_i^*-\bar{\mathbf{F}}_i^*\Big)\bigg\}\Bigg)\le 0,\ \forall\text{feasible }\{\mathbf{F}_i\}_{i=1}^L, \end{aligned} \tag{58}$$

which implies $\{\bar{\mathbf{F}}_i\}_{i=1}^L$ are in fact stationary points for ($\mathsf{P}0$). □

D. Proof of Theorem 3

Proof: First we show that Alg.5 guarantees a monotonically decreasing transmission power sequence.

Denote transmission power function as $\mathsf{P_T}(\mathbf{f})$. Since $\mathbf{G}$ and $\mathbf{W}$ are updated by maximizing $\widetilde{\mathsf{SR}}$, $\widetilde{\mathsf{SR}}(\mathbf{f}^{(j)},\mathbf{G}^{(j)},\mathbf{W}^{(j)}) \geq \widetilde{\mathsf{SR}}(\mathbf{f}^{(j)},\mathbf{G}^{(j)},\mathbf{W}^{(j-1)}) \geq \widetilde{\mathsf{SR}}(\mathbf{f}^{(j)},\mathbf{G}^{(j-1)},\mathbf{W}^{(j-1)}) \geq r_0$, so $\mathbf{f}^{(j)}$ is feasible to (P5) with $\mathbf{G}$ and $\mathbf{W}$ being given as $\mathbf{G}^{(j)}$ and $\mathbf{W}^{(j)}$, which immediately implies $\mathsf{P_T}(\mathbf{f}^{(j+1)}) \leq \mathsf{P_T}(\mathbf{f}^{(j)})$ since $\mathbf{f}^{(j+1)}$ is obtained by optimizing (P6) with $\mathbf{G}^{(j)}$ and $\mathbf{W}^{(j)}$ given.

Since the feasible region of transmitters is bounded due to the power constraints, there exits a subsequence $\{k_j\}_{j=1}^{\infty}$ such that $\{\mathbf{F}_i^{(k_j)}\}_{i=1}^L$ converges. Since $\mathbf{G}$ and $\mathbf{W}$ are obtained by continuous functions of $\{\mathbf{F}_i\}_{i=1}^L$ in (15) and (14), $\left(\{\mathbf{F}_i^{(k_j)}\}_{i=1}^L,\mathbf{W}^{(k_j)},\mathbf{G}^{(k_j)}\right)$ converges. Consequently the existence of limit points of $\left(\{\mathbf{F}_i^{(k)}\}_{i=1}^L,\mathbf{W}^{(k)},\mathbf{G}^{(k)}\right)$ can be proved.

Now suppose that $\left(\{\bar{\mathbf{F}}_i\}_{i=1}^L,\bar{\mathbf{W}},\bar{\mathbf{G}}\right)$ is an arbitrary limit point of $\left(\{\mathbf{F}_i^{(k)}\}_{i=1}^L,\mathbf{W}^{(k)},\mathbf{G}^{(k)}\right)$, which means there exists a subsequence $\{k_j\}$ such that $\left(\{\mathbf{F}_i^{(k_j)}\}_{i=1}^L,\mathbf{W}^{(k_j)},\mathbf{G}^{(k_j)}\right) \overset{j\to\infty}{\longrightarrow} \left(\{\bar{\mathbf{F}}_i\}_{i=1}^L,\bar{\mathbf{W}},\bar{\mathbf{G}}\right)$. Since $\{\mathbf{F}_i^{(k)}\}_{i=1}^L$ is bounded, by possibly restricting to a subsequence, we assume that $\{\mathbf{F}_i^{(k_j+1)}\}_{i=1}^L$ converges to a limit $\left\{\widehat{\mathbf{F}}_i\right\}_{i=1}^L$.

Notice that the following inequalities hold $\forall j$

$$\widetilde{\mathsf{SR}}\left(\{\mathbf{F}_i^{(k_j+1)}\}_{i=1}^L\big|\mathbf{G}^{(k_j)},\mathbf{W}^{(k_j)}\right) \geq r_0, \tag{59a}$$

$$\mathsf{Tr}\{\mathbf{F}_i^{(k_j+1)}\left(\mathbf{K}_i\boldsymbol{\Sigma}_\mathbf{s}\mathbf{K}_i^H+\boldsymbol{\Sigma}_i\right)\left(\mathbf{F}_i^{(k_j+1)}\right)^H)\} \leq P_i,\ \forall i \tag{59b}$$

since $\{\mathbf{F}_i^{(k_j+1)}\}_{i=1}^L$ is obtained by solving (P5) with $\mathbf{G}^{(k_j)}$ and $\mathbf{W}^{(k_j)}$ given. By taking $j\to\infty$ in the above inequalities, we obtain

$$\widetilde{\mathsf{SR}}\left(\{\bar{\mathbf{F}}_i\}_{i=1}^L\big|\bar{\mathbf{G}},\bar{\mathbf{W}}\right) \geq r_0, \tag{60a}$$

$$\mathsf{Tr}\{\widehat{\mathbf{F}}_i\left(\mathbf{K}_i\boldsymbol{\Sigma}_\mathbf{s}\mathbf{K}_i^H+\boldsymbol{\Sigma}_i\right)\widehat{\mathbf{F}}_i^H)\} \leq P_i, i\in\{1,\cdots,L\}. \tag{60b}$$

So $\left\{\widehat{\mathbf{F}}_i\right\}_{i=1}^L$ is feasible to (P5) with $\bar{\mathbf{G}}$ and $\bar{\mathbf{W}}$ given.

For arbitrary $\left\{\mathbf{F}_i\right\}_{i=1}^L$ which is feasible to (P5) with $\bar{\mathbf{G}}$ and $\bar{\mathbf{W}}$ associated, we have

$$\mathsf{P_T}\left(\{\mathbf{F}_i\}_{i=1}^L\right) \geq \mathsf{P_T}\left(\{\mathbf{F}_i^{(k_j+1)}\}_{i=1}^L\right). \tag{61}$$

Taking $j\to\infty$ in the above equation we have

$$\mathsf{P_T}\left(\left\{\mathbf{F}_i\right\}_{i=1}^L\right) \geq \mathsf{P_T}\left(\widehat{\mathbf{F}}_i\}_{i=1}^L\right). \tag{62}$$

The above inequality indicates that $\left\{\widehat{\mathbf{F}}_i\right\}_{i=1}^L$ is actually optimal solution to (P5) with $\bar{\mathbf{W}}$ and $\bar{\mathbf{G}}$ given. At the same time, since the $\mathsf{P_T}$ sequence monotonically converges, we have

$$\mathsf{P_T}\left(\{\widehat{\mathbf{F}}_i\}_{i=1}^L\right) = \mathsf{P_T}\left(\{\bar{\mathbf{F}}_i\}_{i=1}^L\right) = \mathsf{P_T}\left(\{\mathbf{F}_i^{(\infty)}\}_{i=1}^L\right) \tag{63}$$

Combining (63) and (62), we readily see that $\{\bar{\mathbf{F}}_i\}_{i=1}^L$ is actually an optimal solution to the problem (P5) with parameters $\bar{\mathbf{W}}$ and $\bar{\mathbf{G}}$ given.

$$\sum_{i=1}^{L}\left(\mathsf{Tr}\left\{\frac{\partial\mathsf{P_T}}{\partial\mathbf{F}_i^T}\bigg|_{\{\bar{\mathbf{F}}_i\}_{i=1}^L}\left(\mathbf{F}_i-\bar{\mathbf{F}}_i\right)+\frac{\partial\mathsf{P_T}}{\partial\mathbf{F}_i^H}\bigg|_{\{\bar{\mathbf{F}}_i\}_{i=1}^L}\left(\mathbf{F}_i^*-\bar{\mathbf{F}}_i^*\right)\right\}\right)\geq 0, \tag{64}$$

for all feasible $\{\mathbf{F}_i\}_{i=1}^L$ to (P5). Recall that (P1) and (P5) have identical feasibility region of $\{\mathbf{F}_i\}_{i=1}^L$. Thus we have indeed proved that $\{\bar{\mathbf{F}}_i\}_{i=1}^L$ is stationary to (P1). □

References

[1] I. F. Akyildiz, T. Melodia, and K. R. Chowdhury, "A survey on wireless multimedia sensor networks," *Comput. Netw. J.*, vol. 51, no. 4, pp. 921–960, 2007.

[2] A. Marcus and O. Marques, "An eye on visual sensor networks," *IEEE Potentials*, vol. 31, no. 2, pp. 38–43, Mar./Apr. 2012.

[3] I. D. Schizas, G. B. Giannakis, and Z. Q. Luo, "Distributed estimation using reduced-dimensionality sensor observations," *IEEE Trans. Signal Process.*, vol. 55, no. 8, pp. 4284–4299, Aug. 2007.

[4] 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.

[5] A. S. Behbahani, A. M. Eltawil, and H. Jafarkhani, "Linear decentralized estimation of correlated data for power-constrained wireless sensor networks," *IEEE Trans. Signal Process.*, vol. 60, no. 11, pp. 6003–6016, Nov. 2012.

[6] 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.

[7] F. Jiang, J. Chen, A. Lee 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.

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

[9] Y. Liu, T. J. Li, X. Lu, and C. Yuen, "Optimal linear precoding and postcoding for MIMO multi-sensor noisy observation problem," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Sydney, NSW, Australia, Jun. 2014, pp. 5598–5603.

[10] J. Fang, H. Li, Z. Chen, and Y. Gong, "Joint precoder design for distributed transmission of correlated sources in sensor networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 6, pp. 2918–2929, Jun. 2013.

[11] S. Sudevalayam and P. Kulkarni, "Energy harvesting sensor nodes: Survey and implications," *IEEE Commun. Surveys Tuts.*, vol. 13, no. 3, pp. 443–461, 3rd Quart., 2011.

[12] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, Mar. 2008.

[13] G. Dán, M. A. Khan, and V. Fodor, "Characterization of SURF and BRISK interest point distribution for distributed feature extraction in visual sensor networks," *IEEE Trans. Multimedia*, vol. 17, no. 5, pp. 591–602, May 2015.

[14] Z. Hasan, H. Boostanimehr, and V. K. Bhargava, "Green cellular networks: A survey, some research issues and challenges," *IEEE Commun. Surveys Tuts.*, vol. 13, no. 4, pp. 524–540, 4th Quart., 2011.

[15] H. Ren, N. Liu, C. Pan, and L. Hanzo, "Joint fronthaul link selection and transmit precoding for energy efficiency maximization of multiuser MIMO-aided distributed antenna systems," *IEEE Trans. Commun.*, vol. 65, no. 12, pp. 5180–5196, Dec. 2017, doi: 10.1109/TCOMM.2017.2728526.

[16] Z. Wang and L. Vandendorpe, "Subcarrier allocation and precoder design for energy efficient MIMO-OFDMA downlink systems," *IEEE Trans. Commun.*, vol. 65, no. 1, pp. 136–146, Jan. 2017.

[17] Y. Li, Y. Tian, and C. Yang, "Energy-efficient coordinated beamforming under minimal data rate constraint of each user," *IEEE Trans. Veh. Technol.*, vol. 64, no. 6, pp. 2387–2397, Jun. 2015.

[18] O. Tervo, L.-N. Tran, and M. Juntti, "Optimal energy-efficient transmit beamforming for multi-user MISO downlink," *IEEE Trans. Signal Process.*, vol. 63, no. 20, pp. 5574–5588, Oct. 2015.

[19] C. Xing, S. Li, Z. Fei, and J. Kuang, "How to understand linear minimum mean-square-error transceiver design for multiple-input-multiple-output systems from quadratic matrix programming," *IET Commun.*, vol. 7, no. 12, pp. 1231–1242, Aug. 2013.

[20] K. T. Truong, P. Sartori, and R. W. Heath, "Cooperative algorithms for MIMO amplify-and-forward relay networks," *IEEE Trans. Signal Process.*, vol. 61, no. 5, pp. 1272–1287, Mar. 2013.

[21] C. Sun and E. A. Jorswieck, "Multi-stream sum rate maximization for MIMO AF relay networks," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)* Brisbane, QLD, Australia, May 2015, pp. 4434–4438.

[22] S. S. Christensen, R. Agarwal, E. De Carvalho, and J. M. Cioffi, "Weighted sum-rate maximization using weighted MMSE for MIMO-BC beamforming design," *IEEE Trans. Wireless Commun.*, vol. 7, no. 12, pp. 4792–4799, Dec. 2008.

[23] 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.

[24] İ. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecommun.*, vol. 10, no. 6, pp. 585–595, 1999.

[25] M. Razaviyayn, M. Hong, and Z.-Q. Luo, "A unified convergence analysis of block successive minimization methods for nonsmooth optimization," *SIAM J. Optim.*, vol. 23, no. 2, pp. 1126–1153, 2013.

[26] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, NY, USA: Cambridge Univ. Press, 2004.

[27] M. Grant and S. Boyd. (Apr. 2010). *CVX: MATLAB Software for Disciplined Convex Programming*. [Online]. Available: http://cvxr.com/cvx

[28] I. Pólik and T. Terlaky, "Interior point methods for nonlinear optimization," in *Nonlinear Optimization*, G. Di Pillo and F. Schoen, Eds., 1st ed. New York, NY, USA: Springer, 2010.

[29] W. Dinkelbach, "On nonlinear fractional programming," *Manage. Sci.*, vol. 13, no. 7, pp. 492–498, Mar. 1967.

[30] J. Nocedal and S. J. Wright, *Numerical Optimization*. 2nd ed. New York, NY, USA: Springer, 2006.

[31] D. Bertsekas and J. Tsitsiklis, *Parallel and Distributed Computing*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1989.

[32] M. S. Bazaraa, H. D. Sherali, and C. M. Shetty, *Nonlinear Programming: Theory and Algorithms*, 3rd ed. New York, NY, USA: Wiley, 2005.

[33] S. Boyd. *Primal and Dual Decomposition*. Accessed: Jan. 22, 2018. [Online]. Available: http://web.stanford.edu/class/ee364b/lectures

[34] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. New York, NY, USA: Wiley, 2006.

Yang Liu received the B.E. and M.E. degrees in electrical engineering from the Beijing University of Posts and Telecommunications, Beijing, China, in 2007 and 2010, respectively, and the Ph.D. degree in electrical engineering from Lehigh University, Bethlehem, PA, USA, in 2016. His research interests include analog error correction coding, beamforming in MIMO communications and wireless sensor networks.

Jing (Tiffany) Li received the B.S. degree in computer science from Peking University, Beijing, China, and the master's and Ph.D. degrees in electrical engineering from Texas A&M University, College Station, TX, USA. She joined the Electrical and Computer Engineering Department, Lehigh University, in 2003, and was promoted to Associate Professor in 2008. Her research interests include the general area of coding and information theory, wireless communication networks, and data storage systems. She was a recipient of the Lehigh University P. C. Rossin Assistant Professorship in 2005 and the TAMU Ethel Ashworth-Tsutsui Memorial Award for Research in 2001.She was the IEEE COMSOC Data Storage Technical Committee Chair from 2011 to 2012, and an IEEE Distinguished Lecturer from 2012 to 2013. She served as the Symposium Co-Chair in the Chinacom 2015, ICC 2013, WCSP 2012, ICC 2008, Chinacom 2006, GLOBECOM 2005, and Wirelesscom 2005. She also served as an Editor for the IEEE TRANSACTION ON WIRELESS COMMUNICATIONS from 2008 to 2011, an Associate Editor for the IEEE sc Communications Letters from 2004 to 2008, and an Editorial Board Member for the IEEE Communications surveys and Tutorials from 2008 to 2010.