Robust Linear Beamforming in Wireless Sensor Networks

Yang Liu, Jing Li, and Hao Wang, *Member, IEEE*

***Abstract*— A typical wireless sensor network (WSN) has multiple sensors, each of which obtains noisy observation of one common physical event and sends the observation to a fusion center (FC) for post-processing. This paper provides a comprehensive research on robust linear transceiver design in the presence of imperfect channel-state information (CSI) in multi-input multi-output (MIMO) WSNs. For the CSI uncertainty, the classical ellipsoid model is assumed, where channel estimation errors live in an ellipsoid. Two robust beamforming problems are considered: 1) the worst-case mean-square-error (MSE) minimization with limited power and 2) sum-power minimization with guaranteed worst-case MSE. For both problems, centralized and decentralized solutions are developed. Moreover, an interesting fundamental relation between the metrics of signal-to-noise ratio (SNR) and MSE has been presented, which extends our proposed solutions to handle the transceiver design when SNR is considered. Extensive numerical results are presented to confirm our findings.**

***Index Terms*— Wireless sensor network (WSN), robust, linear beamforming, alternative direction method of multipliers (ADMM), decentralized algorithm.**

I. Introduction

WIRELESS sensor networks (WSN) have a wide range of applications, such as environmental monitoring, battle-field surveillance and so on [1]. A wireless sensor network usually has numerous sensors observing one common physical event. The observed data at the sensors are usually noisy. Sensors convey the noisy observation to a fusion center (FC), where post-processing and fusion of the data are performed. Assume that both the sensors and FC are equipped with multiple antennas and linear transceivers to perform communications. Then a central question is how to jointly design the linear transceivers to convey and fuse the data reliably and efficiently.

A. Related Works

A good number of existing works study the linear beamforming design problems in the context of WSN from different aspects [2]–[17], [21], [22]. The seminal work [2] introduces a very generic model of WSN and provides solutions to several special cases. Fang and Li [3] study the linear beamforming scheme for scalar source under one total power constraint. Behbahani *et al.* [4] study the best linear unbiased estimation (BLUE) for scalar source and scalar beamformers with one total power constraint. A closed form BLUE is obtained in [5] for non-constrained power scenario. Behbahani *et al.* [6] and Liu *et al.* [7] have developed efficient beamforming design methods to minimize the mean square error (MSE) suitable to the generic settings proposed by Xiao *et al.* [2]. Besides MSE, throughput is also a critical concern from information viewpoint and linear beamformers are designed to maximize sum-rate in [8]–[10]. Recently, numerous novel applications of WSN empowered with new features have risen. For instance, sequential estimation utilizing Kalman filtering technique for real-time tracking is researched in [11] and [12]. Signal estimation/detection problem in WSN with FC equipped with great number of antennas (i.e. "massive MIMO") is considered in [13] and [14]. Beamforming design in WSN with capability of energy harvesting are studied in [15] and [16]. Collaborations with communication cost between sensors are researched in [16] and [17]. Besides, inspiring work on distributive target detection utilizing WSN have also been reported in [18]–[20].

All these aforementioned researches rely on accurate channel state information (CSI). In fact, perfect CSI assumption is non-trivial and communicating devices usually have to work in the presence of CSI errors. The uncertainty of the CSI may come from different sources, e.g. imperfect channel estimation, limited bandwidth for feedback and the processing/feedback delay for CSI update. Due to the CSI uncertainty, the transceivers designed for specific channel state may become unreliable. Therefore taking into account the CSI errors is important in the transceiver design. Two very recent works [21], [22] have focused on robust beamforming design against uncertain CSI in WSN.

Venkategowda *et al.* [21] and Rostami and Falahati [22] consider robust MSE minimization under ellipsoid and statistical CSI model respectively. Although meaningful and inspiring, the research in [21] and [22] are performed via casting some specialized constraints onto the most generic model in existing literature, e.g. [2]. Specifically, Venkategowda *et al.* [21] consider precoders that separately amplify observation samples (i.e. precoders of sensors are all diagonal matrices instead of arbitrary shapes). Both Venkategowda *et al.* [21] and Rostami and Falahati [22] have considered total sum power constraint. Though the total sum

Manuscript received July 31, 2018; revised January 1, 2019 and February 17, 2019; accepted February 19, 2019. Date of publication March 5, 2019; date of current version June 14, 2019. The associate editor coordinating the review of this paper and approving it for publication was F. Verde. *(Corresponding author: Yang Liu.)*

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

H. Wang is with the School of Information Science and Technology, ShanghaiTech University, Shanghai 201210, China (e-mail: wanghao1@shanghaitech.edu.cn).

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.2019.2903251

power simplifies the problem formulation, it is less practical compared to the separate power constraint at each sensor, considering the fact that sensors are usually geographically distributed. Additionally both Venkategowda *et al.* [21] and Rostami and Falahati [22] have introduced a distortionless (i.e. zero forcing (ZF)) constraint to simply the post-processing at FC, which is generally not considered in classical literature on WSN [2]–[17], [21], [22] since ZF equalizer magnifies noise and harms MSE/SNR [38]. Besides, robust design to minimize power subject to quality of service (QoS) constraint is not discussed in [21] and [22].

B. Contributions

As discussed above, a comprehensive treatment of robust linear beamforming design applicable to the most generic MIMO WSN model is still missing and desirable, which is the very goal of this article. Specifically the contributions of this article is highlighted as follows:

(i) Based on the bounded ellipsoid model of the uncertainty in CSI, we design a centralized algorithm via semi-definite programming (SDP) techniques. Compared to the conventional robust linear beamforming in wireless MIMO amplify-and-forward (AF) relay systems [23]–[25] or MIMO multi-user/interference networks [26], [28]–[30], [32], where the quadratic terms in objective/constraints are generally involved with only one CSI uncertainty and therefore the *S-Procedure* could be utilized straightforwardly, one major difficulty here in coherent-sum multiple-access channel (MAC) of WSN is that the uncertainty of all sensor-FC channels linearly superimpose at the receiver and therefore the worst-case MSE performance could not be easily formulated into an analytic form. To overcome this difficulty, we derive a tight upper bound of the worst-case MSE, which effectively decouples each sensor-FC channel and consequently leads to an SDP formulation, which can be effectively solved via numerical solvers.

(ii) Furthermore, this article succeeds in developing a decentralized solution. Noticing the fact that the centralized SDP formulation has an aggressively growing complexity in the number of sensors and easily becomes prohibitively computation-demanding, we develop a decentralized algorithm based on the alternating direction method of multipliers (ADMM) framework [34]. It is worth noting that decentralized methods still have not yet been fully studied in context of WSN so far [1]–[6], [8], [9], [11]–[17], [21], [22] (except a recent work [10], which develops a distributive solution based on a framework other than ADMM). Our decentralized algorithm decomposes the entire optimization of all sensors' beamformers into independent sub-problems, which can be computed distributively and in parallel. Our distributive reformulation effectively decreases the solution's computation complexity and processing capability requirement, and make it applicable to much larger networks. The strength of this decentralized method has been fully consolidated by our complexity analysis and numerical experiments.

(iii) Besides, we have presented an interesting fundamental relation between the metrics of SNR and MSE for WSN context, which holds whether CSI is perfectly known or not. This result significantly extends the primitive conclusion in [33], where a preliminary point-to-point multiple-input-single-output (MISO) scenario is considered, to the more generic case with multi-antennas deployed at both ends in WSN settings. The significance of this relation lies in that it extends our MSE-oriented solution to the scenarios where SNR objective and/or constraints are considered. This indeed makes our solution a unifying one for robust linear beamforming, whether MSE or SNR is considered. The numerical results in Sec.VII verifies the strength of our proposals when SNR is taken into account.

C. Outline of This Article

The rest of the paper is organized as follows. Section II introduces the system model. In Section III and IV, under the ellipsoidal CSI assumption, robust linear beamforming design to optimize MSE and sum-power are elaborated, with centralized and decentralized solutions discussed. In Section V, a fundamental relation between the metrics of SNR and MSE is presented, based on which the robust transceivers for MSE metric can also be useful for SNR problem. Section VI compares the centralized and decentralized solutions from the aspect of complexity and communication costs. Extensive simulation results are presented in Sec.VII to verify our proposals. Section VIII concludes the whole article.

D. Notations

We use bold lowercase letters to denote complex vectors and bold capital letters to denote complex matrices. $\mathbf{0}$, $\mathbf{O}_{m\times n}$ and $\mathbf{I}_m$ are used to denote zero vectors, zero matrices of dimension $m\times n$, and identity matrices of order m respectively. $\mathbf{A}^T$, $\mathbf{A}^*$ and $\mathbf{A}^H$ are used to denote transpose, conjugate and conjugate transpose (Hermitian transpose) respectively of a complex matrix $\mathbf{A}$. $\mathsf{Tr}\{\cdot\}$ denotes the trace operation of a square matrix. $|\cdot|$ denotes the modulus of a complex scalar. $\|\cdot\|_2$ and $\|\cdot\|_F$ denote the l_2-norm and Frobenius norm of a complex matrix respectively. $\mathsf{vec}(\cdot)$ means column-vectorization operation of a matrix. $\otimes$ denotes the Kronecker product. $\mathsf{Diag}\{\mathbf{A}_1,\cdots,\mathbf{A}_n\}$ denotes the block diagonal matrix with its i-th diagonal block being $\mathbf{A}_i$, $i\in\{1,\cdots,n\}$. $\mathsf{Re}\{x\}$ and $\mathsf{Im}\{x\}$ denote the real and imaginary part of a complex value x, respectively.

II. System Model and Problems

In this section we introduce the system model of WSN as shown in Fig.1. The network comprises of a FC and L sensors. Each sensor is labeled with one distinctive integer from the set $\mathcal{L}\triangleq\{1,\cdots,L\}$. All sensors monitor some common physical event, which is modeled as an unknown complex parameter $\theta\in\mathbb{C}$. Without loss of generality we assume that $\mathsf{E}\{\theta\}=0$ and $\mathsf{E}\{|\theta|^2\}=1$.

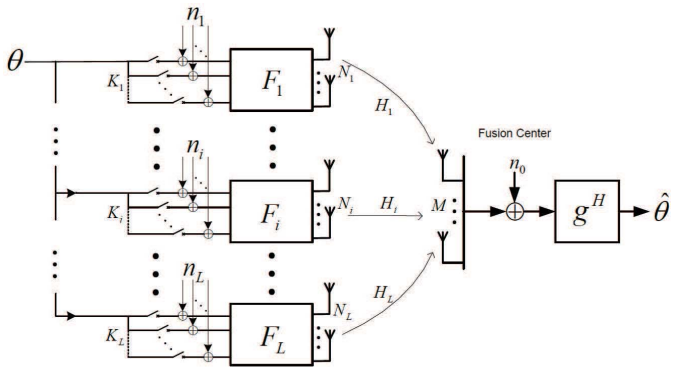

Fig. 1. Models for Wireless Sensor Network.

Suppose that each sensor can obtain K_i samples of the parameter to be measured, which is disturbed by the environmental or thermal noise. The obtained noisy observation by the i-th sensor is presented as

$$\mathbf{x}_i = \mathbf{1}_{K_i}\theta + \mathbf{n}_i, \quad i \in \mathcal{L}, \tag{1}$$

where K_i is number of samples and $\mathbf{n}_i$ represents the observation noise. Without loss of generality we assume that $\mathbf{n}_i$ has zero mean and covariance matrix $\mathsf{E}\{\mathbf{n}_i\mathbf{n}_i^H\} = \mathbf{\Sigma}_i \succ 0$. Each sensor transmits its noisy observation to the FC. In this article we focus on linear precoders and postcoders since linear filters are easy for analysis and practical implementation. Suppose that all sensors are equipped with multi-antennas and linear precoders. We denote the precoding matrix at the i-th sensor as $\mathbf{F}_i \in \mathbb{C}^{N_i \times K_i}$ with N_i being the number of antennas. Assume the FC utilizes a linear filter $\mathbf{g}$ to process the collected signals from sensors. Denote the channel matrix between the i-th sensor to the FC as $\mathbf{H}_i \in \mathbb{C}^{M \times N_i}$, where M is the number of antennas at FC. For brevity, we denote $\boldsymbol{\mathcal{F}} \triangleq \{\mathbf{F}_i\}_{i=1}^L$ and $\boldsymbol{\mathcal{H}} \triangleq \{\mathbf{H}_i\}_{i=1}^L$ in the subsequent discussions.

Here the coherent-sum MAC is assumed to achieve a high time/bandwidth efficiency, which can be practically viewed as a virtual (or distributed) MIMO system. The received signal $\mathbf{r}$ at the FC can be expressed as

$$\mathbf{r} = \sum_{i=1}^{L}\Big(\mathbf{H}_i\mathbf{F}_i\mathbf{x}_i\Big) + \mathbf{n}_0 \tag{2}$$

$$= \sum_{i=1}^{L}\Big(\mathbf{H}_i\mathbf{F}_i\mathbf{1}_{K_i}\Big)\theta + \Big(\sum_{i=1}^{L}\mathbf{H}_i\mathbf{F}_i\mathbf{n}_i + \mathbf{n}_0\Big), \tag{3}$$

where $\mathbf{n}_0$ is the additive noise at the FC. Without loss of generality, we assume $\mathbf{n}_0$ has zero mean and covariance matrix $\sigma_0^2\mathbf{I}_M$. Since sensors are usually geographically distributed, the noise $\mathbf{n}_i$ at different sensors can be assumed mutually uncorrelated.

The FC performs signal recovery via linear filtering and obtains an estimate/detect $\hat{\theta}$ of θ:

$$\hat{\theta} = \mathbf{g}^H\mathbf{r} = \Big(\mathbf{g}^H\sum_{i=1}^{L}\mathbf{H}_i\mathbf{F}_i\mathbf{1}_{K_i}\Big)\theta + \mathbf{g}^H\Big(\sum_{i=1}^{L}\mathbf{H}_i\mathbf{F}_i\mathbf{n}_i + \mathbf{n}_0\Big). \tag{4}$$

According to different natures of the target signal in various WSN applications, different performance metrics are considered. For estimation of signals coming from continuous alphabets, the mean square error (MSE) is a standard metric of general interest, which is given as

$$\begin{aligned}\mathsf{MSE}(\boldsymbol{\mathcal{H}}, \boldsymbol{\mathcal{F}}, \mathbf{g}) &= \mathbb{E}\Big\{\big|\theta - \hat{\theta}\big|^2\Big\} \\ &= \Big|1 - \mathbf{g}^H\Big(\sum_{i=1}^{L}\mathbf{H}_i\mathbf{F}_i\mathbf{1}_{K_i}\Big)\Big|^2 \\ &\quad + \sum_{i=1}^{L}\mathbf{g}^H\mathbf{H}_i\mathbf{F}_i\mathbf{\Sigma}_i\mathbf{F}_i^H\mathbf{H}_i^H\mathbf{g} + \sigma_0^2\|\mathbf{g}\|_2^2, \end{aligned}\tag{5}$$

where the assumption of uncorrelated noise has been invoked.

Besides MSE, signal to noise ratio (SNR) is also a meaningful metric, in which the channel throughput and symbol error rate (SER) can be presented in monotonic functions [36]. SNR is expressed in (6) as follows

$$\begin{aligned}&\mathsf{SNR}(\boldsymbol{\mathcal{H}}, \boldsymbol{\mathcal{F}}, \mathbf{g}) \\ &= \frac{\mathbb{E}\Big\{\Big|\Big(\mathbf{g}^H\sum_{i=1}^{L}\mathbf{H}_i\mathbf{F}_i\mathbf{1}_{K_i}\Big)\theta\Big|^2\Big\}}{\mathbb{E}\Big\{\Big|\mathbf{g}^H\Big(\sum_{i=1}^{L}\mathbf{H}_i\mathbf{F}_i\mathbf{n}_i + \mathbf{n}_0\Big)\Big|^2\Big\}} \\ &= \frac{\mathbf{g}^H\big[\sum_{i=1}^{L}\mathbf{H}_i\mathbf{F}_i\mathbf{1}_{K_i}\big]\big[\sum_{j=1}^{L}\mathbf{H}_j\mathbf{F}_j\mathbf{1}_{K_j}\big]^H\mathbf{g}}{\sigma_0^2\|\mathbf{g}\|_2^2 + \sum_{i=1}^{L}\mathbf{g}^H\mathbf{H}_i\mathbf{F}_i\mathbf{\Sigma}_i\mathbf{F}_i^H\mathbf{H}_i^H\mathbf{g}}, \end{aligned}\tag{6}$$

Power consumption is also a highly concerned metric. The average transmission power of the i-th sensor can be expressed as $\mathbb{E}\{\mathsf{Tr}\{\mathbf{F}_i(\mathbf{1}_{K_i}\theta + \mathbf{n}_i)(\mathbf{1}_{K_i}\theta + \mathbf{n}_i)^H\mathbf{F}_i^H\}\} = \mathsf{Tr}\{\mathbf{F}_i(\mathbf{1}_{K_i}\mathbf{1}_{K_i}^H + \mathbf{\Sigma}_i)\mathbf{F}_i^H\}\}$.

To consider the CSI uncertainty, there exist two pervasive ways to model CSI—the statistical model [26], [27] and the ellipsoid model [28]–[30]. The former model assumes knowledge of the first and second order statistics of the CSI and robust design aims at optimizing the performance in a statistical average sense. The ellipsoid uncertainty assumes that channel state errors live in an ellipsoid and the robust design aims at optimizing the worst-case performance. Moreover the bounded error model can also be utilized to serve as a conservative constraint when statistics of the errors are available, which guarantees that the outage probability will not exceed some predefined value [31]. In this article, we consider ellipsoid CSI model.

III. Robust MSE Minimization Problem

In this section, we consider robust MSE minimization. The channel state can be modeled as [28]–[30]:

$$\mathbf{H}_i = \widehat{\mathbf{H}}_i + \mathbf{E}_i, \quad i \in \mathcal{L}, \tag{7}$$

where $\mathbf{E}_i$ is the CSI error. Defining $\mathbf{e}_i \triangleq \mathsf{vec}(\mathbf{E}_i)$, the region where CSI errors live is an ellipsoid represented as $\{\mathbf{e}_i | \mathbf{e}_i^H\mathbf{Q}_i\mathbf{e}_i \le \epsilon_i^2\}$, where the constant matrix $\mathbf{Q}_i$ is positive definite. To abbreviate notations, we denote $\mathcal{X}_i \triangleq \{\mathbf{H}_i | \widehat{\mathbf{H}}_i + \mathbf{E}_i, \forall \mathbf{E}_i \text{ s.t. } \mathsf{vec}^H(\mathbf{E}_i)\mathbf{Q}_i\mathsf{vec}(\mathbf{E}_i) \le \epsilon_i^2\}$ and $\mathcal{Y} \triangleq \Pi_{i=1}^L\mathcal{X}_i$.

A. Centralized Solution

For ellipsoid CSI uncertainties, the robust transceiver design aims at optimizing the worst-case MSE, which can be formulated as

$$(\mathsf{P1}): \quad \min_{\boldsymbol{\mathcal{F}},\mathbf{g}} \max_{\boldsymbol{\mathcal{H}}\in\mathcal{Y}} \mathsf{MSE}(\boldsymbol{\mathcal{H}},\boldsymbol{\mathcal{F}},\mathbf{g}), \tag{8a}$$

$$\text{s.t. } \mathsf{Tr}\big\{\mathbf{F}_i\big(\mathbf{1}_{K_i}\mathbf{1}_{K_i}^H+\boldsymbol{\Sigma}_i\big)\mathbf{F}_i^H\big\}\big\}\le P_i, \quad i\in\mathcal{L}. \tag{8b}$$

The major difficulty to solve (P1) is the inner maximization of MSE over all possible channel realizations cannot be expressed explicitly. To overcome this difficulty, we turn to minimize an upper bound of the worst-case MSE. For the MSE maximization in the objective, the following inequality stands,

$$\max_{\boldsymbol{\mathcal{H}}} \mathsf{MSE}(\boldsymbol{\mathcal{H}},\boldsymbol{\mathcal{F}},\mathbf{g}) \tag{9}$$

$$= \max_{\boldsymbol{\mathcal{H}}} \Big\{\Big|1-\mathbf{g}^H\Big(\sum_{i=1}^{L}\mathbf{H}_i\mathbf{F}_i\mathbf{1}_{K_i}\Big)\Big|^2+\sigma_0^2\|\mathbf{g}\|_2^2 + \sum_{i=1}^{L}\mathbf{g}^H\mathbf{H}_i\mathbf{F}_i\boldsymbol{\Sigma}_i\mathbf{F}_i^H\mathbf{H}_i^H\mathbf{g}\Big\} \tag{10}$$

$$\le \max_{\boldsymbol{\mathcal{H}}} \Big\{\Big|1-\mathbf{g}^H\big(\sum_{i=1}^{L}\mathbf{H}_i\mathbf{F}_i\mathbf{1}_{K_i}\big)\Big|^2\Big\}+\sigma_0^2\|\mathbf{g}\|_2^2 + \sum_{i=1}^{L}\max_{\mathbf{H}_i} \Big\{\mathbf{g}^H\mathbf{H}_i\mathbf{F}_i\boldsymbol{\Sigma}_i\mathbf{F}_i^H\mathbf{H}_i^H\mathbf{g}\Big\}. \tag{11}$$

To simplify the above equations, we have the following results (which is proved in Appendix A):

Lemma 1: Define the following notations

$$\boldsymbol{s}_i(\boldsymbol{F}_i,\boldsymbol{g}) \triangleq \mathsf{vec}(\boldsymbol{g}^*\boldsymbol{1}_{K_i}^T\boldsymbol{F}_i^T), \tag{12a}$$

$$s_0(\boldsymbol{\mathcal{F}},\boldsymbol{g}) \triangleq 1-\boldsymbol{g}^H\big(\sum_{i=1}^{L}\widehat{\boldsymbol{H}}_i\boldsymbol{F}_i\boldsymbol{1}_{K_i}\big). \tag{12b}$$

Then the first maximization in (11) can be evaluated as

$$\max_{\boldsymbol{\mathcal{H}}\in\mathcal{Y}} \Big\{\Big|1-\boldsymbol{g}^H\big(\sum_{i=1}^{L}\boldsymbol{H}_i\boldsymbol{F}_i\boldsymbol{1}_{K_i}\big)\Big|^2\Big\} = \Big(|s_0|+\sum_{i=1}^{L}\epsilon_i\big\|\boldsymbol{Q}_i^{-\frac{*}{2}}\boldsymbol{s}_i^*\big\|_2\Big)^2.$$

Next we proceed to study the second maximization term in (11). In fact, the value $\max_{\mathbf{H}_i\in\mathcal{X}_i}\mathbf{g}^H\mathbf{H}_i\mathbf{F}_i\boldsymbol{\Sigma}_i\mathbf{F}_i^H\mathbf{H}_i^H\mathbf{g}$ equals the minimal value of p_i^2 with $p_i\ge 0$ such that

$$\big\|\boldsymbol{\Sigma}_i^{\frac{1}{2}}\mathbf{F}_i^H\mathbf{H}_i^H\mathbf{g}\big\|_2^2\le p_i^2, \quad \forall\mathbf{H}_i\in\mathcal{X}_i. \tag{13}$$

The above condition can be further transformed equivalently as follows [45]

$$\begin{bmatrix} p_i & \mathbf{g}^H\mathbf{H}_i\mathbf{F}_i\boldsymbol{\Sigma}_i^{\frac{1}{2}} \\ \boldsymbol{\Sigma}_i^{\frac{1}{2}}\mathbf{F}_i^H\mathbf{H}_i^H\mathbf{g} & p_i\mathbf{I}_{K_i}\end{bmatrix}\succcurlyeq 0, \quad \forall\mathbf{H}_i\in\mathcal{X}_i \tag{14}$$

$$\Leftrightarrow \begin{bmatrix} p_i & \mathbf{g}^H\widehat{\mathbf{H}}_i\mathbf{F}_i\boldsymbol{\Sigma}_i^{\frac{1}{2}} \\ \boldsymbol{\Sigma}_i^{\frac{1}{2}}\mathbf{F}_i^H\widehat{\mathbf{H}}_i^H\mathbf{g} & p_i\mathbf{I}_{K_i}\end{bmatrix} + \begin{bmatrix} 0 & \mathbf{g}^H\mathbf{E}_i\mathbf{F}_i\boldsymbol{\Sigma}_i^{\frac{1}{2}} \\ \boldsymbol{\Sigma}_i^{\frac{1}{2}}\mathbf{F}_i^H\mathbf{E}_i^H\mathbf{g} & \mathbf{O}\end{bmatrix}\succcurlyeq 0, \quad \forall\mathbf{e}_i^H\mathbf{Q}_i\mathbf{e}_i\le\epsilon_i^2, \tag{15}$$

Notice the above inequality should be satisfied by infinite number of channel realizations. The following result is useful to express the above constraint in an analytic manner.

*Lemma 2 (*S-Procedure *[30]): Assume that $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{C}$ are known matrices and $\boldsymbol{A}$ is Hermitian. Then the following inequality*

$$\boldsymbol{A}\succcurlyeq\boldsymbol{B}^H\boldsymbol{D}\boldsymbol{C}+\boldsymbol{C}^H\boldsymbol{D}^H\boldsymbol{B}, \quad \forall\|\boldsymbol{D}\|_2\le\epsilon \tag{16}$$

stands if and only if

$$\exists\lambda\ge 0, \quad \begin{bmatrix}\boldsymbol{A}-\lambda\boldsymbol{B}^H\boldsymbol{B} & -\epsilon\boldsymbol{C}^H \\ -\epsilon\boldsymbol{C} & \lambda\boldsymbol{I}\end{bmatrix}\succcurlyeq\boldsymbol{O}. \tag{17}$$

To utilize Lemma 2, we notice the following identity

$$\mathbf{g}^H\mathbf{E}_i\mathbf{F}_i\boldsymbol{\Sigma}_i^{\frac{1}{2}} = \mathsf{vec}^T\big(\mathbf{g}^H\mathbf{E}_i\mathbf{F}_i\boldsymbol{\Sigma}_i^{\frac{1}{2}}\big) = \mathbf{e}_i^T\Big[\big(\mathbf{F}_i\boldsymbol{\Sigma}_i^{\frac{1}{2}}\big)\otimes\mathbf{g}^*\Big],$$

where the last equality utilizes the relation $\mathsf{vec}\big(\mathbf{ABC}\big) = \big[\mathbf{C}^T\otimes\mathbf{A}\big]\mathsf{vec}\big(\mathbf{B}\big)$. By denoting $\mathbf{w}_i \triangleq \big(\mathbf{F}_i\boldsymbol{\Sigma}_i^{\frac{1}{2}}\big)\otimes\mathbf{g}^*$ the condition (15) can be further transformed into

$$-\begin{bmatrix}1\\ \mathbf{0}\end{bmatrix}\mathbf{e}_i^T\mathbf{Q}_i^{\frac{*}{2}}\big[\mathbf{O},\mathbf{Q}_i^{-\frac{*}{2}}\mathbf{w}_i\big] - \begin{bmatrix}\mathbf{O}\\ \mathbf{w}_i^H\mathbf{Q}_i^{-\frac{*}{2}}\end{bmatrix}\mathbf{Q}^{\frac{*}{2}}\mathbf{e}_i^*\big[1,\mathbf{0}^T\big] \preccurlyeq \begin{bmatrix} p_i & \mathbf{g}^H\widehat{\mathbf{H}}_i\mathbf{F}_i\boldsymbol{\Sigma}_i^{\frac{1}{2}} \\ \boldsymbol{\Sigma}_i^{\frac{1}{2}}\mathbf{F}_i^H\widehat{\mathbf{H}}_i^H\mathbf{g} & p_i\mathbf{I}_{K_i}\end{bmatrix}, \quad \forall\mathbf{e}_i^H\mathbf{Q}_i\mathbf{e}_i\le\epsilon_i^2 \tag{18}$$

Invoking Lemma 2 by identifying the $\mathbf{B}$, $\mathbf{C}$ and $\mathbf{D}$ in Lemma 2 with $\mathbf{B}=\big[\mathbf{O},\mathbf{Q}_i^{-\frac{*}{2}}\mathbf{w}_i\big]$, $\mathbf{C}=[-1,\mathbf{0}^T]$, $\mathbf{D}=\mathbf{Q}_i^{\frac{*}{2}}\mathbf{e}_i^*$ and $\mathbf{A}$ as the right hand side of the inequality in (18), the value $\max_{\mathbf{H}_i\in\mathcal{X}_i}\mathbf{g}^H\mathbf{F}_i\boldsymbol{\Sigma}_i\mathbf{F}_i^H\mathbf{H}_i^H\mathbf{g}$ can be given as the minimal value of p_i^2 satisfying that $\exists\ w_i\ge 0$ such that

$$\mathbf{W}_i\big(\mathbf{F}_i,\mathbf{g},p_i,w_i\big) \tag{19}$$

$$\triangleq \begin{bmatrix} p_i-w_i & \mathbf{g}^H\widehat{\mathbf{H}}_i\mathbf{F}_i\boldsymbol{\Sigma}_i^{\frac{1}{2}} & \mathbf{0}^T \\ \boldsymbol{\Sigma}_i^{\frac{*}{2}}\mathbf{F}_i^H\widehat{\mathbf{H}}_i^H\mathbf{g} & p_i\mathbf{I}_{K_i} & -\epsilon_i\mathbf{w}_i^H\mathbf{Q}_i^{-\frac{*}{2}} \\ \mathbf{0} & -\epsilon_i\mathbf{Q}_i^{-\frac{*}{2}}\mathbf{w}_i & w_i\mathbf{I}_{MN_i}\end{bmatrix}\succcurlyeq 0, \tag{20}$$

By replacing the original worst-case MSE with the upper bound in (11) and following the results in Lemma 1 and the S-procedure transformation in (19), we can formulate the robust MSE minimization problem as follows

$$(\mathsf{P2}): \quad \min_{\substack{\boldsymbol{\mathcal{F}},\mathbf{g},s_0,t,\\ \mathbf{w},\mathbf{p},\mathbf{q}}} \ t, \tag{21a}$$

$$\text{s.t. } \Big(|s_0|+\sum_{i=1}^{L}q_i\Big)^2+\sum_{i=1}^{L}p_i^2+\sigma_0^2\big\|\mathbf{g}\big\|_2\le t^2, \tag{21b}$$

$$s_0 = 1-\mathbf{g}^H\big(\sum_{i=1}^{L}\widehat{\mathbf{H}}_i\mathbf{F}_i\mathbf{1}_{K_i}\big), \tag{21c}$$

$$\epsilon_i\big\|\mathbf{s}_i^T\big(\mathbf{F}_i,\mathbf{g}\big)\mathbf{Q}_i^{-\frac{1}{2}}\big\|_2\le q_i, \quad i\in\mathcal{L}, \tag{21d}$$

$$\mathbf{W}_i\big(\mathbf{F}_i,\mathbf{g},p_i,w_i\big)\succcurlyeq 0, \quad i\in\mathcal{L}, \tag{21e}$$

$$\big\|\mathbf{F}_i\big(\mathbf{1}_{K_i}\mathbf{1}_{K_i}^H+\boldsymbol{\Sigma}_i\big)^{\frac{1}{2}}\big\|_F\le\sqrt{P_i}, \quad i\in\mathcal{L}, \tag{21f}$$

$$p_i\ge 0, \quad w_i\ge 0,\ t\ge 0,\ i\in\mathcal{L} \tag{21g}$$

where $\mathbf{w} \triangleq [w_1, \cdots, w_L]^T$, $\mathbf{p} \triangleq [p_1, \cdots, p_L]^T$ and $\mathbf{q} \triangleq [q_1, \cdots, q_L]^T$. The notations $\mathbf{s}_i(\mathbf{F}_i, \mathbf{g})$ and $\mathbf{W}_i(\mathbf{F}_i, \mathbf{g}, p_i, w_i)$ are defined in Lemma 1 and (19) respectively.

The following lemma, which is proved in Appendix B, further simplifies the problem (P2).

Lemma 3: (P2) can be equivalently written as (P3).

$$
\begin{aligned}
(\mathsf{P3}): \quad \min_{\boldsymbol{\mathcal{F}},\mathbf{g},s_0,t,\atop \mathbf{w},\mathbf{p},\mathbf{q}} \ & t, && (22a)\\
\text{s.t.} \ & \Big(s_0 + \sum_{i=1}^{L} q_i\Big)^2 + \sum_{i=1}^{L} p_i^2 + \sigma_0^2 \|\mathbf{g}\|_2 \le t^2, && (22b)\\
& s_0 = 1 - \mathsf{Re}\{\mathbf{g}^H\big(\sum_{i=1}^{L} \widehat{\mathbf{H}}_i \mathbf{F}_i \mathbf{1}_{K_i}\big)\}, && (22c)\\
& \mathsf{Im}\{\mathbf{g}^H\big(\sum_{i=1}^{L} \widehat{\mathbf{H}}_i \mathbf{F}_i \mathbf{1}_{K_i}\big)\} = 0, && (22d)\\
& s_0 \ge 0, && (22e)\\
& (21d - 21g)
\end{aligned}
$$

Although the problem (P3) is nonconvex for the entire variables jointly, it is convex when $\boldsymbol{\mathcal{F}}$ or $\mathbf{g}$ is fixed. Thus we can solve the problem in an alternative minimization manner, with each step optimizing either $\boldsymbol{\mathcal{F}}$ or $\mathbf{g}$ together with other variables. According to the result in [39], this two-block BCD algorithm can guarantee that any limit point of the solution sequence is stationary to (P3).

It is worth noting that although we assume that the first and second order statistics of the target θ is perfectly known, our solution can also account for the uncertainty in these statistical knowledge. To see this, suppose that $\mathbb{E}\{\theta\} = \zeta$ and $\mathbb{E}\{|\theta|^2\} = \varepsilon^2$. Suppose that we have uncertainty in the knowledge of ζ and ε^2, i.e. $l_0 \le \zeta \le u_0$ and $l_1 \le \varepsilon^2 \le u_1$. Then the MSE can be written as

$$
\begin{aligned}
\mathsf{MSE}(\boldsymbol{\mathcal{H}}, \boldsymbol{\mathcal{F}}, \mathbf{g}) &= \mathbb{E}\Big\{\big|\theta - \hat{\theta}\big|^2\Big\} \\
&= \varepsilon^2 \Big|1 - \mathbf{g}^H\big(\sum_{i=1}^{L} \mathbf{H}_i \mathbf{F}_i \mathbf{1}_{K_i}\big)\Big|^2 \\
&\quad + \sum_{i=1}^{L} \mathbf{g}^H \mathbf{H}_i \mathbf{F}_i \boldsymbol{\Sigma}_i \mathbf{F}_i^H \mathbf{H}_i^H \mathbf{g} + \sigma_0^2 \|\mathbf{g}\|_2^2,
\end{aligned}
$$

Note that MSE function is not dependent on ζ. The worst-case MSE can be given as follows:

$$
\begin{aligned}
&\max_{\substack{\boldsymbol{\mathcal{H}} \in \mathcal{Y}, \\ l_1 \le \varepsilon^2 \le u_1}} \ \mathsf{MSE}(\boldsymbol{\mathcal{H}}, \boldsymbol{\mathcal{F}}, \mathbf{g}, \varepsilon^2) && (23)\\
&= \max_{\boldsymbol{\mathcal{H}} \in \mathcal{Y}} \Bigg\{ u_1 \Big|1 - \mathbf{g}^H\big(\sum_{i=1}^{L} \mathbf{H}_i \mathbf{F}_i \mathbf{1}_{K_i}\big)\Big|^2 && (24)\\
&\quad + \sum_{i=1}^{L} \mathbf{g}^H \mathbf{H}_i \mathbf{F}_i \boldsymbol{\Sigma}_i \mathbf{F}_i^H \mathbf{H}_i^H \mathbf{g} + \sigma_0^2 \|\mathbf{g}\|_2^2 \Bigg\},
\end{aligned}
$$

where the worst-case MSE in (24) can be determined by the same lines as (9-11).

B. Decentralized Solution

One potential shortcoming of the above centralized SDP solution is the complexity, which can easily go prohibitively high when the number of sensors is large (see Tab.II). In the sequel we discuss a decentralized formulation to optimize the transmitters $\boldsymbol{\mathcal{F}}$ (with $\mathbf{g}$ given) distributively based on the ADMM framework [34].

Towards this end, we need first to transfer the problem (P3) into an form more amenable to applying ADMM. Note that the constraints (21d-21f) are already beamformer-decoupled, i.e. each of these constraints merely depends on one specific $\mathbf{F}_i$. In contrast, (22b-22d) are coupled since each of these constraint is relevant to all $\mathbf{F}_i$'s. Firstly we introduce the intermediate variables $u_i = \mathsf{Re}\{\mathbf{g}^H \widehat{\mathbf{H}}_i \mathbf{F}_i \mathbf{1}_{K_i}\}$ and $v_i = \mathsf{Im}\{\mathbf{g}^H \widehat{\mathbf{H}}_i \mathbf{F}_i \mathbf{1}_{K_i}\}$, which transfers the coupling variables from the high-dimension $\{\mathbf{F}_i\}_{i=1}^L$ to scalars $\{u_i\}_1^L$ and $\{v_i\}_1^L$ and significantly reduces the communication cost. Denote $\mathbf{u} \triangleq [u_1, \cdots, u_L]^T$ and $\mathbf{v} \triangleq [v_1, \cdots, v_L]^T$. Secondly, the coupling variables (which we refer as *global variables*) s_0, t, $\{u_i\}_{i=1}^L$, $\{v_i\}_{i=1}^L$, $\{p_i\}_{i=1}^L$ and $\{q_i\}_{i=1}^L$ should have a local "copy" (namely the *local variables*) at each computation unit (e.g. sensor) to enable distributive computation. Therefore we define these local copies of s_0, t, u_j, v_j, p_j and q_j at the i-th computation unit as s_i, t_i, u_{ij}, v_{ij}, p_{ij} and q_{ij} respectively. To simplify notations, we define $\mathbf{s} \triangleq [s_1, \cdots, s_L]^T$, $\mathbf{t} \triangleq [t_1, \cdots, t_L]^T$, $\mathbf{u}_i \triangleq [u_{i1}, \cdots, u_{iL}]^T$, $\mathbf{v}_i \triangleq [v_{i1}, \cdots, v_{iL}]^T$, $\mathbf{p}_i \triangleq [p_{i1}, \cdots, p_{iL}]^T$, $\mathbf{q}_i \triangleq [q_{i1}, \cdots, q_{iL}]^T$. Then the above problem (P3) has been transformed into the following form

$$
\begin{aligned}
(\mathsf{P4}): \quad \min_{\substack{\boldsymbol{\mathcal{F}},\mathbf{w},s_0,t,\mathbf{u},\mathbf{v},\mathbf{p},\mathbf{q} \\ \mathbf{s},\mathbf{t},\{\mathbf{u}_i\}_{i=1}^L,\{\mathbf{v}_i\}_{i=1}^L,\{\mathbf{p}_i\}_{i=1}^L,\{\mathbf{q}_i\}_{i=1}^L}} \ & \frac{1}{L}\sum_{i}^{L} t_i, && (25a)\\
\text{s.t.} \ & \Big\| s_i + \sum_{j=1}^{L} q_{ij}, \big[p_{i1}, \cdots, p_{iL}\big], \\
& \qquad \sigma_0 \mathbf{g}^T \Big\|_2 \le t_i, \quad i \in \mathcal{L}, && (25b)\\
& u_{ii} = \mathsf{Re}\{\mathbf{g}^H \widehat{\mathbf{H}}_i \mathbf{F}_i \mathbf{1}_{K_i}\}, \quad i \in \mathcal{L}, && (25c)\\
& v_{ii} = \mathsf{Im}\{\mathbf{g}^H \widehat{\mathbf{H}}_i \mathbf{F}_i \mathbf{1}_{K_i}\}, \quad i \in \mathcal{L}, && (25d)\\
& s_i = 1 - \sum_{j=1}^{L} u_{ij}, \ \sum_{j=1}^{L} v_{ij} = 0, \ i \in \mathcal{L}, && (25e)\\
& \epsilon_i \big\| \mathbf{s}_i^T(\mathbf{F}_i, \mathbf{g}) \mathbf{Q}_i^{-\frac{1}{2}} \big\|_2 \le q_{ii}, \quad i \in \mathcal{L}, && (25f)\\
& \mathbf{W}_i(\mathbf{F}_i, \mathbf{g}, p_{ii}, w_i) \succcurlyeq 0, \quad i \in \mathcal{L}, && (25g)\\
& \big\| \mathbf{F}_i \big(\mathbf{1}_{K_i} \mathbf{1}_{K_i}^H + \boldsymbol{\Sigma}_i\big)^{\frac{1}{2}} \big\|_F \le \sqrt{P_i}, \quad i \in \mathcal{L}. && (25h)\\
& s_{ij} \ge 0, \quad p_{ij} \ge 0, \ i, j \in \mathcal{L}, && (25i)\\
& s_0 = s_i, \quad t = t_i, \ i \in \mathcal{L}, && (25j)\\
& u_{ij} = u_j, \quad v_{ij} = v_j, \ p_{ij} = p_j, \ q_{ij} = q_j, \ i, j \in \mathcal{L}. && (25k)
\end{aligned}
$$

We pack all the global variables into one vector $\mathbf{x} \triangleq [s, t, \boldsymbol{u}^T, \boldsymbol{v}^T, \boldsymbol{p}^T, \boldsymbol{q}^T]^T$. Similarly, one local copy of $\mathbf{x}$ at the i-th computation unit can be denoted as $\mathbf{z}_i \triangleq [s_i, t_i, \mathbf{u}_i^T, \mathbf{v}_i^T, \mathbf{p}_i^T, \mathbf{q}_i^T]^T$. Besides we define $\mathbf{z} \triangleq [\mathbf{z}_1^T, \cdots, \mathbf{z}_L^T]^T$ as the entire local variables among all L computation units.

Note the constraints (25j) and (25k) actually ensure the connections that each local variable $\mathbf{z}_i$ should be identical with the global variable $\mathbf{x}$. Follow the notations of $\mathbf{x}$ and $\mathbf{z}$ as above, the constraints (25j) and (25k) can actually be expressed in a compact form $\mathbf{z} = \mathbf{J}\mathbf{x}$ with the constant matrix $\mathbf{J}$ being defined as $\mathbf{J} \triangleq [\mathbf{I}_1, \cdots, \mathbf{I}_L]^T$, where each $\mathbf{I}_i$ is an identity matrix with the order of $(4L+2)$ for $i \in \mathcal{L}$.

Follow the above notations, ADMM framework considers the augmented problem (P5) as follows:

$$(\mathsf{P5}): \quad \min_{\boldsymbol{\mathcal{F}},\mathbf{w},\mathbf{x},\mathbf{z}} . \ \frac{1}{L}\sum_{i}^{L} t_i + \frac{\kappa}{2}\|\mathbf{J}\mathbf{x}-\mathbf{z}\|, \quad \text{s.t. } \mathbf{z} = \mathbf{J}\mathbf{x}, \quad (25b-25i) \tag{26}$$

where the penalty parameter κ is a positive constant. It is worth noting that the augmented problem (P5) is indeed equivalent to the original problem (P4).

Before applying ADMM, we first introduce the Lagrangian multipliers λ_i, μ_i, ν_{ij}, τ_{ij}, θ_{ij} and η_{ij} associated with the equality constraints in (25j) and (25k) involving t_i, s_i, p_{ij}, q_{ij}, u_{ij} and v_{ij} respectively and denote them accordingly as $\boldsymbol{\lambda} \triangleq [\lambda_1, \cdots, \lambda_L]^T$, $\boldsymbol{\mu} \triangleq [\mu_1, \cdots, \mu_L]^T$, $\boldsymbol{\nu}_i \triangleq [\nu_{i1}, \cdots, \nu_{iL}]^T$, $\boldsymbol{\tau}_i \triangleq [\tau_{i1}, \cdots, \tau_{iL}]^T$, $\boldsymbol{\theta}_i \triangleq [\theta_{i1}, \cdots, \theta_{iL}]^T$ and $\boldsymbol{\eta}_i \triangleq [\eta_{i1}, \cdots \eta_{iL}]^T$. For easy exposition, we pack all Lagrangian multipliers into one vector $\boldsymbol{\xi} \triangleq [\boldsymbol{\lambda}^T, \boldsymbol{\mu}^T, \boldsymbol{\nu}_1^T, \cdots, \boldsymbol{\nu}_L^T, \boldsymbol{\tau}_1^T, \cdots, \boldsymbol{\tau}_L^T, \boldsymbol{\theta}_1^T, \cdots, \boldsymbol{\theta}_L^T, \boldsymbol{\eta}_1^T, \cdots, \boldsymbol{\eta}_L^T]^T$, which is indeed associated with the constraint $\mathbf{z} = \mathbf{J}\mathbf{x}$.

According to [34], ADMM solves the original problem (P4) via alternatively updating the primal and the dual of the augmented problem (P5), whose dual is (P6) written as follows:

$$(\mathsf{P6}): \quad \max_{\boldsymbol{\xi}} . \left\{ \min_{\substack{\boldsymbol{\mathcal{F}},\mathbf{w},\mathbf{x},\mathbf{z}, \\ (25b-25i)}} . \ \frac{1}{L}\sum_{i}^{L} t_i + \boldsymbol{\xi}^T\big(\mathbf{J}\mathbf{x}-\mathbf{z}\big) + \frac{\kappa}{2}\big\|\mathbf{J}\mathbf{x}-\mathbf{z}\big\|_2^2 \right\}$$

Precisely, in each iteration, the ADMM method solves the primal augmented problem in a Gauss-Seidel manner via separately updating the primal local variables $\{\mathbf{z}, \mathcal{F}, \mathbf{w}\}$ and the primal global variables $\mathbf{x}$, and then updates the Lagrangian multipliers $\boldsymbol{\xi}$.

For the local variable $\{\mathbf{z}, \mathcal{F}, \mathbf{w}\}$ update, $\mathbf{x}$ and $\boldsymbol{\xi}$ are being fixed. Define the feasible region of the localized variable $\{\mathbf{z}_i, \mathbf{f}_i, w_i\}$ as $\mathcal{Z}_i \triangleq \{\mathbf{z}_i, \mathbf{f}_i, w_i | (25b-(25i) \text{ stand}\}, \forall i \in \mathcal{L}$. Then it can be readily seen that the update of local variable $\{\mathbf{z}, \mathcal{F}, \mathbf{w}\}$ can decomposed into L independent smaller subproblems in parallel with the i-th sub-problem given as

$$(\mathsf{P7}^i): \quad \min_{\{\mathbf{z}_i,\mathbf{f}_i,w_i\}\in\mathcal{Z}_i} . \ \frac{1}{L}t_i + \boldsymbol{\xi}_i^T\big(\mathbf{y}_i-\mathbf{z}_i\big) + \frac{\kappa}{2}\big\|\mathbf{y}_i-\mathbf{z}_i\big\|_2^2 \tag{27}$$

where $\mathbf{y}_i \triangleq \big[\mathbf{J}\mathbf{x}\big]_{(4L+2)(i-1)+1:(4L+2)i}$.

When $\{\mathbf{z}, \mathbf{f}, \mathbf{w}\}$ and $\boldsymbol{\xi}$ are being fixed, the update of $\mathbf{x}$ can be obtained in a closed form as

$$\mathbf{x} = (\mathbf{J}^T\mathbf{J})^{-1}\mathbf{J}^T(\mathbf{z} - \kappa^{-1}\boldsymbol{\xi}). \tag{28}$$

$\boldsymbol{\xi}$ is updated in a gradient ascent manner with step size κ

$$\boldsymbol{\xi} := \boldsymbol{\xi} + \kappa(\mathbf{J}\mathbf{x} - \mathbf{z}). \tag{29}$$

Intuitively, the $\{\mathbf{z}, \mathbf{f}, \mathbf{w}\}$ update performs the decentralized optimization with respect to each local variable, the $\mathbf{x}$ update calibrates the global variables via averaging all its local copies (namely the *consensus* step). The steps of ADMM method to solve (P4) with fixed $\mathbf{g}$ are summarized in Alg.1.

Algorithm 1 ADMM to Opt. (P3) w.r.t. $\boldsymbol{\mathcal{F}}$ ($\mathbf{g}$ Is Fixed)

Initialization:Initialize $\mathbf{x}^{(0)}$; $\boldsymbol{\xi}^{(0)} = \mathbf{0}$; $j = 0$;
repeat
  Solve $(\mathsf{P7}^i)$ $i \in \mathcal{L}$ in parallel, obtain $\mathbf{z}^{(j+1)}$;
  Update $\mathbf{x}^{(j+1)}$ by (28);
  Update $\boldsymbol{\xi}^{(j+1)}$ by (29); $j++$;
until *convergence*;

According to [42, Proposition 4.2], optimal solution to (P4) can be obtained as long as $\mathbf{J}^T\mathbf{J}$ is invertible. Notice that $\mathbf{J}^T\mathbf{J} = L\mathbf{I}_{(4L+2)} \succ 0$. Consequently ADMM method will converge to an optimal $\boldsymbol{\mathcal{F}}$ of (P4) with $\mathbf{g}$ fixed.

The algorithm minimizing MSE for ellipsoidal CSI, including centralized and decentralized method, is summarized in Alg.2 as follows

Algorithm 2 MSE Minimization for Bounded CSI Errors

Initialization:Generate $\mathbf{x}^{(0)}$; $\boldsymbol{\xi}^{(0)} = \mathbf{0}$; $j = 0$;
repeat
  fix $\mathbf{g}^{(j)}$, solve (P3) by SDP solver or the ADMM method in Alg.1 to update $\mathbf{f}^{(j+1)}$;
  fix $\mathbf{f}^{(j+1)}$, solve (P3) to obtain $\mathbf{g}^{(j+1)}$;
until *convergence*;

IV. Robust Sum-Power Minimization Problem

In this section, the sum-power problem is considered for ellipsoidal CSI model. The robust design aims at minimizing the average total transmission power while ensuring the MSE in the worst-case is below a target level ϵ_0, i.e.

$$(\mathsf{P8}): \quad \min_{\boldsymbol{\mathcal{F}},\mathbf{g}} . \ \mathbf{f}^H\mathbf{E}\mathbf{f}, \tag{30a}$$

$$\text{s.t. } \max_{\boldsymbol{\mathcal{H}}\in\mathcal{Y}} . \mathsf{MSE}(\boldsymbol{\mathcal{H}}, \boldsymbol{\mathcal{F}}, \mathbf{g}) \le \epsilon_0, \tag{30b}$$

A. Centralized Solution

To tackle the above problem, one key step is to express the non-analytic worst-case MSE constraint (30b) in an explicit form. As discussed previously, one practical solution is to replace the intractable worst-case MSE constraint with the upper bound developed in (11), which can be expressed in an analytic form. Then the sum-power minimization problem becomes

$$(\mathsf{P9}): \quad \min_{\boldsymbol{\mathcal{F}},\mathbf{g},\mathbf{p},s_0} . \ \mathbf{f}^H\mathbf{E}\mathbf{f}, \tag{31a}$$

$$\text{s.t. } \left||s_0| + \sum_{i=1}^{L}\epsilon_i\big\|\mathbf{s}_i\mathbf{Q}_i^{-\frac{1}{2}}\big\|_2\right|^2 + \sum_{i=1}^{L} p_i^2 + \sigma_0^2\big\|\mathbf{g}\big\|_2^2 \le \epsilon_0,$$

$$\big\|\boldsymbol{\Sigma}_i^{\frac{1}{2}}\mathbf{F}_i^H\mathbf{H}_i^H\mathbf{g}\big\|_2^2 \le p_i^2, \quad \forall \mathbf{H}_i \in \mathcal{X}_i, \ i \in \mathcal{L}. \tag{31b}$$

Note that (P9) actually yields a restriction of the original problem (P8), i.e. (P9) indeed shrinks the feasible region of (P8). This can be readily understood via noticing the fact that any feasible solution of (P9) is naturally feasible to (P8), but not vice versa. Therefore the optimal value of (P9) actually serves as an upper bound of that of (P8). As will be shown in Fig.3, the worst-case MSE are usually very close to the optimal value of (P9). This result will convince us that in fact the optimal value of (P9) is usually very close to that of (P8).

Following Lemma 1-3 and similar reasoning in the last section, (P9) can be equivalently written as

$$(\mathsf{P9}'): \quad \min_{\substack{\boldsymbol{\mathcal{F}},\mathbf{g},s_0,\mathbf{t}\\ \mathbf{w},\mathbf{p},\mathbf{q}}} \ \|\mathbf{t}\|_2, \tag{32a}$$

$$\text{s.t.} \ \Big\| s_0+\sum_{i=1}^{L} q_i, \big[p_1,\cdots,p_L\big], \sigma_0\mathbf{g}^T \Big\|_2 \le \sqrt{\epsilon_0}, \tag{32b}$$

$$\big\|\mathbf{F}_i\big(\mathbf{1}_{K_i}\mathbf{1}_{K_i}^H+\boldsymbol{\Sigma}_i\big)^{\frac{1}{2}}\big\|_F \le t_i, \quad i\in\mathcal{L}. \tag{32c}$$

$$(21d),(21e),(21g),(22c),(22d),(22e)$$

where $\mathbf{t}\triangleq[t_1,\cdots,t_L]^T$ and other notations are identical with those defined previously.

Since the problem (P9′) is biconvex in $\boldsymbol{\mathcal{F}}$ and $\mathbf{g}$, we can still alternatively optimize $\boldsymbol{\mathcal{F}}$ and $\mathbf{g}$, as illustrated in Alg.3. Since Alg.3 is still a two-block BCD algorithm, it generates decreasing objective value and can guarantee its limit points of solutions to be stationary to (P9) [39].

B. Decentralized Solution

An ADMM-based decentralized computation can also apply to the sum-power minimization problem. In fact we can rewrite (P9′) (with $\mathbf{g}$ being fixed) into an equivalent form as follows

$$(\mathsf{P10}): \quad \min_{\substack{\boldsymbol{\mathcal{F}},\mathbf{s},\mathbf{t},\\ \{\mathbf{u}_i\}_{i=1}^L,\{\mathbf{v}_i\}_{i=1}^L,\{\mathbf{p}_i\}_{i=1}^L,\{\mathbf{q}_i\}_{i=1}^L}} \ \sum_{i}^{L} t_i^2, \tag{33a}$$

$$\text{s.t.} \ \Big\| s_i+\sum_{j=1}^{L} q_{ij}, \big[p_{i1},\cdots,p_{iL}\big], \sigma_0\mathbf{g}^T \Big\|_2 \le \sqrt{\epsilon_0},$$

$$i\in\mathcal{L}, \tag{33b}$$

$$\big\|\mathbf{F}_i\big(\mathbf{1}_{K_i}\mathbf{1}_{K_i}^H+\boldsymbol{\Sigma}_i\big)^{\frac{1}{2}}\big\|_F \le t_i, \quad i\in\mathcal{L}. \tag{33c}$$

$$(25c-25g),(25i-25k).$$

Here we define the primal variables $\mathbf{s},\mathbf{t},\mathbf{u}_i,\mathbf{v}_i,\mathbf{p}_i,\mathbf{q}_i,\mathbf{z}$ and the Lagrangian multiplies $\boldsymbol{\mu}_i$, $\boldsymbol{\nu}_i$, $\boldsymbol{\tau}_i$, $\boldsymbol{\theta}_i$, $\boldsymbol{\eta}_i$ in the same way as in last section. Define the feasible regions of the localized variables as $\mathcal{Z}_i\triangleq\big\{\mathbf{z}_i,\mathbf{f}_i,w_i\big|(33b-33c),(25c-25g),(25j-25i)\big\}$, $i\in\mathcal{L}$. Then (P10) with given $\mathbf{g}$ can be optimized by ADMM method. The whole procedure of ADMM is literally identical with the steps in Alg.1 with the only exception that the term $\frac{1}{L}\sum_{i=1}^{L}t_i$ in (26) and $\frac{1}{L}t_i$ in (27) are replaced by $\sum_{i=1}^{L}t_i^2$ and t_i^2 respectively. Details are omitted for brevity.

The power-minimization transceiver design for ellipsoidal CSI uncertainty is summarized in Alg.3

Algorithm 3 Sum-Power Min. for Ellipsoidal CSI Model

1 **Initialization**: Generate $\mathbf{x}^{(0)}$; $\boldsymbol{\xi}^{(0)}=\mathbf{0}$; $j=0$;
2 **repeat**
3 fix $\mathbf{g}^{(j)}$, solve (P10) by SDP solver or ADMM method to update $\mathbf{f}^{(j+1)}$;
4 fix $\mathbf{f}^{(j+1)}$, solve (P10) to update $\mathbf{g}^{(j+1)}$;
5 **until** *convergence*;

Too small ϵ_0 will make the power minimization problem infeasible. Practically we can invoke the Alg.2 by feeding with sufficiently large power supplies to obtain a "best" ϵ_0 and regard this value as the best MSE level that we can target.

V. Connecting MSE and SNR

The quality of signal recovery can be evaluated by MSE or SNR. In the following we represent a fundamental relation to connect these two metrics.

Proposition 1: The following inequality stands between MSE and SNR

$$\mathsf{MSE}\ge\Big(1+\mathsf{SNR}\Big)^{-1} \tag{34}$$

Proof: To simplify the following exposition, denote $\mathbf{h}\triangleq\sum_{i=1}^{L}\mathbf{H}_i\mathbf{F}_i\mathbf{1}_{K_i}$, $\boldsymbol{\Sigma}_{\mathbf{n}}\triangleq\sigma_0^2\mathbf{I}_M+\sum_{i=1}^{L}\mathbf{H}_i\mathbf{F}_i\boldsymbol{\Sigma}_i\mathbf{F}_i^H\mathbf{H}_i^H$ and $\widetilde{\mathbf{g}}\triangleq\frac{\mathbf{g}}{\|\mathbf{g}\|_2}$. Then SNR can be equivalently written as

$$\mathsf{SNR}=\frac{\big|\widetilde{\mathbf{g}}^H\mathbf{h}\big|^2}{\widetilde{\mathbf{g}}^H\boldsymbol{\Sigma}_{\mathbf{n}}\widetilde{\mathbf{g}}}. \tag{35}$$

Define $\mathsf{MSE}_{\widetilde{\mathbf{g}}}(\alpha)\triangleq\mathsf{MSE}\big(\boldsymbol{\mathcal{H}},\boldsymbol{\mathcal{F}},\alpha\widetilde{\mathbf{g}}\big)$ where α is a non-negative real value. It can be verified that

$$\mathsf{MSE}_{\widetilde{\mathbf{g}}}(\alpha)=\Big(\big|\mathbf{h}^H\widetilde{\mathbf{g}}\big|^2+\widetilde{\mathbf{g}}^H\boldsymbol{\Sigma}_{\mathbf{n}}\widetilde{\mathbf{g}}\Big)\alpha^2-2\mathsf{Re}\{\mathbf{h}^H\widetilde{\mathbf{g}}\}\alpha+1. \tag{36}$$

Obviously $\mathsf{MSE}_{\widetilde{\mathbf{g}}}(\alpha)$ is a strictly convex function with respect to α. The optimal $\alpha^\star$ yielding the minimal $\mathsf{MSE}_{\widetilde{\mathbf{g}}}(\alpha)$ can be identified as

$$\alpha^\star=\frac{\mathsf{Re}\{\mathbf{h}^H\widetilde{\mathbf{g}}\}}{\big|\mathbf{h}^H\widetilde{\mathbf{g}}\big|^2+\widetilde{\mathbf{g}}^H\boldsymbol{\Sigma}_{\mathbf{n}}\widetilde{\mathbf{g}}} \tag{37}$$

Based on the above discussion, the following inequality holds

$$\begin{aligned}\mathsf{MSE}&=\mathsf{MSE}_{\widetilde{\mathbf{g}}}(\|\mathbf{g}\|_2)\ge\mathsf{MSE}_{\widetilde{\mathbf{g}}}(\alpha^\star)\\ &=1-\frac{\mathsf{Re}^2\{\mathbf{h}^H\widetilde{\mathbf{g}}\}}{\big|\mathbf{h}^H\widetilde{\mathbf{g}}\big|^2+\widetilde{\mathbf{g}}^H\boldsymbol{\Sigma}_{\mathbf{n}}\widetilde{\mathbf{g}}}\ge1-\frac{\big|\mathbf{h}^H\widetilde{\mathbf{g}}\big|^2}{\big|\mathbf{h}^H\widetilde{\mathbf{g}}\big|^2+\widetilde{\mathbf{g}}^H\boldsymbol{\Sigma}_{\mathbf{n}}\widetilde{\mathbf{g}}}\\ &=\frac{1}{1+\mathsf{SNR}}.\end{aligned} \tag{38}$$

The proof is complete. □

Remark 1: Prop.1 reminds us the classical result

$$\mathsf{MSE}^\star=\big(1+\mathsf{SNR}^\star\big)^{-1} \tag{39}$$

e.g. by [35]–[37]. The equality (39) does not help for imperfect CSI scenario. In fact $\mathsf{MSE}^\star$ and $\mathsf{SNR}^\star$ in (39) represent optimal values of MSE and SNR respectively, obtained via using optimal receiver, which is a function of channel state [36]. In the absence of accurate CSI, however, the optimal receiver

TABLE I
COMPLEXITY OF DIFFERENT OPTIMIZATION PROBLEMS

Problem	Computation Complexity
(P3), (P9) w.r.t. $\mathcal{F}$	$\mathcal{O}\Big(L^{1.5}N^{1.5}K(M+K)^{0.5}\big[L^3NK+L^2(N^3K^3+M^2N^3K+N^2K^2M)+L(M^3N^3+MNK^2)\big]\Big)$
(P7^i), (P10^i) w.r.t. $\mathbf{F}_i$	$\mathcal{O}\Big(N^{0.5}(M+K)^{0.5}(KN+L)\big[L^3+L^2(M+KN)+L(KNM+M^2N^2+N^2K^2+MNK)+KM^2N^3+N^3K^3+MN^2K^2+N^3M^3\big]\Big)$
(P3), (P9′) w.r.t. $\mathbf{g}$	$\mathcal{O}\Big(L^{0.5}N^{0.5}(M+K)^{0.5}(M+3L)^{0.5}\big[L^3+L^2(M^2N^2+N^2K^2+MNK)+L(M^3N^3+N^3K^3+M^2N^2K+MNK^2+MN^2K^2)\big]\Big)$

Note: The homogeneous WSN setting is considered: $K_i = K$ and $N_i = N$, $\forall i \in \mathcal{L}$. Only dominant terms of complexity are presented.

TABLE II
MATLAB RUNNING TIME TO SOLVE (P3) TO UPDATE $\mathcal{F}$ (IN SEC.), $K_i = 4$, $N_i = 4$, $M = 4$

Alg.	$L=5$	$L=10$	$L=20$	$L=30$	$L=40$	$L=50$	$L=100$	$L=200$	$L=300$	$L=320$
centralized	5.597	28.08	279.9	2672.1	—	—	—	—	—	—
ADMM	1.594	1.818	2.7180	2.765	3.077	4.958	6.953	13.32	27.52	—

Note: "—" means the problem is too large to be solved.

is unknown and so are $\mathsf{MSE}^\star$ and $\mathsf{SNR}^\star$. Note the equality (39) collapses for arbitrary MSE and SNR associated with non-optimal receiver. Comparatively, (34) always stands, whether the receiver is optimal or not for an arbitrary channel state. The result (34) indeed extends the preliminary conclusion in [33], where a point-to-point multiple-input-single-output (MISO) case is considered, to the more generic scenario in WSN settings, where multi-antennas are deployed at both ends.

Considering the fact that SNR has a much more complicated form than MSE, the significance of Prop.1 lies in that we could circumvent the SNR function by consulting to MSE when the SNR objective/constraint is considered. Specifically to robustly maximize SNR, we can turn to minimize MSE, which equivalently minimizes an upper-bound of SNR. Similarly the constraint $\mathsf{SNR} \geq \gamma$ in a robust power minimization problem can be replaced by its conservative version $\mathsf{MSE} \leq (1+\gamma)^{-1}$ which guarantees the feasibility of the solution, as verified by the numerical results in Sec.VII.

VI. COMPLEXITY AND COMMUNICATION COST FOR DE/CENTRALIZED SOLUTIONS

In previous discussions, both centralized and decentralized solutions have been developed. In practice one should choose wisely between these two alternatives based on the scenario where the wireless sensor network is deployed. In the following we talk about the complexity and communication cost of these two alternative solutions.

A. Complexity

Here we analyze the complexity of the aforementioned solutions in previous sections. For simplicity we just assume that all sensor nodes have identical observation dimensions and antenna numbers, i.e. $K_i = K$ and $N_i = N$ $\forall i \in \mathcal{L}$. According to [45, Sec. 6.6.3], the computation complexity of centralized and decentralized solutions to obtain different variables is provided in Tab.I. As presented in Tab.I, the dominating complexity term in L for centralized solution to (P3) is $L^{4.5}N^{2.5}K^2(M+K)^{0.5}$, which is reduced to $L^4N^{0.5}(M+K)^{0.5}$ via ADMM method. In fact, a more concrete comparison can be obtained via comparing various algorithms' CPU run time, which is demonstrated in Tab.II. As shown by the data, decentralized algorithms can greatly decrease the computation complexity by decomposing the original big problem into smaller ones. Besides, the centralized algorithm also imposes a extremely high demand for computing/storage capability. As reflected in Tab.II the centralized solution easily becomes impractical due to its demanding requirement for memory.

B. Communication Overhead

Decentralized solutions usually have a much higher communication overhead. For decentralized Alg.2/3, x-update (28) (namely the "consensus" step [34]) averages the local copies $\mathbf{z}$ collecting from computing units. The updated $\mathbf{x}$ and $\boldsymbol{\xi}$ should be distributed to computing units. Therefore $6L(2L+1)$ real scalars should be communicated for each ADMM update. Assume that ADMM is performed C_0 times before $\mathbf{g}$ is updated and broadcast. Then a total of $6C_0L(2L+1)+2LM$ real values should be communicated for each iteration of Alg.2/3. In contrast, 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, which results in LKN complex values to transfer totally.

It is worth noting that under specific circumstances, communication cost may become negligible. One typical instance is the use of multi-core processors [47], which are usually equipped by processing hubs in many wireless multimedia sensor networks [1]. According to [47], the communication bandwidth between cores of a typical multi-core platform can reach up to tens of Gigabits per second and thus the information exchange between computing cores is not bottleneck any more. In this case, the proposed decentralized algorithms can serve as a very attractive candidate.

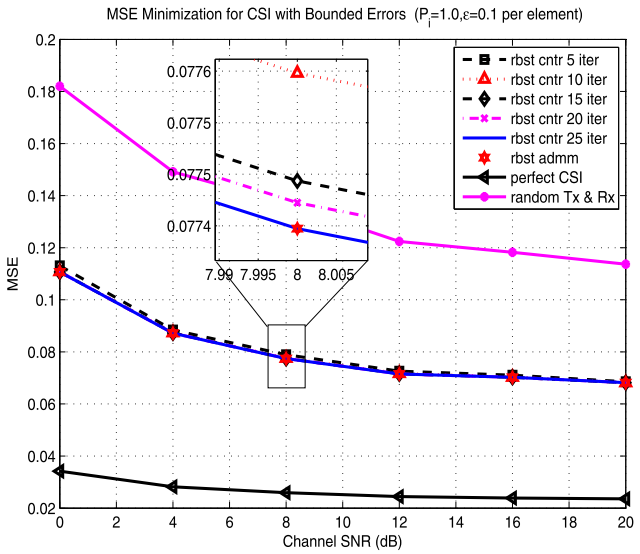

Fig. 2. Alg.2 to minimize MSE for ellipsoidal CSI model.

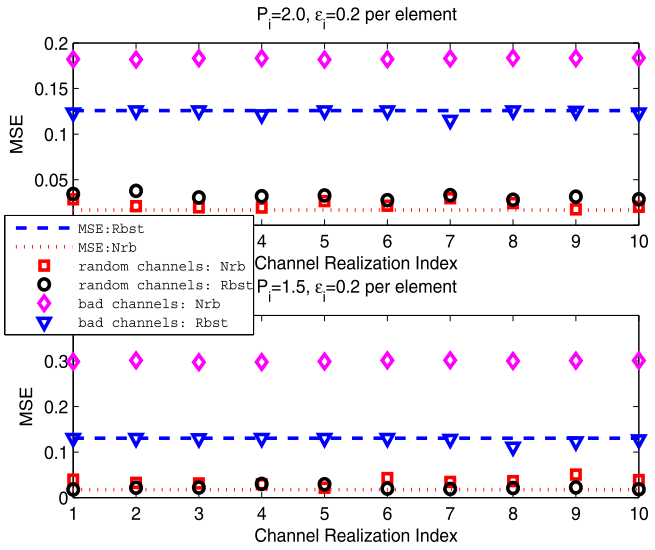

Fig. 3. MSE of robust and non-robust designs for ellipsoidal CSI model.

VII. Numerical Results

In this section, numerical results are presented to testify our proposals in previous sections. In the following experiments, we consider a WSN with $L = 5$ sensors. All sensors are equipped with $N_i = 2$ antennas and the fusion center has $M = 2$ antennas. Each sensor is affected by a colored observation noise, which has a Toeplitz type covariance matrix $\mathbf{\Sigma}_i$ with its (j,k)-th element $[\mathbf{\Sigma}_i]_{j,k} = \pi^{|k-j|}$, $\forall i \in \mathcal{L}$. In our test, we set $\pi = 0.5$. The SNR of signal to the observation noise at each sensor is given as 8dB.

First Alg.2 is tested as shown in Fig.2. In the experiment, we choose $\mathbf{Q}_i$ as identity matrix $\forall i \in \mathcal{L}$. For each specific channel noise level, 50 channel realizations are randomly generated and the average of MSE obtained by Alg.2 is plotted. The solution in [6] with perfect CSI and random transceivers' MSE are also plotted in Fig.2 for comparison. Alg.2 usually converges within 20 iterations and both of its centralized and decentralized implementations yield identical performance.

To further test the robustness of the solutions, we set the channel noise level at 8dB and randomly generate a channel estimate $\{\widehat{\mathbf{H}}_i\}_{i=1}^L$. Both the robust and non-robust solutions are obtained by our Alg.2 and the algorithm in [6] respectively. Based on the generated channel estimate, we randomly generate a group of "erroneous" channels with their CSI errors complying with the ellipsoidal model. For each erroneous CSI sample, we evaluate the MSE associated with robust and non-robust solutions. Moreover, based on each generated CSI error, we utilize the BCA (block coordinate ascent) method to iteratively maximize the MSE with respect to CSI errors (refer to Appendix.C for details) to find "bad" channels. As shown by the dotted lines in Fig.3, the robust solution gives out the worst-case MSE, which is worse than the MSE obtained by non-robust solutions, where only perfect $\{\widehat{\mathbf{H}}_i\}_{i=1}^L$ are considered. However in the case of adverse channels, the MSE of non-robust solutions becomes severely deteriorated and much worse compared to that of the robust ones. It is worth noting that the worst-case MSE obtained by robust solutions are actually very tight in case of "bad" channels.

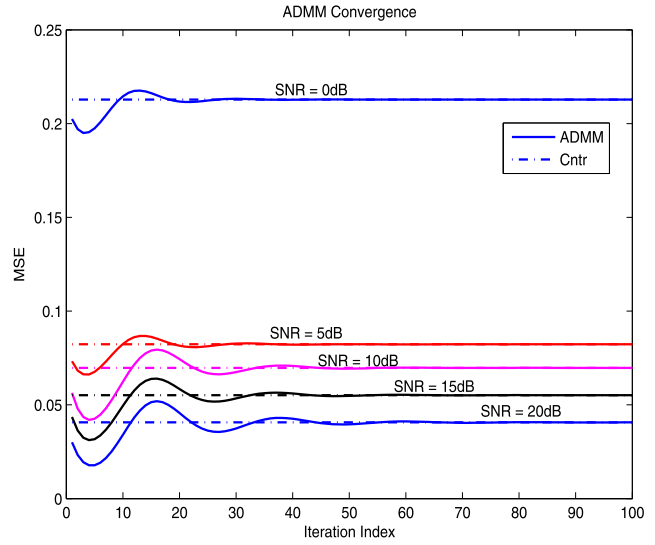

Fig. 4. ADMM method to update sensors' beamformers (line 3 in Alg.2).

In Fig.4, we present ADMM method to optimize the sensors' transmitters distributively. Usually the ADMM algorithm converges fast when $\kappa \in [2, 8]$, which is obtained via our numerical tests. In our experiment, we choose $\kappa = 2$. Different channel SNR levels are tested. The ADMM algorithm usually converges to optima within 70 iterations.

Fig.5 presents the performance of Alg.3 for CSI with ellipsoidal errors. For each channel noise level, 400 random channel estimates are tested and the average result is plotted in Fig.5. In the experiment Alg.2 is first invoked and the obtained optimized MSE times 105% is set as the MSE-target ϵ_0 to invoke Alg.3. The obtained transceivers by Alg.2 are set as the starting point of Alg.3 (tagged as "rbst initial" in Fig.5). It can be seen that Alg.3 can usually get converged within 50 iterations. Both of its centralized and ADMM implementation yield identical sum-powers.

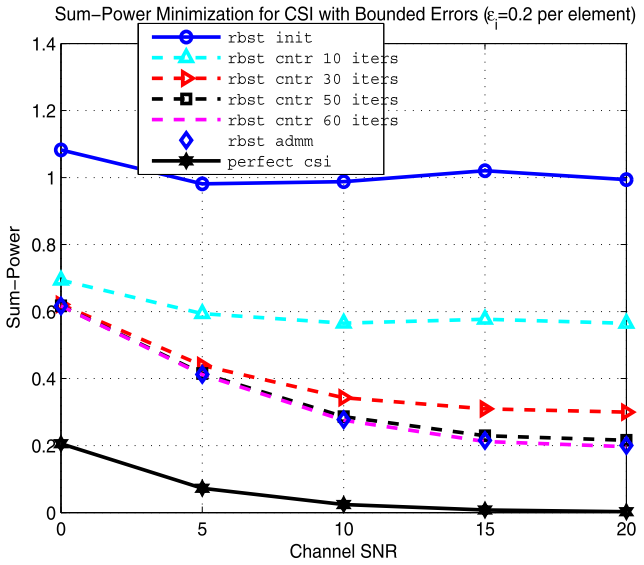

Fig. 5. Alg.3 to minimize sum-power for CSI with ellipsoidal errors.

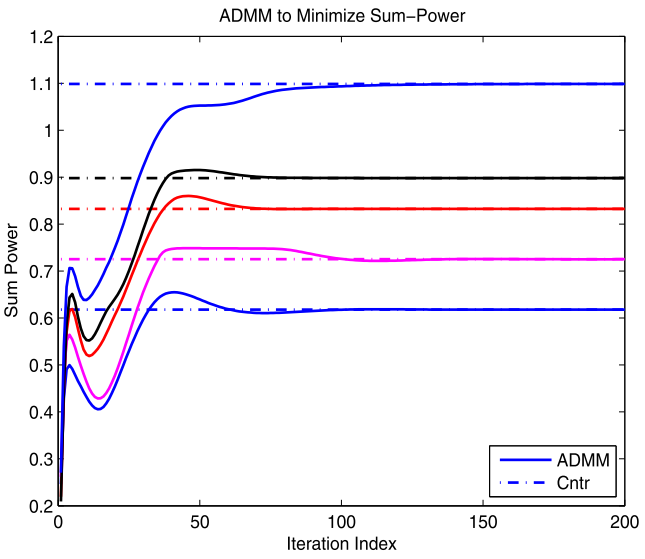

Fig. 7. ADMM method to optimize the transmitters (step 3 of Alg.3).

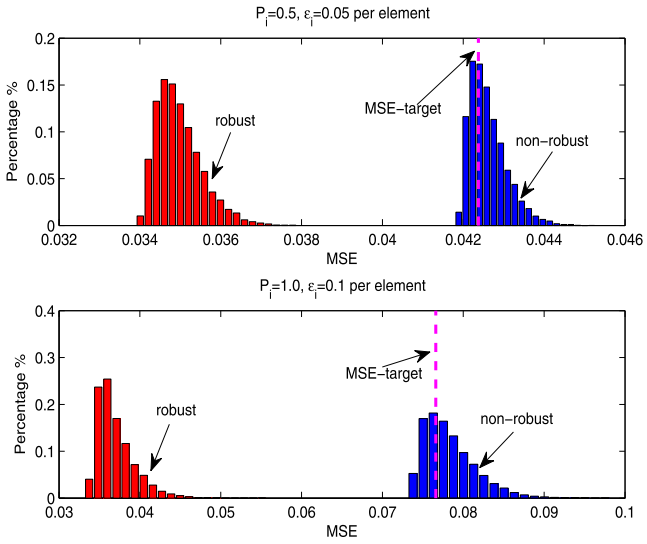

Fig. 6. Robustness of the power-minimization solution for CSI with ellipsoidal errors.

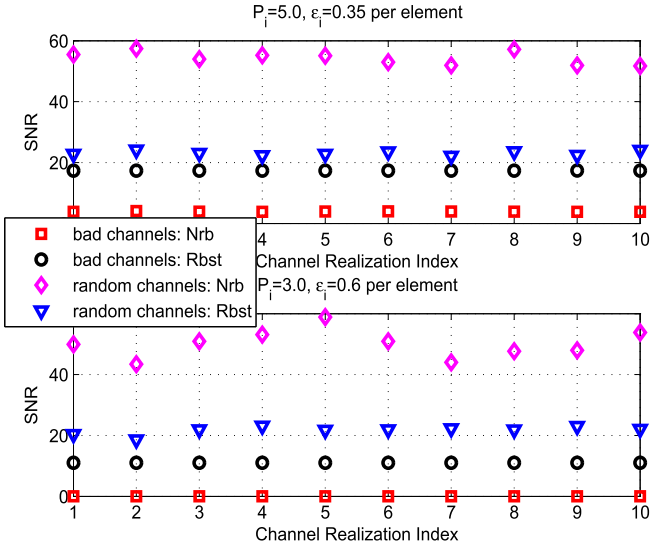

Fig. 8. SNR of robust and non-robust designs for ellipsoidal CSI model.

To verify the power-minimization solutions' robustness, we randomly generate a channel estimate and then obtain its associated robust and non-robust power efficient transceivers. 10000 random (Gaussian distributed with possibly uniform contraction) CSI errors complying ellipsoid model are generated and tested to evaluate the associated MSE. For each erroneous CSI sample, the MSE associated with the robust and non-robust transceivers are calculated. The resultant MSE distributions are plotted in Fig.6, with the MSE-target being also presented as a benchmark. As can be seen more than 60% of the CSI samples yield a worse MSE than the MSE-target for the non-robust scheme. Comparatively the robust transceivers always guarantee the MSE-target. It should be pointed out, we utilize the channel deterioration method in Appendix C to maximize MSE with respective to CSI errors, the MSE-target can always be approached by "bad" channels in almost all test cases. So the MSE-target is actually tight for robust solutions.

The convergence of ADMM implementation of Alg.3 is presneted in Fig.7, where the curves of different colors correspond to different channel estimate realizations that are randomly generated. The penalty parameter κ is chosen as 7 in our experiment. ADMM method converges within 150 iterations.

The computation efficiency of the centralized and decentralized way to update transmitters in Alg.2 is presented in Table II. As can be seen the complexity of the centralized method grows aggressively w.r.t. the number of sensors L and easily exceeds the computation capability of a standard personal computer when L is several tens. In fact, the decentralized method is the only admissible way to get a solution for large-scale WSN.

Additionally we also examine the SNR performance of the robust beamformers. For a randomly generated $\{\widehat{\mathbf{H}}_i\}_{i=1}^{L}$, robust and non-robust power minimization solutions are obtained. We randomly generate 10 channel realizations

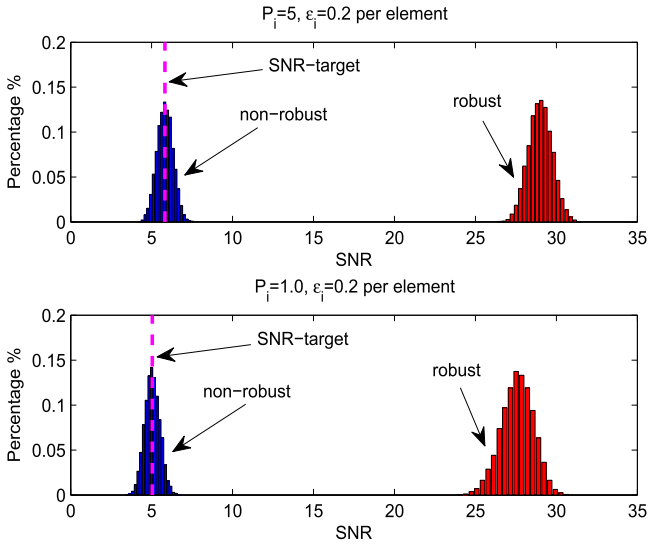

Fig. 9. Robustness of the power-minimization solution with SNR constraint.

according to the ellipsoid model and evaluate SNR for different beamformers. Besides, based on each random channel realization, we utilize a BCD method (as explained in Appendix D) to find worst-case SNR. As illustrated in Fig.8, the non-robust beamformers usually have a better SNR than robust beamformers for random channel realizations. However robust transceivers have an advantageous worst-case SNR. Especially when ϵ_i grows large, the worst-case SNR of non-robust beamformers easily falls close to zaro.

Finally the sum-power minimization with worst-case SNR requirement is also examined. Assume that the worst-case SNR target is γ_0. Then according to the discussion in Sec.V, we can turn to consider the power minimization problem constrained by the worst-case MSE no larger than $(1+\gamma_0)^{-1}$. Thus Alg.3 can be used. The SNR performance of the robust and non-robust solutions are illustrated in Fig.9. 10000 random CSI errors complying ellipsoid model are generated for test. As illustrated in the figure, for non-robust beamformer, nearly 50% of the CSI samples yield a lower SNR than the SNR-target.

VIII. Conclusion

This paper has performed a comprehensive research on robust linear transceiver design in MIMO WSN. Under the assumption of ellipsoidal CSI uncertainty model, MSE and sum-power are optimized. Both centralized and decentralized algorithms are developed. Extensive numerical results are provided to verify our proposals.

Besides, remaining as open issues of the the work at this point, the following questions may serve as a potential research direction in future

i) This article sticks to the model where the target parameter is a scalar. Generalization of the current solution to the more generic case where a parameter vector is considered is desirable. In fact, when a parameter vector $\boldsymbol{\theta}$ is considered, the upper bound derived by (9-11) and Lemma 1 will not be analytic any more. New method to derive analytic and tight upper bound of worst-case MSE should be explored.

ii) Currently the feasibility identification of the sum-power minimization problem (P9) for any given MSE target power ϵ_0 is still an open issue, which can be a meaningful direction for further study.

Appendix

A. Proof of Lemma 1

Proof: The left hand side of the assertion can be evaluated as

$$\Big|1-\mathbf{g}^H\big(\sum_{i=1}^{L}\mathbf{H}_i\mathbf{F}_i\mathbf{1}_{K_i}\big)\Big| \tag{40a}$$

$$=\Big|1-\mathbf{g}^H\big(\sum_{i=1}^{L}\widehat{\mathbf{H}}_i\mathbf{F}_i\mathbf{1}_{K_i}\big)-\mathbf{g}^H\big(\sum_{i=1}^{L}\mathbf{E}_i\mathbf{F}_i\mathbf{1}_{K_i}\big)\Big| \tag{40b}$$

$$=\Big|s_0-\sum_{i=1}^{L}\mathbf{s}_i^T\mathbf{e}_i\Big|=\Big|s_0-\sum_{i=1}^{L}\mathbf{s}_i^T\mathbf{Q}_i^{-\frac{1}{2}}\mathbf{Q}_i^{\frac{1}{2}}\mathbf{e}_i\Big| \tag{40c}$$

$$\le\Big||s_0|+\sum_{i=1}^{L}\epsilon_i\big\|\mathbf{Q}_i^{-\frac{1}{2}}\mathbf{s}_i^*\big\|_2\Big| \tag{40d}$$

where (40b) utilizes the identity $\mathbf{g}^H\mathbf{E}_i\mathbf{F}_i\mathbf{1}_{K_i}=\mathsf{Tr}\{\mathbf{F}_i\mathbf{1}_{K_i}\mathbf{g}^H\mathbf{E}_i\}=\mathsf{vec}^T(\mathbf{g}^*\mathbf{1}_{K_i}^T\mathbf{F}_i^T)\mathbf{e}_i$ and the last step follows Cauchy-Schwartz inequality. Note that (40d) can achieve equality when $\mathbf{e}_i=-e^{j\theta_0}\epsilon_i\|\mathbf{Q}_i^{-\frac{1}{2}}\mathbf{s}_i^*\|_2^{-1}\mathbf{Q}_i^{-1}\mathbf{s}_i^*$, $\forall i\in\mathcal{L}$ with θ_0 being the phase of s_0. □

B. Proof of Lemma 3

Proof: First it is easily seen that (P3) is actually a restrictive version of (P2). In fact, compared to (P2), the constraints (22c), (22d) and $s_0\ge 0$ in (22e) impose an additional restriction that $\mathbf{g}^H(\sum_{i=1}^{L}\widehat{\mathbf{H}}_i\mathbf{F}_i\mathbf{1}_{K_i})$ is a real number and larger than -1.

Next we will prove that at the optimality of (P2), the constraints (22c), (22d) and $s_0\ge 0$ in (22e) actually hold. Recall that $s_0=1-\mathbf{g}^H\big(\sum_{i=1}^{L}\widehat{\mathbf{H}}_i\mathbf{F}_i\mathbf{1}_{K_i}\big)$. Assume that $\{\widetilde{\mathbf{F}}_i\}_{i=1}^L$ and $\widetilde{\mathbf{g}}$ are optimal to (P2). Define $\widetilde{\pi}=-\widetilde{\mathbf{g}}^H\big(\sum_{i=1}^{L}\widehat{\mathbf{H}}_i\widetilde{\mathbf{F}}_i\mathbf{1}_{K_i}\big)$. Thus $\widetilde{s}_0=1+\widetilde{\pi}$.

One critical observation is that the optimal $\{p_i\}_{i=1}^L$ and $\{q_i\}_{i=1}^L$ are independent of phase rotation of $\mathbf{F}_i$ or $\mathbf{g}$. That is if $\widetilde{\mathbf{F}}_i$ (or $\widetilde{\mathbf{g}}$) is changed into $e^{j\theta_i}\widetilde{\mathbf{F}}_i$ (or $e^{j\phi}\widetilde{\mathbf{g}}$) with θ_i (or ϕ) being arbitrary values, the optimal $\{p_i\}_{i=1}^L$ and $\{q_i\}_{i=1}^L$ remain the same. In fact at optimality, $q_i=\big\|\widetilde{\mathbf{g}}^H\big[\big(\widetilde{\mathbf{F}}_i\mathbf{1}_{K_i}\big)^T\otimes\mathbf{I}_M\big]\mathbf{Q}_i^{-\frac{1}{2}}\big\|_2$ and $p_i=\max_{\mathbf{H}_i\in\mathcal{X}_i}\widetilde{\mathbf{g}}^H\mathbf{H}_i\widetilde{\mathbf{F}}_i\boldsymbol{\Sigma}_i\widetilde{\mathbf{F}}_i^H\mathbf{H}_i^H\widetilde{\mathbf{g}}$, $i\in\mathcal{L}$. Obviously the values of $\{p_i\}_{i=1}^L$ and $\{q_i\}_{i=1}^L$ are independent of phase of $\widetilde{\mathbf{F}}_i$ or $\widetilde{\mathbf{g}}$.

First we note $\mathsf{Re}\{\widetilde{\pi}\}\le 0$. In fact, if $\mathsf{Re}\{\widetilde{\pi}\}>0$, we can just take $\bar{\mathbf{F}}_i=-\widetilde{\mathbf{F}}_i$ (or $\bar{\mathbf{g}}=-\widetilde{\mathbf{g}}$) and denote the associated π as $\bar{\pi}$. Then $\bar{\pi}=-\widetilde{\pi}$ Since the optimal $\{p_i\}_{i=1}^L$ and $\{q_i\}_{i=1}^L$ will not change and $|1+\bar{\pi}|^2<|1+\widetilde{\pi}|^2$. Then $\{-\widetilde{\mathbf{F}}_i\}_{i=1}^L$ will strictly decrease the objective of (P2), which contradicts the fact that $\{\widetilde{\mathbf{F}}_i\}_{i=1}^L$ is optimal.

Next we prove that $\mathsf{Re}\{\widetilde{\pi}\}\ge -1$. By contradiction, if $\mathsf{Re}\{\widetilde{\pi}\}<-1$, then set $\bar{\mathbf{F}}_i=(1-\delta)\widetilde{\mathbf{F}}_i$, $\forall i\in\mathcal{L}$ with

δ being a positive value. Then $\bar{\pi} = (1-\delta)\widetilde{\pi}$. Provided δ is sufficiently small, $|1+\bar{\pi}| < |1+\widetilde{\pi}|$. At the same time $\{\bar{\mathbf{F}}_i\}_{k=1}^L$ also decreases the optimal $\{p_i\}_{i=1}^L$ and $\{q_i\}_{i=1}^L$. Thus $\{(1-\delta)\widetilde{\mathbf{F}}_i\}_{i=1}^L$ will strictly decrease the objective, which contradicts the optimality of $\{\widetilde{\mathbf{F}}_i\}_{i=1}^L$.

Now we have proved that $-1 \le \mathsf{Re}\{\widetilde{\pi}\} \le 0$. In the sequel we prove that $\mathsf{Im}\{\widetilde{\pi}\} = 0$. By contradiction, assume that $\mathsf{Im}\{\widetilde{\pi}\} \neq 0$. If $\mathsf{Re}\{\widetilde{\pi}\} > -1$, one can apply a sufficiently small phase rotation θ to all $\widetilde{\mathbf{F}}_i$ such that $\bar{\pi} = e^{j\theta}\widetilde{\pi}$ is rotated in the direction of $(-1,0)$ in the complex plane. Then both $\mathsf{Re}\{1+\widetilde{\pi}\}$ and $|\mathsf{Im}\{1+\widetilde{\pi}\}| = |\mathsf{Im}\{\widetilde{\pi}\}|$ will decrease while $\{p_i\}_{i=1}^L$ and $\{q_i\}_{i=1}^L$ maintain constant, leading to a contradiction. If $\mathsf{Re}\{\widetilde{\pi}\} = -1$, suppose $\mathsf{Im}\{\widetilde{\pi}\} = \pm\epsilon$, $\epsilon > 0$. Apply a common phase rotation to all $\mathbf{F}_i$ so that $\widetilde{\pi}$ is rotated to $\bar{\pi}$, which is a negative real value. Then it can be easily verified that $|\bar{s}_0| = |1+\bar{\pi}| = (\sqrt{1+\epsilon^2}-1) < \epsilon = |1+\widetilde{\pi}| = |\widetilde{s}_0|$, a contradiction has been reached. Thus $\mathsf{Im}\{\widetilde{\pi}\} = 0$.

Therefore we have proved that $\mathsf{Im}\{\widetilde{s}_0\} = 0$ and $\mathsf{Re}\{\widetilde{s}_0\} \ge 0$. Hence (P2) is equivalent to (P3). □

C. Find "Bad" Channels for Worst-Case MSE

When the transceivers $\boldsymbol{\mathcal{F}}$ and $\mathbf{g}$ and the channel estimates $\{\widehat{\mathbf{H}}_i\}_{i=1}^L$ are given, the MSE is a function in channel state errors $\{\mathbf{E}_i\}_{i=1}^L$. To find "bad" channels giving worst-case MSE is indeed to solve the following problem

$$(\mathsf{P11}): \quad \max_{\{\mathbf{e}_i\}_{i=1}^L} . \ \mathsf{MSE}\big(\{\mathbf{e}_i\}_{i=1}^L \big| \boldsymbol{\mathcal{F}}, \{\widehat{\mathbf{H}}_i\}_{i=1}^L, \mathbf{g}\big), \tag{41a}$$

$$\text{s.t. } \mathbf{e}_i^H \mathbf{Q}_i \mathbf{e}_i \le \epsilon_i^2, \quad \forall i \in \mathcal{L} \tag{41b}$$

By defining $\widehat{s}_0 \triangleq 1 - \mathbf{g}^H\big(\sum_{i=1}^L \widehat{\mathbf{H}}_i \mathbf{F}_i \mathbf{1}_{K_i}\big)$, $\mathbf{K}_i \triangleq \big(\mathbf{F}_i \boldsymbol{\Sigma}_i \mathbf{F}_i^H\big)^T \otimes \big(\mathbf{g}\mathbf{g}^H\big)$, $\mathbf{u}_i \triangleq \mathsf{vec}\big(\mathbf{g}\mathbf{1}_{K_i}^T \mathbf{F}_i^H\big)$, the MSE can represented as follows

$$\begin{aligned}&\mathsf{MSE}\Big(\{\mathbf{e}_i\}_{i=1}^L \big| \boldsymbol{\mathcal{F}}, \{\widehat{\mathbf{H}}_i\}_{i=1}^L, \mathbf{g}\Big)\\ &= \Big|\widehat{s}_0 - \sum_{i=1}^L \mathbf{u}_i^H \mathbf{e}_i\Big|^2 + \sigma_0^2 \|\mathbf{g}\|_2^2 + \sum_{i=1}^L \big(\widehat{\mathbf{h}}_i + \mathbf{e}_i\big)^H \mathbf{K}_i \big(\widehat{\mathbf{h}}_i + \mathbf{e}_i\big)\end{aligned} \tag{42}$$

Since (P11) is nonconvex and difficult to solve, we adopt BCA method to iteratively maximize MSE, with each iteration focusing on one separate $\mathbf{e}_i$. By discarding the constant terms and rewriting in a minimization form, the MSE maximization with respect to $\mathbf{e}_i$ is given as

$$(\mathsf{P11}_i): \quad \min_{\mathbf{e}_i} . \ -\mathbf{e}_i^H \mathbf{M}_i \mathbf{e}_i + 2\mathsf{Re}\{\mathbf{v}_i^H \mathbf{e}_i\}, \tag{43a}$$

$$\text{s.t. } \mathbf{e}_i^H \mathbf{Q}_i \mathbf{e}_i \le \epsilon_i^2, \tag{43b}$$

where $\mathbf{M}_i \triangleq \mathbf{K}_i + \mathbf{u}_i\mathbf{u}_i^H$ and $\mathbf{v}_i \triangleq a_i\mathbf{u}_i - \mathbf{K}_i\widehat{\mathbf{h}}_i$ with $a_i \triangleq \widehat{s}_0 - \sum_{j\neq i} \mathbf{u}_j^H \mathbf{e}_j$ and $\widehat{\mathbf{h}}_i \triangleq \mathsf{vec}\{\widehat{\mathbf{H}}_i\}$.

Notice the problem $(\mathsf{P11}_i)$ is obviously strictly feasible, the result of [43, Appendix B.1] is valid to invoke, which states that $(\mathsf{P11}_i)$ has the following relaxation

$$(\mathsf{P12}_i) \quad \min_{\mathbf{X}_i \succeq 0, \mathbf{e}_i} . \ \mathsf{Tr}\big\{-\mathbf{M}_i\mathbf{X}_i\big\} + 2\mathsf{Re}\big\{\mathbf{v}_i^H\mathbf{e}_i\big\}, \tag{44a}$$

$$\text{s.t. } \mathsf{Tr}\big\{\mathbf{Q}_i\mathbf{X}_i\big\} - \epsilon_i^2 \le 0, \tag{44b}$$

$$\begin{bmatrix} \mathbf{X}_i & \mathbf{e}_i \\ \mathbf{e}_i^H & 1 \end{bmatrix} \succeq 0. \tag{44c}$$

where the constraint $\mathbf{X}_i = \mathbf{e}_i\mathbf{e}_i^H$ is omitted in $(\mathsf{P12}_i)$.

Replace the variables $(\mathbf{X}_i, \mathbf{e}_i)$ in $(\mathsf{P12}_i)$ by one matrix variable $\widetilde{\mathbf{X}}_i$ and rewrite it into a SDP form

$$(\mathsf{P12}_i') \quad \min_{\widetilde{\mathbf{X}}_i} . \ \mathsf{Tr}\big\{\mathbf{P}_{i,1}\widetilde{\mathbf{X}}_i\big\}, \tag{45a}$$

$$\text{s.t. } \mathsf{Tr}\big\{\mathbf{P}_{i,2}\widetilde{\mathbf{X}}_i\big\} \le \epsilon_i^2, \tag{45b}$$

$$\mathsf{Tr}\big\{\mathbf{P}_{i,3}\widetilde{\mathbf{X}}_i\big\} = 1, \tag{45c}$$

with the parameter matrices being defined as

$$\mathbf{P}_{i,1} \triangleq \begin{bmatrix} -\mathbf{M}_i & \mathbf{v}_i \\ \mathbf{v}_i^H & 0 \end{bmatrix}, \quad \mathbf{P}_{i,2} \triangleq \begin{bmatrix} \mathbf{Q}_i & \mathbf{0} \\ \mathbf{0}^T & 0 \end{bmatrix}, \quad \mathbf{P}_{i,3} \triangleq \begin{bmatrix} \mathbf{O} & \mathbf{0} \\ \mathbf{0}^T & 1 \end{bmatrix}.$$

Assume that $\widetilde{\mathbf{X}}_i^\star$ is one optimal solution to $(\mathsf{P12}_i')$. Evoking [41, Th. 2.2], we can obtain a vector $\widetilde{\mathbf{x}}$ such that $\mathsf{Tr}\{\mathbf{P}_{i,j}\widetilde{\mathbf{X}}_i^\star\} = \mathsf{Tr}\{\mathbf{P}_{i,j}\widetilde{\mathbf{x}}_i\widetilde{\mathbf{x}}_i^H\}$ for $j = 1,2,3$. Denote $\widetilde{\mathbf{x}}_i = [\widetilde{\mathbf{x}}_{i,1}^T, \widetilde{x}_{i,2}]^T$. Then it can be easily checked that $\mathbf{e}_i^\star = \widetilde{\mathbf{x}}_{i,1}/\widetilde{x}_{i,2}$ is an optimal solution to $(\mathsf{P11}_i)$.

The procedure to find "bad" channels maximizing MSE is summarized in Alg.5.

Algorithm 4 Find "Bad" Channels for Worst Case MSE

```
Initialization:Given F, g and {Ĥ_i}_{i=1}^L;
repeat
    for i = 1, ··· , L; do
        Solve (P12_i') and obtain e_i^⋆ to (P11_i) [41] ;
    end
until convergence;
```

D. Find "Bad" Channels for Worst-Case SNR

This section briefly discusses how to determine the "bad" channels minimizing SNR, when $\boldsymbol{\mathcal{F}}$, $\mathbf{g}$ and $\{\widehat{\mathbf{H}}_i\}_{i=1}^L$ are fixed.

We still adopt BCD method. Denote $\mathbf{R}_{ij} \triangleq \big(\mathbf{F}_j\mathbf{1}_{K_j}\mathbf{1}_{K_i}^T\mathbf{F}_i^H\big)^T \otimes \big(\mathbf{g}\mathbf{g}^H\big)$, $\mathbf{q}_i \triangleq \sum_{j\neq i}\mathbf{R}_{ij}\mathbf{e}_j + \sum_j \mathbf{R}_{ij}\widehat{\mathbf{h}}_j$, $t_i \triangleq \sum_{j,k\neq i}\mathbf{e}_j^H\mathbf{R}_{jk}\mathbf{e}_k + 2\sum_{k=1,j\neq i}^L \mathsf{Re}\{\widehat{\mathbf{h}}_k^H\mathbf{R}_{kj}\mathbf{e}_j\} + \sum_{i,j}\widehat{\mathbf{h}}_i^H\mathbf{R}_{ij}\widehat{\mathbf{h}}_j$, the numerator of SNR can be represented (after manipulations) as

$$\mathbf{e}_i^H\mathbf{R}_{ii}\mathbf{e}_i + 2\mathsf{Re}\big\{\mathbf{q}_i^H\mathbf{e}_i\big\} + t_i. \tag{47}$$

Introduce additional notations $\mathbf{p}_i \triangleq \mathbf{K}_i\widehat{\mathbf{h}}_i$ and $u_i \triangleq \sum_{j\neq i}\mathbf{e}_j^H\mathbf{K}_j\mathbf{e}_j + 2\sum_{j\neq i}^L \mathsf{Re}\{\widehat{\mathbf{h}}_j^H\mathbf{K}_j\mathbf{e}_j\} + \sum_j \widehat{\mathbf{h}}_j^H\mathbf{K}_j\widehat{\mathbf{h}}_j + \sigma_0^2\|\mathbf{g}\|_2^2$, the denominator of SNR can be rewritten as

$$\mathbf{e}_i^H\mathbf{K}_i\mathbf{e}_i + 2\mathsf{Re}\big\{\mathbf{p}_i^H\mathbf{e}_i\big\} + u_i. \tag{48}$$

Then the problem to minimize SNR with respect to $\mathbf{e}_i$ is given as the following problem:

$$(\mathsf{P13}_i): \quad \min_{\mathbf{e}_i} . \ \frac{\mathbf{e}_i^H\mathbf{R}_{ii}\mathbf{e}_i + 2\mathsf{Re}\big\{\mathbf{q}_i^H\mathbf{e}_i\big\} + t_i}{\mathbf{e}_i^H\mathbf{K}_i\mathbf{e}_i + 2\mathsf{Re}\big\{\mathbf{p}_i^H\mathbf{e}_i\big\} + u_i}, \tag{49a}$$

$$\text{s.t. } \mathbf{e}_i^H\mathbf{Q}_i\mathbf{e}_i \le \epsilon_i^2. \tag{49b}$$

Following Charnes-Cooper's transformation and relaxing the rank-one constraint $\mathbf{Z}_i = [\mathbf{e}_i^H, 1]^H[\mathbf{e}_i^H, 1]$, $(\mathsf{P13}_i)$ is transformed into the following relaxed version

$$(\mathsf{P14}_i): \quad \min_{\mathbf{Z}_i, \eta_i} . \ \mathsf{Tr}\big\{\mathbf{T}_{i,1}\mathbf{Z}_i\big\}, \tag{50a}$$

$$\text{s.t. } \mathsf{Tr}\{\mathbf{T}_{i,2}\mathbf{Z}_i\} = 1, \tag{50b}$$
$$\mathsf{Tr}\{\mathbf{T}_{i,3}\mathbf{Z}_i\} \le \epsilon_i^2\eta_i, \tag{50c}$$
$$\mathsf{Tr}\{\mathbf{T}_{i,4}\mathbf{Z}_i\} = \eta_i, \tag{50d}$$
$$\mathbf{Z}_i \succcurlyeq 0, \quad \eta_i \ge 0. \tag{50e}$$

with parameter matrices being defined as

$$\mathbf{T}_{i,1} \triangleq \begin{bmatrix}\mathbf{R}_{ii} & \mathbf{q}_i \\ \mathbf{q}_i^H & t_i\end{bmatrix}, \quad \mathbf{T}_{i,2} \triangleq \begin{bmatrix}\mathbf{K}_i & \mathbf{p}_i \\ \mathbf{p}_i^H & u_i\end{bmatrix}, \tag{51a}$$
$$\mathbf{T}_{i,3} \triangleq \begin{bmatrix}\mathbf{Q}_i & \mathbf{0} \\ \mathbf{0}^T & 0\end{bmatrix}, \quad \mathbf{T}_{i,4} \triangleq \begin{bmatrix}\mathbf{O} & \mathbf{0} \\ \mathbf{0}^T & 1\end{bmatrix}. \tag{51b}$$

Assume that $(\mathbf{Z}_i^\star, \eta_i^\star)$ is an optimal solution to $(\mathsf{P}14_i)$. By [41, Th. 2.3] we can obtain a $\mathbf{z}_i^\star$ such that $\frac{\mathbf{Z}_i^\star}{\eta_i^\star} = \mathbf{z}_i^\star\mathbf{z}_i^{\star H}$. Denote $\mathbf{z}_i^\star \triangleq [\mathbf{z}_{i,1}^T, z_{i,2}]^T$ then it can be proved that $\mathbf{z}_{i,1}^\star/z_{i,2}$ is the optimal $\mathbf{e}_i^\star$ to problem $(\mathsf{P}13_i)$.

Algorithm 5 Find "Bad" Channels for Worst Case SNR

Initialization:Given $\mathcal{F}$, $\mathbf{g}$ and $\{\widehat{\mathbf{H}}_i\}_{i=1}^L$;
repeat
  for $i = 1, \cdots, L$; **do**
    Solve $(\mathsf{P}14_i')$ and obtain $\mathbf{e}_i^\star$ to $(\mathsf{P}13_i)$ [41] ;
  end
until *convergence*;

REFERENCES

[1] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensor networks: A survey," *Comput. Netw.*, vol. 38, no. 4, pp. 393–422, 2002.

[2] 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.

[3] J. Fang and H. Li, "Power constrained distributed estimation with cluster-based sensor collaboration," *IEEE Trans. Wireless Commun.*, vol. 8, no. 7, pp. 3822–3832, Jul. 2009.

[4] 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.

[5] N. K. D. Venkategowda and A. K. Jagannatham, "Optimal minimum variance distortionless precoding (MVDP) for decentralized estimation in MIMO wireless sensor networks" *IEEE Signal Process. Lett.*, vol. 22, no. 6, pp. 696–700, Nov. 2015.

[6] 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.

[7] Y. Liu, J. Li, and X. Lu, "Joint transceiver design for linear MMSE data fusion in coherent MAC wireless sensor networks," *Inf. Fusion*, vol. 37, pp. 37–49, May/Sep. 2017.

[8] Y. Liu, J. Li, and X. Lu, "Multi-terminal joint transceiver design for MIMO systems with contaminated source and individual power constraint," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Honolulu, HI, USA, Jun. 2014, pp. 3087–3091.

[9] X. Wang and Q. Liang, "Efficient sensor selection schemes for wireless sensor networks in microgrid," *IEEE Syst. J.*, vol. 12, no. 1, pp. 539–547, Mar. 2018.

[10] Y. Liu and J. Li, "Linear precoding to optimize throughput, power consumption and energy efficiency in MIMO wireless sensor networks," *IEEE Trans. Commun.*, vol. 66, no. 5, pp. 2122–2136, May 2018.

[11] J. Akhtar and K. Rajawat, "Distributed sequential estimation in wireless sensor networks," *IEEE Trans. Wireless Commun.*, vol. 17, no. 1, pp. 86–100, Apr. 2017.

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

[13] 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.

[14] D. Ciuonzo, P. S. Rossi, and S. Dey, "Massive MIMO channel-aware decision fusion," *IEEE Trans. Signal Process.*, vol. 63, no. 3, pp. 604–619, Feb. 2015.

[15] V. V. Mai, W.-Y. Shin, and K. Ishibashi, "Wireless power transfer for distributed estimation in sensor networks," *IEEE J. Sel. Topics Signal Process.*, vol. 11, no. 3, pp. 549–562, Apr. 2017.

[16] S. Zhang, S. Liu, V. Sharma, and P. K. Varshney, "Optimal sensor collaboration for parameter tracking using energy harvesting sensors," *IEEE Trans. Signal Process.*, vol. 66, no. 12, pp. 3339–3353, Jun. 2018.

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

[18] M. Guerriero, P. Willett, and J. Glaz, "Distributed target detection in sensor networks using scan statistics," *IEEE Trans. Signal Process.*, vol. 57, no. 7, pp. 2629–2639, Jul. 2009.

[19] D. Ciuonzo and P. S. Rossi, "Distributed detection of a non-cooperative target via generalized locallyoptimum approaches," *Inf. Fusion*, no. 36, pp. 261–274, 2017.

[20] D. Ciuonzo, P. S. Rossi, and P. Willett, "Generalized Rao test for decentralized detection of an uncooperative target," *IEEE Signal Process. Lett.*, vol. 24, no. 5, pp. 678–682, May 2017.

[21] N. K. D. Venkategowda, B. B. Narayana, and A. K. Jagannatham, "Precoding for robust decentralized estimation in coherent-MAC-based wireless sensor networks," *IEEE Signal Process. Lett.*, vol. 24, no. 2, pp. 240–244, Feb. 2017.

[22] H. Rostami and A. Falahati, "Precoder design for decentralised estimation over MIMO-WSN based on stochastic models," *IET Commun.*, vol. 12, no. 6, pp. 736–742, 2018.

[23] Q. Li, Q. Zhang, and J. Qin, "Robust beamforming for cognitive multi-antenna relay networks with bounded channel uncertainties," *IEEE Trans. Commun.*, vol. 62, no. 2, pp. 478–487, Aug. 2014.

[24] M. Tao and R. Wang, "Robust relay beamforming for two-way relay networks," *IEEE Commun. Lett.*, vol. 16, no. 7, pp. 1052–1055, Jul. 2012.

[25] Q. Li, Y. Yang, W. K. Ma, M. Lin, J. Ge, and J. Lin, "Robust cooperative beamforming and artificial noise design for physical-layer secrecy in AF multi-antenna multi-relay networks," *IEEE Trans. Signal Process.*, vol. 63, no. 1, pp. 206–220, Jan. 2015.

[26] Q. Zhang, C. He, and L. Jiang, "Per-stream MSE based linear transceiver design for MIMO interference channels with CSI error," *IEEE Trans. Commun.*, vol. 63, no. 5, pp. 1676–1689, May 2015.

[27] T. Bogale, B. Chalise, and L. Vandendorpe, "Robust transceiver optimization for downlink multiuser MIMO system," *IEEE Trans. Signal Process.*, vol. 59, no. 1, pp. 446–453, Jan. 2011.

[28] A. Mutapcic, S. J. Kim, and S. Boyd, "A tractable method for robust downlink beamforming in wireless communications," in *Proc. Conf. Rec. 41st Asilomar Conf. Signals, Syst. Comput.*, Nov. 2007, pp. 1224–1228.

[29] M. Shenouda and T. Davidson, "Convex conic formulations of robust downlink precoder designs with quality of service constraints," *IEEE J. Sel. Topics Signal Process.*, vol. 1, no. 4, pp. 714–724, Dec. 2017.

[30] Y. C. Eldar and N. Merhav, "A competitive minimax approach to robust estimation of random parameters," *IEEE Trans. Signal Process.*, vol. 52, no. 7, pp. 1931–1946, Jul. 2004.

[31] S. A. Vorobyov, H. Chen, and A. B. Gershman, "On the relationship between robust minimum variance beamformers with probabilistic and worst-case distortionless response constraints," *IEEE Trans. Signal Process.*, vol. 56, no. 11, pp. 5719–5724, Nov. 2008.

[32] C. Shen *et al.* "Distributed robust multicell coordinated beamforming with imperfect CSI: An ADMM approach," *IEEE Trans. Signal Process.*, vol. 60, no. 6, pp. 2988–3003, Jun. 2012.

[33] M. B. Shenouda and T. N. Davidson, "Nonlinear and linear broadcasting with QoS requirements: Tractable approaches for bounded channel uncertainties," *IEEE Trans. Signal Process.*, vol. 57, no. 5, pp. 1936–1947, May 2009.

[34] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, Jan. 2011.

[35] J. M. Cioffi. *Signal Processing and Detection*. Accessed: Mar. 20, 2019. [Online]. Available: https://web.stanford.edu/group/cioffi/book

[36] D. P. Palomar and Y. Jiang, "MIMO transceiver design via majorization theory," *Found. Trends Commun. Inf. Theory*, vol. 3, nos. 4–5, pp. 331–551, Nov. 2006.

[37] Y.-P. Lin, S.-M. Phoong, and P. P. Vaidyannathan, *Filter Bank Transceivers for OFDM and DMT Systems*. Cambridge, U.K.: Cambridge Univ. Press, 2010.
[38] D. Tse and P. Viswanath, *Fundamentals Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2005.
[39] L. Grippo and M. Sciandrone, "On the convergence of the block nonlinear Gauss–Seidel method under convex constraints," *Oper. Res. Lett.*, vol. 26, no. 3, pp. 127–136, 2000.
[40] Y. Huang and D. P. Palomar, "Rank-constrained separable semidefinite programming with applications to optimal beamforming," *IEEE Trans. Signal Process.*, vol. 58, no. 2, pp. 664–678, Feb. 2010.
[41] W. Ai, Y. Huang, and S. Zhang, "New results on Hermitian matrix rank-one decomposition," *Math. Program.*, vol. 128, nos. 1–2, pp. 253–283, Jun. 2011.
[42] D. Bertsekas and J. Tsitsiklis, *Parallel and Distributed Computation*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1989.
[43] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, NY, USA: Cambridge Univ. Press, 2004.
[44] I. Polik and T. Terlaky, "Interior point methods for nonlinear optimization," in *Nonlinear Optimization*, G. Di Pillo and F. Schoen, Eds., 1st ed. Berlin, Germany: Springer, 2010.
[45] A. Ben-Tal and A. Nemirovski, *Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications* (MPSSIAM Series on Optimization). Philadelphia, PA, USA: SIAM, 2001.
[46] M. Grant and S. Boyd. (Apr. 2010). *CVX: MATLAB Software for Disciplined Convex Programming*. [Online]. Available: http://cvxr.com/cvx
[47] N. N. Sirhan1 and S. I. Serhan, "Multi-core processors: Concepts and implementations" *Int. J. Comput. Sci. Inf. Technol.*, vol. 10, no. 1, pp. 1–10, Feb. 2018.

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 error correction coding, beamforming in multi-input multi-output 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. In 2003, she joined the Electrical and Computer Engineering Department, Lehigh University, Bethlehem, PA, USA, where she was promoted to an 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 2005 and the TAMU Ethel Ashworth-Tsutsui Memorial Award for Research 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 a Symposium Co-Chair for the Chinacom 2015, the ICC 2013, the WCSP 2012, the ICC 2008, the Chinacom 2006, the GLOBECOM 2005, and the Wirelesscom 2005. She also served as an Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS from 2008 to 2011, as an Associate Editor for the IEEE COMMUNICATIONS LETTERS from 2004 to 2008, and as an Editorial Board Member for the IEEE COMMUNICATIONS SURVEYS AND TUTORIALS from 2008 to 2010.

Hao Wang (M'18) received the B.S. and M.S. degrees in mathematics from Beihang University, Beijing, China, and the Ph.D. degree from the Industrial and Systems Engineering Department, Lehigh University, Bethlehem, PA, USA, in 2015. He was with the ExxonMobil Corporate Strategic Research Laboratory, Mitsubishi Electric Research Laboratories, and GroupM Research and Development. He is currently an Assistant Professor with the School of Information Science and Technology, ShanghaiTech University, Shanghai, China. His interests include nonlinear optimization and machine learning.