Bayesian Learning-based Linear Decentralized Sparse Parameter Estimation in MIMO Wireless Sensor Networks Relying on Imperfect CSI

Kunwar Pritiraj Rajput, *Graduate Student Member, IEEE,* Abhishek Kumar, Suraj Srivastava, *Graduate Student Member, IEEE,* Aditya K. Jagannatham, *Member, IEEE,* and Lajos Hanzo, *Fellow, IEEE*

***Abstract*—Optimal linear minimum mean square error (MMSE) transceiver design techniques are proposed for Bayesian learning (BL)-based sparse parameter vector estimation in a multiple-input multiple-output (MIMO) wireless sensor network (WSN). Our proposed transceiver designs rely on majorization theory and hyperparameter estimates obtained from the BL module for minimizing the mean square error (MSE) of parameter estimation at the fusion center (FC). The linear transceiver design framework is initially proposed for the general scenario with arbitrary SNR sensor observations, followed by a special case with high-SNR sensor observations scenario. Our analysis also incorporates the channel correlation. The MMSE channel estimates are determined for the sensors (SNs), followed by a robust transceiver design procedure that is resilient to the channel state information (CSI) uncertainty arising due to the channel estimation error, an aberration that is unavoidable in practical implementations. Our simulation results demonstrate the improved performance of the proposed BL framework and optimal MMSE transceiver design in sparse parameter estimation relying on realistic imperfect channel estimates over the benchmarks.**

***Index Terms*—Sensor networks, multiple access channel (MAC), Bayesian learning (BL), decentralized estimation, sparse parameter estimation, transceiver design, stochastic CSI uncertainty**

I. Introduction

The Internet of things (IoT), which connects a large number of devices to the network, has enjoyed popularity owing to its ability to support cutting edge applications, such as smart homes/ cities, remote healthcare, industrial automation, among several others. At the heart of a typical IoT implementation is a large number of miniature sensor nodes (SNs) or sensor-equipped devices, which continuously sense the events/ phenomena of interest. After conversion to electrical signals and suitable processing, the SNs transmit their measurements over wireless links to a central entity termed the fusion center (FC) for further processing toward efficient estimation of multiple parameters. Needless to say, the reliable operation of any IoT application critically depends on the accuracy of the parameter estimates at the FC, with erroneous estimates potentially leading to disruptions, and in extreme cases, a complete breakdown of services. This task becomes even more challenging due to the severe bandwidth and power limitations at the SNs of the IoT network. Thus, it is imperative to improve the accuracy of parameter estimation subject to the limited resources of a typical sensor network. To this end, multiple-input multiple-output (MIMO) technology can play a significant role in sensor networks by facilitating simultaneous estimation of multiple parameters at high accuracy via its spatial multiplexing and diversity gain. MIMO transmission, coupled with efficient transmit pre-processing in the form of precoding at the SNs and receiver combining (RC) at the FC, can play a key role in obtaining accurate parameter estimates. Several impressive contributions have addressed various facets of this complex and interesting problem. A brief review of the salient contributions is presented next.

A. Prior work

Linear decentralized estimation schemes for WSNs have been the focus of [1]–[4], where linear transmit precoders (TPCs) are used at the SNs complemented by linear RCs at the FC, for MSE minimization. In such a framework, the SNs typically use an amplify and forward strategy for transmitting their observations to the FC, because such an analog forwarding strategy is known to achieve the best power versus distortion trade-off for Gaussian parameter estimation for transmission over an additive white Gaussian noise (AWGN) channel, as shown in [5]. More importantly, it has also been established in [5], [6], that linear precoding of a Gaussian source is optimal in terms of MSE for transmission over a multiple access channel (MAC) channel, which motivated other researchers to focus on optimal linear processing schemes. The authors of [2] proposed one of the first models for vector parameter estimation in a MIMO sensor network relying on a coherent wireless MAC. However, in order to simplify the analysis, the MIMO channel matrix between each SN and the FC was assumed to be diagonal in nature, which restricts the applicability of the framework

L. Hanzo would like to acknowledge the financial support of the Engineering and Physical Sciences Research Council projects EP/P034284/1 and EP/P003990/1 (COALESCE) as well as of the European Research Council's Advanced Fellow Grant QuantCom (Grant No. 789028). A. K. Jagannatham would like to acknowledge that this research is supported in part by the Science and Engineering Research Board (SERB), Department of Science and Technology, Government of India, in part by the Space Technology Cell, IIT Kanpur, in part by the IIMA IDEA Telecom Centre of Excellence, in part by the Qualcomm Innovation Fellowship, and in part by the Arun Kumar Chair Professorship.

K. P. Rajput, S. Srivastava and A. K. Jagannatham are with the Department of Electrical Engineering, Indian Institute of Technology, Kanpur, Kanpur, 208016, India (e-mail: pratiraj@iitk.ac.in, ssrivast@iitk.ac.in, adityaj@iitk.ac.in.)

A. Kumar is with Qualcomm India Pvt. Ltd., Hyderabad, India (e-mail: akum@qti.qualcomm.com)

L. Hanzo is with the School of Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, U.K. (e-mail: lh@ecs.soton.ac.uk)

TABLE I
CONTRASTING OUR CONTRIBUTIONS TO THE LITERATURE

Feature	[3],[7]	[12]	[17],[21]-[24]	[36]	[37]	[38]	[52]	our work
MIMO sensor network	✓	×	×	✓	✓	✓	✓	✓
Decentralized estimation	✓	✓	×	✓	✓	✓	✓	✓
Coherent MAC	✓	✓	×	✓	✓	✓	✓	✓
Classical parameter estimation	×	×	×	×	×	×	✓	×
Bayesian parameter estimation	✓	✓	×	✓	✓	✓	×	×
Antenna correlation	×	×	×	×	×	×	×	✓
Robust design for CSI uncertainty	×	×	×	✓	✓	✓	×	✓
Joint channel and parameter est.	×	×	×	×	✓	×	×	✓
Closed-form estimate	✓	×	×	✓	✓	×	✓	✓
Arbitrary observation SNR	✓	✓	✓	✓	✓	✓	✓	✓
High observation SNR	×	×	×	✓	×	×	✓	✓
Theoretical analysis for MSE floor	×	×	×	×	×	×	×	✓

proposed therein. This limitation was subsequently overcome by the framework developed in [3] that proposed an efficient iterative procedure for optimal MMSE transceiver design for vector parameter estimation in MIMO sensor networks. Although the algorithm developed in [3] is quite general, it is only applicable in scenarios wherein the number of transmit antennas (TAs) at the SNs is lower than the number of receive antennas (RAs) at the FC or the dimension of the parameter vector of interest. The authors of [7] conceive decentralized and distributed algorithms based on the two-block coordinate descent (2-BCD) framework for achieving the MMSE. Whilst the transceiver designs proposed in [7] are novel, they are based on second order cone programming, which leads to a high computational complexity. Other recent treatises, such as [8]–[13], have also addressed this challenging problem. In [8], the authors propose optimal linear TPC designs maximizing the throughput, power and energy efficiency of a MIMO-aided WSN. The authors of [9] design their TPC for maximizing the transmission rate between the FC and SNs. Linear TPCs have also been developed for systems having a very large number of antennas at the FC [10], [11]. Liu and Chang [12] have proposed distributed estimation schemes both for deterministic as well as random parameters for a hybrid MAC relying on a combination of coherent and orthogonal MACs. Zhan *et al.* [13] have proposed trajectory design for efficient measurement collection from SNs in an unmanned ariel vehicle (UAV)-aided WSN. The authors of [14] have proposed an efficient framework for the detection and estimation of a Gaussian signal in a massive MIMO based WSN. Ciuonzo *et al.* [15] have proposed an interesting decision fusion scheme which reports the local decision of the sensors to the FC over the MAC for an improved global decision, once again in a massive MIMO WSN. This novel decision fusion paradigm is further explored in [16] for wideband collaborative spectrum sensing and sharing in a cognitive radio network.

At this juncture, it is worth noting that many researchers have reported several instances where the observations of the SNs are sparse in nature [17]. For instance, a compressive data gathering framework has been designed for large sensor networks [18], which has been shown to be well-suited for sensing image, audio and other naturally occurring signals that are sparse in a suitable transform domain, as exemplified by image signals in the wavelet domain. Furthermore, the parameter vectors of a narrowband stationary process exhibit low-rate temporal fluctuation, hence they can be readily captured using delta or differential pulse coded modulation (DPCM) for encoding the sparse innovations rather than the observations. This becomes even more critical in an IoT setup, since the data compression arising from exploiting sparsity may lead to significant bandwidth saving [19], [20]. Thus, sparse signal recovery from sensor transmissions has emerged as an important research area. The authors of [21]–[25] propose distributed schemes for joint sparse signal recovery both with and without quantization where each sensor observes the signal of interest and subsequently computes an estimate of the underlying parameter by exchanging its observations exclusively with its single hop neighbours. Although the distributed scheme is robust to node failures, it necessitates inter-sensor communication that potentially imposes a high control/ communication cost on the system. The problem of detecting a sparse signal in presence of non-Gaussian noise is explored in [26], [27]. The authors of [26] have considered generalized-Gaussian sparse parameters and noise in their analysis and conceived a novel scheme for parameter detection using 1-bit quantized measurements to reduce the bandwidth required. Jacques *et al.* have proposed another interesting framework [27] for the detection of sparse signal considering non-Gaussian noise relying on uniformly quantized measurement transmissions. However, to the best of our knowledge, none of the existing contributions consider the problem of sparse parameter estimation in a network wherein multiple SNs observe a spatially correlated sparse signal, followed by optimal pre-processing and recovery at the FC.

Furthermore, most of the above contributions assume the availability of perfect CSI at the FC, which is unrealizable in practice. Therefore, it is vitally important to develop TPC/ RC design techniques relying on imperfect CSI. In this context, the authors of [28]–[31], have considered training-based channel estimation followed by imperfect CSI based transceiver designs. However, these designs do not proactively take the CSI imperfections into account which represent channel estimation error. As a further development Banavar *et al.* [32] design a scheme for distributed estimation relying on partial channel feedback and analyze its performance. The asymptotic performance derived for a large number of SNs communicating over different fading channels and using feedback are also presented in [32]. Based on this framework, the authors of [33], [34] have proposed limited feedback based optimal TPC codebook designs. However, designing such a codebook requires solving complex optimization problems. Zhu *et al.* [35], Venkategowda *et al.* [36], and Liu *et al.* [37] have proposed robust TPC designs for scalar parameter estimation considering different CSI uncertainty models. As a further advance, Rostami and Falahati [38] have proposed robust TPC designs for vector parameter estimation considering CSI uncertainty. However, the TPC designs of [36] and [38] are based on the zero-forcing (ZF) criterion, and hence result in noise enhancement at the FC. Moreover, all the above-mentioned contributions have considered only non-sparse parameter estimation and do not include the realistic effects of antenna correlations at the SNs and FC in their analysis.

To overcome the above limitations we harness the powerful Bayesian learning (BL) framework [39] for sparse signal estimation, explicitly, we propose novel BL-based linear decentralized TPC/ RC design techniques for sparse vector parameter recovery in a sensor network explicitly considering the CSI uncertainty. The key reason behind choosing the BL framework over other existing sparse estimation schemes, such as FOCUSS (FOcal Underdetermined System Solver) [40], BP (Basis Pursuit) [41], and MP (Matching Pursuit) [42] is that the global minima of the BL-cost function guarantees the sparsest representation of the observation using the columns of the dictionary matrix as demonstrated in [39]. Furthermore, its convergence to a fixed point of the log-likelihood is guaranteed from any initialization by virtue of the well-established expectation-maximization (EM) framework. The new contributions of this paper are clearly itemized next and they are also boldly contrasted to the literature in Table-I. The related existing contributions in the literature have also been classified into two broad categories: classical[1] and Bayesian[2] estimation.

B. New Contributions and Organization of the Work

- We propose attractive TPC/ RC designs for sparse parameter estimation in MIMO WSNs relying on imperfect CSI. In order to exploit the sparsity of the unknown parameter vector, a BL-based scheme is conceived for linear decentralized estimation. The key advantage of this BL-based sparse parameter estimation scheme is that it yields the 'maximally-sparse' solution and guaranteed convergence at a low number of iterations.
- First, majorization theory based optimal MMSE TPC/ RC designs are developed for the more general scenario of arbitrary-SNR noisy sensor observations for minimizing the MSE of parameter estimation at the FC. Subsequently, linear transceivers are derived for a scenario of high-SNR sensor observations.
- The effects of transmit and receive antenna correlations have also been incorporated in the parameter estimation process and its effects on the performance are analyzed.
- The MMSE channel estimate is determined by incorporating the channel correlation, followed by characterizing the resultant channel estimation error.
- A robust linear transceiver design is proposed that directly incorporates the CSI uncertainty.
- Finally, simulation results are presented to show the efficacy of the proposed schemes.

The rest of the paper is organized as follows. Section-II describes the MIMO system model of the linear decentralized estimation of a sparse vector parameter. Subsequently, Section-III describes the BL technique and majorization theory based optimal MMSE TPC/ RC design minimizing the MSE at the FC for an arbitrary SNR sensor observations scenario, followed by an interesting scenario with high-SNR sensor observations in Section-III-A. In Section-IV, MMSE-based channel estimation is presented, followed by our robust transceiver design. This is followed by our simulation results in Section-V and concluding remarks in Section-VI. For seamless reading, the proofs of the different propositions are relegated to the Appendices.

A collective description of the notation used in this paper is as follows. The Hermitian and transpose operations are denoted by $(.)^H$ and $(.)^T$, respectively. $\mathrm{Tr}(\mathbf{A})$ denotes the trace of the matrix $\mathbf{A}$, $\mathbf{A} = \mathrm{diag}(\mathbf{a})$ defines a diagonal matrix that contains the elements of the vector $\mathbf{a}$ on its principal diagonal, and $\mathbf{a} = \mathrm{diag}(\mathbf{A})$ is a vector whose elements are the diagonal elements of matrix $\mathbf{A}$. $\mathbf{A} = \mathrm{blkdiag}[\mathbf{A}_1, \mathbf{A}_2, \cdots, \mathbf{A}_L]$ denotes a block-diagonal matrix $\mathbf{A}$ with the matrices $\mathbf{A}_i$'s, $1 \leq i \leq L$, on its principal diagonal. Also, $\mathbb{E}[\cdot]$ represents the expectation operator. $\mathcal{CN}(\boldsymbol{\mu}_x, \mathbf{R}_x)$ denotes the probability density function (pdf) of a symmetric complex Gaussian random vector $\mathbf{x} \in \mathbb{C}^{m\times 1}$ with mean $\boldsymbol{\mu}_x \in \mathbb{C}^{m\times 1}$ and covariance matrix $\mathbf{R}_x \in \mathbb{C}^{m\times m}$. The symbol $\otimes$ denotes the Kronecker product. $\mathbf{I}_n$ represents an $n \times n$ identity matrix, while $\lambda_{\max}$ denotes the maximum eigenvalue of the matrix $\mathbf{X}$.

II. MIMO WSN System and Channel Model

Consider a MIMO WSN comprised of L SNs, with each sensor communicating with the FC over a coherent MAC as shown in Fig 1. Let t_i and r denote the number of TAs at SN i, $1 \leq i \leq L$, and the number of RAs at the FC, respectively. Similar to other contributions [2], [3], the measurement vector $\mathbf{x}_i \in \mathbb{C}^{q_i\times 1}$ of the ith SN can be modeled as

$$\mathbf{x}_i = \mathbf{G}_i\boldsymbol{\theta} + \mathbf{v}_i, \tag{1}$$

where $\mathbf{G}_i \in \mathbb{C}^{q_i\times m}$ represents the observation matrix for the ith SN and the number of observations $q_i,\ 1 \leq i \leq L$, is typically less than the length of the parameter vector [2], which allows us to compress the measurements. The quantity $\mathbf{v}_i \in \mathbb{C}^{q_i\times 1}$ denotes the additive observation noise at the ith SN that has been modeled as $\mathcal{CN}(\mathbf{0}, \mathbf{R}_{v,i})$. The quantity $\boldsymbol{\theta} \in \mathbb{C}^{m\times 1}$ denotes the parameter vector to be estimated. This work considers the parameter vector $\boldsymbol{\theta}$ to be sparse in nature.

[1]In the classical framework, the unknown parameter is considered as a deterministic quantity and the techniques such as Maximum Likelihood Estimate (MLE) [43, Sec. 7.3] and Best Linear Unbiased Estimate (BLUE) are used [43, Sec. 6.4] for estimation.

[2]In Bayesian parameter estimation, the unknown parameter is considered to be random in nature with a known *a priori* distribution. Subsequently, one can employ the MMSE/ MAP [43, Sec. 11.4/11.5] and LMMSE [43, Sec. 12.3] schemes for parameter estimation.

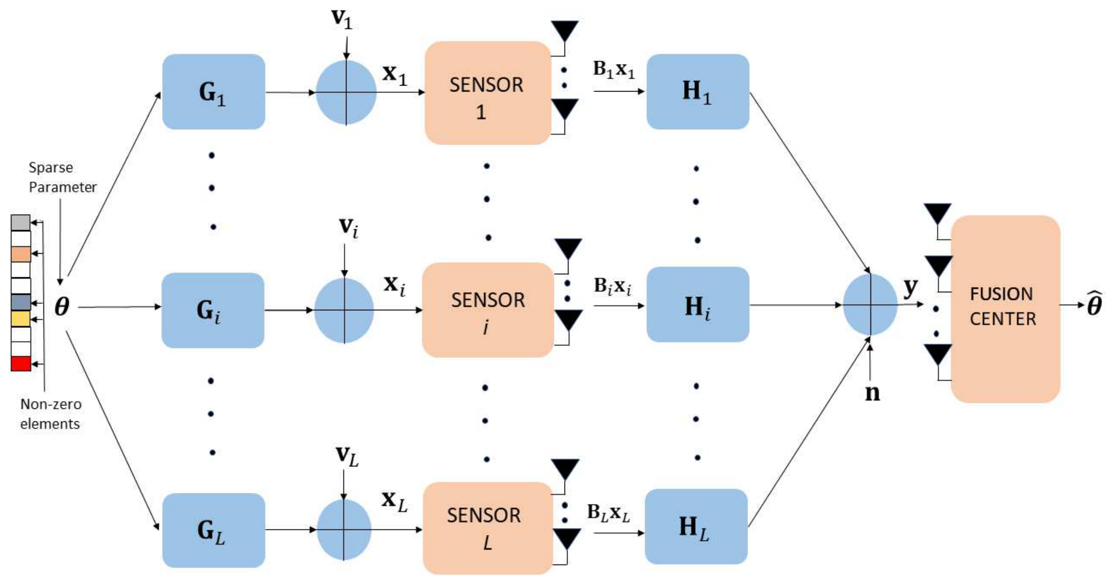


Fig. 1. System model for sparse parameter estimation in the MIMO wireless sensor network

Note that $\boldsymbol{\theta}$ can be either itself sparse or rendered sparse in many scenarios with suitable processing. For instance, since the parameter vectors of a narrowband stationary process exhibit low-rate temporal fluctuation, they can be readily captured using delta or DPCM for encoding the resultant sparse innovations rather than the observations. On the other hand, many naturally occurring signals are sparse in a suitable transform domain [44], as exemplified by image signals in the wavelet domain. In such scenarios, the original non-sparse parameter $\widetilde{\boldsymbol{\theta}}$ can be expressed as $\widetilde{\boldsymbol{\theta}} = \boldsymbol{\Psi}\boldsymbol{\theta}$, where $\boldsymbol{\Psi}$ represents a transformation matrix [44], also known as sparsifying-dictionary, and $\boldsymbol{\theta}$ denotes the sparse representation of $\widetilde{\boldsymbol{\theta}}$. For this case, the measurement vector $\mathbf{x}_i$ of the ith SN can be modeled as $\mathbf{x}_i = \mathbf{G}_i\boldsymbol{\Psi}\boldsymbol{\theta} + \mathbf{v}_i = \widetilde{\mathbf{G}}_i\boldsymbol{\theta} + \mathbf{v}_i$, where $\widetilde{\mathbf{G}}_i = \mathbf{G}_i\boldsymbol{\Psi}$. Therefore, without loss of generality, we proceed with the measurement model considered in (1). The observation vector $\mathbf{x}_i$ obtained at each SN is first passed through a whitening filter $\mathbf{W}_i \in \mathbb{C}^{q_i \times q_i}$, which yields the whitened observation vector $\widetilde{\mathbf{x}}_i \in \mathbb{C}^{q_i \times 1}$ as $\widetilde{\mathbf{x}}_i = \mathbf{W}_i\mathbf{x}_i = \mathbf{W}_i\mathbf{G}_i\boldsymbol{\theta} + \mathbf{W}_i\mathbf{v}_i$. The whitened observation vector $\widetilde{\mathbf{x}}_i$ at the ith SN is then pre-processed by the TPC matrix $\mathbf{B}_i \in \mathbb{C}^{t_i \times q_i}$, followed by transmission to the FC. Finally, the signal vector $\mathbf{y} \in \mathbb{C}^{r \times 1}$ received by the FC over the coherent MAC is given by

$$\mathbf{y} = \sum_{i=1}^{L} \mathbf{H}_i\mathbf{B}_i\mathbf{W}_i\mathbf{G}_i\boldsymbol{\theta} + \sum_{i=1}^{L} \mathbf{H}_i\mathbf{B}_i\mathbf{W}_i\mathbf{v}_i + \mathbf{n}, \qquad (2)$$

where $\mathbf{H}_i \in \mathbb{C}^{r \times t_i}$ denotes the channel matrix between the ith SN and the FC, while the vector $\mathbf{n} \in \mathbb{C}^{r \times 1}$ denotes the AWGN at the FC modeled as $\mathcal{CN}(\mathbf{0}, \mathbf{R}_n)$. In order to incorporate correlation, the MIMO channel matrix $\mathbf{H}_i$ between the ith SN and the FC has been modeled using the standard Kronecker MIMO channel model, as described in [45],

$$\mathbf{H}_i = \left(\mathbf{R}^{Rx}\right)^{1/2} \mathbf{S}_i \left(\mathbf{R}_i^{Tx}\right)^{T/2}, \qquad (3)$$

where the matrices $\mathbf{R}^{Rx} \in \mathbb{C}^{r \times r}$ and $\mathbf{R}_i^{Tx} \in \mathbb{C}^{t_i \times t_i}$ denote the spatial receive and transmit correlation matrices, respectively. The quantity $\mathbf{S}_i \in \mathbb{C}^{r \times t_i}$ represents a stochastic matrix, whose elements are independent and identically distributed (i.i.d) as $\mathcal{CN}(0, 1)$. Thus, when the antenna correlations are negligible, i.e., $\mathbf{R}^{Rx} \approx \mathbf{I}_r$ and $\mathbf{R}_i^{Tx} \approx \mathbf{I}_{t_i}$, the channel model of (3) reduces to an uncorrelated MIMO channel comprising i.i.d. elements. The resultant covariance matrix of the vectorized channel $\mathbf{h}_i = \text{vec}(\mathbf{H}_i)$, denoted by $\mathbf{R}_i \in \mathbb{C}^{rt_i \times rt_i}$, can be obtained as

$$\mathbf{R}_i = \mathbb{E}\left[\mathbf{h}_i\mathbf{h}_i^H\right] = \mathbf{R}_i^{Tx} \otimes \mathbf{R}^{Rx}. \qquad (4)$$

Let the block diagonal TPC, whitened matrix and observation noise covariance matrices, denoted by $\mathbf{B} \in \mathbb{C}^{t \times q}$, $\mathbf{W} \in \mathbb{C}^{q \times q}$ and $\mathbf{R}_v \in \mathbb{C}^{q \times q}$, respectively, with $t = \sum_{i=1}^{L} t_i$ and $q = \sum_{i=1}^{L} q_i$, be defined as

$$\begin{aligned} \mathbf{B} &= \text{blkdiag}[\mathbf{B}_1, \mathbf{B}_2, \cdots, \mathbf{B}_L], \\ \mathbf{R}_v &= \text{blkdiag}[\mathbf{R}_{v,1}, \mathbf{R}_{v,2}, \cdots, \mathbf{R}_{v,L}], \\ \mathbf{W} &= \text{blkdiag}[\mathbf{W}_1, \mathbf{W}_2, \ldots, \mathbf{W}_L], \end{aligned} \qquad (5)$$

while the concatenated observation and channel matrices $\mathbf{G} \in \mathbb{C}^{q \times m}$ and $\mathbf{H} \in \mathbb{C}^{r \times t}$, respectively, be defined as

$$\mathbf{G} = [\mathbf{G}_1^T, \mathbf{G}_2^T, \cdots, \mathbf{G}_L^T]^T, \mathbf{H} = [\mathbf{H}_1, \mathbf{H}_2, \cdots, \mathbf{H}_L]. \qquad (6)$$

The equivalent system model in (2) can be written in a compact form as

$$\mathbf{y} = \mathbf{HBWG}\boldsymbol{\theta} + \mathbf{HBWv} + \mathbf{n}, \qquad (7)$$

where the stacked observation noise vector is $\mathbf{v} = \left[\mathbf{v}_1^T, \mathbf{v}_2^T, \cdots, \mathbf{v}_L^T\right]^T \in \mathbb{C}^{q \times 1}$. Let the stacked observation vector be $\mathbf{x} = \left[\mathbf{x}_1^T, \mathbf{x}_2^T, \cdots, \mathbf{x}_L^T\right]^T \in \mathbb{C}^{q\times 1}$. The average transmit power of the ith SN can be evaluated as

$$\begin{aligned}\mathbb{E}\left[\|\mathbf{B}_i\widetilde{\mathbf{x}}_i\|^2\right] &= \operatorname{Tr}\left(\mathbf{B}_i\mathbb{E}\left[\widetilde{\mathbf{x}}_i\widetilde{\mathbf{x}}_i^H\right]\mathbf{B}_i^H\right)\\ &= \operatorname{Tr}\left(\mathbf{B}_i\left(\mathbf{W}_i\mathbf{G}_i\mathbf{R}_\theta\mathbf{G}_i^H\mathbf{W}_i^H + \mathbf{W}_i\mathbf{R}_{v,i}\mathbf{W}_i^H\right)\mathbf{B}_i^H\right)\\ &= \operatorname{Tr}\left(\mathbf{B}_i\left(\mathbf{W}_i\mathbf{G}_i\mathbf{R}_\theta\mathbf{G}_i^H\mathbf{W}_i^H + \mathbf{I}_{q_i}\right)\mathbf{B}_i^H\right),\end{aligned} \tag{8}$$

where $\mathbf{W}_i\mathbf{R}_{v,i}\mathbf{W}_i^H = \mathbf{I}_{q_i}$ and $\mathbf{R}_\theta \in \mathbb{C}^{m\times m}$ is the prior covariance matrix of the parameter vector $\boldsymbol{\theta}$. From (8), the average total transmit power of the WSN can be expressed as

$$\sum_{i=1}^{L}\mathbb{E}\left[\|\mathbf{B}_i\widetilde{\mathbf{x}}_i\|^2\right] = \operatorname{Tr}\left(\mathbf{B}\left(\mathbf{W}\mathbf{G}\mathbf{R}_\theta\mathbf{G}^H\mathbf{W}^H + \mathbf{I}_q\right)\mathbf{B}^H\right). \tag{9}$$

Let $\hat{\boldsymbol{\theta}}$ denote the estimate of the sparse parameter vector $\boldsymbol{\theta}$ obtained after the subsequent processing at the FC. The error covariance matrix $\mathbf{E} \in \mathbb{C}^{m\times m}$ pertaining to the estimate of the parameter $\boldsymbol{\theta}$, and the resultant MSE are defined as

$$\mathbf{E} = \mathbb{E}\left[\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right)\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right)^H\right],\ \text{MSE} = \operatorname{Tr}(\mathbf{E}). \tag{10}$$

The problem of optimizing the sensor precoding matrices $\mathbf{B}_i,\ 1 \le i \le L$, which minimize the MSE of estimation at the FC, subject to the power constraint, can be formulated as

$$\begin{aligned}&\underset{\mathbf{B}}{\text{minimize}} && \text{MSE}\\ &\text{subject to} && \operatorname{Tr}\left(\mathbf{B}\left(\mathbf{W}\mathbf{G}\mathbf{R}_\theta\mathbf{G}^H\mathbf{W}^H + \mathbf{I}_q\right)\mathbf{B}^H\right) \le P_T,\end{aligned} \tag{11}$$

where P_T denotes the total transmit power budget of the WSN. The BL-based estimation framework of the sparse parameter vector $\boldsymbol{\theta}$ and the optimal MMSE TPC designs for an arbitrary-SNR scenarios is described next.

III. BL-based Sparse Parameter Estimation and TPC Design for Noisy Sensor Observations

The received signal in (7) can be further written in a compact form as

$$\mathbf{y} = \widetilde{\mathbf{F}}\widetilde{\mathbf{G}}\boldsymbol{\theta} + \widetilde{\mathbf{F}}\mathbf{W}\mathbf{v} + \mathbf{n}, \tag{12}$$

where the matrices $\widetilde{\mathbf{G}} \in \mathbb{C}^{q\times m}$ and $\widetilde{\mathbf{F}} = \mathbf{H}\mathbf{B} \in \mathbb{C}^{r\times q}$ are defined as

$$\begin{aligned}\widetilde{\mathbf{G}} &= \left[\widetilde{\mathbf{G}}_1^T, \widetilde{\mathbf{G}}_2^T, \ldots, \widetilde{\mathbf{G}}_L^T\right]^T\\ &= \left[\mathbf{G}_1^T\mathbf{W}_1^T, \mathbf{G}_2^T\mathbf{W}_2^T, \ldots, \mathbf{G}_L^T\mathbf{W}_L^T\right]^T,\end{aligned} \tag{13}$$

$$\widetilde{\mathbf{F}} = \left[\widetilde{\mathbf{F}}_1, \widetilde{\mathbf{F}}_2, \cdots, \widetilde{\mathbf{F}}_L\right] = [\mathbf{H}_1\mathbf{B}_1, \mathbf{H}_2\mathbf{B}_2, \cdots, \mathbf{H}_L\mathbf{B}_L]. \tag{14}$$

The main mathematical steps of the proposed BL-based sparse parameter estimation and TPC design in this scenario are as follows.

1) Gaussian parametrized prior assignment over the sparse parameter vector $\boldsymbol{\theta}$ followed by the estimation of hyperparameters using the EM framework.
2) Estimated hyperparameters are then utilized to derive an upper bound of the MSE at the FC.
3) Finally, the majorization theory based TPC framework is once again developed for minimizing the resultant MSE.

We now derive the above steps in detail. In contrast to other popular sparse signal recovery techniques, such as the FOCUSS [40] and Basis Pursuit (BP) [41], which assume a fixed prior, the proposed BL framework begins with assigning the following parameterized Gaussian prior to the parameter vector $\boldsymbol{\theta}$ [39]

$$p(\boldsymbol{\theta};\boldsymbol{\Gamma}) = \prod_{i=1}^{m}(\pi\gamma_i)^{-1}\exp\left(\frac{-|\boldsymbol{\theta}(i)|^2}{\gamma_i}\right), \tag{15}$$

where $\gamma_i,\ 1 \le i \le m$, denotes the hyperparameter corresponding to the ith component of $\boldsymbol{\theta}$ and $\boldsymbol{\Gamma}$ is the diagonal matrix of hyperparameters that can be written as $\boldsymbol{\Gamma} = \operatorname{diag}(\gamma_1, \gamma_2, \cdots, \gamma_m) \in \mathbb{R}^{m\times m}$. Note that the hyperparameters $\gamma_i,\ 1 \le i \le m$, are unknown and they have to be estimated from the observation $\mathbf{y}$. Since both the likelihood $p(\mathbf{y}|\boldsymbol{\theta})$ and prior $p(\boldsymbol{\theta};\boldsymbol{\Gamma})$, are Gaussian, the *a posteriori* pdf $p(\boldsymbol{\theta}|\mathbf{y};\boldsymbol{\Gamma})$ also becomes Gaussian [43], whose mean and covariance can be computed using closed-form expressions. Thus, the parametrized Gaussian prior assignment makes the proposed EM-based framework tractable for hyperparameter estimation, since it requires to compute the *a posteriori* pdf $p(\boldsymbol{\theta}|\mathbf{y};\boldsymbol{\Gamma})$ of the sparse parameter $\boldsymbol{\theta}$ in the expectation-step (E-step), as described subsequently in this section. Furthermore, as described in [39], for a noiseless scenario, the global-minima of the BL-cost function is achieved at the sparsest solution. Even for a noisy scenario, all the local minima also yield sparse solutions. Therefore, upon convergence of the EM-procedure, many of the γ_i values are driven to zero, effectively forcing the corresponding elements of the parameter vector $\boldsymbol{\theta}$ to zero. The linear MMSE (LMMSE) estimate $\widehat{\boldsymbol{\theta}} \in \mathbb{C}^{m\times 1}$ and the corresponding covariance matrix $\boldsymbol{\Sigma}_\theta \in \mathbb{C}^{m\times m}$ of the parameter $\boldsymbol{\theta}$ is given by

$$\boldsymbol{\mu}_\theta = \boldsymbol{\Sigma}_\theta\widetilde{\mathbf{G}}^H\widetilde{\mathbf{F}}^H\widetilde{\mathbf{R}}_n^{-1}\mathbf{y}, \tag{16}$$

$$\boldsymbol{\Sigma}_\theta = \left(\boldsymbol{\Gamma}^{-1} + \widetilde{\mathbf{G}}^H\widetilde{\mathbf{F}}^H\widetilde{\mathbf{R}}_n^{-1}\widetilde{\mathbf{F}}\mathbf{W}\widetilde{\mathbf{G}}\right)^{-1}, \tag{17}$$

where $\widetilde{\mathbf{R}}_n = \widetilde{\mathbf{F}}\mathbf{W}\mathbf{R}_v\mathbf{W}^H\widetilde{\mathbf{F}}^H + \mathbf{R}_n \in \mathbb{C}^{r\times r}$. Observe from the expressions above that the LMMSE estimate of the sparse parameter $\boldsymbol{\theta}$ depends on the hyperparameter matrix $\boldsymbol{\Gamma}$. Thus, the estimation of the sparse parameter vector $\boldsymbol{\theta}$ boils down to the estimation of the hyperparameter vector $\boldsymbol{\gamma} = [\gamma_1, \gamma_2, \cdots, \gamma_m]^T \in \mathbb{R}^{m\times 1}$. The proposed BL approach maximizes the Bayesian evidence $p(\mathbf{y};\boldsymbol{\Gamma})$ with respect to the hyperparameter $\boldsymbol{\theta}$, which leads to an enhanced sparse estimation performance. The Bayesian evidence $p(\mathbf{y};\boldsymbol{\Gamma})$ is determined by the marginal pdf of

$$p(\mathbf{y};\boldsymbol{\Gamma}) = \frac{1}{(\pi)^r\det(\boldsymbol{\Sigma}_y)}\exp\left(-\mathbf{y}^H\boldsymbol{\Sigma}_y^{-1}\mathbf{y}\right), \tag{18}$$

where $\boldsymbol{\Sigma}_y = \widetilde{\mathbf{R}}_n + \widetilde{\mathbf{F}}\widetilde{\mathbf{G}}\boldsymbol{\Gamma}\widetilde{\mathbf{G}}^H\widetilde{\mathbf{F}}^H \in \mathbb{C}^{r\times r}$ is the covariance matrix of the receive vector $\mathbf{y}$. The corresponding maximum-likelihood (ML) objective is given by

$$\begin{aligned}\widehat{\boldsymbol{\gamma}} &= \arg\max_{\boldsymbol{\gamma}}\log\left[p(\mathbf{y};\boldsymbol{\Gamma})\right]\\ &\equiv \arg\max_{\boldsymbol{\gamma}} -\log\left[\det(\boldsymbol{\Sigma}_y)\right] - \mathbf{y}^H\boldsymbol{\Sigma}_y^{-1}\mathbf{y}.\end{aligned} \tag{19}$$

As discussed in [39], the direct maximization of the above optimization objective with respect to $\boldsymbol{\gamma}$ becomes intractable. Therefore, the proposed BL-framework employs the well-established iterative EM algorithm [46], [47]. The key steps in the EM procedure are outlined below.

Let $\widehat{\boldsymbol{\Gamma}}^{(k-1)}$ and $\mathbf{B}^{(k-1)}$ denote the estimates of the hyper-parameter and TPC matrices in the $(k-1)$st EM iteration[3]. Employing the TPC matrix $\mathbf{B}^{(k-1)}$, one obtains the vector $\mathbf{y}^{(k)}$ received at FC as

$$\mathbf{y}^{(k)} = \mathbf{H}\mathbf{B}^{(k-1)}\mathbf{W}\mathbf{G}\boldsymbol{\theta}^{(k)} + \mathbf{n}^{(k)} = \widetilde{\mathbf{F}}^{(k-1)}\widetilde{\mathbf{G}}\boldsymbol{\theta}^{(k)} + \mathbf{n}^{(k)}, \tag{20}$$

where $\boldsymbol{\theta}^{(k)}$ denotes the sparse parameter vector at time-instant k. Furthermore, it is assumed that the sparsity profile, i.e. the locations of the non-zero elements, of the sparse parameter vector $\boldsymbol{\theta}^{(k)}$ do not change. In the kth iteration, the expectation step (E-step) evaluates the average log-likelihood $\mathcal{L}\left(\boldsymbol{\Gamma} \mid \widehat{\boldsymbol{\Gamma}}^{(k-1)}\right)$ of the complete-data set $\left\{\mathbf{y}^{(k)}, \boldsymbol{\theta}^{(k)}\right\}$ as

$$\mathcal{L}\left(\boldsymbol{\Gamma} \mid \widehat{\boldsymbol{\Gamma}}^{(k-1)}\right) = \mathbb{E}_{\boldsymbol{\theta}^{(k)}|\mathbf{y}^{(k)};\widehat{\boldsymbol{\Gamma}}^{(k-1)}}\left[\log\left[p\left(\mathbf{y}^{(k)}, \boldsymbol{\theta}^{(k)}; \boldsymbol{\Gamma}\right)\right]\right]$$
$$= \mathbb{E}\left[\log\left[p\left(\mathbf{y}^{(k)} \mid \boldsymbol{\theta}^{(k)}\right)\right] + \log\left[p\left(\boldsymbol{\theta}^{(k)}; \boldsymbol{\Gamma}\right)\right]\right]. \tag{21}$$

The first term inside the expectation operator in (21) can be simplified as

$$\log\left[p\left(\mathbf{y}^{(k)} \mid \boldsymbol{\theta}^{(k)}\right)\right]$$
$$= -\kappa_1 - \left(\mathbf{y}^{(k)} - \widetilde{\mathbf{F}}^{(k)}\widetilde{\mathbf{G}}\boldsymbol{\theta}^{(k)}\right)^H \widetilde{\mathbf{R}}_n^{-1}\left(\mathbf{y}^{(k)} - \widetilde{\mathbf{F}}^{(k)}\widetilde{\mathbf{G}}\boldsymbol{\theta}^{(k)}\right), \tag{22}$$

where the constant $\kappa_1 = \log\left[(\pi)^r \det\left(\mathbf{R}_n\right)\right]$. It can be observed that $\log\left[p\left(\mathbf{y}^{(k)} \mid \boldsymbol{\theta}^{(k)}\right)\right]$ is independent of the hyper-parameter vector $\boldsymbol{\gamma}$ and hence can be ignored in the maximization of the subsequent M-step. To evaluate the second term inside the expectation in (21), the *a posteriori* pdf of the parameter $\boldsymbol{\theta}^{(k)}$ can be evaluated as [43]

$$p\left(\boldsymbol{\theta}^{(k)} \mid \mathbf{y}^{(k)}; \widehat{\boldsymbol{\Gamma}}^{(k-1)}\right) = \mathcal{CN}\left(\boldsymbol{\mu}_\theta^{(k)}, \boldsymbol{\Sigma}_\theta^{(k)}\right), \tag{23}$$

where the *a posteriori* mean $\boldsymbol{\mu}_\theta^{(k)} \in \mathbb{C}^{m\times 1}$ and covariance matrix $\boldsymbol{\Sigma}_\theta^{(k)} \in \mathbb{C}^{m\times m}$ can be determined as

$$\boldsymbol{\mu}_\theta^{(k)} = \boldsymbol{\Sigma}_\theta^{(k)}\widetilde{\mathbf{G}}^H\left(\widetilde{\mathbf{F}}^{(k-1)}\right)^H\widetilde{\mathbf{R}}_n^{-1}\mathbf{y}^{(k)}, \tag{24}$$

$$\boldsymbol{\Sigma}_\theta^{(k)} = \left(\left(\widehat{\boldsymbol{\Gamma}}^{(k-1)}\right)^{-1} + \widetilde{\mathbf{G}}^H\left(\widetilde{\mathbf{F}}^{(k-1)}\right)^H\widetilde{\mathbf{R}}_n^{-1}\widetilde{\mathbf{F}}^{(k-1)}\widetilde{\mathbf{G}}\right)^{-1}. \tag{25}$$

[3]The authors appreciate the helpful query from one of the anonymous Reviewers regarding how and when the observations $\mathbf{y}^{(k)}$ are obtained in this setting. We would like to clarify that in this discussion, the terms time-instant and EM iteration are used interchangeably, since one EM iteration is performed at each time-instant. Initially, the TPCs $\mathbf{B}_i^{(0)}$ for each sensor i, are designed using $\widehat{\gamma}_j^{(0)} = 1$ for $1 \leq j \leq m$, followed by receiving observations $\mathbf{y}^{(0)}$ with these precoders. Subsequently, the EM procedure updates the hyperparameters. These updated hyperparameters are employed in the next iteration to design TPCs and collecting measurements. This procedure repeats until convergence is achieved. Algorithm 1 and its graphical description in Fig. 2 clearly illustrate these aspects.

Using the above *a posteriori* pdf of $\boldsymbol{\theta}^{(k)}$, the second term in (21) can be written as

$$\mathbb{E}_{\boldsymbol{\theta}^{(k)}|\mathbf{y}^{(k)};\widehat{\boldsymbol{\Gamma}}^{(k-1)}}\left\{\log\left[p\left(\boldsymbol{\theta}^{(k)}; \boldsymbol{\Gamma}\right)\right]\right\}$$
$$= \sum_{i=1}^{r} -\log\left(\pi\gamma_i\right) - \frac{1}{\gamma_i}\mathbb{E}_{\boldsymbol{\theta}^{(k)}|\mathbf{y}^{(k)};\widehat{\boldsymbol{\Gamma}}^{(k-1)}}\left\{\left|\boldsymbol{\theta}^{(k)}(i)\right|^2\right\}. \tag{26}$$

Furthermore, the maximization problem in the M-step can be written as

$$\widehat{\boldsymbol{\gamma}}^{(k)} = \max_{\boldsymbol{\gamma}} \mathbb{E}_{\boldsymbol{\theta}^{(k)}|\mathbf{y}^{(k)};\widehat{\boldsymbol{\Gamma}}^{(k-1)}}\left\{\log\left[p\left(\boldsymbol{\theta}^{(k)}; \boldsymbol{\Gamma}\right)\right]\right\}. \tag{27}$$

It can be readily observed from (27) that the above maximization problem is decoupled with respect to each γ_i. Thus, it can be solved to obtain the estimates $\widehat{\gamma}_i^{(k)}$ as

$$\widehat{\gamma}_i^{(k)} = \mathbb{E}_{\boldsymbol{\theta}^{(k)}|\mathbf{y}^{(k)};\widehat{\boldsymbol{\Gamma}}^{(k-1)}}\left\{\left|\boldsymbol{\theta}^{(k)}(i)\right|^2\right\} = \boldsymbol{\Sigma}_\theta^{(k)}(i,i) + \mid \boldsymbol{\mu}_\theta^{(k)}(i) \mid^2. \tag{28}$$

The BL-based estimate of the hyperparameter matrix $\boldsymbol{\Gamma}$ in the kth time-instant is given as $\widehat{\boldsymbol{\Gamma}}^{(k)} = \mathrm{diag}\left(\widehat{\gamma}_1^{(k)}, \widehat{\gamma}_2^{(k)}, ..., \widehat{\gamma}_m^{(k)}\right)$, which can be subsequently employed for the TPC design. Note that the *a priori* pdf $p(\boldsymbol{\theta}; \boldsymbol{\Gamma})$ in (15) is Gaussian, furthermore, as described in (22), the likelihood $p\left(\mathbf{y} \mid \boldsymbol{\theta}\right)$, and the *a posteriori* pdf $p\left(\boldsymbol{\theta} \mid \mathbf{y}; \widehat{\boldsymbol{\Gamma}}\right)$ described in (23) both are also Gaussian. For this scenario, exploiting the well-established result given in [43, Sec. 12.3], the estimation error covariance matrix $\mathbf{E}^{(k)}$ of the LMMSE estimate in the kth EM iteration can be derived in a closed-form expression as

$$\mathbf{E}^{(k)} = \left(\left(\widehat{\boldsymbol{\Gamma}}^{(k)}\right)^{-1} + \widetilde{\mathbf{G}}^H\left(\widetilde{\mathbf{F}}^{(k)}\right)^H\widetilde{\mathbf{R}}_n^{-1}\widetilde{\mathbf{F}}^{(k)}\widetilde{\mathbf{G}}\right)^{-1}. \tag{29}$$

Let the singular value decomposition (SVD) of matrix $\widetilde{\mathbf{G}}$ be $\widetilde{\mathbf{G}} = \mathbf{U}_{\tilde{G}}\boldsymbol{\Sigma}_{\tilde{G}}\mathbf{V}_{\tilde{G}}^H$ and EVD of the matrix $\widetilde{\mathbf{R}}_n$ be $\widetilde{\mathbf{R}}_n = \mathbf{U}_n\boldsymbol{\Lambda}_n\mathbf{U}_n^H$. Substituting these in the expression for the error covariance $\mathbf{E}^{(k)}$ in (29) and exploiting the property $\mathbf{W}\mathbf{R}_v\mathbf{W}^H = \mathbf{I}_q$, the resultant MSE can be obtained as shown in (30). Choosing $\widetilde{\mathbf{F}}^{(k)} = \mathbf{U}_n\widetilde{\boldsymbol{\Omega}}^{(k)}\mathbf{U}_{\tilde{G}}^H$, the MSE expression in (30) reduces to

$$\mathrm{MSE}^{(k)}$$
$$= \mathrm{Tr}\Big(\boldsymbol{\Upsilon}^{(k)} + \boldsymbol{\Sigma}_{\tilde{G}}^H\left(\widetilde{\boldsymbol{\Omega}}^{(k)}\right)^H\left(\widetilde{\boldsymbol{\Omega}}^{(k)}\left(\widetilde{\boldsymbol{\Omega}}^{(k)}\right)^H + \boldsymbol{\Lambda}_n\right)^{-1}$$
$$\widetilde{\boldsymbol{\Omega}}^{(k)}\boldsymbol{\Sigma}_{\tilde{G}}\Big)^{-1} = \mathrm{Tr}\left(\boldsymbol{\Upsilon}^{(k)} + \mathbf{D}^{(k)}\right)^{-1}, \tag{31}$$

where $\boldsymbol{\Upsilon}^{(k)} = \mathbf{V}_{\tilde{G}}^H\left(\widehat{\boldsymbol{\Gamma}}^{(k)}\right)^{-1}\mathbf{V}_{\tilde{G}}$, $\mathbf{D}^{(k)} = \boldsymbol{\Sigma}_{\tilde{G}}^H\left(\widetilde{\boldsymbol{\Omega}}^{(k)}\right)^H\left(\widetilde{\boldsymbol{\Omega}}^{(k)}\left(\widetilde{\boldsymbol{\Omega}}^{(k)}\right)^H + \boldsymbol{\Lambda}_n\right)^{-1}\widetilde{\boldsymbol{\Omega}}^{(k)}\boldsymbol{\Sigma}_{\tilde{G}}$ and

$$\widetilde{\boldsymbol{\Omega}}^{(k)} = \begin{bmatrix} \mathrm{diag}(\mathbf{p}^{(k)}) & \mathbf{0}_{m\times(q-m)} \\ \multicolumn{2}{c}{\mathbf{0}_{(r-m)\times q}} \end{bmatrix}. \tag{32}$$

The following result from [48], which is applicable for any two Hermitian matrices $\mathbf{X}$ and $\mathbf{Y}$ of size $n \times n$ has been employed for simplifying the MSE expression in (31)

$$\mathrm{Tr}(\mathbf{X} + \mathbf{Y})^{-1} \leq \sum_{i=1}^{n}\frac{1}{\lambda_1(\mathbf{X}) + \lambda_i(\mathbf{Y})}. \tag{33}$$

$$\mathrm{MSE}^{(k)} = \mathrm{Tr}\left(\left(\widehat{\boldsymbol{\Gamma}}^{(k)}\right)^{-1} + \mathbf{V}_{\tilde{G}}\boldsymbol{\Sigma}_{\tilde{G}}^{H}\mathbf{U}_{\tilde{G}}^{H}\left(\widetilde{\mathbf{F}}^{(k)}\right)^{H}\left(\widetilde{\mathbf{F}}^{(k)}\left(\widetilde{\mathbf{F}}^{(k)}\right)^{H} + \mathbf{U}_n\boldsymbol{\Lambda}_n\mathbf{U}_n^{H}\right)^{-1}\widetilde{\mathbf{F}}^{(k)}\mathbf{U}_{\tilde{G}}\boldsymbol{\Sigma}_{\tilde{G}}\mathbf{V}_{\tilde{G}}^{H}\right)^{-1}$$
$$= \mathrm{Tr}\left(\mathbf{V}_{\tilde{G}}^{H}\left(\widehat{\boldsymbol{\Gamma}}^{(k)}\right)^{-1}\mathbf{V}_{\tilde{G}} + \boldsymbol{\Sigma}_{\tilde{G}}^{H}\mathbf{U}_{\tilde{G}}^{H}\left(\widetilde{\mathbf{F}}^{(k)}\right)^{H}\mathbf{U}_n\left(\mathbf{U}_n^{H}\widetilde{\mathbf{F}}^{(k)}\left(\widetilde{\mathbf{F}}^{(k)}\right)^{H}\mathbf{U}_n + \boldsymbol{\Lambda}_n\right)^{-1}\mathbf{U}_n^{H}\widetilde{\mathbf{F}}^{(k)}\mathbf{U}_{\tilde{G}}\boldsymbol{\Sigma}_{\tilde{G}}\right)^{-1}. \tag{30}$$

Upon using the above result in (31), one can determine the following upper bound for the MSE in the kth EM iteration

$$\mathrm{Tr}\left(\boldsymbol{\Upsilon}^{(k)} + \mathbf{D}^{(k)}\right)^{-1} \leq \sum_{j=1}^{m}\left(\widehat{\gamma}_{\max}^{(k)} + \frac{p_j^{(k)}\sigma_j^2\left(\widetilde{\mathbf{G}}\right)}{p_j^{(k)} + \lambda_j(\mathbf{R}_n)}\right)^{-1}$$
$$= \sum_{j=1}^{m}\left(\frac{p_j^{(k)} + \lambda_j(\mathbf{R}_n)}{p_j^{(k)}\left(\widehat{\gamma}_{\max}^{(k)} + \sigma_j^2\left(\widetilde{\mathbf{G}}\right)\right)} + \widehat{\gamma}_{\max}^{(k)}\lambda_j(\mathbf{R}_n)\right)^{-1}, \tag{34}$$

where $\widehat{\gamma}_{\max}^{(k)} = \max_i\{\widehat{\gamma}_i^{(k)}\}$. Using (9), and replacing the prior covariance matrix $\mathbf{R}_\theta$ by $\widehat{\boldsymbol{\Gamma}}^{(k)}$, one can obtain the expression for the average total transmitted power for this scenario is given as

$$\sum_{i=1}^{L}\mathbb{E}\left[||\mathbf{B}_i\mathbf{x}_i||^2\right]$$
$$= \sum_{i=1}^{L}\mathrm{Tr}\left(\mathbf{B}_i^{(k)}\left(\widetilde{\mathbf{G}}_i\widehat{\boldsymbol{\Gamma}}^{(k)}\widetilde{\mathbf{G}}_i^{H} + \mathbf{I}_{q_i}\right)\left(\mathbf{B}_i^{(k)}\right)^{H}\right). \tag{35}$$

Subsequently, using (13) and (14), one can substitute the matrices $\widetilde{\mathbf{G}}_i = \mathbf{U}_{\tilde{G}_i}\boldsymbol{\Sigma}_{\tilde{G}}\mathbf{V}_{\tilde{G}}$ and $\mathbf{B}_i^{(k)} = \mathbf{H}_i^{\dagger}\widetilde{\mathbf{F}}_i^{(k)} = \mathbf{H}_i^{\dagger}\mathbf{U}_n\boldsymbol{\Omega}^{(k)}\mathbf{U}_{\tilde{G}_i}^{H}$, in the above equation to obtain the expression in (36). Furthermore, defining the matrices $\boldsymbol{\Phi}_i^{(k)} = \mathbf{U}_{\tilde{G}_i}^{H}\mathbf{U}_{\tilde{G}_i}\left(\boldsymbol{\Sigma}_{\tilde{G}}\mathbf{V}_{\tilde{G}}^{H}\widehat{\boldsymbol{\Gamma}}^{(k)}\mathbf{V}_{\tilde{G}}\boldsymbol{\Sigma}_{\tilde{G}}^{H} + \mathbf{I}_{q_i}\right)\mathbf{U}_{\tilde{G}_i}^{H}\mathbf{U}_{\tilde{G}_i} \in \mathbb{C}^{q_i\times q_i}$, and $\boldsymbol{\Psi}_i = \mathbf{U}_n^{H}\left(\mathbf{H}_i^{\dagger}\right)^{H}\mathbf{H}_i^{\dagger}\mathbf{U}_n \in \mathbb{C}^{r\times r}$, the total transmit power can be constrained as

$$\sum_{i=1}^{L}\mathrm{Tr}\left(\boldsymbol{\Psi}_i\widetilde{\boldsymbol{\Omega}}^{(k)}\boldsymbol{\Phi}_i^{(k)}\left(\widetilde{\boldsymbol{\Omega}}^{(k)}\right)^{H}\right)$$
$$\leq \sum_{i=1}^{L}\lambda_{\max}(\boldsymbol{\Psi}_i)\sum_{j=1}^{m}p_j^{(k)}\left[\boldsymbol{\Phi}_i^{(k)}\right]_{jj} \leq P_T. \tag{37}$$

The inequality in (37) relies on the following property $\mathrm{Tr}(\mathbf{XY}) \leq \lambda_{\max}(\mathbf{X})\,\mathrm{Tr}(\mathbf{Y})$ [49], which is once again valid for any Hermitian symmetric matrices $\mathbf{X}$ and $\mathbf{Y}$. Therefore, the optimization problem of MSE minimization can now be formulated using the results in (34) and (37) as follows

$$\begin{aligned} \underset{\mathbf{p}^{(k)}}{\text{minimize}} \quad & \sum_{j=1}^{m}\frac{p_j^{(k)} + \lambda_j(\mathbf{R}_n)}{p_j^{(k)}\left(\widehat{\gamma}_{\max}^{(k)} + \sigma_j^2\left(\widetilde{\mathbf{G}}\right)\right) + \widehat{\gamma}_{\max}^{(k)}\lambda_j(\mathbf{R}_n)} \\ \text{subject to} \quad & \sum_{i=1}^{L}\lambda_{\max}(\boldsymbol{\Psi}_i)\sum_{j=1}^{m}p_j^{(k)}\left[\boldsymbol{\Phi}_i^{(k)}\right]_{jj} \leq P_T. \end{aligned} \tag{38}$$

Using the KKT framework [50], the optimal value $p_j^{(k)}$ and the Lagrange multiplier $\mu_0^{(k)}$ for the above problem can be derived as follows :

$$p_j^{(k)} = \frac{\left(\mu_0^{(k)}\sqrt{\frac{\sigma_j^2(\tilde{\mathbf{G}})\lambda_j(\mathbf{R}_n)}{\sum_{i=1}^{L}\lambda_{\max}(\boldsymbol{\Phi}_i)[\boldsymbol{\Psi}_i]_{jj}}} - \widehat{\gamma}_{\max}^{(k)}\lambda_j(\mathbf{R}_n)\right)^{+}}{(\widehat{\gamma}_{\max}^{(k)} + \sigma_j^2(\tilde{\mathbf{G}}))}, \tag{39}$$

$$\mu_0^{(k)} = \frac{P_T + \sum_{j=1}^{m}\sum_{i=1}^{L}\frac{\lambda_{\max}\left(\boldsymbol{\Phi}_i^{(k)}\right)[\boldsymbol{\Psi}_i]_{jj}\lambda_j(\mathbf{R}_n)\widehat{\gamma}_{\max}^{(k)}}{\left(\widehat{\gamma}_{\max}^{(k)}+\sigma_j^2\left(\tilde{\mathbf{G}}\right)\right)}}{\sum_{j=1}^{m}\sqrt{\frac{\sum_{i=1}^{L}\lambda_{\max}\left(\boldsymbol{\Phi}_i^{(k)}\right)[\boldsymbol{\Psi}_i]_{jj}\lambda_j(\mathbf{R}_n)\sigma_j^2\left(\tilde{\mathbf{G}}\right)}{\left(\widehat{\gamma}_{\max}^{(k)}+\sigma_j^2\left(\tilde{\mathbf{G}}\right)\right)^2}}}. \tag{40}$$

Substituting the optimal values $p_j^{(k)}$ into (32) yields the matrix $\widetilde{\boldsymbol{\Omega}}^{(k)}$, from which one can obtain the matrix $\widetilde{\mathbf{F}}^{(k)}$. Subsequently, the TPC matrix $\mathbf{B}_i^{(k)}$ can be determined using the relationship

$$\mathbf{B}_i^{(k)} = \mathbf{H}_i^{\dagger}\widetilde{\mathbf{F}}_i^{(k)}, \tag{41}$$

where $\mathbf{H}_i^{\dagger}$ represents the pseudo-inverse of the matrix $\mathbf{H}_i$. Note that in a typical WSN, the condition $r >> t_i$ holds true, since the FC can afford a large number of antennas, whereas the SNs are miniature devices. Assuming that the elements of the channel matrix $\mathbf{H}_i$ are i.i.d., it is guaranteed to be of full column rank with a high probability. The BL-based sparse estimate of the parameter vector $\boldsymbol{\theta}^{(k)}$ is obtained as $\widehat{\boldsymbol{\theta}}^{(k)} = \boldsymbol{\mu}_\theta^{(k)}$. A concise summary of the various steps in the BL-based technique proposed for sparse parameter estimation and optimal MMSE TPC design is given in Algorithm-1, with its graphical description is shown in Fig.2. Interestingly, the MSE of estimation for this scenario with noisy sensor observations exhibits a floor at higher values of the SNR at the fusion center, which arises due to amplification of the observation noise. An analytical expression has been derived in Section-III of our technical report [51] for this error floor. The next subsection presents an interesting analysis for the high-SNR observations scenario.

A. BL-based Sparse Parameter Estimation and Optimal MMSE TPC Design for High-SNR Sensor Observations

In the high-SNR scenario [52], [53], where the observation noise vector obeys $\mathbf{v} = \mathbf{0}$ in (7), the received signal vector $\mathbf{y}$ at the FC is given by

$$\mathbf{y} = \mathbf{HBG}\boldsymbol{\theta} + \mathbf{n}. \tag{42}$$

$$
\begin{aligned}
\mathbb{E}\left[||\mathbf{B}_i\mathbf{x}_i||^2\right] &= \sum_{i=1}^{L} \mathrm{Tr}\left(\mathbf{H}_i^{\dagger}\mathbf{U}_n\widetilde{\mathbf{\Omega}}^{(k)}\mathbf{U}_{\tilde{G}_i}^H\left(\mathbf{U}_{\tilde{G}_i}\mathbf{\Sigma}_{\tilde{G}}\mathbf{V}_{\tilde{G}}^H\widehat{\mathbf{\Gamma}}^{(k)}\mathbf{V}_{\tilde{G}}\mathbf{\Sigma}_{\tilde{G}}^H\mathbf{U}_{\tilde{G}_i}^H + \mathbf{I}_{q_i}\right)\mathbf{U}_{\tilde{G}_i}\left(\widetilde{\mathbf{\Omega}}^{(k)}\right)^H\mathbf{U}_n^H\left(\mathbf{H}_i^{\dagger}\right)^H\right) \\
&= \sum_{i=1}^{L}\mathrm{Tr}\left(\mathbf{U}_n^H\left(\mathbf{H}_i^{\dagger}\right)^H\mathbf{H}_i^{\dagger}\mathbf{U}_n\widetilde{\mathbf{\Omega}}^{(k)}\mathbf{U}_{\tilde{G}_i}^H\mathbf{U}_{\tilde{G}_i}\left(\mathbf{\Sigma}_{\tilde{G}}\mathbf{V}_{\tilde{G}}^H\widehat{\mathbf{\Gamma}}^{(k)}\mathbf{V}_{\tilde{G}}\mathbf{\Sigma}_{\tilde{G}}^H + \mathbf{I}_{q_i}\right)\mathbf{U}_{\tilde{G}_i}^H\mathbf{U}_{\tilde{G}_i}\left(\widetilde{\mathbf{\Omega}}^{(k)}\right)^H\right).
\end{aligned} \tag{36}
$$

Algorithm 1 BL-based sparse parameter estimation and optimal MMSE TPC design for the noisy sensor observations scenario described in Section-III

1: **Input**: Observation vector $\mathbf{y}^{(k)}$, noise covariance matrix $\widetilde{\mathbf{R}}_n$, observation noise whitening matrix $\mathbf{W}$, stacked observation matrix $\mathbf{G}$, concatenated channel matrix $\mathbf{H}$, stopping parameters ϵ and k_{max}
2: **Initialization**: $\widehat{\gamma}_i^{(0)} = 1,\ 1 \leq i \leq m \Rightarrow \widehat{\mathbf{\Gamma}}^{(0)} = \mathbf{I}$, set counter $k = 0$ and $\widehat{\mathbf{\Gamma}}^{(-1)} = \mathbf{0}$
3: **while** $\parallel \widehat{\boldsymbol{\gamma}}^{(k)} - \widehat{\boldsymbol{\gamma}}^{(k-1)} \parallel_2 \geq \epsilon$ and $k < k_{max}$ **do**
4: Compute $p_j^{(k)}$ using (39) and $\mathbf{\Omega}^{(k)}$ using (32)
5: Compute matrix $\widetilde{\mathbf{F}}^{(k)} = \mathbf{U}_n\widetilde{\mathbf{\Omega}}^{(k)}\mathbf{U}_{\tilde{G}}^H$
6: Compute TPC $\mathbf{B}_i^{(k)}$ for each sensor i using (41)
7: Obtain the receive vector $\mathbf{y}^{(k)}$ using (20)
8: **E-Step**: Evaluate the *a posteriori* mean $\mathbf{\Sigma}_\theta^{(k)}$ and covariance $\boldsymbol{\mu}_\theta^{(k)}$ using (24) and (25)
9: **M-Step**: Evaluate the hyperparameters $\widehat{\gamma}_i^{(k)}$ using (28)
10: **end while**
11: **Output**: $\widehat{\boldsymbol{\theta}}^{(k)} = \boldsymbol{\mu}_\theta^{(k)}$

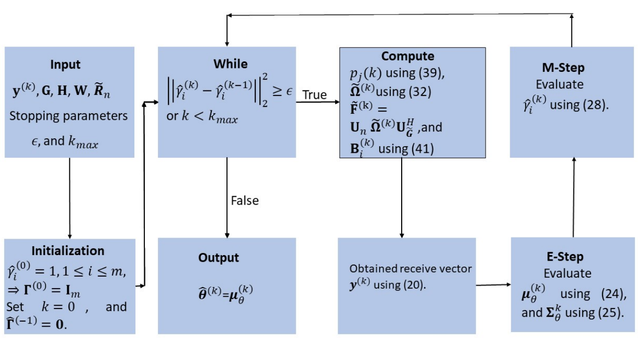


Fig. 2. Graphical illustration of the Algorithm-1.

The BL technique proposed for this high-SNR observation scenario once again assigns the parameterized Gaussian prior $p(\boldsymbol{\theta};\mathbf{\Gamma})$ to the sparse parameter vector $\boldsymbol{\theta}$ as given in (15). The *a posteriori* probability density function of the sparse parameter vector $\boldsymbol{\theta}$ can be evaluated as [43]

$$
p(\boldsymbol{\theta} \mid \mathbf{y};\mathbf{\Gamma}) \sim \mathcal{CN}\left(\boldsymbol{\mu}_\theta, \mathbf{\Sigma}_\theta\right),
$$

where the *a posteriori* mean and covariance matrix, $\boldsymbol{\mu}_\theta \in \mathbb{C}^{m\times 1}$ and $\mathbf{\Sigma}_\theta \in \mathbb{C}^{m\times m}$, respectively, can be determined as

$$
\widehat{\boldsymbol{\theta}} = \mathbf{\Sigma}_\theta\mathbf{G}^H\mathbf{B}^H\mathbf{H}^H\mathbf{R}_n^{-1}\mathbf{y}, \tag{43}
$$

$$
\mathbf{\Sigma}_\theta = \left(\mathbf{\Gamma}^{-1} + \mathbf{G}^H\mathbf{B}^H\mathbf{H}^H\mathbf{R}_n^{-1}\mathbf{H}\mathbf{B}\mathbf{G}\right)^{-1}. \tag{44}
$$

The Bayesian evidence $p(\mathbf{y};\mathbf{\Gamma})$ can once again be maximized using the EM procedure stated in Section-III. The hyperparameter update in the kth EM iteration can be obtained as

$$
\widehat{\gamma}_i^{(k)} = \mathbf{\Sigma}_\theta^{(k)}(i,i) + \mid \boldsymbol{\mu}_\theta^{(k)}(i) \mid^2, \tag{45}
$$

where the posterior mean $\boldsymbol{\mu}_\theta^{(k)}$ and covariance $\mathbf{\Sigma}_\theta^{(k)}$ can be obtained by replacing $\mathbf{y} \rightarrow \mathbf{y}^{(k)}$, $\mathbf{B} \rightarrow \mathbf{B}^{(k-1)}$, and $\mathbf{\Gamma} \rightarrow \widehat{\mathbf{\Gamma}}^{(k-1)}$ in (43) and (44). The BL-based estimate of the hyperparameter matrix $\mathbf{\Gamma}^{(k)}$ is subsequently employed for TPC design in the kth EM iteration. The estimation error covariance matrix $\mathbf{E}^{(k)}$ can be obtained as [43, Sec. 12.3]

$$
\mathbf{E}^{(k)} = \left(\left(\widehat{\mathbf{\Gamma}}^{(k)}\right)^{-1} + \mathbf{G}^H\left(\mathbf{B}^{(k)}\right)^H\mathbf{H}^H\mathbf{R}_n^{-1}\mathbf{H}\mathbf{B}^{(k)}\mathbf{G}\right)^{-1}. \tag{46}
$$

Let the matrix $\widetilde{\mathbf{H}} = \mathbf{H}^H\mathbf{R}_n^{-1}\mathbf{H} \in \mathbb{C}^{t\times t}$ with its eigenvalue decomposition (EVD) given as $\widetilde{\mathbf{H}} = \mathbf{U}_t\mathbf{\Lambda}_t\mathbf{U}_t^H$ and $\mathbf{F}^{(k)} = \mathbf{B}^{(k)}\mathbf{G} \in \mathbb{C}^{t\times m}$. It follows from (5) and (6) that

$$
\begin{aligned}
\mathbf{F}^{(k)} &= \left[\mathbf{F}_1^{(k)}, \mathbf{F}_2^{(k)}, \cdots, \mathbf{F}_L^{(k)}\right] \\
&= \left[\mathbf{B}_1^{(k)}\mathbf{G}_1, \mathbf{B}_2^{(k)}\mathbf{G}_2, \cdots, \mathbf{B}_L^{(k)}\mathbf{G}_L\right].
\end{aligned} \tag{47}
$$

The expression for the error covariance matrix $\mathbf{E}^{(k)}$ can be further simplified as

$$
\mathbf{E}^{(k)} = \left(\left(\widehat{\mathbf{\Gamma}}^{(k)}\right)^{-1} + \left(\mathbf{F}^{(k)}\right)^H\mathbf{U}_t\mathbf{\Lambda}_t{\mathbf{U}_t}^H\mathbf{F}^{(k)}\right)^{-1}. \tag{48}
$$

Choose the TPC matrix $\mathbf{F}^{(k)} = \mathbf{U}_t\mathbf{\Omega}^{(k)}$, where the matrix $\mathbf{\Omega}^{(k)}$ is defined as

$$
\mathbf{\Omega}^{(k)} = \begin{bmatrix}\mathrm{diag}\left(\mathbf{p}^{(k)}\right) \\ \mathbf{0}_{(t-m)\times m}\end{bmatrix} \in \mathbb{R}^{t\times m},\ t \geq m \tag{49}
$$

and $\mathbf{p}^{(k)} = \left[\sqrt{p_1^{(k)}}, \sqrt{p_2^{(k)}}, \cdots, \sqrt{p_m^{(k)}}\right]^T \in \mathbb{R}^{m\times 1}$ with $p_i^{(k)} \geq 0\ \forall i$. An interesting analysis for the scenario when $m > t$ is given in our technical report [51]. Employing the above substitution in (48), the expression of the error covariance matrix $\mathbf{E}^{(k)}$ and that of the MSE cost function $\mathrm{Tr}\left(\mathbf{E}^{(k)}\right)$ can be simplified as

$$
\mathbf{E}^{(k)} = \left(\left(\widehat{\mathbf{\Gamma}}^{(k)}\right)^{-1} + \left(\mathbf{\Omega}^{(k)}\right)^H\mathbf{\Lambda}_t\mathbf{\Omega}^{(k)}\right)^{-1}, \tag{50}
$$

$$
\mathrm{MSE}^{(k)} = \mathrm{Tr}\left(\mathbf{E}^{(k)}\right) = \sum_{i=1}^{m}\left(\frac{1}{\widehat{\gamma}_i^{(k)}} + p_i^{(k)}\lambda_i\left(\widetilde{\mathbf{H}}\right)\right)^{-1}. \tag{51}
$$

The total transmit power of the WSN, formulated in (9) for the high-SNR scenario, can be further simplified as

$$\begin{aligned}\mathrm{Tr}\left(\mathbf{F}^{(k)}\widehat{\boldsymbol{\Gamma}}^{(k)}\left(\mathbf{F}^{(k)}\right)^{H}\right)&=\mathrm{Tr}\left(\boldsymbol{\Omega}^{(k)}\widehat{\boldsymbol{\Gamma}}^{(k)}\left(\boldsymbol{\Omega}^{(k)}\right)^{H}\right)\\&=\sum_{i=1}^{m}p_i^{(k)}\widehat{\gamma}_i^{(k)}.\end{aligned} \tag{52}$$

Hence, the optimization problem of minimizing the estimation MSE of the parameter vector $\boldsymbol{\theta}$, in the coherent MAC-based WSN can be equivalently formulated as

$$\begin{aligned}&\underset{\mathbf{p}^{(k)}}{\text{minimize}} && \sum_{i=1}^{m}\left(\frac{1}{\widehat{\gamma}_i^{(k)}}+p_i^{(k)}\lambda_i\left(\widetilde{\mathbf{H}}\right)\right)^{-1}\\&\text{subject to} && \sum_{i=1}^{m}p_i^{(k)}\widehat{\gamma}_i^{(k)}\leq P_T.\end{aligned} \tag{53}$$

The solution to the above optimization problem can be found using the Karush-Kuhn-Tucker (KKT) framework [50]. The optimal values $p_i^{(k)},\ 1\leq i\leq m$, and the pertinent Lagrange multiplier μ_0 are obtained as

$$p_i^{(k)}=\left[\mu_0^{(k)}\sqrt{\frac{1}{\widehat{\gamma}_i^{(k)}\lambda_i\left(\widetilde{\mathbf{H}}\right)}}-\frac{1}{\widehat{\gamma}_i^{(k)}\lambda_i\left(\widetilde{\mathbf{H}}\right)}\right]^{+}, \tag{54}$$

$$\mu_0^{(k)}=\frac{P_T+\sum_{i=1}^{m}\frac{1}{\lambda_i(\widetilde{\mathbf{H}})}}{\sum_{i=1}^{m}\sqrt{\frac{\widehat{\gamma}_i^{(k)}}{\lambda_i(\widetilde{\mathbf{H}})}}}. \tag{55}$$

Upon substituting the optimal values $p_i^{(k)}$ into (49) we arrive at the matrix $\boldsymbol{\Omega}^{(k)}$, from which one can obtain the matrix $\mathbf{F}^{(k)}$. Subsequently, the TPC matrix $\mathbf{B}_i^{(k)}$ can be determined using the relationship

$$\mathbf{B}_i^{(k)}=\mathbf{F}_i^{(k)}\mathbf{G}_i^{\dagger}, \tag{56}$$

where $\mathbf{G}_i^{\dagger}$ represents the pseudo-inverse of the matrix $\mathbf{G}_i$. Note that as described in several related contributions [2], [36], it is a common practice to compress the SN measurement $\mathbf{x}_i$ prior to transmission. Moreover, when the parameter vector $\boldsymbol{\theta}$ is sparse, as described in the theory of compressive sensing (CS) [44], the measurement $\mathbf{x}_i$ can be significantly compressed, typically, to a much lower dimension than the size of the parameter $\boldsymbol{\theta}$. Thus, in order to achieve this, the observation matrix $\mathbf{G}_i$ for $1\leq i\leq L$ are set with $q_i\leq m$. Furthermore, assuming the elements of the observation matrix $\mathbf{G}_i$ to be i.i.d., it is guaranteed to be of full row rank with high probability. The BL-based sparse estimate of the parameter vector $\boldsymbol{\theta}^{(k)}$ is obtained as $\widehat{\boldsymbol{\theta}}^{(k)}=\boldsymbol{\mu}_{\theta}^{(k)}$. An interesting theoretical result in Section-IV of our technical report [51] shows that the $\mathrm{MSE}^{(k)}$ in (51) is a monotonically decreasing function of $\mathrm{SNR}_{\mathrm{FC}}$. The next section describes the MMSE channel estimation procedure, followed by the robust transceiver design conceived for sparse parameter estimation in the presence of CSI uncertainty.

IV. Channel Estimation and Robust Transceiver Design in MIMO WSNs

A majority of the schemes in the existing literature, such as [3], [7]–[9] have relied on the idealized simplifying assumption that the FC has perfect knowledge of the underlying MIMO channel between the SNs and the FC. With an eye toward practical implementation, in this section we conceive a comprehensive procedure for decentralized parameter estimation using the imperfect CSI obtained via the estimation module. The various key steps of this section are as follows.

1) We begin with estimating the underlying channels between each sensor and the FC using the proposed MMSE channel estimation procedure.
2) Furthermore, employing the imperfect CSI and its error covariance obtained via the estimation module together with the parametrized prior assignment over the sparse parameter vector $\boldsymbol{\theta}$, we develop an EM-based framework for estimating the hyperparameters.
3) Furthermore, the robust transceiver design is proposed that incorporates knowledge of the channel uncertainty and the estimated hyperparameters for minimizing the average MSE.

We now describe the above steps in detail.

A. MMSE CSI Estimation in MIMO WSNs

A flat fading MIMO channel is considered between the FC and each SN in the time-division-duplexing (TDD) mode, wherein the MIMO channel between the FC and the ith SN is denoted by $\mathbf{H}_i^T\in\mathbb{C}^{t_i\times r}$. For the purpose of channel estimation, the FC broadcasts pilot symbols to all the SNs, and the pilot observations of each SN are fed back to the FC, which then estimates the MIMO channel for each SN. Let $\mathbf{x}_p(n)\in\mathbb{C}^{r\times 1},\ 1\leq n\leq N$, denote the pilot vectors transmitted by the FC, where N represents the number of pilot vectors. The pilot vector $\mathbf{y}_i(n)\in\mathbb{C}^{t_i\times 1}$ received by the ith SN is given as

$$\mathbf{y}_i(n)=\mathbf{H}_i^T\mathbf{x}_p(n)+\mathbf{w}_i(n), \tag{57}$$

where $\mathbf{w}_i(n)\in\mathbb{C}^{t_i\times 1}$ denotes the AWGN vector distributed as $\mathcal{CN}\left(\mathbf{0},\mathbf{R}_{w_i}\right)$. The received pilot matrix $\mathbf{Y}_i\in\mathbb{C}^{t_i\times N}$ corresponding to the ith SN, after concatenation of the received pilot vectors $\mathbf{y}_i(n),\ 1\leq n\leq N$, can be modeled as

$$\mathbf{Y}_i=[\mathbf{y}_i(1),\mathbf{y}_i(2),\cdots,\mathbf{y}_i(N)]=\mathbf{H}_i^T\mathbf{X}_p+\mathbf{W}_i, \tag{58}$$

where $\mathbf{X}_p=[\mathbf{x}_p(1),\mathbf{x}_p(2),\cdots,\mathbf{x}_p(N)]\in\mathbb{C}^{r\times N}$ and $\mathbf{W}_i=[\mathbf{w}_i(1),\mathbf{w}_i(2),\cdots,\mathbf{w}_i(N)]\in\mathbb{C}^{t_i\times N}$ denote the concatenated pilot and noise matrices. For the sake of low complexity, an orthogonal pilot matrix satisfying $\mathbf{X}_p\mathbf{X}_p^H=P_p\mathbf{I}$ is chosen, where P_p represents the pilot power. The vectorized channel estimation model corresponding to (58) is obtained by considering the stacked columns of the observation $\mathbf{Y}_i$ as

$$\mathbf{y}_i=\widetilde{\mathbf{X}}_p\mathbf{h}_i+\mathbf{w}_i, \tag{59}$$

where $\mathbf{y}_i=\mathrm{vec}(\mathbf{Y}_i)\in\mathbb{C}^{Nt_i\times 1}$, $\widetilde{\mathbf{X}}_p=(\mathbf{X}_p\otimes\mathbf{I})\in\mathbb{C}^{Nt_i\times rt_i}$, $\mathbf{h}_i=\mathrm{vec}\left(\mathbf{H}_i^T\right)\in\mathbb{C}^{rt_i\times 1}$ and $\mathbf{w}_i=\mathrm{vec}(\mathbf{W}_i)\in\mathbb{C}^{Nt_i\times 1}$. The

$$\begin{aligned}\mathbb{E}\left[\mathbf{E}\right] = &\mathbf{M}^H\Big(\widehat{\mathbf{H}}\mathbf{B}\mathbf{W}\mathbf{G}\mathbf{R}_\theta\mathbf{G}^H\mathbf{W}^H\mathbf{B}^H\widehat{\mathbf{H}}^H + \operatorname{Tr}\left(\mathbf{B}\mathbf{W}\mathbf{G}\mathbf{R}_\theta\mathbf{G}^H\mathbf{W}^H\mathbf{B}^H\right)\mathbf{R}^{Rx}\Big)\mathbf{M} + \mathbf{M}^H\mathbf{R}_n\mathbf{M} + \mathbf{R}_\theta \\ &+ \mathbf{M}^H\Big(\widehat{\mathbf{H}}\mathbf{B}\mathbf{W}\mathbf{R}_v\mathbf{W}^H\mathbf{B}^H\widehat{\mathbf{H}}^H + \operatorname{Tr}\left(\mathbf{B}\mathbf{B}^H\right)\mathbf{R}^{Rx}\Big)\mathbf{M} - \mathbf{M}^H\widehat{\mathbf{H}}\mathbf{B}\mathbf{W}\mathbf{G}\mathbf{R}_\theta - \mathbf{R}_\theta\mathbf{G}^H\mathbf{W}^H\mathbf{B}^H\widehat{\mathbf{H}}^H\mathbf{M} \\ &= \mathbf{M}^H\widehat{\mathbf{H}}\mathbf{B}\mathbf{W}\mathbf{G}\mathbf{R}_\theta\mathbf{G}^H\mathbf{W}^H\mathbf{B}^H\widehat{\mathbf{H}}^H\mathbf{M} - \mathbf{R}_\theta\mathbf{G}^H\mathbf{W}^H\mathbf{B}^H\widehat{\mathbf{H}}^H\mathbf{M} + \mathbf{M}^H\widehat{\mathbf{H}}\mathbf{B}\mathbf{B}^H\widehat{\mathbf{H}}^H\mathbf{M} + \mathbf{R}_\theta \\ &\quad - \mathbf{M}^H\widehat{\mathbf{H}}\mathbf{B}\mathbf{W}\mathbf{G}\mathbf{R}_\theta + \mathbf{M}^H\breve{\mathbf{R}}_n\mathbf{M}. \end{aligned} \tag{67}$$

LMMSE estimate $\widehat{\mathbf{h}}_i$ of the vectorized channel $\mathbf{h}_i$ is obtained as

$$\widehat{\mathbf{h}}_i = \mathbf{R}_i\widetilde{\mathbf{X}}_p^H\left(\widetilde{\mathbf{X}}_p\mathbf{R}_i\widetilde{\mathbf{X}}_p^H + \mathbf{R}_{w_i}\right)^{-1}\mathbf{y}_i, \tag{60}$$

where $\mathbf{R}_i \in \mathbb{C}^{rt_i \times rt_i}$ denotes the covariance matrix of the ith channel $\mathbf{h}_i$.

B. Robust Transceiver Design Subjected to CSI Uncertainty

The error covariance matrix $\mathbf{E}_{h_i}$ corresponding to the LMMSE estimate $\widehat{\mathbf{h}}_i$ can be determined as

$$\mathbf{E}_{h_i} = \mathbb{E}\left\{\left(\mathbf{h}_i - \widehat{\mathbf{h}}_i\right)\left(\mathbf{h}_i - \widehat{\mathbf{h}}_i\right)^H\right\} = \left(\mathbf{R}_i^{-1} + \widetilde{\mathbf{X}}_p^H\mathbf{R}_{w_i}^{-1}\widetilde{\mathbf{X}}_p\right)^{-1}. \tag{61}$$

Using (4) and setting $\mathbf{R}_{w_i} = \sigma_{w_i}^2\mathbf{I}$, where $\sigma_{w_i}^2$ denotes the variance of the noise for the ith SN, the error covariance matrix $\mathbf{E}_{h_i}$ above reduces to

$$\mathbf{E}_{h_i} = \left(\left(\mathbf{R}^{Rx}\right)^{-1} \otimes \left(\mathbf{R}^{Tx,i}\right)^{-1} + \frac{P_p}{\sigma_{w_i}^2}\mathbf{I}_{rt_i}\right)^{-1}. \tag{62}$$

Furthermore, setting $\mathbf{R}^{Tx,i} = \mathbf{I}_{t_i}$, the above expression simplifies to

$$\mathbf{E}_{h_i} = \left(\left(\mathbf{R}^{Rx}\right)^{-1} \otimes \mathbf{I}_{t_i} + \frac{P_p}{\sigma_{w_i}^2}\mathbf{I}_{rt_i}\right)^{-1} = \widetilde{\mathbf{R}}^{Rx} \otimes \mathbf{I}_{t_i}, \tag{63}$$

where $\widetilde{\mathbf{R}}_{Rx} = \left(\left(\mathbf{R}^{Rx}\right)^{-1} + \frac{P_p}{\sigma_{w_i}^2}\mathbf{I}_r\right)^{-1} \in \mathbb{C}^{r\times r}$ denotes the equivalent spatial correlation matrix at the receiver. The relationship between the true and the estimated/ known channel between the FC and the ith SN is given as $\mathbf{H}_i = \widehat{\mathbf{H}}_i + \widetilde{\mathbf{R}}_{Rx}^{1/2}\mathbf{S}_i$, where $\widehat{\mathbf{H}}_i^T$ is obtained by the inverse vectorization operation applied to $\widehat{\mathbf{h}}_i$ found in (60) and the elements of $\mathbf{S}_i$ are assumed to be i.i.d. $\mathcal{CN}(0,1)$. Hence, the resultant distribution of $\mathbf{H}_i$ is $\mathbf{H}_i \sim \mathcal{CN}\left(\widehat{\mathbf{H}}_i, \left(\mathbf{I}_{t_i} \otimes \widetilde{\mathbf{R}}_{Rx}\right)\right)$. Similar to the previous sections, the concatenated channel matrix $\mathbf{H}$ can be derived as

$$\mathbf{H} = \widehat{\mathbf{H}} + \left(\widetilde{\mathbf{R}}_{Rx}\right)^{1/2}\mathbf{S}, \tag{64}$$

where the above quantities are defined as $\widehat{\mathbf{H}} = \left[\widehat{\mathbf{H}}_1, \widehat{\mathbf{H}}_2, \cdots, \widehat{\mathbf{H}}_L\right] \in \mathbb{C}^{r\times t}, \mathbf{S} = \left[\,\mathbf{S}_1, \mathbf{S}_2, \cdots, \mathbf{S}_L\,\right] \in \mathbb{C}^{r\times t}$. Thus, the distribution of $\mathbf{H}$ with CSI uncertainty is given as

$$\mathbf{H} \sim \mathcal{CN}\left(\widehat{\mathbf{H}}, \left(\mathbf{I}_t \otimes \widetilde{\mathbf{R}}_{Rx}\right)\right). \tag{65}$$

The robust transceiver design for the general scenario with arbitrary SNR sensor observations is described next, while that for the special case of high-SNR sensor observations is given in Section-V of our technical report [51]. Starting with the expression for the received vector $\mathbf{y}$ in (12), the estimate $\widehat{\boldsymbol{\theta}}$ of the unknown sparse parameter $\boldsymbol{\theta}$ can be obtained as $\widehat{\boldsymbol{\theta}} = \mathbf{M}^H\mathbf{y}$. The corresponding error covariance matrix is given as

$$\begin{aligned}\mathbf{E} = &\left(\mathbf{M}^H\mathbf{H}\mathbf{B}\mathbf{W}\mathbf{G} - \mathbf{I}\right)\mathbf{R}_\theta\left(\mathbf{G}^H\mathbf{W}^H\mathbf{B}^H\mathbf{H}^H\mathbf{M} - \mathbf{I}\right) + \\ &\mathbf{M}^H\left(\mathbf{H}\mathbf{B}\mathbf{W}\mathbf{R}_v\mathbf{W}^H\mathbf{B}^H\mathbf{H}^H + \mathbf{R}_n\right)\mathbf{M}.\end{aligned} \tag{66}$$

Using the uncertainty model in (65), substituting the channel $\mathbf{H}$ into (66), and employing the result from Lemma A given in Appendix-A, the expression for the resulting average error covariance matrix reduces to (67), which can further be written in a compact form as

$$\begin{aligned}\mathbb{E}\left[\mathbf{E}\right] = &\left(\mathbf{M}^H\widehat{\mathbf{H}}\mathbf{B}\mathbf{W}\mathbf{G} - \mathbf{I}\right)\mathbf{R}_\theta\left(\mathbf{G}^H\mathbf{W}^H\mathbf{B}^H\widehat{\mathbf{H}}^H\mathbf{M} - \mathbf{I}\right) \\ &+ \mathbf{M}^H\left(\widehat{\mathbf{H}}\mathbf{B}\mathbf{B}^H\widehat{\mathbf{H}}^H + \breve{\mathbf{R}}_n + \mathbf{I}\right)\mathbf{M}.\end{aligned} \tag{68}$$

Algorithm 2 MMSE CSI acquisition and BL-based robust sparse parameter estimation for the scenario described in Section-IV

Input: Observation vector $\mathbf{y}^{(k)}$, equivalent noise covariance matrix $\breve{\mathbf{R}}_n$, stacked observation matrix $\mathbf{G}$, pilot vector $\mathbf{y}_i$ for each sensor $1 \le i \le L$, stopping parameters ϵ and k_{max}
Initialization: $\widehat{\gamma}_i^{(0)} = 1,\ 1 \le i \le m \Rightarrow \widehat{\boldsymbol{\Gamma}}^{(0)} = \mathbf{I}$, set counter $k = 0$ and $\widehat{\boldsymbol{\Gamma}}^{(-1)} = \mathbf{0}$
Using with $\mathbf{X}_p\mathbf{X}_p^H = P_p\mathbf{I}$
Compute $\widehat{\mathbf{h}}_i$ using (60)
while $\parallel \widehat{\boldsymbol{\gamma}}^{(k)} - \widehat{\boldsymbol{\gamma}}^{(k-1)} \parallel_2 \ge \epsilon$ and $k < k_{max}$ **do**
  Compute $\widetilde{\boldsymbol{\Omega}}^{(k)}$ using (32)
  Compute the matrix $\widetilde{\mathbf{F}}^{(k)} = \mathbf{U}_n\widetilde{\boldsymbol{\Omega}}^{(k)}\mathbf{U}_{\tilde{G}}^H$
  Compute the TPC $\mathbf{B}_i^{(k)}$ for each sensor i using (41)
  Equivalent Sensing Matrix Calculation: $\widehat{\boldsymbol{\Phi}}^{(k)} = \widehat{\mathbf{H}}\mathbf{B}^{(k)}\mathbf{G}$
  Obtain the receive vector $\mathbf{y}^{(k)}$ using (20)
  E-Step: Evaluate the a *posteriori* covariance and mean

$$\bar{\boldsymbol{\Sigma}}_{\boldsymbol{\theta}}^{(k)} = \left(\widehat{\boldsymbol{\Phi}}^H\left(\breve{\mathbf{R}}_n\right)^{-1}\widehat{\boldsymbol{\Phi}} + \left(\widehat{\boldsymbol{\Gamma}}^{(k)}\right)^{-1}\right)^{-1} \in \mathbb{C}^{m\times m}$$

$$\bar{\boldsymbol{\mu}}_{\boldsymbol{\theta}}^{(k)} = \bar{\boldsymbol{\Sigma}}_{\boldsymbol{\theta}}^{(k)}\widehat{\boldsymbol{\Phi}}^H\left(\breve{\mathbf{R}}_n'\right)^{-1}\mathbf{y}^{(k)} \in \mathbb{C}^{m\times 1}$$

  M-Step: Evaluate the hyperparameter estimates
  for $j = 1 \to m$ **do**
    $\widehat{\gamma}_j^{(k+1)} = \bar{\boldsymbol{\Sigma}}_{\boldsymbol{\theta}}^{(k)}(j,j) + \mid \bar{\boldsymbol{\mu}}_{\boldsymbol{\theta}}^{(k)}(j) \mid^2$
  end for
end while
Output: $\widehat{\boldsymbol{\theta}}^{(k)} = \boldsymbol{\mu}_{\theta}^{(k)}$

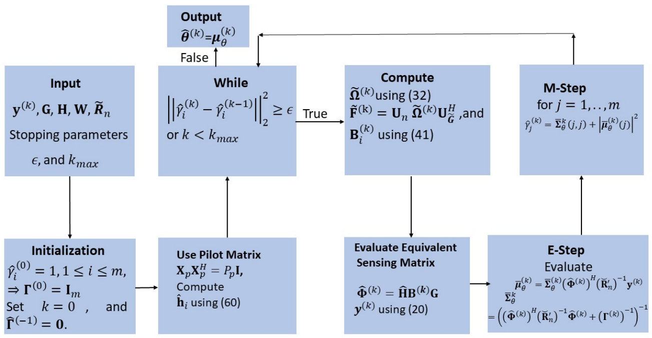


Fig. 3. Graphical illustration of the Algorithm-2.

The matrix $\breve{\mathbf{R}}_n$ represents the equivalent noise covariance matrix, the expression for which can be expressed as

$$\begin{aligned}\breve{\mathbf{R}}_n &= \mathbf{R}_n + \mathrm{Tr}\left(\mathbf{BWGR}_\theta\mathbf{G}^H\mathbf{W}^H\mathbf{B}^H + \mathbf{BB}^H\right)\widetilde{\mathbf{R}}_{Rx} \\ &= \mathbf{R}_n + P_T\widetilde{\mathbf{R}}_{Rx},\end{aligned} \tag{69}$$

where the second equality follows from the fact that for a power-constrained system the quantity $\mathrm{Tr}\left(\mathbf{BWGR}_\theta\mathbf{G}^H\mathbf{W}^H\mathbf{B}^H + \mathbf{BB}^H\right)$ equals P_T. The optimal combiner matrix $\mathbf{M}$ is obtained via differentiating the average error covariance matrix with respect to $\mathbf{M}$, followed by setting the derivative equal to zero. The optimal combiner matrix $\mathbf{M}$ thus determined and the ensuing MSE are given as

$$\begin{aligned}\mathbf{M} =& \left(\widehat{\mathbf{H}}\mathbf{BWGR}_\theta\mathbf{W}^H\mathbf{G}^H\mathbf{B}^H\widehat{\mathbf{H}}^H + \widehat{\mathbf{H}}\mathbf{BB}^H\widehat{\mathbf{H}}^H + \breve{\mathbf{R}}_n\right)^{-1} \\ &\widehat{\mathbf{H}}\mathbf{BWGR}_\theta, \\ \mathrm{MSE} =& \mathrm{Tr}\Bigg(\bigg(\mathbf{R}_\theta^{-1} + \mathbf{G}^H\mathbf{W}^H\mathbf{B}^H\widehat{\mathbf{H}}^H\left(\widehat{\mathbf{H}}\mathbf{BB}^H\widehat{\mathbf{H}}^H + \breve{\mathbf{R}}_n\right)^{-1} \\ &\widehat{\mathbf{H}}\mathbf{BWG}\bigg)^{-1}\Bigg).\end{aligned} \tag{70}$$

The optimization problem minimizing the average MSE given in (70), subject to the total power constraint is given as

$$\begin{aligned}&\underset{\mathbf{B}}{\text{minimize}}\ \mathrm{MSE} \\ &\text{subject to}\ \ \mathrm{Tr}\left(\mathbf{B}\left(\mathbf{WGR}_\theta\mathbf{G}^H\mathbf{W}^H + \mathbf{I}_q\right)\mathbf{B}^H\right) \leq P_T.\end{aligned} \tag{71}$$

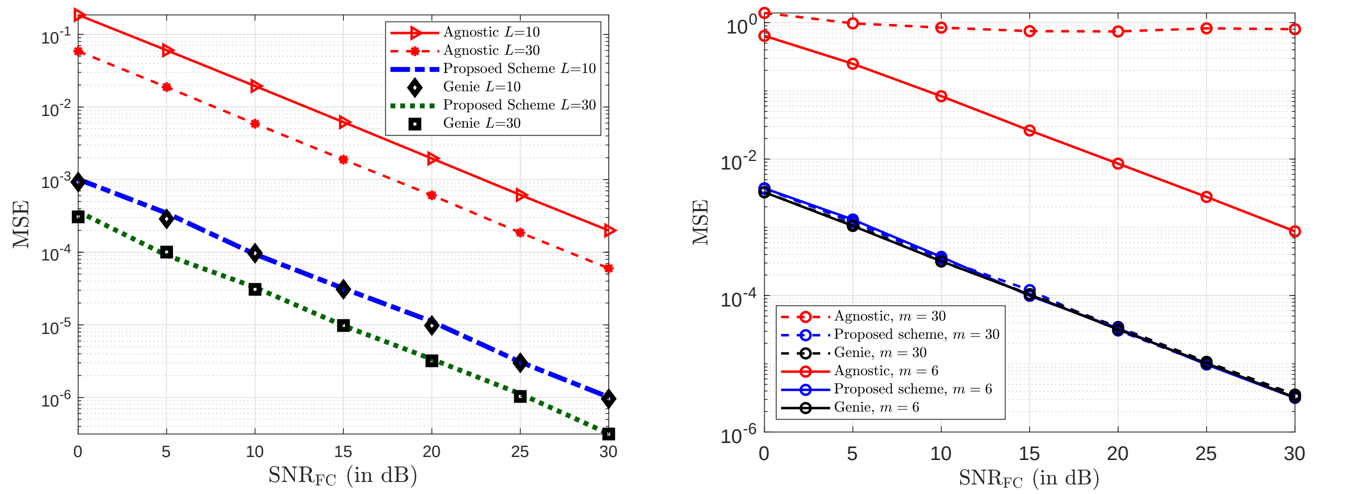
The above optimization problem can be solved using the procedure described in Section-III after replacing $\mathbf{H}$ with $\widehat{\mathbf{H}}$ and $\mathbf{R}_n$ with $\breve{\mathbf{R}}_n$. The step-by-step procedure of our robust TPC design and sparse parameter vector estimation is formulated in Algorithm-2 with its graphical description is shown in Fig.3. Considering a scenario where the perfect knowledge of the noise covariance matrix $\mathbf{R}_n$ is also unavailable, following the procedure described in our technical report [51], one can model the uncertainty in the matrix $\mathbf{R}_n$ as $\mathbf{R}_n = \widehat{\mathbf{R}}_n - \sigma^2\mathbf{I}$, where $\widehat{\mathbf{R}}_n$ denotes the estimate of the noise covariance matrix and σ^2 represents the variance of the measurement error of the noise.

V. Simulation Results

This section presents the results of a Monte Carlo simulation study to illustrate the performance of various algorithms and also to verify our proposed analytical formulations. The spatial correlation matrix at the FC, denoted by $\mathbf{R}^{Rx}$, is set as a standard Toeplitz matrix with elements $\mathbf{R}^{Rx}(j,k) = \rho_r^{|j-k|}$, where the quantity $\rho_r,\ 0 \leq \rho_r \leq 1$, is the receive antenna correlation coefficient. Similarly, the elements of the transmit antenna correlation matrix $\mathbf{R}_i^{Tx}$ are defined as $\mathbf{R}_i^{Tx}(j,k) = \rho_t^{|j-k|}$, where $\rho_t,\ 0 \leq \rho_t \leq 1$, is the transmit antenna correlation coefficient. The antenna correlation coefficient values at the FC and each sensor are set as $\rho_r = 0.6$ and $\rho_t = 0.6$, respectively, unless stated otherwise. The equivalent channel matrix $\mathbf{H}$ defined in (64) for the coherent MAC-based WSN is generated according to (65). The elements of the stacked observation matrix $\mathbf{G}$ are generated as i.i.d symmetric complex Gaussian random variables with zero mean and unit variance. The dimension of the parameter vector $\boldsymbol{\theta}$ and its sparsity-level, i.e the number of non-zero elements in the sparse parameter vector, are set as $m = 6$ and 2, respectively, unless otherwise mentioned explicitly. The number of TA and RA at each SN i and the FC are set as $t_i = 2$ and $r = 6$, respectively. The number of observations is set to $q_i = 6\ ,\forall\ i$. The observation noise and channel noise covariance matrices are set as $\mathbf{R}_v = \sigma_v^2\mathbf{I}_q$ and $\mathbf{R}_n = \sigma_n^2\mathbf{I}_r$, respectively. The SNR at each SN and the FC is defined as $\mathrm{SNR_{OB}} = \frac{1}{\sigma_v^2}$, $\mathrm{SNR_{FC}} = \frac{1}{\sigma_n^2}$, respectively.

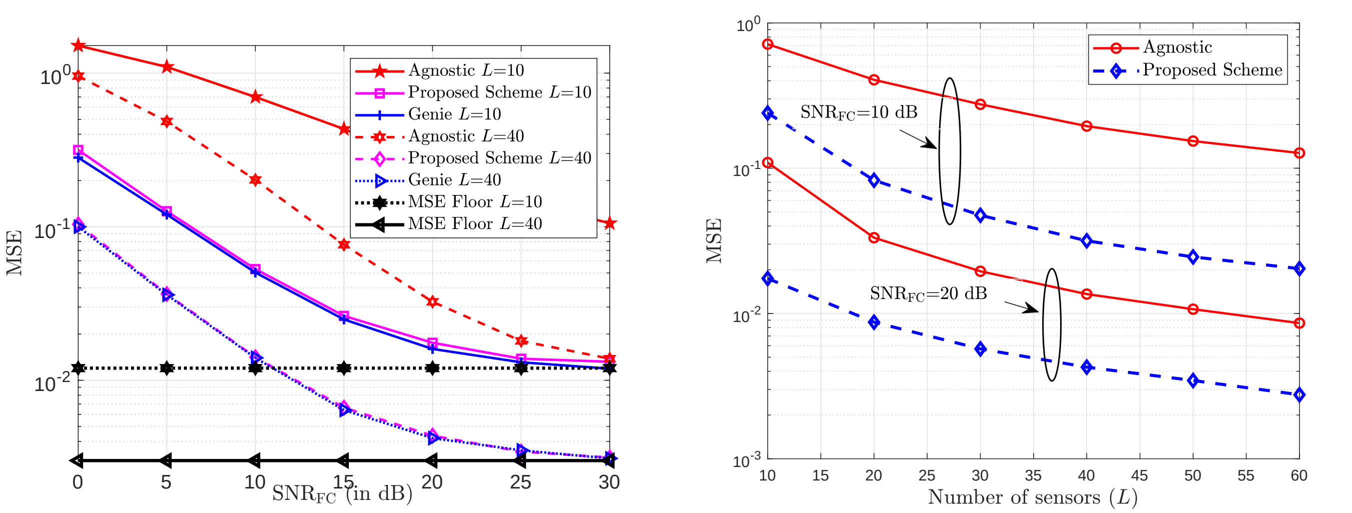
Fig. 4(a) depicts the MSE performance of sparse parameter estimation, for various values of the number of SNs L, against the $\mathrm{SNR_{FC}}$, for noiseless sensor observations. It is shown that the MSE monotonically decreases as the value of $\mathrm{SNR_{FC}}$ increases. The MSE performance of the proposed BL-based sparse parameter estimation scheme coincides with that of a 'Genie' receiver that is assumed to have knowledge of the exact locations of the non-zero coefficient indices of the parameter vector. The performance of the proposed sparsity-oriented BL-based design is also compared to that of the conventional sparsity-'agnostic' estimator. The conventional sparsity agnostic LMMSE estimator sets the a priori covariance $\mathbf{R}_\theta = \mathbf{I}_m$, and thus yields a poor MSE performance. By contrast, the proposed BL-based design considers a parametrized a priori covariance matrix $\mathbf{R}_\theta = \boldsymbol{\Gamma}$, as described in (15), and estimates the associated hyperparameters γ_i using the EM framework developed in (16)-(28). As described in [39], upon convergence, most of the hyperparameters γ_i are actually driven to zero, leading to a sparse estimate of $\boldsymbol{\theta}$. Finally, as the number of SNs L increases, the MSE performance is seen to progressively improve at the FC, which is consistent with the higher number of correlated observations that are available for decentralized parameter estimation. This leads to an improvement in the quality of the estimate of the unknown parameter $\boldsymbol{\theta}$.

Fig. 4(b) also considers a scenario, where the size of the parameter vector $\boldsymbol{\theta}$ is higher than the total number of transmit antennas in the WSN considered, i.e. $m > t$. This is a challenging scenario, since the size of the received vector $\mathbf{y}$ at the FC is still $r = 6$, whereas the length of the sparse parameter vector $\boldsymbol{\theta}$ is $m = 30$. This results in a highly under-determined system, as seen from (42). Interestingly, for this challenging scenario, the performance of the agnostic-design exhibits a



(a) (b)

Fig. 4. (a) MSE versus SNR$_{\text{FC}}$ parametrized by the number of SNs $L \in \{10, 30\}$ considering high-SNR sensor observations (b) MSE versus SNR$_{\text{FC}}$ parametrized by the dimension m of the unknown parameter vector $\boldsymbol{\theta}$, $m \in \{6, 30\}$ considering high-SNR sensor observations.



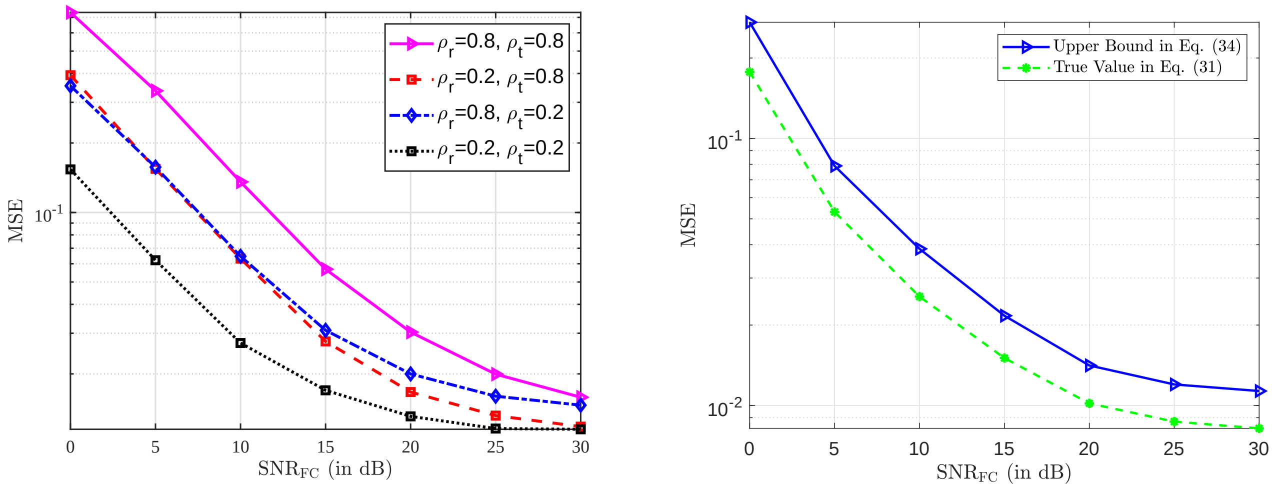
(a) (b)

Fig. 5. (a) MSE versus SNR$_{\text{FC}}$ for $L \in \{10, 40\}$ number of SNs considering arbitrary-SNR noisy sensor observations, and the MSE floors are calculated using (16) in Section-III of our technical report [51] (b) MSE versus the number of SNs (L) for noisy sensor observations and SNR$_{\text{FC}} \in \{10, 20\}$ dB.

floor, as seen in the figure, whereas the the proposed BL-based framework yields a performance comparable to the scenario of $t \geq m$, thanks to the sparse signal recovery guarantees available from the compressed measurements together with the superior convergence properties of the BL-framework, as established in [39].

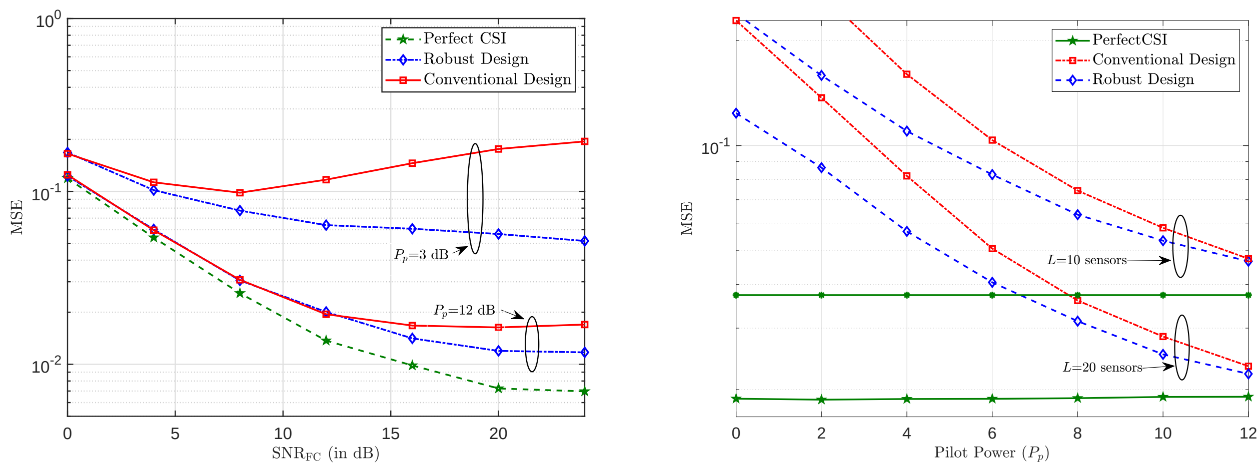
Fig. 5(a) shows the MSE performance of the proposed BL-based sparse parameter estimation scheme developed in Section-III for noisy sensor observations, once again versus SNR$_{\text{FC}}$. Interestingly, whilst the MSE can be seen to progressively decrease in the low-SNR regime, the performance gradually flattens and attains a floor at higher values of SNR$_{\text{FC}}$. An analytical expression for this phenomenon which has been derived in Section-III of our technical report [51]. This is due to the amplification of the observation noise by the TPC matrix $\mathbf{B}$. Similar to the previous scenarios, the proposed scheme is once again seen to match the ideal 'Genie' benchmark associated with known sparsity, which demonstrates the efficacy of the former. A significant performance gain is also observed over the conventional agnostic estimator. Fig. 5(b) demonstrates the MSE performance of the noisy sensor observations scenario versus the number of SNs L. The MSE performance at the FC improves upon increasing L, as it is expected.

Fig. 6(a) examines the effect of channel correlation on the MSE performance by plotting the MSE for different values of the receive and transmit correlation parameters $\rho_r \in \{0.2, 0.8\}$ and $\rho_t \in \{0.2, 0.8\}$, respectively. It is observed that channels



(a) (b)

Fig. 6. (a) MSE versus $\mathrm{SNR}_{\mathrm{FC}}$ for different values of receive and transmit correlation coefficients $\rho_r \in \{0.2, 0.8\}$, $\rho_t \in \{0.2, 0.8\}$ (b) MSE performance comparison of the proposed upper bound and the actual MSE expression.



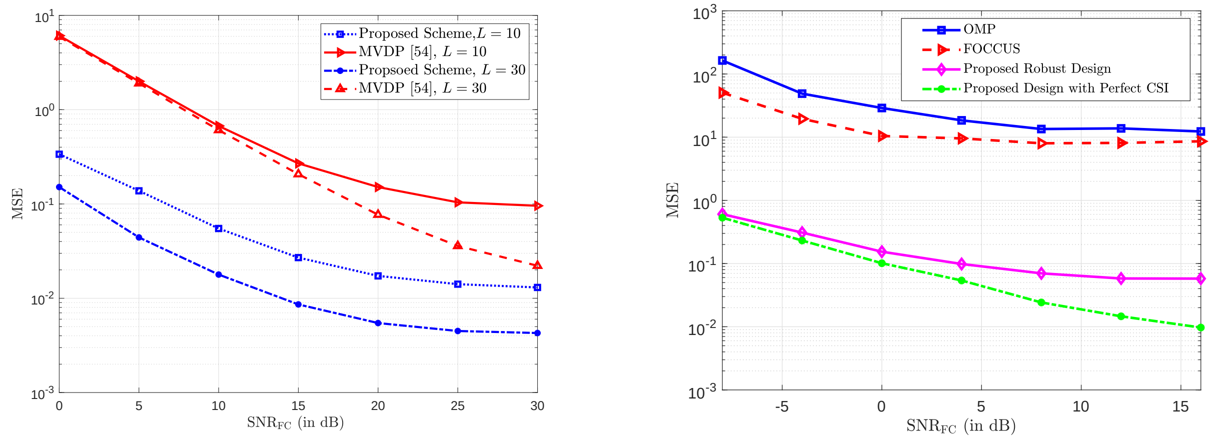
(a) (b)

Fig. 7. (a) MSE versus $\mathrm{SNR}_{\mathrm{FC}}$ for the scheme proposed in Section-IV with $P_T = 20$, $L = 20$, and $\mathrm{SNR}_{\mathrm{OB}} = 10\mathrm{dB}$ (b) MSE versus Pilot power (P_p) for the scheme proposed in Section-IV with $\mathrm{SNR}_{\mathrm{OB}} = 10\mathrm{dB}$, $\mathrm{P}_T = 20$, $\rho_r = 0.6$ and $L \in \{10, 20\}$.

having lower correlation at the transmitter and the receiver yield the best MSE performance. At high SNR, the receive correlation has a greater influence on the MSE. Fig. 6(b) illustrates the tightness of the Weyl's lemma based upper bound of (34) applied to the MSE. For this study, the MSE performance using the precoder $\mathbf{B}_i$ obtained after minimizing the upper bound (34) and the same upon substituting this $\mathbf{B}_i$ into the expression of the true MSE (31) is plotted. It can be readily observed that the proposed upper bound is a tight approximation of the true MSE expression of (31).

Fig. 7(a) depicts the MSE performance of the robust transceiver designed using the MMSE channel estimate and its error covariance, as described in Section-IV, against the $\mathrm{SNR}_{\mathrm{FC}}$. As observed from the figure, the proposed robust design has a better MSE performance than the conventional design that ignores the CSI uncertainty contaminating the channel estimate. The MSE performance of a receiver with perfect CSI is also presented to benchmark the MSE performance. It is important to note that the performance gap between the proposed robust and conventional designs is significant in the low-pilot-power regime, where the uncertainty in the available channel estimate is higher. Naturally, the proposed robust design, which takes the CSI uncertainty into consideration leads to improved performance. Fig. 7(b) shows the MSE performance of parameter estimation against the pilot power (P_p) used for CSI acquisition, as described in Section-IV. It can once again be seen that the proposed robust estimator has a significant performance gain over the agnostic design, especially when the pilot power is low. This reinforces the trend seen in Fig. 7(a).

Fig. 8(a) shows the MSE performance comparison of the proposed BL-based sparse parameter estimation scheme developed in Section-III with that of the minimum variance distortionless precoding (MVDP) scheme of [54]. As expected,



(a) (b)

Fig. 8. (a) MSE vs $\mathrm{SNR_{FC}}$ performance comparison of the proposed scheme in Section-III with the MVDP scheme [54] in the existing literature for $L \in \{10, 30\}$ (b) MSE performance comparison of the proposed robust BL-based sparse signal recovery scheme with other sparse signal recovery schemes like OMP and FOCUSS.

the proposed BL-based scheme outperforms the MVDP, since the latter does not have any mechanism to exploit the sparsity of the unknown parameter vector of interest. For instance, at $\mathrm{SNR_{FC}} = 15$ dB, the proposed BL-based scheme yields approximately 10 dB MSE reduction compared to the MVDP scheme.

Fig. 8(b) compares the MSE performance of the proposed BL-based robust transceiver design to other popular sparse signal recovery techniques, such as the orthogonal matching pursuit (OMP) [42] and FOcal Underdetermined System Solver (FOCUSS) [40] techniques in presence of CSI uncertainty. It can be readily observed that the proposed design significantly outperforms the OMP and FOCUSS techniques. The poor performance of the OMP can be attributed to its sensitivity to the stopping parameter as well as to the sensing matrix $\widehat{\mathbf{\Phi}}^{(k)}$, while the poor performance of the FOCUSS arises due to its convergence deficiencies and sensitivity to the regularization parameter [39]. By contrast, the proposed BL-based design is robust to the sensing matrix $\widehat{\mathbf{\Phi}}^{(k)}$, and the well-established properties of its cost-function of (19), as described in [39], followed by the EM-framework, guarantees convergence to sparse solutions.

VI. Conclusions

A coherent MAC-based MIMO WSN relying on novel BL-based schemes employed for linear decentralized estimation of a sparse parameter vector was proposed. Both ideal noiseless as well as non-ideal noisy SNs were considered, and both, the transmitting as well as receiving antenna correlation of the MIMO channels were incorporated. Furthermore, MMSE channel estimation considering the channel's correlation was conceived, followed by a robust transceiver design, which directly takes into accounts the channel estimation error. Finally, extensive simulation results were presented for characterizing the performance of the proposed schemes. Future research may also consider a quality of service (QoS) based formulation, where the objective is to minimize the total power consumption in the network, while satisfying a given MSE threshold in presence of CSI uncertainty. Furthermore, one can also explore the problem of sparse parameter estimation with quantized measurement transmission with and without CSI uncertainty.

Appendix A
Lemma from [55]

Lemma 1. *Let* $\mathbf{Y} \in \mathbb{C}^{a \times b}$ *be a random matrix with mean* $\bar{\mathbf{Y}} \in \mathbb{C}^{a \times b}$*, covariance matrix* $\sigma_y^2(\mathbf{K} \otimes \mathbf{T})$ *where* $\mathbf{K} \in \mathbb{C}^{b \times b}$ *and* $\mathbf{T} \in \mathbb{C}^{a \times a}$ *are positive definite matrices. Then* $\mathbf{Y}$ *is distributed as*

$$\mathbf{Y} \sim \mathcal{CN}\left[\bar{\mathbf{Y}}, \sigma_y^2(\mathbf{K} \otimes \mathbf{T})\right].$$

For any matrix $\mathbf{A}$ *of appropriate size*

$$\mathbb{E}\left[\mathbf{Y}^H \mathbf{A} \mathbf{Y}\right] = \bar{\mathbf{Y}}^H \mathbf{A} \bar{\mathbf{Y}} + \sigma_y^2 \operatorname{Tr}(\mathbf{T}\mathbf{A})\, \mathbf{K}^T,$$
$$\mathbb{E}\left[\mathbf{Y} \mathbf{A} \mathbf{Y}^H\right] = \bar{\mathbf{Y}} \mathbf{A} \bar{\mathbf{Y}}^H + \sigma_y^2 \operatorname{Tr}\left(\mathbf{A}\mathbf{K}^T\right) \mathbf{T}.$$

References

[1] I. D. Schizas, G. B. Giannakis, and Z. Luo, "Distributed estimation using reduced-dimensionality sensor observations," *IEEE Transactions on Signal Processing*, vol. 55, no. 8, pp. 4284–4299, 2007.

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

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

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

[5] M. Gastpar, "Uncoded transmission is exactly optimal for a simple Gaussian "sensor" network," *IEEE Transactions on Information Theory*, vol. 54, no. 11, pp. 5247–5251, 2008.

[6] M. Gastpar and M. Vetterli, "Source-channel communication in sensor networks," in *Information Processing in Sensor Networks*. Springer, 2003, pp. 162–177.

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

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

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

[10] A. Shirazinia, S. Dey, D. Ciuonzo, P. S. Rossi *et al.*, "Massive MIMO for decentralized estimation over coherent multiple access channels." in *SPAWC*. Citeseer, 2015, pp. 241–245.

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

[12] J. Liu and C. Chung, "Distributed estimation in a wireless sensor network using hybrid MAC," *IEEE Transactions on Vehicular Technology*, vol. 60, no. 7, pp. 3424–3435, 2011.

[13] C. Zhan, Y. Zeng, and R. Zhang, "Trajectory design for distributed estimation in UAV-enabled wireless sensor network," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 10, pp. 10 155–10 159, 2018.

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

[15] D. Ciuonzo, P. S. Rossi, and S. Dey, "Massive MIMO channel-aware decision fusion," *IEEE Transactions on Signal Processing*, vol. 63, no. 3, pp. 604–619, 2015.

[16] I. Dey, D. Ciuonzo, and P. S. Rossi, "Wideband collaborative spectrum sensing using massive MIMO decision fusion," *IEEE Transactions on Wireless Communications*, vol. 19, no. 8, pp. 5246–5260, 2020.

[17] Q. Ling and Z. Tian, "Decentralized sparse signal recovery for compressive sleeping wireless sensor networks," *IEEE Transactions on Signal Processing*, vol. 58, no. 7, pp. 3816–3827, July 2010.

[18] C. Luo, F. Wu, J. Sun, and C. W. Chen, "Compressive data gathering for large-scale wireless sensor networks," in *Proceedings of the 15th Annual International Conference on Mobile Computing and Networking*, ser. MobiCom '09. New York, NY, USA: ACM, 2009, pp. 145–156. [Online]. Available: http://doi.acm.org/10.1145/1614320.1614337

[19] S. Li, L. D. Xu, and X. Wang, "Compressed sensing signal and data acquisition in wireless sensor networks and internet of things," *IEEE Transactions on Industrial Informatics*, vol. 9, no. 4, pp. 2177–2186, Nov 2013.

[20] J. Bobin, J. Starck, and R. Ottensamer, "Compressed sensing in astronomy," *IEEE Journal of Selected Topics in Signal Processing*, vol. 2, no. 5, pp. 718–726, Oct 2008.

[21] X. Wang, G. Li, and P. K. Varshney, "Detection of sparse stochastic signals with quantized measurements in sensor networks," *IEEE Transactions on Signal Processing*, vol. 67, no. 8, pp. 2210–2220, April 2019.

[22] J. Hua and C. Li, "Distributed jointly sparse Bayesian learning with quantized communication," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 4, no. 4, pp. 769–782, Dec 2018.

[23] S. M. Fosson, J. Matamoros, C. Antón-Haro, and E. Magli, "Distributed recovery of jointly sparse signals under communication constraints," *IEEE Transactions on Signal Processing*, vol. 64, no. 13, pp. 3470–3482, July 2016.

[24] S. Khanna and C. R. Murthy, "Communication-efficient decentralized sparse Bayesian learning of joint sparse signals," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 3, no. 3, pp. 617–630, Sep. 2017.

[25] ——, "Decentralized joint-sparse signal recovery: A sparse Bayesian learning approach," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 3, no. 1, pp. 29–45, March 2017.

[26] X. Wang, G. Li, C. Quan, and P. K. Varshney, "Distributed detection of sparse stochastic signals with quantized measurements: The generalized Gaussian case," *IEEE Transactions on Signal Processing*, vol. 67, no. 18, pp. 4886–4898, 2019.

[27] L. Jacques, D. K. Hammond, and J. M. Fadili, "Dequantizing compressed sensing: When oversampling and non- Gaussian constraints combine," *IEEE Transactions on Information Theory*, vol. 57, no. 1, pp. 559–571, 2011.

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

[29] X. Wang and C. Yang, "Decentralized estimation over orthogonal multiple-access fading channels in wireless sensor networks–optimal and suboptimal estimators," *EURASIP Journal on Advances in Signal Processing*, vol. 2011, no. 1, p. 132, 2011.

[30] G. Taricco, "Joint channel and data estimation for wireless sensor networks," *IEEE Wireless Communications Letters*, vol. 1, no. 5, pp. 532–535, 2012.

[31] C.-H. Wu and C.-A. Lin, "Linear coherent distributed estimation over unknown channels," *Signal Processing*, vol. 91, no. 4, pp. 1000–1011, 2011.

[32] M. K. Banavar, C. Tepedelenlioglu, and A. Spanias, "Estimation over fading channels with limited feedback using distributed sensing," *IEEE Transactions on Signal Processing*, vol. 58, no. 1, pp. 414–425, 2010.

[33] C.-H. Wang and S. Dey, "Distortion outage minimization in Nakagami fading using limited feedback," *EURASIP Journal on Advances in Signal Processing*, vol. 2011, no. 1, p. 92, 2011.

[34] M. Fanaei, M. C. Valenti, and N. A. Schmid, "Limited-feedback-based channel-aware power allocation for linear distributed estimation," in *2013 Asilomar Conference on Signals, Systems and Computers*, 2013, pp. 547–551.

[35] J. Zhu, R. S. Blum, X. Lin, and Y. Gu, "Robust transmit beamforming for parameter estimation using distributed sensors," *IEEE Communications Letters*, vol. 20, no. 7, pp. 1329–1332, July 2016.

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

[37] Y. Liu, J. Li, and H. Wang, "Robust linear beamforming in wireless sensor networks," *IEEE Transactions on Communications*, vol. 67, no. 6, pp. 4450–4463, 2019.

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

[39] D. P. Wipf and B. D. Rao, "Sparse Bayesian learning for basis selection," *IEEE Transactions on Signal Processing*, vol. 52, no. 8, pp. 2153–2164, Aug 2004.

[40] I. F. Gorodnitsky and B. D. Rao, "Sparse signal reconstruction from limited data using FOCUSS: A re-weighted minimum norm algorithm," *IEEE Transactions on Signal Processing*, vol. 45, no. 3, pp. 600–616, 1997.

[41] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM review*, vol. 43, no. 1, pp. 129–159, 2001.

[42] S. F. Cotter and B. D. Rao, "Sparse channel estimation via matching pursuit with application to equalization," *IEEE Transactions on Communications*, vol. 50, no. 3, pp. 374–377, 2002.

[43] S. M. Kay, *Fundamentals of statistical signal processing*. Prentice Hall PTR, 1993.

[44] *Compressed Sensing: Theory and Applications*. Cambridge University Press, 2012.

[45] K. Yu, M. Bengtsson, B. Ottersten, D. McNamara, P. Karlsson, and M. Beach, "Modeling of wide-band MIMO radio channels based on NLoS indoor measurements," *IEEE Transactions on Vehicular Technology*, vol. 53, no. 3, pp. 655–665, May 2004.

[46] G. McLachlan and T. Krishnan, *The EM algorithm and extensions*. John Wiley & Sons, 2007, vol. 382.

[47] C. J. Wu, "On the convergence properties of the EM algorithm," *The Annals of statistics*, pp. 95–103, 1983.

[48] R. A. Horn and C. R. Johnson, *Matrix Analysis*, 2nd ed. Cambridge University Press, 2012.

[49] J. B. Lasserre, "A trace inequality for matrix product," *IEEE Transactions on Automatic Control*, vol. 40, no. 8, pp. 1500–1501, 1995.

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

[51] K. P. Rajput, A. Kumar, S. Srivastava, A. K. Jagannatham, and L. Hanzo, "Technical report: Bayesian learning-based linear decentralized sparse parameter estimation in MIMO wireless sensor networks relying on imperfect CSI." IIT Kanpur, Tech. Rep., 2021. [Online]. Available: http://www.iitk.ac.in/mwn/documents/MWNLab_TR_Sparse_WSN_21.pdf.

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

[53] ——, "High SNR linear estimation of vector sources," *IEEE Wireless Communications Letters*, vol. 3, no. 6, pp. 581–584, 2014.

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

[55] A. K. Gupta and D. K. Nagar, *Matrix variate distributions*. Chapman and Hall/CRC, 2018.