Robust Linear Decentralized Tracking of a Time-Varying Sparse Parameter Relying on Imperfect CSI

Kunwar Pritiraj Rajput, *Member, IEEE,* Suraj Srivastava, *Member, IEEE,* Aditya K. Jagannatham, *Senior Member, IEEE,* and Lajos Hanzo, *Life Fellow, IEEE*

***Abstract*—Robust linear decentralized tracking of a time varying sparse parameter is studied in a multiple-input multiple-output (MIMO) wireless sensor network (WSN) under channel state information (CSI) uncertainty. Initially, assuming perfect CSI availability, a novel sparse Bayesian learning-based Kalman filtering (SBL-KF) framework is developed in order to track the time varying sparse parameter. Subsequently, an optimization problem is formulated to minimize the mean square error (MSE) in each time slot, followed by the design of a fast block coordinate descent (FBCD)-based iterative algorithm. A unique aspect of the proposed technique is that it requires only a single iteration per time slot to obtain the transmit precoder (TPC) matrices for all the sensor nodes (SNs) and the receiver combiner (RC) matrix for the fusion center (FC) in an online fashion. The recursive Bayesian Cramer Rao bound (BCRB) is also derived for benchmarking the performance of the proposed linear decentralized estimation (LDE) scheme. Furthermore, for considering a practical scenario having CSI uncertainty, a robust SBL-KF (RSBL-KF) is derived for tracking the unknown parameter vector of interest followed by the conception of a robust transceiver design. Our simulation results show that the schemes designed outperform both the traditional sparsity-agnostic KF and the state-of-the-art sparse reconstruction methods. Furthermore, as compared to the uncertainty-agnostic design, the robust transceiver architecture conceived is shown to provide improved estimation performance, making it eminently suitable for practical applications.**

***Index Terms*—Coherent MAC, Kalman filter, linear decentralized estimation, sparse Bayesian learning, stochastic CSI uncertainty, time varying sparse parameter, wireless sensor network.**

I. Introduction

Wireless sensor networks (WSNs) have the potential to support a wide range of cutting-edge applications, including environmental monitoring [1], smart agriculture [2], smart grid [3], target tracking [4], [5], and so on. As a result, they will play a pivotal role in the Internet of Things (IoT) [6]. The sensor nodes (SNs) in a typical WSN sense/monitor multiple parameter(s) of interest, and subsequently transmit the suitably pre-processed observations over a wireless channel to a fusion center (FC) for further processing. To achieve a high bandwidth efficiency, transmission over a coherent multiple access channel (MAC) is preferred in such systems, which necessitates the development of sophisticated signal processing techniques for accurate parameter estimation at the FC. Such a sensing setup, which relies on pre-processing and post-processing operations at the transmitter and receiver respectively, is termed as linear decentralized estimation (LDE) in the literature [7]. Specifically, one has to design the optimal receive combiner (RC) of the FC having the minimum mean square error (MMSE) along with the corresponding transmit precoder (TPC) matrices to be used at the various SNs. Due to its significance, numerous researchers have studied the problem of LDE for scalar [8]–[12] and vector [7], [13]–[17] parameters, while considering diverse scenarios of interest, when the parameter(s) of interest is (are) deterministic/ random, uncorrelated/ correlated, static/ time varying in nature. Next, a discussion of the most important studies on LDE is provided.

A. Existing literature

The problem of LDE was first explored in their seminal work by Xiao *et al.* [7] where new schemes were proposed for the estimation of both scalar as well as vector parameters. However, their framework considered a diagonal channel between each SN and the FC, which restricts the applicability of their results. This shortcoming was overcome by the alternate minimization based transceiver designs proposed in [15] and [18]. An innovative minimum variance distortionless precoding (MVDP) framework was proposed in [19] for unbiased parameter estimation, which does not require a combiner at the FC. The authors of [20]–[24] studied the problem of LDE in an energy harvesting WSNs where the SNs run on the energy harvested from the radio frequency signals transmitted by the different access points. LDE in massive MIMO WSNs was also explored in [25]–[27] where the FC is equipped with a large number of antennas. A major limitation of the works mentioned above is that they consider a static parameter in their respective models. However, in practice, the physical quantities under observation are typically of time-varying nature, exhibiting temporal correlation. The studies related to the LDE of a time-varying parameter are discussed next.

Leong *et al.* [28] devised LDE techniques for a scalar time-varying parameter with the goal of minimising the MSE at the

Kunwar Pritiraj Rajput, Suraj Srivastava, and Aditya K. Jagannatham are with the Department of Electrical Engineering, Indian Institute of Technology Kanpur, 208016 India, e-mail: {pratiraj, ssrivast, adityaj}@iitk.ac.in.

Lajos 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).

L. Hanzo would like to acknowledge the financial support of the Engineering and Physical Sciences Research Council projects EP/W016605/1 and EP/X01228X/1 as well as of the European Research Council's Advanced Fellow Grant QuantCom (Grant No. 789028)

TABLE I: Boldly and explicitly contrasting our contributions to the literature

Feature	[20]	[26]	[28]	[32]	[36]	[39]	[40]	[41]	[45]	[47]	[48]	[49]	Our work
Multi-sensor MIMO network	✓		✓	✓	✓				✓	✓	✓	✓	✔
Linear decentralized estimation	✓	✓	✓	✓	✓				✓	✓	✓	✓	✔
Coherent MAC	✓	✓		✓	✓				✓	✓	✓	✓	✔
Time varying sparse parameter estimation						✓	✓	✓					✔
Temporal correlation				✓									✔
Amplify and forward	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✔
Per SN power constraint	✓		✓	✓	✓				✓	✓		✓	✔
Recursive BCRB													✔
Robust transceiver design									✓	✓	✓	✓	✔
SBL-KF													✔

FC, while adhering to a total power limitation, or minimising the power consumption, while meeting a certain maximum MSE. The authors of [29] proposed a low-complexity sub-optimal power allocation scheme for the minimization of the outage probability for estimating a dynamic parameter in an orthogonal MAC-based WSN. However, the orthogonal MAC has low bandwidth efficiency [7], hence it is unsuitable for large-scale WSNs. Jiang *et al.* [30] consider the problem of estimating a dynamic parameter under both total and individual sensor power limitations. Closed-form expressions were obtained for the amplitude and phase of the TPC coefficients at each SN for the scenario of a total power constraint, whereas numerical optimization routines were employed for the system with individual SN power constraints. Singh and Rajawat [31] proposed a scheme for the LDE of a vector parameter by considering the MSE as their cost function subject to individual SN power constraints. Furthermore, the authors of [32] conceived both decentralised and distributed sequential linear MMSE (LMMSE) techniques for static and dynamic vector parameter estimation, using time-varying channel and observation matrices. However, the aforementioned solutions are not optimized for realistic sparse parameters, which is the topic of the next section.

The authors of [33]–[35] studied the problem of distributed detection of a sparse stochastic parameter under different assumptions concerning its distribution, considering also the transmission of quantized measurements to the FC. The sparse parameter vector in their study is assumed to be Bernoulli-Gaussian distributed in [33], generalized-Gaussian distributed in [34], and Bernouli-Gaussian distributed with 1-bit quantized measurements at each SN in [35]. Khanna and Murthy in [36] and [37] proposed a pair of different distributed algorithms, which rely on the sparse Bayesian learning (SBL) and alternating direction method of multipliers (ADMM) techniques, respectively, for the estimation of jointly sparse signals in a WSN. It is important to note that, in their work, the sparse parameters of the various SNs are assumed to be different, while sharing a common sparsity profile. Along similar lines, the authors of [38] proposed an innovative scheme for joint sparse signal estimation relying on quantized SN measurement transmissions using an inexact ADMM framework. Novel algorithms are presented in [39] for the estimation of a sparse parameter vector which are derived by combining the Kalman filter (KF) and a pseudo measurement technique based on the l_1 and l_p norms, respectively. A Kalman-like-particle-filter (KLPF)-based time varying sparse parameter estimation scheme was proposed in [40] under realistic communication constraints. In their scheme, the FC transmits its predicted observation to a subset of the selected SNs, and subsequently, each of those SNs transmits its quantized innovation to the FC. Karseras *et al.* [41] proposed a path breaking hierarchical Bayesian Kalman filter (HBKF)-based algorithm for time varying sparse parameter estimation in a WSN. However, the authors of [39]–[41] do not consider a coherent MAC based WSN. Furthermore, they do not develop the MSE-optimal transceiver that can overcome the fading-induced degradation of the wireless channel, while taking into account the power budget of each SN. The authors of [42] have proposed a novel technique for time-varying massive MIMO channel estimation which also exploits the SBL framework. Furthermore, an interesting uplink-aided downlink channel estimation scheme is developed in [43] for a massive MIMO-orthogonal time frequency space (OTFS) system which utilizes the expectation maximization based variational Bayesian (EM-VB) framework. Furthermore, with the exception of [26], all of the contributions covered thus far assume the presence of perfect CSI in their analysis, which is an idealized simplifying assumption that often does not hold in practice.

In order to limit the performance deterioration arising as a result of realistic CSI errors, and thus achieve robust performance, it is imperative to take the CSI uncertainty into account during the design of LDE schemes. There are a few papers, where the authors have designed robust precoders/ combiners to combat the deleterious effects of imperfect CSI, which are reviewed next.

The robust TPC/ RC designed for LDE of a scalar parameter was proposed in [44], [45]. Specifically, the robust LDE scheme of [44] relies on the MVDP framework that does not require a combiner at the FC. Liu *et al.* developed robust decentralised and distributed transceivers for worst-case MSE reduction under individual sensor power limits as well as for overall power minimization under a specific worst-case MSE threshold, in their exposition in [45]. Rostami and Falahati [46] proposed a robust TPC design for vector parameter estimation under infinite and total network power constraints. The authors of [47] and [48] have proposed innovative robust LDE schemes based on majorization theory for non-sparse and sparse parameter estimation, respectively, where closed-form expressions for robust TPCs/ RC matrices were derived. Furthermore, the robust LDE of a temporally correlated parameter vector was considered in [49] under imperfect CSI considering both analog as well as quantized SN measurement transmission.

However, there is no previous study in the literature that deals with the LDE of a dynamic sparse parameter in a coherent MAC-based WSN depending on imperfect CSI. In order to fill this knowledge gap, we develop a novel SBL-KF and fast block coordinate descent (FBCD) based LDE scheme for time-varying sparse vector parameter tracking in the presence of realistic CSI uncertainty. Table I boldly compares the current and previous contributions on a feature-by-feature basis, which are further detailed below. Please observe that rows 8 and 10 are entirely unique to our treatise.

B. Contributions

- To begin with, this paper derives the SBL-KF framework for the LDE of a time-varying sparse parameter considering a scenario relying on the idealized simplifying assumption of full CSI availability, followed by the formulation of a per time slot MSE minimization problem.
- An FBCD algorithm is conceived in order to solve the resultant non-convex problem for transceiver design, which requires only a single iteration per time slot. This significantly reduces the computational complexity in comparison to the conventional version of the BCD algorithm.
- A recursive Bayesian Cramer Rao bound (BCRB) is also derived for benchmarking the performance of the proposed LDE schemes.
- Next, we consider the scenario where only imperfect CSI is available for the channel between each SN and the FC and propose a robust SBK-KF (RSBL-KF) scheme. The per time slot average MSE minimization problem is solved for this system to design a robust transceiver.
- Finally, our extensive simulations verify the proposed SBL-KF- and RSBL-KF-based time varying sparse parameter estimation framework under perfect as well as imperfect CSI scenarios, respectively. The results obtained corroborate the analytical formulations derived in this work and also show the performance improvement over conventional schemes.

C. Organization and Notation

The rest of the paper is structured as follows. Section II presents the system model for LDE in a MIMO WSN, followed by the SBL-KF-based time-varying sparse parameter tracking framework and transceiver design for per time slot MSE minimization in Section III for perfect CSI availability. Section IV derives our stochastic CSI uncertainty model based robust time varying sparse parameter tracking framework. The simulation results are discussed in Section V, and our conclusions are presented in Section VI. The Karush-Kuhn-Tucker (KKT) framework-based Lagrange multiplier computation is harnessed in the Appendix A for computing the TPC for each SN, while the recursive BCRB benchmark is derived in Appendix B.

Throughput the paper, bold lower ($\mathbf{a}$) and uppercase ($\mathbf{A}$) letters denote vectors and matrices, respectively; $\mathbf{a}(i)$ denotes the ith value of the vector $\mathbf{a}$, while the $[\mathbf{A}]_{ij}$ denotes the (i,j)th value of the matrix $\mathbf{A}$; The matrix $\mathbf{A} = \mathcal{D}(a_1, a_2, \ldots, a_N)$ represents a diagonal matrix $\mathbf{A}$ of size $N \times N$ with the elements a_i, for $i = 1, 2, \ldots, N$ on its principal diagonal; The trace and expectation operators are denoted by Tr[.] and $\mathbb{E}[.]$, respectively; The transpose, Hermitian and complex conjugate operations are denoted by $(.)^T$, $(.)^H$ and $(.)^*$, respectively; A q dimensional vector quantity with all its q elements equal to zero is denoted by $\mathbf{0}_{q\times 1}$, while $\mathbf{I}_q$ represents an identity matrix of dimension $q \times q$; The notation $\det(\mathbf{A})$ represents the determinant of a matrix $\mathbf{A}$; $\mathbf{a} \sim \mathcal{CN}(\mathbf{0}, \mathbf{R}_a)$ represents the circularly symmetric complex Gaussian distribution with mean zero and covariance matrix $\mathbf{R}_a$.

II. System Model

Consider a WSN where K sensors are sensing a common sparse parameter of interest, denoted by $\boldsymbol{\theta} \in \mathbb{C}^{q\times 1}$ and each sensor is equipped with N_s transmit antennas (TAs), as shown in Fig. 1. An important point to note is that the unknown parameter vector $\boldsymbol{\theta}$ is time-varying and sparse in nature, whose sparsity level is not known *a priori*. Moreover, the support of the sparse vector is also dynamic in nature, i.e., the set of indices of the non-zero elements of the vector is changing with respect to time. The parameter of interest $\boldsymbol{\theta}[m]$ at time instant m, is assumed to evolve according to the first-order auto-regressive (AR-1) model, [50, Sec. A1.2.4]

$$\boldsymbol{\theta}[m] = \rho\boldsymbol{\theta}[m-1] + \sqrt{1-\rho^2}\mathbf{u}[m], \tag{1}$$

where the quantity ρ denotes the temporal correlation coefficient and $\mathbf{u}[m]$ represents the state innovation noise. This has been shown to be well suited to capture the temporal variation of the parameter in several IoT applications, as evidenced by the pioneering works [30]–[32]. The observation vector corresponding to SN k at time slot (TS) m, denoted by $\mathbf{x}_k[m] \in \mathbb{C}^{l\times 1}$, is modeled as

$$\mathbf{x}_k[m] = \mathbf{A}_k\boldsymbol{\theta}[m] + \mathbf{v}_k[m], \tag{2}$$

where the matrix $\mathbf{A}_k \in \mathbb{C}^{l\times q}$ denotes the kth SN's observation matrix, while $\mathbf{v}_k[m] \in \mathbb{C}^{l\times 1}$ represents the corresponding observation noise vector, which is assumed to be distributed as $\mathcal{CN}(\mathbf{0}, \mathbf{R}_k)$. The quantity l signifies the number of observations taken by each SN. In order to combat the impediments arising due to the fading nature of the wireless channel and also to efficiently utilize the limited power budget, the observation corresponding to each SN is pre-processed using the TPC $\mathbf{P}_k[m] \in \mathbb{C}^{N_s\times l}$. Subsequently, each SN transmits its precoded observations to the FC over a coherent MAC. At the FC, the received vector $\mathbf{y}[m] \in \mathbb{C}^{N_{\text{FC}}\times 1}$ is modeled as

$$\begin{aligned}\mathbf{y}[m] &= \sum_{k=1}^{K}\mathbf{H}_k\mathbf{P}_k[m]\mathbf{x}_k[m] + \mathbf{v}_{\text{FC}}[m]\\ &= \sum_{k=1}^{K}\mathbf{H}_k\mathbf{P}_k[m]\mathbf{A}_k\boldsymbol{\theta}[m] + \sum_{k=1}^{K}\mathbf{H}_k\mathbf{P}_k[m]\mathbf{v}_k[m] + \mathbf{v}_{\text{FC}}[m].\end{aligned} \tag{3}$$

The quantities $\mathbf{H}_k \in \mathbb{C}^{N_{\text{FC}}\times N_s}$ and $\mathbf{v}_{\text{FC}}[m] \sim \mathcal{CN}(\mathbf{0}, \mathbf{R}_{\text{FC}}) \in \mathbb{C}^{N_{\text{FC}}\times 1}$ denote the wireless fading channel between the kth SN as well as the FC and the FC noise for the mth TS, respectively. The quantity N_{FC} denotes the number of antennas

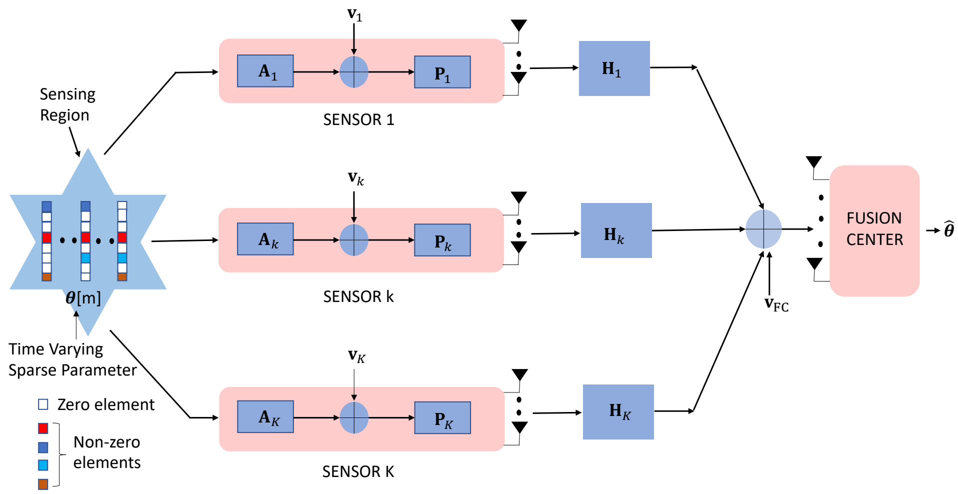

Fig. 1: System model for linear decentralized time varying sparse parameter tracking in a WSN.

at the FC. The expression for the average transmit power corresponding to the kth SN is evaluated as

$$\begin{aligned}\mathbb{E}\left[||\mathbf{P}_k[m]\mathbf{x}_k[m]||_2^2\right] &= \mathbb{E}\left[\text{Tr}\left[\mathbf{P}_k[m]\mathbf{x}_k[m]\mathbf{x}_k^H[m]\mathbf{P}_k^H[m]\right]\right] \\ &= \text{Tr}\left[\mathbf{P}_k[m]\mathbb{E}\left[\mathbf{x}_k[m]\mathbf{x}_k^H[m]\right]\mathbf{P}_k^H[m]\right] \\ &= \text{Tr}\left[\mathbf{P}_k[m]\left[\mathbf{A}_k\mathbf{R}_\theta[m]\mathbf{A}_k^H+\mathbf{R}_k\right]\mathbf{P}_k^H[m]\right], \end{aligned} \tag{4}$$

where $\mathbf{R}_\theta[m] = \mathbb{E}\left[\boldsymbol{\theta}[m]\boldsymbol{\theta}^H[m]\right] \in \mathbb{C}^{q\times q}$. Let $\zeta_k[m]$ represent the kth SN's maximum transmit power budget for the mth TS. The average transmit power of the kth SN can be constrained as

$$\text{Tr}\Big[\mathbf{P}_k[m]\left[\mathbf{A}_k\mathbf{R}_\theta[m]\mathbf{A}_k^H+\mathbf{R}_k\right]\mathbf{P}_k^H[m]\Big] \leq \zeta_k[m]. \tag{5}$$

For the above setup, the next section develops a novel SBL-KF based approach for tracking a time-varying sparse parameter vector and also subsequently, an FBCD-based transceiver design for its efficient estimation.

III. Online Sparse Parameter Tracking and Transceiver Design under Perfect CSI

This section exploits the SBL-KF framework for tracking a time-varying sparse parameter $\boldsymbol{\theta}$, which combines the state-of-the-art SBL and KF techniques.

A. SBL-KF approach for time varying sparse parameter tracking

While the conventional KF efficiently tracks a time-varying parameter, its estimate is not guaranteed to be sparse in nature. On the other hand, SBL has a superior sparse signal estimation capability, wherein the sparsity profile is fixed, i.e. the support of the sparse vector is time-invariant. However, it fails to efficiently track a time-varying sparse parameter, especially in scenarios where the parameters of interest exhibit a significant temporal correlation. Thus, in order to successfully track a time-varying sparse parameter, one is naturally motivated to develop the SBL-KF, which leverages the advantages of both the conventional KF and SBL techniques. This subsection derives the various steps involved in the SBL-KF approach.

Let the set $\boldsymbol{\Psi}_{m-1} = \{\mathbf{y}[0], \mathbf{y}[1], \ldots, \mathbf{y}[m-1]\}$ represent the collection of all the received vectors at the FC spanning from TS 0 to $m-1$. Relying on the previously obtained vectors, the predicted estimate of the underlying sparse parameter vector $\boldsymbol{\theta}[m]$ and the received vector $\mathbf{y}[m]$ at the mth TS is formulated as [50, Sec. 13.3]

$$\begin{aligned}\widehat{\boldsymbol{\theta}}[m|m-1] &= \mathbb{E}\left[\boldsymbol{\theta}[m]|\boldsymbol{\Psi}_{m-1}\right] \\ &= \mathbb{E}\left[\rho\boldsymbol{\theta}[m-1]+\sqrt{1-\rho^2}\mathbf{u}[m]|\boldsymbol{\Psi}_{m-1}\right] \\ &= \rho\widehat{\boldsymbol{\theta}}[m-1|m-1], \end{aligned} \tag{6}$$

$$\begin{aligned}\widehat{\mathbf{y}}[m|m-1] &= \mathbb{E}\left[\mathbf{y}[m]|\boldsymbol{\Psi}_{m-1}\right] \\ &= \sum_{k=1}^{K}\mathbf{H}_k\mathbf{P}_k[m]\mathbf{A}_k\widehat{\boldsymbol{\theta}}[m|m-1], \end{aligned} \tag{7}$$

where the quantities $\widehat{\boldsymbol{\theta}}[m|m-1]$ and $\widehat{\mathbf{y}}[m|m-1]$ denote the best prediction of the sparse parameter vector $\boldsymbol{\theta}[m]$ and the received observation $\mathbf{y}[m]$ at TS m using $\boldsymbol{\Psi}_{m-1}$. Similarly, $\widehat{\boldsymbol{\theta}}[m-1|m-1]$ represents the filtered estimate of $\boldsymbol{\theta}[m-1]$ using $\boldsymbol{\Psi}_{m-1}$. Furthermore, the simplifications in (6) and (7) exploit the fact that the measurement and the FC noises $\mathbf{u}[m]$, $\mathbf{v}_k[m]$, $1 \leq k \leq K$ and $\mathbf{v}_{\text{FC}}[m]$, respectively, are independent of the received vectors in the set $\boldsymbol{\Psi}_{m-1}$.

The proposed SBL-KF technique assigns a parameterized Gaussian prior to the sparse parameter vector $\boldsymbol{\theta}[m]$ and it is given by

$$p(\boldsymbol{\theta}[m];\boldsymbol{\Gamma}[m]) = \prod_{i=1}^{q}\left(\pi\gamma_i[m]\right)^{-1}\exp\left(\frac{-|\boldsymbol{\theta}[m](i)|^2}{\gamma_i[m]}\right), \tag{8}$$

where the quantity $\gamma_i[m]$ that represents the variance of the ith element is unknown. Hence, it is termed a parameterized Gaussian prior, where the unknown hyperparameters are $\gamma_i, \forall i$, which are learnt using the Bayesian learning module developed in this subsection.. Let the hyperparameter matrix $\boldsymbol{\Gamma}[m] \in \mathbb{C}^{q\times q}$ be defined as $\boldsymbol{\Gamma}[m] = \mathcal{D}\left(\gamma_1[m], \gamma_2[m], \ldots, \gamma_q[m]\right)$. The

$$\mathbf{E}[m|m] = \left[\mathbb{E}\left[\widetilde{\boldsymbol{\theta}}[m|m]\widetilde{\boldsymbol{\theta}}[m|m]^H\right]\right] = \mathbf{E}[m|m-1] - \mathbf{E}[m|m-1]\left[\sum_{k=1}^{K}\mathbf{A}_k^H\mathbf{P}_k^H[m]\mathbf{H}_k^H\right]\mathbf{W}^H[m] - \mathbf{W}[m]\left[\sum_{k=1}^{K}\mathbf{H}_k\mathbf{P}_k[m]\right.$$
$$\left.\mathbf{A}_k\right]\mathbf{E}[m|m-1] + \mathbf{W}[m]\left[\mathbf{H}_k\mathbf{P}_k[m]\mathbf{A}_k\mathbf{E}[m|m-1]\mathbf{A}_k^H\mathbf{P}_k^H[m]\mathbf{H}_k^H + \sum_{j=1,k\neq j}^{K}\mathbf{H}_k\mathbf{P}_k[m]\mathbf{A}_k\mathbf{E}[m|m-1]\mathbf{A}_j^H\mathbf{P}_j^H[m]\mathbf{H}_j^H\right.$$
$$\left. + \sum_{k=1}^{K}\mathbf{H}_k\mathbf{P}_k[m]\mathbf{R}_k\mathbf{P}_k^H[m]\mathbf{H}_k^H + \mathbf{R}_{\text{FC}}\right]\mathbf{W}^H[m]. \tag{12}$$

prediction error covariance matrix can be expressed as [50, Sec. 13.3]

$$\mathbf{E}[m|m-1] = \mathbb{E}\Big[\left(\widehat{\boldsymbol{\theta}}[m|m-1] - \boldsymbol{\theta}[m|m-1]\right)\left(\widehat{\boldsymbol{\theta}}[m|m-1] - \boldsymbol{\theta}[m|m-1]\right)^H\Big].$$

Substituting the expressions of the quantities $\widehat{\boldsymbol{\theta}}[m|m-1]$ and $\boldsymbol{\theta}[m|m-1]$ from (6) and (1), respectively, into the above equation, and exploiting the fact that the quantity $\boldsymbol{\theta}[m-1|m-1]$ and the innovation noise $u[m-1]$ are independent, one obtains the relationship of

$$\mathbf{E}[m|m-1] = \rho^2\mathbf{E}[m-1|m-1] + (1-\rho^2)\mathbf{R}_u[m],$$

where $\mathbf{R}_u[m] \in \mathbb{C}^{q\times q}$ denotes the innovation noise covariance matrix for the mth TS. Since both the sparse parameter $\boldsymbol{\theta}[m]$ and the innovation noise $\mathbf{u}[m]$ share the same sparsity profile, i.e. the indices of the zero and non-zero components are identical, we have $\mathbf{R}_u[m] = \boldsymbol{\Gamma}[m]$. Hence, the final expression of the prediction error covariance matrix becomes:

$$\mathbf{E}[m|m-1] = \rho^2\mathbf{E}[m-1|m-1] + (1-\rho^2)\boldsymbol{\Gamma}[m]. \tag{9}$$

The filtered estimate of the unknown sparse parameter vector at the mth TS can be updated as [50, Sec. 13.3]

$$\widehat{\boldsymbol{\theta}}[m|m] = \widehat{\boldsymbol{\theta}}[m|m-1] + \mathbf{W}[m]\left(\mathbf{y}[m] - \widehat{\mathbf{y}}[m|m-1]\right), \tag{10}$$

where $\mathbf{W}[m]$ denotes the Kalman gain matrix for the mth TS, which is given as

$$\mathbf{W}[m] = \mathbf{E}[m|m-1]\left[\sum_{k=1}^{K}\mathbf{A}_k^H\mathbf{P}_k^H[m]\mathbf{H}_k^H\right]\left[\sum_{k=1}^{K}\mathbf{H}_k\mathbf{P}_k[m]\mathbf{A}_k\mathbf{E}[m|m-1]\mathbf{A}_k^H\mathbf{P}_k^H[m]\mathbf{H}_k^H + \mathbf{H}_k\mathbf{P}_k[m]\mathbf{R}_k\mathbf{P}_k^H[m]\mathbf{H}_k^H + \mathbf{R}_{\text{FC}}\right]^{-1}. \tag{11}$$

Furthermore, upon defining $\widetilde{\boldsymbol{\theta}}[m|m] = \widehat{\boldsymbol{\theta}}[m|m] - \boldsymbol{\theta}[m]$, the error covariance matrix $\mathbf{E}[m|m]$ is determined as shown in (12). From (11) and (12), it can be readily observed that MMSE-optimal KF-based tracking requires the knowledge of the unknown hyperparameter matrix $\boldsymbol{\Gamma}[m]$ and suitable TPC $\mathbf{P}_k[m]$ for each sensor. A procedure for the joint design of these two quantities is described next.

The received vector in (3) can also be equivalently written as

$$\mathbf{y}[m] = \widetilde{\mathbf{H}}[m]\boldsymbol{\theta}[m] + \widetilde{\mathbf{v}}[m], \tag{13}$$

where the quantities $\widetilde{\mathbf{H}}[m] \in \mathbb{C}^{N_{\text{FC}}\times q}$, $\widetilde{\mathbf{v}}[m] \sim \mathcal{CN}(\mathbf{0}, \widetilde{\mathbf{R}}_v[m]) \in \mathbb{C}^{N_{\text{FC}}\times 1}$ and the covariance matrix $\widetilde{\mathbf{R}}_v[m] \in \mathbb{C}^{N_{\text{FC}}\times N_{\text{FC}}}$ are defined as

$$\widetilde{\mathbf{H}}[m] = \sum_{k=1}^{K}\mathbf{H}_k\mathbf{P}_k[m]\mathbf{A}_k, \tag{14}$$

$$\widetilde{\mathbf{v}}[m] = \sum_{k=1}^{K}\mathbf{H}_k\mathbf{P}_k[m]\mathbf{v}_k[m] + \mathbf{v}_{\text{FC}}[m], \tag{15}$$

$$\widetilde{\mathbf{R}}_v[m] = \sum_{k=1}^{K}\mathbf{H}_k\mathbf{P}_k[m]\mathbf{R}_k\mathbf{P}_k^H[m]\mathbf{H}_k^H + \mathbf{R}_{\text{FC}}. \tag{16}$$

Due to the Gaussian nature of (13) and (8), the *a posteriori* probability density function (pdf) $p\left(\boldsymbol{\theta}[m]|\mathbf{y}[m];\boldsymbol{\Gamma}[m]\right)$ is also Gaussian in nature and its mean vector and covariance matrix can be expressed in the closed-form of:

$$\boldsymbol{\mu}_\theta[m] = \boldsymbol{\Sigma}_\theta[m]\widetilde{\mathbf{H}}^H[m]\widetilde{\mathbf{R}}_v^{-1}[m]\mathbf{y}[m], \tag{17}$$

$$\boldsymbol{\Sigma}_\theta[m] = \left[\widetilde{\mathbf{H}}^H[m]\widetilde{\mathbf{R}}_v^{-1}[m]\widetilde{\mathbf{H}}[m] + \boldsymbol{\Gamma}^{-1}[m]\right]^{-1}. \tag{18}$$

In the proposed SBL-KF approach, the hyperparameter matrix $\boldsymbol{\Gamma}[m]$ is obtained so that the Bayesian evidence $p(\mathbf{y}[m];\boldsymbol{\Gamma}[m])$, given by

$$\begin{aligned}\log\ p(\mathbf{y}[m];\boldsymbol{\Gamma}[m]) &= -\frac{N_{\text{FC}}}{2}\log(2\pi) - \frac{1}{2}\log\left(\det\left(\boldsymbol{\Sigma}_y[m]\right)\right) \\ &\quad - \mathbf{y}^H[m]\boldsymbol{\Sigma}_y^{-1}[m]\mathbf{y}[m], \\ &\propto -\log\left(\det\left(\boldsymbol{\Sigma}_y[m]\right)\right) \\ &\quad - \mathbf{y}^H[m]\boldsymbol{\Sigma}_y^{-1}[m]\mathbf{y}[m],\end{aligned} \tag{19}$$

is maximized, where $\boldsymbol{\Sigma}_y[m] = \left(\widetilde{\mathbf{H}}[m]\boldsymbol{\Gamma}[m]\widetilde{\mathbf{H}}^H[m] + \widetilde{\mathbf{R}}_v[m]\right) \in \mathbb{C}^{N_{\text{FC}}\times N_{\text{FC}}}$. It can be readily observed that the above optimization objective is non-convex, which renders the problem intractable. We can now harness the EM framework for iteratively maximizing the Bayesian evidence with respect to the hyperparameters $\gamma_i[m]$, $1 \leq i \leq q$. Upon convergence of the SBL algorithm, the hyperparameters corresponding to the sparse locations of the parameter vector $\boldsymbol{\theta}[m]$ tend to zero, leading to a sparse parameter estimate. The key steps of the EM algorithm are described next.

The EM algorithm initially formulates the log-likelihood of the complete information set $\{\mathbf{y}[m], \boldsymbol{\theta}[m]\}$, where $\mathbf{y}[m]$ is the observed output and $\theta[m]$ is the missing information. This is given as $\mathcal{L}\left(\boldsymbol{\Gamma}[m]|\widehat{\boldsymbol{\Gamma}}[m-1]\right)$, where $\widehat{\boldsymbol{\Gamma}}[m-1]$ denotes the estimate of the hyperparameter matrix $\boldsymbol{\Gamma}$ obtained in the

$(m-1)$st TS. In the E-step, the quantity $\mathcal{L}\left(\boldsymbol{\Gamma}[m]|\widehat{\boldsymbol{\Gamma}}[m-1]\right)$ is evaluated as

$$\begin{aligned}
&\mathcal{L}\left(\boldsymbol{\Gamma} \mid \widehat{\boldsymbol{\Gamma}}[m-1]\right)\\
&= \mathbb{E}_{\boldsymbol{\theta}[m]|\mathbf{y}[m];\widehat{\boldsymbol{\Gamma}}[m]}\left[\log\left[p\left(\mathbf{y}[m], \boldsymbol{\theta}[m]; \boldsymbol{\Gamma}[m]\right)\right]\right]\\
&= \mathbb{E}_{\boldsymbol{\theta}[m]|\mathbf{y}[m];\widehat{\boldsymbol{\Gamma}}[m]}\Big[\log\left[p\left(\mathbf{y}[m] \mid \boldsymbol{\theta}[m]\right)\right]\\
&\quad + \log\left[p\left(\boldsymbol{\theta}[m]; \boldsymbol{\Gamma}[m]\right)\right]\Big]. \qquad (20)
\end{aligned}$$

One can simplify the various terms of (20) as discussed next. Note that the quantity $\log\left[p\left(\mathbf{y}[m] \mid \boldsymbol{\theta}[m]\right)\right]$ inside the expectation operator above can be simplified as

$$\begin{aligned}
\log\left[p\left(\mathbf{y}[m] \mid \boldsymbol{\theta}[m]\right)\right] = -\kappa_1 - \left(\mathbf{y}[m] - \widetilde{\mathbf{H}}[m]\boldsymbol{\theta}[m]\right)^H\\
\widetilde{\mathbf{R}}_v^{-1}[m]\left(\mathbf{y}[m] - \widetilde{\mathbf{H}}[m]\boldsymbol{\theta}[m]\right)^H, \quad (21)
\end{aligned}$$

where $\kappa_1 = \log\left[(\pi)^q \det\left(\widetilde{\mathbf{R}}_v[m]\right)\right]$ is a constant which is independent of the hyperparameters. Therefore, the quantity $\log\left[p\left(\mathbf{y}[m] \mid \boldsymbol{\theta}[m]\right)\right]$ is also independent of the hyperparameters, and hence can be ignored in the subsequent maximization step (M-step) with respect to the hyperparameters. On the other hand, the second term inside the expectation in (20) can be simplified as

$$\log\left[p\left(\boldsymbol{\theta}[m]; \boldsymbol{\Gamma}[m]\right)\right] = \sum_{i=1}^{q} -\log(\pi\gamma_i[m]) - \frac{|\boldsymbol{\theta}[m](i)|^2}{\gamma_i[m]}. \quad (22)$$

Furthermore, the *a posteriori* pdf of the parameter $\boldsymbol{\theta}[m]$, which is required for computing the conditional expectation $\mathbb{E}_{\boldsymbol{\theta}[m]|\mathbf{y}[m];\widehat{\boldsymbol{\Gamma}}[m]}[\cdot]$ in (20), can be determined as [50, Sec. 11.3]

$$p\left(\boldsymbol{\theta}[m] \mid \mathbf{y}[m]; \widehat{\boldsymbol{\Gamma}}[m-1]\right) = \mathcal{CN}\left(\boldsymbol{\mu}_\theta[m], \boldsymbol{\Sigma}_\theta[m]\right), \quad (23)$$

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

$$\boldsymbol{\mu}_\theta[m] = \boldsymbol{\Sigma}_\theta[m]\widetilde{\mathbf{H}}^H[m]\widetilde{\mathbf{R}}_v^{-1}[m]\mathbf{y}[m], \quad (24)$$

$$\boldsymbol{\Sigma}_\theta[m] = \left(\left(\widehat{\boldsymbol{\Gamma}}[m-1]\right)^{-1} + \widetilde{\mathbf{H}}^H[m]\widetilde{\mathbf{R}}_v^{-1}[m]\widetilde{\mathbf{H}}[m]\right)^{-1}. \quad (25)$$

Now, upon taking the conditional expectation $\mathbb{E}_{\boldsymbol{\theta}[m]|\mathbf{y}[m];\widehat{\boldsymbol{\Gamma}}[m]}[\cdot]$ in (22), we have

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{\theta}[m]|\mathbf{y}[m];\widehat{\boldsymbol{\Gamma}}[m]}\left\{\log\left[p\left(\boldsymbol{\theta}[m]; \boldsymbol{\Gamma}[m]\right)\right]\right\} = \sum_{i=1}^{q} -\log\left(\pi\gamma_i\right)\\
- \frac{1}{\gamma_i}\mathbb{E}_{\boldsymbol{\theta}[m]|\mathbf{y}[m];\widehat{\boldsymbol{\Gamma}}[m]}\left\{|\theta_i[m]|^2\right\}. \quad (26)
\end{aligned}$$

Finally, the maximization problem in the M-step can be expressed as

$$\begin{aligned}
\widehat{\boldsymbol{\Gamma}}[m] &= \underset{\boldsymbol{\Gamma}}{\operatorname{argmax}}\, \mathbb{E}_{\boldsymbol{\theta}[m]|\mathbf{y}[m];\widehat{\boldsymbol{\Gamma}}[m]}\left\{\log\left[p\left(\boldsymbol{\theta}[m]; \boldsymbol{\Gamma}[m]\right)\right]\right\}\\
&\equiv \underset{\boldsymbol{\Gamma}}{\operatorname{argmax}} - \log(\gamma_i[m]) - \frac{\boldsymbol{\Sigma}_\theta[m](i,i) + |\boldsymbol{\mu}_\theta[m](i)|^2}{\gamma_i[m]}. \quad (27)
\end{aligned}$$

The above maximization problem may then be readily solved by computing the partial derivative with respect to each hyperparameter and setting it equal to zero, which yields the update equation for $\widehat{\gamma}_i[m]$ as

$$\widehat{\gamma}_i[m] = \boldsymbol{\Sigma}_\theta[m](i,i) + \mid \boldsymbol{\mu}_\theta[m](i) \mid^2. \quad (28)$$

Thus, the estimate of the hyperparameter matrix $\boldsymbol{\Gamma}$ becomes $\widehat{\boldsymbol{\Gamma}}[m] = \mathcal{D}\left(\widehat{\gamma}_1[m], \widehat{\gamma}_2[m], ..., \widehat{\gamma}_q[m]\right)$, which is subsequently employed for the TPC/ RC design. The next subsection describes the existing HBKF technique of [51] to track a time varying sparse parameter, which has been employed in our simulations for comparison.

In contrast to the proposed SBL-KF procedure, the HBKF assigns a parameterized Gaussian prior to the innovation $\mathbf{n}[k]$ as

$$p(\mathbf{u}[m]; \boldsymbol{\Xi}[m]) = \prod_{i=1}^{q}\left(\pi\xi_i[m]\right)^{-1}\exp\left(\frac{-|\mathbf{u}[m](i)|^2}{\xi_i[m]}\right), \quad (29)$$

where $\xi_i[m]$, $1 \le i \le q$, denotes the hyperparameters that are to be estimated using the Bayesian learning module. The corresponding hyperparameter matrix is $\boldsymbol{\Xi}[m] = \operatorname{diag}\left[\xi_1[m], \xi_2[m], \ldots, \xi_q[m]\right] \in \mathbb{C}^{q\times q}$. The observation vector $\widehat{\mathbf{y}}[m|m-1]$ predicted at the FC is the same as in (7). Hence, the innovation or observation error vector $\mathbf{y}_e[m]$ is defined as

$$\begin{aligned}
\mathbf{y}_e[m] &= \mathbf{y}[m] - \widehat{\mathbf{y}}[m|m-1]\\
&= \widetilde{\mathbf{H}}[m]\boldsymbol{\theta}[m] + \widetilde{\mathbf{v}}[m] - \widetilde{\mathbf{H}}[m]\boldsymbol{\theta}[m|m-1]\\
&= \rho\widetilde{\mathbf{H}}[m]\left[\boldsymbol{\theta}[m-1] - \widehat{\boldsymbol{\theta}}[m-1|m-1]\right]\\
&\quad + \sqrt{1-\rho^2}\widetilde{\mathbf{H}}[m] + \mathbf{u}[m]\widetilde{\mathbf{v}}[m] \qquad (30)\\
&\approx \sqrt{1-\rho^2}\widetilde{\mathbf{H}}[m]\mathbf{u}[m] + \widetilde{\mathbf{v}}[m]. \qquad (31)
\end{aligned}$$

Equation (30) reduces to (31) under the assumption that the estimation error $\boldsymbol{\theta}[m-1] - \widehat{\boldsymbol{\theta}}[m-1|m-1]$ approaches zero as m increases. Due to the Gaussian nature of noise vectors $\widetilde{\mathbf{v}}[m]$ in (31) and $\mathbf{w}[m]$ in (29), the *a posteriori* pdf $p\left(\mathbf{w}[m]|\mathbf{y}_e[m]; \boldsymbol{\Xi}[m]\right)$ is also Gaussian in nature, whose mean vector and covariance matrix can be evaluated in the following closed-form

$$\boldsymbol{\mu}_u[m] = \sqrt{1-\rho^2}\boldsymbol{\Sigma}_u[m]\widetilde{\mathbf{H}}^H[m]\widetilde{\mathbf{R}}_v^{-1}[m]\mathbf{y}_e[m], \quad (32)$$

$$\boldsymbol{\Sigma}_u[m] = \left[\left(1-\rho^2\right)\widetilde{\mathbf{H}}^H[m]\widetilde{\mathbf{R}}_u^{-1}[m]\widetilde{\mathbf{H}}[m] + \boldsymbol{\Xi}^{-1}[m]\right]^{-1}. \quad (33)$$

Following a procedure similar to the one described for the SBL-KF, the EM-based update of the hyperparameter $\widehat{\xi}_i^{(m)}$ is expressed as

$$\widehat{\xi}_i(m) = \boldsymbol{\Sigma}_u[m](i,i) + \mid \boldsymbol{\mu}_u[m](i) \mid^2, \quad (34)$$

which can be subsequently employed for TPC/RC design in the HBKF-based framework. The detailed procedure of designing the transceiver for MSE minimization in each TS is described in the next subsection. Although, we derive it using the hyperparameter matrix $\boldsymbol{\Gamma}[m]$ obtained from the SBL-KF technique, the formulation is identical for the HBKF framework.

B. Transceiver design procedure

Since the hyperparameter matrix $\boldsymbol{\Gamma}[m]$ of the mth TS is evaluated using the procedure described in the previous subsection, we now exploit this knowledge for designing the transceiver for per time slot MSE minimization, which is formulated as:

$$\begin{aligned}\underset{\mathbf{W}[m],\{\mathbf{P}_k[m]\}_{k=1}^{K}}{\text{minimize}} \quad & \text{Tr}\left[\mathbf{E}[m|m]\right] \\ \text{subject to} \quad & \text{Tr}\left[\mathbf{P}_k[m]\left[\mathbf{A}_k\boldsymbol{\Gamma}[m]\mathbf{A}_k^H+\mathbf{R}_k\right]\mathbf{P}_k^H[m]\right] \\ & \leq \zeta_k[m],\ 1\leq k\leq K. \end{aligned} \tag{35}$$

From the expression of $[\mathbf{E}[m|m]]$ in (12), the optimization objective is coupled in terms of the optimization variables $\mathbf{W}[m], \{\mathbf{P}_k[m]\}_{k=1}^{K}$, thus making the above problem non-convex in nature [15], [18] and hence difficult to solve. To overcome this challenge, one can leverage the BCD-based framework wherein the RC and TPC matrices are computed in an iterative fashion for exploiting the fact that the above optimization problem is convex for the computation of a single matrix at a given time, when the other optimization matrix variables are known. Since it is required to run only a single iteration of the iterative BCD algorithm per TS, it is termed the FBCD algorithm. When all the TPC matrices $\{\mathbf{P}_k[m]\}_{k=1}^{K}$ are known, the optimization problem in (35) reduces to an unconstrained quadratic optimization problem, which can be solved efficiently by differentiating the objective in (35) with respect to $\mathbf{W}[m]$ and setting it equal to zero. As a result, the expression for the MSE-optimal RC matrix is obtained as the Kalman gain matrix $\mathbf{W}[m]$ given in (11). The per slot MSE optimal TPC for each SN k can now be determined using the KKT framework [52, Sec. 5.5.3]. The Lagrangian function for the problem under consideration is given in (36), where the scalar $\lambda_k[m]\geq 0$ represents the Lagrange multiplier corresponding to the kth SN's power constraint. Using the first order optimality condition, the MSE optimal TPC matrix is obtained as

$$\begin{aligned}\mathbf{P}_k[m] = & \Big[\mathbf{H}_k^H\mathbf{W}^H[m]\mathbf{W}[m]\mathbf{H}_k+\lambda_k[m]\mathbf{I}\Big]^{-1}\Bigg[\mathbf{H}_k\mathbf{P}_k^H[m] \\ & \mathbf{E}[m|m-1]\mathbf{A}_k^H-\sum_{j=1}^{k-1}\mathbf{H}_j\mathbf{W}^H[m]\mathbf{W}[m]\mathbf{H}_j^H\mathbf{P}_j[m]\Big(\mathbf{A}_k \\ & \mathbf{E}[m|m-1]\mathbf{A}_j^H\Big)^H-\sum_{j=k+1}^{K}\mathbf{H}_j\mathbf{W}^H[m]\mathbf{W}[m]\mathbf{H}_j^H \\ & \mathbf{P}_j[m-1]\Big(\mathbf{A}_k\mathbf{E}[m|m-1]\mathbf{A}_j^H\Big)^H\Bigg]\Big[\mathbf{A}_k\boldsymbol{\Gamma}[m]\mathbf{A}_k^H+\mathbf{R}_k\Big]^{-1}. \end{aligned} \tag{37}$$

Furthermore, the Lagrange multiplier $\lambda_k[m]$, for each SN k is calculated using the KKT framework [Sec. 5.5.3] [52], and its procedure is described in Appendix A in detail. Algorithm 1 summarizes the proposed SBL-KF and FBCD-based time-varying sparse parameter tracking algorithm, while its graphical representation is seen in Fig. 2.

The proposed SBL-KF and FBCD-based time-varying sparse parameter tracking algorithm is initialized as follows.

Algorithm 1 SBL-KF and FBCD based time varying sparse parameter tracking

---

Input: $\{\mathbf{y}[m],\widetilde{\mathbf{R}}_v[m]\}\ \forall m$, $\{\mathbf{H}_k\}_{k=1}^{K}$, $\{\mathbf{A}_k\}_{k=1}^{K}$.
Initialization: Precoders $\{\mathbf{P}_k[0]\}_{k=1}^{K}\sim\mathcal{CN}(\mathbf{0},\mathbf{I})$, $\widehat{\boldsymbol{\theta}}[-1|-1]=\mathbf{0}_{q\times 1}$, $\boldsymbol{\Gamma}[-1]$ obtained using SBL on $\mathbf{y}[0]$, $\mathbf{E}[-1|-1]=\boldsymbol{\Gamma}[-1]$, set time slot index $m=0$
for $m=0,1,2,\cdots$ **do**
  E-Step: Evaluate $\boldsymbol{\mu}_\theta[m]$ and $\boldsymbol{\Sigma}_\theta[m]$ using (24) and (25)
  M-Step: Evaluate $\widehat{\gamma}_j[m]$ using (28)
  Use (6), (7) and (9) for prediction
  Compute the RC matrix $\mathbf{W}[m]$ using (11)
  Update the filtered estimated $\widehat{\boldsymbol{\theta}}[m|m]$ using (10)
  Compute the TPC $\mathbf{P}_k[m+1],\forall k$ using (37)
end for
Output: Filtered estimate $\widehat{\boldsymbol{\theta}}[m|m]$

The elements of the initial precoders for each SN k, denoted by $\mathbf{P}_k[0]$, are initialized randomly as i.i.d. $\mathcal{CN}(0,1)$ random variables. The initial estimate $\widehat{\boldsymbol{\theta}}[-1|-1]$ of the sparse parameter vector is set to $\mathbf{0}_{q\times 1}$. Furthermore, the initial estimate of the hyperparameter matrix $\boldsymbol{\Gamma}[-1]$ is obtained by running the SBL procedure using the initial observation $\mathbf{y}[0]$. Additionally, our proposed framework initializes the error covariance matrix $\mathbf{E}[-1|-1]$ with the hyperparameter matrix $\boldsymbol{\Gamma}[-1]$, which significantly reduces the MSE for the initial TSs, as observed in our simulation results in Section-V. The next theorem presents the centralized BCRB benchmark for the proposed LDE scheme for a time-varying sparse parameter.

Theorem 1. *Considering a centralized scenario, where all the observations are available at the FC without any distortion, the BCRB for the MSE of the proposed SBL-KF based LDE scheme is formulated as*

$$\text{MSE}_{\text{BCRB}}\triangleq\mathbb{E}\left[||\widehat{\boldsymbol{\theta}}[m]-\boldsymbol{\theta}[m]||_2^2\right]\geq\text{Tr}\left[\mathbf{B}^{-1}[m]\right], \tag{38}$$

where the Bayesian Fisher information matrix (BFIM) $\mathbf{B}[m]\in\mathbb{C}^{q\times q}$ *can be computed recursively as*

$$\mathbf{B}[m]=\Big(\rho^2\mathbf{B}^{-1}[m-1]+(1-\rho^2)\boldsymbol{\Gamma}[m]\Big)^{-1}+\mathbf{A}^H\mathbf{R}^{-1}\mathbf{A}. \tag{39}$$

The stacked observation matrix $\mathbf{A}\in\mathbb{C}^{lK\times q}$ *and the measurement noise covariance matrix* $\mathbf{R}\in\mathbb{C}^{lK\times lK}$ *are given as*

$$\mathbf{A}=\left[\mathbf{A}_1^T,\mathbf{A}_2^T,\ldots,\mathbf{A}_S^T\right]^T, \tag{40}$$

$$\mathbf{R}=\mathcal{D}\left(\mathbf{R}_1,\mathbf{R}_2,\cdots,\mathbf{R}_S\right). \tag{41}$$

Proof. Given in Appendix B. □

The next section develops robust transceiver designs for the LDE of a time-varying sparse parameter in the presence of imperfect CSI.

IV. Robust Transceiver Design

In practical scenarios, it is challenging to obtain perfect knowledge of the CSI between each SN and the FC due to several factors, such as the limited pilot overhead, limited

$$\mathcal{L}\left[\{\mathbf{P}_k[m]\}_{k=1}^{K}, \{\lambda_s[m]\}_{k=1}^{K}\right] = \text{Tr}\Bigg[\mathbf{E}[m|m-1] - \mathbf{E}[m|m-1]\left[\sum_{k=1}^{K}\mathbf{A}_k^H\mathbf{P}_k^H[m]\mathbf{H}_k^H\right]\mathbf{W}^H[m] - \mathbf{W}[m] \left[\sum_{k=1}^{K}\mathbf{H}_k\mathbf{P}_k[m]\mathbf{A}_k\right]\mathbf{E}[m|m-1] + \mathbf{W}[m]\Bigg[\mathbf{H}_k\mathbf{P}_k[m]\mathbf{A}_k\mathbf{E}[m|m-1]\mathbf{A}_k^H\mathbf{P}_k^H[m]\mathbf{H}_k^H + \sum_{j=1,l\neq j}^{K}\mathbf{H}_k\mathbf{P}_k[m]\mathbf{A}_k \mathbf{E}[m|m-1]\mathbf{A}_j^H\mathbf{P}_j^H[m]\mathbf{H}_j^H + \sum_{k=1}^{K}\mathbf{H}_k\mathbf{P}_k[m]\mathbf{R}_k\mathbf{P}_k^H[m]\mathbf{H}_k^H + \mathbf{R}_{\text{FC}}\Bigg]\mathbf{W}^H[m]\Bigg] + \sum_{k=1}^{K}\Bigg[\lambda_k[m]\Big[\text{Tr}\Big[\mathbf{P}_k[m]\Big(\mathbf{A}_k \mathbf{\Gamma}[m]\mathbf{A}_k^H + \mathbf{R}_k\Big)\mathbf{P}_k^H[m]\Big] - \zeta_k[m]\Big]\Bigg], \tag{36}$$

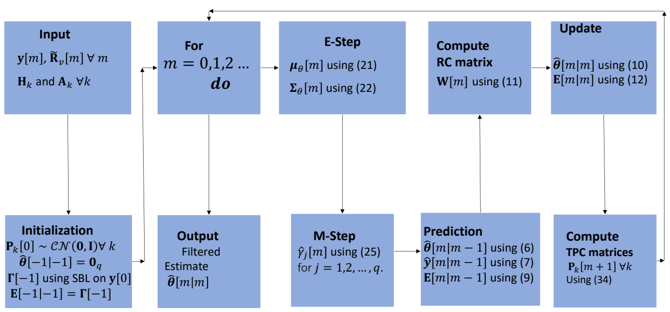

Fig. 2: Graphical representation of algorithm 1.

feedback, quantization error etc. Hence, in a practical system, it is imperative to take the CSI uncertainty into account for achieving robust parameter tracking. To this end, similar to [49], [53], let the channel between each SN and the FC be modeled as

$$\mathbf{H}_k = \widehat{\mathbf{H}}_k + \Delta\mathbf{H}_k, \tag{42}$$

where $\widehat{\mathbf{H}}_k$ denotes the available channel estimate and $\Delta\mathbf{H}_k$ represents the estimation error matrix, whose elements are assumed to be distributed as $\mathcal{CN}\left(0, \sigma_H^2\right)$. The result given in the following lemma will be used in this section.

Lemma 1. *Consider a matrix* $\mathbf{X} = \widehat{\mathbf{X}} + \Delta\mathbf{X} \in \mathbb{C}^{r\times t}$*, where the matrix* $\widehat{\mathbf{X}}$ *is is fixed, while the elements of the matrix* $\Delta\mathbf{X}$ *are distributed as* $\mathcal{CN}(0, \sigma^2)$*. Then, for any arbitrary matrix* $\mathbf{Z}$ *of appropriate dimension, it follows that [49]*

$$\mathbb{E}\left[\mathbf{X}\mathbf{Z}\mathbf{Z}^H\mathbf{X}^H\right] = \widehat{\mathbf{X}}\mathbf{Z}\mathbf{Z}^H\widehat{\mathbf{X}}^H + \sigma_H^2\text{Tr}\left[\mathbf{Z}\mathbf{Z}^H\right]\mathbf{I}_r. \tag{43}$$

Next, we discuss the RSBL-KF for parameter tracking relying on imperfect CSI.

A. RSBL-KF

Upon considering the scenario of realistic CSI uncertainty, during the mth TS, the prediction equation of the sparse parameter vector $\boldsymbol{\theta}[m]$ is identical to (6), while the prediction for the received vector $\mathbf{y}[m]$, defined as $\widehat{\mathbf{y}}_{\text{R}}[m|m-1] = \mathbb{E}\left[\mathbf{y}[m]|\mathbf{\Psi}_{m-1}\right]$, can be formulated as

$$\widehat{\mathbf{y}}_{\text{R}}[m|m-1] = \mathbb{E}_{\Delta\mathbf{H}_k}\left[\sum_{k=1}^{K}\mathbf{H}_k\mathbf{P}_k[m]\mathbf{A}_k\widehat{\boldsymbol{\theta}}[m|m-1]\right] = \sum_{k=1}^{K}\widehat{\mathbf{H}}_k\mathbf{P}_k[m]\mathbf{A}_k\widehat{\boldsymbol{\theta}}[m|m-1]. \tag{44}$$

Substituting the expression of $\mathbf{H}_k$ from (42) into (12), followed by employing the result given in Lemma 1, one obtains the following expression for the average error covariance matrix defined as $\bar{\mathbf{E}}[m|m] = \mathbb{E}_{\Delta\mathbf{H}_k}\left[\mathbf{E}[m|m]\right]$:

$$\bar{\mathbf{E}}[m|m] = \mathbf{E}[m|m-1] - \mathbf{E}[m|m-1]\left[\sum_{k=1}^{K}\mathbf{A}_k^H\mathbf{P}_k^H[m]\widehat{\mathbf{H}}_k^H\right] \mathbf{W}^H[m] - \mathbf{W}[m]\left[\sum_{k=1}^{K}\widehat{\mathbf{H}}_k\mathbf{P}_k[m]\mathbf{A}_k\right]\mathbf{E}[m|m-1] + \mathbf{W}[m] \Bigg[\widehat{\mathbf{H}}_k\mathbf{P}_k[m]\mathbf{A}_k\mathbf{E}[m|m-1]\mathbf{A}_k^H\mathbf{P}_k^H[m]\widehat{\mathbf{H}}_k^H + \sigma_H^2\text{Tr}\Big[\mathbf{P}_k[m]\mathbf{A}_k \mathbf{E}[m|m-1]\mathbf{A}_k^H\mathbf{P}_k^H[m]\Big] + \sum_{j=1,s\neq j}^{K}\widehat{\mathbf{H}}_k\mathbf{P}_k[m]\mathbf{A}_k\mathbf{E}[m|m-1] \mathbf{A}_j^H\mathbf{P}_j^H[m]\widehat{\mathbf{H}}_j^H + \sum_{k=1}^{K}\widehat{\mathbf{H}}_k\mathbf{P}_k[m]\mathbf{R}_k\mathbf{P}_k^H[m]\widehat{\mathbf{H}}_k^H + \sigma_H^2 \text{Tr}\big[\mathbf{P}_k[m]\mathbf{R}_k\mathbf{P}_k^H[m]\big] + \mathbf{R}_{\text{FC}}\Bigg]\mathbf{W}^H[m]. \tag{45}$$

Once again, invoking the FBCD framework while considering the TPC matrices $\{\mathbf{P}_k[m]\}_{k=1}^{K}$ to be known, the optimization problem of determining the robust RC matrix for minimizing the average MSE per TS is given by

$$\underset{\mathbf{W}[m]}{\text{minimize}} \quad \text{Tr}\left[\bar{\mathbf{E}}[m|m]\right]. \tag{46}$$

Observe from (45), that the above optimization problem is an unconstrained convex quadratic cost minimization problem in terms of the RC matrix $\mathbf{W}[m]$. The robust RC matrix for the mth TS is obtained by differentiating the aforementioned objective function with respect to $\mathbf{W}[m]$ and equating it to zero, yielding

$$\mathbf{W}[m] = \mathbf{E}[m|m-1]\left[\sum_{k=1}^{K}\mathbf{A}_k^H\mathbf{P}_k^H[m]\widehat{\mathbf{H}}_k^H\right]\left[\sum_{k=1}^{K}\widehat{\mathbf{H}}_k\mathbf{P}_k[m]\mathbf{A}_k\mathbf{E}[m|m-1]\mathbf{A}_k^H\mathbf{P}_k^H[m]\widehat{\mathbf{H}}_k^H + \sigma_H^2\text{Tr}\Big[\mathbf{P}_k[m]\mathbf{A}_k\mathbf{E}[m|m-1]\mathbf{A}_k^H\mathbf{P}_k^H[m]\Big] + \widehat{\mathbf{H}}_k\mathbf{P}_k[m]\mathbf{R}_k\mathbf{P}_k^H[m]\widehat{\mathbf{H}}_k^H + \sigma_H^2\text{Tr}\Big[\mathbf{P}_k[m]\mathbf{R}_k\mathbf{P}_k^H[m]\Big] + \mathbf{R}_{\text{FC}}\right]^{-1}. \quad (47)$$

Once the robust RC matrix is obtained, the optimization problem of finding the TPC matrix corresponding to the kth SN with an objective to minimize the MSE for the mth TS can be formulated as

$$\begin{aligned}\underset{\mathbf{P}_k[m]}{\text{minimize}} \quad & \text{Tr}\left[\bar{\mathbf{E}}[m|m]\right] \\ \text{subject to} \quad & \text{Tr}\left[\mathbf{P}_k[m]\left[\mathbf{A}_k\mathbf{\Gamma}[m]\mathbf{A}_k^H + \mathbf{R}_k\right]\mathbf{P}_k^H[m]\right] \leq \zeta_k[m].\end{aligned} \quad (48)$$

The above optimization problem can be solved by following similar lines to (35), and the MSE-optimal robust TPC matrix for the kth SN at mth TS is given as

$$\mathbf{P}_k[m] = \Big[\widehat{\mathbf{H}}_k^H\mathbf{W}^H[m]\mathbf{W}[m]\widehat{\mathbf{H}}_k + \sigma_H^2\text{Tr}\left[\mathbf{W}^H[m]\mathbf{W}[m]\right] + \lambda_k[m]\mathbf{I}\Big]^{-1}\Bigg[\mathbf{H}_k\mathbf{P}_k^H[m]\mathbf{E}[m|m-1]\mathbf{A}_k^H - \sum_{j=1}^{k-1}\mathbf{H}_j\mathbf{W}^H[m]\mathbf{W}[m]\mathbf{H}_j^H\mathbf{P}_j[m]\left(\mathbf{A}_k\mathbf{E}[m|m-1]\mathbf{A}_j^H\right)^H - \sum_{j=k+1}^{K}\mathbf{H}_j\mathbf{W}^H[m]\mathbf{W}[m]\mathbf{H}_j^H\mathbf{P}_j[m-1]\left(\mathbf{A}_k\mathbf{E}[m|m-1]\mathbf{A}_j^H\right)^H\Bigg]\Big[\mathbf{A}_k\mathbf{\Gamma}[m]\mathbf{A}_k^H + \mathbf{R}_k\Big]^{-1}, \quad (49)$$

where the optimal value of the Lagrange multiplier $\lambda_k[m]$ can once again be obtained using the procedure described in Appendix-A. The next section presents our simulation results characterizing the performance of the proposed SBL-KF and FBCD-based LDE schemes for time-varying sparse parameter estimation.

V. Simulation Results

The number of SNs in the MIMO WSN is set to $K = 10$. The number of observations taken by each SN is $l = 5$, the number of TAs at each SN is $N_s = 5$, while the number of receive antennas at the FC is considered to be $N_{\text{FC}} = 5$. The dimension of the unknown time varying sparse parameter $\boldsymbol{\theta}$ is $q = 10$ with the number of non-zero locations equal to 2. Without loss of generality, the observation noise covariance matrix $\mathbf{R}_k$ for each SN k and the FC noise covariance matrix $\mathbf{R}_{\text{FC}}$ are considered as $\mathbf{R}_s = \sigma_s^2\mathbf{I}_l$ and $\mathbf{R}_{\text{FC}} = \sigma_f^2\mathbf{I}_{\text{FC}}$, where the quantities σ_k^2 and σ_f^2 denote the variance of the observation and FC noises, respectively. The signal-to-noise ratio (SNR) at the FC is defined as $\text{SNR}_{\text{FC}} = \frac{1}{\sigma_f^2}$. Similarly, the observation

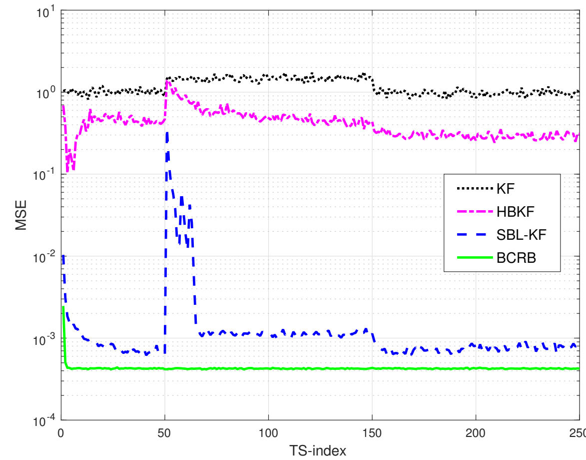

Fig. 3: MSE performance comparison vs. TS-index between the conventional KF, HBKF and the proposed SBL-KF schemes along with BCRB.

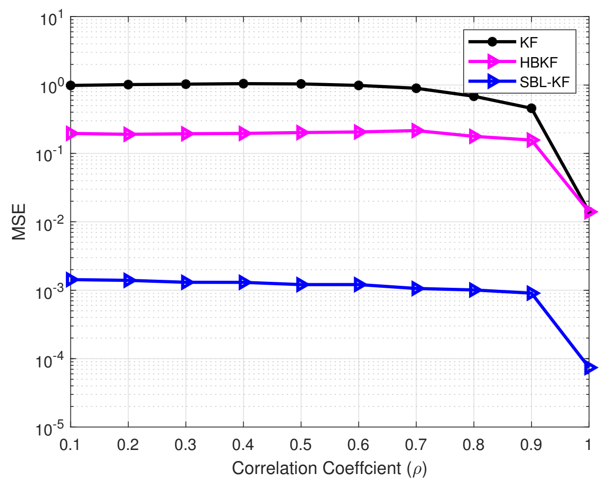

Fig. 4: MSE vs. the correlation coefficient (ρ) for the HBKF and SBL-KF schemes.

SNR denoted by SNR_{OB} is set to $\text{SNR}_{\text{OB}} = \frac{1}{\sigma_k^2}$. Each element of the channel and observation matrices for all the SNs is generated as $\mathcal{CN}(0,1)$.

Fig. 3 plots the MSE as a function of the TS-index for the conventional KF, HBKF and the proposed SBL-KF schemes. The time-varying sparse parameter is generated as follows. For TS $m \leq 49$, the support of the sparse parameter vector is represented by the set $\{2,4\}$, whereas at the 50th TS, the support is expanded to $\{1,2,4\}$. Finally, from $m \geq 150$, the support is set as $\{1,2\}$. Thus, we observe that the sparsity profile of the sparse parameter vector varies slowly with time. As is evident from the figure, the proposed SBL-KF technique with TPC designed using the FBCD procedure capable of successfully tracking the time-varying parameter. Furthermore, the conventional KF framework, which is not designed for sparse parameter estimation, is unable to track it accurately. The tracking performance of the HBKF, which was briefly

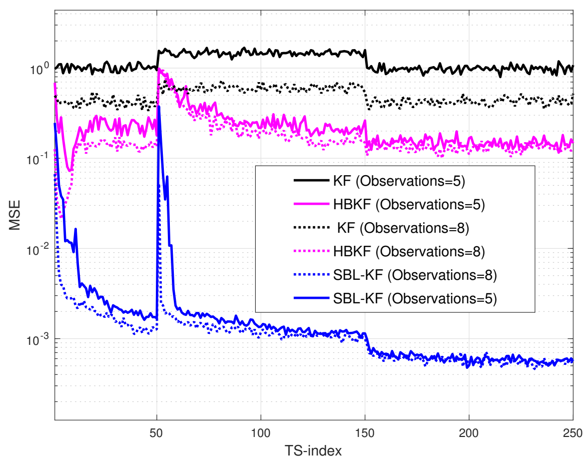

Fig. 5: MSE vs. TS-index for $N_{\text{FC}} = \{5, 8\}$ observations at the FC for the HBKF and the proposed SBL-KF schemes.

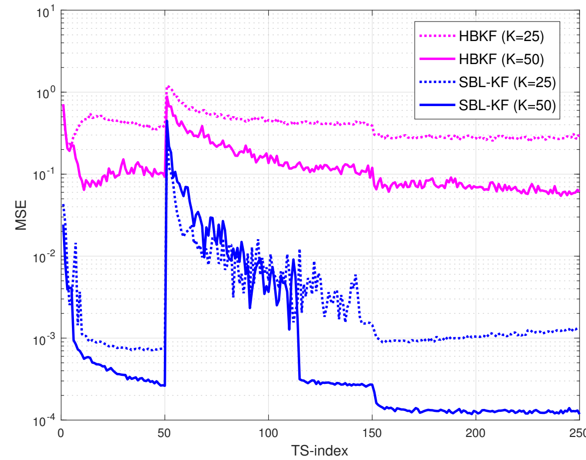

Fig. 6: MSE vs. TS-index for $K = 25$ and 50 SNs in the WSN for perfect CSI.

described in Section-III-A, is also shown for comparison. It is visually clear that the performance of HBKF is poor in comparison to that of the proposed SBL-KF framework, which arises due to the approximation made in Eq. (31) for the HBKF procedure. Furthermore, observe that the initial MSE obtained using the proposed SBL-KF framework is significantly lower than that of the HBKF. This is due to the fact that the SBL-KF exploits the initialization procedure described at the end of Section-III-B, which leads to its improved parameter tracking performance. Finally, it can also be observed that the proposed SBL-KF scheme performs very close to the recursive BCRB that corresponds to the best estimation performance that any LDE scheme can achieve. Hence, it demonstrates the efficacy of the proposed LDE scheme.

Fig. 4 illustrates the MSE performance of the competing schemes by varying the temporal correlation coefficient ρ of the sparse parameter vector. Observe that the performance of the proposed SBL-KF technique is consistent right across the entire range of correlation coefficients ρ. On the other hand, upon increasing ρ, the MSE of the state-of-the-art HBKF technique approaches that obtained using the conventional KF, which implies that the former cannot exploit the sparsity, when the temporal correlation is high. This is due to the fact that as $\rho \longrightarrow 1$, it can inferred from (32) and (33) that the mean $\boldsymbol{\mu}_u[m] \longrightarrow \mathbf{0}$ and $\boldsymbol{\Sigma}_u[m] \longrightarrow \mathbf{I}$, which leads to inaccurate hyperparameter estimation in (34), ultimately resulting its poor tracking performance.

Fig. 5 depicts the MSE versus TS-index for $N_{\text{FC}} = \{5, 8\}$ observations at the FC, which equals the number of FC antennas N_{FC}. When $N_{\text{FC}} < q$, the system is underdetermined, since the number of observations is lower than the size of the unknown parameter, which makes it challenging to estimate the unknown quantity. Thus, the conventional estimation techniques fail in this scenario. Interestingly however, since the unknown parameter is sparse in nature, the SBL technique that proactively exploits the sparsity is still able to successfully estimate the underlying parameter even in such an underdetermined system. Furthermore, the MSE performance improves upon increasing number of observations, a trend that is along expected lines.

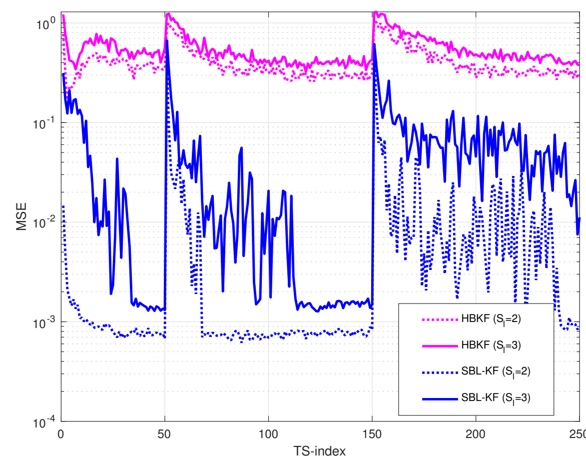

Fig. 7: MSE vs. TS-index parametrized by the sparsity profile.

Fig. 6 shows the MSE vs. TS-index for $K \in \{25, 50\}$ SNs in the WSN. As time elapses the tracking performance improves, which reinforces the trend seen in the previous figures. Additionally, increasing the number of SNs in the system leads to more observations becoming available at the FC, which ultimately results in the MSE reduction witnessed. Furthermore, in Fig. 7, the MSE performance is plotted for values of the cardinality of the support $S_l \in \{2, 3\}$ of the parameter $\boldsymbol{\theta}$ as a function of the TS-index. When $S_l = 3$, for TSs $m \leq 49$, only the $\{1, 2, 4\}$th elements of $\boldsymbol{\theta}$ are non-zero. For TSs $50 \leq m \leq 149$, the support of the parameter vector is given by the set $\{2, 4, 6\}$. For TSs $m \geq 150$, the support changes to $\{1, 4, 6\}$. Similarly, for the scenario with $S_l = 2$, the number of non-zero elements at any given TS is consistently equal to 2, with the support changing at TS

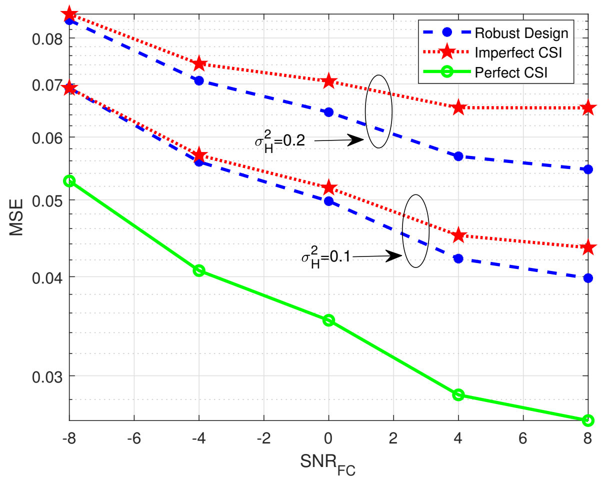

Fig. 8: MSE vs. the $\mathrm{SNR_{FC}}$ for different values of the channel uncertainty variance σ_H^2.

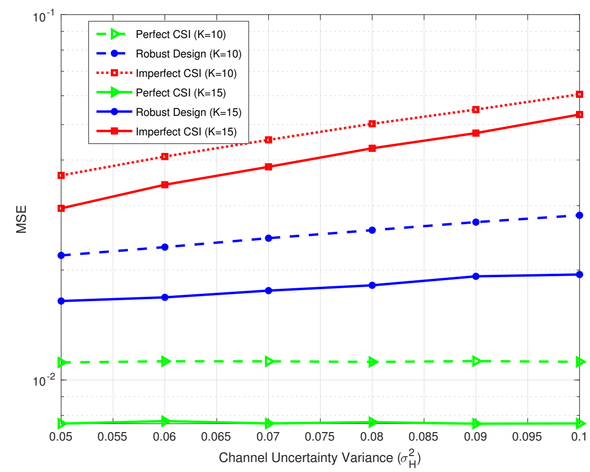

Fig. 9: MSE vs. the channel uncertainty variance σ_H^2 for different number of SNs K.

$m = 50, 150$. It can be observed that as S_l increases, the MSE performance deteriorates, again a performance trend that is along expected lines.

Fig. 8 depicts the MSE performance at TS $m = 15$ of the RSBL-KF scheme proposed in Section-IV for scenarios having CSI uncertainty, versus $\mathrm{SNR_{FC}}$. The values of the CSI uncertainty variance σ_H^2 are set to $\{0.1, 0.2\}$. The imperfect CSI plot corresponds to the system wherein the TPCs and RC, $\{\mathbf{P}_k[m]\}_{k=1}^K$ and $\mathbf{W}[m]$, respectively, are designed using the available channel estimates, while ignoring the CSI error. Finally, the perfect CSI performance corresponds to a scenario with no channel estimation error and serves as the best-case MSE performance bound of the preceding two systems. It is clear that the RSBL-KF attains a better MSE performance than the imperfect CSI-based design, which ignores the CSI uncertainty. Hence, the latter scheme exhibits poor estimation performance. Furthermore, upon increasing $\mathrm{SNR_{FC}}$, the MSE difference between the robust and uncertainty-agnostic designs becomes significant, demonstrating the large performance gains realised by utilising the robust design.

The MSE of the proposed RSBL-KF scheme is shown in Fig. 9 as a function of the channel uncertainty variance σ_H^2 for $K \in \{10, 15\}$ SNs. Upon increasing σ_H^2, the MSE performance of the uncertainty-agnostic design degrades dramatically, but that of the robust design conceived remains rather consistent. This clearly demonstrates the benefit of using the robust sparse parameter estimate approach in real-world scenarios in the face of realistic CSI uncertainty. The MSE performance of the ideal CSI-based design is displayed again as a performance reference for the other estimators, which is flat since it is independent of σ_H^2. It can also be observed that deploying more SNs in the WSN results in an even higher improvement in the MSE of decentralized estimation.

VI. Conclusion

A novel SBL-KF and FBCD-based LDE scheme was conceived for tracking a time-varying sparse parameter considering also a practical scenario having realistic CSI uncertainty. The proposed framework yields a significantly improved estimation performance over the conventional KF, which does not guarantee a sparse estimate. Additionally, the proposed SBL-KF and FBCD-based LDE scheme outperforms the existing HBKF framework for sparse time-varying parameter estimation. Along similar lines, the RSBL-KF scheme is also derived for a scenario associated with imperfect CSI, which was shown to yield an improved MSE performance in comparison to the uncertainty-agnostic design that does not account for the CSI uncertainty. Our future research may explore the problem of LDE of a time-varying sparse parameter relying on quantized measurement transmission, which can in turn lead to a significant reduction of the bandwidth resources.

Appendix A
Lagrange Multiplier ($\lambda_k[m]$) Calculation for Each SN k of the WSN

This Appendix describes the procedure of determining the Lagrange multiplier $\lambda_k[m]$ of each SN k for the TPC design problem in (36) of Section-III relying on perfect CSI. Let the matrices $\mathbf{D}_k[m] \in \mathbb{C}^{N_s \times N_s}$ and $\mathbf{F}_k[m] \in \mathbb{C}^{N_s \times q}$ be defined as

$$\mathbf{D}_k[m] = \mathbf{H}_k^H \mathbf{W}^H[m] \mathbf{W}[m] \mathbf{H}_k, \tag{50}$$

$$\begin{aligned} \mathbf{F}_k[m] = &\Bigg[\mathbf{H}_k \mathbf{P}_k^H[m] \mathbf{E}[m|m-1] \mathbf{A}_k^H - \sum_{j=1}^{k-1} \mathbf{H}_j \mathbf{W}^H[m] \\ &\mathbf{W}[m] \mathbf{H}_j^H \mathbf{P}_j[m] \left(\mathbf{A}_k \mathbf{E}[m|m-1] \mathbf{A}_j^H \right)^H - \sum_{j=k+1}^{K} \mathbf{H}_j \mathbf{W}^H[m] \\ &\mathbf{W}[m] \mathbf{H}_j^H \mathbf{P}_j[k-1] \left(\mathbf{A}_k \mathbf{E}[m|m-1] \mathbf{A}_j^H \right)^H \Bigg] \\ &\left[\mathbf{A}_k \mathbf{\Gamma}[m] \mathbf{A}_k^H + \mathbf{R}_k \right]^{-1}. \end{aligned} \tag{51}$$

Hence, the expression for the TPC in (37) for the kth SN of the mth TS can be recast as

$$\mathbf{P}_k[m] = \left[\mathbf{D}_k[m] + \lambda_k[m]\mathbf{I}_{N_s}\right]\mathbf{F}_k[m]. \quad (52)$$

Next, enforcing the complementary slackness condition of the KKT framework, one obtains

$$\lambda_k[m]\Big[\text{Tr}\left[\mathbf{P}_k[m]\left(\mathbf{A}_k\boldsymbol{\Gamma}[m]\mathbf{A}_k^H + \mathbf{R}_k\right)\mathbf{P}_k^H[m]\right] - \zeta_k[m]\Big] = 0. \quad (53)$$

Upon substituting $\mathbf{P}_k[m]$ from (52) into (53) yields

$$\text{Tr}\Bigg[\left[\mathbf{D}_k[m] + \lambda_k[m]\mathbf{I}_{N_s}\right]^{-1}\mathbf{F}_k[m]\left(\mathbf{A}_k\boldsymbol{\Gamma}[m]\mathbf{A}_k^H + \mathbf{R}_k\right) \mathbf{F}_k^H[m]\left[\mathbf{D}_k[m] + \lambda_k[m]\mathbf{I}_{N_s}\right]^{-1}\Bigg] = \zeta_k[m]. \quad (54)$$

Furthermore, replacing the matrix $\mathbf{D}_k[m]$ by its eigenvalue decomposition, which is defined as $\mathbf{D}_k[m] = \mathbf{U}_k[m]\boldsymbol{\Lambda}_k[m]\mathbf{U}_k^H[m]$, the above equation can be recast as

$$\text{Tr}\Bigg[\left[\boldsymbol{\Lambda}_k[m] + \lambda_k[m]\mathbf{I}_{N_s}\right]^{-2} \underbrace{\mathbf{U}_k^H[m]\mathbf{F}_k[m]\left(\mathbf{A}_k\boldsymbol{\Gamma}[m]\mathbf{A}_k^H + \mathbf{R}_k\right)\mathbf{F}_k^H[m]\mathbf{U}_k[m]}_{\boldsymbol{\Sigma}_k[m]}\Bigg] = \zeta_k[m]. \quad (55)$$

It can also be equivalently written as

$$\sum_{i=1}^{N_s}\frac{[\boldsymbol{\Sigma}_k[m]]_{ii}}{\left([\boldsymbol{\Lambda}_k[m]]_{ii} + \lambda_k[m]\right)^2} = \zeta_k[m]. \quad (56)$$

The Lagrange multiplier $\lambda_k[m]$ of the kth SN can be obtained by solving the above equation using the bisection approach.

Appendix B
Proof of Theorem 1

This Appendix provides the detailed proof of Theorem 1. Starting with the centralized scenario, the overall observation vector, denoted by $\mathbf{x}[m] \in \mathbb{C}^{lK\times 1}$, can be obtained by stacking the observation vectors corresponding to each SN k as follows:

$$\mathbf{x}[m] = \left[\mathbf{x}_1^T[m], \mathbf{x}_2^T[m], \ldots, \mathbf{x}_K^T[m]\right]^T = \mathbf{A}\boldsymbol{\theta}[m] + \mathbf{v}[m], \quad (57)$$

where the stacked measurement noise vector $\mathbf{v}[m] \in \mathbb{C}^{lK\times 1}$ is $\mathbf{v}[m] = \left[\mathbf{v}_1^T[m], \mathbf{v}_2^T[m], \ldots, \mathbf{v}_K^T[m]\right]^T$.

Next, the BFIM matrix $\mathbf{B}[m]$ is given as [54]

$$\mathbf{B}[m] = \mathbf{C}_{22}[m] - \mathbf{C}_{21}[m]\Big[\mathbf{B}[m-1] + \mathbf{C}_{11}[m]\Big]^{-1}\mathbf{C}_{12}[m], \quad (58)$$

where the different $(q \times q)$-dimensional matrices above are defined as

$$\mathbf{C}_{11}[m] = -\mathbb{E}\left[\frac{\partial^2\mathcal{L}\left(\boldsymbol{\theta}[m]|\boldsymbol{\theta}[m-1]\right)}{\partial\boldsymbol{\theta}[m-1]\partial\boldsymbol{\theta}^H[k-1]}\right], \quad (59)$$

$$\mathbf{C}_{12}[m] = -\mathbb{E}\left[\frac{\partial^2\mathcal{L}\left(\boldsymbol{\theta}[m]|\boldsymbol{\theta}[m-1]\right)}{\partial\boldsymbol{\theta}[m-1]\partial\boldsymbol{\theta}^H[m]}\right] = \mathbf{C}_{21}^H[m], \quad (60)$$

$$\mathbf{C}_{22}[m] = -\mathbb{E}\left[\frac{\partial^2\mathcal{L}\left(\boldsymbol{\theta}[m]|\boldsymbol{\theta}[m-1]\right)}{\partial\boldsymbol{\theta}[m]\partial\boldsymbol{\theta}^H[m]}\right] - \mathbb{E}\left[\frac{\partial^2\mathcal{L}\left(\mathbf{x}[m]|\boldsymbol{\theta}[m]\right)}{\partial\boldsymbol{\theta}[m]\partial\boldsymbol{\theta}^H[m]}\right]. \quad (61)$$

Using (1) and (57), the log-likelihood functions $\mathcal{L}\left(\boldsymbol{\theta}[m]|\boldsymbol{\theta}[m-1]\right)$ and $\mathcal{L}\left(\mathbf{x}[m]|\boldsymbol{\theta}[m]\right)$, respectively, are defined as

$$\mathcal{L}\left(\boldsymbol{\theta}[m]|\boldsymbol{\theta}[m-1]\right) = k_1 - \frac{\left(\boldsymbol{\theta}[m] - \rho\boldsymbol{\theta}[m-1]\right)^H\boldsymbol{\Gamma}^{-1}[m]\left(\boldsymbol{\theta}[m] - \rho\boldsymbol{\theta}[m-1]\right)}{2(1-\rho^2)}, \quad (62)$$

$$\mathcal{L}\left(\mathbf{x}[m]|\boldsymbol{\theta}[m]\right) = k_2 - \frac{\left(\mathbf{x}[m] - \mathbf{A}\boldsymbol{\theta}[m]\right)^H\mathbf{R}^{-1}[m]\left(\mathbf{x}[m] - \mathbf{A}\boldsymbol{\theta}[m]\right)}{2}, \quad (63)$$

where we have $k_1 = -\frac{1}{2}\log\left[(2\pi)^q(1-\rho^2)\det\left(\boldsymbol{\Gamma}[m]\right)\right]$ and $k_2 = -\frac{1}{2}\log\left[(2\pi)^{lS}\det\left(\mathbf{R}\right)\right]$. Upon substituting $\mathcal{L}\left(\boldsymbol{\theta}[m]|\boldsymbol{\theta}[m-1]\right)$ and $\mathcal{L}\left(\mathbf{x}[m]|\boldsymbol{\theta}[m]\right)$, respectively, from (62) and (63), into (59)-(61), one can determine the expressions of the matrices $\mathbf{C}_{11}[m]$, $\mathbf{C}_{12}[m]$, $\mathbf{C}_{21}[m]$ and $\mathbf{C}_{22}[m]$. Furthermore,upon substituting these matrices in (58) followed by invoking the Woodbury matrix identity [50, Sec. A1.1.3], one obtains the following compact expression for the BFIM

$$\mathbf{B}[m] = \Big(\rho^2\mathbf{B}^{-1}th + \left(1-\rho^2\right)\boldsymbol{\Gamma}[m]\Big)^{-1} + \mathbf{A}^H\mathbf{R}^{-1}\mathbf{A}. \quad (64)$$

Hence, the BCRB for any given TS m is derived as

$$\text{MSE}_{\text{BCRB}} \geq \text{Tr}\left[\mathbf{B}^{-1}[m]\right]. \quad (65)$$

References

[1] M. Zhang and X. Li, "Drone-enabled Internet-of-Things relay for environmental monitoring in remote areas without public networks," *IEEE Internet of Things Journal*, vol. 7, no. 8, pp. 7648–7662, 2020.

[2] S. K. Roy, S. Misra, N. S. Raghuwanshi, and S. K. Das, "AgriSens: IoT-based dynamic irrigation scheduling system for water management of irrigated crops," *IEEE Internet of Things Journal*, vol. 8, no. 6, pp. 5023–5030, 2021.

[3] J.-A. Jiang, J.-C. Wang, H.-S. Wu, C.-H. Lee, C.-Y. Chou, L.-C. Wu, and Y.-C. Yang, "A novel sensor placement strategy for an IoT-based power grid monitoring system," *IEEE Internet of Things Journal*, vol. 7, no. 8, pp. 7773–7782, 2020.

[4] Q. Wang, J. Gu, H. Huang, Y. Zhao, and M. Guizani, "A resource-efficient online target detection system with autonomous drone-assisted IoT," *IEEE Internet of Things Journal*, vol. 9, no. 15, pp. 13 755–13 766, 2022.

[5] Y. Yuan, W. Yi, and W. Choi, "Dynamic sensor scheduling for target tracking in wireless sensor networks with cost minimization objective," *IEEE Internet of Things Journal*, pp. 1–1, 2022.

[6] F. Guo, F. R. Yu, H. Zhang, X. Li, H. Ji, and V. C. M. Leung, "Enabling massive IoT toward 6G: A comprehensive survey," *IEEE Internet of Things Journal*, vol. 8, no. 15, pp. 11 891–11 915, 2021.

[7] 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.

[8] J.-J. Xiao, S. Cui, Z.-Q. Luo, and A. Goldsmith, "Power scheduling of universal decentralized estimation in sensor networks," *IEEE Transactions on Signal Processing*, vol. 54, no. 2, pp. 413–422, 2006.

[9] Z.-Q. Luo, "Universal decentralized estimation in a bandwidth constrained sensor network," *IEEE Transactions on Information Theory*, vol. 51, no. 6, pp. 2210–2219, 2005.

[10] J.-J. Xiao and Z.-Q. Luo, "Universal decentralized detection in a bandwidth-constrained sensor network," *IEEE Transactions on Signal Processing*, vol. 53, no. 8, pp. 2617–2624, 2005.

[11] ——, "Decentralized estimation in an inhomogeneous sensing environment," *IEEE Transactions on Information Theory*, vol. 51, no. 10, pp. 3564–3575, 2005.

[12] M. F. A. Ahmed, T. Y. Al-Naffouri, M. Alouini, and G. Turkiyyah, "The effect of correlated observations on the performance of distributed estimation," *IEEE Transactions on Signal Processing*, vol. 61, no. 24, pp. 6264–6275, Dec 2013.

[13] A. S. Behbahani, A. M. Eltawil, and H. Jafarkhani, "Linear estimation of correlated vector sources for wireless sensor networks with fusion center," *IEEE Wireless Communications Letters*, vol. 1, no. 4, pp. 400–403, 2012.

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

[15] 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.

[16] ——, "Decentralized estimation under correlated noise," *IEEE Transactions on Signal Processing*, vol. 62, no. 21, pp. 5603–5614, 2014.

[17] S. Liu, S. Kar, M. Fardad, and P. K. Varshney, "Optimized sensor collaboration for estimation of temporally correlated parameters," *IEEE Transactions on Signal Processing*, vol. 64, no. 24, pp. 6613–6626, 2016.

[18] 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.

[19] N. K. D. 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.

[20] N. K. D. Venkategowda, H. Lee, and I. Lee, "Joint transceiver designs for MSE minimization in MIMO wireless powered sensor networks," *IEEE Transactions on Wireless Communications*, vol. 17, no. 8, pp. 5120–5131, 2018.

[21] Y.-W. P. Hong, T.-C. Hsu, and P. Chennakesavula, "Wireless power transfer for distributed estimation in wireless passive sensor networks," *IEEE Transactions on Signal Processing*, vol. 64, no. 20, pp. 5382–5395, 2016.

[22] C. Psomas and I. Krikidis, "Energy beamforming in wireless powered mmwave sensor networks," *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 2, pp. 424–438, 2019.

[23] G. Pan, H. Lei, Y. Yuan, and Z. Ding, "Performance analysis and optimization for SWIPT wireless sensor networks," *IEEE Transactions on Communications*, vol. 65, no. 5, pp. 2291–2302, 2017.

[24] A. Hentati, J.-F. Frigon, and W. Ajib, "Energy harvesting wireless sensor networks with channel estimation: Delay and packet loss performance analysis," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 2, pp. 1956–1969, 2020.

[25] F. Jiang, J. Chen, A. L. Swindlehurst, and J. A. Lpez-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.

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

[27] 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.

[28] A. S. Leong, S. Dey, and J. S. Evans, "Asymptotics and power allocation for state estimation over fading channels," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 47, no. 1, pp. 611–633, 2011.

[29] A. S. Leong, S. Dey, G. N. Nair, and P. Sharma, "Power allocation for outage minimization in state estimation over fading channels," *IEEE Transactions on Signal Processing*, vol. 59, no. 7, pp. 3382–3397, 2011.

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

[31] R. R. Singh and K. Rajawat, "Online precoder design for parameter tracking in wireless sensor networks," in *2015 IEEE 26th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, 2015, pp. 2061–2065.

[32] J. Akhtar and K. Rajawat, "Distributed sequential estimation in wireless sensor networks," *IEEE Transactions on Wireless Communications*, vol. 17, no. 1, pp. 86–100, 2018.

[33] 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, 2019.

[34] 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.

[35] C. Li, Y. He, X. Wang, G. Li, and P. K. Varshney, "Distributed detection of sparse stochastic signals via fusion of 1-bit local likelihood ratios," *IEEE Signal Processing Letters*, vol. 26, no. 12, pp. 1738–1742, 2019.

[36] 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, 2017.

[37] ——, "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, 2017.

[38] 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, 2018.

[39] A. Carmi, P. Gurfil, and D. Kanevsky, "Methods for sparse signal recovery using Kalman filtering with embedded pseudo-measurement norms and quasi-norms," *IEEE Transactions on Signal Processing*, vol. 58, no. 4, pp. 2405–2409, 2010.

[40] M. Shamaiah and H. Vikalo, "Estimating time-varying sparse signals under communication constraints," *IEEE Transactions on Signal Processing*, vol. 59, no. 6, pp. 2961–2964, 2011.

[41] E. Karseras, K. Leung, and W. Dai, "Hierarchical Bayesian Kalman filters for wireless sensor networks," in *21st European Signal Processing Conference (EUSIPCO 2013)*, 2013, pp. 1–5.

[42] J. Ma, S. Zhang, H. Li, F. Gao, and S. Jin, "Sparse Bayesian learning for the time-varying massive MIMO channels: Acquisition and tracking," *IEEE Transactions on Communications*, vol. 67, no. 3, pp. 1925–1938, 2019.

[43] Y. Liu, S. Zhang, F. Gao, J. Ma, and X. Wang, "Uplink-aided high mobility downlink channel estimation over massive MIMO-OTFS system," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 9, pp. 1994–2009, 2020.

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

[45] 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.

[46] 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.

[47] K. P. Rajput, Y. Verma, N. K. D. Venkategowda, A. K. Jagannatham, and P. K. Varshney, "Robust linear transceiver designs for vector parameter estimation in MIMO wireless sensor networks under CSI uncertainty," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 8, pp. 7347–7362, 2021.

[48] K. P. Rajput, A. Kumar, S. Srivastava, A. K. Jagannatham, and L. Hanzo, "Bayesian learning-based linear decentralized sparse parameter estimation in MIMO wireless sensor networks relying on imperfect CSI," *IEEE Transactions on Communications*, vol. 69, no. 9, pp. 6236–6250, 2021.

[49] K. P. Rajput, M. F. Ahmed, N. K. D. Venkategowda, A. K. Jagannatham, G. Sharma, and L. Hanzo, "Robust decentralized and distributed estimation of a correlated parameter vector in MIMO-OFDM wireless sensor networks," *IEEE Transactions on Communications*, vol. 69, no. 10, pp. 6894–6908, 2021.

[50] S. M. Kay, *Fundamentals of Statistical Signal Processing:Estimation Theory*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1993.

[51] E. Karseras, K. Leung, and W. Dai, "Tracking dynamic sparse signals using hierarchical Bayesian Kalman filters," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2013, pp. 6546–6550.

[52] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.

[53] A. Haqiqatnejad, F. Kayhan, and B. Ottersten, "Robust SINR-constrained symbol-level multiuser precoding with imperfect channel knowledge," *IEEE Transactions on Signal Processing*, vol. 68, pp. 1837–1852, 2020.

[54] H. L. V. Trees and K. L. Bell, *Bayesian bounds for parameter estimation and nonlinear filtering/tracking*. Wiley-IEEE press New York, 2007.