IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received August 31, 2019, accepted October 3, 2019, date of publication October 15, 2019, date of current version October 28, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2947433

Statistical Inference-Based Distributed Blind Estimation in Wireless Sensor Networks

WENTAO ZHANG, LI CHEN, MIN QIN, AND WEIDONG WANG
Department of Electronic Engineering and Information Science, University of Science and Technology of China, Hefei 230027, China
Corresponding author: Li Chen (chenli87@ustc.edu.cn)

This work was supported by the National Science and Technology Major Project of China MIIT under Grant 2018ZX03001001-003.

ABSTRACT To realize the Internet of Things, one of the essential elements is wireless sensor networks which can sense the physical conditions of the environment. The ubiquitous sensing is achieved by a large number of spatially dispersed sensors and distributed estimation technology. However, the low-cost sensors are insufficient to support conventional distributed estimation schemes. Since most conventional schemes include channel training process, the resource consumption of which is enormous. Thus, one key challenge in designing a feasible distributed estimation scheme is to reduce resource consumption from channel training. We tackle the challenge by proposing a distributed blind estimation scheme. The proposed scheme consists of two components: random transmission and statistical inference. Specifically, assuming sensors contain only two states that are active and inactive. The random transmission strategy turns the sensing value into a parameter to govern the sensor states. At the fusion center, statistical inference method is used to recover the sensing value. The specific design of the inference method involves the distribution approximation and clustering, which are accomplished by Gaussian mixture model and expectation-maximization principle. By the proposed scheme, the channel information is no longer needed in distributed estimation. Therefore, it is more energy-efficient and more applicable to the complicated wireless environment compared with conventional schemes. Besides, we investigate the impacts of the number of sensors and quantization on the estimation performance. Finally, simulation results demonstrate the effectiveness of the proposed blind estimation scheme.

INDEX TERMS Distributed estimation, expectation maximum, Gaussian mixture model, random transmission, statistical inference, wireless sensor network.

I. INTRODUCTION

Ubiquitous sensing enabled by wireless sensor networks (WSN) has attracted increasing attention because of its various application areas including environment monitoring [1], health management [2], traffic monitoring [3] and industrial control [4], etc. The WSN develops rapidly since it contains the following advantages: the distributed processing of a large amount of collected information can improve the accuracy of monitoring and reduce the accuracy requirements for a single sensor; the redundant sensors make the system robust; a large number of sensors can increase the coverage of the monitored area. In the various applications, distributed estimation is one of the critical technologies since it can provide accurate estimates of the parameters of the phenomenon [5].

The associate editor coordinating the review of this manuscript and approving it for publication was Zhangbing Zhou.

A. RELATED WORKS

In large WSNs, one factor which affects the distributed estimation performance is the form of the sensor measurements (digital and analog). For the analog approach, the measurements are transmitted directly or via analog modulation to the fusion center (FC). For the digital approach, the sensors quantize the measurements first and then transmit the quantized measurements to the FC. Compared to the digital form, the analog form provides better estimation performance since it retains more information of the phenomenon [6]. However, the analog form is not practical since the WSNs are generally bandwidth-constrained. Several researchers have studied the quantization scheme in WSN [7]–[9]. For a heterogeneous WSN, a suboptimal bit allocation scheme was proposed in [7] aiming to minimizes MSE while satisfying a network bandwidth constraint. Utilizing identical one-bit quantizers that minimize Cramér-Rao lower bound (CRLB) was introduced in [8]. Another quantizer proposed in [9] was designed to maximize Bayesian Fisher information.

Another vital issue for distributed estimation is Multiple Access Channel (MAC). Three types of MAC commonly used for distributed estimation are coherent MAC, orthogonal MAC, and hybrid MAC. For the orthogonal MAC used in [10], the sensors transmit their measurements to the FC via orthogonal channels, which can be realized with the orthogonal frequency-division multiplexing or the orthogonal time-division multiplexing. For the coherent MAC, the measurements are coherently transmitted to the FC [11]. More recently, the hybrid MAC is proposed in [12], where all the sensors are divided into groups, and the coherent MAC is used within one group, whereas the orthogonal MAC is used across different groups. Compared to the coherent MAC, orthogonal MAC is more favorable for implementation since it requires no synchronization among sensors [13].

Many classics distributed estimation schemes have been proposed to improve the distributed estimation performance. Cui *et al.* [14] utilized Best Linear Unbiased Estimator (BLUE) to estimate the sensing value and discussed the power scheduling problem. Liner Minimum Mean Square Error (LMMSE) scheme was utilized in [11], [12], [15], [16]. Xiao *et al.* [11] considered the joint estimation of a source vector by linear source-channel coding. Liu and Chung [12] investigated the hybrid MAC for distributed estimation. Considering the unknown fading channels, Senol and Tepedelenlioglu [15] studied the impact of imperfect channel estimation and presented the optimum number of sensors given the total power and noise statistics. Also, Alam *et al.* [16] considered that the receiving signals were jointly corrupted by impulsive noise and channel fading, then LMMSE was used to recover the Gaussian source signal. Furthermore, distributed estimation schemes based on Maximum Likelihood Estimation (MLE) were proposed in [17]–[20]. Xiao *et al.* [17] considered the unreliable communication links from the sensors to the FC and proposed an MLE based on distributed average consensus method to make the system more robust. Wang and Yang [18] proposed optimal and suboptimal MLE to combat both the measured noise and communication errors led by channel fading. To further address the complexity and implementation issues of the optimal MLE, Aysal and Barner [19] proposed two, fast, practical and straightforward suboptimal solutions. Wang and Yang [20] proposed a robust distributed MLE which improves the estimation performance with imperfect channel estimation. They also found the optimal length of the training sequence.

More recently, several new challenges and latest advances have emerged in the problem of distributed estimation. The first challenge is that when sensors defect, the sensing value will be severely distorted. The distorted value causes a sever decline in estimation performance. Joint distributed detection-estimation schemes have been proposed to solve the challenge [21], [22]. Specifically, they assumed that when sensors defect, these sensors only generate pure-noise measurements, which implies that the desired signal is not observable at these sensors. Zhou *et al.* [21] proposed a learning-based distributed procedure, called the mixed detection-estimation (MDE) algorithm, based on iterative closed-loop interactions between validity learning (detection) and target estimation. The detection step reassesses the validity of the local measurements at each sensor node, and the target estimation step reconstructs the desired signal. Zhu and Sun [22] proposed a joint detection and estimation fusion scheme with correlated sensor quantized data. Heterogenous bandwidth constraint is another challenge in WSN where each sensor node has different or adaptive quantization rate. Sani and Vosoughi [23] proposed a linear distributed estimation scheme to address two critical problems pertaining to bandwidth-constrained distributed estimation in a heterogenous sensor network: (1) given a network bandwidth constraint, they investigated the quantization rate allocation schemes to minimize Mean Square Error (MSE) at the FC; (2) given a target MSE at the FC, they explored the quantization rate allocation schemes to minimize the required network bandwidth. Unlabeled sensing is a novel scheme for parameter estimation where each sensor acquires a noisy version of the signal and the data at the FC are unlabeled [24]–[26]. It is a practical scheme to reduce communication cost and latency. A permutation matrix is utilized to accomplish parameter estimation.

Traditional distributed estimation schemes essentially rely on known and tractable mathematical models. However, the blind estimation scheme makes the ready-made mathematical models out-of-operation. Data-driven machine learning is a powerful tool which provides an alternative technique of adaptive modeling and parameter estimation relying on learning from data [27]. Generally, machine learning algorithms are classified into three categories, i.e. supervised learning, unsupervised learning and reinforcement learning. The existing research most relevant to the current work is the applications of the unsupervised learning algorithm. Recently, K-means as an effective unsupervised clustering technique is widely applied to achieve blind detector [28]–[30]. In [28], a coding-aided K-means clustering (CKMC) blind detector for space shift keying (SSK) multiple-input multiple-output (MIMO) systems is proposed where the training of CSI is not required. Improved K-means blind detectors are proposed in [29], [30] to avoid the error floor effects caused by bad initial cluster centers. Density-based spatial clustering applications with noise (DBSCAN) algorithm is utilized to address the problem of anomaly detection [31]. GMM-EM [32] is another powerful unsupervised learning algorithm that has been used for density estimation and data clustering. In [33], the measurements of the sensor are statistically modeled by a Gaussian mixture model, and a distributed expectation maximization algorithm is proposed to estimate the model parameters clustering the measurements.

B. MOTIVATION

Among previous works on distributed estimation, one part of them did not consider the impact of the fading channel (known as error-free transmission) and the others assumed that either universal channel models or instantaneous channel coefficients are given. In other words, the Channel State Information (CSI) or the complete Channel Density Information (CDI) are known in these works. However, transmitting training sequences to obtain the CSI or giving the CDI has the following significant shortcomings: 1) it increases communication cost and power consumption; 2) the training sequences are needed to be sent frequently when the channels are fast time-varying; 3) the deviation between the channel model and the real channel seriously affects the estimation performance. Therefore, a blind estimation without channel information is somehow indispensable and more preferable in practice.

Data-driven unsupervised learning algorithms provide potential approach for distributed blind estimation. These unsupervised clustering algorithms have their own property. The performance of K-means algorithm tends to be affected by skewed data distributions, i.e., imbalanced data [34]. However, imbalanced data will appear in distributed estimation. DBCAN is sensitive to the training parameter, however, if the data and scale are not well understood, choosing meaningful parameters can be difficult. Besides, a significant limitation of K-means and DBSCAN is that the data point is deterministically assigned to one and only one cluster, however, there is an overlap between the true data clusters in reality. The classification ambiguity will occur in the overlapping region. GMM-EM clustering algorithm is capable of addressing the above classification ambiguity problem by soft assigning the data to different clusters. Through the GMM-EM clustering, the data point is no longer deterministically assigned to one cluster; in contrast, the data is assigned to more than one cluster according to a series of probability values. Also, GMM-EM clustering is capable of processing imbalanced data. Hence, GMM-EM algorithm is the most suitable method for the distributed blind estimation.

C. CONTRIBUTIONS AND ORGANIZATION

In this paper, inspired by joint detection-estimation and unlabeled sensing, we propose a blind estimation scheme based on statistical inference to recover the sensing value without any channel information. In the proposed scheme, quantization is utilized since the bandwidth of the transmission channel is assumed to be strictly constrained. The design of quantization is in accordance with probability quantizer [35], [36], where for a given input, there are only two candidate symbols. The quantized symbol indicates the sensor state (active or inactive). In our design, the sensing value turns into a parameter to govern the sensor states (the quantized symbols). Orthogonal MAC is adopted since the blind distributed estimation is more feasible without interference. Corrupted by fading channels and additive noise, the received signals are regarded as unlabeled data. The GMM-EM algorithm is used to detect the sensor state by clustering the received signals. Finally, we obtain the estimated value according to the results of the clustering. The proposed estimation scheme has the advantages of low complexity and energy-efficient. The main contributions of the work are summarized as follows

- *Random transmission strategy:* Under the condition of without any channel information, we design a random transmission strategy to turn the sensing value into a parameter to govern the sensors states. Mathematically, the states are distributed as Bernoulli random variables (r.v.). We prove that when the number of sensors is large enough, the proportion of the active sensors can approximate the normalized measurements. Therefore, the original estimation problem for the sensing value is equivalent to estimate the number of active sensors for a given sensor amount. Then we investigate the relation between the number of sensors and the estimation performance. Moreover, the impact of quantization on estimation performance is studied.
- *Statistical inference principle:* After the random transmission, the FC is designed to estimate the sensing value (equivalent to estimate the proportion of the active sensors). To this end, statistical inference method based on GMM-EM [37] is utilized to approximate the distribution of the received signals and clustering the signals.

The remainder of the paper is organized as follows. Section II introduces the system model and problem formulation. The random transmission strategy and the statistical inference principle are presented in Section III and Section IV, respectively. Section V gives a summary of the proposed distributed blind estimation scheme. Section VI provides the simulation results, followed by concluding remarks in Section VII.

II. SYSTEM MODEL AND PROBLEM FORMULATION

A. SYSTEM MODEL

The system model we consider comprises an FC and N sensors, as shown in Fig. 1. All the sensors and the FC

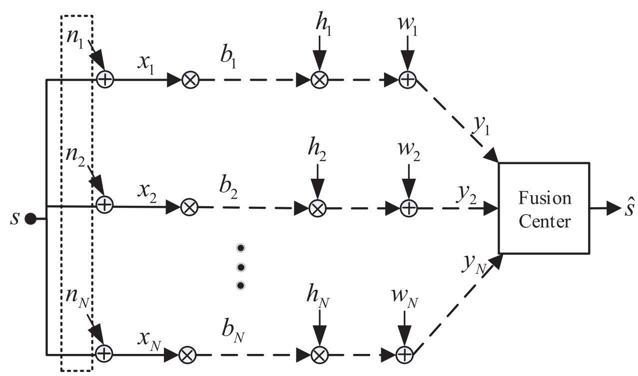


FIGURE 1. System model. s denotes the sensing value, n denotes the measure noise, x denotes the measurement, b denotes the quantized symbol, h denotes the transmission channel, w denotes the received noise at the fusion center, y denotes the received signal, and $\hat{s}$ denotes the estimated value.

are equipped with single-antenna. The N sensors measure a common scalar sensing value s. Generally, the environment varies more slowly than communication channels. Hence it is reasonable to assume s is determinate. The N measurements corrupted by noise are denoted as

$$x_i = s + n_i, \quad i = 1, \ldots, N, \tag{1}$$

where $\{n_i,\ i = 1, \ldots, N\}$ are additive white Gaussian noises (AWGNs) following i.i.d. $\mathcal{N}\left(0, \sigma_n^2\right)$ distribution. Thereby the measurements $\{x_i,\ i = 1, \ldots, N\}$ follow i.i.d. $\mathcal{N}\left(s, \sigma_n^2\right)$ distribution and they are assumed to be limited within $(0, W)$ since the sensors generally have an inherent measuring range.

Due to the inherent bandwidth and energy limitations, the measurements $\{x_i,\ i = 1, \ldots, N\}$ are quantized. The quantized symbols are denoted as $\{b_i,\ i = 1, \ldots, N\}$ which indicate the sensor states (active or inactive). Therefore, for a given input, the quantized output is one of two candidate symbols.

We assume that the N sensors transmit the quantized symbols to the FC via N orthogonal multiple access fading channels. Channel coefficients are assumed to be i.i.d, but the distribution is unknown. Let h_i represents the channel coefficient from sensor i to the FC. We assume pair-wise synchronization between each sensor and the FC. However, synchronization among all the sensors is not required. The received signal at the FC from the sensor i, denoted as y_i, is given by

$$y_i = h_i b_i + w_i, \quad i = 1, \ldots, N, \tag{2}$$

where the received noises $\{w_i, i = 1, \ldots, N\}$ are AWGNs following i.i.d. $\mathcal{N}\left(0, \sigma_w^2\right)$.

B. PROBLEM FORMULATION

First of all, we review one of the conventional estimation schemes MLE. The fundamental process and elements of MLE are illustrated in Fig. 2, where $p\left(\cdot|\cdot\right)$ denotes the conditional probability distribution function. Utilizing the MLE, the likelihood function (i.e., the conditional probability function $p\left(y|\, s\right)$) is necessary, since the correlation between the sensing value s and the received signals y is expressed by $p\left(y|\, s\right)$. We obtain the estimated value $\hat{s}$ by maximizing $p\left(y|\, s\right)$. Notice that CSI (h) or CDI (h) is essential to calculate the conditional probability function $p\left(y|\, s\right)$. Assume that there is no quantization at the sensors, and the FC has the knowledge of CSI. Then the FC can simply perform the ML estimate of s by

$$\begin{aligned}
&\hat{s}_{\text{MLE_CSI}} \\
&= \arg\max_s \sum_{i=1}^{N} \log p\left(y_i|\, s\right) \\
&= \arg\max_s \log \frac{1}{\left(2\pi\sigma_w^2\right)^{N/2}} \exp\left[-\frac{\sum\limits_{i=1}^{N}\left(y_i - h_i s\right)^2}{2\sigma_w^2}\right].
\end{aligned} \tag{3}$$

Another case is that the FC has complete CDI. To simplify the problem, we assume that the Probability Distribution Function (PDF) of the channel is $\mathcal{N}\left(0, 1\right)$, then the estimated value $\hat{s}$ is given by

$$\begin{aligned}
&\hat{s}_{\text{MLE_CDI}} \\
&= \arg\max_s \sum_{i=1}^{N} \log p\left(y_i|\, s\right) \\
&= \arg\max \log \frac{1}{\left[2\pi\left(s^2 + \sigma_w^2\right)\right]^{N/2}} \exp\left[-\frac{\sum\limits_{i=1}^{N} y_i^2}{2\left(s^2 + \sigma_w^2\right)}\right].
\end{aligned} \tag{4}$$

As far as we know, most of the conventional schemes for distributed estimation assume that the CSI is perfectly obtained or the complete CDI is given. To avoid the dependence on the channel information, a blind estimation scheme that requires no channel information (CSI or CDI) is indispensable.

The blind estimation scheme faces the following two challenges. First, as shown in Fig.2, the value of the received signal is jointly determined by the sensing value s and the channel h. The conditional probability function $p\left(y|\, s\right)$ gives the correlation between s and y. However, $p\left(y|\, s\right)$ is unacquirable without CSI or CDI. Therefore, the sensing value can not be estimated from the received signals alone. Hence the first challenge is to break the conventional transmission model such that the sensing value can be estimated from the received signals alone. Second, the blind estimation algorithm aims to estimate the sensing value s without any channel information. It requires that the blind estimation algorithm ought to be feasible under any channel condition. Therefore, the second challenge is to provide a unified framework to estimate the sensing value. The two challenges to realize the blind estimation design are summarized as:

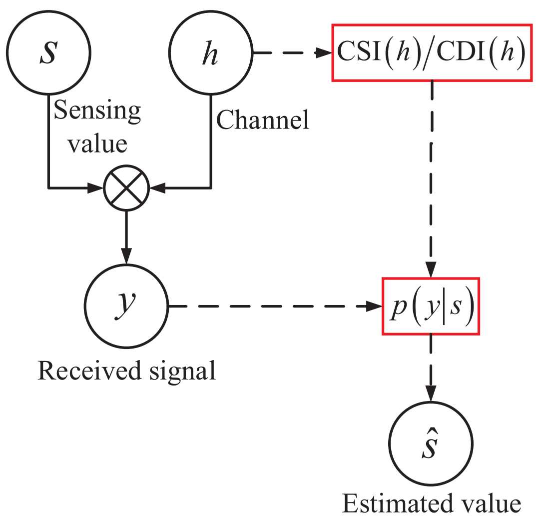


FIGURE 2. Maximum likelihood estimation scheme without quantization, where the received signal y is the product of the sensing value s and the channel h, and $p\left(y|\, s\right)$ denotes likelihood function.

1) Propose a novel transmission strategy to avoid the dependence on the channel information;

2) Establish a unified estimation framework to guarantee the estimation is valid under any channel condition.

III. RANDOM TRANSMISSION STRATEGY

A potential consideration of the novel transmission is illustrated in Fig. 3, where $p(\cdot|\cdot)$ denotes the conditional probability distribution function and $stat(\cdot)$ denotes statistical characteristics, like sample mean and sample size. The objective of the novel transmission strategy is to turn the sensing value s into a parameter and pass the parameter to the FC. Specifically, let all the quantized symbols be denoted by a vector $\mathbf{b} = [b_1, b_2, \ldots, b_N]^T$. We consider randomizing $\mathbf{b}$ such that distribution function $p(\mathbf{b})$ is parameterized by s. Then the sensors transmit the random vector $\mathbf{b}$ to the FC instead of s. Therefore, it is called random transmission. Statistical characteristics of the received signals $stat(y|\mathbf{b})$ imply the correlation between the received signals y and the random vector $\mathbf{b}$.

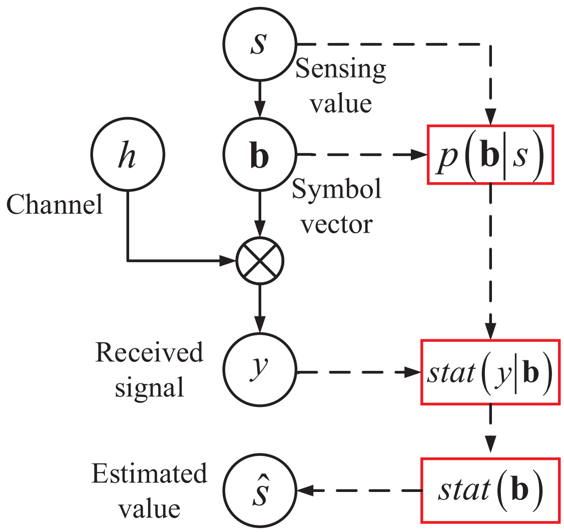


FIGURE 3. Blind estimation scheme where $p(\cdot|\cdot)$ denotes the conditional probability distribution function and $stat(\cdot)$ denotes statistical characteristics.

In the following, we first present the one-bit random transmission strategy in details and reveal the correlation between the sensing value s and the received signals y applying the random transmission. In addition, we investigate the relation between the estimation error and the number of sensors. To further study the effect of quantization on the estimation performance, we extend the random transmission from one-bit quantization to multiple-bit quantization situation.

A. ONE-BIT RANDOM TRANSMISSION

The one-bit random transmission scheme consists of three steps: Range normalization, Generate random threshold and Quantization. In the following, we give the three steps in details.

- *Step 1 (Range normalization):* All the measurements $\{x_i, \ i = 1, \ldots, N\}$ are divided by the max boundary value W. The normalized measurements are denoted as $\{\theta_i, i = 1, \ldots, N\}$;
- *Step 2 (Random threshold generating):* During one transmission, a sensor generates a random quantization threshold denoted as ρ_i. The random threshold uniformly distributes in the range of (0,1) and all the N thresholds are independently generated among the sensors.
- *Step 3 (Quantization):* At arbitrary sensor i, the normalized measurement value θ_i is compared with the random quantization threshold ρ_i. The sensor state is in accordance with the following rules: if $\theta_i > \rho_i$, the sensor i is active, otherwise inactive. Then the sensor sate determines the quantized symbol: if the sensor i is active, $b_i = a$ and if the sensor i is inactive, $b_i = 0$.

One-bit random transmission turns the sensing value s into the distribution parameter of the random vector $\mathbf{b}$. Let m be the number of active sensors. The relation between the estimated value $\hat{s}$ and the number of the active sensors is given in Proposition 1.

Proposition 1: By one-bit random transmission, the estimated value $\hat{s}$ is given by

$$\hat{s} = \frac{m \cdot W}{N}. \tag{5}$$

where $\lim_{N\to\infty} \frac{m}{N} \to \theta$ and θ is the normalized measurement without the measure noise.

Proof: See Appendix A. □

Remark 1: Proposition 1 reveals that the estimated value is determined by the number of active sensors for a given sensor amount. Therefore, the random transmission scheme changes the original estimation problem for the sensing value s to estimate the number of the active sensors m.

The relation between the sensing value s and the received signal y is further analyzed. Fortunately, under the one-bit quantization and the assumption of i.i.d channel, the received signals present two distinct distributions theoretically. Specifically, if the sensor i is active, the received signal can be written as $y_i = a{\cdot}h_i{+}w_i$. If the sensor i is inactive, the received signal can be written as $y_i = 0 \cdot h_i + w_i = w_i$, namely the received signal only consists of the noise w_i. Therefore all the received signals can be divided into two clusters. One cluster represents the collection of the active sensors, and the other one represents the collection of the inactive sensors. The size of the clusters reflects the relation between s and y.

According to Proposition 1, we obtain the estimated value $\hat{s}$. An explicit relation between the estimation error $|s - \hat{s}|$ and the number of sensors N is shown in Proposition 2.

Proposition 2: Given a target value σ of the average estimation error, to guarantee the probability $P(|s - \hat{s}| < \sigma)$ greater than a threshold probability C, the number of sensors

N should satisfy the following condition

$$N > \frac{\left(W\tilde{\beta}^{*}\right)^{2}\theta(1-\theta)}{\sigma^{2}} \tag{6}$$

where, W is the max boundary value of x, $\tilde{\beta}^{*}$ is the root of $\Re(x) - \Re(-x) = C$ and $\Re(x)$ is the standard normal distribution.

Proof: See Appendix B. □

Remark 2: According to proposition 2, we find that to guarantee a certain system performance, the required number of sensors is determined by the threshold probability C and the target estimation error σ. High estimation accuracy or high estimation probability require more sensors.

B. MULTIPLE-BIT RANDOM TRANSMISSION

To further investigate the effect of quantization on the estimation performance, we extend the random transmission from one-bit to multiple-bit situation. We consider a uniform quantizer. Let $\left\{\Delta = 1/\left(2^{M}-1\right), M = 2, 3, \ldots\right\}$ be the quantization resolutions and define the quantization lattice in $\mathbb{R}$ by

$$\Lambda = \left\{j\Delta : j = 0, 1, \ldots, 2^{M-1}\right\}. \tag{7}$$

It is assumed that all the normalized measurements are within the same interval $(t\Delta, (t+1)\Delta)$, where $t \in \left\{0, 1, \ldots, 2^{M}-2\right\}$. The quantizer is a function that maps the normalized measurements to some point in Λ. In the following, we give the abridged steps of multiple-bit random transmission, which is similar to the one-bit situation.

- *Step 1 (Range normalization):* Same as one-bit random transmission and the normalized measurements denoted as $\{\theta_i, i = 1, \ldots, N\}$;
- *Step 2 (Random threshold generating):* All the sensors independently generate a random threshold ρ_i which uniformly distributes in the range of $(t\Delta, (t+1)\Delta)$.
- *Step 3 (Quantization):* The normalized measurement value θ_i is compared with the random threshold ρ_i. Then the comparison result determines the quantized symbol: if $\theta_i > \rho_i$, $b_i = a(t+1)\Delta$, otherwise $b_i = at\Delta$.

Similar to the one-bit situation, we divide the sensors into active ones and inactive ones. The quantized symbol $b_i = a(t+1)\Delta$ corresponds to the active sensor while $b_i = at\Delta$ corresponds to the inactive sensor. Let m be the number of active sensors. The relation between $\hat{s}$ and m is given in Proposition 3. Then Proposition 4 studies the effect of quantization on the estimation performance.

Proposition 3: By multiple-bit random transmission, the estimated value $\hat{s}$ is given by

$$\hat{s} = \frac{m/N + t}{2^{M}-1} \cdot W, \tag{8}$$

where $t \in \left\{0, 1, \ldots, 2^{M}-2\right\}$ is the index of the quantization interval. Notice that when $M = 1$, there is $t = 0$ and the formula of (8) is identical with (5).

Proof: See Appendix A. □

Proposition 4: Applying the multiple-bit random transmission strategy with a given threshold probability C, the probability distribution of the estimation error $|s - \hat{s}|$ satisfies

$$P\left\{|\hat{s} - s| < \frac{W\tilde{\beta}^{*}\sqrt{\phi(1-\phi)}}{(2^{M}-1)\sqrt{N}}\right\} = C, \tag{9}$$

where $\tilde{\beta}^{*}$ is the root of $\Re\left(\tilde{\beta}\right) - \Re\left(-\tilde{\beta}\right) = C$, $\Re(x)$ is the standard normal distribution, $\phi = \left(2^{M}-1\right)\theta - \left\lfloor\left(2^{M}-1\right)\theta\right\rfloor$, $\lfloor\cdot\rfloor$ is the round down operator, and $\theta = s/W$.

Proof: See Appendix C. □

Remark 3: Let $\Omega = \frac{W\tilde{\beta}^{*}\sqrt{\phi(1-\phi)}}{(2^{M}-1)\sqrt{N}}$ be the estimation error bound. The smaller Ω indicates better estimation performance. It is worth noting that the sensing value s affects Ω. For the same M, different s results in different Ω. As M increases, the impact will be negligible.

Given W, N, C and M, the estimation error bound Ω is determined by the term $O(\phi) = \sqrt{\phi(1-\phi)}$ with constraint $0 < \phi < 1$. It is easy to see that $O(\phi)$ is symmetric and convex. When $\phi = 0.5$, $O(\phi)$ achieves a global maximum, and when ϕ is close to 0 or 1, $O(\phi)$ approaches a global minimum. As shown in Fig. 4, the position of θ in the quantization interval $(t\Delta, (t+1)\Delta)$ determines the value of ϕ. When θ is near the quantization interval boundary, ϕ is close to 0 or 1, and when θ is in the middle of the quantization interval, ϕ is equal to 0.5. Therefore the sensing value s affects the estimation error bound Ω. However, since ϕ is in the range of (0,1), the difference of Ω brought by $O(\phi)$ is limited. As M increases, the value of Ω is mainly determined by the denominator. Hence the impact of different sensing value will be negligible.

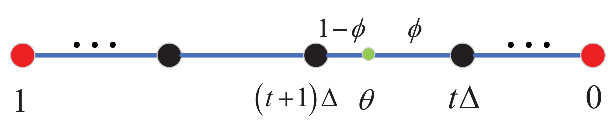


FIGURE 4. Multiple-bit uniform quantizer.

IV. STATISTICAL INFERENCE PRINCIPLE

By random transmission, the N sensors send a random vector $\mathbf{b}$ whose elements are binary. Therefore the corresponding received signals y can be regarded as labeled data. The elements of $\mathbf{b}$ are the labels of the received signals. However, the labels are invisible to the FC. Therefore, the objective is to obtain the labels and cluster the received signals. The cluster size can represent the number of sensors in different states. Then we obtain the estimated value $\hat{s}$ according to Proposition 1.

We propose a unified estimation frame based on statistical inference involving GMM and EM algorithm. GMM as a well-known clustering algorithm perfectly matches our design objective. GMM can approximate an arbitrary distribution with sufficient Gaussian components which guarantees the algorithm valid under any channel conditions. Utilizing GMM to model the distribution of the received signals, the mixing weights of Gaussian components represent the sizes of the data clusters. Hence, the weights of Gaussian components are used to estimate the sensing value s. The EM algorithm which used to estimate the parameters of the GMM will be introduced next. Finally, The GMM-EM based estimation algorithm is given.

A. GAUSSIAN MIXTURE MODEL

Gaussian mixture models are used for approximating multi-modal distributions. The PDF of a GMM with K components can be expressed as

$$p(x) = \sum_{k=1}^{K} \pi_k \mathcal{N}\left(x | \mu_k, \sigma_k^2\right), \tag{10}$$

where $\{\pi_i, i = 1, \ldots, K\}$ are the mixing weights with constraints $0 \leq \pi_i \leq 1$ and $\sum_{i=1}^{K} \pi_i = 1$, $\left\{\mathcal{N}\left(\mu_i, \sigma_i^2\right), i = 1, \ldots, K\right\}$ are the gaussian components, μ_i and σ_i^2 are the mean and the variance of the Gaussian component, respectively.

B. EM ALGORITHM FRAME

Assume that we have a probabilistic model consisting of observed variables $\mathbf{v}$, hidden variables $\mathbf{h}$ and unknown parameters Φ. Generally, the unknown parameters are estimated by solving the maximum log likelihood (ML) problem which can be expressed as

$$\hat{\Phi} = \arg\max_{\Phi} \log p(\mathbf{v} | \Phi) \tag{11}$$

However, directly solving (11) is intractable because it is costly to marginalize out the hidden variables $\mathbf{h}$. Alternatively, a lower bound of $\log p(\mathbf{v} | \Phi)$ is defined as

$$\mathcal{L}(\mathbf{v}, q(\mathbf{h}) | \Phi) = \log p(\mathbf{v} | \Phi) - D_{\mathrm{KL}}\left(q(\mathbf{h}) \| p(\mathbf{h} | \mathbf{v}; \Phi)\right) \tag{12}$$

where q is an arbitrary probability distribution of $\mathbf{h}$, Kullback-Leibler (KL) divergence describes the "similarity" between $q(\mathbf{h})$ and $p(\mathbf{h} | \mathbf{v}; \theta)$. Because the KL divergence is always nonnegative so that the lower bound has at most the same value as the log-probability. The lower bound is equal to the log-probability if and only if $D_{\mathrm{KL}}\left(q(\mathbf{h}) \| p(\mathbf{h} | \mathbf{v}; \Phi)\right) = 0$, by letting $q(\mathbf{h}) = p(\mathbf{h} | \mathbf{v}; \Phi)$. Furthermore, the more canonical definition of lower bound $\mathcal{L}(\mathbf{v}, q(\mathbf{h}) | \Phi)$ is derived as

$$\mathcal{L}(\mathbf{v}, q(\mathbf{h}) | \Phi) = \mathbb{E}_{\mathbf{h} \sim q}\left[\log p(\mathbf{v}, \mathbf{h} | \Phi)\right] + H(q(\mathbf{h})) \tag{13}$$

where $\mathbb{E}_{\mathbf{h} \sim q}\left[\log p(\mathbf{v}, \mathbf{h} | \Phi)\right]$ is defined as Q function. Then the ML problem of (12) can be solved by maximizing the lower bound $\mathcal{L}(\mathbf{v}, q(\mathbf{h}) | \Phi)$.

The EM algorithm estimates the unknown parameters by maximizing the lower bound $\mathcal{L}$. The maximization of the lower bound can be solved by repeating the follow two steps until convergence, where the subscript (t) denotes the iteration index:

Expectation-Step (E-Step):

$$q^{(t)}(\mathbf{h}) = p\left(\mathbf{h} | \mathbf{v}; \Phi^{(t-1)}\right)$$
$$\Rightarrow Q\left(\Phi, \Phi^{(t-1)}\right) = \mathbb{E}_{\mathbf{h} \sim q^{(t)}(\mathbf{h})}\left[\log p(\mathbf{v}, \mathbf{h} | \Phi)\right], \tag{14}$$

Maximization-Step (M-Step):

$$\Phi^{(t)} = \arg\max_{\Phi} Q\left(\Phi, \Phi^{(t-1)}\right). \tag{15}$$

C. GMM-EM BASED ESTIMATION

By a sufficient number of Gaussian functions, and adjusting their means and covariances as well as the weights in the linear combination, almost any continuous density can be approximated to arbitrary accuracy [38]. Therefore, we consider utilizing a GMM with K ($K \geq 2$) components to approximate the distribution of the received signals as

$$p(y) = \sum_{k=1}^{K} \pi_k \mathcal{N}\left(y | \mu_k, \sigma_k^2\right), \tag{16}$$

where $\{\pi_i, i = 1, \ldots, K\}$ are the mixing weights with constraints $0 \leq \pi_i \leq 1$ and $\sum_{i=1}^{K} \pi_i = 1$, $\left\{\mathcal{N}\left(\mu_i, \sigma_i^2\right), i = 1, \ldots, K\right\}$ are the gaussian components, μ_i and σ_i^2 are the mean and the variance of the Gaussian component respectively. Let the first Gaussian component denote the distribution of the signals from the inactive sensors, and the superposition of the other Gaussian components approximates the distribution of the signals from the active sensors. We can rewrite (16) as

$$p(y) = \pi_1 \mathcal{N}\left(y | 0, \sigma_1^2\right) + \sum_{k=2}^{K} \pi_k N\left(y | \mu_k, \sigma_k^2\right). \tag{17}$$

The unknown parameter set of the GMM is denoted as $\Phi = \left\{\pi_1, \ldots, \pi_K, \mu_2, \ldots, \mu_K, \sigma_1^2, \ldots, \sigma_K^2\right\}$. Generally, the unknown parameters can be found by solving the following ML problem:

$$\hat{\Phi} = \arg\max_{\Phi} p(y_1, y_2, \ldots, y_N | \Phi) \tag{18}$$

However, the lack of information concerning the labels of received signals makes the typical MLE of (18) intractable. The missing labels, so-called hidden variables, indicate the received signals assigning to which Gaussian components.

The EM algorithm is commonly used to find the ML solutions for the probabilistic models with hidden variables. The followings are the derivation of E-step and M-step in details.

(a) E-step: Firstly, we introduce N missing labels denoted as $\{\mathbf{z}_i = [z_{i1}, z_{i2}, \ldots, z_{iK}], i = 1, \ldots, N\}$. The labels are K-dimensional binary random vectors, having the "one-hot" form that only one particular element is equal to one and the others are equal to zero. The particular element indicates the received signal belongs to which Gaussian component.

Thus the likelihood function of the complete-data $\{y_i, \mathbf{z}_i\}$ is derived as

$$p(y_1, \mathbf{z}_1, \ldots, y_N, \mathbf{z}_N | \Phi) = \prod_{i=1}^{N} \left(\pi_1 \cdot \mathcal{N}\left(y_i | 0, \sigma_1^2\right)\right)^{z_{i1}} \prod_{k=2}^{K} \left(\pi_k \cdot \mathcal{N}\left(y_i | 0, \sigma_1^2\right)\right)^{z_{ik}} \quad (19)$$

Then the log-likelihood function of (19) is given by

$$\log p(y_1, \mathbf{z}_1, \cdots, y_N, \mathbf{z}_N | \Phi) = \sum_{i=1}^{N}\sum_{k=1}^{K} z_{ik} \log \pi_k + \sum_{i=1}^{N} z_{i1} \log \mathcal{N}\left(y_i | 0, \sigma_1^2\right) + \sum_{i=1}^{N}\sum_{k=2}^{K} z_{ik}\left[\log \mathcal{N}\left(y_i | \mu_k, \sigma_k^2\right)\right] \quad (20)$$

The expectation of hidden variables is derived as

$$\mathbb{E}[z_{ik}] = \gamma(z_{ik}) = p\left(z_{ij} | y_i; \Phi\right) = \frac{p(z_{ik}=1, y_i; \Phi)}{p(y_i; \Phi)} = \frac{p(y_i | z_{ik}=1, \Phi)\, p(z_{ik}=1)}{\sum_{k=1}^{K} p(y_i | z_{ik}=1, \Phi)\, p(z_{ik}=1)}, \quad i=1,\ldots,N; \quad k=1,\ldots,K. \quad (21)$$

Thus the Q function is derived as

$$Q\left(\Phi, \Phi^g\right) = \mathbb{E}\left[\log p(y_1, \mathbf{z}_1, \cdots, y_N, \mathbf{z}_N | \Phi) | y_1, \ldots, y_N, \Phi^g\right] = \sum_{i=1}^{N}\sum_{k=1}^{K} \gamma(z_{ik}) \log \pi_k + \sum_{i=1}^{N} \gamma(z_{i1}) \log \mathcal{N}\left(y_i | 0, \sigma_1^2\right) + \sum_{i=1}^{N}\sum_{k=2}^{K} \gamma(z_{ik})\left[\log \mathcal{N}\left(y_i | \mu_k, \sigma_k^2\right)\right] \quad (22)$$

(b) M-step: The GMM parameters are optimized separately by maximizing the Q function as following

$$\Phi^n = \arg\max_{\Phi} Q\left(\Phi, \Phi^g\right), \quad (23)$$

where Φ^n denotes the updated parameters set and Φ^g is the old parameters set.

Proposition 5: The optimal values of the parameters by maximizing the Q function are given as follows:

$$\pi_k^* = \frac{\sum_{i=1}^{N} \gamma(z_{ik})}{N}, \quad k=1,\ldots,K, \quad (24)$$

$$\mu_k^* = \frac{\sum_{i=1}^{N} \gamma(z_{ik}) \cdot y_i}{\sum_{i=1}^{N} \gamma(z_{ik})}, \quad k=2,\ldots,K, \quad (25)$$

$$\left(\sigma_1^2\right)^* = \frac{\sum_{i=1}^{N} \gamma(z_{i1})\, y_i^2}{\sum_{i=1}^{N} \gamma(z_{i1})} \quad (26)$$

$$\left(\sigma_k^2\right)^* = \frac{\sum_{i=1}^{N} \gamma(z_{ik})(y_i-\mu_k)^2}{\sum_{i=1}^{N} \gamma(z_{ik})}, \quad k=2,\ldots,K. \quad (27)$$

Proof: See Appendix D. □

Remark 4 (Special Case - Gaussian Channel): There is a special case that the channels follow i.i.d $\mathcal{N}\left(\mu_h, \sigma_h^2\right)$. The distributions of noises from the FC are assumed i.i.d $\mathcal{N}\left(0, \sigma_w^2\right)$. Thus, the received signals from the active sensors follow $\mathcal{N}\left(a\mu_h + \mu_w, a^2\sigma_h^2 + \sigma_w^2\right)$, the received signals from the inactive sensors follow $\mathcal{N}\left(0, \sigma_w^2\right)$. The distribution of all the received signals can be simply modeled as a two-component GMM.

Once the GMM-EM has run to completion, the proportion of the active sensors $\frac{m}{N}$ is approximated by

$$\frac{m}{N} \simeq 1 - \pi_1^*. \quad (28)$$

Utilizing the GMM-EM algorithm, we obtain a series of probability values (21) corresponding to the different clusters. For a specific received signal, if the probability value of the first term is close to 0.5, we can not exactly infer the state of the sensor from this received signal, in other words, this received signal will cause classification ambiguity which will degrade the estimate accuracy. The probability values of these specific received signals are reassigned when we estimate the proportion of the active sensors. Specifically, when the GMM converges, we distinguish the ambiguous data in accordance with the following rules: if $|\gamma(z_{i1}) - 0.5| < \tau$, the data y_i is judged as an ambiguous data where τ is the judgment threshold. The ambiguous data set denotes as $\mathcal{D}$, and the size of $\mathcal{D}$ is $L_{\mathcal{D}}$. Then we calculate auxiliary probability values by

$$p_k = \frac{\sum_{i \subset \{y_i, i=1,\ldots,N\} \setminus \mathcal{D}} \gamma(z_{ik})}{N - L_{\mathcal{D}}}, \quad k=1,\ldots,K, \quad (29)$$

which are reassigned to the ambiguous data. Thus a more accurate approximation to the proportion of the active sensors $\frac{m}{M}$ is given by

$$\frac{m}{N} \simeq 1 - \frac{\sum_{i \subset \{y_i, i=1,\ldots,N\} \setminus \mathcal{D}} \gamma(z_{i1}) + L_{\mathcal{D}} p_1}{N} \quad (30)$$

V. SUMMARY OF THE BLIND ESTIMATION SCHEME

The implementation of the blind estimation scheme is shown in Fig. 5. The scheme consists of two main modules Random Transmission and Statistical Inference. At the sensors, the sensing value s turns into a parameter to govern the sensor states which is indicated by the random symbol vector $\mathbf{b}$. The specific procedure of the Random Transmission is give in Algorithm 1. Then the random symbol vector $\mathbf{b}$ is transmitted to the FC via fading channels h_i, $i = 1, \ldots, N$. The received signals y_i, $i = 1, \ldots, N$ are divided into two clusters based on GMM-EM estimation. Finally, the sensing value is estimated based on the size of the clusters. Algorithm 2 summarizes the resultant procedure of the Statistical Inference principle at the FC.

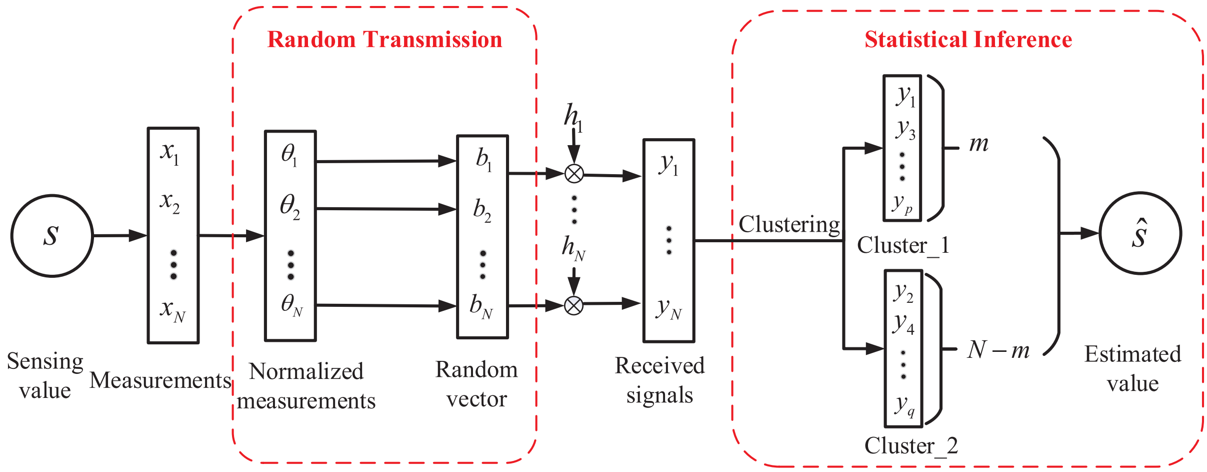


FIGURE 5. System design of the distributed blind estimation scheme.

Algorithm 1 Random Transmission at Sensors

Input: Measurements $x_i, i = 1, \ldots, N$ and max boundary value W.
Output: Random symbol vector $\mathbf{b}$.
Calculate normalized measurements θ_i according to formula (32);
Generate random threshold $\rho_i, i = 1, \ldots, N$ within the range of $(t\Delta, (t+1)\Delta)$;
for $i = 1$ *to* N **do**
  if $\theta_i > \rho_i$ **then**
    $b_i = a(t+1)\Delta$;
  else
    $b_i = at\Delta$;
  end
end
Return: Random symbol vector $\mathbf{b}$.

Algorithm 2 Statistical Inference at FC

Input: Received signals $y_i, i = 1, \ldots, N$, the number of Gaussian components K and max boundary value W.
Output: Estimated value $\hat{s}$.
Initialization: Initialize GMM parameter set Φ^0;
EM Iteration:
while *Not Convergence* **do**
  E-step: Calculate the expectation of hidden variables according to (21) and Q function according to (22);
  M-step: Update the mixing weight π_k, the mean μ_k and variance σ_k^2 according to (24), (25), (26) and (27), respectively;
end
Approximate the proportion of active sensors:
Approximate $\frac{m}{N}$ according to (28) or (30)
Return: Estimated value $\hat{s} = (\frac{m}{N})_{approximation} W$.

VI. SIMULATION RESULTS

Simulation results are presented in this section. The simulation parameters are set as Table I. The proposed blind estimation (BE) scheme performs in the two cases that are the special case and the general case. In the special case, the channels from the sensors to the FC are Gaussian channels (GC) whose the mean and the variance are random in every estimation. In the general case, the channels are unknown channels (UC). To simplify the simulation, we sample channel coefficients from a uniform distribution in the general case.

First, we give the GMM approximate performance in the special and the general case, respectively. Then we evaluate the relation between the number of sensors and the estimation performance. Next, we illustrate the impact of quantization on the estimation performance. Finally, we evaluate the performance of the BE scheme with varying values of received SNR. The results are compared to the MLE with perfect CSI in (3) and k-means clustering algorithm. The MLE operates directly on the analog measurements and over a Gaussian channel whose the mean and the variance are random in every estimation. The cluster number of K-means is set to 2.

TABLE 1. Simulation parameters settings.

Parameters	Values
Quantization Gain: a	10
Measured SNR: s^2/σ_n^2	20 dB
Received SNR: y^2/σ_w^2	0 - 26 dB
sensing value Range: $(0, W)$	0 - 2

A. GMM APPROXIMATE PERFORMANCE

Consider a scenario with following specific parameters value: $N = 700$, received SNR $= 10$ dB, and the values of other parameters are given in Table I. In the special case, we apply a two-component GMM to approximate the distribution of the received signals. Fig. 6(a) brings out the normalized histogram of the received signals and the curve of the trained two-component GMM. Note that the curve closely approximates the normalized histogram. Then the simulation is extended to the general case. A three-component GMM is trained to approximate the distribution of the received signals. The simulation results of Fig. 6(b) reveal that the

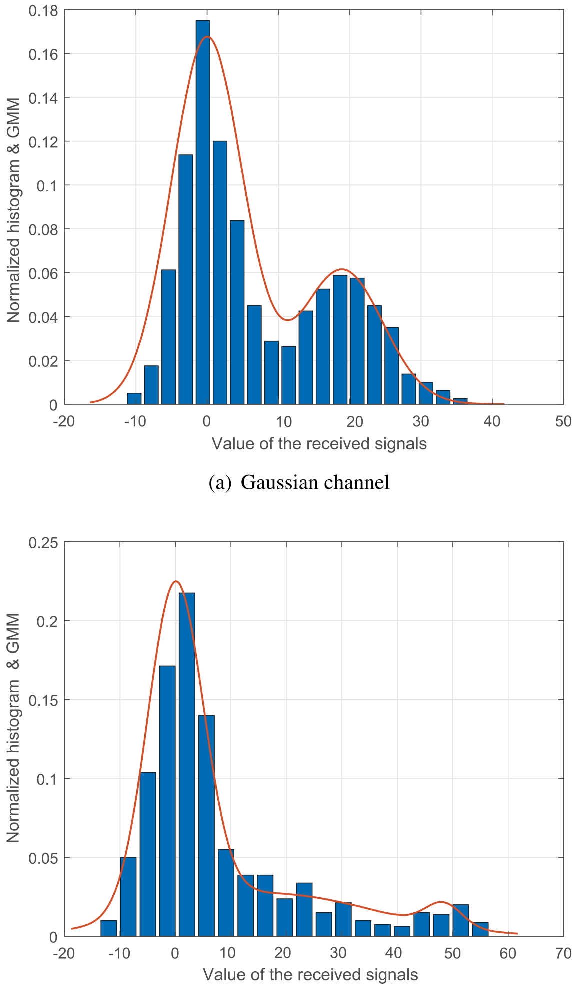


FIGURE 6. GMM approximation v.s. Normalized histogram of received signals.

three-component GMM can approximate the distribution of received signals when the channel model is unknown, where the means of three Gaussian components are around 0, 25 and 50, respectively.

B. EFFECT OF THE NUMBER OF SENSORS

The relation between the number of sensors N and the system performance is evaluated in this subsection. Given a target value of the estimation error σ and a threshold probability C, we get the condition that N should satisfy according to Proposition 2. In the simulation, σ is set to 0.05 and 0.1, and C is set to 0.7, 0.75, 0.8, 0.85, 0.9, 0.95, and the sensing value s is set to 1.6, and the received SNR is set to 15dB. The required minimum values of N are given in Fig. 7 whose legend is "Theory in (6)". Based on theoretical minimum N, we run the blind estimation process 5000 times to obtain a series of simulation probability, where true $\frac{m}{M}$ and approximate $\frac{m}{M}$ ((28) and (30)) are used, respectively. From the figure, we see when utilizing the true $\frac{m}{M}$, the curve of simulation is close to

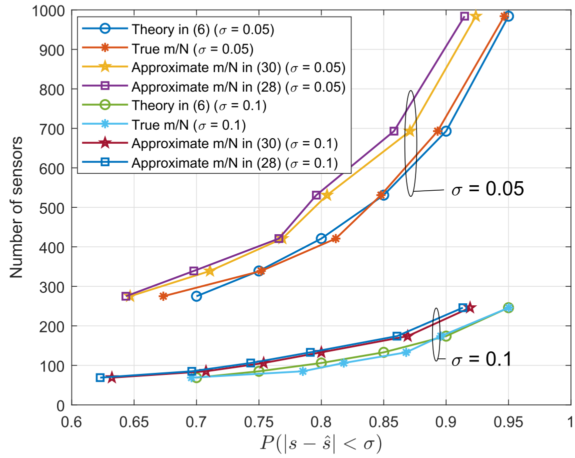


FIGURE 7. Relation between the number of sensor and system performance.

the theoretical curve, whereas there is a gap when utilizing the approximate $\frac{m}{M}$. The gap indicates that more sensors are required to achieve the theoretical performance. When $\sigma = 0.1$, it requires about 30 additional sensors to approach the theoretical performance, and when $\sigma = 0.05$, it requires about 100 additional sensors. Moreover, we can see that the approximate $\frac{m}{M}$ (30) outperforms the approximate $\frac{m}{M}$ (28).

C. IMPACT OF QUANTIZATION

As mentioned in Remark 3, the smaller estimation bound Ω indicates better estimation performance. In this subsection, we give some numerical results shown in Fig. 8 to evaluate the impact of quantization, where $\tilde{\beta}^*$ is set to 1, N is set to 200, and θ is set to 0.1, 0.5, 0.85. From the figure, we see that Ω declines rapidly as M increases. Besides, when $M < 6$, the values of Ω are significantly different for the same M, and when $M \geq 6$, the difference becomes negligible.

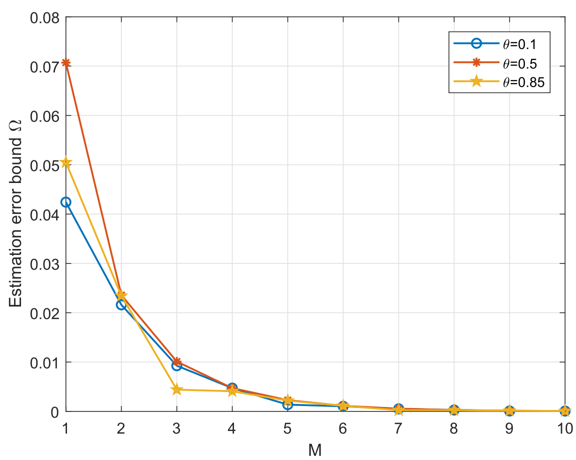


FIGURE 8. Estimation error bound Ω v.s. quantization bit M.

D. EFFECT OF RECEIVED SNR

The effect of received SNR is studied in this subsection. First, the performance metric is the average estimation error denoted as

$$\mathrm{Avg}\left[\left|s-\hat{s}\right|\right]=\frac{\sum_{i=1}^{n}\left|s_i-\hat{s}_i\right|}{n}, \quad (31)$$

where n is the total time of estimations. Fig. 9 presents the curves of the average estimation error versus the received SNR with different N and algorithms. The results show that at the low SNR region, as the SNR increases the average estimation error decreases obviously utilizing BE schemes. When the SNR increases (exceeds about 15dB in the special case and 20dB in the general case), the performance of BE can not be improved by increasing the received SNR. We infer that when the received SNR is large enough, the performance bottleneck of the BE scheme is no longer the noise but the number of sensors. Besides, we notice that when the SNR is low (below about 3dB), the performance of BE (GMM-EM) under the unknown channel is better than that under the Gaussian channel since at the low SNR region, a K-component ($K > 2$) GMM outperforms a two-component GMM when approximating the distribution of received signals. With perfect CSI and analog transmission, the MLE outperform the BE, and the average estimation error of the MLE obviously degrades as the received SNR increase. Moreover, we see that the blind estimation utilizing GMM-EM outperforms that utilizing K-means since the soft assignment clustering is more applicable to the distributed estimation problem and GMM-EM algorithm has the ability to cluster unbalanced data set.

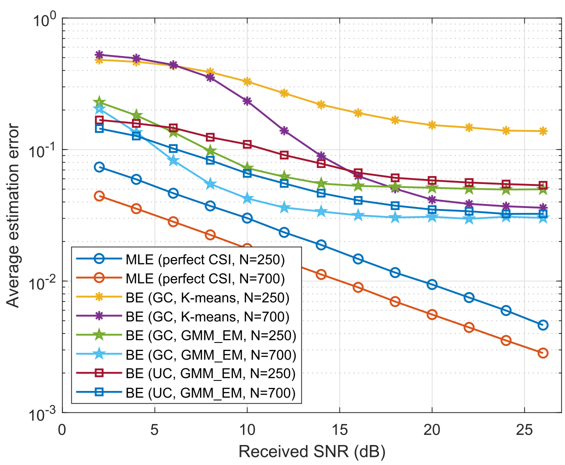


FIGURE 9. **Average estimation error v.s. received SNR.**

VII. CONCLUSION AND FUTURE WORK

The work has presented a novel distributed blind estimation scheme based on statistical inference method. The blind estimation scheme which consists of the random transmission and the statistical inference has made it possible to estimate a sensing value without any channel information. By the random transmission strategy, we have turned the sensing value into a statistical parameter to govern the sensor states which are indicated by the quantized symbols. A unified estimation frame has been proposed based on the GMM-EM statistical inference method to recover the sensing value. To evaluate the proposed blind estimation performance, we investigate the impact of the number of sensors and the quantization. Simulation results have demonstrated the estimation accuracy of the proposed blind estimation method grows as the increasing numbers of sensors and received SNR. The blind estimation scheme has significantly reduced operational complexity and energy consumption for an individual sensor. Future work has the following directions: (1) employing non-uniform or adaptive quantizer for random transmission; (2) extending scalar estimation to vector estimation; (3) power control in the blind estimation.

APPENDICES A

Here, we give the unified proof of Proposition 1 and Proposition 3.

Consider a M-bits random transmission, where $M = 1, 2, 3, \ldots$. The quantization resolution is $\Delta = 1/\left(2^M - 1\right)$. By the step 1 of the random transmission, the normalized measurements are expressed as

$$\theta_i=\frac{x_i}{W}=\frac{s}{W}+\frac{n_i}{W}, \quad i=1,\ldots,N. \quad (32)$$

Since the noise $\{n_i,\ i=1,\ldots,N\}$ follow i.i.d. $\mathcal{N}\left(0,\sigma_n^2\right)$, the normalized measurements $\{\theta_i,\ i=1,\ldots,N\}$ follow $\mathcal{N}\left(\frac{s}{W},\frac{\sigma_n^2}{W^2}\right)$.

By the step 3, the distribution of i-th quantized symbol is parameterized by the normalized measurement value θ_i as follows

$$b_i=\begin{cases} a(t+1)\Delta, & \text{w.p } \phi_i \\ at\Delta, & \text{w.p } 1-\phi_i \end{cases}, \quad i=1,\ldots,N, \quad (33)$$

where a is the quantization gain, $\phi_i = \left(2^M-1\right)\theta_i - \left\lfloor\left(2^M-1\right)\theta_i\right\rfloor$ where $\phi_i \in (0,1), \forall i$ and $\lfloor\cdot\rfloor$ is the round down operator. t is determined by the interval in which θ_i falls, that is $t\Delta \le \theta_i \le (t+1)\Delta,\ t \in \left\{0,1,\ldots,2^M-2\right\}$. For the special case of $M=1$, there are $\phi_i=\theta_i$ and $t=0$.

Let $X_i = 1$ represent $b_i = a(t+1)\Delta$ and $X_i = 0$ represent $b_i = at\Delta$. Thus X_i is a Bernoulli trial with success probability ϕ_i. Then the number of active sensors is $m = \sum_{i}^{N} X_i$. When there is no measure noise, the expectations of $\{X_i, i=1,\ldots,N\}$ are given as

$$\mathbb{E}\left[X_i\right]=\phi=\left(2^M-1\right)\theta-\left\lfloor\left(2^M-1\right)\theta\right\rfloor, \quad i=1,\ldots,N, \quad (34)$$

where θ is the normalized measurement when there is no noise. Since $\{X_i, i=1,\ldots,N\}$ are independent, the expectation of m is derived as

$$\begin{aligned}\mathbb{E}[m] &= \mathbb{E}\left[X_1\right]+\ldots+\mathbb{E}\left[X_N\right] \\ &= N\cdot\left\{\left(2^M-1\right)\theta-\left\lfloor\left(2^M-1\right)\theta\right\rfloor\right\}. \end{aligned} \quad (35)$$

According to Bernoulli's Law of Large Numbers, for every $\varepsilon > 0$ we have

$$\lim_{N\to\infty} P\left\{\left|\frac{m-\mathbb{E}[m]}{N}\right| < \varepsilon\right\} \to 1$$
$$\Rightarrow \lim_{N\to\infty} P\left\{\left|\frac{m}{N}-\phi\right| < \varepsilon\right\} \to 1 \quad (36)$$

When the measure noise exists, the probabilities of activating the sensors are different. The expectations of $\{X_i, i = 1, \ldots, N\}$ are

$$\mathbb{E}[X_i]=\phi_i=\left(2^M-1\right)\theta_i-\left\lfloor\left(2^M-1\right)\theta_i\right\rfloor, \quad i = 1, \ldots, N. \quad (37)$$

Then the expectation of m is given by

$$\mathrm{E}[m] = \sum_{i=}^{N}\phi_i$$
$$= N\cdot\phi+\left(2^M-1\right)\frac{\sum_{i=1}^{N} n_i}{W}-\left\lfloor\left(2^M-1\right)\frac{\sum_{i=1}^{N} n_i}{W}\right\rfloor \quad (38)$$

Utilizing Bernoullis Law of Large Numbers again, for every $\varepsilon > 0$ we have

$$\lim_{N\to\infty} P\left\{\left|\frac{m-\mathrm{E}[m]}{N}\right| < \varepsilon\right\} \to 1$$
$$\Rightarrow \lim_{N\to\infty} P\left\{\left|\frac{m}{N}-\phi-\frac{\zeta}{NW}-\left\lfloor\frac{\zeta}{NW}\right\rfloor\right| < \varepsilon\right\} \to 1, \quad (39)$$

where $\zeta = \left(2^M-1\right)\sum_{i=1}^{N} n_i$. When $N \to \infty$, $\frac{\zeta}{NW} \to 0$. Then (39) can be further derived as

$$\lim_{N\to\infty} P\left\{\left|\frac{m}{N}-\phi\right| < \varepsilon\right\} \to 1 \quad (40)$$

The results of (36) and (40) reveal that when N is large enough, the proportion of the active sensors $\frac{m}{N}$ is close to ϕ, denoted as

$$\lim_{N\to\infty}\frac{m}{N} \to \phi. \quad (41)$$

Then we have

$$\hat{\phi}=\frac{m}{N} \Rightarrow \left(2^M-1\right)\hat{\theta}-\left\lfloor\left(2^M-1\right)\hat{\theta}\right\rfloor=\frac{m}{N} \quad (42)$$

Since θ is within $(t\Delta, (t+1)\Delta)$, then we have

$$\hat{\theta}=\left(\frac{m}{N}+t\right)\Delta=\frac{m/N+t}{2^M-1} \quad (43)$$

Then the estimated value is given by

$$\hat{s}=\hat{\theta}\cdot W=\frac{m/N+t}{2^M-1}\cdot W \quad (44)$$

Notice that when $M = 1$, there is $t = 0$, then we have

$$\hat{s}=\frac{m\cdot W}{N} \quad (45)$$

APPENDIX B

Given a target value σ of the average estimation error and the threshold probability C, we have

$$\left\{\left|\hat{s}-s\right| < \sigma\right\} \geq C$$
$$\Rightarrow P\left\{\left|\frac{mW}{N}-\theta W\right| < \sigma\right\} \geq C$$
$$\Rightarrow P\left(\left|\frac{m}{N}-\theta\right| < \frac{\sigma}{W}\right\} \geq C \quad (46)$$

According to Central Limit Theorem, for any given β and when $N \to \infty$, we have

$$P\left\{\frac{m-N\cdot\theta}{\sqrt{N\cdot\theta(1-\theta)}} < \beta\right\} \to \Re(\beta), \quad (47)$$

where, $\Re(\beta)$ is the standard normal distribution defined by

$$\Re(\beta)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\beta} e^{-\frac{1}{2}t^2}dt. \quad (48)$$

Consider the absolute error, we have

$$P\left\{\left|\frac{m-N\cdot\theta}{\sqrt{N\theta(1-\theta)}}\right| < \beta\right\} \to \Re(\beta)-\Re(-\beta). \quad (49)$$

Normalized form of (46) denoted as

$$P\left\{\left|\frac{m-N\cdot\theta}{\sqrt{N\theta(1-\theta)}}\right| < \frac{\sigma\sqrt{N}}{W\sqrt{\theta(1-\theta)}}\right\} \geq C \quad (50)$$

Let $\tilde{\beta}=\frac{\sigma\sqrt{N}}{W\sqrt{\theta(1-\theta)}}$, then we have

$$P\left\{\left|\frac{m-N\cdot\theta}{\sqrt{N\theta(1-\theta)}}\right| < \tilde{\beta}\right\} \to \Re\left(\tilde{\beta}\right)-\Re\left(-\tilde{\beta}\right) \quad (51)$$

For the given threshold probability C, let $\tilde{\beta}^*$ be the root of $\Re\left(\tilde{\beta}\right)-\Re\left(-\tilde{\beta}\right)=C$. In order to guarantee the probability of (51) is greater than C, $\tilde{\beta} \geq \tilde{\beta}^*$ is required. It holds that

$$\frac{\sigma\sqrt{N}}{W\sqrt{\theta(1-\theta)}} \geq \tilde{\beta}^* \quad (52)$$

Then we get

$$N > \frac{\left(W\tilde{\beta}^*\right)^2\theta(1-\theta)}{\sigma^2} \quad (53)$$

APPENDIX C

According to (52) given in Appendix B and replacing θ with ϕ, we have

$$\frac{\sigma\sqrt{N}}{W\sqrt{\phi(1-\phi)}} \geq \tilde{\beta}^*, \quad (54)$$

where $\phi=\left(2^M-1\right)\theta-\left\lfloor\left(2^M-1\right)\theta\right\rfloor$. Then we get

$$\sigma \geq \frac{W\tilde{\beta}^*\sqrt{\phi(1-\phi)}}{\sqrt{N}} \quad (55)$$

which satisfies (46) given in Appendix B. When the equality in (55) holds, we have

$$P\left\{\left|\frac{m}{N}-\phi\right|<\frac{\tilde{\beta}^*\sqrt{\phi\,(1-\phi)}}{\sqrt{N}}\right\}=C$$
$$\Rightarrow P\left\{\left|\left(\frac{m}{N}+t\right)\Delta-(\phi+t)\,\Delta\right|<\frac{\Delta\tilde{\beta}^*\sqrt{\phi\,(1-\phi)}}{\sqrt{N}}\right\}=C$$
$$\Rightarrow P\left\{\left|\hat{\theta}-\theta\right|<\frac{\Delta\tilde{\beta}^*\sqrt{\phi\,(1-\phi)}}{\sqrt{N}}\right\}=C. \quad (56)$$

Then the estimation error for multiple-bit situation satisfies

$$P\left\{\left|\hat{s}-s\right|<\frac{W\tilde{\beta}^*\sqrt{\phi\,(1-\phi)}}{(2^M-1)\sqrt{N}}\right\}=C, \quad (57)$$

where $\tilde{\beta}^*$ is the root of $\Re\left(\tilde{\beta}\right)-\Re\left(-\tilde{\beta}\right)=C$.

APPENDIX D

Firstly, we give the the updating formula of π by optimize the following problem with the constraint $\sum\limits_{k=1}^{K}\pi_k=1$,

$$\pi_k^*=\arg\max_{\pi_k} L\,(\pi)$$
$$=\arg\max_{\pi_k}\sum_{i=1}^{N}\sum_{k=1}^{K}\gamma\,(z_{ik})\log\pi_k-(\sum_{k=1}^{K}\pi_k-1),$$
$$k=1,\ldots,K. \quad (58)$$

The first partial derivative of $L\,(\pi_k)$ to π_k can be expressed as

$$\frac{\partial L\,(\pi_k)}{\partial\pi_k}=\sum_{i=1}^{N}\frac{\gamma\,(z_{ik})}{\pi_k}-N,\quad k=1,\ldots,K, \quad (59)$$

Let the first partial derivative of $L\,(\pi_k)$ is equal to zero, the optimal value of π_k solving (58) is given by

$$\pi_k^*=\frac{\sum\limits_{i=1}^{N}\gamma\,(z_{ik})}{N},\quad k=1,\ldots,K, \quad (60)$$

Then, the updating formula of the $\mu_k, k=2,\ldots,K$ is derived by considering those terms in (22) which are related to the μ_k only, the corresponding optimization problem is shown as

$$\mu_k^*=\arg\max_{\mu_k} L\,(\mu_k)$$
$$=\arg\max_{\mu_k}\sum_{i=1}^{N}\gamma\,(z_{ik})\cdot\left[-\frac{1}{2\sigma_k^2}(y_i-\mu_k)^2\right],$$
$$k=2,\ldots,K \quad (61)$$

The first partial derivative of $L\,(\mu_k)$ to μ_k can be expressed as

$$\frac{\partial L\,(\mu_k)}{\partial\mu_k}=\sum_{i=1}^{N}-\gamma\,(z_{ik})\cdot\left[\frac{(y_i-\mu_k)}{2\sigma_k^2}\right] \quad (62)$$

Let the first partial derivative of $L\,(\mu_k)$ is equal to zero, then the optimal value of μ_k solving (61)) is given by

$$\mu_k^*=\frac{\sum\limits_{i=1}^{N}\gamma\,(z_{ik})\cdot y_i}{\sum\limits_{i=1}^{N}\gamma\,(z_{ik})},\quad k=2,\ldots,K, \quad (63)$$

Next, we derive the the updating formulas of σ_k^2, $k=1,\ldots,K$ by solving the optimization problems given as follows:

$$\left(\sigma_k^2\right)^*=\arg\max_{\sigma_k^2} L\left(\sigma_k^2\right)$$
$$=\arg\max_{\sigma_k^2}\sum_{i=1}^{N}\gamma\,(z_{ik})\cdot\left[-\frac{1}{2}\log\sigma_k^2-\frac{1}{2\sigma_k^2}(y_i-\mu_k)^2\right]$$
$$k=1,\ldots,K \quad (64)$$

The first partial derivative of $L\left(\sigma_k^2\right)$ to σ_k^2 can be expressed as

$$\frac{\partial L\left(\sigma_k^2\right)}{\partial\sigma_k^2}=\sum_{i=1}^{N}\gamma\,(z_{ik})\cdot\left[-\frac{1}{2\sigma_k^2}+\frac{(y_i-\mu_k)^2}{2(\sigma_k^2)^2}\right],$$
$$k=1,\ldots,K \quad (65)$$

Let the first partial derivative of $L\left(\sigma_k^2\right)$ is equal to zero, then the optimal value of σ_k^2 solving (64) is given by

$$\left(\sigma_k^2\right)^*=\frac{\sum\limits_{i=1}^{N}\gamma\,(z_{ik})\,(y_i-\mu_k)^2}{\sum\limits_{i=1}^{N}\gamma\,(z_{ik})},\quad k=1,\ldots,K. \quad (66)$$

where $\mu_1=0$, thus

$$\left(\sigma_1^2\right)^*=\frac{\sum\limits_{i=1}^{N}\gamma\,(z_{i1})\,y_i^2}{\sum\limits_{i=1}^{N}\gamma\,(z_{i1})} \quad (67)$$

REFERENCES

[1] A. R. Jaladi, K. Khithani, P. Pawar, K. Malvi, and G. Sahoo, "Environmental monitoring using wireless sensor networks (WSN) based on IoT," *Int. Res. J. Eng. Technol*, vol. 4, no. 1, pp. 1371–1378, 2017.

[2] H. Alemdar and C. Ersoy, "Wireless sensor networks for healthcare: A survey," *Comput. Netw.*, vol. 54, no. 15, pp. 2688–2710, Oct. 2010.

[3] K. Nellore and G. P. Hancke, "A survey on urban traffic management system using wireless sensor networks," *Sensors*, vol. 16, no. 2, p. 157, 2016.

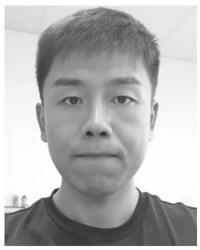
[4] J. Aponte-Luis, J. Gómez-Galán, F. Gómez-Bravo, M. Sánchez-Raya, J. Alcina-Espigado, and P. Teixido-Rovira, "An efficient wireless sensor network for industrial monitoring and control," *Sensors*, vol. 18, no. 1, p. 182, 2018.

[5] Y. Xu, V. Gupta, and C. Fischione, "Distributed estimation," in *Academic Press Library in Signal Processing*, vol. 4. Amsterdam, The Netherlands: Elsevier, 2014, pp. 675–706.

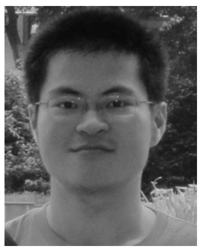
[6] W. Bajwa, A. Sayeed, and R. Nowak, "Matched source-channel communication for field estimation in wireless sensor network," in *Proc. 4th Int. Symp. Inf. Process. Sensor Netw. (IPSN)*, Apr. 2005, pp. 332–339.

[7] J. Li and G. AlRegib, "Rate-constrained distributed estimation in wireless sensor networks," *IEEE Trans. Signal Process.*, vol. 55, no. 5, pp. 1634–1643, May 2007.

[8] S. Kar, H. Chen, and P. K. Varshney, "Optimal identical binary quantizer design for distributed estimation," *IEEE Trans. Signal Process.*, vol. 60, no. 7, pp. 3896–3901, Jul. 2012.
[9] A. Vempaty, H. He, B. Chen, and P. K. Varshney, "On quantizer design for distributed Bayesian estimation in sensor networks," *IEEE Trans. Signal Process.*, vol. 62, no. 20, pp. 5359–5369, Oct. 2014.
[10] J.-J. Xiao, A. Ribeiro, Z.-Q. Luo, and G. B. Giannakis, "Distributed compression-estimation using wireless sensor networks," *IEEE Signal Process. Mag.*, vol. 23, no. 4, pp. 27–41, Jul. 2006.
[11] J. J. Xiao, S. Cui, Z. Q. Luo, and A. J. Goldsmith, "Linear coherent decentralized estimation," *IEEE Trans. Signal Process.*, vol. 56, no. 2, pp. 757–770, Feb. 2008.
[12] J.-C. Liu and C.-D. Chung, "Distributed estimation in a wireless sensor network using hybrid MAC," *IEEE Trans. Veh. Technol.*, vol. 60, no. 7, pp. 3424–3435, Sep. 2011.
[13] J. Su and H. H. Nguyen, "Distributed estimation in wireless sensor networks with semi-orthogonal MAC," *EURASIP J. Wireless Commun. Netw.*, vol. 2016, no. 1, 2016, Art. no. 214.
[14] S. Cui, J. J. Xiao, A. J. Goldsmith, Z. Q. Luo, and H. V. Poor, "Estimation diversity and energy efficiency in distributed sensing," *IEEE Trans. Signal Process.*, vol. 55, no. 9, pp. 4683–4695, Sep. 2007.
[15] H. Senol and C. Tepedelenlioglu, "Performance of distributed estimation over unknown parallel fading channels," *IEEE Trans. Signal Process.*, vol. 56, no. 12, pp. 6057–6068, Dec. 2008.
[16] M. S. Alam, G. Kaddoum, and B. L. Agba, "Performance analysis of distributed wireless sensor networks for Gaussian source estimation in the presence of impulsive noise," *IEEE Signal Process Lett.*, vol. 25, no. 6, pp. 803–807, Jun. 2018.
[17] L. Xiao, S. Boyd, and S. Lall, "A scheme for robust distributed sensor fusion based on average consensus," in *Proc. 4th Int. Symp. Inf. Process. Sensor Netw. (IPSN)*, Apr. 2005, pp. 63–70.
[18] X. Wang and C. Yang, "Optimal and suboptimal decentralized estimation underlying fading channels in WSNs," in *Proc. IEEE Global Telecommun. Conf. IEEE (GLOBECOM)*, Nov. 2007, pp. 2863–2867.
[19] T. C. Aysal and K. E. Barner, "Constrained decentralized estimation over noisy channels for sensor networks," *IEEE Trans. Signal Process.*, vol. 56, no. 4, pp. 1398–1410, Apr. 2008.
[20] M. Wang and C. Yang, "Distributed estimation in wireless sensor networks with imperfect channel estimation," in *Proc. 9th Int. Conf. Signal Process.*, Oct. 2008, pp. 2649–2652.
[21] Q. Zhou, D. Li, S. Kar, L. M. Huie, H. V. Poor, and S. Cui, "Learning-based distributed detection-estimation in sensor networks with unknown sensor defects," *IEEE Trans. Signal Process.*, vol. 65, no. 1, pp. 130–145, Jan. 2017.
[22] H. Zhu and R. Sun, "Joint detection and estimation fusion in the presence of correlated sensor quantized data," in *Proc. 20th Int. Conf. Inf. Fusion (Fusion)*, Jul. 2017, pp. 1–6.
[23] A. Sani and A. Vosoughi, "On distributed linear estimation with observation model uncertainties," *IEEE Trans. Signal Process.*, vol. 66, no. 12, pp. 3212–3227, Jun. 2018.
[24] J. Zhu, H. Cao, C. Song, and Z. Xu, "Parameter estimation via unlabeled sensing using distributed sensors," *IEEE Commun. Lett.*, vol. 21, no. 10, pp. 2130–2133, Oct. 2017.
[25] G. Wang, J. Zhu, R. S. Blum, P. Willett, S. Marano, V. Matta, and P. Braca, "Signal amplitude estimation and detection from unlabeled binary quantized samples," *IEEE Trans. Signal Process.*, vol. 66, no. 16, pp. 4291–4303, Aug. 2018.
[26] L. Peng, X. Song, M. C. Tsakiris, H. Choi, L. Kneip, and Y. Shi, "Algebraically-initialized expectation maximization for header-free communication," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2019, pp. 5182–5186.
[27] J. Wang, C. Jiang, H. Zhang, Y. Ren, K.-C. Chen, and L. Hanzo, "Thirty years of machine learning: The road to Pareto-optimal next-generation wireless networks," 2019, *arXiv:1902.01946*. [Online]. Available: https://arxiv.org/abs/1902.01946
[28] H.-W. Liang, W.-H. Chung, and S.-Y. Kuo, "Coding-aided K-means clustering blind transceiver for space shift keying MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 15, no. 1, pp. 103–115, Jan. 2016.
[29] L. You, P. Yang, Y. Xiao, S. Rong, D. Ke, and S. Li, "Blind detection for spatial modulation systems based on clustering," *IEEE Commun. Lett.*, vol. 21, no. 11, pp. 2392–2395, Nov. 2017.
[30] L. Zhang and M. Jin, "A constrained clustering-based blind detector for spatial modulation," *IEEE Commun. Lett.*, vol. 23, no. 7, pp. 1170–1173, Jul. 2019.
[31] H. S. Emadi and S. M. Mazinani, "A novel anomaly detection algorithm using DBSCAN and SVM in wireless sensor networks," *Wireless Pers. Commun.*, vol. 98, no. 2, pp. 2025–2035, 2017.
[32] G. J. McLachlan and K. E. Basford, *Mixture Models: Inference and Applications to Clustering*, vol. 84. New York, NY, USA: Marcel Dekker, 1988.
[33] B. Safarinejadian, M. B. Menhaj, and M. Karrari, "Distributed unsupervised Gaussian mixture learning for density estimation in sensor networks," *IEEE Trans. Instrum. Meas.*, vol. 59, no. 9, pp. 2250–2260, Sep. 2010.
[34] C. N. S. Kumar, K. N. Rao, A. Govardhan, and K. S. Reddy, "Imbalanced k-means: An algorithm to cluster imbalanced-distributed data," *Int. J. Eng. Tech. Res.*, vol. 2, no. 2, pp. 114–122, 2014.
[35] S. Zhu and B. Chen, "Quantized consensus by the ADMM: Probabilistic versus deterministic quantizers," *IEEE Trans. Signal Process.*, vol. 64, no. 7, pp. 1700–1713, Apr. 2016.
[36] A. T. Suresh, F. X. Yu, S. Kumar, and H. B. McMahan, "Distributed mean estimation with limited communication," in *Proc. 34th Int. Conf. Mach. Learn.*, vol. 70, 2017, pp. 3329–3337.
[37] T. K. Moon, "The expectation-maximization algorithm," *IEEE Signal Process. Mag.*, vol. 13, no. 6, pp. 47–60, Nov. 1996.
[38] N. Kostantinos, "Gaussian mixtures and their applications to signal processing," in *Advanced Signal Processing Handbook*. Boca Raton, FL, USA: CRC Press, 2000.



WENTAO ZHANG received the B.S. degree from the Department of Electronic Engineering, South China University of Technology, Canton, China, in 2015. He is currently pursuing the Ph.D. degree in communication and information engineering with the University of Science and Technology of China, Hefei. His current research interests include wireless sensor network, machine learning, and deep learning.



LI CHEN received the B.E. degree in electrical and information engineering from the Harbin Institute of Technology, Harbin, China, in 2009, and the Ph.D. degree in electrical engineering from the University of Science and Technology of China, Hefei, China, in 2014, where he is currently a Faculty Member with the Department of Electronic Engineering and Information Science. His research interests include energy efficiency in wireless communications and wireless optical communications. He received the Chinese Academy of Sciences President Award, in 2014.

MIN QIN received the B.S. degree from the Department of Electronic Engineering and Information Science (EEIS), University of Science and Technology of China (USTC), Hefei, China, in 2014, where he is currently pursuing the Ph.D. degree in communication and information engineering. His current research interests include wireless channel estimation, mobile edge computing, and optimization theory.

WEIDONG WANG received the B.S. degree from the Beijing University of Aeronautics and Astronautics, Beijing, China, in 1989, and the M.S. degree from the University of Science and Technology of China, Hefei, China, in 1993, where he is currently a Full Professor with the Department of Electronic Engineering and Information Systems. His research interests include wireless communication, microwave and millimeter, and radar technology. He is a member of the Committee of Optoelectronic Technology, Chinese Society of Astronautics.