

RESEARCH

Open Access

Decentralized estimation over orthogonal multiple-access fading channels in wireless sensor networks—optimal and suboptimal estimators

Xin Wang* and Chenyang Yang

Abstract

We study optimal and suboptimal decentralized estimators in wireless sensor networks over orthogonal multiple-access fading channels in this paper. Considering multiple-bit quantization for digital transmission, we develop maximum likelihood estimators (MLEs) with both known and unknown channel state information (CSI). When training symbols are available, we derive a MLE that is a special case of the MLE with unknown CSI. It implicitly uses the training symbols to estimate CSI and exploits channel estimation in an optimal way and performs the best in realistic scenarios where CSI needs to be estimated and transmission energy is constrained. To reduce the computational complexity of the MLE with unknown CSI, we propose a suboptimal estimator. These optimal and suboptimal estimators exploit both signal- and data-level redundant information to combat the observation noise and the communication errors. Simulation results show that the proposed estimators are superior to the existing approaches, and the suboptimal estimator performs closely to the optimal MLE.

Keywords: Decentralized estimation, maximum likelihood estimation, fading channels, wireless sensor network

1 Introduction

Wireless sensor networks (WSNs) consist of a number of sensors deployed in a field to collect information, for example, measuring physical parameters such as temperature and humidity. Since the sensors are usually powered by batteries and have very limited processing and communication abilities [1], the parameters are often estimated in a decentralized way. In typical WSNs for decentralized estimation, there exists a fusion center (FC). The sensors transmit their locally processed observations to the FC, and the FC generates the final estimation based on the received signals [2].

Both observation noise and communication errors deteriorate the performance of decentralized estimation. Traditional fusion-based estimators are able to minimize the mean square error (MSE) of the parameter estimation by assuming perfect communication links (see [3] and references therein). They reduce the observation noise by exploiting the redundant observations provided by multiple sensors. However, their performance

degrades dramatically when communication errors cannot be ignored or corrected. On the other hand, various wireless communication technologies aiming at achieving transmission capacity or improving reliability do not minimize the MSE of the parameter estimation. For example, although diversity combining reduces the bit error rate (BER), it requires that the signals transmitted from multiple sensors are identical, which is not true in the context of WSNs due to the observation noise at sensors. This motivates to optimize estimator at the FC under realistic observation and channel models, which minimizes the MSE of parameter estimation.

The bandwidth and energy constraints are two critical issues for the design of WSNs. When the strict bandwidth constraint is taken into account, the decentralized estimation when the sensors only transmit one bit for each observation, that is, using binary quantization, is studied in [4-9]. When communication channels are noiseless, a maximum likelihood estimator (MLE) is introduced and optimal quantization is discussed in [4]. A universal and isotropic quantization rule is proposed in [6], and adaptive binary quantization methods are studied in [7,8]. When channels are noisy, the MLE in

* Correspondence: athody@vip.sina.com
School of Electronics and Information Engineering, Beihang University,
Beijing 100191, China

additive white Gaussian noise (AWGN) channels is studied and several low complexity suboptimal estimators are derived in [9]. It has been found that the binary quantization is sufficient for decentralized estimation at low observation signal-to-noise ratio (SNR), but more bits are required for each observation at high observation SNR [4].

When the energy constraint and general multi-level quantizers are considered, various issues of the decentralized estimation are studied under different channels. When communications are error free, the quantization at the sensors is designed in [10-12]. The optimal trade-off between the number of active sensors and the quantization bit rate of each sensor is investigated under total energy constraint in [13]. In binary symmetrical channels (BSCs), the power scheduling is proposed to reduce the estimation MSE when the best linear unbiased estimator (BLUE) and a quasi-BLUE, where quantization noise is taken into account, are used at the FC [14]. Nonetheless, to the best of the authors' knowledge, the optimal decentralized estimator using multiple-bit quantization in fading channels is still unavailable. Although the MLE proposed in AWGN channels [9] can be applied for fading channels if the channel state information (CSI) is known at the FC, it only considers binary quantization.

Besides the decentralized estimation based on digital communications, the estimation based on analog communications receives considerable attentions due to the important conclusions drawn from the studies for the multi-terminal coding problem [15,16]. The most popular scheme is amplify-and-forward (AF) transmission, which is proved to be optimal in quadratic Gaussian sensor networks under multiple-access channels (MACs) with AWGN [17]. The power scheduling and energy efficiency of AF transmission are studied under AWGN channels in [18], where AF transmission is shown to be more energy efficient than digital communications. However, in fading channels, AF transmission is no longer optimal in orthogonal MACs [19-21]. The outage laws of the estimation diversity with AF transmission in fading channels are studied in [20] and [21] in different asymptotic regimes. These studies, especially the results in [19], indicate that the separate source-channel coding scheme is optimal in fading channels with orthogonal multiple-access protocols, which outperforms AF transmission, a simple joint source-channel coding scheme.

In this paper, we develop optimal and suboptimal decentralized estimators for a deterministic parameter considering digital communication. The observations of the sensors are quantized, coded and modulated, and then transmitted to the FC over Rayleigh fading orthogonal MACs. Because the binary quantization is only

applicable at low observation SNR levels [4,13], a general multi-bit quantizer is considered.

We strive for deriving MLEs and feasible suboptimal estimator when different local processing and communication strategies are used. To this end, we first present a general message function to represent various quantization and transmission schemes. We then derive the MLE for an unknown parameter with known CSI at the FC.

In typical WSNs, the sensors usually cannot transmit too many training symbols for the receiver to estimate channel coefficients because of both energy and bandwidth constraints. Therefore, we will consider realistic scenarios that the CSI is unknown at the FC when no or only a few training symbols are available. It is known that channel information has a large impact on the structure and the performance of decentralized estimation. In orthogonal MACs, most of the existing works assume that perfect CSI is available at the FC. Recently, the impact of channel estimation errors on the decentralized detection in WSNs is studied in [22], and its impact on the decentralized estimation when using AF transmission is investigated in [23]. However, the decentralized estimation with unknown CSI for digital communications has still not been well understood.

Our contributions are summarized as follows. We develop the decentralized MLEs with known and unknown CSI at the FC over orthogonal MACs with Rayleigh fading. The performance of the MLE with known CSI can serve as a practical performance lower bound of the decentralized estimation, whereas the MLE with unknown CSI is more realistic. For the special cases of error-free communications or noiseless observations, we show that the MLEs degenerate into the well-known centralized fusion estimator—BLUE—or a maximal ratio combiner (MRC)-based estimator when CSI is known and a subspace-based estimator when CSI is unknown. This indicates that our estimators exploit both data-level redundancy and signal-level redundancy provided by multiple sensors. To provide feasible estimator with affordable complexity, we propose a suboptimal algorithm, which can be viewed as a modified expectation-maximization (EM) algorithm [24].

The rest of the paper is organized as follows. Section 2 describes the system models. Section 3 presents the MLEs with known and unknown CSI and their special cases, and Section 4 introduces the suboptimal estimator. In Section 5, we analyze the asymptotic performance and complexity of the presented MLEs and discuss the codebook issue. Simulation results are provided in Section 6, and the conclusions are given in Section 7.

2 System model

We consider a typical kind of WSNs that consists of N sensors and a FC to measure an unknown deterministic parameter θ , where there are no inter-sensor communications among the sensors. The sensors process their observations for the parameter θ before transmission. For digital communications, the processing includes quantization, channel coding and modulation. For analog communications, the processing may simply be amplifying the observations before transmission. A messaging function $\mathbf{c}(x)$ is used to describe the local processing. Though we can use $\mathbf{c}(x)$ for both digital and analog communication systems, we focus on digital transmission since the popular analog transmission scheme, AF, has been shown to be not optimal in fading channels [19-21].

2.1 Observation model

The observation for the unknown parameter provided by the i th sensor is

$$x_i = \theta + n_{s,i}, \quad i = 1, \dots, N, \quad (1)$$

where $n_{s,i}$ is the independent and identically distributed (i.i.d.) Gaussian observation noise with zero mean and variance σ_s^2 , and θ is bounded within a dynamic range $[-V, +V]$.

2.2 Quantization, coding, and modulation

We use the messaging function $\mathbf{c}(x)|\mathbb{R} \rightarrow \mathbb{C}^L$ to represent all the processing at the sensors including quantization, coding and modulation, which maps the observations to the transmit symbols. To facilitate analysis, the energy of the transmit symbols is normalized to 1, that is,

$$\mathbf{c}(x)^H \mathbf{c}(x) = 1, \quad \forall x \in \mathbb{R}. \quad (2)$$

We consider uniform quantization by regarding θ as a uniformly distributed parameter. Uniform quantization is the Lloyd-Max quantizer that minimizes the quantization distortion of uniformly distributed sources [25,26]. For an M -level uniform quantizer, define the dynamic range of the quantizer as $[-W, +W]$, and then all the possible quantized values of the observations can be written as

$$S_m = m\Delta - W, \quad m = 0, \dots, M - 1, \quad (3)$$

where $\Delta = 2W/(M - 1)$ is the quantization interval.

The observations are rounded to the nearest S_m , so that $\mathbf{c}(x)$ is a piecewise constant function described as

$$\mathbf{c}(x) = \begin{cases} \mathbf{c}_0, & -\infty < x \leq S_0 + \frac{\Delta}{2} \\ \mathbf{c}_m, & S_m - \frac{\Delta}{2} < x \leq S_m + \frac{\Delta}{2}, \\ \mathbf{c}_{M-1}, & S_{M-1} - \frac{\Delta}{2} < x < +\infty \end{cases} \quad (4)$$

where $\mathbf{c}_m = [c_{m,1}, \dots, c_{m,L}]^T$ is the L symbols corresponding to the quantized observation S_m , $m = 0, \dots, M - 1$.

Under the assumption that W is much larger than the dynamic range of θ , the probability that $|x_i| > W$ can be ignored. Then, $\mathbf{c}(x)$ is simplified as

$$\mathbf{c}(x) = \mathbf{c}_m, \quad S_m - \frac{\Delta}{2} < x \leq S_m + \frac{\Delta}{2}. \quad (5)$$

Define the transmission codebook as

$$\mathbf{C}_t = [\mathbf{c}_0, \dots, \mathbf{c}_{M-1}] \in \mathbb{C}^{L \times M}, \quad (6)$$

which can be used to describe any coding and modulation scheme following the M -level quantization.

The sensors can use various codes such as natural binary codes to represent the quantized observations. In this paper, our focus is to design decentralized estimators; therefore, we will not address the transmission codebook optimization for parameter estimation.

2.3 Received signals

Since we consider orthogonal MACs, the FC can perfectly separate and synchronize to the received signals from different sensors. Assume that the channels are block fading, that is, the channel coefficients are invariant during the period that the sensors transmit L symbols representing one observation. After matched filtering and symbol-rate sampling, the L received samples corresponding to the L transmitted symbols from the i th sensor can be expressed as

$$\mathbf{y}_i = \sqrt{\mathcal{E}_d} \mathbf{h}_i \mathbf{c}(x_i) + \mathbf{n}_{c,i}, \quad i = 1, \dots, N, \quad (7)$$

where $\mathbf{y}_i = [y_{i,1}, \dots, y_{i,L}]^T$, \mathbf{h}_i is the channel coefficient, which is i.i.d. and subjected to complex Gaussian distribution with zero mean and unit variance, $\mathbf{n}_{c,i}$ is a vector of thermal noise at the receiver subjecting to complex Gaussian distribution with zero mean and covariance matrix $\sigma_c^2 \mathbf{I}$, and \mathcal{E}_d is the transmission energy for each observation.

3 Optimal estimators with or without CSI

In this section, we derive MLEs when CSI is known or unknown at the receiver of the FC, respectively. To understand how they deal with both the communication errors and the observation noises, we study two special cases. The MLE using training symbols in the transmission codebook is also studied as a special form of the MLE with unknown CSI.

3.1 MLE with known CSI

Given θ , the received signals from different sensors are statistically independent. If the CSI is known at the receiver of the FC, the log-likelihood function is

$$\begin{aligned} \log p(\mathbf{Y}|\mathbf{h}, \theta) &= \sum_{i=1}^N \log p(y_i|h_i, \theta) \\ &= \sum_{i=1}^N \log \left(\int_{-\infty}^{+\infty} p(y_i|h_i, x)p(x|\theta) dx \right), \end{aligned} \quad (8)$$

where $\mathbf{Y} = [y_1, \dots, y_N]$, $\mathbf{h} = [h_1, \dots, h_N]^T$ is the channel coefficients vector, and $p(x|\theta)$ is the conditional probability density function (PDF) of the observation given θ . Following the observation model shown in (1), we have

$$p(x|\theta) = \frac{1}{\sqrt{2\pi}\sigma_s} \exp\left(-\frac{(x-\theta)^2}{2\sigma_s^2}\right). \quad (9)$$

According to the received signal model shown in (7), the PDF of the received signals given CSI and the observation of the sensors is

$$p(y_i|h_i, x) = \frac{1}{(\pi\sigma_c^2)^L} \exp\left(-\frac{\|y_i - \sqrt{\mathcal{E}_d}h_i\mathbf{c}(x)\|_2^2}{\sigma_c^2}\right) \quad (10)$$

where $\|z\|_2 = (\mathbf{z}^H\mathbf{z})^{1/2}$ is l_2 norm of vector \mathbf{z} .

Substituting (9) and (10) into (8), we obtain the log-likelihood function for estimating θ , which can be used for any messaging function $\mathbf{c}(x)$, no matter when it describes analog or digital communications.

For digital communications, $\mathbf{c}(x)$ is a piecewise constant function as shown in (4). To simplify the analysis, we use its approximate form shown in (5) in the rest of this paper. After substituting (5) into (10) and then to (8), we have

$$\log p(\mathbf{Y}|\mathbf{h}, \theta) = \sum_{i=1}^N \log \left(\sum_{m=0}^{M-1} p(y_i|h_i, \mathbf{c}_m)p(S_m|\theta) \right), \quad (11)$$

where $p(y_i|h_i, \mathbf{c}_m)$ is the PDF of the received signals given the CSI and the transmitted symbols of the sensors, which is

$$p(y_i|h_i, \mathbf{c}_m) = \frac{1}{(\pi\sigma_c^2)^L} \exp\left(-\frac{\|y_i - \sqrt{\mathcal{E}_d}h_i\mathbf{c}_m\|_2^2}{\sigma_c^2}\right) \quad (12)$$

and $p(S_m|\theta)$ is the probability mass function (PMF) of the quantized observation given θ , which is

$$p(S_m|\theta) = Q\left(\frac{S_m - \frac{\Delta}{2} - \theta}{\sigma_s}\right) - Q\left(\frac{S_m + \frac{\Delta}{2} - \theta}{\sigma_s}\right), \quad (13)$$

where $Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty \exp\left(-\frac{t^2}{2}\right) dt$.

The MLE is obtained by maximizing the log-likelihood function shown in (11).

3.1.1 Special case when $\sigma_s^2 \rightarrow 0$

When the observation SNR tends to infinity, the observations of the sensors are perfect, that is, $x_i = \theta, \forall i = 1, \dots, N$. The PDF of the observation x_i given θ degrades to

$$p(x|\theta) = \delta(x - \theta), \quad (14)$$

where $\delta(x)$ is the Dirac-delta function.

In this case, the log-likelihood function for both analog and digital communications has the same form, which can be obtained by substituting (14) into (8). After ignoring all terms that do not affect the estimation, the log-likelihood function is simplified as

$$\log p(\mathbf{Y}|\mathbf{h}, \theta) = -\sum_{i=1}^N \frac{\|y_i - \sqrt{\mathcal{E}_d}h_i\mathbf{c}(\theta)\|_2^2}{\sigma_c^2}, \quad (15)$$

where $\mathbf{c}(\theta)$ is the transmitted symbols when the observations of the sensors are θ .

For digital communications, $\mathbf{c}(\theta)$ is a code word of \mathbf{C}_t and is a piecewise constant function. Therefore, we cannot get θ by taking partial derivative of (15). Instead, we first regard $\mathbf{c}(\theta)$ as the parameter to be estimated and obtain the MLE for estimating $\mathbf{c}(\theta)$. Then, we use it as a decision variable to detect the transmitted symbols and reconstruct θ according to the quantization rule with the decision results.

The log-likelihood function in (15) is concave with respect to (w.r.t.) $\mathbf{c}(\theta)$, and its only maximum is obtained by solving the equation $\partial \log p(\mathbf{Y}|\mathbf{h}, \theta)/\partial \mathbf{c}(\theta) = 0$, which is

$$\hat{\mathbf{c}}(\theta) = \frac{1}{\sqrt{\mathcal{E}_d} \sum_{j=1}^N |h_j|^2} \sum_{i=1}^N h_i^* y_i. \quad (16)$$

It follows that when the observations are perfect, the structure of the MLE is the MRC concatenated with data demodulation and parameter reconstruction. This is no surprise since in this case, all the signals transmitted by different sensors are identical; thus, the receiver at the FC is able to apply the conventional diversity technology to reduce the communication errors.

3.1.2 Special case when $\sigma_c^2 \rightarrow 0$

When the communications are perfect, $\mathbf{y}_i = \sqrt{\mathcal{E}_d}h_i\mathbf{c}_{m_i}$. It means that \mathbf{y}_i merely depends on \mathbf{c}_{m_i} or equivalently depends on S_{m_i} . Then, the log-likelihood function becomes a function of the quantized observation S_{m_i} .

The log-likelihood function with perfect communications becomes

$$\log p(\mathbf{Y}|\mathbf{h}, \theta) \rightarrow \log p(\mathbf{S}|\mathbf{h}, \theta) = \sum_{i=1}^N \log \left(Q\left(\frac{S_{m_i} - \frac{\Delta}{2} - \theta}{\sigma_s}\right) - Q\left(\frac{S_{m_i} + \frac{\Delta}{2} - \theta}{\sigma_s}\right) \right), \quad (17)$$

where $\mathbf{S} = [S_{m_1}, \dots, S_{m_N}]^T$.

By taking the derivative of (17) to be 0, we obtain the likelihood equation

$$\sum_{i=1}^N \frac{\exp\left(-\frac{(S_{m_i} - \frac{\Delta}{2} - \theta)^2}{2\sigma_s^2}\right) - \exp\left(-\frac{(S_{m_i} - \frac{\Delta}{2} + \theta)^2}{2\sigma_s^2}\right)}{Q\left(\frac{S_{m_i} - \frac{\Delta}{2} - \theta}{\sigma_s}\right) - Q\left(\frac{S_{m_i} - \frac{\Delta}{2} + \theta}{\sigma_s}\right)} = 0 \quad (18)$$

Generally, this likelihood equation has no closed-form solution. Nonetheless, the closed-form solution can be obtained when the quantization noise is very small, that is, $\Delta \rightarrow 0$. Under this condition, $S_{m_i} \rightarrow x_i$ and (18) becomes

$$\lim_{\Delta \rightarrow 0} \frac{\partial \log p(\mathbf{S}|\mathbf{h}, \theta)}{\partial \theta} = \sum_{i=1}^N \frac{x_i - \theta}{\sigma_s^2} = 0. \quad (19)$$

The MLE obtained from (19) is

$$\hat{\theta} = \frac{1}{N} \sum_{i=1}^N x_i. \quad (20)$$

It is also no surprise to see that the MLE reduces to BLUE, which is often applied in centralized estimation [14], where the FC can obtain all raw observations of the sensors.

3.2 MLE with unknown CSI

In practical WSNs, the FC usually has no CSI, and the sensors can transmit training symbols to facilitate channel estimation. The training symbols can be incorporated into the message function $\mathbf{c}(x)$. Then, the MLE with training symbols available is a special form of the MLE with unknown CSI. We will derive the MLE with unknown CSI with general $\mathbf{c}(x)$ in the following and derive that with training symbols in $\mathbf{c}(x)$ in next subsection.

When CSI is unknown at the FC, the log-likelihood function is

$$\log p(\mathbf{Y}|\theta) = \sum_{i=1}^N \left(\log \left(\int_{-\infty}^{+\infty} p(\mathbf{y}_i|x) p(x|\theta) dx \right) \right), \quad (21)$$

which has a similar form to the likelihood function with known CSI shown in (8).

According to the received signal model, given x , \mathbf{y}_i subjects to zero mean complex Gaussian distribution, that is,

$$p(\mathbf{y}_i|x) = \frac{1}{\pi^L \det \mathbf{R}_y} \exp(-\mathbf{y}_i^H \mathbf{R}_y^{-1} \mathbf{y}_i), \quad (22)$$

where \mathbf{R}_y is the covariance matrix of \mathbf{y}_i , which is

$$\mathbf{R}_y = \sigma_c^2 \mathbf{I} + \mathcal{E}_d \mathbf{c}(x) \mathbf{c}(x)^H. \quad (23)$$

Since the energy of the transmit symbols is normalized as shown in (2), we have

$$\mathbf{R}_y \mathbf{c}(x) = (\sigma_c^2 \mathbf{I} + \mathcal{E}_d \mathbf{c}(x) \mathbf{c}(x)^H) \mathbf{c}(x) = (\sigma_c^2 + \mathcal{E}_d) \mathbf{c}(x). \quad (24)$$

Therefore, $\mathbf{c}(x)$ is an eigenvector of \mathbf{R}_y , and the corresponding eigenvalue is $(\sigma_c^2 + \mathcal{E}_d)$.

For any vector orthogonal to $\mathbf{c}(x)$, denoted as $\mathbf{c}^\perp(x)$, we have

$$\mathbf{R}_y \mathbf{c}^\perp(x) = (\sigma_c^2 \mathbf{I} + \mathcal{E}_d \mathbf{c}(x) \mathbf{c}(x)^H) \mathbf{c}^\perp(x) = \sigma_c^2 \mathbf{c}^\perp(x). \quad (25)$$

Therefore, the eigenvalues corresponding to the remaining $L - 1$ eigenvectors are all σ_c^2 . The determinant of \mathbf{R}_y is

$$\det \mathbf{R}_y = (\mathcal{E}_d + \sigma_c^2) \sigma_c^{2(L-1)}. \quad (26)$$

Following the Matrix Inversion Lemma [27], we have

$$\mathbf{R}_y^{-1} = \frac{1}{\sigma_c^2} \mathbf{I} - \frac{\mathcal{E}_d}{\sigma_c^2 (\mathcal{E}_d + \sigma_c^2)} \mathbf{c}(x) \mathbf{c}(x)^H. \quad (27)$$

Substituting (26) and (27) into (22), we have

$$p(\mathbf{y}_i|x) = \alpha \exp \left(-\frac{\|\mathbf{y}_i\|_2^2}{\sigma_c^2} + \frac{\mathcal{E}_d \mathbf{y}_i^H \mathbf{c}(x) \mathbf{c}(x)^H \mathbf{y}_i}{\sigma_c^2 (\mathcal{E}_d + \sigma_c^2)} \right) = \alpha \exp \left(-\frac{\|\mathbf{y}_i\|_2^2}{\sigma_c^2} + \frac{\mathcal{E}_d |\mathbf{y}_i^H \mathbf{c}(x)|^2}{\sigma_c^2 (\mathcal{E}_d + \sigma_c^2)} \right), \quad (28)$$

where α is a constant.

Upon substituting (28) and (9) into (21), the log-likelihood function becomes

$$\log p(\mathbf{Y}|\theta) = \sum_{i=1}^N \log \left(\int_{-\infty}^{+\infty} \exp \left(-\frac{(x-\theta)^2}{2\sigma_s^2} - \frac{\|\mathbf{y}_i\|_2^2}{\sigma_c^2} + \frac{\mathcal{E}_d |\mathbf{y}_i^H \mathbf{c}(x)|^2}{\sigma_c^2 (\mathcal{E}_d + \sigma_c^2)} \right) dx \right). \quad (29)$$

Then the MLE is obtained as

$$\hat{\theta} = \arg \max_{\theta} \log p(\mathbf{Y}|\theta). \quad (30)$$

When considering digital communications, by substituting (5) into (29), the log-likelihood function is obtained as

$$\log p(\mathbf{Y}|\theta) = \sum_{i=1}^N \log \left(\sum_{m=1}^M p(\mathbf{y}_i|\mathbf{c}_m) p(S_{m_i}|\theta) \right), \quad (31)$$

where $p(S_m|\theta)$ is shown in (13), and

$$p(\mathbf{y}_i|\mathbf{c}_m) = \alpha \exp\left(-\frac{\|\mathbf{y}_i\|_2^2}{\sigma_c^2} + \frac{\mathcal{E}_d|\mathbf{y}_i^H\mathbf{c}_m|^2}{\sigma_c^2(\mathcal{E}_d + \sigma_c^2)}\right). \quad (32)$$

3.2.1 Special case when $\sigma_s^2 \rightarrow 0$

Similarly to the log-likelihood function with known CSI, the log-likelihood function with unknown CSI for perfect observations has the same form for both analog and digital communications.

Upon substituting (14) into (21) and ignoring all terms that do not affect the estimation, the log-likelihood function becomes

$$\begin{aligned} \log p(\mathbf{Y}|\theta) &= \sum_{i=1}^N |\mathbf{y}_i^H \mathbf{c}(\theta)|^2 \\ &= \mathbf{c}(\theta)^H \left(\sum_{i=1}^N \mathbf{y}_i \mathbf{y}_i^H \right) \mathbf{c}(\theta). \end{aligned} \quad (33)$$

Again, since $\mathbf{c}(\theta)$ is underivable for digital communications, we regard $\mathbf{c}(\theta)$ as the parameter to be estimated. Recall that the energy of $\mathbf{c}(\theta)$ is normalized. Then, the problem that finds $\mathbf{c}(\theta)$ to maximize (33) is a solvable quadratically constrained quadratic program (QCQP) [28]:

$$\begin{aligned} \max_{\mathbf{c}(\theta)} & \mathbf{c}(\theta)^H \left(\sum_{i=1}^N \mathbf{y}_i \mathbf{y}_i^H \right) \mathbf{c}(\theta) \\ \text{s.t.} & \|\mathbf{c}(\theta)\|_2^2 = 1. \end{aligned} \quad (34)$$

The solution of (34) can be obtained as

$$\hat{\mathbf{c}}(\theta) = \mathbf{v}_{\max} \left(\sum_{i=1}^N \mathbf{y}_i \mathbf{y}_i^H \right), \quad (35)$$

where $\mathbf{v}_{\max}(\mathbf{M})$ is the eigenvector corresponding to the maximal eigenvalue of the matrix \mathbf{M} .

This shows that when CSI is unknown at the FC in the case of noise-free observations, the MLE becomes a subspace-based estimator.

3.2.2 Special case when $\sigma_c^2 \rightarrow 0$

When the communication SNR tends to infinity, the receiver of the FC can recover the quantized observations of the sensors with error free if a proper codebook, which will be discussed in Section 5.3, is applied. Then, the MLE with unknown CSI also degenerates into the BLUE shown in (20).

3.3 MLE with unknown CSI using training symbols

Define \mathbf{c}_p as a vector of L_p training symbols for the receiver to estimate the channels, which is predesigned and is known at both the transmitter and receiver. Each transmission for an observation will begin with transmitting the training symbols, followed by the data

symbols, which is defined as $\mathbf{c}_d(x)$. In this case, the messaging function becomes

$$\mathbf{c}(x) = \begin{pmatrix} \mathbf{c}_p \\ \mathbf{c}_d(x) \end{pmatrix}. \quad (36)$$

Substituting (36) into signal model shown in (7), the received signal \mathbf{y}_i can be decomposed into two parts that correspond to \mathbf{c}_p and $\mathbf{c}_d(x)$, respectively. The received signal from the i th sensor corresponding to \mathbf{c}_p is

$$\mathbf{y}_{i,p} = \sqrt{\mathcal{E}_d} h_i \mathbf{c}_p + \mathbf{n}_{cp,i}, \quad (37)$$

and the received signal from the i th sensor corresponding to $\mathbf{c}_d(x)$ is

$$\mathbf{y}_{i,d} = \sqrt{\mathcal{E}_d} h_i \mathbf{c}_d(x) + \mathbf{n}_{cd,i}, \quad (38)$$

where both $\mathbf{n}_{cp, i}$ and $\mathbf{n}_{cd, i}$ are vectors of thermal noise at the receiver. Note that $\mathbf{y}_{i, p}$ is independent from the observation x_i .

We let $\mathbf{c}_p^H \mathbf{c}_p = L_p/L$ and $\mathbf{c}_d(x)^H \mathbf{c}_d(x) = 1 - L_p/L$ in order to satisfy the normalization condition of $\mathbf{c}(x)$. Ignoring all the terms that do not affect the estimation, we obtain the log-likelihood function as

$$\log p(\mathbf{Y}|\theta) = \sum_{i=1}^N \log \left(\int_{-\infty}^{+\infty} \exp\left(-\frac{(x-\theta)^2}{2\sigma_x^2} + \beta |\mathbf{y}_{i,d}^H \mathbf{c}_d(x)|^2 + 2\beta \Re\{\mathbf{c}_p^H \mathbf{y}_{i,p} \mathbf{c}_d^H(x)\}\right) dx \right). \quad (39)$$

where $\mathbf{y}_{i, p}$ and $\mathbf{y}_{i, d}$ are, respectively, the received signals corresponding to the training symbols and the data symbols, and β is a constant.

Now we show that $\mathbf{c}_p^H \mathbf{y}_{i,p}$ in (39) can be regarded as the minimum mean square error (MMSE) estimate for the channel coefficient h_i with a constant factor. Since both h_i and the receiver thermal noise are complex Gaussian distributed, the MMSE estimate of h_i is equivalent to linear MMSE estimate, that is,

$$\begin{aligned} \hat{h}_i &= (\mathbf{R}_{y_p}^{-1} \mathbf{r}_{yh})^H \mathbf{y}_{i,p} \\ &= \frac{\sqrt{\mathcal{E}_d} L}{L\sigma_c^2 + L_p \mathcal{E}_d} \mathbf{c}_p^H \mathbf{y}_{i,p}, \end{aligned} \quad (40)$$

where $\mathbf{r}_{yh} = E[\mathbf{y}_{i,p} h_i^*] = \sqrt{\mathcal{E}_d} \mathbf{c}_p$, and \mathbf{R}_{y_p} is the covariance matrix of $\mathbf{y}_{i, p}$, which is

$$\mathbf{R}_{y_p} = \mathcal{E}_d \mathbf{c}_p \mathbf{c}_p^H + \sigma_c^2 \mathbf{I}. \quad (41)$$

Let $\kappa = \frac{\sqrt{\mathcal{E}_d} L}{L\sigma_c^2 + L_p \mathcal{E}_d}$, then we have $\mathbf{c}_p^H \mathbf{y}_{i,p} = \hat{h}_i / \kappa$. Substituting it into (39), we obtain

$$\log p(\mathbf{Y}|\theta) = \sum_{i=1}^N \log \left(\int_{-\infty}^{+\infty} \exp\left(-\frac{(x-\theta)^2}{2\sigma_x^2} + \beta |\mathbf{y}_{i,d}^H \mathbf{c}_d(x)|^2 + \frac{2\beta}{\kappa} \Re\{\hat{h}_i \mathbf{y}_{i,d}^H \mathbf{c}_d(x)\}\right) dx \right). \quad (42)$$

In the sequel, we will show that the MLE in this case is equivalent to a two-stage estimator. During the first

stage, the FC uses (40) to obtain the MMSE estimate of h_i . During the second stage, the FC conducts the MLE using \hat{h}_i . The channel estimate can be modeled as $\hat{h}_i = h_i + \varepsilon_{h_i}$, where ε_{h_i} is the estimation error subjecting to the complex Gaussian distribution with zero mean, and its variance is equal to the MSE of the linear MMSE estimator of h_i , which is [29]

$$E[(h_i - \hat{h}_i)^2] = E[h_i h_i^*] - \mathbf{r}_{yh}^H \mathbf{R}_{yp}^{-1} \mathbf{r}_{yh} = \frac{L\sigma_c^2}{L\sigma_c^2 + L_p \mathcal{E}_d}, \quad (43)$$

where \mathbf{R}_{yp}^{-1} can be obtained following Matrix Inversion Lemma [27].

Substituting \hat{h}_i into (7), the received signal of the data symbols becomes

$$\mathbf{y}_{i,d} = \sqrt{\mathcal{E}_d} \hat{h}_i \mathbf{c}_d(x) - \sqrt{\mathcal{E}_d} \varepsilon_{h_i} \mathbf{c}_d(x) + \mathbf{n}_{ci,d}, \quad (44)$$

where $\mathbf{n}_{ci,d}$ is the receiver thermal noise.

By deriving the conditional PDF $p(\mathbf{y}_{i,d} | \hat{h}_i, x)$ from (44), we can obtain a log-likelihood function that is exactly the same as that shown in (42). This implies that the MLE with unknown CSI exploits the available training symbols implicitly to provide an optimal channel estimate and then uses it to provide the optimal estimation of θ .

Note that the log-likelihood function in (42) is different from the log-likelihood function that uses the estimated CSI as the true value of CSI, which is

$$\log p(\mathbf{Y} | h_i = \hat{h}_i, \theta) = \sum_{i=1}^N \log \left(\int_{-\infty}^{+\infty} \exp \left(-\frac{(x-\theta)^2}{2\sigma_s^2} + \frac{2\sqrt{\mathcal{E}_d}}{\sigma_s^2} \Re\{\hat{h}_i \mathbf{y}_{i,d}^H \mathbf{c}_d(x)\} \right) dx \right). \quad (45)$$

By maximizing (45), we obtain a coherent estimator since there only exists the coherent term $\Re\{\hat{h}_i \mathbf{y}_{i,d}^H \mathbf{c}_d(x)\}$ in this log-likelihood function. By contrast, there exists a coherent term $\Re\{\hat{h}_i \mathbf{y}_{i,d}^H \mathbf{c}_d(x)\}$ as well as a non-coherent term $|\mathbf{y}_{i,d}^H \mathbf{c}_d(x)|^2$ in the log-likelihood function in (42). This means that the MLE obtained from (42) uses the channel estimate as a ‘‘partial’’ CSI that accounts for the channel estimation errors. The true value of the channel coefficients contained in the channel estimate corresponds to the coherent term in the log-likelihood function, whereas the uncertainty in the channel estimate, that is, the estimation errors, leads to the non-coherent term. We will compare the performance of the two estimators through simulations in Section 6.

4 Suboptimal estimator

In the previous section, we developed the MLE with known CSI, which is not feasible in real-world systems since perfect CSI cannot be provided especially in WSN with strict energy constraint. Nevertheless, its

performance can serve as a practical lower bound when both the observation noise and the communication errors are in presence.

The MLE with unknown CSI is more practical, but is too complex for application. Nonetheless, its structure provides some useful hints to derive low complexity estimator. In the following, we derive a suboptimal algorithm for the case with unknown CSI.

We first consider an approximation of the PMF, $p(S_m | \theta)$. Following the Lagrange Mean Value Theorem [30], there exists ξ in an interval $[\frac{S_m - \frac{\Delta}{2} - \theta}{\sigma_s}, \frac{S_m + \frac{\Delta}{2} - \theta}{\sigma_s}]$ that satisfies

$$p(S_m | \theta) = -Q'(\xi) \frac{\Delta}{\sigma_s} = \frac{\Delta}{\sqrt{2\pi}\sigma_s} \exp \left(-\frac{\xi^2}{2} \right). \quad (46)$$

If the quantization interval Δ is small enough, we can let ξ equal to the middle value of the interval, that is, $\xi = (S_m - \theta)/\sigma_s$, and obtain an approximate expression of the PMF as

$$p(S_m | \theta) \approx p_A(S_m | \theta) \triangleq \frac{\Delta}{\sqrt{2\pi}\sigma_s} \exp \left(-\frac{(S_m - \theta)^2}{2\sigma_s^2} \right) \quad (47)$$

Substituting (47) into (31) and taking its partial derivative with respect to θ , the likelihood equation is

$$\frac{\partial \log p(\mathbf{Y} | \theta)}{\partial \theta} = \sum_{i=1}^N \frac{\sum_{m=0}^{M-1} p(\mathbf{y}_i | \mathbf{c}_m) \frac{\partial p_A(S_m | \theta)}{\partial \theta}}{\sum_{m=0}^{M-1} p(\mathbf{y}_i | \mathbf{c}_m) p_A(S_m | \theta)} = 0, \quad (48)$$

where $\frac{\partial p_A(S_m | \theta)}{\partial \theta}$ can be derived as

$$\frac{\partial p_A(S_m | \theta)}{\partial \theta} = \frac{S_m - \theta}{\sigma_s^2} \cdot \frac{\Delta}{\sqrt{2\pi}\sigma_s} \exp \left(-\frac{(S_m - \theta)^2}{2\sigma_s^2} \right) = \frac{S_m - \theta}{\sigma_s^2} p_A(S_m | \theta). \quad (49)$$

Substituting (49) into (48), the likelihood equation can be simplified as

$$\theta = \frac{1}{N} \sum_{i=1}^N \left(\frac{\sum_{m=0}^{M-1} p(\mathbf{y}_i | \mathbf{c}_m) p_A(S_m | \theta) S_m}{\sum_{m=0}^{M-1} p(\mathbf{y}_i | \mathbf{c}_m) p_A(S_m | \theta)} \right), \quad (50)$$

which is the necessary condition for the MLE.

Unfortunately, we cannot obtain an explicit estimator for θ from this equation because the right-hand side of the likelihood equation also contains θ . However, considering the property of the conditional PDF, we can rewrite (50) as

$$\begin{aligned} \theta &= \frac{1}{N} \sum_{i=1}^N \left(\sum_{m=0}^{M-1} p(S_m | \mathbf{y}_i, \theta) S_m \right) \\ &= \frac{1}{N} \sum_{i=1}^N E[S_m | \mathbf{y}_i, \theta]. \end{aligned} \quad (51)$$

The term inside the sum of the right-hand side of the likelihood equation shown in (51) is actually the MMSE estimator of S_{m_i} for a given θ . This indicates that we can regard the MLE as a two-stage estimator.

During the first stage, it estimates S_{m_i} with the received signals from each sensor. During the second stage, it combines \hat{S}_{m_i} by a sample mean estimator.

We present a suboptimal estimator with a similar two-stage structure. This estimator can be viewed as a modified EM algorithm [24] since its two-stage structure is similar to the EM algorithm. Because the likelihood function shown in (31) has multiple extrema and the equation shown in (50) is only a necessary condition, the initial value of the iterative computation is critical to the convergence of the iterative algorithm. To obtain a good initial value, the suboptimal estimator estimates S_{m_i} by assuming it to be uniformly distributed. Furthermore, since the estimation quality of the first stage is available, we use BLUE to obtain $\hat{\theta}$ for exploiting the quality information instead of using the MLE in the M-step as in the standard EM algorithm.

During the first stage of the iterative computation, the suboptimal algorithm estimates S_{m_i} under MMSE criterion. This estimator requires *a priori* probability of S_{m_i} that depends on the unknown parameter θ . The initial distribution of S_{m_i} is set to be uniform distribution, that is, the estimate for *a priori* PDF of S_{m_i} $\hat{p}(S_{m_i}) = \frac{1}{M}$. After a temporary estimate of θ had been obtained, we use $p(S_{m_i}|\hat{\theta})$ to update $\hat{p}(S_{m_i})$. The MMSE estimator during the first stage is

$$\hat{S}_{m_i} = E[S_{m_i}|y_i] = \sum_{m_i=0}^{M-1} p(S_{m_i}|y_i)S_{m_i}, \quad (52)$$

where

$$p(S_{m_i}|y_i) = \frac{p(y_i|S_{m_i})\hat{p}(S_{m_i})}{p(y_i)} = \frac{p(y_i|S_{m_i})\hat{p}(S_{m_i})}{\sum_{m_i=0}^{M-1} p(y_i|S_{m_i})\hat{p}(S_{m_i})} \quad (53)$$

Because there is a one-to-one and onto mapping between S_m and \mathbf{c}_m , $p(y_i|S_{m_i})$ is equal to $p(y_i|\mathbf{c}_{m_i})$, which is shown in (32). After replacing $p(y_i|S_{m_i})$ in (53) with $p(y_i|\mathbf{c}_{m_i})$ and substituting it into (52), we have

$$\hat{S}_{m_i} = \frac{\sum_{m_i=0}^{M-1} p(y_i|\mathbf{c}_{m_i})\hat{p}(S_{m_i})S_{m_i}}{\sum_{m_i=0}^{M-1} p(y_i|\mathbf{c}_{m_i})\hat{p}(S_{m_i})}. \quad (54)$$

Now we derive the mean and variance of \hat{S}_{m_i} , which will be used in the BLUE of θ .

If $\hat{p}(S_{m_i})$ equals to its true value, the MMSE estimator in (54) is unbiased because

$$\begin{aligned} E[\hat{S}_{m_i}] &= \int_{\mathcal{C}^L} \hat{S}_{m_i} p(y_i) dy_i \\ &= \int_{\mathcal{C}^L} \frac{\sum_{m=0}^{M-1} p(y_i|\mathbf{c}_m)\hat{p}(S_m)S_m}{\sum_{m=0}^{M-1} p(y_i|\mathbf{c}_m)\hat{p}(S_m)} \sum_{m=0}^{M-1} p(y_i|\mathbf{c}_m)\hat{p}(S_m) dy_i \\ &= \int_{\mathcal{C}^L} \sum_{m=0}^{M-1} p(y_i|\mathbf{c}_m)\hat{p}(S_m)S_m dy_i \\ &= \sum_{m=0}^{M-1} \hat{p}(S_m)S_m \\ &= E[S_{m_i}]. \end{aligned} \quad (55)$$

However, $\hat{p}(S_{m_i})$ in our algorithm is not the true value since we use $\hat{\theta}$ instead of θ to get it. Therefore, the MMSE estimate may be biased. Because it is hard to obtain this bias in practical systems, we regard the MMSE estimator as an unbiased estimate in our suboptimal algorithm and evaluate the resulting performance loss via simulations later.

The variance of the MMSE estimate can be derived as

$$\begin{aligned} \text{Var}[\hat{S}_{m_i}|y_i] &= E[S_{m_i}^2|y_i] - E^2[S_{m_i}|y_i] \\ &= \frac{\sum_{m_i=0}^M S_{m_i}^2 p(y_i|\mathbf{c}_{m_i})\hat{p}(S_{m_i})}{\sum_{m_i=0}^M p(y_i|\mathbf{c}_{m_i})\hat{p}(S_{m_i})} - \hat{S}_{m_i}^2. \end{aligned} \quad (56)$$

Then, the BLUE for estimating θ is

$$\hat{\theta} = \left(\sum_{j=1}^N \frac{1}{\sigma_s^2 + \text{Var}[\hat{S}_{m_j}|y_j]} \right)^{-1} \sum_{i=1}^N \frac{\hat{S}_{m_i}}{\sigma_s^2 + \text{Var}[\hat{S}_{m_i}|y_i]} \quad (57)$$

Let k denote the index of the iteration, the iterative algorithm performed at the FC can be summarized as follows:

(S1) When $k = 1$, set $\hat{p}(S_{m_i})^{(k)} = 1/M$ as the initial value.

(S2) Compute $\hat{S}_{m_i}^{(k)}$, $i = 1, \dots, N$, and its variance with (54) and (56).

(S3) Substitute $\hat{S}_{m_i}^{(k)}$ and its variance into (57) to get $\hat{\theta}^{(k)}$.

(S4) Update $\hat{p}(S_{m_i})$ using $p_A(S_{m_i}|\hat{\theta})$, i.e., $\hat{p}(S_{m_i})^{(k+1)} = p_A(S_{m_i}|\hat{\theta}^{(k)})$.

(S5) Repeat step (S2) ~ (S4) to obtain $\hat{\theta}^{(k+1)}$ until the algorithm converges or a predetermined number of iterations is reached.

Note that this suboptimal algorithm differs from the one proposed in [9], which applies maximal *a posteriori* (MAP) criterion to detect binary observations of sensors and then uses the results as the true values of the observations in a MLE derived in noise-free channels. Our suboptimal algorithm inherits the structure of the MLE developed in fading channels, which gives "soft" estimates of the quantized observations at first, and

combines them with a linear optimal estimator afterward. By conducting these two stages iteratively, the estimation accuracy is improved rapidly. Although the suboptimal algorithm may converge to local optimal solutions due to the non-convexity of the original optimization problem, it still performs fairly well as will be shown in the simulation results. The convergence behavior of the algorithm will be studied in Section 5.4.

5 Performance analysis and discussion

5.1 Asymptotic performance w.r.t. number of the sensors

Now we discuss the asymptotic performance of the MLEs w.r.t. the number of sensors N by studying the Fisher information as well as the Cramér-Rao lower bound (CRLB) of the estimators.

We first consider the MLE with unknown CSI, where the channel coefficients are i.i.d. random variables. In this case, given θ , the received signals from different sensors are i.i.d. among each other; thus, the Fisher information, defined as $\mathbf{I}_N(\theta) = -E \left[\frac{\partial^2 \log p(\mathbf{Y}|\theta)}{\partial \theta^2} \right]$, linearly increases with the number of the sensors. Therefore, the CRLB, which is the reciprocal of the Fisher information, decreases at a speed of $1/N$, which is the same as the BLUE lower bound of centralized estimation [14].

When CSI is available at the FC, the received signals are no longer identical distributed. In this case, the Fisher information depends on the channel realizations. In the sequel, we will show that the mathematical expectation of the Fisher information over \mathbf{h} is always lower than that with unknown CSI, which means that the knowledge about the channels provides more information to improve the estimation quality.

Denote the Fisher information with known CSI as $\mathbf{I}_C(\theta)$, which depends on the channel coefficient vector \mathbf{h} . Considering that $p(\mathbf{Y}|\theta) = E_{\mathbf{h}}[p(\mathbf{Y}|\mathbf{h}, \theta)]$, we have

$$\begin{aligned} E_{\mathbf{h}}[\mathbf{I}_C(\theta)] &= -E_{\mathbf{h}} \left[E_{\mathbf{Y}} \left[\frac{\partial^2 \log p(\mathbf{Y}|\mathbf{h}, \theta)}{\partial \theta^2} \right] \right] \\ &= E_{\mathbf{h}} \left[\int_{\mathcal{C}^{N \times L}} \frac{\left(\frac{\partial p(\mathbf{Y}|\mathbf{h}, \theta)}{\partial \theta} \right)^2}{p(\mathbf{Y}|\mathbf{h}, \theta)} d\mathbf{Y} \right]. \end{aligned} \quad (58)$$

The terms in the integration of (58) are convex in $p(\mathbf{Y}|\mathbf{h}, \theta)$ because

$$\frac{\partial^2}{\partial p(\mathbf{Y}|\mathbf{h}, \theta)^2} \left(\frac{\left(\frac{\partial p(\mathbf{Y}|\mathbf{h}, \theta)}{\partial \theta} \right)^2}{p(\mathbf{Y}|\mathbf{h}, \theta)} \right) = \frac{2 \left(\frac{\partial p(\mathbf{Y}|\mathbf{h}, \theta)}{\partial \theta} \right)^2}{p(\mathbf{Y}|\mathbf{h}, \theta)^3} \geq 0. \quad (59)$$

Since the integration can be viewed as a nonnegative weighted summation, which will preserve the convexity of the functions [28], (58) is a convex function of $p(\mathbf{Y}|\mathbf{h}, \theta)$. Following Jensen's inequality and the convexity of

(58), we have

$$\begin{aligned} E_{\mathbf{h}}[\mathbf{I}_C(\theta)] &\geq \int_{\mathcal{C}^{N \times L}} \frac{\left(\frac{\partial E_{\mathbf{h}}[p(\mathbf{Y}|\mathbf{h}, \theta)]}{\partial \theta} \right)^2}{E_{\mathbf{h}}[p(\mathbf{Y}|\mathbf{h}, \theta)]} d\mathbf{Y} \\ &= \int_{\mathcal{C}^{N \times L}} \frac{\left(\frac{\partial p(\mathbf{Y}|\theta)}{\partial \theta} \right)^2}{p(\mathbf{Y}|\theta)} d\mathbf{Y} \\ &= \mathbf{I}_N(\theta). \end{aligned} \quad (60)$$

Therefore, the asymptotic performance of the MLE with known CSI is superior to that of the MLE with unknown CSI, where the CRLB of the latter decreases at the speed of $1/N$.

5.2 Computational complexity

5.2.1 MLE with known CSI

Since the parameter being estimated is a scalar, one-dimensional searching algorithms can be used to obtain the maximum of the log-likelihood function. However, because the log-likelihood function shown in (11) is non-concave and has multiple extrema, we need to find all its local maxima to get the global maximum.

Exhaustive searching method can be used to find the global maximum. In order to make the MSE introduced by discrete searching neglectable, we let the searching step size be less than Δ/N ; thus, we need to compute the value of the likelihood function at least $M \times N$ times to obtain the MLE.

The FC applies (11), (12) and (13) to compute the values of the likelihood function with different θ . The exponential term in (12) is independent from θ ; thus, it can be computed before searching and be stored for future use.

Given θ , we still need to compute $p(S_m|\theta)$, $m = 0, \dots, M - 1$, which complexity is $O(M)$, then to obtain each value of the likelihood function with M additions and M multiplications. Therefore, the computational complexity for getting one value of $\log p(\mathbf{Y}|\mathbf{h}, \theta)$ is $O(MN)$.

After considering the operations required by the exhaustive searching, the overall complexity of the MLE is $O(M^2N^2)$.

5.2.2 MLE with unknown CSI

The difference between the MLEs with known and unknown CSI is that $p(\mathbf{y}_i|\mathbf{c}_m)$ is used in MLE with unknown CSI instead of $p(\mathbf{y}_i|h_i, \mathbf{c}_m)$. Since $p(\mathbf{y}_i|\mathbf{c}_m)$ can also be computed before the searching, this difference has no impact on the complexity of the MLE with unknown CSI. The computational complexity of the MLE with unknown CSI is also $O(M^2N^2)$.

5.2.3 Suboptimal estimator

For each iteration of the suboptimal estimator, we need to get \hat{S}_m and its variance with (54) and (56) and then

obtain the estimate of θ with (57). The complexity is similar to that of computing the log-likelihood function, which is $O(MN)$. If the algorithm converges after I_t iterations, the complexity of the suboptimal estimator will be $O(I_t MN)$.

5.3 Discussion about transmission codebook issues

As we have discussed, the transmission codebooks can represent various quantization, coding and modulation schemes as well as the training symbols. Here, we discuss the impact of the codebooks on the decentralized MLEs.

We rewrite the conditional PDF with known CSI shown in (10) as

$$p(\mathbf{y}_i|h_i, x) = \frac{1}{(\pi\sigma_c^2)^L} \exp\left(-\frac{\mathcal{E}_d|h_i|^2}{\sigma_c^2}\right) \exp\left(-\frac{\|\mathbf{y}_i\|_2^2}{\sigma_c^2} + \frac{2\sqrt{\mathcal{E}_d}\Re\{h_i\mathbf{y}_i^H\mathbf{c}(x)\}}{\sigma_c^2}\right). \quad (61)$$

Comparing the conditional PDF with unknown CSI $p(\mathbf{y}_i|x)$ shown in (28) with $p(\mathbf{y}_i|h_i, x)$ shown in (61), we see that both PDFs depend on the correlation between the received signals \mathbf{y}_i and the transmitted symbols $\mathbf{c}(x)$. With known CSI, the optimal estimator is a coherent algorithm, since (61) relies on the real part of the correlation, $\mathbf{y}_i^H\mathbf{c}(x)$. With unknown CSI, the optimal estimator is a non-coherent algorithm, since (28) depends on the square norm of $\mathbf{y}_i^H\mathbf{c}(x)$. Because $\mathbf{y}_i^H\mathbf{c}(x) = \sqrt{\mathcal{E}_d}h_i^*\mathbf{c}^H(x_i)\mathbf{c}(x) + \mathbf{n}_{c,i}^H\mathbf{c}(x)$, both MLEs depend on the cross-correlation of the transmit symbols $\mathbf{c}^H(x_i)\mathbf{c}(x)$.

If there exist two transmit symbols \mathbf{c}_m and \mathbf{c}_n in the transmission codebook that have the same norm, that is,

$$\mathbf{c}_m = \mathbf{c}_n e^{j\phi}, \quad (62)$$

then $p(\mathbf{y}_i|x)$ will have two identical extrema since the MLE with unknown CSI only depends on $|\mathbf{y}_i^H\mathbf{c}(x)|^2$. Such a phase ambiguity will lead to severe performance degradation to the decentralized estimator. Therefore, the autocorrelation matrix of the codebook plays a critical role on the performance of the MLE, especially when CSI is unknown.

Many transmission schemes have this phase ambiguity problem, for example, when the natural binary code and BPSK are applied to represent each quantized observation and to modulate. For any \mathbf{c}_m in such a transmission codebook, defined as \mathbf{C}_{tm} , there exists $\mathbf{c}_{m'}$ in \mathbf{C}_{tm} that satisfies $\mathbf{c}_{m'} = -\mathbf{c}_m$. Therefore, \mathbf{C}_{tm} is not a proper codebook. Another example is AF, the messaging function of which is $\mathbf{c}(x) = Gx$, where G is the amplification gain. The MLE with unknown CSI is unable to distinguish x from $-x$ when using this messaging function.

In order to handle the phase ambiguity problem inherent in the codebook \mathbf{C}_{tm} , we can simply insert training symbols into the transmit symbols. Though

heuristic, this approach provides fairly good performance because the MLE exploits the training symbols to estimate the channel coefficients implicitly as we have shown. Moreover, since from the later simulations we see that the MLE without CSI and without training symbols does not perform well, we need to insert training symbols when we apply the decentralized estimator.

Since the MLEs are associated with the autocorrelation matrix of the transmission codebook, this allows us to enhance the performance of the estimators by systematically designing the codebook. Nonetheless, this is out of the scope of this paper. Some preliminary results for optimizing the transmission codebooks are shown in [31].

5.4 Convergence of the suboptimal estimator

For an iterative algorithm $\theta^{(k+1)} = T(\theta^{(k)})$, we call that the algorithm is convergent if the distance between $\theta^{(k+1)}$ and a fixed point of $T(\theta)$ is smaller than the distance between $\theta^{(k)}$ and this fixed point, where the fixed points of $T(\theta)$ are the points that satisfy equation $\theta = T(\theta)$. This means that after each iteration, the output of the algorithm is closer to a fixed point.

Define Φ as a fixed point of $T(\theta)$ in (φ_1, φ_2) . The algorithm is convergent if $|\theta^{(k+1)} - \Phi| < |\theta^{(k)} - \Phi|$ for all $\theta^{(k)} \in (\varphi_1, \varphi_2)$.

In the following, we first study the convergence behavior of an iterative algorithm obtained directly from the likelihood equation (50) due to the mathematical tractability, where $T(\theta)$ is defined as the right-hand side of equation (50). The iteration algorithm of the suboptimal estimator can be regarded as a modified version of this algorithm, which will be discussed afterward.

To simplify the notation, we rewrite $T(\theta)$ as a function of $\frac{\partial \log p(\mathbf{Y}|\theta)}{\partial \theta}$. From Eqs. (48), (49) and (50), we have

$$T(\theta) = \frac{\sigma_s^2}{N} \frac{\partial \log p(\mathbf{Y}|\theta)}{\partial \theta} + \theta. \quad (63)$$

Since the iterative function shown in (63) is derived from the likelihood equation, all stationary points of the log-likelihood function are fixed points of $T(\theta)$. Denote Φ_n , $n = 1, 2, \dots$, as the local maxima of the log-likelihood function, which are sorted in ascending order. Since the log-likelihood function is a continuous function of θ , there exists a minimum between two adjacent maxima. The minimum between Φ_n and Φ_{n+1} is defined as φ_n . We will show in the following that in each interval $(\varphi_{n-1}, \varphi_n)$, the algorithm converges to Φ_i after ignoring the effect of the non-extremal stationary points of log-likelihood function.

Assume that there is no non-extremal stationary point in $(\varphi_{n-1}, \varphi_n)$. Because Φ_n is a maximum, the sign of $\frac{\partial \log p(\mathbf{Y}|\theta^{(k)})}{\partial \theta^{(k)}}$ is always different from the sign of $(\theta^{(k)} - \Phi_n)$

for all $\varphi_{n-1} < \theta^{(k)} < \varphi_n$. Following the corollary shown in Appendix, the algorithm is convergent if

$$\frac{\sigma_s^2}{N} \frac{\partial^2 \log p(\mathbf{Y}|\theta)}{\partial \theta^2} > -2, \quad \forall \theta \in (\varphi_{n-1}, \varphi_n). \quad (64)$$

Taking the second-order partial derivative of $\log p(\mathbf{Y}|\theta)$, we have

$$\frac{\sigma_s^2}{N} \frac{\partial^2 \log p(\mathbf{Y}|\theta)}{\partial \theta^2} = \frac{1}{N\sigma_s^2} \sum_{i=1}^N \left(\sum_{m=0}^{M-1} S_m^2 \frac{p(\mathbf{y}_i|\mathbf{c}_m)p(S_m|\theta)}{\sum_{m=0}^{M-1} p(\mathbf{y}_i|\mathbf{c}_m)p(S_m|\theta)} - \left(\sum_{m=0}^{M-1} S_m \frac{p(\mathbf{y}_i|\mathbf{c}_m)p(S_m|\theta)}{\sum_{m=0}^{M-1} p(\mathbf{y}_i|\mathbf{c}_m)p(S_m|\theta)} \right)^2 \right) - 1. \quad (65)$$

By defining

$$f_{m,i} = \frac{p(\mathbf{y}_i|\mathbf{c}_m)p(S_m|\theta)}{\sum_{m=0}^{M-1} p(\mathbf{y}_i|\mathbf{c}_m)p(S_m|\theta)}, \quad (66)$$

we have $f_{m,i} \geq 0$ and $\sum_{m=0}^{M-1} f_{m,i} = 1$. Therefore, $f_{m,i}$, $m = 0, \dots, M-1$ can be regarded as a PMF. Then, the term in (65) can be rewritten as

$$\sum_{m=0}^{M-1} S_m^2 \frac{p(\mathbf{y}_i|\mathbf{c}_m)p(S_m|\theta)}{\sum_{m=0}^{M-1} p(\mathbf{y}_i|\mathbf{c}_m)p(S_m|\theta)} - \left(\sum_{m=0}^{M-1} S_m \frac{p(\mathbf{y}_i|\mathbf{c}_m)p(S_m|\theta)}{\sum_{m=0}^{M-1} p(\mathbf{y}_i|\mathbf{c}_m)p(S_m|\theta)} \right)^2 = \sum_{m=0}^{M-1} S_m^2 f_{m,i} - \left(\sum_{m=0}^{M-1} S_m f_{m,i} \right)^2 \geq 0, \quad (67)$$

and consequently,

$$\frac{\sigma_s^2}{N} \frac{\partial^2 \log p(\mathbf{Y}|\theta)}{\partial \theta^2} \geq -1, \quad (68)$$

which satisfies (64). Therefore, the iterative algorithm is convergent.

Now we discuss the non-minimum stationary points of the log-likelihood function. Considering a minimum φ_n , for any $\theta \in (\varphi_n, \varphi_{n+1})$, the sign of $\frac{\partial \log p(\mathbf{Y}|\theta)}{\partial \theta}$ is the same as that of $(\theta - \varphi_n)$ on both sides of φ_n , which does not satisfy the sufficient and necessary condition shown in Appendix. Therefore, the algorithm does not converge to φ_n unless $\theta^{(k)}$ exactly equals φ_n . Any disturbance will perturb $\theta^{(k+1)}$ far from this minimum point. As to any non-extremal stationary point $\bar{\theta}$, the sign of $\frac{\partial \log p(\mathbf{Y}|\theta)}{\partial \theta}$ is the same as that of $(\theta - \bar{\theta})$ at one side of this point. The disturbance with proper direction will also make $\theta^{(k+1)}$ far from this point.

When the communication SNR tends to infinity, that is, $\sigma_c \rightarrow 0$, there is only one $p(\mathbf{y}_i|\mathbf{c}_m)$, $m = 0, \dots, M-1$, that can be positive. All other $p(\mathbf{y}_i|\mathbf{c}_m)$ tend to 0. By substituting this into (65), we have $\frac{\sigma_s^2}{N} \frac{\partial^2 \log p(\mathbf{Y}|\theta)}{\partial \theta^2} = -1$. It is not hard to verify that in this case, $|\theta^{(k+1)} - \varphi_m| = 0$ for any $\theta^{(k)}$. It means that the iterative algorithm converges to a local maximum of the log-likelihood function exactly after one iteration.

At practical communication SNR levels, $\frac{\sigma_s^2}{N} \frac{\partial^2 \log p(\mathbf{Y}|\theta)}{\partial \theta^2} > -1$, which will affect on the convergent speed of the algorithm.

Now we consider the iterative algorithm of the suboptimal estimator. Similar to the previous discussion, we rewrite the suboptimal algorithm (57) as a function of $p(\mathbf{y}_i|\theta)$ and its partial derivatives. After taking the first- and second-order partial derivatives of $p(\mathbf{y}_i|\theta)$ and comparing them with (54), (56) and (57), the suboptimal estimator can be rewritten as

$$\theta^{(k+1)} = \frac{\sigma_s^2 \sum_{i=1}^N w_i(\theta^{(k)}) \frac{\partial p(\mathbf{y}_i|\theta^{(k)})}{\partial \theta^{(k)}}}{\sum_{j=1}^N w_j(\theta^{(k)})} + \theta^{(k)}, \quad (69)$$

where

$$w_i(\theta) = \left(2 + \sigma_s^2 \frac{\partial^2 p(\mathbf{y}_i|\theta)}{\partial \theta^2} \right)^{-1}. \quad (70)$$

This estimator has the same form as the algorithm defined by (63). Therefore, following the same argument, we can show that a sufficient condition that the suboptimal estimator be convergent is

$$\frac{\partial}{\partial \theta} \frac{\sigma_s^2 \sum_{i=1}^N w_i(\theta) \frac{\partial p(\mathbf{y}_i|\theta)}{\partial \theta}}{\sum_{j=1}^N w_j(\theta)} > -2, \quad (71)$$

where

$$\begin{aligned} & \frac{\partial}{\partial \theta} \frac{\sigma_s^2 \sum_{i=1}^N w_i(\theta) \frac{\partial p(\mathbf{y}_i|\theta)}{\partial \theta}}{\sum_{j=1}^N w_j(\theta)} \\ &= \frac{\sigma_s^2 \left(\sum_{i=1}^N \sum_{j=1}^N w_i(\theta) w_j(\theta) \frac{\partial^2 p(\mathbf{y}_i|\theta)}{\partial \theta^2} + \sum_{i=1}^N \sum_{j=1}^N w_i(\theta) w_j(\theta) \frac{\partial^2 p(\mathbf{y}_j|\theta)}{\partial \theta^2} - \sum_{i=1}^N \sum_{j=1}^N w_i(\theta) w_j(\theta) \frac{\partial p(\mathbf{y}_i|\theta)}{\partial \theta} \frac{\partial p(\mathbf{y}_j|\theta)}{\partial \theta} \right)}{\left(\sum_{j=1}^N w_j(\theta) \right)^2} \\ &= \frac{\sigma_s^2 \sum_{i=1}^N w_i(\theta) \frac{\partial^2 p(\mathbf{y}_i|\theta)}{\partial \theta^2}}{\sum_{j=1}^N w_j(\theta)} \end{aligned} \quad (72)$$

By letting $N = 1$, we can obtain from (68) that for all i , $\frac{\sigma_s^2}{N} \frac{\partial^2 p(\mathbf{y}_i|\theta)}{\partial \theta^2} \geq -1$, and all $w_i(\theta) > 0$. Therefore,

$$\frac{\partial}{\partial \theta} \frac{\sigma_s^2 \sum_{i=1}^N w_i(\theta) \frac{\partial p(\mathbf{y}_i|\theta)}{\partial \theta}}{\sum_{j=1}^N w_j(\theta)} \geq -1, \quad (73)$$

which satisfies the condition (71).

When the communication SNR tends to infinity, all $\sigma_s^2 \frac{\partial^2 p(\mathbf{y}_i|\theta)}{\partial \theta^2}$ tend to -1 as discussed. The estimator shown in (57) degenerates into the algorithm shown in (63). It is also convergent to a local maximum of the log-likelihood function exactly after one iteration.

At practical communication SNR levels, we can see from (72) that $\frac{\partial^2 p(\mathbf{y}_i|\theta)}{\partial \theta^2}$ is weighted by itself since $w_i(\theta)$ depends on $\frac{\partial^2 p(\mathbf{y}_i|\theta)}{\partial \theta^2}$. A larger $\frac{\partial^2 p(\mathbf{y}_i|\theta)}{\partial \theta^2}$ will make the weight $w_i(\theta)$ smaller. Therefore, the value of the partial derivative in (73) is closer to -1 compared with the iterative algorithm defined with (63) given \mathbf{y}_i and $\hat{\theta}^{(k)}$, which increases the speed of convergence.

6 Simulation results

We use the Monte Carlo method to evaluate the performance of the estimators. In each trail, the parameter θ is generated from a uniformly distributed source within its dynamic range. We use the MSE of estimating θ , that is, $E[(\theta - \hat{\theta})^2]$, as a performance metric. The observation SNR considered in simulations is defined as [12]

$$\gamma_s = 20 \log_{10} \left(\frac{W}{\sigma_s} \right). \quad (74)$$

We use \mathcal{E}_d , the energy consumed by each sensor to transmit one observation, to define the communication SNR in order to fairly compare the energy efficiency of the estimators with different transmission schemes. The communication SNR is defined as

$$\gamma_c = 10 \log_{10} \left(\frac{\mathcal{E}_d}{N_0} \right). \quad (75)$$

An $M = 16$ level uniform quantizer is considered, where each quantized value can be represented by a $K = 4$ bit binary sequence. We do not consider the binary quantizer, which only performs well in low observation SNR.

The codebooks used in the simulations are summarized in Table 1. Considering the general features of WSNs, that is, usually short data packets are transmitted and each sensor is of low cost, we use a simple error control coding (ECC) scheme, the cyclic redundancy check (CRC) codes with generation polynomial $G(x) = x^4 + x + 1$, as an example of the coded transmission. The codebook is denoted as \mathbf{C}_{tc} . For comparison, uncoded transmission is also evaluated, where natural binary code is applied to represent each quantization, which codebook is denoted as \mathbf{C}_{tn} . We consider BPSK modulation for all codebooks. Because the code length of the uncoded transmission is shorter than that of the coded transmission, the energy to transmit each symbol will be higher for a given \mathcal{E}_d . Due to the phase ambiguity problem discussed in Section 5.3, we also consider the codebook with training symbols \mathbf{C}_{tp} .

When CSI is known at the FC, we evaluate the performance of the MLE with codebook \mathbf{C}_{tn} . The simulation results are marked as “MLE CSI” in the legend. When CSI is unknown and the codebook is still \mathbf{C}_{tn} , the legends for MLE and the suboptimal estimator are “MLE NoCSI” and “Subopt NoCSI,” respectively. When

CSI is unknown and the codebook is \mathbf{C}_{tp} , where 2 or 5 training symbols are inserted, the simulation results are marked as “MLE NoCSI TS2/5” and “Subopt NoCSI TS2/5.” We also evaluate the performance of the MLE with a near-optimal codebook obtained in [31], which is marked as “MLE NoCSI OPT.” As discussed in Section 3.2, the FC can use the training symbols to estimate the CSI and use the estimated CSI as the known CSI to estimate θ . We evaluate this estimator with the codebook \mathbf{C}_{tp} , which is marked as “MLE EstCH TS2/5.”

To demonstrate the performance gain of the proposed estimators, two traditional fusion-based estimators and AF transmission are simulated. In the fusion-based estimators, the FC first demodulates the transmitted data from each sensor, then reconstructs the observation of each sensor from the demodulated symbols following the rule of quantization and finally combines these estimated observations with BLUE fusion rule to produce the estimate of θ . When ECCs are applied at the sensors, the receiver at the FC will exploit its error detection ability to discard the data that cannot pass the error check. The fusion-based estimators using codebook \mathbf{C}_{tn} and \mathbf{C}_{tc} are denoted as “Fusion-NoECC” and “Fusion-CRC” in the legends of the figures, respectively. For AF, the amplification gain G is designed to make the average transmission power of the sensors equals to that of the digital communication schemes. We also use the MLE at the FC to estimate θ , which is marked as “AF” in the legend.

The MSE of the Quasi-BLUE [14] is shown as the performance lower bound with legend “Q-BLUE Bound.” This MSE is obtained in perfect communication scenarios with the same M -level quantizer as other estimators.

6.1 Convergence of the suboptimal estimator

We first study the convergence of the suboptimal estimator. Figure 1 depicts the MSEs of the suboptimal estimator as a function of the number of iterations. As discussed in 5.4, at high communication SNR levels, the MSE of the suboptimal estimator is convergent after one iteration, that is, the MSE does not decrease with the iterations any more. At low communication SNR levels, the convergent speed becomes lower.

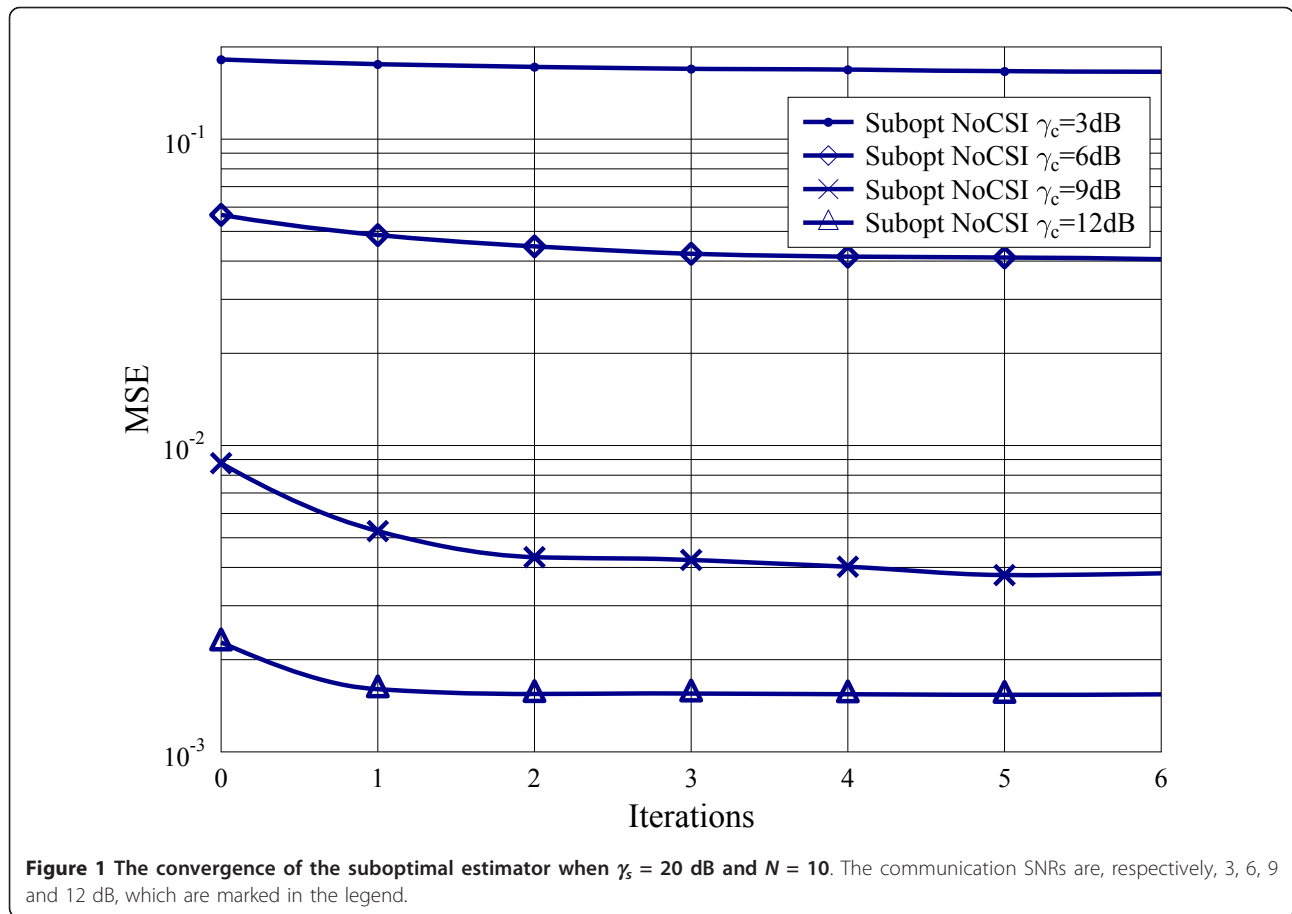
6.2 MSE versus the communication SNR

Figure 2 depicts the MSEs of the estimators with known and unknown CSI.

When CSI is known at the FC, it is shown from Figure 2a that the MLE outperforms the fusion-based estimators. The MSE of the MLE approaches to the Quasi-BLUE lower bound rapidly with the increasing of the communication SNR. As expected, the MLE with AF transmission, marked as *AF*, is inferior to the MLE with digital communication using 4-bits

Table 1 The summary of the codebooks considered

Codebook	Error control coding	Training symbols	Modulation
\mathbf{C}_{tn}	No	No	BPSK
\mathbf{C}_{tc}	CRC	No	BPSK
\mathbf{C}_{tp}	No	2 or 5	BPSK



quantization, marked as *MLE CSI*. This justifies the conclusions in [19-21], which show that AF is not optimal in fading channels.

According to the performance analysis for BPSK modulation in Rayleigh fading channels [32], the BER of the transmission scheme with codebook C_{tn} exceeds 0.15 when $\gamma_s < 3$ dB. ECC can improve the transmission performance for high communication SNR, but it causes more errors for low SNR. For the transmission schemes using CRC, the BER is even worse because long codes will reduce the transmission energy per symbol. For such a high BER, the fusion-based estimators do not perform well. Most of the demodulated data will be dropped due to the error check; thus, the fusion-based estimators do not have enough information to exploit, which finally leads to the worse MSE performance.

When CSI is unknown at the FC, the MSEs of the MLE with unknown CSI and with two different ways of using training symbols for channel estimation are shown in Figure 2b. One is the MLE obtained from the log-likelihood function in (42), and the other is the estimator obtained from (45), which uses the estimated channel coefficients as their true values. As expected, our

MLE shown in (42) performs better, because it takes into account the uncertainty of the channel estimation.

Because there exists phase ambiguity in the schemes with C_{tn} and AF transmission, simulation results show that the MSEs of the MLE and suboptimal estimator using these two transmission schemes are very large and do not decrease when γ_c increases. Therefore, they are not shown on the figures.

When we insert training symbols, the performance of the MLE with unknown CSI improves significantly, but it is still much worse than that of the MLE with known CSI at low communication SNR levels. It is interesting to see that using more training symbols does not improve the performance of the MLE as expected, because inserting training symbols will reduce the energy for the data symbols when the energy for transmitting an observation is fixed. Our simulations show that the best performance is obtained when $L_p = 2$. This is consistent with the observation of [33], where the optimal L_p equals to \sqrt{K} .

As discussed, inserting training symbols is a heuristic way to improve the performance. It is shown in the figure that a codebook designed by using optimization

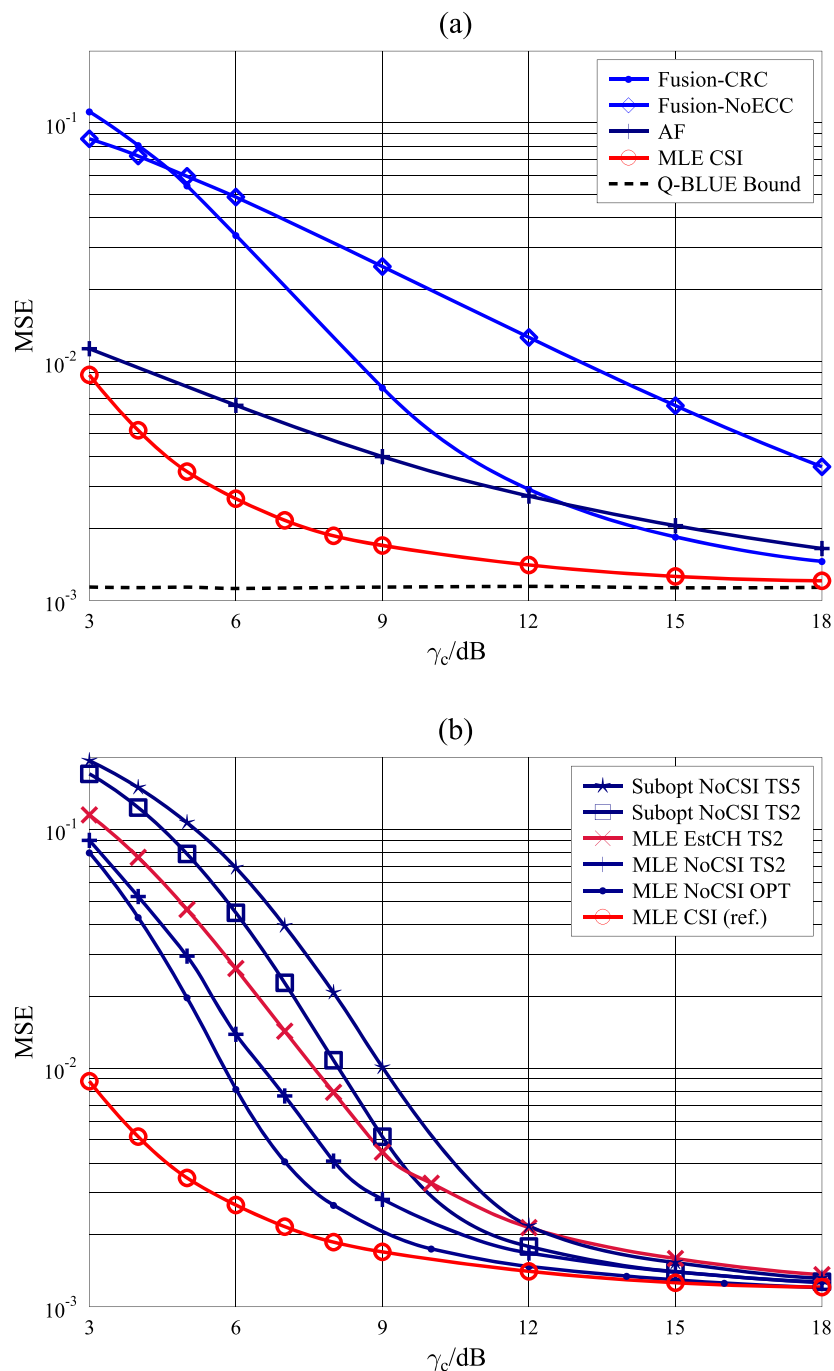


Figure 2 The MSEs of the estimators with known CSI and unknown CSI as a function of communication SNR when $N = 10$ and $\gamma_s = 20$ dB.

method outperforms all the codebooks with training symbols.

6.3 MSE versus the number of sensors

Figure 3 shows the MSEs of the estimators with known CSI and unknown CSI as a function of the number of sensors N . We see that the MSEs of all the estimators

decrease at the speed of $1/N$ for large enough N , but the MSEs do not approach the lower bound due to the communication errors. This validates our asymptotic performance analysis for the MLEs both with known CSI and with unknown CSI in 5.1. Moreover, we observe that the proposed estimators perform much better than the fusion-based estimators. It means that the

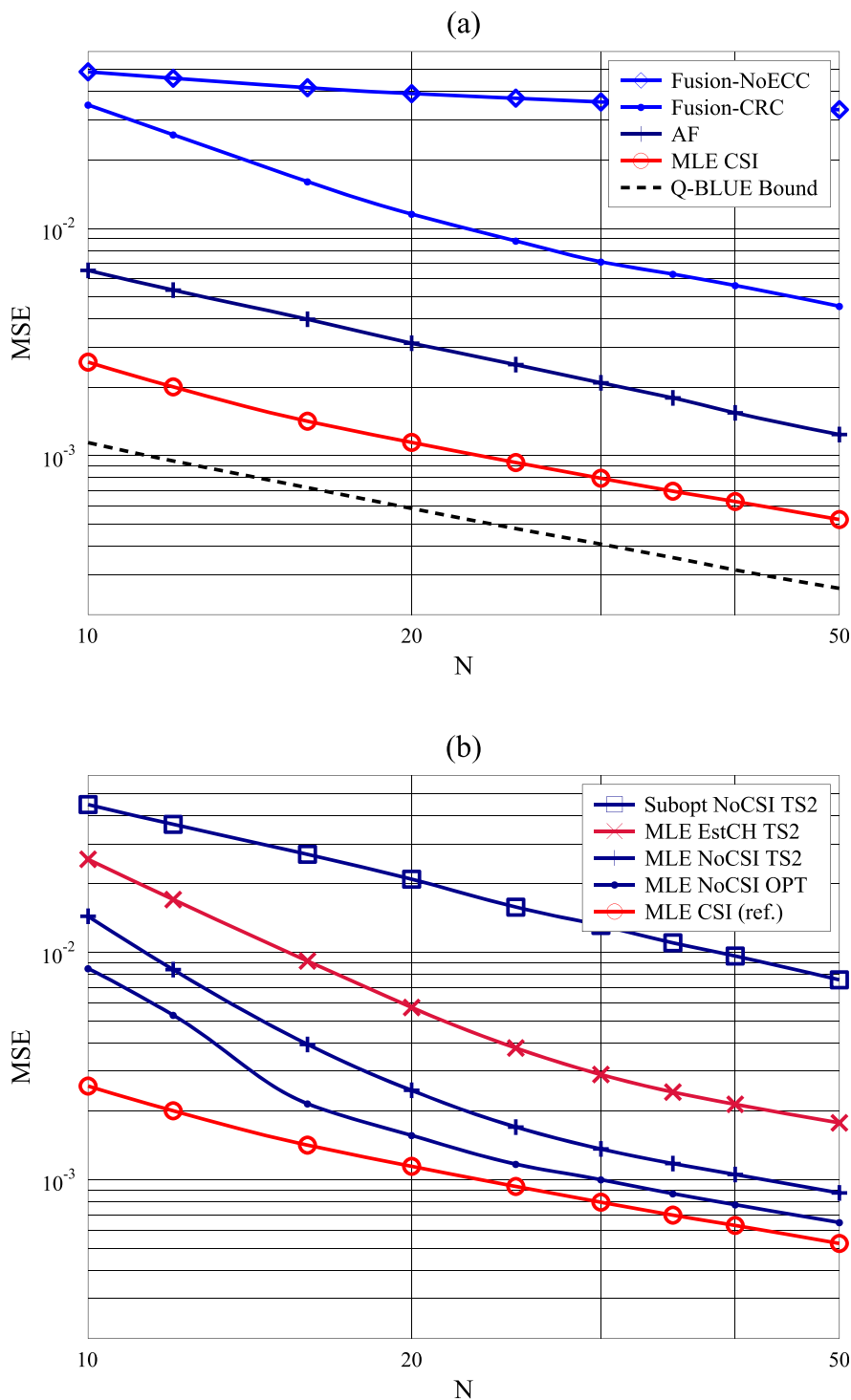


Figure 3 The MSEs of the estimators with known and unknown CSI, where $\gamma_c = 6$ dB and $\gamma_s = 20$ dB.

networks with conventional approaches must activate more sensors to achieve the same MSE performance as those with our estimators, which will lead to low energy and bandwidth efficiency.

7 Conclusion

In this paper, we studied decentralized estimation for a deterministic parameter using digital communications over orthogonal multiple-access fading channels with a

multiple-bit quantizer. By introducing a general messaging function, the proposed estimators can be applied for various quantization, coding and modulation schemes, including AF transmission, binary quantization and with or without training symbols.

We derived the MLEs with both known and unknown CSI. The MLE with known CSI can serve as a practical performance lower bound of existing decentralized estimators. It is shown that the MLE with multi-level quantization outperforms the MLE with AF as well as the fusion-based estimators.

The MLE with unknown CSI is more realistic. Without training symbols, it does not perform well due to the phase ambiguity. When inserting training symbols before data symbols, it estimates channel coefficients implicitly and exploits the channel estimates in an optimal way. Under the energy constraint, only a few symbols are beneficial for training channels, while more training symbols will lead to performance degradation. To design an estimator with affordable complexity, we developed a suboptimal estimator that converges rapidly. The proposed estimator performs well. It exhibits similar performance as the MLE at high SNRs and has minor performance loss at low SNRs.

Appendix

Proposition: For an iterative algorithm $\theta^{(k+1)} = T(\theta^{(k)})$ with a form that $T(\theta) = f(\theta) + \theta$, this algorithm converges to a fixed point Φ of $T(\theta)$ if and only if

$$-2(\theta - \Phi) < f(\theta) < 0, \quad \forall \theta - \Phi > 0, \quad (76)$$

and

$$0 < f(\theta) < 2(\Phi - \theta), \quad \forall \theta - \Phi < 0. \quad (77)$$

Proof: We first prove that (76) and (77) are sufficient conditions. For the function $T(\theta) = f(\theta) + \theta$ and its fixed point Φ , we have

$$|\theta^{(k+1)} - \Phi| = |T(\theta^{(k)}) - \Phi| = |f(\theta^{(k)}) + \theta^{(k)} - \Phi| \quad (78)$$

When $\theta^{(k)} - \Phi > 0$, substituting (76) into (78), we have

$$|\theta^{(k+1)} - \Phi| = |f(\theta^{(k)}) + \theta^{(k)} - \Phi| < |\theta^{(k)} - \Phi|, \quad (79)$$

which shows that the algorithm is convergent. When $\theta^{(k)} - \Phi < 0$, substituting (77) into (78), we also obtain the inequality shown in (79). Therefore, (76) and (77) are sufficient conditions of the convergence.

Now we prove that they are also necessary conditions. If the algorithm is convergent, we have

$$|\theta^{(k+1)} - \Phi| = |f(\theta^{(k)}) + \theta^{(k)} - \Phi| < |\theta^{(k)} - \Phi|. \quad (80)$$

When $\theta^{(k)} - \Phi > 0$, (80) can be rewritten as

$$\begin{aligned} |f(\theta^{(k)}) + \theta^{(k)} - \Phi| &< \theta^{(k)} - \Phi \\ \Rightarrow -(\theta^{(k)} - \Phi) &< f(\theta^{(k)}) + \theta^{(k)} - \Phi < (\theta^{(k)} - \Phi). \end{aligned} \quad (81)$$

After the simplifications, we can obtain (76) from (81).

Similarly, when $\theta^{(k)} - \Phi < 0$, (77) can be obtained following the same procedure. Therefore, (76) and (77) are necessary conditions. \square

Corollary: A sufficient condition that the algorithm converges to Φ is $f(\theta)(\theta - \Phi) < 0$, $\forall \theta \neq \Phi$, and $f'(\theta) > -2$.

Proof: Since Φ is a fixed point of $T(\theta)$, we have $\Phi = T(\Phi) = f(\Phi) + \Phi$, thus $f(\Phi) = 0$. When $\theta^{(k)} - \Phi > 0$ and $f'(\theta) > -2$, we have

$$f(\theta^{(k)}) = \int_{\Phi}^{\theta^{(k)}} f'(\theta) d\theta \geq \int_{\Phi}^{\theta^{(k)}} -2 d\theta = -2(\theta^{(k)} - \Phi). \quad (82)$$

When $\theta^{(k)} - \Phi < 0$ and $f'(\theta) > -2$, we have

$$f(\theta^{(k)}) = \int_{\theta^{(k)}}^{\Phi} -f'(\theta) d\theta \leq \int_{\theta^{(k)}}^{\Phi} 2 d\theta = 2(\Phi - \theta^{(k)}). \quad (83)$$

Therefore, the first inequality in (76) and the second inequality in (77) are satisfied. From the condition $f(\theta)(\theta - \Phi) < 0$, it is not hard to find that the second inequality in (76) and the first inequality in (77) are also satisfied. Thus, the iterative algorithm is convergent following Proposition. \square

Acknowledgements

This work was supported by the National Nature Science Foundation of China under Grant 60672103. Parts of this work were presented at IEEE Globecom'07, Washington, DC, United States, Nov. 2007.

Competing interests

The authors declare that they have no competing interests.

Received: 26 November 2010 Accepted: 12 December 2011

Published: 12 December 2011

References

1. IF Akyildiz, W Su, Y Sankarasubramaniam, E Cayirci, Wireless sensor networks: a survey. *Comput Netw.* **38**(4), 393–422 (2002). doi:10.1016/S1389-1286(01)00302-4
2. J-J Xiao, A Ribeiro, Z-Q Luo, GB Giannakis, Distributed compression-estimation using wireless sensor networks. *IEEE Signal Process Mag.* **23**(7), 27–41 (2006)
3. XR Li, Y Zhu, J Wang, C Han, Optimal linear estimation fusion—part I: unified fusion rules. *IEEE Trans Inf Theory* **49**(9), 2192–2208 (2003). doi:10.1109/TIT.2003.815774
4. A Ribeiro, GB Giannakis, Bandwidth-constrained distributed estimation for wireless sensor networks—part I: Gaussian case. *IEEE Trans Signal Process.* **54**(3), 1131–1143 (2006)
5. A Ribeiro, GB Giannakis, Bandwidth-constrained distributed estimation for wireless sensor networks—part II: unknown probability density function. *IEEE Trans Signal Process.* **54**(7), 2784–2796 (2006)

6. Z-Q Luo, An isotropic universal decentralized estimation scheme for a bandwidth constrained ad hoc sensor network. *IEEE J Sel Areas Commun.* **23**(4), 735–744 (2005)
7. H Li, J Fang, Distributed adaptive quantization and estimation for wireless sensor networks. *IEEE Signal Process Lett.* **14**(10), 669–672 (2007)
8. J Fang, H Li, Distributed adaptive quantization for wireless sensor networks: from delta modulation to maximum likelihood. *IEEE Trans Signal Process.* **56**(10), 5246–5257 (2008)
9. T Aysal, K Barner, Constrained decentralized estimation over noisy channels for sensor networks. *IEEE Trans Signal Process.* **56**(4), 1398–1410 (2008)
10. WM Lam, AR Reibman, Design of quantizers for decentralized estimation systems. *IEEE Trans Commun.* **41**(11), 1602–1605 (1993). doi:10.1109/26.241739
11. HC Papadopoulos, GW Wornell, AV Oppenheim, Sequential signal encoding from noisy measurements using quantizers with dynamic bias control. *IEEE Trans Inf Theory.* **47**(3), 978–1002 (2001). doi:10.1109/18.915654
12. J-J Xiao, Z-Q Luo, Decentralized estimation in an inhomogeneous sensing environment. *IEEE Trans Inf Theory* **51**(10), 3564–3575 (2005). doi:10.1109/TIT.2005.855580
13. J Li, G AlRegib, Distributed estimation in energy-constrained wireless sensor networks. *IEEE Trans Signal Process.* **57**(10), 3746–3758 (2009)
14. J-J Xiao, S Cui, Z-Q Luo, AJ Goldsmith, Power scheduling of universal decentralized estimation in sensor networks. *IEEE Trans Signal Process.* **54**(2), 413–422 (2006)
15. M Gastpar, To code or not to code, PhD Dissertation, Ecole Polytechnique Fédérale de Lausanne, EPFL. (2002)
16. M Gastpar, M Vetterli, Source-Channel Communication in Sensor Networks. *Lecture Notes in Computer Science.* **2634**, 162–177 (2003)
17. M Gastpar, Uncoded transmission is exactly optimal for a simple Gaussian “sensor” network, in *2007 Information Theory and Applications Workshop* 5247–5251 (Jan. 2007)
18. S Cui, J-J Xiao, AJ Goldsmith, Z-Q Luo, HV Poor, Energy-efficient joint estimation in sensor networks: Analog versus digital. in *IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP' 05. IV*, 745–748 (2005)
19. J-J Xiao, Z-Q Luo, Multiterminal source-channel communication over an orthogonal multiple access channel. *IEEE Trans Inf Theory* **53**(9), 3255–3264 (2007)
20. S Cui, J-J Xiao, AJ Goldsmith, Z-Q Luo, HV Poor, Estimation diversity and energy efficiency in distributed sensing. *IEEE Trans Signal Process.* **55**(9), 4683–4695 (2007)
21. K Bai, H Senol, C Tepedelenlioglu, Outage scaling laws and diversity for distributed estimation over parallel fading channels *IEEE Trans. Signal Process.* **57**(8), 3182–3192 (2009)
22. HR Admadi, A Vosoughi, Impact of channel estimation error on decentralized detection in bandwidth constrained wireless sensor networks, in *IEEE Military Communications Conference, MILCOM' 08*, 1–7 (Nov. 2008)
23. H Senol, C Tepedelenlioglu, Performance of distributed estimation over unknown parallel fading channels. *IEEE Trans Signal Process.* **56**(12), 6057–6068 (2008)
24. AP Dempster, NM Laird, DB Rubin, Maximum likelihood from incomplete data via the EM algorithm. *J R Stat Soc Ser B (Methodological)* **39**(1), 1–38 (1977)
25. J Max, Quantizing for minimum distortion. *IRE Trans Inf Theory* **6**(1), 7–12 (1960). doi:10.1109/TIT.1960.1057548
26. SP Lloyd, Least squares quantization in pcm. *IEEE Trans Inf Theory* **28**(2), 129–137 (1982). doi:10.1109/TIT.1982.1056489
27. [Online]. Available: http://en.wikipedia.org/wiki/Woodbury_matrix_identity
28. S Boyd, L Vandenberghe, *Convex Optimization* (Cambridge University Press, Cambridge, 2004)
29. SM Kay, *Fundamentals of Statistical Signal Processing, vol. I: Estimation Theory* (Prentice Hall PTR, New Jersey, 1993)
30. [Online]. Available: http://en.wikipedia.org/wiki/Mean_value_theorem
31. X Wang, C Yang, Optimal transmission codebook design in fading channels for decentralized estimation in wireless sensor networks, in *IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP' 09*, 2293–2296 (Apr. 2009)
32. JG Proakis, *Digital Communications*, 4th edn. (The McGraw-Hill Companies, Inc., New York, 2001)
33. M Wang, C Yang, Distributed estimation in wireless sensor networks with imperfect channel estimation, in *9th International Conference on Signal Processing, ICSP' 08.* **3**, 2649–2652 (Oct. 2008)

doi:10.1186/1687-6180-2011-132

Cite this article as: Wang and Yang: Decentralized estimation over orthogonal multiple-access fading channels in wireless sensor networks—optimal and suboptimal estimators. *EURASIP Journal on Advances in Signal Processing* 2011 **2011**:132.

Submit your manuscript to a SpringerOpen® journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com