Distributed Remote Estimation Over the Collision Channel With and Without Local Communication

Xu Zhang, *Member, IEEE*, Marcos M. Vasconcelos, *Member, IEEE*, Wei Cui, and Urbashi Mitra, *Fellow, IEEE*

***Abstract*—Internet of Things networks are the large-scale distributed systems consisting of a massive number of simple devices communicating, typically, over a shared wireless medium. This new paradigm requires novel ways of coordinating access to limited communication resources without introducing unreasonable delays. Herein, the optimal design of a remote estimation system with n sensors communicating with a fusion center via a collision channel of limited capacity $k \leq n$ is considered. In particular, for independent and identically distributed observations with a symmetric probability density function, we show that the problem of minimizing the mean-squared error with respect to a threshold strategy is quasi-convex. When coordination among sensors via a local communication network is available, the online learning of possibly unknown parameters of the probabilistic model is possible, enabling each sensor to optimize its own threshold autonomously. We propose two strategies for remote estimation with local communication: 1) one strategy swiftly reaches the performance of the optimal decentralized threshold policy and 2) the second strategy approaches the performance of the optimal centralized scheme with a slower convergence rate. A hybrid scheme that combines the best of both approaches is proposed, offering fast convergence and excellent performance.**

***Index Terms*—Decision theory, estimation, multi-agent systems, networked control systems, optimization.**

I. Introduction

INTERNET of Things (IoT) networks are systems comprised of a large number of low-cost devices intermittently transmitting small bundles of delay-sensitive data to access points or neighboring devices [1]. Specific IoT applications suffer from limited communication bandwidth. For instance, low power wide-area (LPWA) networks are a class of IoT systems explicitly designed to enable machine-to-machine communications characterized by infrequent transmitting nodes operating under strict low power, complexity, and bandwidth constraints [2]. LPWA communication protocols, such as LoRa and SigFox, are widely used to provide connectivity in static and mobile sensor networks [3]. However, these systems operate in *unlicensed* spectrum bands, which means that multiple simultaneous transmissions may lead to undesirable packet collisions and loss in performance.

Our goal is to develop new techniques for medium access control for IoT systems resilient to packet collisions. We consider the design of an LPWA network system for remote monitoring, where a large number of sensors communicate with a base station/fusion center under a strict constraint on bandwidth. For that purpose, we enable autonomous and distributed optimal allocation of limited communication resources by using event-triggered communication via a threshold strategy, e.g., [4] and references therein. In the absence of local coordination among sensors, event-triggered communication strategies improve the system performance by forcing the sensors to transmit only their most informative measurements. When local communication is available, by exchanging information at most with their immediate neighbors, each sensor can tune its thresholds to mitigate the negative effect of collisions even under incomplete knowledge of its measurements' underlying statistical model.

In this article, we study the remote sensing system depicted in Fig. 1, where n sensors observing independent and identically distributed continuous random variables communicate with a fusion center over a collision channel. The channel can only support the reliable transmission of at most k packets, where $k \leq n$. If the number of simultaneous transmissions is larger than k, a collision occurs. We are interested in the design of event-triggered transmission strategies to control channel access in a distributed way with the goal of optimizing estimation performance.

Our abstraction for a sensor network of multiple identical sensors communicating with a fusion center over a finite capacity collision channel using low complexity threshold policies conforms with the requirements of LPWA networks. The central

Manuscript received December 21, 2020; revised May 23, 2021; accepted June 16, 2021. Date of publication July 27, 2021; date of current version May 26, 2022. This work was supported in part by ONR under Grant N00014-15-1-2550, by NSF under Grant CNS-1213128, Grant CCF-1718560, Grant CCF-1410009, and Grant CPS-1446901, and in part by AFOSR under Grant FA9550-12-1-0215. Recommended by Associate Editor Y. Yuan. *(Corresponding author: Marcos M. Vasconcelos.)*

Xu Zhang and Wei Cui are with the School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China (e-mail: connorzx@bit.edu.cn; cuiwei@bit.edu.cn).

Marcos M. Vasconcelos is with the Commonwealth Cyber Initiative and the Bradley Department of Electrical Engineering, Virginia Tech, Arlington, VA 22203 USA (e-mail: marcosv@vt.edu).

Urbashi Mitra is with the Ming Hsieh Department of Electrical Engineering, University of Southern California, Los Angeles, CA 90089 USA (e-mail: ubli@usc.edu).

Digital Object Identifier 10.1109/TCNS.2021.3100405

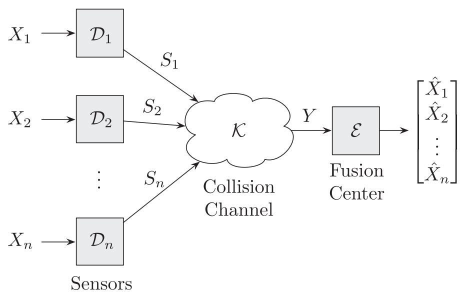

Fig. 1. System diagram for remote estimation over the collision channel.

insight here is that, in many sensing applications, the communication of uninformative measurements can be sacrificed without significant loss in performance, freeing resources to the remaining sensors in the network. This cooperation among sensors is the centerpiece of this article, which seeks to lay the foundations of a new framework for distributed communication protocols under assumptions of the observations' probabilistic model.

The optimal design of remote estimation systems has been of great interest in the past decade, and there exists rich literature on these systems under different technical assumptions. The optimization of transmission policies for a system involving a single sensor subject to a limited number of transmissions over a finite horizon was studied in [5]. Instead of limiting the number of transmissions, the authors of [6] considered the problem of minimizing an objective function consisting of the mean-squared error (MSE) plus transmission cost. Both [5] and [6] showed that symmetric threshold strategies are optimal under symmetry conditions of the probabilistic model of the measurements. Those results were later generalized in [7], which obtained similar structural results for a system with an energy harvesting sensor. An in-depth comprehensive survey of those and other earlier results can be found in [8]. More recently, connections between remote estimation and the notion of age of information have been established in [9].

Unlike the works mentioned above, this article considers the system with multiple sensors under a limited number of simultaneous transmissions. Remote sensing systems with multiple sensors sharing a collision channel were first considered in [10], which showed that the optimal transmission policies for symmetrically distributed observations were characterized by *asymmetric* thresholds. A similar sensing system with discrete observations was considered in [11]. Remote estimation of autoregressive Markov processes over the collision channel has also been considered in [12], in which symmetric threshold policies are used in a random access scheme. Notably, Chen *et al.* [12] also allude to the connection between remote estimation and age of information in a multisensor setting, thereby extending it [9]. Recent works in this area have incorporated reinforcement learning into sensor scheduling and remote estimation, when certain parameters of the system are unknown [13]–[15]. Another set of results considered other relevant issues concerning privacy [16], adversarial jamming [17], packet drops [18], and energy management [19], [20].

The system considered here is a multisensor system communicating over a collision channel of capacity $k \geq 1$. Instead of focusing on obtaining structural results, we concentrate on designing optimal thresholds for a symmetric system, corresponding to a symmetric team-decision problem. The sensors observations have identical statistical properties, i.e., they are drawn according to the same distribution. The symmetry assumption allows us to obtain a tractable quasi-convex optimization problem when the sensors observations are symmetric around its mean. In contrast, asymmetric formulations inevitably would lead to a nonconvex problem for which optimal solutions would be difficult to obtain. The performance of the optimal decentralized system is compared against a centralized scheme where only the sensors with the k most informative measurements are allowed to transmit. This is a generalization of the observation-driven scheduling problem for remote sensing studied in [21], in which a scheduler collects the measurements from all sensors, and chooses a single one to be transmitted to the destination. For independent observations distributed according to a symmetric probability density, [21] showed that a person-by-person optimal policy consists of sending the measurement with the largest magnitude to the fusion center. Remarkably, the performance of the optimal decentralized system without coordination for Gaussian observations is very close to the performance of the centralized scheme, which requires all the sensors to exchange their observations over a local communication network.

To close the existing gap in performance among the optimal decentralized scheme and the centralized one, we must allow for local communication. The additional gain in performance comes at a price in communication delay, since the sensors need to exchange messages locally to perform a distributed quantile regression task [22]–[24]. For a moderate to large number of sensors, distributed quantile estimation requires hundreds of local communication rounds; however, we propose a smart initialization mechanism based on average consensus [25], which improves the convergence time by orders of magnitude, allowing us to close the optimality gap swiftly. Moreover, we provide numerical evidence that our scheme performs very well even when the density is symmetric but unknown. Finally, we provide an upper bound on the expected switching-time between the average-consensus and the quantile-regression schemes.

In this article, we assume that each sensor's observations are drawn from a distribution with a symmetric pdf f_X. In Sections II and III, full knowledge of f_X is required. In Section IV-A, we assume that f_X belongs to a parametric class of densities with unknown parameters, e.g., $\mathcal{N}(0, \sigma^2)$. In Section IV-B and C, we assume that f_X is unknown.

Preliminary versions of Theorems 1 and 2 have appeared in [26], which only considered the scheme without local communication and contained partial proofs. In this article, we present complete proofs of the previous results, and additionally propose an exact centralized lower bound to Problem 1 plus three decentralized schemes based on local communication together with corresponding analysis and simulations.

The main contributions of this article are as follows.

1) In the absence of local communication, we study the design of a globally optimal threshold communication

strategy under a symmetry assumption of the probability distribution of the observations. We show that the MSE is a strictly quasi-convex function of the threshold, which is amenable to low complexity numerical optimization schemes. We provide numerical evidence that this optimal threshold policy is robust to perturbations of the underlying probabilistic model that violate the symmetry assumption.

2) In the presence of local communication, the sensors can coordinate to learn a common threshold strategy when the underlying probability distribution is not completely specified, or possibly completely unknown. We propose a consensus-based scheme for zero-mean Gaussian distributions and unknown variance, and a quantile estimation scheme for unknown zero-mean distributions. In the consensus-based scheme, each sensor estimates the unknown variance and then computes its threshold to determine locally whether to transmit or not. This scheme swiftly reaches the performance of the optimal decentralized scheme. In the quantile estimation scheme, each sensor estimates locally the kth largest observation among all observations and uses it as the threshold to decide whether to transmit or not. This scheme approaches the performance of the best known centralized scheme albeit with a slow convergence rate.
3) We propose a hybrid scheme that uses the consensus scheme via a Gaussian approximation to bootstrap the quantile estimation scheme. This scheme achieves both fast convergence and asymptotic performance close to the centralized policy. We provide a bound on the switching-time between the two schemes, and an example, which shows that the scheme is robust to distribution mismatch.

II. Problem Formulation

This section establishes the problem setup for a decentralized, remote estimation system over a collision channel of limited capacity. Consider the system diagram shown in Fig. 1. There are n sensors and a fusion center $\mathcal{E}$, which are connected by a collision channel $\mathcal{K}$. The ith sensor observes a zero-mean random variable X_i, $i \in \{1,\dots,n\}$. The random variables $\{X_i\}_{i=1}^n$ are independent and identically distributed (i.i.d.), and admit a probability density function (pdf) $f_X(x)$, such that $f_X(x) > 0$ for $x \in \mathbb{R}$. Each sensor decides whether to transmit its observed measurement to the fusion center or remain silent according to a threshold strategy.

Definition 1 (Threshold Strategy): Let $D_i \in \{0,1\}$ be the binary decision variable of the ith sensor, where $D_i = 1$ denotes that the sensor decides to transmit its measurement, and $D_i = 0$ denotes that the sensor decides to remain silent. A threshold strategy for the ith sensor is a function $\mathcal{D}_i : \mathbb{R} \to \{0,1\}$ such that

$$\mathcal{D}_i(x) \stackrel{\text{def}}{=} \mathbf{1}(|x| \geq T) \tag{1}$$

where $T \in [0, +\infty)$ denotes the threshold parameter and $\mathbf{1}(\mathfrak{S})$ denotes the indicator function of the statement $\mathfrak{S}$.

Remark 1: This formulation is an instance of a symmetric stochastic team [27]. This particular class of team decision problems is often more tractable because the optimization is over a single policy. It also allows to study the performance and robustness of optimal strategies with respect to the number of sensors. In some cases, it is possible to characterize the system's performance in the regime when the number of sensors is infinite, a particularly relevant feature for IoT applications.

After making a decision, each sensor produces a channel input packet, S_i, defined as follows:

$$S_i \stackrel{\text{def}}{=} \begin{cases} (i, X_i), & \text{if } D_i = 1 \\ \varnothing, & \text{if } D_i = 0 \end{cases}, \quad i \in \{1,\dots,n\}. \tag{2}$$

Remark 2: We assume that if a sensor decides to transmit, its unique identification number i is transmitted along with its measurement. This is done so that the receiver can identify the origin of the successfully received communication packets without ambiguity.

The collection of n sensors shares a collision channel $\mathcal{K}$ of limited capacity k, defined as follows:

Definition 2 (Collision Channel of Capacity k): The collision channel of capacity k allows the communication of at most $k \leq n$ simultaneous packets. Let $\mathbb{D} \stackrel{\text{def}}{=} \{i \mid D_i = 1\}$ denote the set of indices of all transmitting sensors. The output of the collision channel Y is given by the following:

$$Y \stackrel{\text{def}}{=} \begin{cases} \varnothing, & \text{if } |\mathbb{D}| = 0 \\ \{(i, X_i) \mid i \in \mathbb{D}\}, & \text{if } 1 \leq |\mathbb{D}| \leq k \\ (\mathfrak{C}, \mathbb{D}), & \text{otherwise.} \end{cases} \tag{3}$$

The special symbol $\mathfrak{C}$ denotes that a collision occurred and $\varnothing$ denotes that the channel is idle.

Assumption 1: When a collision occurs, we assume that the fusion center can decode the indices of the transmitting sensors.

Our purpose is to solve the following estimation problem over the collision channel under the normalized mean squared error (NMSE) criterion.

Problem 1: Assuming that each sensor uses a threshold strategy of the form of (1), given the number of sensors n, the pdf of the sensors' observations f_X, and the capacity of the collision channel k find a threshold T that minimizes the NMSE

$$\mathcal{J}_{n,k}(T) \stackrel{\text{def}}{=} \frac{1}{n} \mathbf{E}\left[\sum_{i=1}^{n} \left(X_i - \hat{X}_i\right)^2\right] \tag{4}$$

where the estimates $\hat{X}_i$ are given by the following:

$$\hat{X}_i \stackrel{\text{def}}{=} \mathbf{E}[X_i \mid Y], \quad i \in \{1,\dots,n\}. \tag{5}$$

III. Optimal Decentralized Scheme Without Local Communication

A. Quasi-Convexity of Problem 1

Assuming that there is no local communication among the sensors, and that the distribution of the observations is symmetric, Problem 1 can be solved exactly. We begin by deriving alternative expressions for (4) and (5). We will then show the quasi-convexity of Problem 1, which can thus be solved using low complexity numerical procedures.

Lemma 1: Provided the pdf f_X is symmetric; given the set of the decision variables $\mathbb{D} \stackrel{\text{def}}{=} \{i \mid D_i = 1\}$, the output of the estimator can be rewritten as

$$\hat{X}_i \stackrel{\text{w.p.1}}{=} \begin{cases} X_i, & \text{if } |\mathbb{D}| \leq k \text{ and } i \in \mathbb{D} \\ 0, & \text{otherwise} \end{cases} \quad i \in \{1, \ldots, n\}. \quad (6)$$

Proof: We will compute the conditional expectation in (5) for every possible output of the collision channel. When there is no collision and X_i was transmitted, i.e., $|\mathbb{D}| \leq k$ and $D_i = 1$, we have $(i, X_i) \in Y$, which implies that

$$\hat{X}_i = \mathbf{E}[X_i \mid Y] \stackrel{\text{w.p.1}}{=} X_i. \quad (7)$$

When a collision occurs and X_i was transmitted, i.e., $|\mathbb{D}| > k$ and $D_i = 1$, we have $Y = (\mathfrak{C}, \mathbb{D})$ and know $i \in \mathbb{D}$. Assumption 1 implies that

$$\hat{X}_i \stackrel{(a)}{=} \mathbf{E}[X_i \mid D_i = 1] = \mathbf{E}[X_i \mid |X_i| \geq T] \stackrel{(b)}{=} 0 \quad (8)$$

where (a) is due to $\{X\}_{i=1}^n$ being a collection of independent random variables, and (b) is due to the symmetry of f_X.

When X_i is not transmitted, the index i does not appear in the channel output Y, which implies that $D_i = 0$. In this case

$$\hat{X}_i = \mathbf{E}[X_i \mid D_i = 0] = \mathbf{E}[X_i \mid |X_i| < T] \stackrel{(c)}{=} 0 \quad (9)$$

where (c) is due to the symmetry of the pdf f_X.

Lemma 2: Let $\{X_i\}_{i=1}^n$ be an i.i.d. sequence distributed according to a symmetric pdf f_X. The objective function in Problem 1 can be expressed as follows:

$$\mathcal{J}_{n,k}(T) = \mathbf{E}\left[X^2\right] - \mathbf{E}\left[X^2 \mathbf{1}(|X| \geq T)\right] F_{n,k}(T) \quad (10)$$

where

$$F_{n,k}(T) \stackrel{\text{def}}{=} \sum_{\ell=0}^{k-1} \binom{n-1}{\ell} (1 - q(T))^\ell \, q(T)^{n-1-\ell} \quad (11)$$

and

$$q(T) \stackrel{\text{def}}{=} \mathbf{P}(|X| < T). \quad (12)$$

Proof: See Appendix A.

Theorem 1: The cost function $\mathcal{J}_{n,k}(T)$ in (10) is strictly quasi-convex and admits a unique optimal threshold $T^\star$ such that

$$T^\star = \arg\min_{T \geq 0} \mathcal{J}_{n,k}(T). \quad (13)$$

Proof: See Appendix B.

Remark 3: Theorem 1 holds for any symmetric pdf, regardless of the number of modes of the distribution. We highlight that proving quasi-convexity is typically a nontrivial task, and existing methods rely on composition rules of operations that preserve quasi-convexity, which are not available in our case. From an algorithmic standpoint, quasi-convexity is a property as desirable as convexity. Although a closed-form expression to $T^\star$ is unlikely to exist, we can compute it via iterative numerical methods. Due to the continuity and quasi-convexity of $\mathcal{J}_{n,k}(T)$ (established in Appendix B), we can use disciplined quasi-convex programming to compute the optimal threshold [28].

When using numerical optimization solvers, it is important to properly initialize the interval to be searched, especially when the support of the pdf f_X is unbounded. Next, we will provide an interval initialization by analyzing the 0–1 phase transition property of $F_{n,k}$. By inspection of (10), when T is such that $F_{n,k}(T) \approx 0$, the cost is $\mathcal{J}_{n,k}(T) \approx \mathbf{E}[X^2]$; when T is such that $F_{n,k}(T) \approx 1$, then $\mathcal{J}_{n,k}(T) \approx \mathbf{E}[X^2 \mathbf{1}(|X| < T)]$, which is nondecreasing in T. Therefore, the optimal $T^\star$ occurs in the interval when $F_{n,k}$ transitions from 0 to 1.

Lemma 3: Let $T^\star$ be the optimal threshold for the cost function $\mathcal{J}_{n,k}(T)$ in (10). Then

$$T^\star \geq q^{-1}\left(1 - \frac{k}{n}\right). \quad (14)$$

Proof: See Appendix C.

Lemma 4: Let $d > 0$. Then the following inequality holds: For $T > q^{-1}(\zeta)$:

$$F_{n,k}(T) \geq 1 - 10^{-d} \quad (15)$$

where ζ is the unique zero of

$$g_{n,k,d}(q) \stackrel{\text{def}}{=} k - (n-1)(1-q) - k \ln\left(\frac{k}{n-1}\right) + k \ln(1-q) + d \ln 10 \quad (16)$$

in the interval $(1 - k/(n-1), 1]$.

Proof: See Appendix D.

Theorem 2: There exists $\bar{\zeta} > 0$ such that

$$q^{-1}\left(1 - \frac{k}{n}\right) \leq T^\star \leq q^{-1}(\bar{\zeta}). \quad (17)$$

Proof: The proof follows from Lemmas 3 and 4 and the fact that one can always set d as large as necessary to guarantee that the upper bound includes the optimal threshold.

Remark 4: Theorem 2 provides an interval that is guaranteed to contain the optimal solution. Typically, $d = 2$ or 3 will suffice. The significance of Theorem 2 is that it let us avoid initializing a numerical solver where $\mathcal{J}_{n,k}(T)$ is flat, which may lead to falsely declaring that a local minimum has been found, and failing to find the unique global minimum guaranteed by Theorem 1.

B. Centralized Lower Bound to Problem 1

When the goal is to minimize the MSE of zero-mean independent variables such as in Problem 1, the best known centralized strategy consists of transmitting the k largest measurements in magnitude to the fusion center [21]. The performance of this strategy serves as a lower bound to decentralized communication strategies over the collision channel with capacity k. For the "top-k" strategy, the value of the cost function is given by

$$\mathcal{J}_{n,k}^L \stackrel{\text{def}}{=} \frac{1}{n} \sum_{i=k+1}^{n} \mathbf{E}\left[Z_{(i)}^2\right] \quad (18)$$

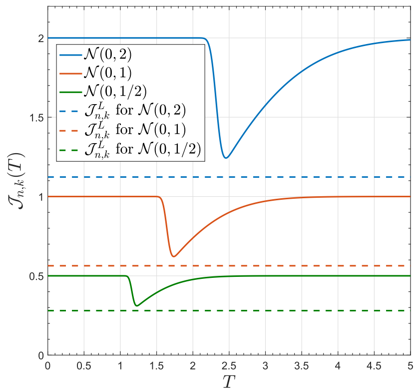

Fig. 2. Cost function $\mathcal{J}_{n,k}(T)$ as a function of the threshold T with $n = 1000$ sensors and a collision channel with capacity $k = 100$ packets for Gaussian observations of different variances. The dashed horizontal lines represent the corresponding centralized lower bounds $\mathcal{J}^L_{n,k}$ given in (18).

where $Z_i \stackrel{\text{def}}{=} |X_i|$, and $Z_{(i)}$ is defined as the ith largest value in $\{Z_\ell\}_{\ell=1}^n$ such that

$$Z_{(n)} \leq Z_{(n-1)} \leq \cdots \leq Z_{(1)}. \tag{19}$$

From results on ordered statistics [29], the second moment of $Z_{(i)}$ is given by the following:

$$\mathbf{E}\left[Z_{(i)}^2\right] = \frac{\int_0^\infty z^2 F_Z(z)^{n-i}\left(1 - F_Z(z)\right)^{i-1} f_Z(z)dz}{\mathrm{B}(n-i+1, i)} \tag{20}$$

where f_Z and F_Z are the pdf and cdf of Z, respectively, and $\mathrm{B}(\cdot,\cdot)$ denotes the beta function. Since $Z = |X|$, we have

$$F_Z(z) = 2F_X(z) - 1, \quad z \geq 0 \tag{21}$$

and

$$f_Z(z) = 2f_X(z), \quad z \geq 0. \tag{22}$$

This lower bound is used as benchmark in the examples shown in this article. The gap between the performance of the optimal threshold policy and the value of $\mathcal{J}^L_{n,k}$ corresponds to the loss due to decentralization.

C. Numerical Results

Fig. 2 shows the normalized MSE $\mathcal{J}_{n,k}(T)$ for a system with $n = 1000$ sensors and a collision channel of capacity $k = 100$ making Gaussian observations with different variances. We can observe the quasi-convexity property, and compare the performance of the optimal decentralized scheme $\mathcal{J}_{n,k}(T)$ to the centralized lower bound $\mathcal{J}^L_{n,k}$. From this figure, we can also observe that $\mathcal{J}_{n,k}(T)$ is flat at regions away from the optimal threshold $T^\star$. This observation reinforces the need for Theorem 2 and proper initialization of the numerical solvers used to compute $T^\star$. Table I shows the search intervals computed using Theorem 2 with $d = 3$, assuming Gaussian observations $\mathcal{N}(0, \sigma^2)$.

For a system with $n = 1000$ sensors, Fig. 3 displays the dependency of the optimal MSE $\mathcal{J}_{n,k}(T^\star)$ and the lower bound $\mathcal{J}^L_{n,k}$ as function of the capacity of the collision channel k for standard Gaussian observations, $X_i \sim \mathcal{N}(0,1)$. As the capacity k increases, more measurements are successfully received at the fusion center, and the NMSE decreases.

TABLE I
SEARCH INTERVAL FOR GAUSSIAN OBSERVATIONS. THE UPPER BOUND ON $T^\star$ IS FOUND USING LEMMA 4 WITH $d = 3$

σ^2	$n = 10, k = 1$	$n = 100, k = 10$	$n = 1000, k = 100$
1	$[1.64, 07.27]$	$[1.64, 4.01]$	$[1.64, 3.24]$
2	$[2.33, 10.28]$	$[2.33, 5.68]$	$[2.33, 4.59]$
3	$[2.85, 12.59]$	$[2.85, 6.95]$	$[2.85, 5.62]$
4	$[3.29, 14.54]$	$[3.29, 8.03]$	$[3.29, 6.48]$
5	$[3.68, 16.26]$	$[3.68, 8.98]$	$[3.68, 7.25]$

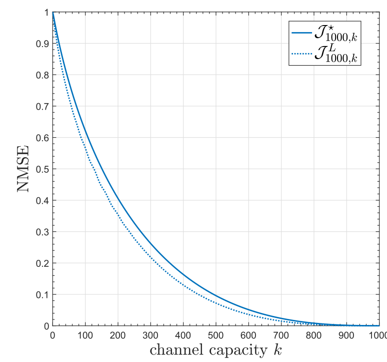

Fig. 3. Optimal cost $\mathcal{J}_{n,k}(T^\star)$ and the lower bound $\mathcal{J}^L_{n,k}$ as functions of the capacity of the collision channel k (performance of the optimal decentralized and centralized schemes) with $n = 1000$. The observations at the sensors are i.i.d. according to a standard Gaussian distribution, $X \sim \mathcal{N}(0,1)$.

We can also observe that the optimal choice for the threshold successfully mitigates the occurrence of collisions. Consequently, the decentralized scheme performs remarkably close to the centralized scheme. The gap between the solid (decentralized) and the dotted (centralized) curves is the performance loss due to decentralization.

The pdf's symmetry, identical distributions and known means are assumptions required to obtain our technical results. However, our optimal symmetric threshold policies are surprisingly robust in the absence of all three conditions mentioned above. Consider the following numerical example, where each sensor makes observations distributed according $X_i \sim \mathcal{N}(\mu_i, 1)$, where $\mu_i \in [-\epsilon, +\epsilon]$. We simulate the remote estimation system for 100 sample paths. For each sample path, a vector $\mathbf{d}$, where each component d_i is uniformly distributed over the interval $[-1, 1]$ is independently generated and kept fixed. Then, the mean $\mu_i = \epsilon d_i$ is computed. The value of the mean is unknown to the sensors and the estimator, which use a strategy designed under the assumption that $\mu_i = 0$, $i \in \{1, \ldots, n\}$. In this case, the observations are statistically independent; however, the pdfs are neither symmetric nor identically distributed. Adjusting the constant ϵ from 0 to 1, we can vary the degree of asymmetry.

We measure the NMSE using a Monte Carlo simulation for each point of each sample path, using 10^4 observation samples for each of the n sensors. In Fig. 4, we see that the performance of the system employing a symmetric threshold strategy designed for a nominal symmetric and i.i.d. system is robust to variations in the mean for $\epsilon \leq 0.25$. That robustness is consistent across

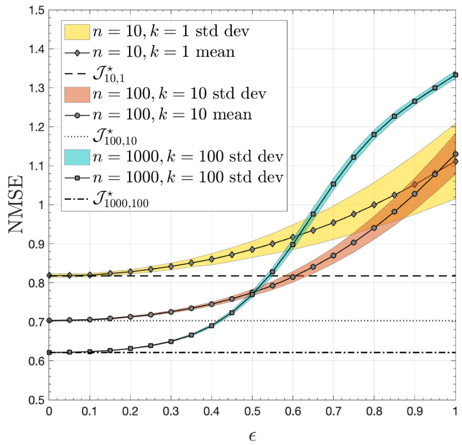

Fig. 4. Monte-Carlo Simulation for the remote estimation system for random perturbations on the mean $\mu_i \sim \mathcal{U}[-\epsilon, \epsilon]$ as a function of ϵ with $n = 10100, 1000$ sensors and capacity $k = 1, 10, 100$.

systems with a broad range of number of sensors. Clearly, when ϵ increases, the performance of the system degrades. The purpose of this simulation is to show that the system can be used, in practice, under moderate perturbations in the symmetry of the probabilistic model.

IV. Decentralized Schemes With Local Communication

Consider a connected undirected graph $\mathbb{G} = \mathbb{N}, \mathbb{E}$ with n nodes, each node represents a sensor observing an independent random variable as before. Here, $\mathbb{N} = \{1, \ldots, n\}$ denotes the set of sensors and $\mathbb{E} \subset \mathbb{N} \times \mathbb{N}$ denotes the set of edges between nodes. Let $\mathbb{N}_i$ denote the set of neighbors of the ith sensor, and $d_i \stackrel{\text{def}}{=} |\mathbb{N}_i|$. By local communication, we mean that if $(i, j) \in \mathbb{E}$, sensors i and j can communicate with each other for a given number of rounds before making their final decisions on whether to attempt a transmission to the fusion center or not. Each round of local communication has short range and represents one unit of accrued delay in communication between the sensors and fusion center.

A. Average Consensus-Based Decentralized Scheme

In many scenarios, we may not have access to one or more parameters of the pdf f_X although we know that the distribution is of a certain type, e.g., we may know that the distribution is Gaussian, but its mean and/or variance is unknown. By means of local communication among the sensors, we enable them to estimate the unavailable parameters such that the optimal threshold $T^\star$ may be computed in a decentralized way. This is done at the expense of some delay in communication with the fusion center. Using a distributed averaging algorithm, we describe how the method works in the Gaussian case. A similar procedure would also work for other symmetric densities such as Laplace, Uniform, Cauchy, and Logistic, among others.

Assume that the sensors make i.i.d. observations drawn from a zero-mean Gaussian distribution with known mean and *unknown* variance, e.g., $\mathcal{N}(0, \sigma^2)$. The observation from sensor i is denoted x_i. An unbiased estimator $\hat{\sigma}_n$ for the variance of a dataset $\{x_1, \ldots, x_n\}$ is given by $\hat{\sigma}_n \stackrel{\text{def}}{=} n^{-1} \sum_i x_i^2$.

The general case with a nonzero mean can also be easily handled, but requires an additional estimator for the sample mean at each sensor, and that each sensor transmits its sample mean estimate to the estimator.

Let $y_i(t)$ denote the local estimate of the sample variance at the ith sensor after the tth round of local communication. We initialize the local estimates by setting $y_i(0) \stackrel{\text{def}}{=} x_i^2, i = \{1, \ldots, n\}$. Using a distributed averaging scheme, each sensor estimates $\bar{y} \stackrel{\text{def}}{=} n^{-1} \sum_i^n y_i(0) = \hat{\sigma}_n$. On the tth round of local communication each sensor performs the following steps:

1) *Distributed Variance Estimation*: Each node updates its local estimate based on the local estimates of its neighbors according to the *Metropolis*[1] update rule [30]:

$$y_i(t+1) = y_i(t) + \sum_{j \in \mathbb{N}_i} \frac{1}{\max\{d_i, d_j\}} \left(y_j(t) - y_i(t)\right) \tag{23}$$

for $i \in \{1, \ldots, n\}$.

2) *Threshold Computation*. Using the techniques introduced in Section III and assuming that $X \sim \mathcal{N}(0, y_i(t))$, each node solves

$$T_i^\star(t) \stackrel{\text{def}}{=} \arg\min_{T \geq 0} \mathcal{J}_{n,k}(T) \tag{24}$$

where $\mathcal{J}_{n,k}(T)$ is given by (10).

If at time t the sensors use the thresholds $\{T_i^\star(t)\}_{i=1}^n$, the decision variables $u_i(t)$ are computed as follows:

$$u_i(t) = \mathbf{1}\left(|x_i| \geq T_i^\star(t)\right). \tag{25}$$

The instantaneous performance of this approximate scheme is given by the following:

$$\mathcal{J}_\pi\left(\boldsymbol{x}, \boldsymbol{u}(t)\right) \stackrel{\text{def}}{=} \begin{cases} \frac{1}{n}\sum_{i=1}^n x_i^2\left(1 - u_i(t)\right), & \text{if } \sum_{i=1}^n u_i(t) \leq k \\ \frac{1}{n}\sum_{i=1}^n x_i^2, & \text{if } \sum_{i=1}^n u_i(t) > k \end{cases} \tag{26}$$

where the notation $\pi = C$ is used to denote that the sensors are using a *consensus-based* scheme.

Fig. 5 (top) shows the empirical performance of the system obtained by generating 100 independent sample paths (one for each realization of the observation vector $\boldsymbol{x} \stackrel{\text{def}}{=} (x_1, \ldots, x_n)$). The mean and the standard deviation of the data $\mathcal{J}_C(\boldsymbol{x}, \boldsymbol{u}(t))$ are plotted. The underlying local communication graph $\mathbb{G}$ is sampled from the ensemble of geometric graphs with connectivity radius $r = \sqrt{\log_2 n/n}$, which is known to result in a connected graph with high probability [31]. One key observation here is that the mean of the sample paths converges to a value below the performance of the optimal scheme $\mathcal{J}^\star_{n,k}$. The reason behind this is that the thresholds are adapted to the observed data $\boldsymbol{x}$, which is the same data used to compute the empirical performance, resulting in a downward biased estimation of the true optimal solution $\mathcal{J}^\star_{n,k}$ [32]. Another observation is the fact

[1] Any other averaging scheme would be equally applicable, but possibly leading to different convergence rates.

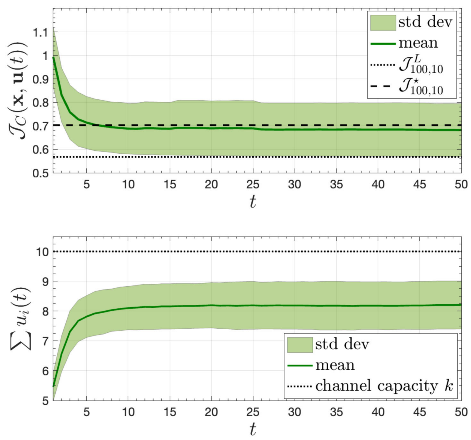

Fig. 5. Top figure shows the empirical performance of our scheme based on average consensus for 100 sample paths $\mathcal{J}_C(\boldsymbol{x}, \boldsymbol{u}(t))$ as a function of time t for a system with $n = 100$ sensors and channel capacity $k = 10$. The figure in the bottom shows the corresponding total number of transmitting sensors at a given time t.

that this strategy is conservative in terms of average number of transmissions relative to the channel capacity. Fig. 5 (bottom) shows that the average number of transmissions converges to 6.2, which is approximately 38% less than the maximum available capacity of $k = 10$. Moreover, the *maximum* number of simultaneous transmissions in the sample paths that are within the standard deviation (shown in the shaded green region) is 8, from which we conclude that the occurrence of collisions is effectively mitigated while achieving empirical performance very close to the theoretical optimal. We conjecture that this additional gap observed between the maximum number of simultaneous transmissions and the capacity is responsible for the robustness observed in Fig. 4.

B. Top-K Strategy Based on Decentralized Quantile Estimation

When local communication among sensors is available, nothing prevents the sensors to coordinate and attempt to implement the centralized top-k scheme outlined in Section III-B. Therefore, one possibility consists of each sensor exchanging messages to compute a local estimate of the kth ordered statistics $z_{(k)}$ as defined in (19) and use it as a threshold. Ideally, if the estimates are perfect, only the sensors holding the measurements within the top-k largest magnitudes will transmit. This strategy seeks to maximally exploit the available communication resources to achieve the best performance.

We will use a distributed subgradient method to estimate the sample quantile corresponding to the kth ordered statistics. Let $z_i \stackrel{\text{def}}{=} |x_i|$, $i \in \{1, \ldots, n\}$ and the pth sample quantile be defined as

$$\theta_p \stackrel{\text{def}}{=} \inf \left\{ \xi \mid \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}(z_i \leq \xi) \geq p \right\}. \tag{27}$$

Proposition 1 (Relationship Between Sample Quantiles and Ordered Statistics): Let $\{z_i\}_{i=1}^n$ be a sequence of realizations of the i.i.d. sequence of continuous random variables $\{Z_i\}_{i=1}^n$ and its corresponding nonincreasing reordering $\{z_{(i)}\}_{i=1}^n$. Then

$$p \in \left(\frac{n-k}{n}, \frac{n-k+1}{n} \right) \Rightarrow \theta_p = z_{(k)}. \tag{28}$$

Proof: The proof of this result relies on the fact that $\{z_i\}_{i=1}^n$ are realizations of continuous random variables, and therefore are distinct with probability 1. Setting p in the interval of (28) guarantees that the minimum data point ξ such that $\sum_{i=1}^{n} \mathbf{1}(z_i \leq \xi) \geq np$ holds is exactly $z_{(k)}$. ■

Let $w_i(t)$ denote the estimate of $z_{(k)}$ for the ith sensor at tth iteration and set $w_i(0) \stackrel{\text{def}}{=} z_i$, $i \in \{1, \ldots, n\}$. Let $\eta(t)$ be a deterministic step-size sequence, which is chosen as $\eta(t) = \alpha / t^{\tau}$, where $\alpha > 0$ and $\tau \in (0.5, 1]$.

On the tth round of local communication, we perform the following iteration:

$$w_i(t+1) = w_i(t) + \sum_{j \in \mathbb{N}_i} \frac{1}{\max\{d_i, d_j\}} (w_j(t) - w_i(t)) - \eta(t) s_i\left(z_i, w_i(t)\right) \tag{29}$$

where

$$s_i\left(z_i, w_i(t)\right) \stackrel{\text{def}}{=} \begin{cases} 1-p & z_i < w_i(t) \\ -p & z_i \geq w_i(t). \end{cases} \tag{30}$$

If at time t the ith sensor uses its quantile estimate $w_i(t)$ as a threshold, the decision variables $u_i(t)$ are computed as follows:

$$u_i(t) = \mathbf{1}\left(|x_i| \geq w_i(t)\right). \tag{31}$$

The instantaneous performance of this scheme is computed according to (26) with $\pi = Q$ denoting the fact that the sensors are using the *quantile-based* scheme.

Theorem 3: Let $p \in ((n-k)/n, (n-k+1)/n)$. If the local communication graph $\mathbb{G}$ is connected, then

$$\lim_{t \to +\infty} w_i(t) = z_{(k)}, \quad i \in \{1, \ldots, n\}. \tag{32}$$

Proof: From [22, Sec. 1.3, pp. 7–9], we know that

$$\theta_p = \arg\min_{\xi \in \mathbb{R}} \sum_{i=1}^{n} \rho_p\left(z_i - \xi\right) \tag{33}$$

where

$$\rho_p(x) \stackrel{\text{def}}{=} \begin{cases} (p-1)x & x < 0 \\ px & x \geq 0. \end{cases} \tag{34}$$

Equation (33) is a nonsmooth convex optimization problem, which can be distributed assuming that z_i is the local variable available only to sensor i. Therefore, the optimal solution can be obtained by interleaving the subgradient method with an average-consensus iteration [33]. Consider the iteration given by the following:

$$w_i(t+1) = \sum_{j \in \mathbb{N}_i} a_{ij} w_j(t) - \eta(t) s_i\left(z_i, w_i(t)\right) \tag{35}$$

where a_{ij} are the averaging coefficients, and $s_i(z_i, w_i(t))$ is a subgradient of $\rho_p z_i - \xi$ with respect to ξ at $w_i(t)$, e.g.,

$$s_i\left(z_i, w_i(t)\right) \overset{\text{def}}{=} \begin{cases} 1-p & z_i < w_i(t) \\ -p & z_i \geq w_i(t). \end{cases} \tag{36}$$

Let a_{ij} be the Metropolis averaging coefficients [34]. Under the assumptions that the step-size $\eta(t)$ is *square-summable, but not summable*, and that the local communication graph $\mathbb{G}$ is connected, (35) is guaranteed to converge to the optimal solution [30], i.e., θ_p, which, from Proposition 1, is equal to $z_{(k)}$ for $p \in ((n-k)/n, (n-k+1)/n)$.

One consequence of Theorem 3 is that for a large enough delay in communication, the performance of the scheme based on sample quantile estimation converges to the bounded interval, which is specified by the following result.

Corollary 1: Let $p \in ((n-k)/n, (n-k+1)/n)$. Then exists a number $M > 0$ such that for $t \geq M$,

$$\frac{1}{n} \sum_{i=k+1}^{n} Z_{(i)}^2 \leq \mathcal{J}_Q\left(\boldsymbol{X}, \boldsymbol{U}(t)\right) \leq \frac{1}{n} \sum_{i=k}^{n} Z_{(i)}^2 \tag{37}$$

Proof: From Theorem 3, we have

$$\lim_{t \to +\infty} w_i(t) = z_{(k)}, \quad i \in \{1, \ldots, n\}. \tag{38}$$

From the definition of limit, there exists a positive number

$$\varepsilon \overset{\text{def}}{=} \min\left\{z_{(k-1)} - z_{(k)}, z_{(k)} - z_{(k+1)}\right\} \tag{39}$$

and a sufficiently large number M such that

$$\left|w_i(t) - z_{(k)}\right| < \varepsilon, \quad t \geq M. \tag{40}$$

This implies that after M rounds of local communication, the thresholds $w_i(t)$ will lie in $(z_{(k+1)}, z_{(k-1)})$ for all $i \in \{1, \ldots, n\}$. Furthermore, for $t \geq M$, the number of transmissions will be either k or $k-1$. Therefore, either the top k or $k-1$ largest measurements will be sent to the remote estimator, resulting in the following inequality:

$$\frac{1}{n} \sum_{i=k+1}^{n} z_{(i)}^2 \leq \mathcal{J}_Q\left(\boldsymbol{x}, \boldsymbol{u}(t)\right) \leq \frac{1}{n} \sum_{i=k}^{n} z_{(i)}^2. \tag{41}$$

Fig. 6 illustrates the performance of the distributed quantile estimation scheme by computing the mean of 100 sample paths $\mathcal{J}_Q(\boldsymbol{x}, \boldsymbol{u}(t))$. The underlying graph is the same used in the simulation results in Section IV-A and the observations are standard Gaussian random variables. The chosen step-size sequence is $\eta(t) = 1/t^{0.51}$. Comparing Figs. 5 and 6, the asymptotic performance of the quantile estimation scheme is superior to the performance of the average consensus-based decentralized scheme.

However, the convergence rate of the quantile estimation scheme is considerably slower than the consensus-based system. There is a simple, intuitive argument for these performance

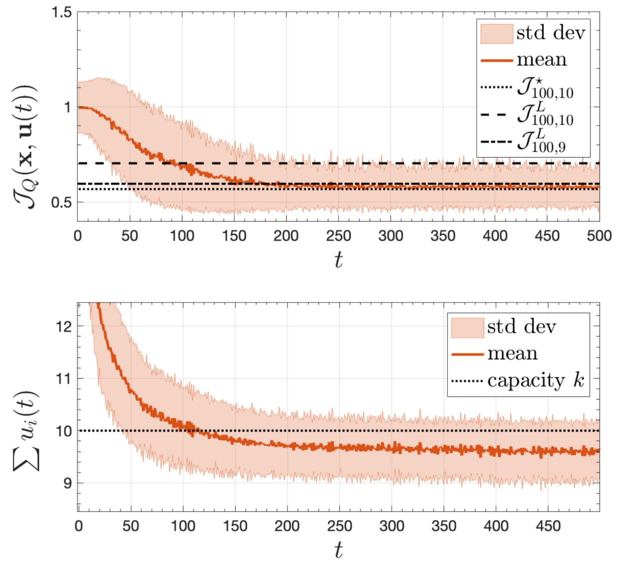

Fig. 6. Top figure shows the empirical performance of our scheme based on quantile estimation for 100 sample paths $\mathcal{J}_Q(\boldsymbol{x}, \boldsymbol{u}(t))$ as a function of time t for a system with $n = 100$ sensors and channel capacity $k = 10$. The figure in the bottom shows the corresponding total number of transmitting sensors at a given time t.

differences: The quantile estimation scheme's asymptotic performance is essentially the centralized method's performance. The sensors must exchange much more information at this performance level than what is needed to estimate the distribution's variance via average consensus. Moreover, the quantile estimation scheme seeks to eliminate the occurrence of collisions in the long run while at the same time approaching the channel's capacity. Notice that through quantile estimation, we operate at approximately 95% of the channel capacity. The average consensus-based strategy is conservative and works at 62% of the capacity limit. Hence, the existence of a performance gap.

C. Fast Quantile Estimation Decentralized Scheme

In this section, we introduce a hybrid scheme with a faster convergence rate and better or equal performance than both schemes presented so far. We will only assume that the pdf of the sensors' observations is zero-mean and symmetric, but is otherwise unknown. Let R be an integer such that when $t < R$, we use the consensus-based method in Section IV-A, which has a faster convergence rate; when $t = R$, each node uses the threshold computed by solving the optimization problem in (24) to initialize the quantile estimation scheme, i.e.,

$$w_i(R) = T_i^\star(R), \quad i \in \{1, \ldots, n\}. \tag{42}$$

After that, we use the quantile-based scheme as described in Section IV-B, which converges to a lower asymptotic cost. The instantaneous cost is given by $\mathcal{J}_F(\boldsymbol{x}, \boldsymbol{u}(t))$ defined as follows:

$$\mathcal{J}_F\left(\boldsymbol{x}, \boldsymbol{u}(t)\right) \overset{\text{def}}{=} \begin{cases} \mathcal{J}_C\left(\boldsymbol{x}, \boldsymbol{u}(t)\right), & \text{if } t < R \\ \mathcal{J}_Q\left(\boldsymbol{x}, \boldsymbol{u}(t)\right), & \text{if } t \geq R. \end{cases} \tag{43}$$

The switching-time R is chosen as the time when all of the local thresholds at the sensors $T_i^\star(t)$ are close to the threshold computed with access to entire dataset $\bar{T}^\star$. Since the convergence-time depends on the data, which is random, we define the *expected threshold agreement time*.

Definition 3: Let $\delta > 0$, the *expected threshold agreement time* is defined as

$$R(\delta) \overset{\text{def}}{=} \mathbf{E}\left[\min\{t \in \mathbb{Z}_{\geq 0} \mid \max_{i\in\{1,\ldots,n\}} |T_i^\star(t) - \bar{T}^\star| < \delta\ \}\right] \tag{44}$$

where the expectation is taken over the sensors' observations $X_1, \ldots, X_n$.

Intuitively, the threshold agreement time $R(\delta)$ is a function of the graph's connectivity (through the averaging matrix $\mathbf{A}$), the cost function $\mathcal{J}$, and the probability distribution of $X_1, \ldots, X_n$.

The next result provides a useful upper bound to the expected threshold agreement time, which can be also used as a switching-time.

Theorem 4: Let $\mathbf{A}$ be the averaging matrix used along with the local communication graph $\mathbb{G}$, the vector $Y(0) = \text{vec}(X_1^2, \ldots, X_n^2)$, and the function $\tau : \mathbb{R} \to \mathbb{R}$ be defined as

$$\tau(\sigma^2) \overset{\text{def}}{=} \arg\min_{T\geq 0} \mathcal{J}_{n,k}(T) \tag{45}$$

assuming that $X_i \sim \mathcal{N}(0, \sigma^2),\ i \in \{1, \ldots, n\}$. Then the following upper bound to the expected threshold agreement time holds:

$$R(\delta) \leq \mathbf{E}\left[\min\left\{t \in \mathbb{Z}_{\geq 0} \,\middle|\, \left\| \left(\mathbf{A}^t - \frac{1}{n}\mathbf{1}\mathbf{1}^T\right) Y(0)\right\|_\infty < \frac{\delta}{|\tau'(n^{-1}\mathbf{1}^T Y(0))|}\right\}\right] \tag{46}$$

where τ' is the derivative of (45) with respect to σ^2 (i.e., the sensitivity function), and the expectation is taken with respect to $X_1, \ldots, X_n$.

Proof: Consider the function $\tau(\sigma^2)$ defined in (45), which is continuously differentiable and concave on $(0, +\infty)$ (cf. Lemma 6 in Appendix E), the following holds for all $i \in \{1, \ldots, n\}$:

$$|T_i^\star(t) - \bar{T}^\star| = |\tau(y_i(t)) - \tau(\bar{y})| \leq |\tau'(\bar{y})| \cdot |y_i(t) - \bar{y}| \tag{47}$$

where $\bar{y} = n^{-1}\sum y_i(0)$, and $y_i(t)$ is the ith sensor's estimate of $\bar{y}$ at time t. Thus,

$$\max_{i\in\{1,\ldots,n\}} |T_i^\star(t) - \bar{T}^\star| \leq |\tau'(\bar{y})| \cdot \|\boldsymbol{y}(t) - \mathbf{1}\bar{y}\|_\infty = |\tau'(\bar{y})| \cdot \left\|\left(\mathbf{A}^t - \frac{1}{n}\mathbf{1}\mathbf{1}^T\right)\boldsymbol{y}(0)\right\|_\infty. \tag{48}$$

Let $\mathbb{S}(\delta)$ and $\mathbb{C}(\delta)$ be defined as

$$\mathbb{S}(\delta) \overset{\text{def}}{=} \left\{t \in \mathbb{Z}_{\geq 0} \,\middle|\, \left\|\left(\mathbf{A}^t - \frac{1}{n}\mathbf{1}\mathbf{1}^T\right)\boldsymbol{y}(0)\right\|_\infty < \delta/|\tau'(\bar{y})|\right\} \tag{49}$$

and

$$\mathbb{C}(\delta) \overset{\text{def}}{=} \left\{t \in \mathbb{Z}_{\geq 0} \,\middle|\, \max_{i\in\{1,\ldots,n\}} |T_i^\star(t) - \bar{T}^\star| < \delta\right\}. \tag{50}$$

Notice that $\mathbb{S}(\delta) \subseteq \mathbb{C}(\delta) \Rightarrow \min \mathbb{C}(\delta) \leq \min \mathbb{S}(\delta)$. Taking the expectation with respect to $X_1, \ldots, X_n$ yields

$$R(\delta) \leq \mathbf{E}[\min \mathbb{S}(\delta)] \overset{\text{def}}{=} \bar{R}(\delta). \tag{51}$$

TABLE II
SWITCHING-TIME $\bar{R}_{n,k}$ FOR DIFFERENT VALUES OF δ

δ	$\bar{R}_{10,1}$	$\bar{R}_{100,10}$	$\bar{R}_{1000,100}$
10^{-1}	15.35 ± 0.39	21.84 ± 0.47	41.39 ± 1.16
10^{-2}	34.86 ± 0.51	63.85 ± 0.92	218.28 ± 3.09
10^{-3}	56.34 ± 0.57	112.38 ± 1.17	441.94 ± 3.62
10^{-4}	76.73 ± 0.59	162.10 ± 1.34	668.71 ± 4.01
10^{-5}	97.21 ± 0.53	213.10 ± 1.43	903.29 ± 4.07

For the graphs that we have been using in our numerical results, we have computed switching-times based on the upper-bound of Theorem 4. Table II contains the switching-times $\bar{R}_{n,k}$ corresponding to different values of δ for systems with n sensors and channel capacity k. The underlying graphs $\mathbb{G}$ are random Geometric graphs with $n = 10100$ and 1000 nodes and are fixed[2]. The expectation in the upper-bound is approximated using 1000 Monte-Carlo simulations, for which the mean and its 95% confidence interval are provided.

D. Mismatched Distributions and Illustrative Example

To broaden our schemes' applicability, consider the case when the distribution is not necessarily Gaussian. Still, we use a Gaussian approximation, i.e., we perform the local threshold design, assuming that the measurements are drawn from a Gaussian distribution with unknown variance. In this "mismatched" distribution scenario, we use consensus to estimate the variance of the distribution. At each iteration, each sensor solves (13) assuming the distribution is $\mathcal{N}(0, y_i(t))$. At a certain point $t = R$, we initialize the quantile estimation scheme using (42).

The parameter δ used in the fast quantile estimation scheme is chosen to be 10^{-1}, which implies in a switching time $R = 22$. Assuming the measurements are drawn from a Laplacian distribution, $X_i \sim \mathcal{L}(0, 1),\ i \in \{1, \ldots, n\}$, we compare the performance of the regular quantile estimation scheme with its accelerated counterpart. The numerical results shown in Fig. 7 show that the fast quantile estimation scheme is approximately two orders of magnitude faster, even when the design is done based on mismatched distributions. The reason why the hybrid scheme is so effective, is that the consensus scheme quickly synchronizes the local estimates at the sensors to a threshold close to the kth ordered statistics, accelerating the overall convergence of the quantile-based scheme. This synchronization also leads to smoother sample paths, whereas the sample paths of the pure quantile estimation scheme display large oscillations when different local estimates are approaching the kth ordered statistics (cf. shaded regions in Fig. 7). Finally, since the quantile estimation scheme is independent of the distributions, the asymptotic performance is unaffected by the mismatch. That is the reason why the hybrid scheme is able to achieve the desirable features of both schemes.

V. CONCLUSION

In the first part of this article, we have studied the design of threshold strategies for a remote estimation system over the

[2]All data and code discussed in this article is available at GitHub https://github.com/mullervasconcelos/collision-quantile.git

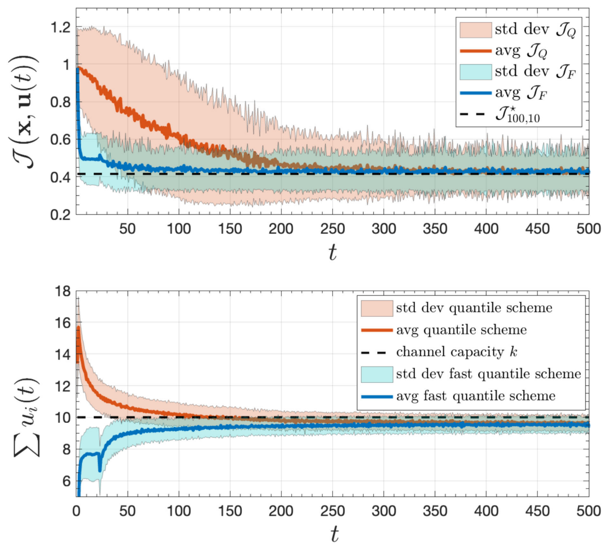

Fig. 7. Performance of the quantile estimation scheme and its accelerated counterpart. The observations are i.i.d. according to a Laplacian distribution, $X_i \sim \mathcal{L}(0,1)$, $i \in \{1,\ldots,n\}$. The number of sensors is $n = 100$, and capacity is $k = 10$.

collision channel without local communication. We showed that when the observations across sensors are i.i.d. and the pdf is symmetric, there exists a unique optimal threshold for the NMSE criterion. Our numerical results show that the optimal threshold strategy has a performance reasonably close to a clairvoyant centralized lower bound and is robust to moderate perturbations on the symmetry assumptions.

In the second part of this article, we assume the measurements are i.i.d. according to symmetric probability distribution belonging to a parametric class with unknown parameters. Under this partial distribution knowledge, local communication among sensors enables distributed learning of the optimal thresholds. We presented three schemes as follows:

1) A fast converging consensus-based design, where each sensor estimates missing distribution parameters from data and then solve a local optimization problem.
2) A distributed quantile estimation strategy that implements the optimal centralized policy at the price of a slower convergence rate.
3) A hybrid scheme with fast convergence and excellent performance.

Many open questions stem from this work. One of them is the characterization of the system robustness concerning perturbations of the probabilistic model. For example, we would like to characterize the system's performance loss when the imperfect local parameter estimates are used instead of the nominal values. Another important open question pertains to the possibility of correlation among sensors' observations and properties of the objective function, existence, and uniqueness of optimal solutions. More importantly, whether the use of asymmetric threshold strategies would imply substantial gains in performance. Finally, the convergence rate analysis of the consensus and quantile-based schemes and their network scaling properties are still open questions.

APPENDIX A
PROOF OF LEMMA 2

We begin by defining the event $\mathfrak{E}_{i,k} \overset{\text{def}}{=} \{\sum_{\ell\neq i} D_\ell \leq k-1\}$. Using the law of total expectation, the objective function in (4) becomes

$$\mathcal{J}_{n,k}(T) = \frac{1}{n}\sum_{i=1}^{n}\Big[\mathbf{E}[(X_i-\hat{X}_i)^2 \mid D_i=0]\mathbf{P}(D_i=0) + \mathbf{E}[(X_i-\hat{X}_i)^2 \mid D_i=1, |\mathbb{D}|>k]\mathbf{P}(D_i=1,|\mathbb{D}|>k) + \mathbf{E}[(X_i-\hat{X}_i)^2 \mid D_i=1, |\mathbb{D}|\leq k]\mathbf{P}(D_i=1,|\mathbb{D}|\leq k)\Big]. \tag{52}$$

The independence of D_i and $\{D_\ell\}_{\ell\neq i}$ yields

$$\mathcal{J}_{n,k}(T) = \frac{1}{n}\sum_{i=1}^{n}\Big[\mathbf{E}[(X_i-\hat{X}_i)^2 \mid D_i=0]\mathbf{P}(D_i=0) + \mathbf{E}[(X_i-\hat{X}_i)^2 \mid D_i=1, \mathfrak{E}^c_{i,k}]\mathbf{P}(D_i=1)\mathbf{P}(\mathfrak{E}^c_{i,k}) + \mathbf{E}[(X_i-\hat{X}_i)^2 \mid D_i=1, \mathfrak{E}_{i,k}]\mathbf{P}(D_i=1)\mathbf{P}(\mathfrak{E}_{i,k})\Big]. \tag{53}$$

Substituting (6) in (53), the cost function becomes

$$\mathcal{J}_{n,k}(T) = \frac{1}{n}\sum_{i=1}^{n}\big[\mathbf{E}\left[X_i^2\right] - \mathbf{E}\left[X_i^2\mathbf{1}(D_i=1)\right]\mathbf{P}(\mathfrak{E}_{i,k})\big]. \tag{54}$$

Since $\{X_i\}_{i=1}^n$ are i.i.d., after some algebra we have

$$\mathcal{J}_{n,k}(T) = \mathbf{E}[X^2] - \mathbf{E}[X^2\mathbf{1}(|X|\geq T)]F_{n,k}(T). \tag{55}$$

APPENDIX B
PROOF OF THEOREM 1

From [29], the derivative of $F_{n,k}(T)$ is

$$F'_{n,k}(T) = 2\,k\binom{n-1}{k}q(T)^{n-1-k}\left(1-q(T)\right)^{k-1}f_X(T). \tag{56}$$

We shall show that there is a unique $T^\star$ such that the derivative $\mathcal{J}'_{n,k}(T)$ is zero for $T>0$, where

$$\mathcal{J}'_{n,k}(T) = 2\,T^2 f_X(T)F_{n,k}(T) - \mathbf{E}\left[X^2\mathbf{1}(|X|\geq T)\right]F'_{n,k}(T). \tag{57}$$

Due to the fact that $f_X(T)>0$ for $T>0$, (56) implies that $F'_{n,k}(T)>0$ for $T>0$. So we have that

$$\frac{\mathcal{J}'_{n,k}(T)}{F'_{n,k}(T)} = \underbrace{\left[\frac{2\,T^2 f_X(T)F_{n,k}(T)}{F'_{n,k}(T)} - \mathbf{E}\left[X^2\mathbf{1}(|X|\geq T)\right]\right]}_{\overset{\text{def}}{=}h(T)}. \tag{58}$$

Incorporating (11) and (56) into $h(T)$ yields

$$h(T) = T^2 q(T) \sum_{j=0}^{k-1} \frac{(k-1)!(n-1-k)!}{j!(n-1-j)!} \left(\frac{q(T)}{1-q(T)}\right)^{k-j-1} - \mathbf{E}\left[X^2 \mathbf{1}(|X| \geq T)\right]. \quad (59)$$

Since $q(T) \in (0,1)$ is strictly increasing for $T > 0$, $1/(1-q(T)) \in (0,\infty)$ is also strictly increasing for $T > 0$. The product of two positive and strictly increasing functions is a positive strictly increasing function, which implies that $q(T)/(1-q(T)) \in (0,+\infty)$ is a strictly increasing function.

Since $k-j-1 \geq 0$ for $j \in \{0,\ldots,k-1\}$, we obtain that $[q(T)/(1-q(T))]^{k-j-1} \in (0,+\infty)$ is a nondecreasing function. The sum of nondecreasing functions is a non-decreasing function, which implies that

$$\sum_{j=0}^{k-1} \left[\frac{(k-1)!(n-1-k)!}{j!(n-1-j)!} \left(\frac{q(T)}{1-q(T)}\right)^{k-j-1}\right] \quad (60)$$

is a positive nondecreasing function. Using the fact that the product of a positive strictly increasing and a positive nondecreasing function is a strictly increasing function, we get

$$T^2 q(T) \sum_{j=0}^{k-1} \left[\frac{(k-1)!(n-1-k)!}{j!(n-1-j)!} \left(\frac{q(T)}{1-q(T)}\right)^{k-j-1}\right] \quad (61)$$

which is a positive strictly increasing function.

Combining with the fact that $-\mathbf{E}[X^2\mathbf{1}(|X| \geq T)]$ is strictly increasing continuous function for $T > 0$, we conclude $h(T)$ is a strictly increasing continuous function of T for $T > 0$. When $T \to 0^+$, we have $\mathbf{E}[X^2\mathbf{1}(|X| \geq T)] > 0$ and $q(T) \to 0$, which implies $\inf_T h(T) = \lim_{T\to 0^+} h(T) < 0$. When $T \to +\infty$, we have $\mathbf{E}[X^2\mathbf{1}(|X| \geq T)] \to 0$ and $q(T) \to 1$, which implies $\sup_T h(T) = \lim_{T\to+\infty} h(T) > 0$.

Therefore, there exists only one $T^\star \in (0,+\infty)$ such that $h(T^\star) = 0$. Since $f_{n,k}(T) > 0$ and $\mathcal{J}'_{n,k}(T) = f_{n,k}(T)h(T)$, there is a unique $T^\star$ that minimizes $\mathcal{J}_{n,k}(T)$ for $T \in (0,+\infty)$. Combining with the fact that $\mathcal{J}_{n,k}(T)$ is a continuous function, $T^\star$ is the optimal threshold.

APPENDIX C
PROOF OF LEMMA 3

Recall that $h(T)$ in (59) is a strictly increasing continuous function of T, establishing the inequality $T^\star > T_c$ is equivalent to showing that $h(T_c) < h(T^\star) = 0$, where $T_c \stackrel{\text{def}}{=} q^{-1}1 - \text{k/n}$. Using (59) yields

$$h(T_c) = \frac{T_c^2}{n} \sum_{j=0}^{k-1} \frac{k!(n-1-k)!}{j!(n-1-j)!} \left(\frac{n-k}{k}\right)^{k-j} - \mathbf{E}\left[X^2\mathbf{1}(|X| \geq T_c)\right] \quad (62)$$

$$< \frac{T_c^2}{n} \sum_{j=0}^{k-1} \frac{k!(n-1-k)!}{j!(n-1-j)!} \left(\frac{n-k}{k}\right)^{k-j} - \frac{kT_c^2}{n} \leq 0 \quad (63)$$

where the first inequality follows from

$$\mathbf{E}\left[X^2\mathbf{1}(|X| \geq T_c)\right] = \int_{T_c}^{+\infty} x^2 f_Z(x)dx \quad (64)$$

$$> T_c^2(1 - F_Z(T_c)) = \frac{kT_c^2}{n} \quad (65)$$

and the last inequality follows from

$$\sum_{j=0}^{k-1} \frac{k!(n-1-k)!}{j!(n-1-j)!} \left(\frac{n-k}{k}\right)^{k-j} \leq k. \quad (66)$$

APPENDIX D
PROOF OF LEMMA 4

Lemma 5 (Chernoff Bound [35]): Let $D_1,\ldots,D_n$ be independent Bernoulli random variables such that $\mathbf{P}(D_i = 1) = 1 - q_i$. Let $S_n = \sum_{i=1}^n D_i$ and $\mu = \mathbf{E}[S_n]$. Then the following inequality holds: For $\delta > 0$

$$\mathbf{P}(S_n \geq (1+\delta)\mu) \leq \left(\frac{e^\delta}{(1+\delta)^{(1+\delta)}}\right)^\mu. \quad (67)$$

We begin to prove Lemma 4. Since $\{D_i\}$ is an i.i.d. sequence of Bernoulli random variables with probability $\mathbf{P}(D_i = 1) = 1 - q(T)$, we have $\mu = (n-1)(1-q(T))$. Let $\delta = \frac{k}{\mu} - 1$, and notice that

$$\delta > 0 \Leftrightarrow q(T) > 1 - k/(n-1). \quad (68)$$

Then, the following holds:

$$F_{n,k}(T) = 1 - \mathbf{P}\left(\sum_{i=1}^{n-1} D_i \geq k\right) \overset{(a)}{\geq} 1 - \left(\frac{e^\delta}{(1+\delta)^{(1+\delta)}}\right)^\mu \quad (69)$$

where (a) follows from Lemma 5.

Let $d > 0$. We would like to find $\bar{T}$ such that $F_{n,k}(T) \geq 1 - 10^{-d}$, for $T \geq \bar{T}$. Therefore, we would like to solve

$$\left(\frac{e^\delta}{(1+\delta)^{(1+\delta)}}\right)^\mu = 10^{-d} \quad (70)$$

which is equivalent to finding a zero of $g_{n,k,d}(q)$ defined in (16). We will show that $g_{n,k,d}(q)$ always admits a unique zero in the interval $(1-k/(n-1), 1]$.

Since $g_{n,k,d}(q)$ is differentiable, we have

$$g'_{n,k,d}(q) = (n-1) - \frac{k}{1-q} < 0 \quad (71)$$

which means that $g_{n,k,d}(q)$ is strictly decreasing on the interval $(1-k/(n-1), 1)$. Moreover, when $q \to 1 - k/(n-1)$, we have $g_{n,k,d}(q) \to d\ln 10 > 0$. When $q \to 1$, we have $g_{n,k,d}(q) \to -\infty$. Thus, since $g_{n,k,d}(q)$ is continuous and strictly decreasing on $(1-k/(n-1), 1)$, there is a unique zero ζ in the interval $(1-k/(n-1), 1]$. Let $\bar{T} \stackrel{\text{def}}{=} q^{-1}(\zeta)$.

APPENDIX E

Lemma 6: Assuming that $X \sim \mathcal{N}(0,\sigma^2)$, define the function $\tau : \mathbb{R}_{\geq 0} \to \mathbb{R}$ as in (45). The function τ is concave. More specifically: There exists a constant $c > 0$ such that

$$\tau(\sigma^2) = c\sqrt{2\sigma^2}. \tag{72}$$

Proof: Recalling that $\tau(\sigma^2)$ satisfies $h(\tau(\sigma^2)) = 0$, where h is given by (59). Define the function $r : \mathbb{R}_{\geq 0} \to \mathbb{R}$ as

$$r(x) = x^2 v_1(x) \sum_{j=0}^{k-1} \frac{(k-1)!(n-1-k)!}{j!(n-1-j)!} \left(\frac{v_1(x)}{1 - v_1(x)}\right)^{k-j-1} - v_2(x) \tag{73}$$

where

$$v_1(x) \overset{\text{def}}{=} \frac{2}{\sqrt{\pi}} \int_0^x e^{-u^2} du \text{ and } v_2(x) \overset{\text{def}}{=} \frac{2}{\sqrt{\pi}} \int_x^{+\infty} u^2 e^{-u^2} du. \tag{74}$$

We can show that $r(x)$ is a strictly increasing continuous function of x and there exists a unique constant $c > 0$ such that $r(c) = 0$. Moreover, for $X \sim \mathcal{N}(0, \sigma^2)$, we have

$$r\left(\frac{\tau(\sigma^2)}{\sqrt{2\sigma^2}}\right) = \frac{h(\tau(\sigma^2))}{2\sigma^2} = 0. \tag{75}$$

Since there is a unique zero c such that $r(c) = 0$, we have $\tau(\sigma^2) = c\sqrt{2\sigma^2}$.

References

[1] A. Bakshi, L. Chen, K. Srinivasan, C. E. Koksal, and A. Eryilmaz, "EMIT: An efficient MAC paradigm for the Internet of Things," *IEEE Trans. Netw.*, vol. 27, no. 4, pp. 1572–1583, Aug. 2019.

[2] M. R. Palattella *et al.*, "Internet of things in the 5G era: Enablers, architecture, and business models," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 3, pp. 510–527, Mar. 2016.

[3] H. Lee and K. Ke, "Monitoring of large-area IoT sensors using a LoRa wireless mesh network system: Design and evaluation," *IEEE Trans. Instrum. Meas.*, vol. 67, no. 9, pp. 2177–2187, Sep. 2018.

[4] S. Weerakkody, Y. Mo, B. Sinopoli, D. Han, and L. Shi, "Multi-sensor scheduling for state estimation with event-based, stochastic triggers," *IEEE Trans. Autom. Control*, vol. 61, no. 9, pp. 2695–2701, Sep. 2016.

[5] O. C. Imer and T. Basar, "Optimal estimation with limited measurements," *Int. J. Syst., Control Commun.*, vol. 2, no. 1, pp. 5–29, 2010.

[6] G. M. Lipsa and N. C. Martins, "Remote state estimation with communication costs for first-order LTI systems," *IEEE Trans. Autom. Control*, vol. 56, no. 9, pp. 2013–2025, Sep. 2011.

[7] A. Nayyar, T. Basar, D. Teneketzis, and V. Veeravalli, "Optimal strategies for communication and remote estimation with an energy harvesting sensor," *IEEE Trans. Autom. Control*, vol. 58, no. 9, pp. 2246–2260, Sep. 2013.

[8] M. M. Vasconcelos and N. C. Martins, "A survey on remote estimation problems," in*Proc. Princ. Cyber-Phys. Syst., Interdiscipl. Approach*, 2020, pp. 81–103.

[9] J. Yun, C. Joo, and A. Eryilmaz, "Optimal real-time monitoring of an information source under communication costs," in *Proc. IEEE Conf. Decis. Control*, 2018, pp. 4767–4772.

[10] M. M. Vasconcelos and N. C. Martins, "Optimal estimation over the collision channel," *IEEE Trans. Autom. Control*, vol. 62, no. 1, pp. 321–336, Jan. 2017.

[11] M. M. Vasconcelos and N. C. Martins, "Optimal remote estimation of discrete random variables over the collision channel," *IEEE Trans. Autom. Control*, vol. 64, no. 4, pp. 1519–1534, Apr. 2019.

[12] X. Chen, X. Liao, and S. S. Bidokhti, "Real-time sampling and estimation on random access channels: Age of information and beyond," in *Proc. IEEE Conf. Comput. Commun.*, 2021, pp. 1–10.

[13] Y. Li, J. Wu, and T. Chen, "Transmit power control and remote state estimation with sensor networks: A Bayesian inference approach," *Automatica*, vol. 97, pp. 292–300, 2018.

[14] S. Wu, X. Ren, Q. Jia, K. H. Johansson, and L. Shi, "Learning optimal scheduling policy for remote state estimation under uncertain channel condition," *IEEE Control Netw. Syst.*, vol. 7, no. 2, pp. 579–591, Jun. 2020.

[15] A. S. Leong, A. Ramaswamy, D. E. Quevedo, H. Karl, and L. Shi, "Deep reinforcement learning for wireless sensor scheduling in cyber-physical systems," *Automatica*, vol. 113, 2020, Art. no. 108759.

[16] A. S. Leong, S. Dey, and D. E. Quevedo, "Transmission scheduling for remote state estimation and control with an energy harvesting sensor," *Automatica*, vol. 91, pp. 54–60, 2018.

[17] X. Gao, E. Akyol, and T. Basar, "Communication scheduling and remote estimation with adversarial intervention," *IEEE/CAA J. Automatica Sinica*, vol. 6, no. 1, pp. 32–44, Jan. 2019.

[18] S. Wu, X. Ren, S. Dey, and L. Shi, "Optimal scheduling of multiple sensors over shared channels with packet transmission constraint," *Automatica*, vol. 96, pp. 22–31, 2018.

[19] T. Iwaki, Y. Wu, J. Wu, H. Sandberg, and K. H. Johansson, "Wireless sensor network scheduling for remote estimation under energy constraints," in *Proc. IEEE 56th Annu. Conf. Decis. Control*, 2017, pp. 3362–3367.

[20] T. Farjam, F. Fardno, and T. Charalambous, "Power allocation of sensor transmission for remote estimation over an unknown Gilbert-Elliott channel," in *Proc. Eur. Control Conf.*, 2020, pp. 1461–1467.

[21] M. M. Vasconcelos and U. Mitra, "Observation-driven scheduling for remote estimation of two Gaussian random variables," *IEEE Control Netw. Syst.*, vol. 7, no. 1, pp. 232–244, Mar. 2020.

[22] R. Koenker, *Quantile Regression, Ser. Econometric Society Monographs.* Cambridge, U.K.: Cambridge Univ. Press, 2005.

[23] J. Lee, C. Tepedelenlioglu, A. Spanias, and G. Muniraju, "Distributed quantiles estimation of sensor network measurements," *Int. J. Smart Secur. Technol.*, vol. 2, no. 7, no. 2, pp. 38–61, 2020.

[24] H. Wang and C. Li, "Distributed quantile regression over sensor networks," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 4, no. 2, pp. 338–348, Jun. 2018.

[25] L. Xiao and S. Boyd, "Fast linear iterations for distributed averaging," *Syst. Control Lett.*, vol. 53, no. 1, pp. 65–78, 2004.

[26] X. Zhang, M. M. Vasconcelos, W. Cui, and U. Mitra, "An optimal symmetric threshold strategy for remote estimation over the collision channel," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2020, pp. 9195–9199.

[27] S. Sanjari, N. Saldi, and S. Yüksel, "Optimality of independently randomized symmetric policies for exchangeable stochastic teams with infinitely many decision makers," 2020, *arXiv:2008.11570*.

[28] A. Agrawal and S. Boyd, "Disciplined quasiconvex programming," *Optim. Lett.*, vol. 14, no. 7, pp. 1643–1657, 2020.

[29] B. C. Arnold and N. Balakrishnan, *Relations, Bounds and Approximations for Order Statistics*. New York, NY, USA: Springer-Verlag, 1989, vol. 53.

[30] A. Nedić, A. Olshevsky, and M. G. Rabbat, "Network topology and communication-computation tradeoffs in decentralized optimization," *Proc. IEEE*, vol. 106, no. 5, pp. 953–976, May 2018.

[31] M. Penrose, *Random Geometric Graphs*. London, U.K.: Oxford Univ. Press, 2003.

[32] A. Shapiro, D. Dentcheva, and A. Ruszczyński, *Lectures on stochastic programming: Modeling and Theory*. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 2014.

[33] J. Tsitsiklis, D. Bertsekas, and M. Athans, "Distributed asynchronous deterministic and stochastic gradient optimization algorithms," *IEEE Trans. Autom. Control*, vol. 31, no. 9, pp. 803–812, Sep. 1986.

[34] 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.*, 2005, pp. 63–70.

[35] M. Mitzenmacher and E. Upfal, *Probability and Computing: Randomization and Probabilistic Techniques in Algorithms and Data Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 2017.

Xu Zhang (Member, IEEE) was born in Hebei, China, in 1992. He received the B.S. degree in electronic engineering from Beijing Institute of Technology, Beijing, China, in 2015. He is currently working toward the Ph.D. degree in the school of Information and Electronics, Beijing Institute of Technology, Beijing, China.

His research interests include compressed sensing, array signal processing, and distributed remote estimation.

Marcos M. Vasconcelos (Member, IEEE) received the Ph.D. degree in electrical engineering from the University of Maryland, College Park, MD, USA, in 2016.

From 2016 to 2020, he was a Postdoctoral Research Associate with the Ming Hsieh Department of Electrical Engineering, University of Southern California, Los Angeles, CA, USA. He is currently a Research Assistant Professor with the Commonwealth Cyber Initiative and the Bradley Department of Electrical and Computer Engineering at Virginia Tech. His research interests include networked control and estimation, communication networks, game theory, multiagent optimization, distributed machine learning, and systems biology.

Wei Cui received the B.S. degree in physics and Ph.D. degree in electronics engineering from Beijing Institute of Technology, Beijing, China, in 1998 and 2003, respectively.

From 2003 to 2005, he was a Postdoctoral Researcher with the School of Electronic and Information Engineering, Beijing Jiaotong University, Beijing, China. Since then, he has been with the Beijing Institute of Technology, where he is currently a Professor with the School of Information and Electronics. His research interests include adaptive signal processing, array signal processing, sparse signal processing, and their various applications.

Urbashi Mitra (Fellow, IEEE) received the B.S. and M.S. degrees from the University of California at Berkeley, Berkeley, CA, USA, and the Ph.D. degree from Princeton University, Princeton, NJ, USA, all in electrical engineering.

She is currently the Gordon S. Marshall Professor in engineering with the University of Southern California, Los Angeles, CA, USA with appointments in electrical engineering and computer science. Her research interests include wireless communications, communication and sensor networks, biological communication systems, detection and estimation and the interface of communication, as well as sensing and control.

Dr. Mitra is the inaugural Editor-in-Chief for the IEEE TRANSACTIONS ON MOLECULAR, BIOLOGICAL AND MULTI-SCALE COMMUNICATIONS. She was a member of the IEEE Information Theory Society's Board of Governors (2002–2007, 2012–2017), the IEEE Signal Processing Society Technical Committee on Signal Processing for Communications and Networks (2012–2017), the IEEE Signal Processing Society's Awards Board (2017–2018), and the Vice Chair of the IEEE Communications Society, Communication Theory Working Group (2017–2018). She is the recipient of the 2017 IEEE Communications Society Women in Communications Engineering Technical Achievement Award, a 2016 U.K. Royal Academy of Engineering Distinguished Visiting Professorship, a 2016 U.S. Fulbright Scholar Award, a 2016–2017 U.K. Leverhulme Trust Visiting Professorship, 2015–2016 IEEE Communications Society Distinguished Lecturer, 2012 Globecom Signal Processing for Communications Symposium Best Paper Award, 2012 US National Academy of Engineering Lillian Gilbreth Lectureship, the 2009 DCOSS Applications & Systems Best Paper Award, Texas Instruments Visiting Professor (Fall 2002, Rice University), 2001 Okawa Foundation Award, 2000 OSU College of Engineering Lumley Award for Research, 1997 OSU College of Engineering MacQuigg Award for Teaching, and a 1996 National Science Foundation CAREER Award.