On Universal Distributed Estimation of Noisy Fields with One–bit Sensors*

Ye Wang, Nan Ma, Manqi Zhao, Prakash Ishwar, and Venkatesh Saligrama[†]

Abstract— This paper formulates and studies a general distributed field reconstruction problem using a dense network of noisy one–bit randomized scalar quantizers in the presence of additive observation noise of unknown distribution. A constructive quantization, coding, and field reconstruction scheme is developed and an upper–bound to the associated mean squared error (MSE) at any point and any snapshot is derived in terms of the local spatio–temporal smoothness properties of the underlying field. It is shown that when the noise, sensor placement pattern, and the sensor schedule satisfy certain minimal technical requirements, it is possible to drive the MSE to zero with increasing sensor density at points of field continuity while ensuring that the per–sensor bitrate and sensing–related network overhead rate simultaneously go to zero. The proposed scheme achieves the order–optimal MSE versus sensor density scaling behavior for the class of spatially constant spatio–temporal fields.

I. INTRODUCTION AND MOTIVATION

We study the problem of reconstructing a temporal sequence of unknown spatial data fields in a bounded geographical region of interest at a data fusion center from finite bit–rate messages generated by a dense noncooperative network of noisy low–resolution sensors (at known locations) that are statistically identical (exchangeable) with respect to the sensing operation. The interchangeability assumption reflects the property of an unsorted collection of inexpensive mass–produced sensors that behave in a statistically identical fashion. We view each data field as an unknown deterministic function over the geographical space of interest and make only the minimal assumption that they have a known bounded maximum dynamic range. The sensor observations are corrupted by bounded, zero–mean additive noise which is independent across sensors with arbitrary dependencies across field snapshots and has an arbitrary but unknown distribution but a known maximum dynamic range. The sensors are equipped with binary analog–to–digital converters (ADCs) (comparators) with random thresholds that are independent across sensors with arbitrary dependencies across snapshots and are uniformly distributed over a known dynamic range. These modeling assumptions partially account for certain real–world scenarios that include (i) the unavailability of good initial statistical models for data fields in yet to be well studied natural phenomena,

(ii) unknown additive sensing/observation noise sources, (iii) additive model perturbation errors, (iv) substantial variation of preset comparator thresholds accompanying the mass–manufacture of low–precision sensors, (v) significant temperature fluctuations across snapshots affecting hardware characteristics, and (vi) the use of intentional dither signals for randomized scalar quantization.

In this work, a data fusion center is any point of data aggregation and/or processing in the sensor network and can be real or virtual. For instance, sensors can be dynamically organized into clusters with different sensors assuming the role of a fusion center at different times [1]. This work does not explicitly address physical–layer network data transport issues. In particular, we do not consider joint source–channel coding strategies. Instead, to conform with the existing base of digital communication architectures, the effective communication links are abstracted into a network of reliable but finite–rate bit–pipes from each sensor to the data fusion center. In practice, sensor data can be moved to the fusion center through a variety of physical–layer transport mechanisms, example, a stationary base–station with directional antenna arrays, a mobile data collector, and passive sensor querying mechanisms involving, for instance, laser–beams and modulating mirrors [2]. We acknowledge that separating the distributed field reconstruction problem into efficient data acquisition and efficient data transport parts through a finite–rate reliable bit–pipe abstraction may be suboptimal [3, p. 449], [4], [5]. For instance, in some scenarios multihop communication is not needed and the characteristics of the field, the communication channel, and the distortion–metric are "matched" to one another. In such a scenario, uncoded "analog" transmission can offer huge performance gains if the synchronization of sensor transmissions can be orchestrated at the physical layer to achieve beamforming gains and the network channel state information is available to the transmitting sensors [4]. For the joint source–channel aspects of this and related problems, see [6]–[9]. For networking issues such as sensor scheduling, quality of service, and energy efficiency also see [10] and references therein.

Building upon prior results in [11], [12], and [13], we develop a simple coding and field reconstruction scheme based on noisy one–bit samples of noisy observations and characterize the associated scaling behavior of the MSE of field reconstruction with sensor density in terms of the local and global moduli of continuity of the underlying sequence of fields for fixed, positive, and equal sensor coding rates (bits per sensor per snapshot). These achievable results reveal that for bounded, zero–mean additive observation noise of

† Y. Wang, N. Ma, M. Zhao, P. Ishwar, V. Saligrama are with the Department of Electrical and Computer Engineering, Boston University, Boston, MA {yw,nanma,mqzhao,pi,srv}@bu.edu

unknown distribution, the MSE at every point of continuity of every field can be made to go to zero as sensor density increases while simultaneously sending the per–sensor bitrate and any sensing–related network rate overheads (e.g., sensor addresses) to zero. The rate of decay of field reconstruction MSE at a given location is related to the local modulus of continuity of the field at the given location and time. This is possible if the sensor placement and sampling schedule satisfy a certain uniformity property and if the field estimate at any given spatial location is formed using the observations from increasingly many sensors that are located within a vanishingly smaller neighborhood of the location. Specializing these results to the case of spatially constant fields yields an achievable MSE decay rate of $O(1/L)$ where L is the sensor network size.[1] A Cramér–Rao lower–bound on the MSE for parameter estimation establishes that the $O(1/L)$ MSE scaling behavior is order–optimal. Since in our problem formulation, the per–sensor bitrate is held fixed and equal across sensors, in a scaling sense, the MSE decreases inversely with the total network rate.

These results are consistent with the information–theoretic, total network rate versus MSE scaling results for the CEO problem which was first introduced in [14] and thereafter studied extensively in the information theory literature (see [15], [16] and references therein). However, it should be noted that information–theoretic rate–distortion studies of this and related distributed field reconstruction (multiterminal source coding) problems typically consider stationary ergodic stochastic fields with complete knowledge of the field and observation–noise statistics, block vector–quantization (VQ) and binning operations, and time–averaged (as opposed to worst–case) expected distortion criteria. In VQ, sensors are allowed to collect long blocks of real–valued field samples (of infinite resolution) from multiple field snapshots before a discrete, finite bit–rate VQ operation. The fields are often assumed to be spatially constant and independent and identically distributed (iid) across time (frequently Gaussian) and the observation noise is often assumed to be additive with a known distribution (frequently Gaussian) as in the CEO problem. It should also be noted that the MSE scaling results for the CEO problem in [15] are with respect to the total network rate where the number of agents (or sensors) has already been sent to infinity while maintaining the total network rate a finite value. Recent information–theoretic results for stationary fields under zero observation noise have been developed in [17], [18].

Previous estimation–theoretic studies of one–bit distributed field reconstruction have focused on reconstructing a single field snapshot and have either (i) assumed zero observation noise [11], [12], or (ii) assumed a spatially constant field (equivalent to scalar parameter estimation) with a one–bit communication as opposed to a one–bit sensing constraint. Our proposed system integrates the desirable field sensing and reconstruction properties of these apparently different one–bit field estimation schemes and establishes the statistical and performance equivalence of these approaches. An important hardware implication of this paper is that noisy op–amps (noisy one–bit ADCs) are adequate for high–resolution distributed field reconstruction. This should be be contrasted with the framework in [13] which implicitly requires sensors to have the ability to quantize their observations to an arbitrarily high bit resolution. A side contribution of this paper is the holistic treatment of the general distributed field–reconstruction problem in terms of (i) the field characteristics, (ii) sensor placement characteristics, (iii) sensor observation, quantization, and coding constraints with associated sensing hardware implications, (iv) transmission and sensing–related network overhead rates, and (v) reconstruction and performance criteria. We have attempted to explicitly indicate and keep track of what information is known/available/used where and what is not.

The randomized scalar quantization model for the sensor comparators not only captures poor sensing capabilities but is also an enabling factor in the high–fidelity reconstruction of signals from quantized noisy observations. As shown in [19] in an information–theoretic setting and alluded to in [12], using *identical* deterministic scalar–quantization (SQ) in all sensors results in the MSE performance being fundamentally limited by the precision of SQ, irrespective of increasing sensor density, even in the absence of sensor observation noise.[2] However, our results further clarify that having "diversity" in the scalar quantizers, achieved, for example, through the means of an intentional random dither, noisy threshold, or other mechanisms, can achieve MSE performance that tends to zero as the density of sensors goes to infinity. Randomization enables high–precision signal reconstruction because zero–mean positive and negative fluctuations around a signal value can be reliably "averaged out" when there are enough independent noisy observations of the signal value. This observation is also corroborated by the findings reported in the following related studies [11]–[13], [20], [21].

The rest of this paper is organized as follows. The main problem description with all technical modeling assumptions is presented in Section II. The main technical results of this paper are summarized and discussed in Section III. Section IV describes the proposed constructive distributed coding and field reconstruction scheme and the analysis of MSE performance. In Section V, we discuss the close connections between the work in [12], [13], and the present work, and establish the fundamental statistical and performance equivalence of the core techniques in these studies. Finally, in Section VI we present concluding remarks and comment on ongoing work and future research directions.

---

[1]Landau's asymptotic notation: $f(L) = O(g(L)) \Leftrightarrow \limsup_{L\to\infty} |f(L)/g(L)| < \infty$; $f(L) = \Omega(g(L)) \Leftrightarrow g(L) = O(f(L))$; $f(L) = \Theta(g(L)) \Leftrightarrow f(L) = O(g(L))$ and $g(L) = O(f(L))$.

[2]However, VQ with binning does not suffer from this limitation as shown in [17], [18].
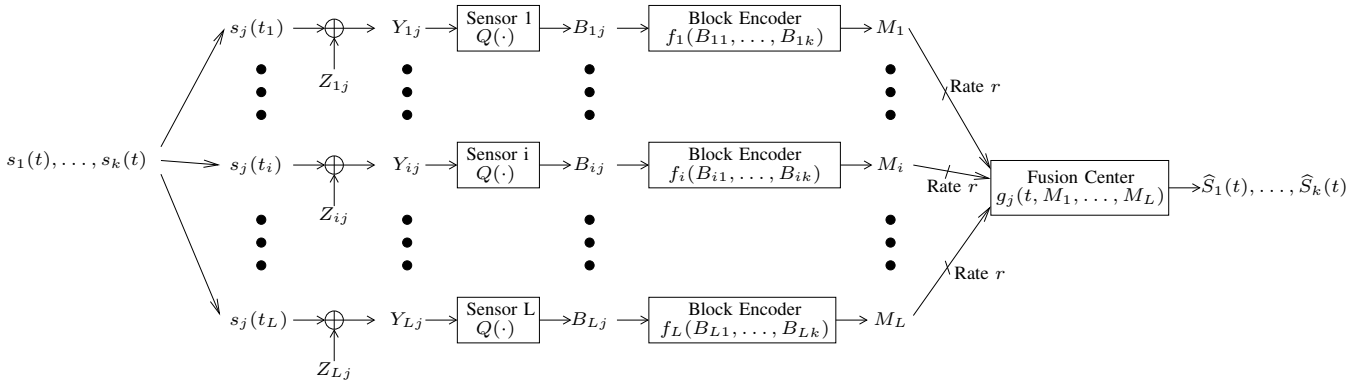
Fig. 1. **Block diagram of a distributed field reconstruction sensor–network using randomized 1–bit SQ with block–coding.** *Sensor* i *quantizes its noisy observations,* $Y_{i1}, \ldots, Y_{ik}$, *to the binary values* $B_{i1}, \ldots, B_{ik}$. *The sensor then generates the message* $M_i \in \{1, \ldots, 2^{kr}\}$ *based on these quantized values. These messages* $\{M_i\}$ *are then relayed to the fusion center where the field estimates* $\widehat{S}_j(t)$ *are produced.*

II. DISTRIBUTED FIELD RECONSTRUCTION SENSOR–NETWORK (DFRS) SETUP

A. Field Model

We consider a sequence of spatial fields occurring at k discrete–time snapshots.[3] Each field is modeled as a continuous[4] bounded function,

$$s_j : G \to \mathbb{R} : \forall t \in G, \ \forall j \in \{1, \ldots, k\}, \ |s_j(t)| \le a < +\infty,$$

where $G \subseteq \mathbb{R}^d$ is a known geographical region of interest in d–dimensional real space and a is a known bound on the maximum field dynamic range. Although the results of this paper hold for any G which is bounded and is the closure of its nonempty interior, for simplicity and clarity of exposition, we will assume $G = [0, 1]^d$, the d–dimensional unit–hypercube, in the sequel. Distances are measured with respect to a norm $\| \cdot \|$. Since the fields are continuous functions on the compact set G, they are in fact uniformly continuous on G.

Results on the fidelity of the field reconstruction will be described in terms of the local and global moduli of continuity associated with the field:

Definition 2.1: (Local modulus of continuity) The local modulus of continuity $\omega_j : [0, \infty) \times G \to [0, \infty)$ of the function $s_j(t)$ at the point $t \in G$ is defined as

$$\omega_j(\delta, t) \triangleq \sup_{\{t' \in G : \|t - t'\| \le \delta\}} |s_j(t) - s_j(t')|,$$

where $\| \cdot \|$ is the underlying norm[5] for $\mathbb{R}^d$. Note that for all

[3]If the spatio–temporal field is temporally bandlimited then the field values at intermediate time points can be interpolated from the estimates at discrete time snapshots if the temporal sampling rate is (strictly) higher than the temporal Nyquist rate of the field. The associated MSE will be no larger than the maximum MSE of the estimates across the discrete–time snapshots times a proportionality constant.

[4]More generally, our results can be extended to arbitrary, measurable functions. For such functions the pointwise MSE bounds given in Section III-A still hold. The estimates at the points of continuity will have MSE tending to 0 as the network size scales. However, the points of discontinuity may have a finite, but non–zero MSE floor.

[5]For asymptotic results in which $\delta \longrightarrow 0$, any norm on $\mathbb{R}^d$ would suffice since all norms on any finite–dimensional Banach space are equivalent [22, Theorem 23.6, p. 177]. However, for definiteness, it may be helpful to think of the Euclidean 2–norm.

$t \in G$, $\omega_j(\delta, t)$ is a nondecreasing function of δ and that it $\longrightarrow 0$ as $\delta \longrightarrow 0$ since $s_j(t)$ is continuous at each point t in G.

Definition 2.2: (Global modulus of continuity) The global modulus of continuity $\widetilde{\omega}_j : [0, \infty) \to [0, \infty)$ of the function $s_j(t)$ is defined as

$$\widetilde{\omega}_j(\delta) \triangleq \sup_{t \in G} \omega_j(\delta, t).$$

Again note that $\widetilde{\omega}_j(\delta)$ is a nondecreasing function of δ and that it $\longrightarrow 0$ as $\delta \longrightarrow 0$ since $s_j(t)$ is uniformly continuous over G. Note that bounded, continuous functions on a compact set are uniformly continuous.

The global and local moduli of continuity of a spatial field respectively reflect the degree of global and local spatial smoothness of the field with smaller values, for a fixed value of δ, corresponding to greater smoothness. For example, for a spatially constant field, that is, for all $t \in G$, $s_j(t) = s_j$ (a constant), we have $\widetilde{\omega}_j(\delta) = 0$ for all $\delta \ge 0$. For $d = 1$ and fields with a uniformly bounded derivative, that is, for all $t \in G$, $\sup_{t \in G} |d(s_j(t))/dt| = \Delta < +\infty$, $\widetilde{\omega}_j(\delta) = \Delta \cdot \delta$. More generally, for a Lipschitz–γ spatial function $s_j(t)$, we have $\widetilde{\omega}_j(\delta) \propto \delta^\gamma$. Closed–form analytical expressions of moduli of continuity may not be available for arbitrary fields but bounds often are. Sometimes bounds that are tight in the limit as $\delta \longrightarrow 0$ are also available. From Definitions 2.1, 2.2, and the boundedness of the field dynamic range, it also follows that for all $\delta \ge 0$, for all $t \in G$, and for all $i \in \{1, \ldots, k\}$, we have

$$0 \le \omega_j(\delta, t) \le \widetilde{\omega}_j(\delta) \le 2a < +\infty.$$

B. Sensor Placement

We assume that we have a dense, noncooperative network of L sensors at distinct spatial locations over G, such that they satisfy the "nearly uniform with partition condition" with partitioning parameter m as described in Definition 2.3. The actual deployment can be otherwise arbitrary as long as this condition is met. We use $T \triangleq \{t_1, \ldots, t_L\}$ to denote the collection of sensors and t_i to denote the location of sensor i.

Definition 2.3: (Nearly uniform with partitioning by factor m) A set T of L distinct spatial sampling locations in G is said to be nearly uniformly partitionable by factor m (integer $m \geq 1$), if T can be the partitioned into m^d disjoint subsets $T_1, \ldots, T_{m^d}$ each of cardinality $\left(\frac{L}{m^d}\right)$, where for each partition and $\forall N \in \mathbb{Z} : 1 \leq N \leq \left(\frac{\sqrt[d]{L}}{m}\right)$, there are at least N^d distinct sensors of that subset located within a norm–distance $\left(\frac{Nm}{\sqrt[d]{L}-1}\right)$ of any $t \in G$.

As an example, for the $\|\cdot\|_\infty$ norm, and $G = [0,1]^d$, an exactly uniform rectangular grid of sensors on G can be properly partitioned to be nearly uniform with partitioning. More precisely, for any desired m, chose an auxiliary value l that is a multiple of m and set $L = l^d$. Distribute the L sensors in uniform grid over $G = [0,1]^d$ with intersensor spacing of $\left(\frac{1}{l-1}\right)$ given by the sensor location set

$$
T = \left\{ \left(\frac{x_1}{l-1}, \ldots, \frac{x_d}{l-1}\right) : \right.
$$
$$
\left. x_1, \ldots, x_d \in \{0, 1, \ldots, l-1\} \right\}.
$$

The first partition of this set is given by taking the subset of sensors forming a uniform grid with intersensor spacing of $\left(\frac{m}{l-1}\right)$ given by

$$
T_1 = \left\{ \left(\frac{mx_1}{l-1}, \ldots, \frac{mx_d}{l-1}\right) : \right.
$$
$$
\left. x_1, \ldots, x_d \in \left\{0, 1, \ldots, \left(\frac{l}{m}-1\right)\right\}\right\}.
$$

The remaining $(m^d - 1)$ partitions are given by shifting this pattern over the space G to include the remaining sensors in the other partitions. Figure 2 illustrates this sensor deployment and the nature of the partitioning for the case $d = 2$ with $L = 36$ sensors and partitioning factor of $m = 2$.
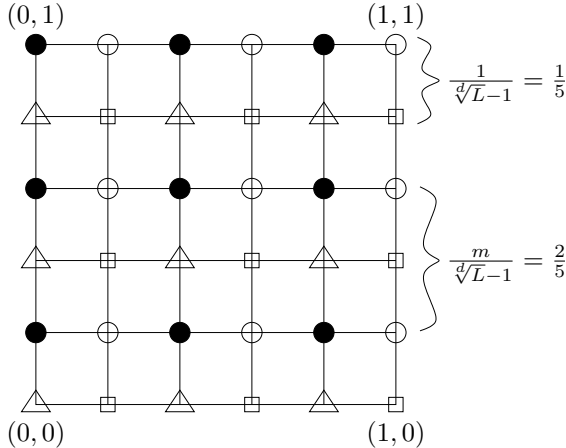

Fig. 2. **Example deployment** *for case* $d = 2$ *with* $L = 36$ *sensors satisfying the nearly uniformly partitionable condition for factor* $m = 2$. *The four different shapes represent the* $m^d = 4$ *partitioned subsets; the black circles being the first subset and the others produced as shift versions of it.*

In this work, we do not address the optimality of sensor placement. One formal deterministic approach to address

such issues is the theory of low–discrepancy sampling sequences that has been used in pseudo–random number generation, quasi–Monte Carlo simulation, and computer graphics [23], [24]. In practical situations, it may not be possible to precisely control the placement of the sensors unless the sensors have mobility. See [25] for a study of a simple probabilistic sensor placement model. We assume that sensors have been scattered in a manner that guarantees the nearly uniform placement property, at least for all L sufficiently large, and that the fusion center has knowledge of the sensor locations. The results of Sections III-A and IV will show how the design parameter N can be picked in relation to L so that the field reconstruction MSE goes to zero as L increases to infinity.

C. Sensor Observation and Coding Models

1) Sensor Observation Noise: The sensor observations are corrupted by bounded, zero–mean additive noise which is independent across sensors, but can be arbitrarily correlated across field snapshots[6]. The noise has an unknown joint cumulative distribution function (cdf) $F_{\mathbf{Z}}(\mathbf{z})$ that can be arbitrary within the zero–mean, boundedness and independence constraints already stated. The maximum dynamic range of the noise $b \in [0, +\infty)$ is known. For all $i \in \{1, \ldots, L\}$ and all $j \in \{1, \ldots, k\}$,

$$
Y_{ij} = s_j(t_i) + Z_{ij},
$$

where $\{Z_{ij}\} \sim$ cdf $F_{\mathbf{Z}}(\mathbf{z})$. We use $\mathcal{F}$ to denote the set of all joint cdfs corresponding to joint probability measures that are factorizable (independent across sensors) into L joint zero–mean probability measures (corresponding to each distinct vector of Z_{ij}'s affecting the same sensor) on $\mathbb{R}^k$ with support $[-b, +b]^k$. Note that $|Y_{ij}| \leq |s_j(t_i)| + |Z_{ij}| \leq c \triangleq (a + b)$.

2) Randomized 1–bit SQ with Block Coding: Due to severe precision and reliability limitations, each sensor $i \in \{1, \ldots, L\}$, has access to only to a vector of unreliable quantized binary samples $\mathbf{B}_i \triangleq (B_{i1}, \ldots, B_{ik})$ for processing and coding and not direct access to the real–valued noisy observations $Y_{i1}, \ldots, Y_{ik}$. The quantized binary sample B_{ij} is generated from the corresponding noisy observation Y_{ij} through a randomized mapping $Q_{ij} : [-c, c] \to \{0, 1\}$: for each $i \in \{1, \ldots, L\}$ and each $j \in \{1, \ldots, k\}$,

$$
B_{ij} = Q_{ij}(Y_{ij}),
$$

where we assume that the mappings Q_{ij} are independent across sensors i, but can be arbitrarily correlated across snapshots j. We denote the conditional marginal statistics of the quantized samples by $p_{B_{ij}|Y_{ij}}(y) \triangleq \mathbb{P}(B_{ij} = 1|Y_{ij} = y)$. We are specifically interested in cases where $p_{B_{ij}|Y_{ij}}(y)$ is an affine function of y since it allows estimates of the fields to be made from the B_{ij}'s without knowledge of the noise distribution (see Section A). Specifically we consider the conditional distribution

$$
p_{B_{ij}|Y_{ij}}(y) = \mathbb{P}(y > X_{ij}) = \left(\frac{y+c}{2c}\right).
$$

[6]The measurement snapshot timers of all the participating sensors are assumed to be synchronized.

This conditional distribution can be achieved by a quantization method which is based on comparing the noisy observation with a random uniformly distributed threshold given by

$$B_{ij} = Q_{ij}^{Th}(Y_{ij}) \triangleq \mathbf{1}(Y_{ij} > X_{ij}), \qquad (2.1)$$

where the X_{ij}'s are random (independent across i, but arbitrarily correlated across j) thresholds distributed uniformly over $[-c, c]$ and $\mathbf{1}(\cdot)$ denotes the indicator function:

$$\mathbf{1}(Y_{ij} > X_{ij}) = \begin{cases} 1 & \text{if } Y_{ij} > X_{ij}, \\ 0 & \text{otherwise.} \end{cases}$$

This uniform random–threshold 1–bit SQ model partially accounts for some practical scenarios that include (i) comparators with a floating threshold voltage, (ii) substantial variation of preset comparator thresholds accompanying the mass–manufacture of low–precision sensors, (iii) significant environmental fluctuations that affect the precision of the comparator hardware, or generally (iv) unreliable comparators with considerable sensing noise and jitter. An alternative justification is that the random thresholds are intentionally inserted as a random dither. Scenario (i) can be accommodated by independence across snapshots, scenario (ii) can be accommodated by complete correlation (fixed) across snapshots, and scenarios (iii) and (iv) can be accommodated by arbitrary correlation across snapshots.
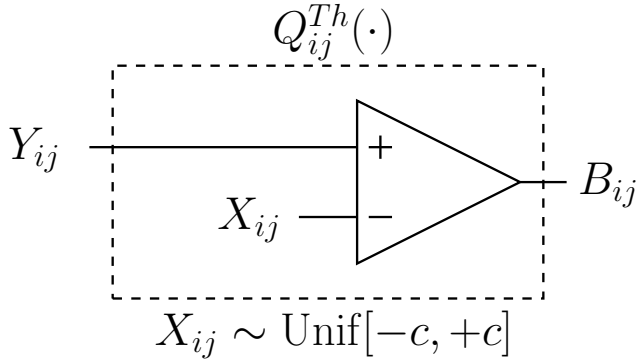

Fig. 3. **Quantizer Hardware Example:** *The given model of for the $Q_{ij}^{Th}(\cdot)$ function from (2.1) can be implemented by a comparator with a uniformly distributed threshold. These thresholds are independent across sensors, but arbitrarily correlated across snapshots, allowing many scenarios to be accommodated.*

Each sensor i utilizes a block encoder to "compress" its vector of k quantized samples $\mathbf{B}_i$ to a message $M_i \in \{1, 2, \ldots, 2^{kr}\}$ before transmitting to the fusion center. The block encoder and message for sensor i are given by

$$f_i : \{0,1\}^k \to \{1, 2, \ldots, 2^{kr}\}, \quad M_i = f_i(B_{i1}, \ldots, B_{ik}),$$

where r is the coding rate in bits per sensor per snapshot. For $r \geq 1$ compression is trivial since $\mathbf{B}_i$ can assume no more than 2^k distinct values which can be indexed using k bits.

D. Transmission and Field Reconstruction

Our problem setup abstracts the underlying transmission network of sensors as a network of bit pipes that are capable of reliably delivering these L messages (the payloads) and the network addresses of the message origination nodes (the headers) to the fusion center. This enables the fusion center to correctly associate the spatial location information with the corresponding sensor field–measurement information for reliable field reconstruction. This can be achieved by having each sensor append a $\log(L)$ bits long label to its message. This results in a total sensor–location rate–overhead of $r_{ohd} = (L/k)\log(L)$ bits per snapshot on the network information transport costs. This overhead will be negligible if $k \gg L \log(L)$. If the underlying sequence of fields are spatially constant, then, the sensor location information is clearly not needed at the fusion center (see Section IV).

The fusion center forms the estimates of the k fields based on the sensor messages using the reconstruction functions

$$g_j : G \times \{1, 2, \ldots, 2^{kr}\}^L \to [-a, a], \ \forall j \in \{1, \ldots, k\}.$$

The estimate of field j at point $t \in G$ is given by

$$\widehat{S}_j(t) = g_j(t, M_1, \ldots, M_L).$$

Definition 2.4: (Rate–r DFRS) A rate–r DFRS based on randomized 1–bit SQ with block coding is defined by the set of rate–r encoder functions $\{f_i(\cdot)\}_{i=1}^L$ and the set of reconstruction functions $\{g_j(\cdot)\}_{j=1}^k$.

Figure 1 depicts a rate–r DFRS using randomized 1–bit SQ with block coding.

1) Performance Criterion:

Definition 2.5: (Pointwise MSE) The pointwise MSE of the estimate of field j at location $t \in G$, for a given rate–r DFRS and a specific noise joint cdf $F_{\mathbf{Z}}(\mathbf{z}) \in \mathcal{F}$, is given by

$$D_j(t; F_{\mathbf{Z}}) = \mathbb{E}[(\widehat{S}_j(t) - s_j(t))^2].$$

Since we are interested in schemes that will work for *any* noise cdf in $\mathcal{F}$, we consider the maximization of $D_j(t; F_{\mathbf{Z}})$ over all possible $F_{\mathbf{Z}} \in \mathcal{F}$. We also consider the maximization over all fields and all locations in G since we want to reconstruct every point of every field with high fidelity.

Definition 2.6: (Worst–case MSE) The worst–case MSE D is given by

$$D = \max_{j \in \{1, \ldots, k\}} \sup_{t \in G} \sup_{F_{\mathbf{Z}} \in \mathcal{F}} D_j(t; F_{\mathbf{Z}}).$$

Our objective is to understand the scaling behavior of MSE with L, k, and r. The next section summarizes our partial results in this direction.

III. MAIN RESULTS

A. Achievable MSE Performance

Our first result gives an upper bound on the MSE achievable through a constructive DFRS based on randomized 1–bit SQ with block coding for all rates $r = 1/(m^d)$, where $m \geq 1$ is an integer. The actual scheme will be described in Section IV. The MSE analysis appears within the proof of the theorem detailed in Appendix A. This achievable MSE upper bound can be made to decrease to zero as sensor–density

goes to infinity without knowledge of the local or global smoothness properties of the sequence of fields. Furthermore, this scheme is universal in the sense that it does not assume knowledge of $F_{\mathbf{Z}}(\mathbf{z})$ beyond membership to $\mathcal{F}$.

Theorem 3.1: (Achievable MSE performance: Randomized 1–bit SQ and $r = 1/(m^d)$) For all rates $r = 1/(m^d)$, where $m \geq 1$ is an integer, there exists a rate–r DFRS based on randomized 1–bit SQ with block coding (described in Section IV) such that for all $t \in G$, $j \in \{1, \ldots, k\}$, $F_{\mathbf{Z}}(\mathbf{z}) \in \mathcal{F}$, and positive integer $N \leq \sqrt[d]{L}$,

$$
\begin{aligned}
D_j(t; F_{\mathbf{Z}}) &\leq \omega_j^2 \left(\left(\frac{Nm}{\sqrt[d]{L} - 1} \right), t \right) + \left(\frac{c^2}{N^d} \right) \\
&\leq \widetilde{\omega}_j^2 \left(\frac{Nm}{\sqrt[d]{L} - 1} \right) + \left(\frac{c^2}{N^d} \right),
\end{aligned}
$$

where N is a tunable design parameter.

Note that Theorem 3.1 holds for all arbitrary fields. The modulus of continuity terms in the local (first) and global (second) upper bounds of Theorem 3.1 are due to the bias and the $\frac{c^2}{N^d}$ term is due to the variance of the field estimates (see (4.5) in Section IV). From Theorem 3.1 and the properties of moduli of continuity (see Section II-A), it follows that for the coding and reconstruction scheme of Section IV, as $L \longrightarrow \infty$, the estimate $\widehat{S}_j(t)$ can converge, in a mean square sense, to $s_j(t)$ for all $t \in G$, provided that

$$
\text{(i) } N^d \longrightarrow \infty, \text{ and (ii) } \left(\frac{Nm}{\sqrt[d]{L} - 1} \right) \longrightarrow 0. \qquad (3.2)
$$

When both conditions in (3.2) hold, the convergence of $\widehat{S}_j(t)$ to $s_j(t)$ is uniform because $s_j(t)$ is uniformly continuous on G. It also follows that the worst–case MSE scaling behavior (see Definition 2.6)

$$
D \leq \max_{j \in \{1, \ldots, k\}} \left\{ \widetilde{\omega}_j^2 \left(\frac{Nm}{\sqrt[d]{L} - 1} \right) + \left(\frac{c^2}{N^d} \right) \right\} \qquad (3.3)
$$

is achievable and that $D \longrightarrow 0$ as N and L scale as in (3.2). **Implications:** These results allow us to make per sensor per snapshot bit rate r, worst–case MSE D, and sensor message ID overheads (given by $(L/k) \log(L)$ bits) simultaneously smaller than any desired $\epsilon > 0$. First, we can choose a sufficiently large time sharing factor m such that the rate $r = (1/m^d) < \epsilon$. Then we can choose a sufficiently large number of sensors L and corresponding system parameter N such that the bound on D given by (3.3) is made less than ϵ. Finally, we can look at a sufficiently large number of snapshots k such that network message overheads $(L/k) \log(L) < \epsilon$.

In the constructive coding and field reconstruction scheme of Section IV, the estimate for each point $t \in G$ is formed using messages from N^d sensors that lie within a radius of $\frac{Nm}{\sqrt[d]{L} - 1}$ from t. The nearly uniform with partitioning sensor placement condition (see Definition 2.3) ensures that this is possible. Therefore, if N and L scale according to (3.2), then each point estimate will be formed using *increasingly many* messages ($N^d \longrightarrow \infty$) from sensors located within a *vanishingly small* radius, $\frac{Nm}{\sqrt[d]{L} - 1} \longrightarrow 0$, of the point being

estimated. Since the variance term $\frac{c^2}{N^d}$ in the upper bound of Theorem 3.1 can decrease no faster than $O(1/L)$, the decay of the global MSE upper bound, in the proposed constructive scheme, can be no faster than $O(1/L)$ as the sensor density grows. However, N needs to grow at a slower rate than $\sqrt[d]{L}$ for $\frac{Nm}{\sqrt[d]{L} - 1}$ to decay to 0. When $\widetilde{\omega}_j(\cdot)$ is not identically zero, the rate of growth of N with L that minimizes the global MSE upper bound of Theorem 3.1 is determined by the following condition

$$
\widetilde{\omega}_j^2 \left(\frac{Nm}{\sqrt[d]{L} - 1} \right) = \Theta \left(\frac{1}{N^d} \right).
$$

For certain classes of signals for which the global modulus of continuity has a closed form, the optimum growth rate can be explicitly determined. For instance, if $d = 1$ and $\widetilde{\omega}_j(\delta) = \Delta \cdot \delta$ (Lipschitz–1 fields), $N_{opt}(L) = \Theta(L^{\frac{2}{3}})$ for which MSE $= O(1/L^{\frac{2}{3}})$.

Corollary 3.1: (Achievable MSE performance: Randomized 1–bit SQ, $r = 1/(m^d)$, and constant fields) If for all $t \in G$ and all $j \in \{1, \ldots, k\}$, we have $s_j(t) = s_j$, or equivalently, for all $\delta \geq 0$ and all $j \in \{1, \ldots, k\}$, $\widetilde{\omega}_j(\delta) = 0$, then the result given by (3.3) reduces to

$$
D \leq \left(\frac{c^2}{L} \right),
$$

where N^d is set to L to minimize the bound. Furthermore, all field–estimates given by (4.5) are unbiased.

Note that here the spatial locations and density of sensors are irrelevant: the MSE behavior is governed purely by the number of sensors L and not their spatial distribution. Also note that for a sequence of spatially constant fields, the estimates for all fields are unbiased. This is because the first term in the upper bound of Theorem 3.1 is an upper bound on the bias of the proposed estimator (4.5) (see the bias analysis result in Appendix A).

B. A Fundamental Lower Bound on MSE

We state a fundamental lower bound on the MSE performance for any rate r DFRS that produces unbiased estimates for the case of spatially constant fields. This lower bound provides a converse result to the achievable result stated in Corollary 3.1. In our achievable MSE results, we give universal upper bounds that hold for noise distributions $F_{\mathbf{Z}} \in \mathcal{F}$. Thus as a converse result, we are interested lower bounding the minimax performance, that is the MSE performance of the optimal estimator used under the worst–case noise. Our bound also applies to general randomized 1–bit SQ functions $Q_{ij}(\cdot)$ including those based on uniform random thresholds $Q_{ij}^{Th}(\cdot)$ (see (2.1)).

Theorem 3.2: (Lower bound on MSE: Unbiased estimators for constant fields) For a sequence of spatially constant fields and any DFRS which produces unbiased field estimates, there exists a joint cdf $F_{\mathbf{Z}} \in \mathcal{F}$ such that for noise distributed according to $F_{\mathbf{Z}}$ the MSE is bound lower bounded by

$$
\mathbb{E}[(\widehat{S}_j - s_j)^2] \geq \left(\frac{C_j}{L} \right), \quad \text{for all } j \in \{1, \ldots, k\},
$$

where C_j is finite, non–zero, and does not depend on L.

Proof: Since $\{s_j\} \rightarrow \{Y_{ij}\} \rightarrow \{B_{ij}\} \rightarrow \{M_i\}$ forms a Markov chain, the estimates based on the sensor messages $\{M_1, \ldots, M_L\}$ cannot have a lower MSE than estimates based on the noisy observations $\{Y_{ij}\}$. Consider a well–behaved, non–trivial, joint cdf $F_{\mathbf{Z}} \in \mathcal{F}$ such that the Z_{ij} are iid and the conditional probabilities of Y_{ij} given the fields satisfy the regularity conditions necessary for the Cramér-Rao bound [26] to be applied. Applying the Cramér-Rao bound yields that MSE of each field estimate based on $\{Y_{ij}\}$ is lower bounded by $\frac{C_j}{L}$ where C_j is finite, non–zero, and does not depend on L. ∎

Theorem 3.2 states that for the case of spatially constant fields, the worst–case MSE $D = \Omega(1/L)$ for any DFRS based on randomized 1–bit SQ with block coding and produces unbiased estimates for each field. Corollary 3.1 states that the scheme of Section IV when applied to the spatially constant fields case produces unbiased estimates of each field while achieving a worst–case MSE $D = O(1/L)$. Thus, for spatially constant fields, the proposed scheme achieves the order optimal performance in terms of the scaling behavior of L, the size of the network.

IV. Proposed Constructive Distributed Coding and Field Reconstruction Scheme

In this section we present the proposed DFRS scheme that was alluded to in Section III. In this scheme, sensors create the quantized binary samples $\{X_{ij}\}$ from its observations $\{Y_{ij}\}$ through comparisons with the random thresholds $\{X_{ij}\}$, as described in Section II-C.2: $B_{ij} = Q_{ij}^{Th}(Y_{ij})$ as in (2.1). In order to achieve compression at rates $r = 1/(m^d)$, a simple "time–sharing" based compression method is used. Each sensor, instead of transmitting all of its k bits (binary quantized observations) $\mathbf{B}_i = (B_{i1}, \ldots, B_{ik})$, transmits only $kr = k/(m^d)$ of its bits and the remaining bits are dropped. The partitioning of the collection sensors into disjoint subsets with the properties described in Definition 2.3 is used in order to properly determine which bits sensors should drop or keep. In the first snapshot, the sensors belonging to the subset T_1 keep their bits and all other sensors drop their bits. In the jth snapshot, the sensors belonging to the subset $T_{(j \bmod m^d)+1}$ keep their bits and all other sensors drop their bits. Therefore by Definition 2.3, for each snapshot and for any integer $N \le \frac{\sqrt[d]{L}}{m}$ there are N^d binary observations of the field made by sensors within a distance of $\frac{Nm}{\sqrt[d]{L}-1}$ any point in G.

For notational simplicity, the reconstruction function $\widehat{S}_j(t) = g_j(t, M_1, \ldots, M_L)$ will be described directly in terms of t and $\mathbf{B}_i$.

The fusion center computes $\widehat{S}_j(t)$ using the binary messages from the N^d sensors that are closest to the spatial location t. Here, closeness is measured with respect to the underlying norm $\|\cdot\|$. Note that $N \le \frac{\sqrt[d]{L}}{m}$ is a tunable parameter and in principle, the fusion center can choose different values of N for different t and j. However, due to the absence of information about the local smoothness properties of the fields at the fusion center, we assume that

the value of N is the same for all t and j. Let $T_{\{j\}}^N(t) \subseteq T$ denote the subset of spatial locations of the N^d sensors that are closest (with ties resolved by some predetermined ordering) to the spatial location t and belonging to the set $T_{(j \bmod m^d)+1}$, the set of sensors that recorded a bit for field j. An alternative approach (used in [12]) is for the fusion center to first compute $\widehat{S}_j(t)$ at the L sensor locations in T and then interpolate the results using a zero–order hold or other higher–order interpolation algorithms such as those based on cubic B–splines. The resulting MSE will be of the same order. We use the former approach because its analysis is more compact. The nearly uniform sensor placement property (see Definition 2.3) ensures that

$$\forall t_l \in T_j^N(t), \|t_l - t\| \le \left(\frac{Nm}{\sqrt[d]{L}-1} \right). \quad (4.4)$$

The estimate for $s_j(t)$ is then given by

$$\begin{aligned}
\widehat{S}_j(t) &= g_j(t, M_1, \ldots, M_L) \\
&\triangleq 2c \left[\frac{1}{N^d} \sum_{\{l:t_l \in T_{\{j\}}^N(t)\}} B_{lj} \right] - c, \quad (4.5)
\end{aligned}$$

which is the sample mean of the N^d binary quantized samples normalized and shifted into the dynamic range of noisy field observations $[-c, c]$. Appendix A proves that the MSE of this constructive coding and reconstruction scheme is upper bounded by the result described in Theorem 3.1.

V. Connections to Prior One-bit Estimation Approaches

This section clarifies the close connection between the methods and results in [12], [13], and the present work. It is shown that the apparently different randomized 1–bit field estimation schemes in these studies are in fact statistically and MSE performance–wise equivalent. The general framework of the present work integrates the desirable field sensing and reconstruction properties and insights of the earlier studies and provides a unified view of the general problem. For this discussion, we will drop the snapshot indices j from our notation since $k = 1$ in the earlier studies.

A. One–Bit Randomized–Dithering

The problem setup considered in [12] is closely related to the randomized 1–bit SQ problem setup studied here. The problem setup in [12] corresponds to setting $d = k = 1$, $G = [0, 1]$, $t_i = i/L$ and $Z_{ij} = 0$ for all i and j, requiring all B_{ij} to belong to $\{-1, +1\}$ instead of $\{0, 1\}$, and requiring that the cdf of the random thresholds X_{ij} satisfy certain technical conditions described in [12, Section II.A]. These technical conditions include the uniform distribution (considered in this paper) as a special case. An important conceptual difference is that in [12] the sensor quantization noise is viewed as a randomized dither signal which is intentionally added to the observations and that the dither cdf is known (it need not be uniform). The reconstruction explicitly exploits the knowledge of the dither statistics.
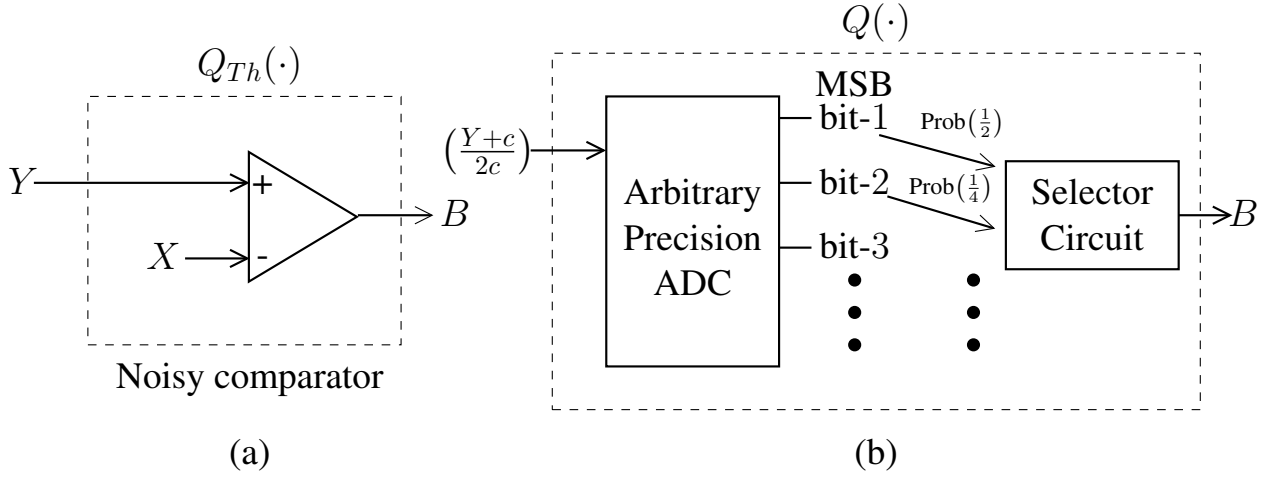
Fig. 4. **The $Q_{ij}^{Th}(\cdot)$ function in (2.1) and the $Q(\cdot)$ function of [13]** *suggest markedly different hardware implementations. The former naturally suggests (a), where the binary quantized value is produced by a simple comparison to a random threshold X. The latter suggests (b), where an arbitrarily–precise ADC circuitry probabilistically selects an arbitrary bit of the observed value. Interestingly, these two implementations produce statistically equivalent quantized outputs B given identical inputs Y.*

Specifically, the noiseless observation Y_i, at sensor i, and the corresponding quantized binary sample B_i become

$$Y_i = s(t_i),$$
$$B_i = Q(Y_i) \triangleq \mathrm{sgn}(Y_i + X_i),$$

where X_i is iid dithering noise with a known distribution $p_X(\cdot)$ which satisfies certain technical assumptions as given in [12, Section II.A]. Note that taking the sign of the sum of the observation and random dither X_i is equivalent to comparing with the threshold $-X_i$. Thus the quantization function $Q(\cdot)$ of [12] is equivalent[7] to a comparator with a random threshold that is distributed according to $p_X(-x)$. The quantization function $Q_{ij}^{Th}(\cdot)$ in (2.1) can be viewed as a special case of this where $p_X(-x)$ is the uniform distribution over $[-c, c]$. The constructive scheme of Section IV shows that $Q_{ij}^{Th}(\cdot)$ can in fact be used even on noisy field observations with an additive noise of *unknown* distribution.

B. Parameter Estimation with One–Bit Messages

The parameter estimation problem in [13] corresponds to the special case of spatially constant fields with $k = 1$ ($s(t_i) = s$ for all i where the index j is omitted since $j = k = 1$) which is addressed by Corollary 3.1. We summarize below the key features of the randomized binary quantizer proposed in [13] and show that the randomized 1–bit SQ function $Q(\cdot)$ of [13] is statistically and MSE performance–wise equivalent to the uniform random threshold quantizer $Q_{ij}^{Th}(\cdot)$ in (2.1). However, the $Q(\cdot)$ function of [13] implicitly requires sensors of arbitrarily high precision, a property that is undesirable for sensor hardware implementations.

In [13], each sensor i first shifts and scales it observation Y_i into interval $[0, 1]$ creating the value $\widetilde{Y}_i \triangleq \left(\frac{Y_i + c}{2c}\right)$. Next, each sensor i generates an auxiliary random variable α_i,

which is iid across sensors and is geometrically distributed over the set of all positive integers: $\mathbb{P}(\alpha_i = j) = 2^{-j}$ for all $j \in \{1, 2, 3, \ldots, \infty\}$. The final quantized binary sample B_i reported by sensor i is given by the α_i^{th} bit in the binary expansion of $\widetilde{Y}_i$:

$$B_i = Q(Y_i) \triangleq B(\widetilde{Y}_i, \alpha_i),$$
$$\text{where } \widetilde{Y}_i = \sum_{j=1}^{\infty} B(\widetilde{Y}_i, j) 2^{-j}. \qquad (5.6)$$

Here, $B(\widetilde{Y}_i, j)$ denotes the jth bit of $\widetilde{Y}_i$. For example, if $\widetilde{Y}_i = 0.375$, then the first four bits of its binary expansion are given by $B(\widetilde{Y}_i, 1) = 0$, $B(\widetilde{Y}_i, 2) = 1$, $B(\widetilde{Y}_i, 3) = 1$, and $B(\widetilde{Y}_i, 4) = 0$. If $\alpha_i = 3$, then sensor i reports $B_i = 1$. This method for generating binary sensor messages requires sensors to have the operational ability to quantize an observed real number (the normalized values $\widetilde{Y}_i$) to an arbitrarily high bit–resolution. Note that the binary values B_i are iid across all sensors and that its expected value is given by

$$
\begin{aligned}
\mathbb{E}[B_i] &= \mathbb{E}_{\widetilde{Y}_i}[\mathbb{E}_{B_i}[B_i | \widetilde{Y}_i]] \\
&= \mathbb{E}_{\widetilde{Y}_i}\left[\sum_{j=1}^{\infty} B(\widetilde{Y}_i, j) 2^{-j}\right] \\
&= \mathbb{E}_{\widetilde{Y}_i}[\widetilde{Y}_i] \\
&= \mathbb{E}\left[\frac{Y_i + c}{2c}\right] \qquad (5.7) \\
&= \frac{\mathbb{E}[s + Z_i] + c}{2c} = \left(\frac{s + c}{2c}\right). \qquad (5.8)
\end{aligned}
$$

In sharp contrast to the $Q(\cdot)$ function described above, which requires sensors to have the operational ability to resolve any arbitrary bit in the binary expansion of their normalized observations, $Q_{ij}^{Th}(\cdot)$ requires only a noisy comparator. Despite the markedly different operational imple-

---

[7]The sign function maps to $\{-1, +1\}$ whereas a threshold comparator maps to $\{0,1\}$. However, the replacement of the -1 symbol with the 0 symbol is unimportant from an estimation viewpoint.

mentations of $Q(\cdot)$ and $Q_{ij}^{Th}(\cdot)$ (see (5.6), (2.1), and Fig. 4 which depicts hardware implementations) they are in fact statistically identical: the binary quantized values B_i generated by the two schemes have the same $p_{B_{ij}|Y_{ij}}(\cdot)$ and $p_{B|s}(\cdot)$ functions where $p_{B_{ij}|Y_{ij}}(\cdot)$ is the conditional expectation of B_i given $Y_i = y_i$ and $p_{B|s}(\cdot)$ is the unconditional expectation of B_i parameterized by the underlying field value $s(t_i) = s$. These expectations have been evaluated in (5.7), (5.8), (A.1) and (A.2), and we see that for both functions

$$\mathbb{E}[B_i|Y_i = y_i] = p_{B_{ij}|Y_{ij}}(y_i) = \left(\frac{y_i + c}{2c}\right), \text{ and}$$
$$\mathbb{E}[B_i] = p_{B|s}(s(t_i)) = \left(\frac{s(t_i) + c}{2c}\right).$$

This statistical equivalence allows the two quantization functions $Q(\cdot)$ and $Q_{ij}^{Th}(\cdot)$ to be interchanged without affecting the estimation performance.

VI. CONCLUDING REMARKS

The results of this work show that for the distributed field reconstruction problem, for every point of continuity of every field snapshot, it is possible to drive the MSE to zero with increasing sensor density while ensuring that the per–sensor bitrate and sensing–related network overhead rate simultaneously go to zero. This can be achieved with noisy threshold (one–bit) comparators with the minimal knowledge of signal and noise dynamic ranges provided that the noise samples are zero–mean, and independent across sensors and the underlying field, and the sensor placement and sampling schedule satisfy a certain uniformity property. The rate of decay of MSE with increasing sensor density is related to the the local and global smoothness characteristics of the underlying fields and is order–optimal for the class of spatially constant fields. This work has further clarified the utility of randomization for signal acquisition in a distributed sensor network context and has attempted to systematically account for sensor placement and hardware issues in addition to the typical constraints encountered in related studies. Cramér–Rao lower–bounds for general spatially–varying fields in terms of local and global field smoothness characteristics as well as randomized sensor deployment strategies with probabilistic guarantees for near–uniform sensor deployment together with an analysis of associated sensor over–provisioning costs is part of our ongoing work.

APPENDIX A
PROOF OF THEOREM 3.1

First, note that the expected value of the binary message B_{ij} is given by

$$
\begin{aligned}
\mathbb{E}[B_{ij}] &= \mathbb{E}[\mathbf{1}(Y_{ij} > X_{ij})] \\
&= \mathbb{E}_{Y_{ij}}[\mathbb{E}_{X_{ij}}[\mathbf{1}(Y_{ij} > X_{ij})|Y_{ij}]] \\
&= \mathbb{E}_{Y_{ij}}[\mathbb{P}(X_{ij} < Y_{ij}|Y_{ij})] \\
&= \mathbb{E}_{Y_{ij}}\left[\frac{Y_{ij} + c}{2c}\right] \quad (A.1) \\
&= \frac{\mathbb{E}[s_j(t_i) + Z_{ij}] + c}{2c} \\
&= \left(\frac{s_j(t_i) + c}{2c}\right), \quad (A.2)
\end{aligned}
$$

which is the value of the field $s_j(\cdot)$ at location t_i shifted and normalized to the interval $[0, 1]$. Note that this result holds for any $F_{\mathbf{Z}}(\mathbf{z}) \in \mathcal{F}$.

Using (A.2) we can bound the bias and the variance of the estimator $\widehat{S}_j(t)$. The bound on the MSE follows from the bounds on these values since, for any estimator of a non-random parameter, we have

$$\mathrm{MSE}\left(\widehat{S}_j(t)\right) = \mathrm{bias}^2\left(\widehat{S}_j(t)\right) + \mathrm{var}\left(\widehat{S}_j(t)\right). \quad (A.3)$$

We bound the magnitude of bias of the estimate $\widehat{S}(t)$ in the following way

$$
\begin{aligned}
\left|\mathrm{bias}\left(\widehat{S}_j(t)\right)\right| &= \left|\mathbb{E}\left[\widehat{S}_j(t) - s_j(t)\right]\right| \\
&= \left|\mathbb{E}\left[2c\left[\frac{1}{N^d}\sum_{\{i:t_i \in T_{\{j\}}^N(t)\}} B_{ij}\right]\right. \right. \\
&\qquad \left.\left. -c - s_j(t)\right]\right| \\
&= \left|2c\left[\frac{1}{N^d}\sum_{\{i:t_i \in T_{\{j\}}^N(t)\}} \mathbb{E}[B_{ij}]\right] \right. \\
&\qquad \left. -c - s_j(t)\right| \\
&\stackrel{(i)}{=} \left|2c\left[\frac{1}{N^d}\sum_{t_i \in T_{\{j\}}^N(t)} \left(\frac{s_j(t_i) + c}{2c}\right)\right] \right. \\
&\qquad \left. -c - s_j(t)\right| \\
&= \left|\frac{1}{N^d}\sum_{t_i \in T_{\{j\}}^N(t)} (s_j(t_i) - s_j(t))\right|
\end{aligned}
$$

$$\leq \quad \frac{1}{N^d} \sum_{t_i \in T^N_{\{j\}}(t)} |s_j(t_i) - s_j(t)|$$

$$\overset{(ii)}{\leq} \quad \frac{1}{N^d} \sum_{t_i \in T^N_{\{j\}}(t)} \omega_j \left(\|t - t_i\|, t \right)$$

$$\overset{(iii)}{\leq} \quad \frac{1}{N^d} \sum_{t_i \in T^N_{\{j\}}(t)} \omega_j \left(\left(\frac{Nm}{\sqrt[d]{L} - 1} \right), t \right)$$

$$= \quad \omega_j \left(\left(\frac{Nm}{\sqrt[d]{L} - 1} \right), t \right)$$

$$\overset{(iv)}{\leq} \quad \widetilde{\omega}_j \left(\frac{Nm}{\sqrt[d]{L} - 1} \right), \tag{A.4}$$

where (i) follows from (A.2), (ii) and (iv) follow from Definitions 2.2 and 2.1, and (iii) follows from (4.4) and because the local modulus of continuity is a nondecreasing function of its first argument for each fixed value of its second argument.

The variance of the estimate is bounded by

$$\text{var}[\widehat{S}_j(t)] = \text{var}\left[2c \left[\frac{1}{N^d} \sum_{\{i:t_i \in T^N_{\{j\}}(t)\}} B_{ij} \right] - c \right]$$

$$= \left(\frac{4c^2}{N^{2d}} \right) \sum_{\{i:t_i \in T^N_{\{j\}}(t)\}} \text{var}[B_{ij}] \tag{A.5}$$

$$\leq \left(\frac{4c^2}{N^{2d}} \right) \left(\frac{N^d}{4} \right) = \left(\frac{c^2}{N^d} \right), \tag{A.6}$$

where we used standard properties of variance and the fact that $\{B_{ij}\}$ are independent to obtain (A.5), and we used the fact the variance of a random variable that takes values only in the set $\{0, 1\}$ is bounded by $(1/4)$ to obtain (A.6).

Combining these bounds for the bias and variance given in (A.4) and (A.6) of the estimator and using the identity in (A.3), we get the desired bound on the MSE for all $t \in G$, $j \in \{1, \ldots, k\}$, and $F_{\mathbf{Z}}(\mathbf{z}) \in \mathcal{F}$. ∎

REFERENCES

[1] J. Chou, D. Petrovic, and K. Ramchandran, "Tracking and exploiting correlations in dense sensor networks," in *Proc. Annual Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, Nov. 2002.

[2] J. M. Kahn, R. H. Katz, and K. S. J. Pister, "Next century challenges: Mobile networking for "Smart Dust"," in *Proc. ACM International Conference on Mobile Computing and Networking (MOBICOM)*, Seattle, WA, Aug. 1999, pp. 271–278. [Online]. Available: citeseer.nj.nec.com/kahn99next.html

[3] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 1st ed. New York, NY: Wiley–Interscience, 1991.

[4] M. Gastpar and M. Vetterli, "Source–channel communication in sensor networks," *Lecture Notes in Computer Science*, vol. 2634, pp. 162–177, Apr. 2003.

[5] M. Gastpar, B. Rimoldi, and M. Vetterli, "To Code, or not to code: Lossy source–channel communication revisited," *IEEE Trans. Info. Theory*, vol. IT–49, pp. 1147–1158, May 2003.

[6] R. Nowak, U. Mitra, and R. Willet, "Estimating inhomogenous fields using wireless sensor networks," *IEEE J. Sel. Areas Commun.*, vol. 22, no. 6, pp. 999–1006, Aug. 2004.

[7] K. Liu, H. El–Gamal, and A. Sayeed, "On optimal parametric field estimation in sensor networks," in *Proc. IEEE/SP 13th Workshop on Statistical Signal Processing*, Jul. 2005, pp. 1170–1175.

[8] W. Bajwa, A. Sayeed, and R. Nowak, "Matched source–channel communication for field estimation in wireless sensor networks," in *Proc. Fourth Inttl. Symposium Information Processing in Sensor Networks*, Apr. 2005, pp. 332–339.

[9] N. Liu and S. Ulukus, "Optimal distortion–power tradeoffs in Gaussian sensor networks," in *Proc. IEEE International Symposium on Information Theory*, Seattle, WA, USA, Jul. 2006, pp. 1534–1538.

[10] Q. Zhao, A. Swami, and L. Tong, "The interplay between signal processing and networking in sensor networks," *IEEE Signal Processing Magazine*, vol. 23, no. 4, pp. 84–93, Jul. 2006.

[11] E. Masry and S. Cambanis, "Consistent estimation of continuous–time signals from nonlinear transformations of noisy samples," *IEEE Trans. Info. Theory*, vol. IT–27, pp. 84–96, Jan. 1981.

[12] E. Masry, "The reconstruction of analog signals from the sign of their noisy samples," *IEEE Trans. Info. Theory*, vol. IT–27, no. 6, pp. 735–745, Nov. 1981.

[13] Zhi–Quan Luo, "Universal decentralized estimation in a bandwidth constrained sensor network," *IEEE Trans. Info. Theory*, vol. IT–51, pp. 2210–2219, Jun. 2005.

[14] T. Berger, Z. Zhang, and H. Viswanathan, " The CEO problem [multiterminal source coding]," *IEEE Trans. Info. Theory*, vol. IT–42, pp. 887–902, May. 1996.

[15] H. Viswanathan and T. Berger, "The quadratic Gaussian CEO problem," *IEEE Trans. Info. Theory*, vol. IT–43, pp. 1549–1559, Sep. 1997.

[16] V. Prabhakaran, D. Tse, and K. Ramchandran, "Rate-region of the quadratic Gaussian CEO problem," in *Proc. IEEE International Symposium on Information Theory*, Chicago, IL, Jun. 2004, p. 119.

[17] A. Kashyap, L. A. Lastras-Montano, C. Xia, and Z. Liu, "Distributed source coding in dense sensor networks," in *Proc. Data Compression Conference*, Snowbird, UT, Mar. 2005, pp. 13–22.

[18] D. L. Neuhoff and S. S. Pradhan, "An upper bound to the rate of ideal distributed lossy source coding of densely sampled data," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, Toulouse, France, May 2006, pp. 1137–1140.

[19] D. Marco, E. J. Duarte-Melo, M. Liu, and D. Neuhoff, "On the many–to–one transport capacity of a dense wireless sensor network and the compressibility of its data," in *Information Processing in Sensor Networks, Proceedings of the Second International Workshop, Palo Alto, CA, USA, April 22-23, 2003*, ser. Lecture Notes in Computer Science edited by L. J. Guibas and F. Zhao, Springer, New York, 2003, Apr., pp. 1–16.

[20] P. Ishwar, A. Kumar, and K. Ramchandran, "Distributed sampling for dense sensor networks: A "bit-conservation" principle," in *Information Processing in Sensor Networks, Proceedings of the Second International Workshop, Palo Alto, CA, USA, April 22-23, 2003*, ser. Lecture Notes in Computer Science edited by L. J. Guibas and F. Zhao, Springer, New York, 2003, pp. 17–31.

[21] A. Kumar, P. Ishwar, and K. Ramchandran, "On distributed sampling of smooth non-bandlimited fields," in *Proc. Third Inttl. Symposium Information Processing in Sensor Networks*. New York, NY: ACM Press, 2004, pp. 89–98.

[22] C. D. Aliprantis and O. Burkinshaw, *Principles of Real Analysis*. San Diego, CA: Academic Press, 1990.

[23] H. Niederreiter, *Random Number Generation and Quasi-Monte Carlo Methods*. Society for Industrial and Applied Mathematics, 1992.

[24] M. Drmota and R. F. Tichy, *Sequences, discrepancies, and applications*. Springer, Berlin, 1997.

[25] M. Leoncini, G. Resta, and P. Santi, "Analysis of a wireless sensor dropping problem in wide-area environmental monitoring," in *Proc. Fourth IEEE-ACM Symposium on Information Processing in Sensor Networks*, Apr. 2005, pp. 239–245.

[26] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory*, 1st ed. Upper Saddle River, NJ: Prentice–Hall, 1993, vol. 1.