Decentralized Inference over Multiple Access Channels

Ke Liu, Hesham El Gamal†, and Akbar Sayeed

Abstract

The problem of decentralized inference over multiple access fading channels is considered from a joint source-channel coding perspective. In our setting, spatially distributed wireless sensor nodes collect observation data about a hidden source and communicate relevant statistics to a receiver node that generates the desired inference (estimation or detection) about unknown source parameters. We assume a homogeneous sensor signal field in which information about the source parameters is distributed in space (across sensors) and time in an independent and identically distributed (i.i.d.) manner. Our study characterizes the asymptotic performance of an identical source-channel mapping for i.i.d. sensor observation data in which the same encoder is used at all the nodes. This scheme generalizes many existing methods including uncoded transmission and type-based multiple access. Sufficient conditions are obtained under which our identical mapping approach achieves the genie-aided scaling law (in the number of nodes) associated with a noiseless channel, even when the nodes transmit with asymptotically vanishing power. Our analysis also elucidates the critical impact of channel state information on the achievable performance. In particular, it identifies scenarios in which identical mapping fails and a simple non-identical mapping scheme is shown to improve performance. Numerical examples are provided to illustrate the theoretical principles derived in the paper.

Index Terms

Decentralized inference, identical mapping, uncoded transmission, the method of types, sensor networks.

Dr. Liu (kel@qualcomm.com) and Prof. El Gamal (helgamal@ece.osu.edu) are with Department of Electrical and Computer Engineering, Ohio State University, 2015 Neil Ave, Columbus, OH 43210. Prof. Sayeed (akbar@engr.wisc.edu) is with Department of Electrical and Computer Engineering, University of Wisconsin-Madison, 1415 Engineering Drive, Madison, WI 53706.

I. INTRODUCTION

We consider a decentralized inference problem where a receiver node tries to reconstruct a hidden source by collecting information from a plurality of sensor nodes over a *multiple access channel* (MAC). This is an abstraction of many important applications of sensor networks that involve detection or estimation of relevant signal field parameters. Our emphasis on the MAC scenario is motivated by the shared nature of wireless medium that provides the physical communication environment for sensor networks.

The problem of decentralized inference over a wireless MAC involves the interplay between sensing and communication which inherently calls for utilizing tools from statistical decision theory and communications theory. Earlier related works have concentrated on the so-called *multi-terminal inference* problem which imposes rate constraints on messages, or exchanges between the sensor nodes and the fusion center, to abstract the impact of communication channels (see [1] and reference wherein). Notably, works on the "CEO problem" [2], [3], [4] characterize the asymptotic performance of decentralized inference in the limit of large number of nodes, given a total **rate** constraint. These studies, though insightful, represent a suboptimal approach corresponding to a separate optimization of the source and channel encoders. Closely related to decentralized inference is the problem of *data collection* where the distributed observations are recovered at the receiver node. The joint source-channel coding in [5] gives an improved achievable region for communicating correlated sources over the MAC, while the capacity region remains unsolved. The problem of decentralized estimation of a single Gaussian source over multiple access channels was studied in [6], [7] from a joint source-channel coding perspective, and extended to multiple Gaussian sources in [8]. In particular, these works emphasize the superiority of *uncoded transmission*, as opposed to a separated source-channel coding approach, from the viewpoint of the distortion scaling with the number of nodes.

In this paper, we develop a paradigm for decentralized inference that exploits the MAC structure from a joint source-channel coding perspective. We assume that the spatio-temporal sensor observation data are generated by an unknown source and we are interested in generating inference about some relevant source parameters. Thus, in general, the information retrieval requirements are less stringent in our inference problem as compared to the data collection problem [5]. We assume a homogeneous signal field in which source information is encoded into spatio-temporal sensor data in an independent and identically distributed (i.i.d.) fashion. In practice, homogeneity may only apply to a part of the sensor field, and the i.i.d. assumption can be approximately satisfied by sufficiently separated sensors. Our work reveals the impact of power scaling, channel fading statistics, and channel state information on the achievable performance in the asymptotic scenario with a large number of sensor nodes. More specifically, we show that finite total power is sufficient for achieving the genie-aided distortion scaling (i.e., noiseless channel) when a joint source-channel coding approach is used. Here, we focus on the *identical source-channel mapping* in which every node uses the same encoder. The identical mapping is attractive for scalable sensor networks due to its modular structure and was recently shown to be optimal in certain decentralized detection scenarios. It is worth noting that identical mapping includes uncoded transmission as an important special case in which the encoder map is simply the **identity map** [7].

The following summarizes our main results and contributions.

1) We obtain a sufficient condition that ensures the asymptotic optimality of the identical mapping scheme. Intuitively, this condition requires the source information to be embedded in the conditional mean of the encoder output. We, therefore, refer to this condition as the *mean condition*.
2) We use the mean condition to establish the following results: a) For **non-zero mean** fading channels, identical mapping achieves the genie-aided linear decay in the mean square estimation error and exponential decay in the detection error probability (with the number of sensor nodes), b) For **zero mean** fading channels, the genie-aided scaling is still achievable if the channel coefficients are known at the transmitter, c) When the channel coefficients are only known at the receiver, the performance of identical mapping degrades significantly, exhibiting no better than double logarithmic decay in the mean square estimation error. Interestingly, these scaling laws can be attained under the assumption that the *power density* (or power per node) diminishes as the number of nodes grows.
3) We propose a simple non-identical mapping scheme that achieves a logarithmic decay in the mean square estimation error with zero-mean fading when no channel state information is available.
4) When the source statistical knowledge is not available to the sensor nodes, we propose the Type-Based-Multiple-Access (TBMA) approach (see Section III-B) as a universal encoding

map that satisfies the scaling law optimality condition for arbitrary discrete sources. This result generalizes our earlier works and the works of Mergen and Tong [10], [11], [12] where the critical impact of the channel mean on the performance has been observed in some specific contexts.

The rest of paper is organized as follows.[1] The next section introduces the mathematical model for decentralized inference over the MAC. Section III investigates the scaling laws and the optimality conditions for identical mapping. Examples and numerical results are presented in Section IV. Concluding remarks are presented in Section V. All the proofs are relegated to the appendices to enhance the flow of the paper.

II. System Model

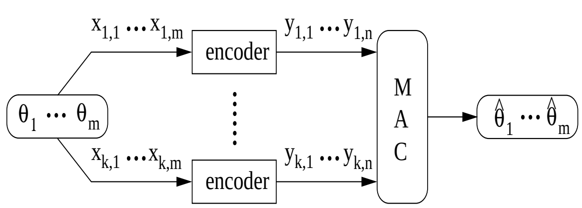

Fig. 1. A schematic illustrating decentralized inference over multiple access channels.

Fig. 1 illustrates our formulation for decentralized inference over multiple access channels. For convenience, all the signals are real-valued. The source symbols are denoted by $\{\theta_j\}$, $\theta_j \in \Theta$ where j denotes the time index. The observation data $x_{i,j}$'s are conditionally i.i.d. across time slots and spatial nodes, that is, for $i_1 \neq i_2$ or $j_1 \neq j_2$,

$$P_{\theta_{j_1},\theta_{j_2}}(x_{i_1,j_1}, x_{i_2,j_2}) = P_{\theta_{j_1}}(x_{i_1,j_1}) P_{\theta_{j_2}}(x_{i_2,j_2}). \tag{1}$$

Consider a block encoder at node-i which selects an n-dim codeword vector $\mathbf{y}_i = (y_{i,1}, \ldots, y_{i,n})^T$ for every m-dim observation vector $\mathbf{x}_i = (x_{i,1}, \ldots, x_{i,m})^T$ (corresponding to the source vector $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_m)^T$). The *bandwidth expansion factor* is therefore given by n/m channel uses

[1] The following notational convention is assumed throughout the paper. Lower-case bold symbols denote column vectors, while upper-case bold symbols denote matrices. Transposition is denoted by $(\cdot)^T$. The usual L^2 norm is denoted by $\|\cdot\|_2$. Notation $\mathbb{E}_{\boldsymbol{\theta}}$ (or $P_{\boldsymbol{\theta}}$) denotes the expectation operator (or probability measure) conditioned on $\boldsymbol{\theta}$.

per source symbol. We assume no inter-node communications so that one node cannot access the observation data at the other nodes. In this paper we shall primarily focus on the identical mapping (IM) approach in which the same encoder (or data processing) is applied at every node.[2] The received signal over the noisy MAC can be written as

$$z_j = \sqrt{\frac{P_{tot}(k)}{k}} \sum_{i=1}^{k} a_{i,j} y_{i,j} + w_j, \tag{2}$$

or in a vector form

$$\mathbf{z} = \sqrt{\frac{P_{tot}(k)}{k}} \sum_{i=1}^{k} \mathbf{A}_i \mathbf{y}_i + \mathbf{w} \tag{3}$$

where $\mathbf{A}_i = \text{diag}(a_{i,1}, \ldots, a_{i,n})$, the $a_{i,j}$'s are the unit power i.i.d channel coefficients, $\mathbf{w} = (w_1, \ldots, w_n)$ is the zero-mean white Gaussian noise vector (i.e., $w \sim N(0,1)$), and $P_{tot}(k)$ is the total power (as a function of the number of sensor nodes k). This formulation allows for a general modeling of network power scaling. The slowest scaling considered in our work, i.e., $P_{tot}(k) \sim O(1)$, corresponds to *finite total power*, which has diminishing power density.

In decentralized inference, the essential goal of the encoder/decoder design is to facilitate reconstructing the source $\boldsymbol{\theta}$ from the received signal at the fusion center with minimal distortion. For continuous source alphabet the inference problem is usually referred to as *decentralized estimation*, whose performance is often measured by the *mean-square error* (MSE)

$$\text{MSE}_{\boldsymbol{\theta}} = \mathbb{E}_{\boldsymbol{\theta}} \|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|_2^2 = \mathbb{E}_{\boldsymbol{\theta}} \operatorname{tr}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T. \tag{4}$$

For finite source alphabet the problem corresponds to *decentralized detection* whose performance is gauged by the *detection error probability* (DEP)

$$P_{e,\boldsymbol{\theta}} = P_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}} \neq \boldsymbol{\theta}). \tag{5}$$

III. Asymptotic Analysis of Identical Mapping

As a benchmark, we borrow from the classical literature on statistical inference the idealistic *centralized inference* scenario. Assuming perfect, genie-aided, access to the observed data, the inference problem achieves a $1/k$ decay in MSE (for estimation) or an e^{-k} decay in DEP (for detection) [13], [14], [15].

[2]For technical clarity the identical encoder considered here is independent of the number of nodes.

In the CEO problem formulation [2], [3], [4], the MAC is replaced by a total rate constraint. The overall inference task is, therefore, divided into separate source and channel coding, i.e., the observed data is first compressed into discrete messages which are then channel-coded to ensure reliable transmission. This approach gives a $1/\log(kP_{tot})$ scaling in MSE due to the MAC rate constraint $\log(kP_{tot})$ [3]. Unfortunately, this implies the necessity of an exponential power scaling $P_{tot} \sim e^k$ in order to achieve the genie-aided (i.e., noiseless channel) benchmark.

In this section, we study the spatial asymptotics of decentralized inference using a joint source-channel approach that exploits the intrinsic MAC structure. We show that finite total power is sufficient for achieving the genie-aided performance in many non-trivial scenarios. Our arguments are based on the identical mapping (IM) approach. Interestingly, IM can be viewed as a generalization of the uncoded transmission scheme proposed for the Gaussian source case [7]. In the following, we would unify, whenever possible, the estimation and detection sub-problems in order to emphasize the common intuition shared by both. We define the channel state information (CSI) to be the instantaneous knowledge of the MAC fading coefficients, i.e., $a_{i,j}$'s. The statistical information about the problem, such as the statistics of the data and channel coefficients, is assumed to be known at the transmitter and receiver unless explicitly noted otherwise.

A. Sufficient Conditions for Genie-Aided Scaling

We first study the case when no CSI is available to the transmitter nor the receiver. It follows from the independence of the source and channel that $\mathbb{E}_{\boldsymbol{\theta}} \mathbf{A}_i \mathbf{y}_i = \mu_a \mathbf{d}_{\boldsymbol{\theta}}$ where $\mu_a = \mathbb{E}\, a_{i,j}$ and $\mathbf{d}_{\boldsymbol{\theta}} = \mathbb{E}_{\boldsymbol{\theta}}\, \mathbf{y}_i$. We rewrite (3) as

$$\tilde{\mathbf{z}} = \frac{\mathbf{z}}{\sqrt{P_{tot}}} = \frac{1}{\sqrt{k}} \sum_{i=1}^{k} (\mathbf{A}_i \mathbf{y}_i - \mu_a \mathbf{d}_{\boldsymbol{\theta}}) + \sqrt{k} \mu_a \mathbf{d}_{\boldsymbol{\theta}} + \frac{1}{\sqrt{P_{tot}}} \mathbf{w}. \tag{6}$$

The covariance matrix of $\mathbf{A}_i \mathbf{y}_i$ admits the expression

$$\begin{aligned} \mathbf{\Sigma}_{\boldsymbol{\theta}} &= \mathbb{E}_{\boldsymbol{\theta}} (\mathbf{A}_i \mathbf{y}_i - \mu_a \mathbf{d}_{\boldsymbol{\theta}})(\mathbf{A}_i \mathbf{y}_i - \mu_a \mathbf{d}_{\boldsymbol{\theta}})^T \\ &= \mathbb{E}_{\boldsymbol{\theta}}\, \mathbf{A}_i \mathbf{y}_i \mathbf{y}_i^T \mathbf{A}_i^T - \mu_a^2 \mathbf{d}_{\boldsymbol{\theta}} \mathbf{d}_{\boldsymbol{\theta}}^T \\ &= \mathbf{R}_{\mathbf{a}} \odot \mathbf{R}_{\boldsymbol{\theta}} - \mu_a^2 \mathbf{d}_{\boldsymbol{\theta}} \mathbf{d}_{\boldsymbol{\theta}}^T \end{aligned} \tag{7}$$

where $\mathbf{R}_{\mathbf{a}}$ is the correlation matrix of the channel vector $\mathbf{a}_i = (a_{i1}, \ldots, a_{in})^T$, $\mathbf{R}_{\boldsymbol{\theta}}$ is the correlation matrix of the codeword vector, and the notation $\odot$ stands for the element-wise multiplication.

The conditional mean of the codewords, i.e., $\mathbf{d}_{\boldsymbol{\theta}} = \mathbb{E}_{\boldsymbol{\theta}}\, \mathbf{y}_i$, can be seen as a deterministic map that associates every source vector $\boldsymbol{\theta} \in \Theta^m$ with a corresponding vector $\mathbf{d}_{\boldsymbol{\theta}}$ in $\mathbb{R}^n$. The cornerstone in our approach is to observe that if *the mean map $\mathbf{d}_{\boldsymbol{\theta}}$ encodes significant information about the source, then one could "invert" the map to recover $\boldsymbol{\theta}$.*

A simple estimator of $\mathbf{d}_{\boldsymbol{\theta}}$ is suggested by (6). More specifically

$$\hat{\mathbf{z}} = \frac{\tilde{\mathbf{z}}}{\sqrt{k}\mu_a} = \mathbf{d}_{\boldsymbol{\theta}} + \frac{1}{k\mu_a}\sum_{i=1}^{k}(\mathbf{A}_i\mathbf{y}_i - \mu_a\mathbf{d}_{\boldsymbol{\theta}}) + \frac{1}{\mu_a\sqrt{kP_{tot}(k)}}\mathbf{w}, \tag{8}$$

provided that $\mu_a \neq 0$, that is, the fading channel coefficients has a non-zero mean. The decoder then chooses the source parameter $\hat{\boldsymbol{\theta}}$ according to[3]

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \|\hat{\mathbf{z}} - \mathbf{d}_{\boldsymbol{\theta}}\|_2. \tag{9}$$

The error associated with this method comes from the noisy perturbation in $\hat{\mathbf{z}}$ whose variance decays as $1/k$, i.e.,

$$\mathbb{E}_{\boldsymbol{\theta}}(\hat{\mathbf{z}} - \mathbf{d}_{\boldsymbol{\theta}})(\hat{\mathbf{z}} - \mathbf{d}_{\boldsymbol{\theta}})^T = \frac{1}{k\mu_a^2}\boldsymbol{\Sigma}_{\boldsymbol{\theta}} + \frac{1}{k\mu_a^2 P_{tot}(k)}\mathbf{I}_n. \tag{10}$$

Such a $1/k$ decay in the distortion of $\mathbf{d}_{\boldsymbol{\theta}}$ translates into $1/k$ decay in MSE or e^{-k} decay in DEP. A more refined argument is given by the next result whose proof is relegated to Appendix I.

Theorem 1: Under non-zero mean channel fading ($\mu_a \neq 0$), the identical mapping achieves the centralized benchmark ($1/k$ decay in the MSE for estimation or e^{-k} decay in the DEP for detection), provided that

1) The total power scales at least as $P_{tot}(k) \sim O(1)$.
2) The mean map $\mathbf{d}_{\boldsymbol{\theta}} : \Theta^m \to \mathbb{R}^n$ is one-to-one.
3) For the estimation problem, $\mathbf{d}_{\boldsymbol{\theta}}$ is an embedding (a function which is a homeomorphism onto its image). In particular, the derivative $\partial_{\boldsymbol{\theta}}\mathbf{d}_{\boldsymbol{\theta}}$ is of full rank.
4) For the detection problem, the codeword symbols and channel coefficients (viewed as random variables) satisfy certain technical conditions in terms of moment generating function (see Appendix I.), which exclude "heavy-tailed" distributions (This is a class of distributions such that $P(|X| > r) \sim 1/r^\alpha$ for some $\alpha > 0$. Note that the moment generating function does not exist for heavy-tailed distributions.).

[3]Note that this decision rule applies to both estimation and detection.

Theorem 1 shows that the genie-aided scaling is achievable under a finite total power by IM (unlike separate source-channel strategies which require exponential power scaling). The main idea is that there is no attempt here to retrieve the observed data and the receiver aims only to recover the statistical mean of the codewords. In this setting, the MAC is naturally matched to our objective of averaging-out the measurement and channel noise. Of course, a non-zero channel fading is crucial and the mean map of the codes shall satisfy the conditions in Theorem 1, collectively referred to as the mean condition in the sequel.

Next we further elucidate the effect of power scaling on asymptotic efficiency of the IM approach. Here the asymptotic efficiency measures the pre-scaling constant c: c/k for MSE or e^{-kc} for DEP (the latter is often called detection error exponent). For the estimation problem we apply the Cramer-Rao analysis (see, e.g., [13], [15]) to evaluate the asymptotics of the *Fisher information matrix* (FIM) associated with IM. In the following result we assume continuous random variables with smooth probability density functions.

Lemma 1: Let $P_{tot}(\infty) = \lim_{k\to\infty} P_{tot}(k)$. Denote by h_∞ the Gaussian density function of $N(\mathbf{0}, \frac{1}{P_{tot}(\infty)}\mathbf{I}_n)$. (If $P_{tot}(\infty) = \infty$, then h_∞ is the Delta function.) Let $\psi(\mathbf{t})$ denote the characteristic function of $\mathbf{A}\mathbf{y} - \mu_a \mathbf{d}_{\boldsymbol{\theta}}$. Assume[4]

$$\int |\mathbf{t}|^s \sup_k |\psi(\mathbf{t}/\sqrt{k})^k| d\mathbf{t} < \infty, \quad s = 0, 1. \tag{11}$$

Then

$$\liminf_{k\to\infty} \frac{1}{k} \operatorname{tr} \mathbf{J}_k \geq \mu_a^2 \operatorname{tr}\Big\{ (\partial_{\boldsymbol{\theta}} \mathbf{d}_{\boldsymbol{\theta}})^T \cdot \big[\int \frac{(\partial_{\mathbf{u}} h_{\boldsymbol{\theta}} * h_\infty)^T (\partial_{\mathbf{u}} h_{\boldsymbol{\theta}} * h_\infty)}{h_{\boldsymbol{\theta}} * h_\infty} d\mathbf{u} \big] \cdot \partial_{\boldsymbol{\theta}} \mathbf{d}_{\boldsymbol{\theta}} \Big\} \tag{12}$$

where $*$ denotes convolution, and $h_{\boldsymbol{\theta}}$ denotes the Gaussian density function of $N(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\theta}})$. In particular, if $P_{tot}(\infty) = \infty$, the right hand side of (12) simplifies to

$$\mu_a^2 \operatorname{tr}\big\{ (\partial_{\boldsymbol{\theta}} \mathbf{d}_{\boldsymbol{\theta}})^T \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} (\partial_{\boldsymbol{\theta}} \mathbf{d}_{\boldsymbol{\theta}}) \big\}. \tag{13}$$

Proof: See Appendix II. □

According to the Cramer-Rao bound (MSE$_{\boldsymbol{\theta}} \geq \operatorname{tr} \mathbf{J}_k^{-1}$), the pre-scaling constant can be

[4] The condition in (11) is contrived to ensure the convergence of density functions in connection with the use of Central Limit Theorem in our proof. In cases where such convergence can be verified directly or indirectly via other means, this technical condition can be dropped.

characterized as[5]

$$k \cdot \mathrm{MSE}_{\boldsymbol{\theta}} \geq k \operatorname{tr} \mathbf{J}_k^{-1} \geq \frac{1}{\frac{1}{k} \operatorname{tr} \mathbf{J}_k}. \tag{14}$$

Lemma 1 provides a conservative assessment of the estimation efficiency (only a lower bound), although the stronger result that equality holds in (12) is the most desirable. (In fact, the equality result does hold for Gaussian signals.) This difficulty is due to passing the limit into the integral in our convergence argument (see Appendix II). One may attempt to establish the desired convergence by imposing certain restrictions on the signal statistics. However, not being able to state a clean and insightful interpretation of these technical conditions, we leave as a conjecture the stronger statement that the equality indeed holds in (12) for a wide range of scenarios. Having said that, choosing an encoder which maximizes the lower bound in (12) still appears as a plausible design approach. Here, we observe that any power scaling law that ensures an unbounded growth of the total power with the number of sensor nodes is sufficient to achieve the genie-aided lower bound.

For the detection problem we present an approximate analysis to illustrate key intuitions behind the IM scheme. The summation term in (6) can be approximated as Gaussian by appealing to the intuition behind the Central Limit Theorem, and hence

$$\tilde{\mathbf{z}} \approx N(\sqrt{k}\mu_a \mathbf{d}_{\boldsymbol{\theta}}, \boldsymbol{\Sigma}_{\boldsymbol{\theta}} + \frac{1}{P_{tot}(k)}\mathbf{I}). \tag{15}$$

The detection task can be thought as discerning Gaussian distributions whose mean and covariance are parameterized by the source symbol. As seen in (15), the mean of the approximated Gaussian distribution scales as $\sqrt{k}$, while the covariance, being finite, is eventually dominated by the mean in terms of the contribution to the detection performance. The pairwise detection error probability (for the new Gaussian problem) can be upper-bounded by [13]

$$P(\boldsymbol{\theta} \to \boldsymbol{\theta}') \leq \exp\{-\frac{1}{8} k \mu_a^2 \Delta^2\} \tag{16}$$

where

$$\Delta^2 = (\mathbf{d}_{\boldsymbol{\theta}} - \mathbf{d}_{\boldsymbol{\theta}'})^T (\boldsymbol{\Sigma}_{\boldsymbol{\theta}} + \frac{1}{P_{tot}(\infty)}\mathbf{I})^{-1} (\mathbf{d}_{\boldsymbol{\theta}} - \mathbf{d}_{\boldsymbol{\theta}'}) \tag{17}$$

measures the (normalized) distance between two competing decision points. Hence, we note the resemblance of (17) with its estimation counterparts in (12) and (13). Under the mean

[5] $\operatorname{tr} \mathbf{Q}^{-1} \geq (\operatorname{tr} \mathbf{Q})^{-1}$ for a positive definite matrix $\mathbf{Q}$.

condition (i.e., $\mathbf{d}_{\boldsymbol{\theta}}$ is one-to-one) the above argument establishes the exponential decay in DEP. Furthermore, the error exponent is proportional to the pairwise distance Δ. Therefore, a natural design criterion is to *maximize the minimal distance* Δ *among all the pairs*, i.e.,

$$\Delta_{min}^2 = \min_{\boldsymbol{\theta},\boldsymbol{\theta}'} \Big\{ (\mathbf{d}_{\boldsymbol{\theta}} - \mathbf{d}_{\boldsymbol{\theta}'})^T (\boldsymbol{\Sigma}_{\boldsymbol{\theta}} + \frac{1}{P_{tot}(\infty)} \mathbf{I})^{-1} (\mathbf{d}_{\boldsymbol{\theta}} - \mathbf{d}_{\boldsymbol{\theta}'}) \Big\}. \tag{18}$$

The matrix $\boldsymbol{\Sigma}_{\boldsymbol{\theta}} + \frac{1}{P_{tot}(\infty)}\mathbf{I}$ in (18) captures the joint impact of the measurement and the channel noise. The first term, $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$, is the covariance matrix corresponding to the measurement noise. The second term $\frac{1}{P_{tot}}\mathbf{I}$ reflects the effective channel noise. When the channel noise is dominant, i.e., $\operatorname{tr} \boldsymbol{\Sigma}_{\boldsymbol{\theta}} \ll 1/P_{tot}(\infty)$, the criterion (18) can be interpreted as to maximize the minimum Euclidean distance between the codeword "mean"

$$\Delta_{1,min}^2 = \min_{\boldsymbol{\theta},\boldsymbol{\theta}'} \|\mathbf{d}_{\boldsymbol{\theta}} - \mathbf{d}_{\boldsymbol{\theta}'}\|^2. \tag{19}$$

On the other hand, when the measurement noise is dominant, i.e., $\operatorname{tr} \boldsymbol{\Sigma}_{\boldsymbol{\theta}} \gg 1/P_{tot}(\infty)$, the design criterion must take into account the codeword "covariance"

$$\Delta_{2,min}^2 = \min_{\boldsymbol{\theta},\boldsymbol{\theta}'} (\mathbf{d}_{\boldsymbol{\theta}} - \mathbf{d}_{\boldsymbol{\theta}'})^T \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} (\mathbf{d}_{\boldsymbol{\theta}} - \mathbf{d}_{\boldsymbol{\theta}'}). \tag{20}$$

The above analysis has assumed no CSI at the transmitter nor the receiver. In this case, the condition $\mu_a \neq 0$ is critical to the scaling-law optimality of identical mapping. When channel fading has zero mean ($\mu_a = 0$), the identical mapping (and as a special case uncoded transmission) exhibits poor performance. This motivates an interesting question: *how much benefit can CSI produce in this scenario?*

Theorem 2: Assume $\mu_a = 0$ and the mean condition is satisfied by the encoder map.

1) Transmitter CSI: IM combined with transmitter *beam-forming* achieves the genie-aided scaling laws. Furthermore, imperfect CSI, along with a finite total power, suffices as long as the equivalent beam-forming channel has non-zero mean.
2) Receiver only CSI: IM inference performance is lower bounded by $\frac{1}{\log\log k}$ MSE decay for decentralized estimation or $\frac{1}{\log k}$ DEP decay for decentralized detection, provided that the maxima statistic of channel fading distribution, i.e., $M_k = \max_{i \leq i \leq k} |a_i|$, satisfies

$$\limsup_{k\to\infty} M_k/\sqrt{k} = 0 \quad \text{a.s.} \tag{21}$$

Proof: See Appendix III. □

Observing the very slow rate of decay corresponding to $1/\log\log k$, one can conclude from Theorem 2 that IM fails in the $\mu_a = 0$ and no transmitter CSI scenarios. In these cases non-identical encoders may achieve better performance. One such scheme is proposed in Appendix IV. It is based on quantization and time-division multiple-access (TDMA) and achieves a $1/\log k$ decay in the estimation MSE.

B. Universal Mapping Based on Types

Encoder maps that satisfy the mean condition are generally specific to the source distribution in the inference problem. This may imply that such encoders require *intelligent sensors* informed a-priori with the necessary statistical knowledge about the source. A challenging question is whether the same performance can be achieved by using *dumb sensors* with no (or minimal) information about the source statistics. For discrete sources, we present a universal solution based on *the method of types* (Readers are referred to [17], [18] for an excellent introduction to the area.). The type-based decentralized inference has been the focus of recent works [19], [10], [20], [21], [11], [12] which provide a detailed performance analysis in different contexts. Here we approach the type-based scheme from the IM framework developed in this paper. To simplify the discussion we focus on a one-shot model where a single source parameter θ corresponds to m consecutive i.i.d. observations.

Let $\mathcal{A}$ denote the finite alphabet of the observation data. The *type* of a sequence $\mathbf{x} = (x_1, \ldots, x_m)$ is a vector $\mathbf{T}$ consisting of the relative frequency of all the alphabet symbols

$$\mathbf{T}(a) = N(a|\mathbf{x})/m, \quad \forall a \in \mathcal{A}, \tag{22}$$

where $N(a|\mathbf{x})$ denotes the number of occurrences for symbol a in the sequence $\mathbf{x}$. It is worth noting that different sequences may share a common type, for example, (1,0,1,1) and (0,1,1,1) have the same type $(3/4, 1/4)$, that is, three occurrences for ‘1’ and one for ‘0’.

The extraction of type statistics can be seen as an encoder map whose conditional mean is precisely the distribution P_θ, i.e. $\mathbf{d}_\theta = \mathbf{P}_\theta$:

$$\mathbb{E}\,\mathbf{T}(a) = \mathbb{E}_\theta\, N(a|\mathbf{x})/m = \mathbb{E}_\theta\, \frac{1}{m}\sum_{i=1}^{m} I(x_i = a) = P_\theta(a) \tag{23}$$

where $I(x_i = a)$ is the indicator function of x_i being a and $\mathbf{P}_\theta = (P_\theta(a_i))_{a_i \in \mathcal{A}}$. Clearly, the mean map $\mathbf{d}_\theta : \theta \to \mathbf{P}_\theta$ contains the essential information about the source θ, and hence, satisfies the mean condition. Although further performance characterization can be obtained by exploiting the particular properties of types as done in [10], [11], [12], many important features of type-based schemes can be concisely understood by observing that type-encoding is a universal map that inherits the scaling laws properties of identical mapping (without requiring prior knowledge about the source statistics).

IV. Numerical Examples

A. Decentralized Estimation

We consider the one-shot source model

$$x_i = \theta + \eta_i, \tag{24}$$

with the i.i.d measurement noise $\eta_i \sim N(0,1)$. Previous works [6], [7] have shown that uncoded-analog transmission achieves the optimal $1/k$ MSE decay. However, analog transmission has a negative impact on amplifier design due to the large dynamic range of the signal. This motivates the use of quantization to limit the peak-to-average power ratio. For simplicity of presentation, we focus on the *one-bit quantization* where the transmit signal is either 1 or -1. More specifically, we set a quantization threshold δ such that $y = 1$ if $x \geq \delta$ or 0 otherwise. Since $x \sim N(\theta, 1)$, the quantizer induces an θ-parameterized, binary distribution on y with

$$p = P_{\boldsymbol{\theta}}(y = 1) = P_{\boldsymbol{\theta}}(x \geq \delta) = Q(\delta - \theta) \tag{25}$$

where $Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} dt$ is the Gaussian tail integral. The mean and variance of this distribution are given by

$$d_\theta = 2p - 1 = 2Q(\delta - \theta) - 1 \tag{26}$$

$$\sigma_\theta^2 = 4p(1-p) = 4Q(\delta - \theta)Q(\theta - \delta). \tag{27}$$

One can readily see that this one-bit quantization approach satisfies the mean conditions: 1) d_θ is one-to-one in θ; and 2) its derivative is of full rank

$$\partial_{\boldsymbol{\theta}} d_{\boldsymbol{\theta}} = \frac{\sqrt{2}}{\sqrt{\pi}} e^{-\frac{1}{2}(\delta-\theta)^2} \neq 0. \tag{28}$$

Therefore, by Theorem 1, this approach achieves the genie-aided scaling laws with only finite total power. Relating to sensor networks, this result provides an encouraging example where simple hardware (identical, one-bit quantizer at every node) can deliver exceptional performance.

Furthermore, the choice of δ has a significant impact on the asymptotic efficiency of the estimator. Inspired by Lemma 1, we focus on maximizing the limiting (normalized) FIM

$$\frac{(\partial_\theta d_\theta)^2}{\sigma_\theta^2} = \frac{e^{-(\delta-\theta)^2}}{2\pi Q(\delta-\theta)Q(\theta-\delta)}, \tag{29}$$

which nevertheless shows that there does not exist a uniform choice of δ which maximizes the FIM for all θ. To remedy this difficulty one can set a prior distribution on θ, e.g., $\theta \sim N(0,1)$, to get

$$\begin{aligned} \mathbb{E}_\theta \frac{(\partial_\theta d_\theta)^2}{\sigma_\theta^2} &= \mathbb{E}_\theta \frac{e^{-(\delta-\theta)^2}}{2\pi Q(\delta-\theta)Q(\theta-\delta)} \\ &\overset{(a)}{\geq} \mathbb{E}_\theta \frac{2}{\pi} e^{-(\delta-\theta)^2} \overset{(b)}{=} \frac{2\sqrt{3}}{3\pi} e^{-\frac{1}{3}\delta^2} \end{aligned} \tag{30}$$

where (a) is a consequence of the boundedness of σ_θ^2 (i.e., $p(1-p) \leq 1/4$), and (b) follows from the Gaussian quadratic integration. Note that the lower-bound in (30) is maximized by $\delta = 0$ corresponding to symmetric one-bit quantizer.

Fig. 2 compares the performance of the one-bit quantizer with that of uncoded analog transmission, with the same fixed total power of 10 dB. We can see that both schemes achieve the optimal $1/k$ scaling, as predicted by Theorem 1. Fig. 2 also demonstrates the impact of the quantizer threshold: the performance degrades as δ moves away from 0. In Fig. 3, the static channel is replaced by a Gaussian fading channel an imperfect transmitter CSI. We model the CSI as $b_i = a_i + \epsilon_i$ whose signal-to-noise ratio measures the quality of the CSI. The transmitter tries to align a phase-coherent transmission using sign(b_i). The figure demonstrates the effectiveness of beam-forming; i.e., a 10 dB noisy CSI seems to cause a very small performance degradation as compared with the perfect CSI scenario.

B. Decentralized Detection

We consider binary detection in which the source symbol θ is either 0 or 1. The binary observation sequence is generated as

$$x_{ij} = \theta_{ij} \oplus e_{ij} \tag{31}$$

or, $\mathbf{x} = \boldsymbol{\theta} \oplus \mathbf{e}$ in vector form, where $\oplus$ denotes the binary addition and the i.i.d. measurement noise e_{ij} has Bernoulli distribution $B(p)$ with p being the probability of 1. Each source symbol θ induces two Bernoulli distributions, namely, $x \sim P_0 = B(p)$ for $\theta = 0$ and $x \sim P_1 = B(1-p)$

for $\theta = 1$. We assume BPSK modulation in which the codeword symbol 1 is mapped to the real-valued transmit signal 1, and 0 to -1.

The simplest code is the *repetition code* whose mean and covariance, assuming coding rate $1/l$, can be computed as

$$\mathbf{d}_0 = -\mathbf{d}_1 = (2p-1)\mathbf{1}, \quad \mathbf{\Sigma}_0 = \mathbf{\Sigma}_1 = 4p(1-p)\mathbf{1}\mathbf{1}^T \tag{32}$$

where $\mathbf{1}$ is a column vector of l ones. According to (18), the coding distance is given by

$$\begin{aligned}\Delta^2 &= 2(1-2p)\mathbf{1}^T \frac{1}{\sqrt{l}}\mathbf{1}\Big(\frac{1}{4p(1-p)l + \frac{1}{P_{tot}}}\Big)\frac{1}{\sqrt{l}}\mathbf{1}^T 2(1-2p)\mathbf{1} \\ &= \frac{(1-2p)^2}{p(1-p) + \frac{1}{4lP_{tot}}}.\end{aligned} \tag{33}$$

In the channel noise dominated regime (i.e., $p \approx 0$ or a small P_{tot}), the repetition code increases the minimum distance by l-fold

$$\Delta_1^2 = 4(1-2p)^2 l P_{tot}. \tag{34}$$

But, when the measurement noise dominates (i.e., $P_{tot}(k) \to \infty$), increasing l does not bring in much gain

$$\Delta^2 \to \Delta_2^2 = \frac{(1-2p)^2}{p(1-p)}. \tag{35}$$

In the more general case, a linear (n, m)-code can be characterized by its $n \times m$ *generation matrix* $\mathbf{G}$

$$\mathbf{c} = \mathbf{G} \cdot \mathbf{x} = \mathbf{G} \cdot (\boldsymbol{\theta} \oplus \mathbf{e}) = (\mathbf{G} \cdot \boldsymbol{\theta}) \oplus (\mathbf{G} \cdot \mathbf{e}) \tag{36}$$

where $\cdot$ denotes binary multiplication. The codeword $\mathbf{c} \in \{0,1\}^n$ is then BPSK-modulated into the (real-valued) transmit signal as

$$\mathbf{y} = 2\mathbf{c} - 1. \tag{37}$$

It is seen in (36) that the codeword distribution (due to the random $\mathbf{e}$) depends on the value of $\mathbf{G} \cdot \boldsymbol{\theta}$. For two different source parameters $\boldsymbol{\theta} \neq \boldsymbol{\theta}'$, if $\mathbf{G} \cdot \boldsymbol{\theta} = \mathbf{G} \cdot \boldsymbol{\theta}'$, then they both would correspond to the same codeword distribution, thus indistinguishable by the receiver. This simple argument actually shows the following.

Lemma 2: The $n \times m$ generation matrix $\mathbf{G}$ must be full-rank.

Fig. 4–6 compare the $(8, 4)$ extended Hamming code with the $(8, 4)$ repetition code. In Fig. 4, the measurement noise is small and the Hamming code outperforms the repetition code. In fact, it has Δ^2_{min} being 11.63 versus 7.12 for the repetition code. This example also shows that coding across i.i.d. source symbols can lead to better performance. But, as the measurement noise increases, the Hamming code gradually loses out to the repetition code (see Fig. 5 and Fig. 6), which shows "good" codes in the traditional sense may not be so in the inference context. In fact, these codes can "amplify" the measurement noise by generating large perturbation in the encoder output. Although this may help combating channel noise, it also increases the codeword covariance $\mathbf{\Sigma}_{\boldsymbol{\theta}}$, resulting in a worse performance according to (18).

V. CONCLUSION

We have studied the scaling laws of decentralized inference under the framework of identical mapping which exploits the inherent spatial summation provided by multiple-access channels. Our theoretical analysis established the achievability of the genie-aided performance, assuming the availability of *some* transmitter CSI. The beam-forming gain allowed our identical mapping approach to achieve the genie-aided performance with only a fixed total power. Identical mapping, on the other hand, was argued to perform poorly in zero-mean fading channels with no transmitter CSI. In this case, we proposed a new TDMA-based protocol and established its significant performance gain, as compared to the identical mapping approach.

APPENDIX I
PROOF OF THEOREM 1

Clearly, the mean map $\mathbf{d}_{\boldsymbol{\theta}}$ must be one-to-one in order to avoid the ambiguity of inverting. We divide the proof into the estimation and the detection part.

A. Estimation– Continuous Θ

To simplify discussion (without much loss in generality), we assume Θ to be an open and bounded subset of $\mathbb{R}$. Consequently, the parameter space Θ^m can be seen as a m-manifold whose image under the mean map, denoted by $\mathcal{M}$, is a diffeomorphic submanifold embedded in $\mathbb{R}^n$. The decision rule precisely corresponds to the canonical projection which finds the closest point (in Euclidean distance) on $\mathcal{M}$ to the noisy estimate $\hat{\mathbf{z}}$. The Tubular Neighborhood Theorem in differential topology ensures the existence of an open neighborhood of $\mathcal{M}$ within which the projection point is unique [22].

Given a fixed $\boldsymbol{\theta}$, find an open ball of radius-r centered around the point $\mathbf{d}_{\boldsymbol{\theta}} \in \mathcal{M}$ and contained in the tubular neighborhood of $\mathcal{M}$ in $\mathbb{R}^n$. If $\hat{\mathbf{z}}$ lies within the ball, it has a unique projection onto $\mathcal{M}$, which reflects a perturbation $\Delta\boldsymbol{\theta}$ under the mean map. Choosing r sufficiently small, one has

$$\Delta\mathbf{d}_{\boldsymbol{\theta}} = \hat{\mathbf{z}} - \mathbf{d}_{\boldsymbol{\theta}} = \partial_{\boldsymbol{\theta}}\mathbf{d}_{\boldsymbol{\theta}} \cdot \Delta\boldsymbol{\theta}, \tag{38}$$

which implies

$$\begin{aligned} \mathrm{MSE}_{\mathbf{d}_{\boldsymbol{\theta}}} &= \mathbb{E}_{\boldsymbol{\theta}}\, \Delta\mathbf{d}_{\boldsymbol{\theta}}(\Delta\mathbf{d}_{\boldsymbol{\theta}})^T = \partial_{\boldsymbol{\theta}}\mathbf{d}_{\boldsymbol{\theta}} \cdot \mathbb{E}_{\boldsymbol{\theta}}\, \Delta\boldsymbol{\theta}\Delta\boldsymbol{\theta}^T \cdot (\partial_{\boldsymbol{\theta}}\mathbf{d}_{\boldsymbol{\theta}})^T \\ &= (\partial_{\boldsymbol{\theta}}\mathbf{d}_{\boldsymbol{\theta}})\mathrm{MSE}_{\boldsymbol{\theta}}(\partial_{\boldsymbol{\theta}}\mathbf{d}_{\boldsymbol{\theta}})^T. \end{aligned} \tag{39}$$

Thus, in order to translate the MSE scaling in $\mathbf{d}_{\boldsymbol{\theta}}$ back to the corresponding scaling in $\boldsymbol{\theta}$, the $n \times m$ derivative matrix $\partial_{\boldsymbol{\theta}}\mathbf{d}_{\boldsymbol{\theta}}$ must be **left** invertible, or equivalently, having full rank m. In this case, one has

$$\mathrm{MSE}_{\boldsymbol{\theta}} = (\partial_{\boldsymbol{\theta}}\mathbf{d}_{\boldsymbol{\theta}})^{\dagger}\mathrm{MSE}_{\mathbf{d}_{\boldsymbol{\theta}}}\big((\partial_{\boldsymbol{\theta}}\mathbf{d}_{\boldsymbol{\theta}})^{\dagger}\big)^T \tag{40}$$

where the pseudo-inverse is given by

$$(\partial_{\boldsymbol{\theta}}\mathbf{d}_{\boldsymbol{\theta}})^{\dagger} = \big((\partial_{\boldsymbol{\theta}}\mathbf{d}_{\boldsymbol{\theta}})^T(\partial_{\boldsymbol{\theta}}\mathbf{d}_{\boldsymbol{\theta}})\big)^{-1}(\partial_{\boldsymbol{\theta}}\mathbf{d}_{\boldsymbol{\theta}})^T. \tag{41}$$

However, the value of $\hat{\mathbf{z}}$ may escape the ball due to the random perturbation. The probability of such an "undesirable" event can be upper-bounded using the Chebyshov's inequality

$$\Pr_{\boldsymbol{\theta}}(\|\hat{\mathbf{z}} - \mathbf{d}_{\boldsymbol{\theta}}\|_2 \geq r) \leq \frac{\mathbb{E}_{\boldsymbol{\theta}}\,\|\hat{\mathbf{z}} - \mathbf{d}_{\boldsymbol{\theta}}\|_2^2}{r^2} \sim \frac{1}{k}, \tag{42}$$

which, together with the boundedness of Θ, proves the $1/k$ scaling in $\mathrm{MSE}_{\boldsymbol{\theta}}$:

$$\begin{aligned} \mathrm{MSE}_{\boldsymbol{\theta}} &\leq (\partial_{\boldsymbol{\theta}}\mathbf{d}_{\boldsymbol{\theta}})^{\dagger}\mathrm{MSE}_{\mathbf{d}_{\boldsymbol{\theta}}}\big((\partial_{\boldsymbol{\theta}}\mathbf{d}_{\boldsymbol{\theta}})^{\dagger}\big)^T \\ &\qquad + c \cdot P_{\boldsymbol{\theta}}(\|\hat{\mathbf{d}}_{\boldsymbol{\theta}} - \mathbf{d}_{\boldsymbol{\theta}}\|_2 \geq r) \\ &\sim \frac{1}{k}. \end{aligned} \tag{43}$$

B. Detection– Finite Discrete Θ

It is very difficult to directly analyze the Voronoi regions associated with the decision rule in (9). We instead pursue an alternative bounding technique.

Let r be the half of the minimum distance among any two different decision points

$$r = \frac{1}{2}\min_{\boldsymbol{\theta},\boldsymbol{\theta}'} \|\mathbf{d}_{\boldsymbol{\theta}} - \mathbf{d}_{\boldsymbol{\theta}'}\|_2. \tag{44}$$

Thus, the radius-r balls around each $\mathbf{d}_{\boldsymbol{\theta}}$ are disjoint. When an error event occurs, $\hat{\mathbf{z}}$ must escape the ball around the true value $\mathbf{d}_{\boldsymbol{\theta}}$. Therefore

$$\begin{aligned} P_{e,\boldsymbol{\theta}} &\leq P_{\boldsymbol{\theta}}(\|\hat{\mathbf{z}} - \mathbf{d}_{\boldsymbol{\theta}}\|_2 > r) \\ &= P_{\boldsymbol{\theta}}(\|\frac{1}{k}\sum_{i=1}^{k}(\mathbf{A}_i\mathbf{y}_i - \mu_a\mathbf{d}_{\boldsymbol{\theta}}) + \frac{1}{\sqrt{kP_{tot}}}\mathbf{w}\|_2 > \mu_a r) \\ &\leq \sum_{j=1}^{n} P_{\boldsymbol{\theta}}(|\frac{1}{k}\sum_{i=1}^{k}[a_{i,j}y_{i,j} - \mu_a d_{\boldsymbol{\theta},j}] + \frac{1}{\sqrt{kP_{tot}}}w_j| > \frac{\mu_a r}{\sqrt{n}}). \end{aligned} \tag{45}$$

We write the Gaussian channel noise $w_j = \frac{1}{\sqrt{k}}\sum_{i=1}^{k} w_{i,j}$ as a Gaussian sum and let $V_i = a_{i,j}y_{i,j} - \mu_a d_{\boldsymbol{\theta},j}$. The goal is to establish the exponential decay of a sum of i.i.d. random variables, that is,

$$P_{\boldsymbol{\theta}}(|\sum_{i=1}^{k}[V_i + \frac{1}{\sqrt{P_{tot}(k)}}w_{i,j}]| > kc) \sim e^{-k} \tag{46}$$

where $c = \frac{\mu_a r}{\sqrt{n}} > 0$ is a constant.

We proceed by using the Chernoff bounding technique. For $s > 0$ one has

$$\begin{aligned} P = P_{\boldsymbol{\theta}}(\sum_{i=1}^{k}[V_i + \frac{1}{\sqrt{P_{tot}(k)}}w_{i,j}] > kc) \\ &\leq \mathbb{E}\, e^{s\left[\sum_{i=1}^{k}(V_i + \frac{1}{\sqrt{P_{tot}}}w_{i,j}) - kc\right]} \\ &= \left(\mathbb{E}\, e^{sV_i}\right)^k \cdot e^{\frac{k}{2P_{tot}(k)}s^2} \cdot e^{-skc} \\ &= \exp\{-k[sc - \phi(s) - \frac{1}{2P_{tot}(k)}s^2]\} \end{aligned} \tag{47}$$

where $\phi(s) = \log(\mathbb{E}\, e^{sV_i})$ is the (log) moment generating function of V_i. Let $\psi(s) = sc - \phi(s) - \frac{1}{2P_{tot}(k)}s^2$. The exponential decay in P can be obtained if the following holds

$$\exists s_0 > 0, \quad \text{such that}, \quad \psi(s_0) > 0. \tag{48}$$

In fact, the above condition is satisfied by variety of distributions. For examples, bounded random variables as well as Gaussian variables have $\phi(s) \leq Ls^2$ for some $L > 0$, which implies that $\psi(s) \geq sc - (L + \frac{1}{2P_{tot}(k)})s^2$. Choosing $s_0 = c/(2L + \frac{1}{P_{tot}(k)})$, one has for all k

$$\psi(s_0) \geq \frac{c^2}{4L + \frac{2}{P_{tot}(k)}} \geq \frac{c^2}{4L + \frac{2}{P_{tot}(1)}} > 0 \tag{49}$$

where $P_{tot}(k) \geq P_{tot}(1)$ by monotonicity of the total power.

A similar argument applies to the other direction, that is,

$$P' = P_{\boldsymbol{\theta}}(\sum_{i=1}^{k}[V_i + \frac{1}{\sqrt{P_{tot}(k)}}w_{i,j}] < -kc), \tag{50}$$

except that $\psi(s)$ changes to $sc - \phi(-s) - \frac{1}{2P_{tot}(k)}s^2$.

APPENDIX II
PROOF OF LEMMA 1

Let $f_{\boldsymbol{\theta},k}$ denote the density function of $\tilde{\mathbf{z}}$ in (6). One has

$$f_{\boldsymbol{\theta},k}(\mathbf{u}) = (h_{\boldsymbol{\theta},k} * h_k)(\mathbf{u} - \sqrt{k}\mu_a \mathbf{d}_{\boldsymbol{\theta}}) \tag{51}$$

where $h_{\boldsymbol{\theta},k}$ is the density function of $\frac{1}{\sqrt{k}}\sum_{i=1}^{k}(\mathbf{A}_i\mathbf{y}_i - \mu_a\mathbf{d}_{\boldsymbol{\theta}})$, and h_k denotes the Gaussian density function of $N(\mathbf{0}, \frac{1}{P_{tot}(k)}\mathbf{I}_n)$. Let $\psi(\mathbf{t})$ be the characteristic function of $\mathbf{Ay} - \mu_a\mathbf{d}_{\boldsymbol{\theta}}$. The characteristic function associated with $h_{\boldsymbol{\theta},k}$ is given by $\psi_k(\mathbf{t}) = \psi(\mathbf{t}/\sqrt{k})^k$. Since $\int |\psi_k(\mathbf{t})| d\mathbf{t} < \infty$ by (11), $h_{\boldsymbol{\theta},k}$ is given by the inverse Fourier transform of its corresponding characteristic function [16]

$$h_{\boldsymbol{\theta},k}(\mathbf{u}) = \frac{1}{2\pi}\int e^{-i\mathbf{t}\cdot\mathbf{u}}\psi_k(\mathbf{t})d\mathbf{t}. \tag{52}$$

It follows from the Central Limit Theorem that the characteristic function ψ_k converges to that of a Gaussian distribution. Given (52) and the finiteness condition in (11) ($s = 0$), the Dominated Convergence Theorem implies that

$$h_{\boldsymbol{\theta},k} \to h_{\boldsymbol{\theta}} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\theta}}). \tag{53}$$

Similarly, one can show the convergence in the derivatives of $h_{\boldsymbol{\theta},k}$ using the finiteness condition in (11) ($s = 1$).

We now begin the FIM calculation.

$$\begin{aligned} \mathbf{J}_k &= -\,\mathbb{E}_{\tilde{\mathbf{z}}}\,\partial^2_{\boldsymbol{\theta}\boldsymbol{\theta}} \ln f_{\boldsymbol{\theta},k}(\tilde{\mathbf{z}}) = -\int \partial_{\boldsymbol{\theta}}\Big(\frac{\partial_{\boldsymbol{\theta}} f_{\boldsymbol{\theta},k}}{f_{\boldsymbol{\theta},k}}\Big) f_{\boldsymbol{\theta},k} \\ &= -\int \frac{(\partial^2_{\boldsymbol{\theta}\boldsymbol{\theta}} f_{\boldsymbol{\theta},k}) f_{\boldsymbol{\theta},k} - (\partial_{\boldsymbol{\theta}} f_{\partial,k})^T(\partial_{\boldsymbol{\theta}} f_{\boldsymbol{\theta},k})}{f_{\boldsymbol{\theta},k}} \\ &= \int \frac{(\partial_{\boldsymbol{\theta}} f_{\boldsymbol{\theta},k})^T(\partial_{\boldsymbol{\theta}} f_{\boldsymbol{\theta},k})}{f_{\boldsymbol{\theta},k}}. \end{aligned} \tag{54}$$

Using the chain rule, one has

$$\begin{aligned}\partial_{\boldsymbol\theta} f_{\boldsymbol\theta,k}(\mathbf{u}) &= \partial_{\boldsymbol\theta}\int h_{\boldsymbol\theta,k}(\mathbf{u}-\sqrt{k}\mu_a\mathbf{d}_{\boldsymbol\theta}-\mathbf{v})h_k(\mathbf{v})d\mathbf{v}\\ &= \int\Big[\partial_{\boldsymbol\theta} h_{\boldsymbol\theta,k}(\mathbf{u}-\sqrt{k}\mu_a\mathbf{d}_{\boldsymbol\theta}-\mathbf{v})-\\ &\quad \sqrt{k}\partial_{\mathbf{u}} h_{\boldsymbol\theta,k}(\mathbf{u}-\sqrt{k}\mu_a\mathbf{d}_{\boldsymbol\theta}-\mathbf{v})\mu_a\partial_{\boldsymbol\theta}\mathbf{d}_{\boldsymbol\theta}\Big]h_k(\mathbf{v})d\mathbf{v}\\ &= (g_{\boldsymbol\theta,k}*h_k)(\mathbf{u}-\sqrt{k}\mu_a\mathbf{d}_{\boldsymbol\theta})\end{aligned} \tag{55}$$

where the shorthand notation $g_{\boldsymbol\theta,k}(\mathbf{u}) = \partial_{\boldsymbol\theta} h_{\boldsymbol\theta,k}(\mathbf{u})-\sqrt{k}\partial_{\mathbf{u}} h_{\boldsymbol\theta,k}(\mathbf{u})\mu_a\partial_{\boldsymbol\theta}\mathbf{d}_{\boldsymbol\theta}$. Remove the translation $\mathbf{u}\to(\mathbf{u}-\sqrt{k}\mu_a\mathbf{d}_{\boldsymbol\theta})$ to get

$$\mathbf{J}_k = \int\frac{(g_{\boldsymbol\theta,k}*h_k)^T\cdot(g_{\boldsymbol\theta,k}*h_k)}{h_{\boldsymbol\theta,k}*h_k}. \tag{56}$$

Taking the limit as $k\to\infty$ and applying Fatou's Lemma [23],[6] one has

$$\begin{aligned}&\liminf_{k\to\infty}\operatorname{tr}\frac{1}{k}\mathbf{J}_k\\ &= \liminf_{k\to\infty}\operatorname{tr}\int\frac{\left(\frac{g_{\boldsymbol\theta,k}*h_k}{\sqrt{k}}\right)^T\left(\frac{g_{\boldsymbol\theta,k}*h_k}{\sqrt{k}}\right)}{h_{\boldsymbol\theta,k}*h_k}\\ &\geq \operatorname{tr}\int\liminf_{k\to\infty}\frac{\left(\frac{g_{\boldsymbol\theta,k}*h_k}{\sqrt{k}}\right)^T\left(\frac{g_{\boldsymbol\theta,k}*h_k}{\sqrt{k}}\right)}{h_{\boldsymbol\theta,k}*h_k}\\ &= \mu_a^2\int\frac{(\partial_{\boldsymbol\theta}\mathbf{d}_{\boldsymbol\theta})^T[(\partial_{\mathbf{u}} h_{\boldsymbol\theta}*h_\infty)^T(\partial_{\mathbf{u}} h_{\boldsymbol\theta}*h_\infty)]\partial_{\boldsymbol\theta}\mathbf{d}_{\boldsymbol\theta}}{h_{\boldsymbol\theta}*h_\infty}\end{aligned} \tag{57}$$

which establishes (12).

When $P_{tot}(k)\to\infty$, the limiting h_∞ is the identity operator (or the Dirac-δ) with respect to convolution. Since

$$\begin{aligned}\partial_{\mathbf{u}} h_{\boldsymbol\theta}(\mathbf{u}) &= \partial_{\mathbf{u}}\Big(\frac{1}{\sqrt{\det(2\pi\boldsymbol\Sigma_{\boldsymbol\theta})}}\exp\{-\frac{1}{2}\mathbf{u}^T\boldsymbol\Sigma_{\boldsymbol\theta}^{-1}\mathbf{u}\}\Big)\\ &= -h_{\boldsymbol\theta}(\mathbf{u})\mathbf{u}^T\boldsymbol\Sigma_{\boldsymbol\theta}^{-1},\end{aligned} \tag{58}$$

the result in (13) follows immediately from

$$\int\frac{(\partial_{\mathbf{u}} h_{\boldsymbol\theta})^T\partial_{\mathbf{u}} h_{\boldsymbol\theta}}{h_{\boldsymbol\theta}} = \int h_{\boldsymbol\theta}(\mathbf{u})\boldsymbol\Sigma_{\boldsymbol\theta}^{-1}\mathbf{u}\mathbf{u}^T\boldsymbol\Sigma_{\boldsymbol\theta}^{-1}d\mathbf{u} = \boldsymbol\Sigma_{\boldsymbol\theta}^{-1}. \tag{59}$$

[6] $\liminf\int f_k \geq \int\liminf f_k$.

APPENDIX III
PROOF OF THEOREM 2

The proof of 1) is essentially a slight adaptation of Theorem 1 to the beam-forming channel a_i'. One only needs to check $\mathbb{E}\, a_i' = \mathbb{E}\,|a| \neq 0$ given the beam-forming phase rotation $a^*/|a|$ applied at the transmitter.

We now focus on receiver CSI only. In this case, the knowledge on channel coefficients is incorporated in the performance measure. For example, $\text{MSE}_{\boldsymbol{\theta}} = \mathbb{E}_a[\text{MSE}_{\boldsymbol{\theta}}|a]$. In the following, we characterize such a conditional measure.

To simplify the presentation we consider scalar signals, that is, setting $n = m = 1$,

$$\tilde{z} = \frac{1}{\sqrt{k}} \sum_{i=1}^{k} a_i(y_i - d_{\boldsymbol{\theta}}) + \frac{1}{\sqrt{k}} d_{\boldsymbol{\theta}} \sum_{i=1}^{k} a_i + \frac{1}{P_{tot}} w \tag{60}$$

where the zero-mean term $y_i - d_{\boldsymbol{\theta}}$ has variance $\sigma_{\boldsymbol{\theta}}^2$.

The desired lower bounds in the theorem can be deduced from, for sufficiently large k,

$$\frac{1}{\sqrt{k}} d_{\boldsymbol{\theta}} \sum_{i=1}^{k} a_i < d_{\boldsymbol{\theta}} \sqrt{\log \log k}, \tag{61}$$

which is a direct application of the following result related to the Brownian motion and laws of the iterated logarithm [16].

Theorem 3 (Theorem 9.7 in [16]): If $X_1, X_2, \ldots$ are i.i.d. with $\mathbb{E}\, X_i = 0$ and $\mathbb{E}\, X_i^2 = 1$ then

$$\limsup_{k \to \infty} \frac{\sum_{i=1}^{k} X_i}{\sqrt{2k \log \log k}} = 1 \quad \text{a.s.} \tag{62}$$

The rest of proof follows the same line as Theorem 1, except one technicality— we need to show that[7]

$$\frac{1}{\sqrt{k}} \sum_{i=1}^{k} a_i(y_i - d_{\boldsymbol{\theta}}) \to N(0, \sigma_{\boldsymbol{\theta}}^2). \tag{63}$$

For such, we rely on the Lindeberg-Feller central limit theorem on triangular arrays.

Theorem 4 (see Theorem 4.5 in [16]): For each k, let $X_{k,i}$, $1 \leq i \leq k$, be independent random variables with $\mathbb{E}\, X_{k,i} = 0$. Denote by $\mathbb{E}[X; A]$ the expectation over subset, i.e., $\int_A X dP$. Suppose

$$\sum_{i=1}^{k} \mathbb{E}\, X_{k,i}^2 \to \sigma^2 > 0, \tag{64}$$

[7] Because of the conditional nature of the problem the a_i's are treated as deterministic values.

$$\lim_{k\to\infty}\sum_{i=1}^{k}\mathbb{E}[|X_{k,i}|^2; |X_{k,i}| > \epsilon] = 0 \quad \forall\epsilon > 0. \tag{65}$$

Then $\sum_{i=1}^{k} X_{k,i} \to N(0, \sigma^2)$ as $k \to \infty$.

Setting $X_{k,i} = \frac{1}{\sqrt{k}} a_i(y_i - d_{\boldsymbol{\theta}})$, we check the two conditions in the Lindeberg-Feller array as follows. By the strong law of large numbers one has

$$\sum_{i=1}^{k}\mathbb{E}_{\boldsymbol{\theta}}\Big[\frac{a_i}{\sqrt{k}}(y_i - d_{\boldsymbol{\theta}})\Big]^2 = \sigma^2_{\boldsymbol{\theta}}\frac{\sum_{i=1}^{k} a_i^2}{k} \to \sigma^2_{\boldsymbol{\theta}} \quad \text{a.s.} \tag{66}$$

Moreover,

$$\begin{aligned}
&\sum_{i=1}^{k}\mathbb{E}_{\boldsymbol{\theta}}\Big[[\frac{a_i}{\sqrt{k}}(y_i - d_{\boldsymbol{\theta}})]^2; |\frac{a_i}{\sqrt{k}}(y_i - d_{\boldsymbol{\theta}})| > \epsilon\Big] \\
&= \sum_{i=1}^{k}\frac{a_i^2}{k}\,\mathbb{E}_{\boldsymbol{\theta}}[(y_i - d_{\boldsymbol{\theta}})^2; |y_i - d_{\boldsymbol{\theta}}| > \frac{\epsilon\sqrt{k}}{|a_i|}] \\
&\le \sum_{i=1}^{k}\frac{a_i^2}{k}\,\mathbb{E}_{\boldsymbol{\theta}}[(y_i - d_{\boldsymbol{\theta}})^2; |y_i - d_{\boldsymbol{\theta}}| > \frac{\epsilon\sqrt{k}}{\max_i |a_i|}] \\
&= \frac{\sum_{i=1}^{k} a_i^2}{k}\,\mathbb{E}_{\boldsymbol{\theta}}[(y - d_{\boldsymbol{\theta}})^2; |y - d_{\boldsymbol{\theta}}| > \frac{\epsilon\sqrt{k}}{\max_i |a_i|}] \\
&\to 0
\end{aligned} \tag{67}$$

provided that $\limsup_{k\to\infty}\max_i |a_i|/\sqrt{k} = 0$ a.s.

APPENDIX IV

A TDMA-BASED NON-IDENTICAL MAPPING SCHEME

Assume a quantization size of Q bits for every source observation X_{ij}. We choose a *nominal* bit-budget B for each node

$$B = \frac{m}{k}\log P_{tot}(k), \tag{68}$$

which can accommodate $D = B/Q$ of source observations.

The encoding operates on a block of m observation symbols, where m is appropriately chosen with respect to k such that D contains at least one symbol. The total transmission time m is equally divided among the k nodes, and a node uses full power $P_{tot}(k)$ when it transmits. Although each node can only send the (quantized) data measurement corresponding to D hidden source parameters, they can collaborate with each other to cover all the m source parameters. One simple scheme is as follows (choosing $m = k$). Node-1 covers source parameters 1 to D, node-2 covers 2 to $D+1$, and so on. In other words, each node sends a length-D block of data in a cyclic fashion so that a total of r observation data are transmitted for every source parameter

$$r = \frac{kD}{m} = \frac{\log P_{tot}}{Q}. \tag{69}$$

However, not all transmission would be successful due to channel outage. Let γ be the probability of a successful transmission, which occurs when the instantaneous channel capacity is larger than the bit budget

$$\gamma = P(B \leq \frac{m}{k} \log(|a|^2 P_{tot})) = P(|a|^2 \geq 1). \tag{70}$$

Thus, on average, a γr number of data measurements can be reliably received for each hidden parameter. This implies a $1/\gamma r \sim 1/\log P_{tot}(k)$ decay in estimation MSE, which translates to $1/\log k$ scaling when $P_{tot} \sim k$.

REFERENCES

[1] T. S. Han and S. Amari, "Statistical inference under multiterminal data compression," *IEEE Trans. Inform. Theory*, vol. 44, no. 6, pp. 2300–2324, Oct. 1998.

[2] T. Berger, Z. Zhang, and H. Viswanathan, "The CEO problem," *IEEE Trans. Inform. Theory*, vol. IT-42, pp. 887–902, May 1996.

[3] H. Viswanathan and T. Berger, "The quadratic Gaussian CEO problem," *IEEE Trans. Inform. Theory*, vol. 43, no. 5, pp. 1549–1559, Sept. 1997.

[4] Y. Oohama, "The rate-distortion function for the quadratic Gaussian CEO problem," *IEEE Trans. Inform. Theory*, vol. 44, no. 3, pp. 1057–1070, May 1998.

[5] T. M. Cover, A. A. E. Gamal, and M. Salehi, "Multiple access channels with arbitrarily correlated sources," *IEEE Trans. Inform. Theory*, vol. IT-26, pp. 648–657, Nov. 1980.

[6] M. Gastpar and M. Vetterli, "Scaling laws for homogeneous sensor networks," in *the 41st Annual Allerton Conference on Communication, Control, and Computing*, Monticello, IL, USA, Oct. 2003.

[7] M. Gastpar, B. Rimoldi, and M. Vetterli, "To code, or not to code: Lossy source-channel communication revisited," *IEEE Trans. Inform. Theory*, vol. 49, no. 5, pp. 1147–1158, May 2003.

[8] M. Gastpar and M. Vetterli, "Power, spatio-temporal bandwidth, and distortion in large sensor networks," *IEEE J. Select. Areas Commun.*, vol. 23, no. 4, pp. 745–754, Apr. 2005.

[9] J.-F. Chamberland and V. V. Veeravalli, "Asymptotic results for decentralized detection in power constrained wireless sensor networks," *IEEE J. Select. Areas Commun.*, vol. 22, no. 6, pp. 1007–1015, Aug. 2004.

[10] K. Liu and A. M. Sayeed, "Optimal detection strategies for wireless sensor networks," in *the 42nd Annual Allerton Conference on Communication, Control, and Computing*, Monticello, IL, USA, Oct. 2004.

[11] G. Mergen and L. Tong, "Estimation over multiaccess channels," submitted to IEEE Trans. Signal Processing, July 2004.

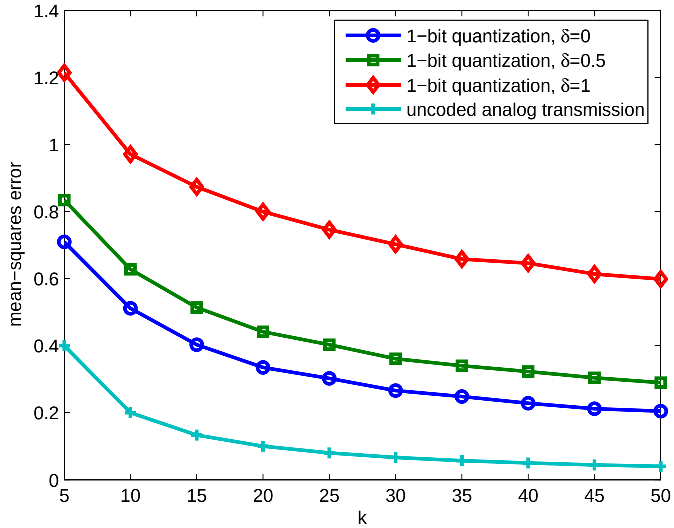

Fig. 2. Estimation performance of one-bit quantizers. Static MAC and $P_{tot} = 10dB$.

[12] G. Mergen, V. Naware, and L. Tong, "Asymptotic detection performance of type-based multiple access over multiaccess fading channels," submitted to IEEE Trans. Signal Processing, 2004.

[13] L. L. Scharf, *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis*. New York: Addison-Wesley, 1990.

[14] H. V. Poor, *An Introduction to Signal Detection and Estimation*, 2nd ed. New York: Springer-Verlag, 1994.

[15] J. Shao, *Mathematical Statistics*, 2nd ed. New York: Springer, 2003.

[16] R. Durrett, *Probability*, 2nd ed. Belmont, CA: Wadsworth, 1996.

[17] I. Csiszár, "The method of types," *IEEE Trans. Inform. Theory*, vol. 44, no. 6, pp. 2505–2523, Oct. 1998.

[18] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

[19] K. Liu and A. M. Sayeed, "Asymptotically optimal decentralized type-based detection in wireless sensor networks," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'04)*, vol. 3, Montreal, Quebec, Canada, May 2004, pp. 873–876.

[20] K. Liu and A. Sayeed, "On asymptotic optimality of decentralized detection in sensor networks," University of Wisconsin-Madison, Madison, WI, USA, Technical Report ECE-04-2, Feb. 2004.

[21] A. Giridhar and P. R. Kumar, "Computing and communicating functions over sensor networks," *IEEE J. Select. Areas Commun.*, vol. 23, no. 4, pp. 755–764, Apr. 2005.

[22] V. Guillemin and A. Pollack, *Differential Topology*. New Jersey: Prentice-Hall, 1974.

[23] G. B. Folland, *Real Analysis: Modern Techiques and Their Applications*, 2nd ed. New York: Wiley, 1999.

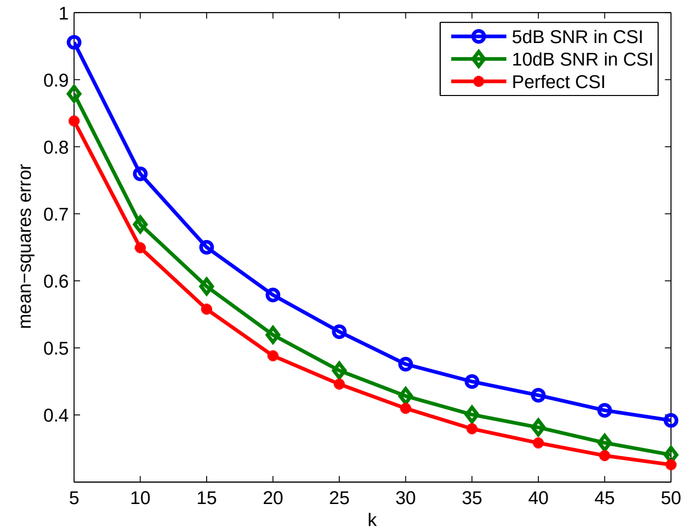

Fig. 3. Performance of one-bit quantizer ($\delta = 0$) under noisy CSI feedback. Gaussian fading MAC and $P_{tot} = 10dB$.

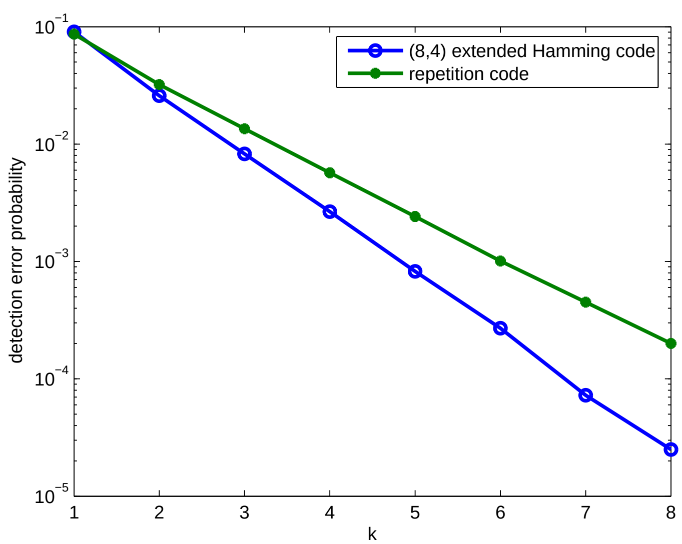

Fig. 4. Performance comparison between the $(8, 4)$ extended Hamming code and the repetition code. The Bernoulli measurement noise has $p = 0.01$ and the total transmit power $P_{tot} = 0dB$.

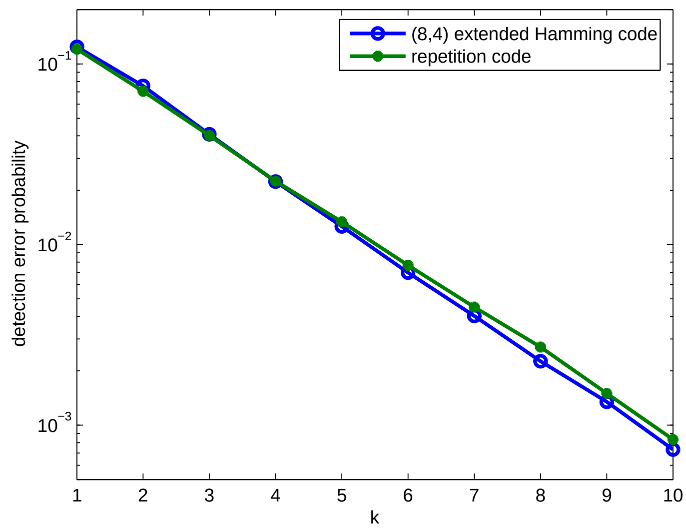

Fig. 5. Performance comparison between the $(8, 4)$ extended Hamming code and the repetition code. The Bernoulli measurement noise has $p = 0.05$ and the total transmit power $P_{tot} = 0dB$.

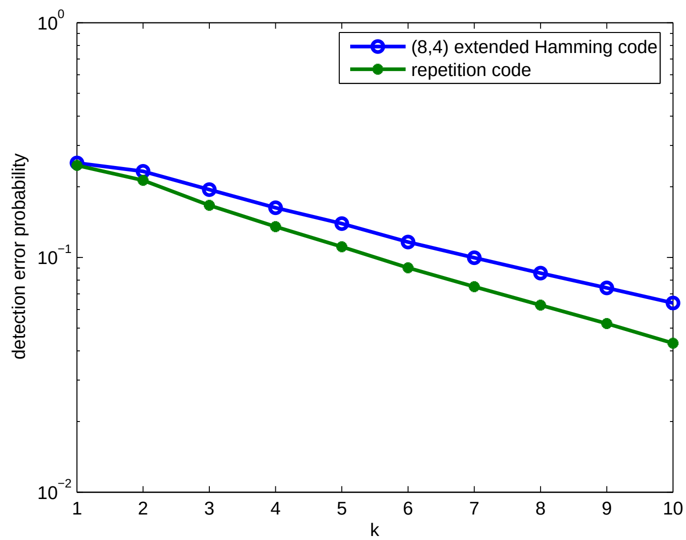

Fig. 6. Performance comparison between the $(8, 4)$ extended Hamming code and the repetition code. The Bernoulli measurement noise has $p = 0.2$ and the total transmit power $P_{tot} = 0dB$.