The Sensing Capacity of Sensor Networks

Yaron Rachlin, *Member, IEEE*, Rohit Negi, *Member, IEEE*, and Pradeep K. Khosla, *Fellow, IEEE*

***Abstract*—This paper demonstrates fundamental limits of sensor networks for detection problems where the number of hypotheses is exponentially large. Such problems characterize many important applications including detection and classification of targets in a geographical area using a network of seismic sensors, and detecting complex substances with a chemical sensor array. We refer to such applications as large-scale detection problems. Using the insight that these problems share fundamental similarities with the problem of communicating over a noisy channel, we define the "sensing capacity" and lower bound it for a number of sensor network models. The sensing capacity expression differs significantly from the channel capacity due to the fact that for a fixed sensor configuration, codewords are dependent and nonidentically distributed. The sensing capacity provides a bound on the minimal number of sensors required to detect the state of an environment to within a desired accuracy. The results differ significantly from classical detection theory, and provide an intriguing connection between sensor networks and communications. In addition, we discuss the insight that sensing capacity provides for the problem of sensor selection.**

***Index Terms*—Detection theory, sensing capacity, sensor networks, sensor selection.**

I. Introduction

In many sensing applications, such as pollution monitoring and border security, the phenomena under observation has a large scale that exceeds the range of any one sensor. As a result, collecting measurements from multiple sensors is essential to the sensing task. Sensor networks are systems that collect information about the state of an environment using multiple sensors. Obtaining information about an environment can be cast as either a 'detection' or an 'estimation' problem. In estimation problems such as the problem of estimating a continuous field to within a desired distortion, the state of the environment is continuous. In detection problems, such as binary hypothesis testing, the state of the environment is represented as a finite set of hypotheses. In this paper we study the problem of 'large-scale detection' where the state of the environment belongs to an exponentially large, structured set of hypotheses. Large-scale detection problems include many applications where a sensor network is deployed to monitor large-scale phenomena. We exploit the structure of large-scale detection problems to demonstrate a fundamental information-theoretic relationship between the number of sensor measurements and ability of a sensor network to detect the state of the environment to within a desired accuracy.

We obtain our results by drawing on an analogy between sensor networks and channel encoders. For a fixed sensor configuration, each state of the environment induces a corresponding set of sensor outputs. This set of sensor outputs can be viewed as a noise-corrupted 'codeword,' which must be 'decoded' in order to detect the state of the environment. Thus, the sensor network acts as a channel encoder. In order to motivate this perspective, we examine the following large-scale detection problems.

Robotic mapping is the first large-scale application we consider [1]. In mapping, robots collect sensor measurements to map an unknown environment for the purpose of navigation. [2] introduced occupancy grids, one of the most popular approaches to this problem. In occupancy grids, the world is modeled as a discrete grid, where each grid location has a value corresponding to the state of the environment. For example, in a binary a grid a '0' can indicate free space while a '1' can indicate an obstacle. A robot traversing an unknown environment collects sensor measurements that encode the state of the environment. For example, a robot using a sonar sensor emits a wide acoustic pulse and measures the time until a reflected pulse is sensed. These readings are ambiguous, since one cannot infer the precise location of the obstacle that caused the reflection from a single sensor reading. In addition, sonar readings are noisy. As a result, multiple sensor measurements must be used to distinguish among an exponentially large number of possible grid states. The sequence of sonar readings can be viewed as a noise-corrupted codeword corresponding to the state of the grid. While robotic mapping systems have been successfully implemented in practice, little can be said about their theoretical performance. Theoretical understanding could shed light on the number of sensor measurements required to map an unknown environment. In addition, theory can provide insight into questions about sensor selection. Is it better to use cheap, low power, wide angle sensors or expensive, high power, narrow angle sensors? A theoretical framework could provide general insight into such sensor selection questions.

Video surveillance is another large scale detection problem. [3] used multi-camera sensor networks to detect and track objects across multiple areas, and [4] uses multiple cameras to localize moving objects in a room. The region under surveillance can be viewed as a three-dimensional grid. For example each grid position can have a binary value, representing motion or lack of motion in that grid position. As in the previous example, the number of states of this grid is exponential in the number of grid blocks. Each camera observes a subset of grid blocks, but introduces ambiguity by reducing a three-dimensional volume to a two-dimensional (2D) image. As a result multiple camera

Manuscript received November 13, 2006; revised September 16, 2009; accepted June 09, 2010. Date of current version February 18, 2011. The material in this paper was presented at the Information Theory Workshop, San Antonio, TX, October 2004.

Y. Rachlin is with the C. S. Draper Laboratory, Cambridge, MA 02139 USA (e-mail: yaron.rachlin@alumni.cmu.edu).

R. Negi and P. K. Khosla are with the Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA 15213 USA (e-mail: negi@ece.cmu.edu; pkk@ece.cmu.edu).

Communicated by L. Tong, Associate Editor for Detection and Estimation.

Digital Object Identifier 10.1109/TIT.2010.2103733

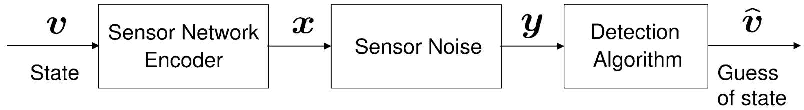

Fig. 1. Sensor network model.

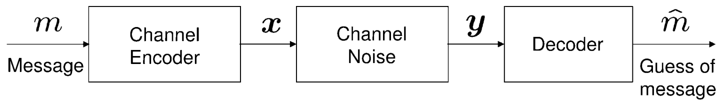

Fig. 2. Communication channel model.

images must be combined to detect the state of the environment. The set of images encode the grid state. While practical systems for surveillance applications are deployed, a theoretical framework for understanding performance limits for such problems is not available.

Identifying a complex chemical substance is a third example of a large-scale detection problem. In this application the output of chemical sensor arrays, consisting of heterogeneous chemical sensors, is used to distinguish among a large number of substances [5]. Each substance can be modeled as a mixture of constituent chemicals at various discrete concentration levels [6], resulting in an exponentially large set of possible states. Each chemical sensor in the array reacts to a subset of chemicals. For example, sensors can output a voltage proportional to a weighted sum of the concentrations of a subset of chemicals. The output of a chemical sensor array encodes the state of the sample being sensed. As in the previous two examples, theory could provide insight into the practical design of such sensor arrays.

Target detection and classification in a geographical area is an important class of applications for sensor networks [7], and a final motivating example of a large-scale detection problem. We consider the problem of detection and classification using seismic sensors, as demonstrated in [7], [8]. The environment can be modeled as a discrete grid, where each position can contain targets of multiple types. The number of target configurations is exponential in the number of grid blocks. Seismic sensors are scattered randomly on this grid, and sense the vibrations of targets over subsets of the grid. The intensity of vibration is dependent on the target's distance from the sensor, and therefore a single sensor cannot distinguish between many targets far away and a single target nearby. The set of seismic sensor outputs encode the location and class of targets in the field.

All of the examples considered above share the following common elements. The state can be modeled as a discrete vector or grid, and the number of states is exponentially large. Sensors output noise-corrupted functions of subsets of the vector or grid. These sensor measurements must be fused to detect the state of the environment. In this paper we analyze the fundamental limits of this process by using the insight that the problem of large-scale detection and the problem of communicating over a noisy channel share essential similarities.

II. Sensor Networks and Communication Channels

The examples described in Section I motivate the sensor network model shown in Fig. 1. A discrete target vector $\boldsymbol{v}$ represents the state of the environment. In this paper, the term 'state' and 'target vector' are used interchangeably. A fixed sensor configuration encodes the state as a vector of noiseless sensor outputs that form the codeword $\boldsymbol{x}$. The observed sensor measurements are written as $\boldsymbol{y}$, a noise-corrupted version of $\boldsymbol{x}$. Finally, a detection algorithm uses $\boldsymbol{y}$ to compute a guess of the state of the environment $\hat{\boldsymbol{v}}$.

The sensor model shown in Fig. 1 is similar to the classical communication channel model shown in Fig. 2. The target vector $\boldsymbol{v}$ corresponds to the message m being sent. The sensor network acts as a channel encoder, producing the codeword $\boldsymbol{x}$. Finally, a detection algorithm acts as a channel decoder on the noise corrupted codeword $\boldsymbol{y}$. Shannon's celebrated Channel capacity results provide limits for the communications channel [9]. Is there such a limit for sensor networks? [10] proposed the idea of a sensing capacity, but the existence of a positive sensing capacity, and therefore the practical value of this idea, remained an open question. Using the definition presented in [10], the sensing capacity is zero, and therefore lacks practical value. The infeasibility of positive sensing capacity using this definition is due to the observation, discussed below, that in sensor networks codewords are dependent. Motivated by this earlier work, we introduced a definition of sensing capacity that allowed for detection to within a tolerable distortion, and demonstrated a model for which this sensing capacity can be strictly positive [11]. This result demonstrates that the sensing capacity plays a role in our sensor network model analogous to the role of channel capacity in a communications channel. However, because the models differ in significant ways, the notions of channel capacity and sensing capacity also differ.

The most important difference between the sensor network model and a communication channel model is at the encoder. In communications, the content of the message and its codeword representation can be decoupled. Further, the channel encoder can implement any mapping between message and codeword. As a result, two highly similar messages can be differentiated with arbitrarily high accuracy. In contrast, a sensor network encoder is highly constrained, since limited range sensors react to phenomena in the environment, and the same sensor configuration encodes all states of the environment. Due to these constraints on the structure of the encoder, similar states are likely to have similar codewords. As a result, two highly similar states of the environment cannot be distinguished with arbitrarily high accuracy in the presence of noise. While similarities between the sensor network model and the channel model motivate the application of insights about communications from information theory, significant differences between the two models require care in applying such insights in order to understand the impact of these differences on the final theoretical results.

Section III provides an overview of the main results presented in this paper, and reviews related work. Section IV presents sensing capacity results for nonspatial (e.g., chemical) sensing applications, while Section V demonstrates sensing capacity results for a sensor network model that accounts for spatial locality in sensor observations. Section VI concludes the paper and discusses future work.

III. Main Results and Related Work

We review the main theoretical results presented in this paper. In Section IV we introduce a simple but useful sensor network model that can be used to model sensing applications such as chemical sensing applications and computer network monitoring. For this model, we define and bound the sensing capacity. The sensing capacity bound differs significantly from the standard channel capacity results, and requires novel arguments to account for the constrained encoding of a sensor network. This is an important observation due to the use of mutual information as a sensor selection heuristic [12]. Our result shows that this is not the correct metric for the large-scale detection applications considered in this paper. Extensions are presented to account for nonbinary target vectors, target sparsity, and heterogeneous sensors. Plotting the sensing capacity bound, we demonstrate interesting tradeoffs. For example, perhaps counter-intuitively, sensors of shorter range can achieve a desired detection accuracy with fewer measurements than sensors of longer range. Finally, we compare our sensing capacity bound to simulated sensor network performance.

In Section V we introduce a sensor network model that accounts for contiguity in a sensor's field of view. Contiguity is an essential aspect of many classes of sensors. For example, cameras observe localized regions and seismic sensors sense vibrations from nearby targets. We demonstrate sensing capacity bounds that account for such sensors by extending results about Markov types [13], and use convex optimization to compute these bounds. The first result in Section V assumes the state of the environment is modeled as a one-dimensional vector. In Section V.D we extend this result to the case where the state of the environment is modeled as a grid. While a one-dimensional vector can model sensor network applications such as border security and traffic monitoring, results about two dimensions significantly increase the type of applications described by our models.

The performance of sensor networks is limited by both sensing resources and nonsensing resources such as communications, computation, and power. One set of results has been obtained by considering the limitations that communications requirements impose on a sensor network. [14] extends the results in [15] to account for the different traffic models that arise in a sensor network. [16] studies network transport capacity for the case of regular sensor networks. [17] studies the impact of computational constraints and power on the communication efficiency of sensor networks. [18] has considered the interaction between transmission rates and power constraints. Another set of results has been obtained by extending results from compression to sensor networks. Distributed source coding [19], [20] provides limits on the compression of separately encoded correlated sources. [21] applies these results to sensor networks. [22] provides an overview of this area of research. This work focuses on compressing correlated sensor observations to reduce the communication bandwidth required. The distributed nature of the compression is the object of analysis in that work.

In contrast to the work mentioned above, we focus directly on the limits of detecting the state of the environment using noisy sensor observations. The notion of sensing capacity characterizes the limits that sensing (e.g., sensor type, range, and noise) imposes on the attainable accuracy of detection. We do not examine the compression of sensor observations, or the resources required to communicate sensor observations to a point in the network. Instead, we focus on the limits of detection accuracy assuming complete availability of noisy sensor observations. Thus, our large-scale detection problem is quite unlike a distributed source coding problem. An easy way to distinguish between the two is to consider the case where the sensor network has infinite communication and computation resources. In that case, the distributed source coding problem becomes irrelevant, since each sensor can communicate its observations in their entirety to a computer, which can then perform centralized compression. However, even in this scenario, there will exist fundamental limits on the accuracy of detection given the available sensing resources.

Our work is most closely related to work on detection and classification in sensor networks. [23] describes a large body of work on distributed detection where the number of hypotheses is small. [24], [25] extend this work to consider a decentralized binary detection problem with noisy communication links to obtain error exponents. [26] analyzes the performance of various classification schemes for classifying a Gaussian source. This is an m-ary problem where the number of hypotheses is small. [27] analyzes the performance suboptimal classification schemes for classifying multiple targets. While the number of hypotheses is exponential in the number of targets, the large-scale detection problem of a large number of targets is not considered. [28] considers the problem of sensor placement for detecting the location of one or few targets in a grid. This problem is most closely related to the large-scale detection problems addressed in this paper. However, due to restrictions on the numbers of targets, the number of hypotheses remains small in comparison to a large-scale detection problem. A coding-based approach was used to bound the minimum number of sensors required for discrimination, and to propose structured sensor configurations. However, sensors were noiseless, and of limited type, and no notion of sensing capacity was considered. In contrast to existing to existing work on detection and classification in sensor networks, we demonstrate fundamental performance limits for large-scale detection problems.

The problem of estimating a continuous field using a sensor network is an active area of research. [29] considers the relationship of transport capacity and the rate distortion function of a continuous random processes. [30] proves limits on the estimation of an inhomogeneous random fields using sensor that collect noisy point samples. Other work on the problem of estimating a continuous random field includes [31]–[34]. [35] considers the estimation of continuous parameters of a set of underlying random processes through a noisy communications channel. The results presented in this paper consider the detection of a discrete state of an environment. We do not consider extensions to environments with a continuous state.

Recent work [36], [37] extended the definition of sensing capacity introduced in [11] and used it to bound the sensing capacity of a fundamentally different model. We describe these differences in the context of our model in Section IV-D.

IV. Sensing Capacity of the Arbitrary Connections Model

In this section we define and analyze the sensing capacity of the arbitrary connections model, a simple but useful model introduced in [11]. The graphical nature of this model is inspired by a general graphical model for sensor networks introduced in [38], the first publication of which the authors are aware that introduced the idea of modeling sensor networks as a graphical model. We denote random variables and functions by upper-case letters, and instantiations or constants by lower-case letters. Bold-font denotes vectors. $\log(\cdot)$ has base-2. Sets are denoted using calligraphic script. $D(P\|Q)$ denotes the Kullback-Leibler distance and $H(P)$ denotes entropy of a random variable with probability distribution P. $H(Q|P)$ is the conditional entropy of a random variable with conditional probability distribution Q given another random variable with probability distribution P.

A. Arbitrary Connections Model

Fig. 3 shows an example of the arbitrary connections model. The state of the environment is modeled as a k-dimensional binary target vector $\boldsymbol{v}$. Each position in the vector may represent the presence of a target in an actual region in space, or may have other interpretations, such as the presence of a specific chemical in a sample. The possible target vectors are denoted $\boldsymbol{v_i}, i \in \{1, \ldots, 2^k\}$. We say that: a certain $\boldsymbol{v}$ has occurred" if that vector represents the true state. We define a sensor network $s(k,n)$ as a graph showing the connections of n sensors to k positions in the target vector. The sensor network has n identical sensors. Sensor ℓ makes c connections to the k spatial positions. A sensor can connect to a spatial position more than once, and therefore each sensor is connected to at most c spatial positions. We refer to such sensors as having a range c. Ideally, each sensor produces a value x that is an *arbitrary function of the targets* which it senses, $x_\ell = \Psi(v_{t_1}, \ldots, v_{t_c})$. $x \in \mathcal{X}$, where $\mathcal{X}$ is finite and defined by the sensing function. Thus, the "ideal output vector" of the sensor network $\boldsymbol{x}$ depends on the sensor connections, and on the target vector $\boldsymbol{v}$ that occurs. We denote the ideal sensor and sensor network outputs when $\boldsymbol{v_i}$ occurs as x_i and $\boldsymbol{x_i}$, respectively. We assume that each sensor output y is corrupted by noise, so that the conditional p.m.f. $P_{Y|X}(y|x)$ determines the output. $y \in \mathcal{Y}$, where $\mathcal{Y}$ is finite and defined by the noise model. Since the sensors are identical, $P_{Y|X}$ is the same for all the sensors. Further, we assume that the noise is independent in the sensors, so that the 'sensor output vector' $\boldsymbol{y}$ relates to the ideal output $\boldsymbol{x}$ as $P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x}) = \prod_{\ell=1}^{n} P_{Y|X}(y_\ell|x_\ell)$. Given the noise corrupted output $\boldsymbol{y}$ of the sensor network, we detect

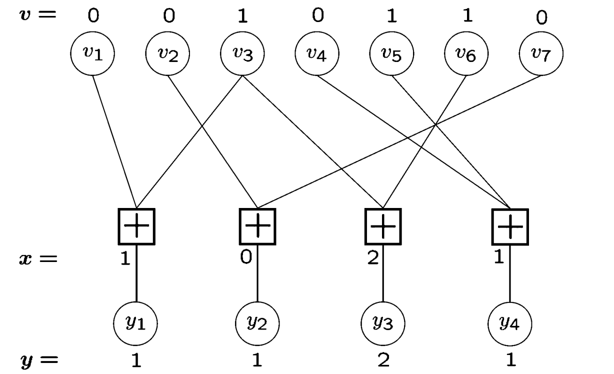

Fig. 3. Arbitrary connections model with $k = 7, n = 4, c = 2$, and a sum sensing function.

the target vector $\boldsymbol{v}$ which occurred by using a detector $g(\boldsymbol{y})$. Because of the constrained encoding of a sensor network, we allow the decoder a distortion of $D \in [0,1]$. Denoting $d_{\mathrm{H}}(\boldsymbol{v_i}, \boldsymbol{v_j})$ as the Hamming distance between two target vectors, the tolerable distortion region of $\boldsymbol{v_i}$ is $\mathcal{D}_{\boldsymbol{v_i}} = \{j : \frac{1}{k} d_{\mathrm{H}}(\boldsymbol{v_i}, \boldsymbol{v_j}) < D\}$. Given that $\boldsymbol{v_i}$ occurred, the detector is in error if $g(\boldsymbol{y}) \notin \mathcal{D}_{\boldsymbol{v_i}}$.

Fig. 3 shows the target vector $\boldsymbol{v} = (0,0,1,0,1,1,0)$ indicating 3 targets among the 7 target positions. In this example, the sensing function Ψ is a sum that indicates the number of positions which contain a target, $x_\ell = \sum_{u=1}^{c} v_{t_u}$, so that $x \in \mathcal{X} = \{0, 1, \ldots, c\}$. Such a function could model a chemical sensor that is sensitive to a subset of chemicals and whose output is linearly proportional to the number of such chemicals present in the sample. More complex, e.g., nonlinear, relationships between chemicals and sensor output require a different choice of Ψ. In the figure, each sensor senses two target positions, and the sensors encode the target vector as $\boldsymbol{x} = (1,0,2,1)$. However, due to noise, the observed vector of sensor outputs is $\boldsymbol{y} = (1,1,2,1)$. The target vector $\boldsymbol{v}' = (0,1,1,0,1,1,0)$, which differs from $\boldsymbol{v}$ in one target position, is encoded as $\boldsymbol{x} = (1,1,2,1)$. As a result a detection algorithm can easily confuse $\boldsymbol{v}'$ for $\boldsymbol{v}$, demonstrating the limitation imposed by the constrained encoding of a sensor network. The Hamming distance between $\boldsymbol{v}$ and $\boldsymbol{v}'$ corresponds to the sum of the false positives and false negatives, and therefore has a natural interpretation in the detection problems we consider.

The arbitrary connections model describes large-scale detection problems that do not have a spatial aspect. Examples of such applications include the detection of complex chemical and computer network monitoring. A group testing problem, such as disease detection in a population where individual samples can be combined, is another such application. In addition to practical utility, this model is easy to analyze and provides useful insights into large-scale detection problems.

The model presented in this section differs from the one used in [38]. Two important differences include restrictions in [38] on the number of sensors that observe each target position, and the use of a factor graph model where the factors are posterior probabilities of targets conditioned on sensors. The arbitrary connections model only models the probabilities of sensor outputs given targets. Our model is similar to the model introduced in [39], though in our model, sensors are limited in the number of targets to which they can connect. This difference, enables our use of the method of types [13] to prove our bounds.

B. Sensing Capacity Definitions

How many sensor measurements must a sensor network collect to detect the a target vector to within a desired distortion? To answer this question we define the idea of a 'sensing capacity.' The probability of error of a sensor network given that target vector $\boldsymbol{v_i}$ occurred is $P_{e,i,s} = \Pr(\text{error}|i,s) = \sum_{\boldsymbol{y}\in\mathcal{Y}^n}\Pr(g(\boldsymbol{y}) \notin \mathcal{D}_{\boldsymbol{v_i}}|\boldsymbol{v_i},s,\boldsymbol{x_i},\boldsymbol{y})P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x_i})$. The expected probability of error for a sensor network is $P_{e,s} = \sum_i P_{e,i,s}P_{\boldsymbol{V}}(\boldsymbol{v_i})$. The rate R of a sensor network is defined as the ratio of target positions being sensed to the number of sensor measurements, $R = \frac{k}{n}$.

Sensing Capacity Definition: The *sensing capacity* of a sensor network, $C(D)$, is defined as the maximum rate $R^\star$ such that below this rate there exists a sequence of sensor networks $s(\lceil nR\rceil, n)$ whose expected probability of error across all target vectors goes to zero with increasing n, that is, $P_{e,s} \to 0$ as $n \to \infty$ at a fixed rate R.

Is $C(D)$ nonzero? One of the main contributions of the theorem presented in this section is to demonstrate that the sensing capacity can be strictly positive for the arbitrary connections model. The proof of the result presented in this section relies on the method of types [13], and the statement of our theorem requires the definition of *types* and *joint types*. The type of target vector $\boldsymbol{v_i}, \boldsymbol{\gamma} = (\gamma_0, \gamma_1)$, is a histogram of the number of 0's and 1's in $\boldsymbol{v_i}$. Here, γ_0 denotes the fraction of zeros in $\boldsymbol{v_i}$ and γ_1 is similarly defined. We consider a sensor that independently (i.e., with replacement), and with uniform probability, connects to c out of k target positions. For a target vector $\boldsymbol{v_i}$ of type $\boldsymbol{\gamma}$, the probability that this sensor's noiseless output is x_i can be written as a function of the target vector's type as follows:

$$P_{X_i}(X_i = x_i) = \sum_{\substack{\boldsymbol{a}\in\{0,1\}^c\\ \Psi(\boldsymbol{a})=x_i}} \prod_{m=1}^{c} \gamma_{a_m} \doteq P^{\boldsymbol{\gamma}}(x). \tag{1}$$

The joint type of two target vectors is $\boldsymbol{\lambda} = (\lambda_{00}, \lambda_{01}, \lambda_{10}, \lambda_{11})$. Here, λ_{01} is the fraction of positions in $\boldsymbol{v_i}, \boldsymbol{v_j}$ where $\boldsymbol{v_i}$ has bit '0' while $\boldsymbol{v_j}$ has bit '1'. Similarly, we define $\lambda_{00}, \lambda_{10}, \lambda_{11}$. Again, we consider a sensor that independently, and with uniform probability, connects to the c out k target positions. The probability that the sensor's noiseless output is x_i for target vector $\boldsymbol{v_i}$, while its output is x_j for target vector $\boldsymbol{v_j}$, can be written in terms of the of joint type of the two target vectors, $\boldsymbol{\lambda}$, as follows:

$$\begin{aligned} P_{X_iX_j}(X_i = x_i, X_j = x_j) &= \sum_{\substack{\boldsymbol{a},\boldsymbol{b}\in\{0,1\}^c\\ \Psi(\boldsymbol{a})=x_i,\Psi(\boldsymbol{b})=x_j}} \prod_{m=1}^{c} \lambda_{a_m b_m} \\ &\doteq P^{\boldsymbol{\lambda}}(x_i, x_j). \end{aligned} \tag{2}$$

We give specific examples of the quantities introduced above by considering randomly generated sensors like the ones shown in Fig. 3, where $c = 2$ and sensors count the number of targets present in the target positions that they sense. Given a target vector, a randomly generated sensor will output '2' only if both of its connections connect to positions with a '1.' For a vector of type $\boldsymbol{\gamma}$, this occurs with probability $(\boldsymbol{\gamma}_1)^2$. Table I describes the complete output p.m.f. $P_{X_i}(x_i)$ for such a randomly generated sensor, given that a vector of type $\boldsymbol{\gamma}$ occurred. Given two target vectors $\boldsymbol{v_i}, \boldsymbol{v_j}$ of joint type $\boldsymbol{\lambda}$, a sensor will output '0' for both target vectors only if both its connections are connected to target positions that have a '0' bit in both these target vectors. This happens with probability $(\lambda_{00})^2$. Table II lists the complete joint p.m.f. $P_{X_iX_j}(x_i, x_j)$ of a randomly generated sensor for two target vectors with a joint type $\boldsymbol{\lambda}$.

We define two probability distributions that we use to state Theorem 1. As in the previous paragraphs, we consider a sensor that randomly connects to c target positions. The first probability distribution is the joint distribution of the ideal sensor output X_i, which is the ideal sensor output when $\boldsymbol{v_i}$ of type $\boldsymbol{\gamma}$ occurs, and Y, which is related to X_i through the sensor noise model

$$P^{\boldsymbol{\gamma}}_{X_iY}(x_i, y) = P^{\boldsymbol{\gamma}}(x_i)P_{Y|X}(y|x_i)$$

The second probability distribution is the joint distribution of the ideal sensor output X_i and Y, where Y is related to a different ideal sensor output X_j using the sensor noise model. $\boldsymbol{v_i}$ and $\boldsymbol{v_j}$ have joint type $\boldsymbol{\lambda}$

$$Q^{\boldsymbol{\lambda}}_{X_iY}(x_i, y) = \sum_{a\in\mathcal{X}} P^{\boldsymbol{\lambda}}(x_i, x_j = a)P_{Y|X}(y|x_j = a).$$

C. Sensing Capacity Lower Bound

Theorem 1: (Sensing Capacity Achievability Theorem for the Arbitrary Connections Model): The sensing capacity at distortion D is bounded as

$$C(D) \geq C_{LB}(D) = \min_{\substack{\boldsymbol{\lambda}\\ \lambda_{01}+\lambda_{10}\geq D\\ \lambda_{00}+\lambda_{01}=\gamma_0\\ \lambda_{10}+\lambda_{11}=\gamma_1}} \frac{D\left(P^{\boldsymbol{\gamma}}_{X_iY}\|Q^{\boldsymbol{\lambda}}_{X_iY}\right)}{H(\boldsymbol{\lambda}) - H(\boldsymbol{\gamma})} \tag{3}$$

where $\boldsymbol{\gamma} = (0.5, 0.5)$ and $\boldsymbol{\lambda} = (\lambda_{00}, \lambda_{01}, \lambda_{10}, \lambda_{11})$ is an arbitrary probability mass functions.

The proof of Theorem 1 broadly follows the proof of channel capacity provided by Gallager [40], by analyzing a union bound of pair-wise error probabilities, averaged over randomly generated sensor networks. However, it differs from [40] in several important ways. In our sensor network model, the codewords are dependent and nonidentically distributed. To prove our bound, we group the exponential number of pair-wise error terms into a polynomial number of terms using the method of types.

Proof: We use a random coding argument to bound the sensing capacity for the arbitrary connections model. Instead of constructing a sequence of sensor network directly, we bound the expected probability of error, where the expectation is over an ensemble of randomly generated sensor networks. The sensor networks are generated as follows. Each sensor connects to c randomly chosen target positions out of the k possible positions. The connections are made independently, and are chosen with replacement. Therefore a sensor can choose the same target position more than once. Using this probabilistic model for sensor

TABLE I
DISTRIBUTION OF X_i IN TERMS OF THE TYPE $\boldsymbol{\gamma}$ OF $\boldsymbol{v_i}$ WHEN $c = 2$

X_i	$X_i = 0$	$X_i = 1$	$X_i = 2$
P_{X_i}	$(\gamma_0)^2$	$2\gamma_0\gamma_1$	$(\gamma_1)^2$

TABLE II
JOINT DISTRIBUTION OF X_j AND X_i IN TERMS OF THE JOINT TYPE $\boldsymbol{\lambda}$ OF $\boldsymbol{v_i}, \boldsymbol{v_j}$ WHEN $c = 2$

$P_{X_iX_j}$	$X_j = 0$	$X_j = 1$	$X_j = 2$
$X_i = 0$	$(\lambda_{00})^2$	$2\lambda_{00}\lambda_{01}$	$(\lambda_{01})^2$
$X_i = 1$	$2\lambda_{00}\lambda_{10}$	$2(\lambda_{10}\lambda_{01} + \lambda_{00}\lambda_{11})$	$2\lambda_{01}\lambda_{11}$
$X_i = 2$	$(\lambda_{10})^2$	$2\lambda_{10}\lambda_{11}$	$(\lambda_{11})^2$

network generation, we write the expected probability of error, averaged over the sensor network ensemble as $P_e = \mathrm{E}_S[P_{e,S}]$.

For a fixed sensor network s there is a known and fixed correspondence between target vectors $\boldsymbol{v_i}$ and $\boldsymbol{x_i}$. When we take the expectation over all such randomly generated sensor networks, the ideal sensor outputs associated with each target vector become random. We denote the random vector which occurs when $\boldsymbol{v_i}$ is the target vector as $\boldsymbol{X_i}$. In the random sensor network model, each sensor forms its connections independently of the other sensors. As a result, conditioned on the occurrence of $\boldsymbol{v_i}$ of type $\boldsymbol{\gamma}, P_{\boldsymbol{X_i}}(\boldsymbol{x_i}) = \prod_{\ell=1}^{n} P_{X_i}(x_{i\ell})$. As shown in (1), $P_{Xi}(x_i) = P^{\boldsymbol{\gamma}}(x_i)$. As as result, we can state $P_{\boldsymbol{X_i}}(\boldsymbol{x_i}) = \prod_{\ell=1}^{n} P^{\boldsymbol{\gamma}}(x_{i\ell}) \doteq P^{\boldsymbol{\gamma}}(\boldsymbol{x_i})$. Since a sensor network produces a codeword that is a function of the target vector, codeword distribution depends on the occurring target vector. We now consider a pair of random vectors $\boldsymbol{X_i}$ and $\boldsymbol{X_j}$, associated with a pair of target vectors, $\boldsymbol{v_i}$ and $\boldsymbol{v_j}$. Using (2) and the fact that each sensor form its connection independently, $P_{\boldsymbol{X_i X_j}}(\boldsymbol{x_i}, \boldsymbol{x_j}) = \prod_{\ell=1}^{n} P_{X_i X_j}(x_{i\ell}, x_{j\ell}) = \prod_{\ell=1}^{n} P^{\boldsymbol{\lambda}}(x_{i\ell}, x_{j\ell}) \doteq P^{\boldsymbol{\lambda}}(\boldsymbol{x_i}, \boldsymbol{x_j})$. This shows that $\boldsymbol{X_i}$ and $\boldsymbol{X_j}$ are *not independent*, since the sensor connections produce a dependency between the two ideal output vectors. Therefore, the 'codewords' $\mathcal{C} = \{\boldsymbol{X}_1, \ldots, \boldsymbol{X}_{2^k}\}$ of the sensor network are nonidentical and dependent on each other, unlike channel codes in classical information theory.

We assume a maximum-likelihood detector $g(\boldsymbol{y}) = \arg\max_j P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x_j})$. Repeating a couple of definitions for the sake of clarity, $d_{\mathrm{H}}(\boldsymbol{v_i}, \boldsymbol{v_j})$ is the Hamming distance between two target vectors, and the tolerable distortion region of $\boldsymbol{v_i}$ is $\mathcal{D}_{\boldsymbol{v_i}} = \{j : \frac{1}{k} d_{\mathrm{H}}(\boldsymbol{v_i}, \boldsymbol{v_j}) < D\}$. Given that $\boldsymbol{v_i}$ occurred, the detector is in error if $g(\boldsymbol{y}) \notin \mathcal{D}_{\boldsymbol{v_i}}$. We assume that target positions are generated i.i.d. with the probability of a target being present equal to $\frac{1}{2}$. As a result, all target vectors of length k have probability $\frac{1}{2^k}$. The probability of error averaged over the sensor network ensemble can be written

$$P_e = \mathrm{E}_{\boldsymbol{VY}\mathcal{C}}[\Pr(g(\boldsymbol{Y}) \notin \mathcal{D}_{\boldsymbol{V}} | \boldsymbol{V}, \mathcal{C}, \boldsymbol{Y})]. \tag{4}$$

Using the fact that we are taking the expectation of a probability, we bound P_e as follows:

$$P_e \leq \mathrm{E}_{\boldsymbol{VY}\mathcal{C}}\left[\sum_w \Pr(g(\boldsymbol{Y}) \in \mathcal{S}_w | \boldsymbol{V}, \mathcal{C}, \boldsymbol{Y})^\rho\right] \tag{5}$$

where $\rho \in [0, 1]$, and $\{\mathcal{S}_1, \mathcal{S}_2, \ldots\}$ is a partition of the complement of $\mathcal{D}_{\boldsymbol{V}}$, denoted $\mathcal{D}_{\boldsymbol{V}}^C$. Using the union bound, we upper bound the probability $\Pr(g(\boldsymbol{Y}) \in \mathcal{S}_w | \boldsymbol{V}, \mathcal{C}, \boldsymbol{Y})$ as follows:

$$P_e \leq \mathrm{E}_{\boldsymbol{VY}\mathcal{C}}\left[\sum_w \left(\sum_{j \in \mathcal{S}_w} \Pr(g(\boldsymbol{Y}) = j | \boldsymbol{V}, \mathcal{C}, \boldsymbol{Y})\right)^\rho\right]. \tag{6}$$

The term $\Pr(g(\boldsymbol{Y}) = j | \boldsymbol{V}, \mathcal{C}, \boldsymbol{Y})$ is a pairwise error term that depends only on the codewords $\boldsymbol{X_i}$ and $\boldsymbol{X_j}$. Using this observation, the fact that x^ρ is a concave function for $\rho \in [0, 1]$, and Jensen's inequality, we obtain

$$P_e \leq \mathrm{E}_{\boldsymbol{VYX_i}}\left[\sum_w \left(\sum_{j \in \mathcal{S}_w} \mathrm{E}_{\boldsymbol{X_j}|\boldsymbol{X_i}}[\Pr(g(\boldsymbol{Y}) = \boldsymbol{v_j} | \boldsymbol{V}, \boldsymbol{X_i}, \boldsymbol{X_j}, \boldsymbol{Y})]\right)^\rho\right]. \tag{7}$$

The term $\Pr(g(\boldsymbol{Y}) = \boldsymbol{v_j} | \boldsymbol{V}, \boldsymbol{X_i}, \boldsymbol{X_j}, \boldsymbol{Y})$ is a one-zero function, equaling one when $g(\boldsymbol{Y}) = \boldsymbol{v_j}$ and zero otherwise. Using our assumption that g is an ML detector we upper bound this probability as follows:

$$\begin{aligned} P_e \leq \frac{1}{2^k} \sum_i \sum_{\boldsymbol{x_i} \in \mathcal{X}^n} \sum_{\boldsymbol{y} \in \mathcal{Y}^n} & P_{\boldsymbol{X_i}}(\boldsymbol{x_i}) P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x_i}) \\ & \cdot \sum_w \left(\sum_{j \in \mathcal{S}_w} \sum_{\boldsymbol{x_j} \in \mathcal{X}^n} P_{\boldsymbol{X_j}|\boldsymbol{X_i}}(\boldsymbol{x_j}|\boldsymbol{x_i}) \right. \\ & \left. \times \left(\frac{P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x_j})}{P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x_i})}\right)^{\frac{1}{1+\rho}}\right)^\rho. \end{aligned} \tag{8}$$

The bound in (8) has an exponentially large number of terms, which we reduce to a polynomial number of terms using the method of types. The distributions in (8) can be completely specified by the type $\boldsymbol{\gamma}$ and joint type $\boldsymbol{\lambda}$ rather than the specific i, j pair of target vectors. This allows us to group terms in the summation by their types, which enables us to take advantage of the fact that the number of types is polynomial in k. We group the summation over i according to the type of $\boldsymbol{v_i}$. We group the summation over j by the joint type of $\boldsymbol{v_i}$ and $\boldsymbol{v_j}$. To do this, we choose each $\mathcal{S}_w$ to be a distinct joint type $\boldsymbol{\lambda}$, and let w index the set $S_{\boldsymbol{\gamma}}(D)$ of all $\boldsymbol{\lambda}$ that are the joint type of $\boldsymbol{v_i}$ and $\boldsymbol{v_j} \in \mathcal{D}_{\boldsymbol{v_i}}^C$. After grouping according to types, we write (8) as

$$\begin{aligned} P_e \leq \frac{1}{2^k} \sum_{\boldsymbol{\gamma}} \alpha(\boldsymbol{\gamma}, k) \sum_{\boldsymbol{x_i} \in \mathcal{X}^n} \sum_{\boldsymbol{y} \in \mathcal{Y}^n} & P^{\boldsymbol{\gamma}}(\boldsymbol{x_i}) P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x_i}) \\ & \cdot \sum_{\boldsymbol{\lambda} \in S_{\boldsymbol{\gamma}}(D)} \left(\beta(\boldsymbol{\lambda}, k) \sum_{\boldsymbol{x_j} \in \mathcal{X}^n} P^{\boldsymbol{\lambda}}(\boldsymbol{x_j}|\boldsymbol{x_i}) \right. \\ & \left. \times \left(\frac{P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x_j})}{P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x_i})}\right)^{\frac{1}{1+\rho}}\right)^\rho \end{aligned} \tag{9}$$

where $\alpha(\boldsymbol{\gamma}, k)$ is the number of target vectors $\boldsymbol{v_i}$ of length k and type $\boldsymbol{\gamma}$, and where $\beta(\boldsymbol{\lambda}, k)$ is the number of target vectors $\boldsymbol{v_j}$ of length k and joint type $\boldsymbol{\lambda}$ with a target vector $\boldsymbol{v_i}$ of type $\boldsymbol{\gamma}$. The set $S_{\boldsymbol{\gamma}}(D)$ is defined as

$$S_{\boldsymbol{\gamma}}(D) = \{\boldsymbol{\lambda} : \lambda_{01} + \lambda_{10} \geq D, \lambda_{00} + \lambda_{01} = \boldsymbol{\gamma}_0, \lambda_{10} + \lambda_{11} = \boldsymbol{\gamma}_1\}. \quad (10)$$

The condition of $\lambda_{01} + \lambda_{10} \geq D$ restricts the set to $\boldsymbol{\lambda}$ between $\boldsymbol{v_i}$ and $\boldsymbol{v_j}$ where the distortion between $\boldsymbol{v_i}$ and $\boldsymbol{v_j}$ exceeds D. This restricts our summation to the set of $\boldsymbol{v_j}$ that result in an error. The conditions $\lambda_{00} + \lambda_{01} = \boldsymbol{\gamma}_0$ and $\lambda_{10} + \lambda_{11} = \boldsymbol{\gamma}_1$ ensures that $\boldsymbol{\lambda}$ is consistent with $\boldsymbol{\gamma}$. Using standard results from the method of types [13] about the number of binary vectors of a given type, we obtain the bound $\alpha(\boldsymbol{\gamma}, k) \leq 2^{kH(\boldsymbol{\gamma})}$. The number of vectors with a given joint type is bounded as

$$\beta(\boldsymbol{\lambda}, k) = \binom{k\gamma_0}{k\lambda_{00}} \binom{k\gamma_1}{k\lambda_{11}} \leq 2^{k(H(\boldsymbol{\lambda}) - H(\boldsymbol{\gamma}))} \quad (11)$$

Combining (9) with the bounds on α and β, and using the conditional independence of sensor outputs, we obtain

$$P_e \leq \sum_{\boldsymbol{\gamma}} \sum_{\boldsymbol{\lambda} \in S_{\boldsymbol{\gamma}}(D)} 2^{-k(1-H(\boldsymbol{\gamma}))} 2^{k\rho(H(\boldsymbol{\lambda}) - H(\boldsymbol{\gamma}))} 2^{-nE(\rho, \boldsymbol{\lambda})} \quad (12)$$

where $E(\rho, \boldsymbol{\lambda})$ is defined as

$$E(\rho, \boldsymbol{\lambda}) = -\log \left(\sum_{a_i \in \mathcal{X}} \sum_{b \in \mathcal{Y}} P^{\boldsymbol{\gamma}}(a_i) P_{Y|X}(b|a_i)^{\frac{1}{1+\rho}} \cdot \left(\sum_{a_j \in \mathcal{X}} P^{\boldsymbol{\lambda}}(a_j|a_i) P_{Y|X}(b|a_j)^{\frac{1}{1+\rho}} \right)^{\rho} \right). \quad (13)$$

Since the number of types $\boldsymbol{\gamma}$ and joint types $\boldsymbol{\lambda}$ are upper bounded by $(k+1)^2$ and $(k+1)^4$, respectively, and $k = \lceil nR \rceil$, implying $k < nR + 1$, (12) is bounded as

$$P_e \leq 2^{-n\left(E_r(R,D) + \frac{o(\log(n))}{n}\right)} \quad (14)$$

where $o(\log(n))$ grows logarithmically in n, and where $E_r(R, D)$ is defined as

$$E_r(R, D) = \min_{\boldsymbol{\gamma}} \min_{\boldsymbol{\lambda} \in S_{\boldsymbol{\gamma}}(D)} \max_{0 \leq \rho \leq 1} (E(\rho, \boldsymbol{\lambda}) + R(1 - H(\boldsymbol{\gamma})) - \rho R(H(\boldsymbol{\lambda}) - H(\boldsymbol{\gamma}))). \quad (15)$$

The average error probability $P_e \to 0$ as $n \to \infty$ if $E_r(R, D) > 0$. Observing that $E(0, \boldsymbol{\lambda}) = 0 \quad \forall \boldsymbol{\lambda}$, we let ρ go to zero, rather than optimizing it, thus resulting in a lower bound on $E_r(R, D)$. In the above expression, this implies that in order for R to be achievable $\frac{E(\rho, \boldsymbol{\lambda})}{\rho} + R\frac{1-H(\boldsymbol{\gamma})}{\rho} - R(H(\boldsymbol{\lambda}) - H(\boldsymbol{\gamma}))$ must be positive for all types and joint types as $\rho \to 0$.

For $H(\boldsymbol{\gamma}) \neq 1$, $\frac{1-H(\boldsymbol{\gamma})}{\rho} \to \infty$ as $\rho \to 0$. For such a $\boldsymbol{\gamma}$, $P_e \to 0$ since $E_r(R, D)$ is positive for all rates R. Since we seek to bound R for which $E_r(R, D)$ is positive for all types and joint types, we let $\boldsymbol{\gamma} = (0.5, 0.5)$. This implies that as $\rho \to 0, R$ is achievable when the derivative of $E(\rho, \boldsymbol{\lambda})$ with respect to ρ at $\rho = 0$ is greater than $R(H(\boldsymbol{\lambda}) - H(\boldsymbol{\gamma}))$. It can be easily shown that, $\partial E(\rho, \boldsymbol{\lambda})/\partial \rho|_{\rho=0} = D(P^{\boldsymbol{\gamma}}_{X_iY} | Q^{\boldsymbol{\lambda}}_{X_iY})$. Using this derivative in the analysis above, we see that the achievable rates R are bounded as shown here.

$$R \leq \min_{\substack{\boldsymbol{\lambda} \\ \lambda_{01}+\lambda_{10} \geq D \\ \lambda_{00}+\lambda_{01}=\gamma_0 \\ \lambda_{10}+\lambda_{11}=\gamma_1}} \frac{D\left(P^{\boldsymbol{\gamma}}_{X_iY} \| Q^{\boldsymbol{\lambda}}_{X_iY}\right)}{H(\boldsymbol{\lambda}) - H(\boldsymbol{\gamma})} \quad (16)$$

where $\boldsymbol{\gamma} = (0.5, 0.5)$, and $\boldsymbol{\lambda}$ is an arbitrary p.m.f. since $n \to \infty$. Therefore, the right-hand side (RHS) of (16) is a lower bound on $C(D)$. ■

D. Discussion of Theorem

The most striking difference between the result shown in Theorem 1, and Shannon's channel capacity results is that the bound on the sensing capacity is not a mutual information. If the 'codewords' $\boldsymbol{X_i}$ were independent, the Kullback-Leibler distance would reduce to the mutual information between X_i and its noisy version Y. To see this, consider the case where target vectors $\boldsymbol{v_i}$ and $\boldsymbol{v_j}$ have types $\boldsymbol{\gamma}_i$ and $\boldsymbol{\gamma}_j$. Assume the joint type $\boldsymbol{\lambda}$ of the two vectors is a product of these two types, $\lambda_{ab} = \gamma_{ia}\gamma_{jb}$. For such a $\boldsymbol{\lambda}$, (2) can be used to show that the codewords are independent, $P^{\boldsymbol{\lambda}}(x_i, x_j) = P^{\boldsymbol{\gamma}_i}(x_i)P^{\boldsymbol{\gamma}_j}(x_j)$. In this case, $Q^{\boldsymbol{\lambda}}_{X_iY}(x_i, y) = P^{\boldsymbol{\gamma}_i}(x_i)P_Y(y)$ and the numerator in Theorem 1 would reduce to $I(X_i; Y)$. The fact that our result reduces to the mutual information only when the codewords are independent is important because of the frequent use of mutual information as a sensor selection metric (e.g., [12]), and indicates that the mutual information is not the correct notion of information for the large-scale detection applications considered in this paper.

The difference between channel capacity and sensing capacity arises due to different codeword geometries. In proofs of the achievability of channel capacity, since a codeword can be arbitrarily assigned to a message in communications, codewords are distributed uniformly. In a sensor network, the codeword distribution depends on the state of the environment (the target vector). Further, since in our model sensors are of limited range (i.e., $\frac{c}{k} \to 0$ as $k \to \infty$) the mapping between environment state and codewords is constrained. As a result, detection of a target vector with distortion equal to zero is impossible to achieve with a vanishing error probability. The following argument demonstrates that if distortion is equal to zero, then probability of error going to zero is not possible due to the finite number of connections per sensor. The probability of error P_e can be lower bounded as follows:

$$P_e \geq \frac{1}{2} \Pr[\text{ Target location 1 missed by all connections}] = \frac{1}{2}\left(1 - \frac{1}{k}\right)^{cn} = \frac{1}{2}\left(1 - \frac{1}{k}\right)^{\frac{kc}{R}}.$$

As k goes to infinity, $\frac{1}{2}(1 - \frac{1}{k})^{\frac{kc}{R}}$ converges to $\frac{1}{2}e^{-\frac{c}{R}}$. This demonstrates that the probability of error is lower bounded by a function that does not decrease even as k and n increase to infinity at a fixed rate R. To drive the lower bound on the error probability to zero for a fixed c requires the rate to decrease to zero.

Our results can be understood in the context of results on the error probability of linear codes. Consider sensors with a sensing function of addition modulo two. For such a sensing function, the sensor network can be modeled as a low-density generator-matrix code. Such codes are duals of LDPC codes [41], and are known to have a ratio of codeword distance to codeword length decreasing to zero with codeword length [42]. In LDPC codes, changing a single message bit changes δn of the n codeword bits, for some positive $\delta \in (0,1]$. For our sensor network model, changing a single message bit changes ϵn codeword bits, where $\epsilon \rightarrow 0$ for large codewords. As a result of constraints on the encoder, codewords are clustered, with similar target vectors encoded as similar codewords. Therefore, similar target vectors are more likely to be confused due to noise than dissimilar target vectors. The Kullback-Leibler distance in Theorem 1 is the appropriate information measure for such a codeword geometry. The minimization over the joint type appears because the "closest" target vectors dominate the error probability. Thus, the sensing capacity is similar to classical channel capacity, with differences arising due to the nonidentical, dependent codeword distribution.

Theorem 1 provides a basis for comparing sensors of different type. For example, consider two types of sensors that differ only in their sensing function, where one sensors can count up to c targets, while the other sensor can only differentiate between 0 and ≥ 1 targets. Which type of sensor yields a higher sensing capacity? We use our bound to demonstrate a numerical answer to such questions in Section IV-F. Our bound can also be used gain analytical insight. Consider two types of sensor with sensing functions Ψ^1 and Ψ^2. We assume that both sensor types are noiseless and have range c. When the sensing function Ψ^2 is a function of the sensing function Ψ^1, the sensing capacity of the first type of sensor exceeds or equals that of the second type. This can be shown by using (1) and (2) in the bound shown in (3), and applying the log sum inequality [43].

Recently, [36] extended our definition of sensing capacity to apply to a different model where the environment is modeled as a real vector, and the sensing functions are linear, with fixed SNR. Importantly, the range of the sensors used in [36] grows proportionally with the size of the environment, in contrast to our model where sensor range is fixed. The bounds obtained in these papers differ from our results. We attribute these differences to the significantly different modeling assumptions.

E. Extensions

Section IV-A introduced a sensor network model where each sensor is allowed to make arbitrary connections to the target vector. In several applications, more complex sensor network models will be appropriate. This section describes extensions of the arbitrary connection model. Extensions that account for contiguity in sensor connections require a new model and are discussed in Section V. The first extension considers nonbinary target vectors. Binary target vectors indicate the presence or absence of targets at the spatial positions. A target vector over a general finite alphabet may indicate, in addition to the presence of targets, the class of a target. Alternatively, the entries of nonbinary vectors can indicate levels of intensity or concentration. Assuming a nonbinary target vector, we can define types and joint types over an alphabet $\mathcal{V}$, and apply the same analysis as before to obtain the sensing capacity bound here.

$$C(D) \geq C_{LB}(D) = \min_{\substack{\boldsymbol{\lambda} \\ \sum_{a\neq b}\lambda_{ab}\geq D \\ \sum_b \lambda_{ab}=\gamma_a}} \frac{D\left(P^{\boldsymbol{\gamma}}_{X_iY} \| Q^{\boldsymbol{\lambda}}_{X_iY}\right)}{H(\boldsymbol{\lambda}) - H(\boldsymbol{\gamma})} \tag{17}$$

where $\boldsymbol{\gamma} = (\gamma_a = \frac{1}{|\mathcal{V}|}, a \in \mathcal{V})$, while $\boldsymbol{\lambda} = (\lambda_{ab},\ a, b \in \mathcal{V})$ is an arbitrary probability mass function.

The second extension allows the following *a priori* distribution over target vectors. Assume that each target position is generated i.i.d. with probability P_V over the alphabet $\mathcal{V}$. This may model the fact that targets are sparsely present. The previous proof can be extended by using a maximum-*a-posteriori* (MAP) detector, instead of the ML detector considered earlier. To state the resulting sensing capacity bound, we introduce the following definitions: $\boldsymbol{\lambda} = (\lambda_{ab}, a, b \in \mathcal{V})$ is an arbitrary probability mass function, $\boldsymbol{\gamma}_i = P_V, \boldsymbol{\gamma}_j$ is the marginal of $\boldsymbol{\lambda}$ calculated as $\gamma_{jb} = \sum_a \lambda_{ab}, \mathcal{R}(D) = \{\boldsymbol{\lambda} : H(\boldsymbol{\lambda}) - H(\boldsymbol{\gamma}_j) - D(\boldsymbol{\gamma}_j | P_V) > 0, \sum_{a\neq b} \lambda_{ab} \geq D, \sum_b \lambda_{ab} = \gamma_{ia}\}$. If the set $\mathcal{R}(D)$ is empty, the sensing capacity is infinite since all rates are achievable. Otherwise, the sensing capacity is bounded as follows:

$$\begin{aligned} C(D) &\geq C_{LB}(D) \\ &= \min_{\boldsymbol{\lambda}\in\mathcal{R}(D)} \frac{D\left(P^{\boldsymbol{\gamma}_i}_{X_iY} \| Q^{\boldsymbol{\lambda}}_{X_iY}\right)}{H(\boldsymbol{\lambda}) - H(\boldsymbol{\gamma}_j) - D(\boldsymbol{\gamma}_j \| P_V)}. \end{aligned} \tag{18}$$

This extension can be used to model sparsity where the number of targets present is close to δk, where $\delta \in (0,1)$ is a constant. If we consider a definition of sparsity where the number of targets present is ϵk, with $\epsilon \rightarrow 0$ for large k, then the sensing capacity goes to infinity. The reason for this is that due to our definition of distortion, all target vectors will be indistinguishable for any $D > 0$, and no sensors are necessary to achieve the desired distortion. The bound shown in (18) will be plotted in Section IV-F to demonstrate the effect of sparsity on the sensing capacity.

A third extension accounts for heterogenous sensors, where each class of sensor possibly has a different range c, noise model $P_{Y|X}$, and/or sensing function Ψ. Let the sensor of class l be used with a given relative frequency α_l. The number of sensors of type l is $\alpha_l n$ and $\sum_l \alpha_l = 1$. For such a model the sensing capacity bound is as follows:

$$\begin{aligned} C(D) &\geq C_{LB}(D) \\ &= \min_{\substack{\boldsymbol{\lambda} \\ \sum_{a\neq b}\lambda_{ab}\geq D \\ \sum_b \lambda_{ab}=\gamma_a}} \frac{\sum_l \alpha_l D\left(P^{\boldsymbol{\gamma},l}_{X_iY} \| Q^{\boldsymbol{\lambda},l}_{X_iY}\right)}{H(\boldsymbol{\lambda}) - H(\boldsymbol{\gamma})} \end{aligned} \tag{19}$$

where $\boldsymbol{\gamma} = (\gamma_a = \frac{1}{|\mathcal{V}|}, a \in \mathcal{V}), \boldsymbol{\lambda} = (\lambda_{ab},\ a, b \in \mathcal{V})$ is an arbitrary probability mass function. The key modification to the original proof that is used to obtain this extension is to use the conditional independence of sensor outputs to factor the probabilities in (9) according to sensor type. The mixture

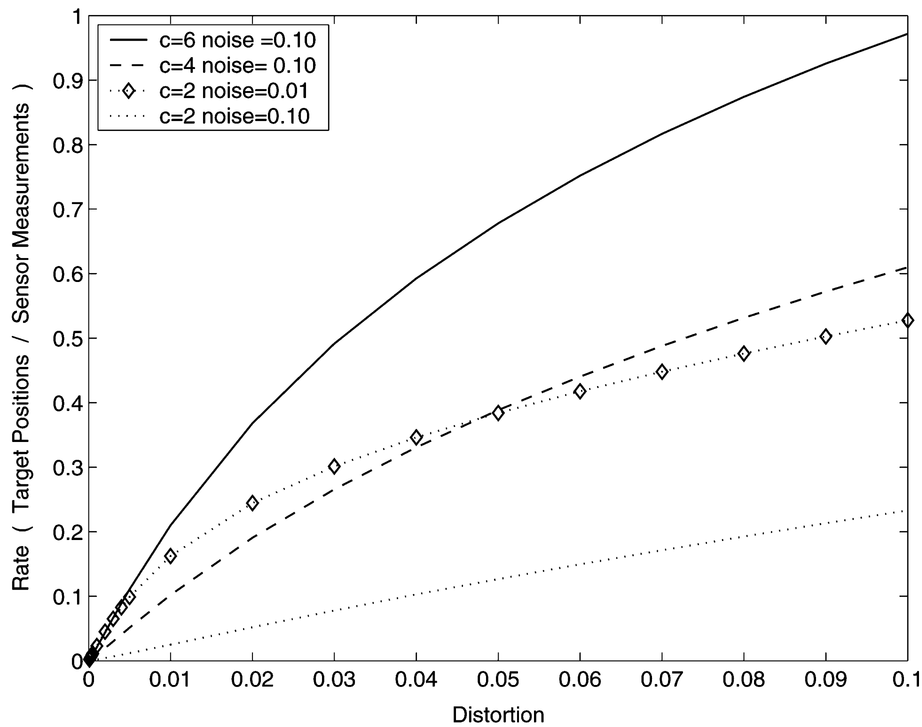

Fig. 4. $C_{LB}(D)$ of arbitrary connections model for sensors of varying noise levels and range.

of sensors used in practical applications is constrained by factors such as cost, availability, and energy consumption. Due to these external, application-dependent constraints, optimization over the α_l's requires assumptions that are beyond the scope of this paper. The extensions described in (18) and (19) can be combined in an extension that accounts for both sensor heterogeneity and target sparsity.

F. Numerical Results

We compute the capacity bound $C_{LB}(D)$ in Theorem 1 for various distortions, noise levels, and sensor ranges. A sensor of range c is connected to c target positions. We assume that the sensing function Ψ simply counts the number of target positions in the sensor range with a target present. The sensor noise model assumes that the probability of counting error decays exponentially with the error magnitude. In the figures, '$\text{Noise} = p'$ indicates that for a sensor $P(Y \neq X) = p$ with $\mathcal{Y} = \mathcal{X}$ assumed. In Fig. 4, we demonstrate $C_{LB}(D)$ for various sensor noise levels and ranges. We compute this bound by systematically sampling the space of possible $\boldsymbol{\lambda}$. While $\boldsymbol{\lambda}$ is a four-dimensional vector, because of constraints we need to sample only two dimensions in order search over all valid $\boldsymbol{\lambda}$. In all cases, $C_{LB}(D)$ approaches 0 as D approaches 0. This occurs because similar target vectors have similar codewords due to dependence in the codeword distribution. The relative magnitude of the bounds for sensors of various c and noise levels describes tradeoffs among sensor types that can be captured by our result. Some tradeoffs are intuitive. For example, lower noise sensor of range c have a higher sensing capacity than higher noise sensors of the same range. Other tradeoffs are more complex. For example the tradeoff between shorter and longer range sensors depends on the desired distortion. Sensors of range 4 and noise 0.10 result in a higher sensing capacity than sensors of range 2 and noise 0.01 for distortion above 0.047. The opposite is true for distortions below 0.047. Thus, the bound presented in (1) describes a complex tradeoffs between sensor noise, sensor range, and the desired detection accuracy.

Fig. 5 shows $C_{LB}(D)$ at $D = 0.1$ as a function of sensor noise level for sensors of various range and sensing functions. This figure demonstrates that the strategy of simple sensor replication, which is a popular practical method for reducing error probability, can be inefficient. For example, for sensors of range 4 and a sum sensing function, a rate of 0.61 is achievable at noise level 0.1. If each sensor with noise 0.1 is replicated three times and majority decoding is used, the noise can be reduced to $3 \times (0.1)^2 \times 0.9 + (0.1)^3 = 0.028$. For a noise level of $0.028, C_{LB}(0.1)$ equals 0.91 for a sensor of range 4 and a sum sensing function. However, due to sensor replication, the rate is reduced to $0.91/3 = 0.303$. This rate is significantly lower than the rate of 0.61 for sensors of noise 0.1 achievable by using our random sensor network construction. Thus, the bound indicates that cooperative sensor strategies can require significantly fewer sensor measurements than sensor replication. Fig. 5 also shows $C_{LB}(D)$ at $D = 0.1$ for sensors with $c = 4$ and a weighted sum sensing function with weights $\{1, 0.5, 0.25, 0.1\}$. This sensing function has a higher sensing capacity than sensors with the same range and an unweighted sum sensing function across all noise levels. We conjecture that this occurs because a weighted sum can distinguish among more target configurations than an unweighted sum. Interestingly, the gap between the two sensing functions increases with increasing noise.

While the weighted sum sensing function results in a higher rate than a simple sum sensing function, a sensor network designer might be interested in a normalized rate, where the rate is

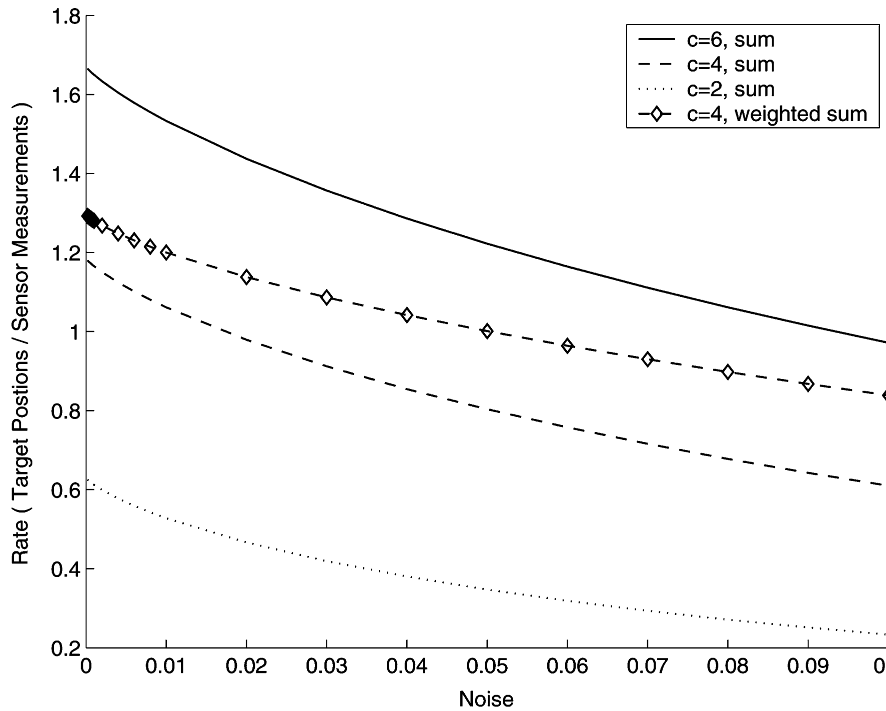

Fig. 5. $C_{LB}(0.1)$ of arbitrary connections model for sensors of varying noise levels, range, and sensing function.

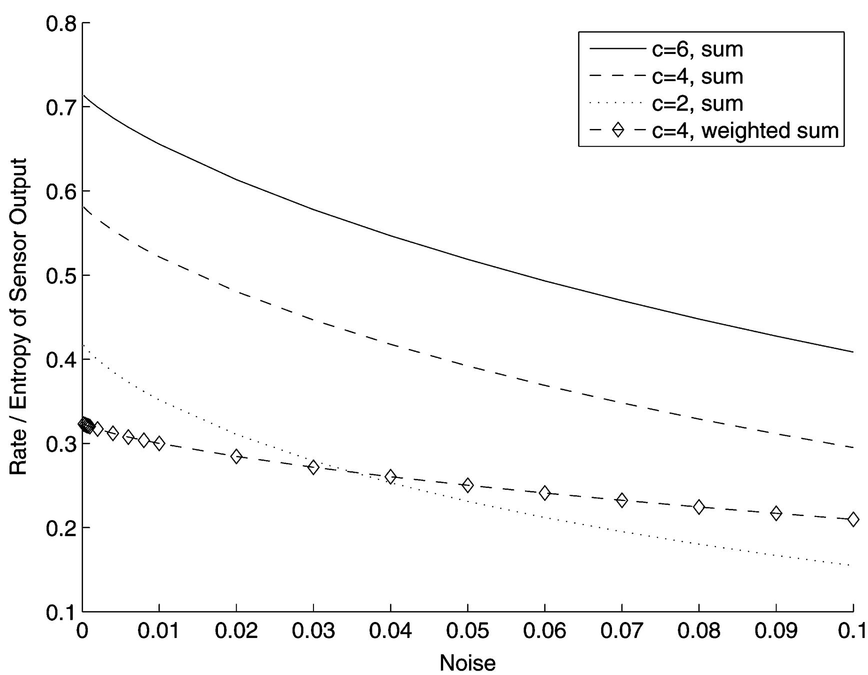

Fig. 6. $C_{LB}(0.1)$ of arbitrary connections model normalized by the entropy of the sensor output, for sensors of varying noise levels, range, and sensing function.

divided by the number of bits a sensor produces. Fig. 6 presents the curves shown in Fig. 5 normalized by the entropy of the sensor output $H(Y)$, assuming the typical target vector occurs, i.e., $\boldsymbol{\gamma} = (0.5, 0.5)$. This graph shows that to achieve a desired distortion, the weighted sum sensing function requires the transmission of more bits than a simple sum sensing function. This presents a sensor network designer with an interesting tradeoff. For example, if sensors are expensive, and communications is not a constraint, a designer would prefer sensors with the higher resolution weighted sum function. A similar observation does not hold about the simple sum function for sensors of varying range. In this graph, sum sensors of higher range, can achieve higher rates per sensor and per sensor-bit than sum sensors of shorter range.

Fig. 7 contains plots of (18) for various values of target vector sparsity. We computed these bounds for sensors with a sum sensing function, $c = 16$, and a noise probability of 0.05. This figure demonstrates that the effect of target sparsity on the

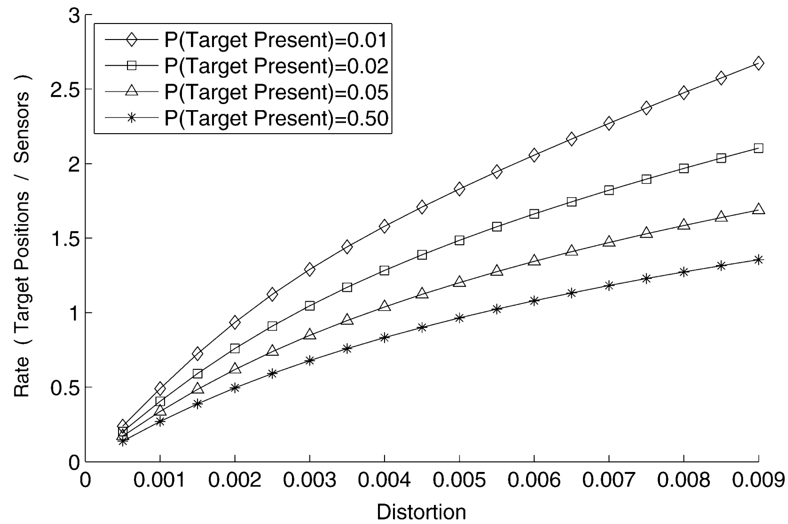

Fig. 7. $C_{LB}(D)$ of arbitrary connections model for target vectors of varying sparsity.

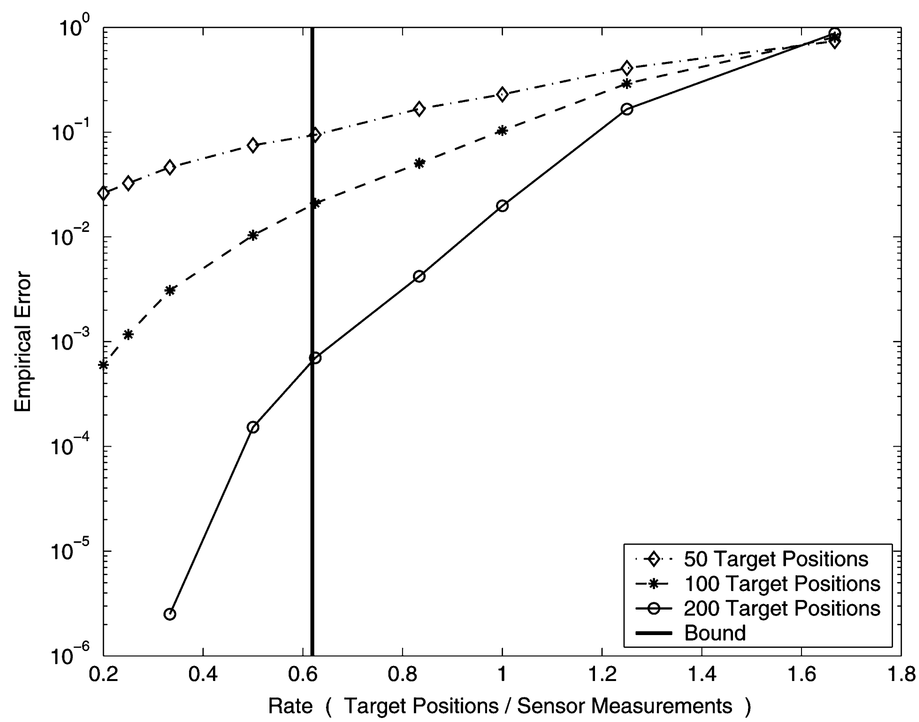

Fig. 8. Average empirical error rate of belief propagation based detection for varying rates, and the corresponding sensing capacity bound.

sensing capacity is significant. For a target probability of 0.01, a rate of 2.27 at a distortion of 0.007 is achievable. For a target probability of 0.5, which corresponds to the capacity bound in Theorem 1, the achievable rates are bounded by 1.18. This demonstrates that knowledge of target sparsity can be exploited to deploy a significantly smaller number of sensors to achieve a desired distortion.

Using the loopy belief propagation algorithm [44] we empirically examined sensor probability of error as a function of rate. We generated sensor networks of various rates by setting the number of targets, and varying the number of sensors. We chose the number of connections to be $c = 4$, the distortion level to be 0.1, and the noise level to be 0.1 (i.e., $P(Y \neq X) = 0.1$, with $\mathcal{Y} = \mathcal{X}$). As in the previous section, we assume that the probability of error decays exponentially with error magnitude. We empirically evaluated the average error rate obtained in decoding target vectors in a randomly generated set of sensor networks. We plotted the average error rate for each rate value, and for various numbers of targets as shown in Fig. 8. As the number of targets increase, the transition from high error to low error rate becomes increasingly sharp. However, all the error curves are well below the achievable rate $C_{LB}(0.1) = 0.62$. We conjecture that this occurs because belief propagation is suboptimal for graphs with cycles.

V. Sensing Capacity of Contiguous Connections Model

In this section, we bound the sensing capacity of a sensor network model that models contiguity in a sensor's connections.

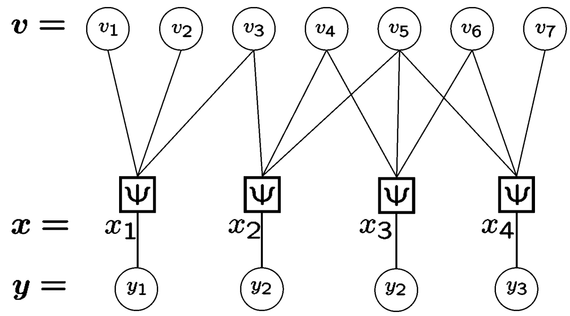

Fig. 9. Sensor network model with $k = 7, n = 3, c = 3$, contiguous connections, and a sensing function corresponding to the weighted sum of the observed targets.

Fig. 9 shows an example of such a model. Sensor ℓ is connected to exactly c *contiguous* positions out of the k spatial positions. In contrast, the arbitrary connections model analyzed in the previous section did not account for localized sensor observations since each sensor could sense any c (not necessarily contiguous) spatial positions.

A. Higher Order Types

The statement of the result for contiguous models requires higher order types [13]. We introduce *circular c-order types* and *circular c-order joint types*. We define the circular c-order type of a binary sequence (i.e., a target vector) as a 2^c dimensional vector, $\boldsymbol{\gamma}$, where each entry in the vector corresponds to the frequency of occurrence of one of the possible subsequences of length c. A circular sequence is one in which the last element of the sequence precedes the first element of the sequence. The total number of subsequences of length c that can occur in a circular sequence of length k is k. For example, for a binary target vector and $c = 2, \boldsymbol{\gamma} = (\gamma_{00}, \gamma_{01}, \gamma_{10}, \gamma_{11})$. While it is possible to prove our bound using noncircular types as shown in [45], circular types lead to the same asymptotic result with the benefit of significantly simpler notation. The notational simplicity arises out of the fact that the lower order circular types are precise marginals of the higher order circular types. Although all the types in this section are circular, we will omit the word 'circular' when referring to types in the remainder of this section for brevity.

We denote the set of all c-order types over the alphabet $\mathcal{V}^c$ for target vectors of length k as $\mathcal{P}_k(\mathcal{V}^c)$. Since each sensor independently chooses a block of c contiguous spatial positions, the distribution of its ideal output X_i depends only on the c-order type $\boldsymbol{\gamma}$ of the target vector $\boldsymbol{v_i}$ which occurs. For a sensing function Ψ and a target vector $\boldsymbol{v_i}$ of type $\boldsymbol{\gamma}$

$$P_{X_i}(X_i = x_i) = \sum_{\substack{\boldsymbol{a} \in \mathcal{V}^c \\ \Psi(\boldsymbol{a}) = x_i}} \gamma_{\boldsymbol{a}} \doteq P^{\boldsymbol{\gamma}}(x). \tag{20}$$

Next, we note that the joint distribution $P_{\boldsymbol{X_i X_j}}$ depends on the *c-order joint type* $\boldsymbol{\lambda}$ of the ith and jth target vectors $\boldsymbol{v_i}, \boldsymbol{v_j}$. $\boldsymbol{\lambda}$ is the vector of $\lambda_{(\boldsymbol{a})(\boldsymbol{b})}$, the fraction of positions in $\boldsymbol{v_i}, \boldsymbol{v_j}$ where $\boldsymbol{v_i}$ has a bit subsequence $\boldsymbol{a}$ while $\boldsymbol{v_j}$ has a bit subsequence $\boldsymbol{b}$. For example, when $c = 2$ and $\mathcal{V} = \{0, 1\}, \boldsymbol{\lambda} = (\lambda_{(00)(00)}, \ldots, \lambda_{(11)(11)})$. We denote the set of all c-order joint types over the alphabet $\mathcal{V}^c$ for target vectors of length k as $\mathcal{P}_k(\mathcal{V}^c, \mathcal{V}^c)$. Each $\boldsymbol{\lambda} \in \mathcal{P}_k(\mathcal{V}^c, \mathcal{V}^c)$ must satisfy the normalization constraint that the sum over all entries of λ equals one. Since the joint type $\boldsymbol{\lambda}$ also defines the type $\boldsymbol{\gamma}$ of $\boldsymbol{v_i}$, for all $\{\boldsymbol{a}\} \in \mathcal{V}^c$ we must have $\gamma_{\boldsymbol{a}} = \sum_{\boldsymbol{b} \in \mathcal{V}^c} \lambda_{(\boldsymbol{a})(\boldsymbol{b})}$. Taking advantage of the fact that for circular types, lower order types are precise marginals of higher order types, we denote $\lambda_{(a)(b)} = \sum_{\boldsymbol{a'} \in \mathcal{V}^{c-1}} \sum_{\boldsymbol{b'} \in \mathcal{V}^{c-1}} \lambda_{(a\boldsymbol{a'})(b\boldsymbol{b'})}$. $\lambda_{(a)(b)}$ is the normalized count of locations where target vector i has value a while target vector j has value b. Since each sensor depends only on the c contiguous targets bits which it senses, $P_{\boldsymbol{X_i X_j}}$ depends only on the joint type $\boldsymbol{\lambda}$. For target vectors $\boldsymbol{v_i}, \boldsymbol{v_j}$ of c-order joint type $\boldsymbol{\lambda}$,

$$P_{X_i X_j}(X_i = x_i, X_j = x_j) = \sum_{\substack{\boldsymbol{a}, \boldsymbol{b} \in \mathcal{V}^c \\ \Psi(\boldsymbol{a}) = x_i \\ \Psi(\boldsymbol{b}) = x_j}} \lambda_{(\boldsymbol{a})(\boldsymbol{b})} \doteq P^{\boldsymbol{\lambda}}(x_i, x_j). \tag{21}$$

TABLE III
$\boldsymbol{\lambda}$ WITH $c = 2$ FOR $\boldsymbol{v_i} = 01101000$ AND $\boldsymbol{v_j} = 01000111$

$\lambda_{(ab)(cd)}$	$cd = 00$	$cd = 01$	$cd = 10$	$cd = 11$
$ab = 00$	0	0	1/8	2/8
$ab = 01$	1/8	1/8	0	0
$ab = 10$	1/8	1/8	0	0
$ab = 11$	0	0	1/8	0

For example, for binary target vectors and $c = 2$, vectors 00000000, 01101000, 01000111 have $\boldsymbol{\gamma} = (1, 0, 0, 0), (3/8, 2/8, 2/8, 1/8), (2/8, 2/8, 2/8, 2/8)$, respectively. Table III contains the 2-order joint type of two target vectors. Consider a sensor network where each sensor is randomly connected to $c = 2$ contiguous spatial positions. We assume that Ψ outputs the number of targets which the sensor observes. Thus, each sensor has an ideal output alphabet $\mathcal{X} = \{0, 1, 2\}$. For target vectors of type $\boldsymbol{\gamma}, P(X_i = 0) = \gamma_{00}, P(X_i = 1) = \gamma_{01} + \gamma_{10}, P(X_i = 2) = \gamma_{11}$, respectively. Given two target vectors $\boldsymbol{v_i}, \boldsymbol{v_j}$ of joint type $\boldsymbol{\lambda}$, a sensor will output '0' for both target vectors only if both of its connections see a '0' bit in both target vectors. This happens with probability $\lambda_{(00)(00)}$. Table IV lists the joint p.m.f. $P_{X_i X_j}(x_i, x_j) = P^{\boldsymbol{\lambda}}(x_i, x_j)$ for all output pairs x_i, x_j corresponding to joint type $\boldsymbol{\lambda}$. The table shows that X_i, X_j are not independent, in general.

To prove Theorem 1, we bounded the number of target vectors $\boldsymbol{v_j}$ that have a given joint type with a target vector $\boldsymbol{v_i}$ in (11). To prove a sensing capacity bound for the contiguous connections model we prove a bound on the number of target vectors $\boldsymbol{v_j}$ that have a joint c-order type $\boldsymbol{\lambda}$ with a target vector of c-order type $\boldsymbol{\gamma}$ in the lemma below. Before proceeding, we introduce the following notation. The set of length k target vectors of c-order type $\boldsymbol{\gamma}$ is denoted $\mathcal{T}^k_{\boldsymbol{\gamma}}$. The set of pairs of length k target vectors of joint type $\boldsymbol{\lambda}$ is denoted $\mathcal{T}^k_{\boldsymbol{\lambda}}$. The set of length k target vectors that have joint c-order type $\boldsymbol{\lambda}$ with a given vector of type $\boldsymbol{\gamma}$, is denoted $\mathcal{T}^k_{\boldsymbol{\lambda}|\boldsymbol{\gamma}}$.

Lemma 1 (Bound on $|\mathcal{T}^k_{\boldsymbol{\lambda}|\boldsymbol{\gamma}}|$): The number of binary vectors of length k with c-order joint type $\boldsymbol{\lambda}$ for a given vector of c-order type $\boldsymbol{\gamma}$, denoted $|\mathcal{T}^k_{\boldsymbol{\lambda}|\boldsymbol{\gamma}}|$, is bounded as follows:

$$|\mathcal{T}^k_{\boldsymbol{\lambda}|\boldsymbol{\gamma}}| \leq C(k) 2^{k(H(\tilde{\boldsymbol{\lambda}}|\boldsymbol{\lambda}') - H(\tilde{\gamma}|\boldsymbol{\gamma}'))} \tag{22}$$

TABLE IV
JOINT DISTRIBUTION OF X_j AND X_i IN TERMS OF THE JOINT TYPE $\boldsymbol{\lambda}$ OF $\boldsymbol{v_j}$ AND $\boldsymbol{v_i}$, WITH $c = 2$

$P_{X_iX_j}$	$X_j = 0$	$X_j = 1$	$X_j = 2$
$X_i = 0$	$\lambda_{(00)(00)}$	$\lambda_{(00)(01)} + \lambda_{(00)(10)}$	$\lambda_{(00)(11)}$
$X_i = 1$	$\lambda_{(10)(00)} + \lambda_{(01)(00)}$	$\lambda_{(01)(01)} + \lambda_{(01)(10)} + \lambda_{(10)(01)} + \lambda_{(10)(10)}$	$\lambda_{(10)(11)} + \lambda_{(01)(11)}$
$X_i = 2$	$\lambda_{(11)(00)}$	$\lambda_{(11)(01)} + \lambda_{(11)(10)}$	$\lambda_{(11)(11)}$

$C(k) = 2^{2(c-1)}k^{2^{c-1}}(k+1)^{2^{2(c-1)}}$ and $\boldsymbol{\lambda}' = \{\lambda_{(\boldsymbol{a})(\boldsymbol{b})}, \forall \boldsymbol{a}, \boldsymbol{b} \in \mathcal{V}^{c-1}\}$ is a probability mass function defined as $\lambda'_{(\boldsymbol{a})(\boldsymbol{b})} = \sum_{a,b\in\mathcal{V}} \lambda_{(\boldsymbol{a}a)(\boldsymbol{b}b)}$. $\tilde{\boldsymbol{\lambda}} = \{\tilde{\lambda}_{(\boldsymbol{a}a)(\boldsymbol{b}b)}, \quad \forall \boldsymbol{a}, \boldsymbol{b} \in \mathcal{V}^{c-1}, \quad \forall a, b \in \mathcal{V}\}$ is a conditional probability mass function defined as $\tilde{\boldsymbol{\lambda}}_{(\boldsymbol{a}a)(\boldsymbol{b}b)} = \frac{\lambda_{(\boldsymbol{a}a)(\boldsymbol{b}b)}}{\lambda_{(\boldsymbol{a})(\boldsymbol{b})}}$. $\boldsymbol{\gamma}' = \{\gamma'_{\boldsymbol{a}}, \quad \forall \boldsymbol{a} \in \mathcal{V}^{c-1}\}$ is probability mass function defined as $\gamma'_{\boldsymbol{a}} = \sum_{a\in\mathcal{V}} \gamma_{\boldsymbol{a}a}$. $\tilde{\boldsymbol{\gamma}} = \{\tilde{\gamma}_{\boldsymbol{a}a}, \quad \forall \boldsymbol{a} \in \mathcal{V}^{c-1}, \quad \forall a \in \mathcal{V}\}$ is a conditional probability mass function defined as $\tilde{\boldsymbol{\gamma}}_{\boldsymbol{a}a} = \frac{\gamma_{\boldsymbol{a}a}}{\gamma_{\boldsymbol{a}}}$.

Proof: To bound $|\mathcal{T}^k_{\boldsymbol{\lambda}|\boldsymbol{\gamma}}|$, we begin by bounding $|\mathcal{T}^k_{\boldsymbol{\gamma}}|$. The c-order type $\boldsymbol{\gamma}$ of a binary vector specifies a 2-order type (referred to as a Markov type) of a vector who entries are in an alphabet of cardinality 2^{c-1}. Consider the vector 011010001. Denoting a pair of bits using $\{0,1,2,3\}$, as we move from left to right over this vector, one bit at a time, the sequence obtained is 132120012. The 3-order type $\boldsymbol{\gamma}$ specifies the 2-order type over this new vector. For example, the fraction of times 1 transitions to 3 is equal to γ_{011} and the fraction of times 1 transitions to 2 is equal to γ_{010}. Any c-order type over a binary sequence can thus be mapped to a 2-order type over a sequence with symbols in an alphabet of cardinality 2^{c-1}. [46] proves bounds on the number of sequences that correspond to a 2-order circular type over a sequence with an alphabet $\mathcal{V}$. Given our mapping from a c-order type to a 2-order type, we can apply this result to obtain the following bound:

$$|\mathcal{T}^k_{\boldsymbol{\gamma}}| \geq C_1(k)2^{k(H(\boldsymbol{\gamma})-H(\boldsymbol{\gamma}'))} = C_1(k)2^{kH(\tilde{\gamma}|\gamma')} \tag{23}$$

where $C_1(k) = k^{-2^{c-1}}(k+1)^{-2^{2(c-1)}}$. We now bound $|\mathcal{T}^k_{\boldsymbol{\lambda}}|$ using a similar argument. The c-order joint type $\boldsymbol{\lambda}$ of a pair of binary vectors specifies a 2-order type of a single vector whose entries are symbols from an alphabet of cardinality $2^{2(c-1)}$. We consider an example for a 3-order joint type, and for vectors $\boldsymbol{v} = 011010001$ and $\boldsymbol{v}' = 101011011$. We can rewrite these vectors as a single vector whose entries at location i are defined by the pair of entries v_i, v'_i and the subsequent pair of entries v_{i+1}, v'_{i+1}. These four entries, combined as $v_iv_{i+1}v'_iv'_{i+1}$ are mapped to a symbol in an alphabet of cardinality 2^4 by reading the entries as a binary number (i.e., $0000 = 0, 0001 = 1, \ldots$). In this manner, $\boldsymbol{v}$ and $\boldsymbol{v}'$ are mapped to a vector $(6, 14, 10, 9, 11, 2, 1, 7, 11)$. The 3-order joint type $\boldsymbol{\lambda}$ specifies the 2-order type over this new vector. For example, the fraction of times 1 transitions to 7 is equal to $\lambda_{(001)(011)}$ and the fraction of times 2 transitions to 1 is equal to $\lambda_{(000)(101)}$. Any c-order joint type over a binary sequence can thus be mapped to a 2-order type over a sequence with symbols in an alphabet of cardinality $2^{2(c-1)}$. We use the results of [46] again. Given our mapping from a c-order joint type to a 2-order type, we can apply this result to obtain the following bound:

$$|\mathcal{T}^k_{\boldsymbol{\lambda}}| \leq C_2 2^{k(H(\boldsymbol{\lambda})-H(\boldsymbol{\lambda}'))} = C_2 2^{kH(\tilde{\boldsymbol{\lambda}}|\boldsymbol{\lambda}')} \tag{24}$$

where $C_2 = 2^{2(c-1)}$. We observe that $|\mathcal{T}^k_{\boldsymbol{\lambda}|\boldsymbol{\gamma}}|$ depends only on the type $\boldsymbol{\gamma}$ of the vector on which we are conditioning, and not on the actual vector. Therefore, $|\mathcal{T}^k_{\boldsymbol{\lambda}|\boldsymbol{\gamma}}| = \frac{|\mathcal{T}^k_{\boldsymbol{\lambda}}|}{|\mathcal{T}^k_{\boldsymbol{\gamma}}|}$. Using (23) and (24), we obtain the following bound:

$$|\mathcal{T}^k_{\boldsymbol{\lambda}|\boldsymbol{\gamma}}| \leq C(k)2^{k(H(\tilde{\boldsymbol{\lambda}}|\boldsymbol{\lambda}')-H(\tilde{\gamma}|\gamma'))} \tag{25}$$

where $C(k) = C_1^{-1}(k)C_2$ ∎

B. Sensing Capacity Lower Bound

We define $P^{\boldsymbol{\gamma}}_{X_iY}Q^{\boldsymbol{\lambda}}_{X_iY}$ as they were defined for the arbitrary connections model bounds, with the only difference arising to the use of c-order types instead of types.

Theorem 2: (Sensing Capacity Achievability Theorem for the Contiguous Connections Model): The sensing capacity at distortion D satisfies

$$C(D) \geq C_{LB}(D) = \min_{\substack{\boldsymbol{\lambda} \\ \lambda_{(0)(1)}+\lambda_{(1)(0)} \geq D}} \frac{D\left(P^{\boldsymbol{\gamma}}_{X_iY} \| Q^{\boldsymbol{\lambda}}_{X_iY}\right)}{H(\tilde{\boldsymbol{\lambda}}|\boldsymbol{\lambda}') - H(\tilde{\boldsymbol{\gamma}}|\boldsymbol{\gamma}')} \tag{26}$$

where $\boldsymbol{\lambda} \in \mathcal{P}(\{0,1\}^c, \{0,1\}^c), \boldsymbol{\gamma}_{\boldsymbol{a}} = \sum_{\boldsymbol{b}\in\{0,1\}^c} \lambda_{(\boldsymbol{a})(\boldsymbol{b})}$, and $H(\tilde{\boldsymbol{\gamma}}|\boldsymbol{\gamma}') = 1$.

If we specialize this result to the case of $c = 1$, this theorem provides a bound that coincides with our bound for the arbitrary connections model. The proof of the sensing capacity lower bound is similar for the arbitrary and contiguous connections models. The main differences in the proofs arise due to the contiguity of sensor field of view, which necessitates the use of c-order types. Extensions demonstrated in Section IV-E for the arbitrary connections model can be easily applied to the contiguous connections model.

Proof Outline: The proof of Theorem 2 is essentially identical to the proof of Theorem 1, with types and joint types replaced by c-order types and joint types. The use of these higher order types requires counting arguments described in Lemma 1. For c-order types, we bound α in (9) as follows:

$$\alpha(\boldsymbol{\gamma}, k) = |\mathcal{T}^k_{\boldsymbol{\gamma}}| \leq 2^{kH(\tilde{\gamma}|\boldsymbol{\gamma}')}. \tag{27}$$

For c-order joint types, we bound $\beta(\boldsymbol{\lambda}, k) = |\mathcal{T}^k_{\boldsymbol{\lambda}|\boldsymbol{\gamma}}|$ in (9) using Lemma 1. The set $S_{\boldsymbol{\gamma}}(D)$ is defined as

$$S_{\boldsymbol{\gamma}}(D) = \left\{\boldsymbol{\lambda} : \lambda_{(0)(1)} + \lambda_{(1)(0)} \geq D, \boldsymbol{\gamma}_{\boldsymbol{a}} = \sum_{\boldsymbol{b}\in\{0,1\}^c} \lambda_{(\boldsymbol{a})(\boldsymbol{b})}\right\}. \tag{28}$$

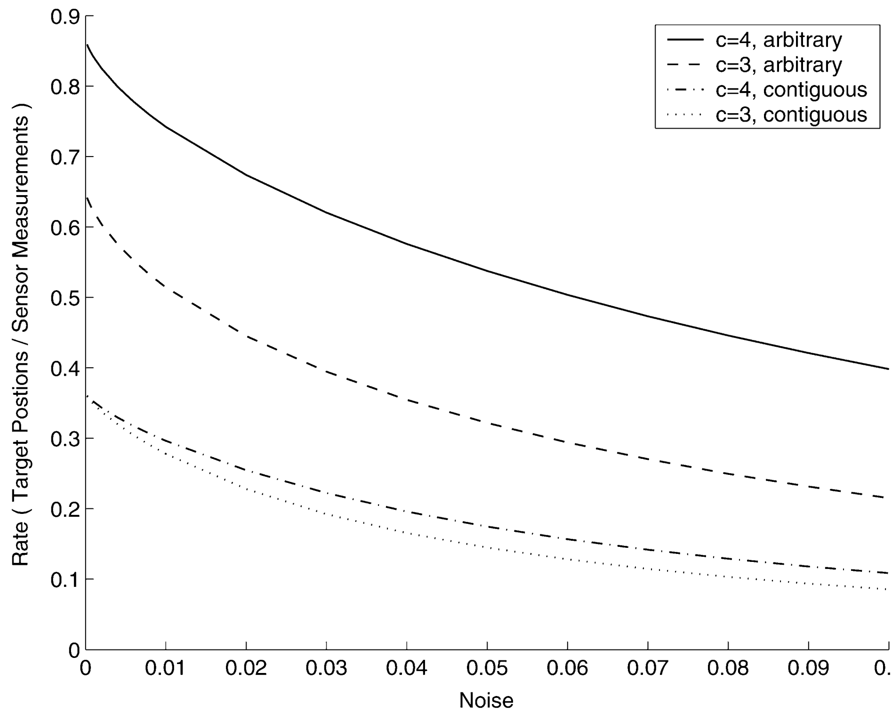

Fig. 10. $C_{LB}(0.025)$ for localized and nonlocalized sensors.

Given these new bounds and definitions, and the substitution of c-order types for types, the proof of Theorem 1 can be applied directly to prove Theorem 2.

C. Numerical Results

In Fig. 10, we compare $C_{LB}(D = 0.025)$ for sensor networks with localized (i.e., contiguous connections model) and nonlocalized (i.e., arbitrary connections model) sensing. We assume that the sensing function Ψ is a weighted additive function, with weights $\{1, 0.5, 0.25, 0.1\}$ for $c = 4$ and $\{1, 0.5, 0.25\}$ for $c = 3$. The sensor noise model used throughout this section assumes that the probability of error decays exponentially with the error magnitude. In the figures, '$\text{Noise} = p$' indicates that for a sensor, $P(Y \neq X) = p$, with $\mathcal{Y} = \mathcal{X}$ assumed. Contiguous sensor field of view causes a significant reduction in sensing capacity. We conjecture that this effect is similar to the inferior performance of channel codes that have finite memory, such as convolutional codes, as opposed to LDPC codes. Further, it is interesting to note that the gap in sensing capacity between sensors of range $c = 3$ and $c = 4$ is larger for the arbitrary connections model than the contiguous connections model.

To compute the bound shown in Theorem 2, we solve a sequence of convex optimization problems. Rather than computing the bound directly, we find the largest R for which the minimum of $f(\boldsymbol{\lambda}) = D(P^{\boldsymbol{\gamma}}_{X_iY} \| Q^{\boldsymbol{\lambda}}_{X_iY}) - R(H(\tilde{\boldsymbol{\lambda}}|\boldsymbol{\lambda}') - H(\tilde{\boldsymbol{\lambda}}|\boldsymbol{\gamma}'))$ over all valid $\boldsymbol{\lambda}$ is greater than 0. Minimizing $f(\boldsymbol{\lambda})$ is a convex optimization problem since $f(\boldsymbol{\lambda})$ is convex in $\boldsymbol{\lambda}$ and the set of valid $\boldsymbol{\lambda}$ is convex. Since $H(\tilde{\boldsymbol{\gamma}}|\boldsymbol{\gamma}') = 1$, the convexity of $f(\boldsymbol{\lambda})$ in $\boldsymbol{\lambda}$ can be proven using the log-sum inequality and the concavity of entropy.

D. Extension to Two-Dimensional Fields

The sensing capacity bounds obtained in this section can be extended from discrete target vectors to two dimensional 'target fields.' This extension requires the introduction of two dimensional types. Such types are histograms over the set of possible two dimensional patterns. We first analyzed the sensing capacity for a contiguous connections model in [47].

Fig. 11 shows an example of our sensor network model. The state of the environment is modeled as a $k \times k$ grid with k^2 spatial positions. Each discrete position may contain no target or one target, and therefore the target configuration is represented by a k^2-bit target field $\boldsymbol{f}$. The possible target fields are denoted $\boldsymbol{f_i}, i \in \{1, \ldots, 2^{k^2}\}$. Target fields occur with equal probability. The sensor network has n identical sensors. Sensor ℓ located at grid block F_h senses a set of discrete target positions in the grid within a Euclidean distance c of the sensor's grid location. We assume circular boundary conditions, where locations on the edge of the grid are considered adjacent to grid locations on the opposite edge of the grid. Fig. 11 depicts sensors with range $c = 1$. $\mathcal{S}_{c,h}$ is the coverage of a sensor located at grid block F_h with range c, and is defined as the set of target positions within distance c of the sensor. For example, in Fig. 11, sensor 0 is located at F_7 and has range $c = 1$. Its coverage is $\mathcal{S}_{1,7} = \{F_6, F_7, F_8, F_2, F_{12}\}$. Each sensor outputs a value $x \in \mathcal{X}$ that is an arbitrary function of the targets which it senses, $x = \Psi(f_{t_1}, \ldots, f_{t_{|\mathcal{S}_{c,h}|}})$. Due to the assumption of circular boundary conditions, the number of targets sensed by a sensor depends only on the sensor range and not on a sensor's location. We write the number of target positions sensed by a sensor of range c as $|\mathcal{S}_c|$. We assume a simple model for randomly generating sensor networks, where each sensor chooses a region of

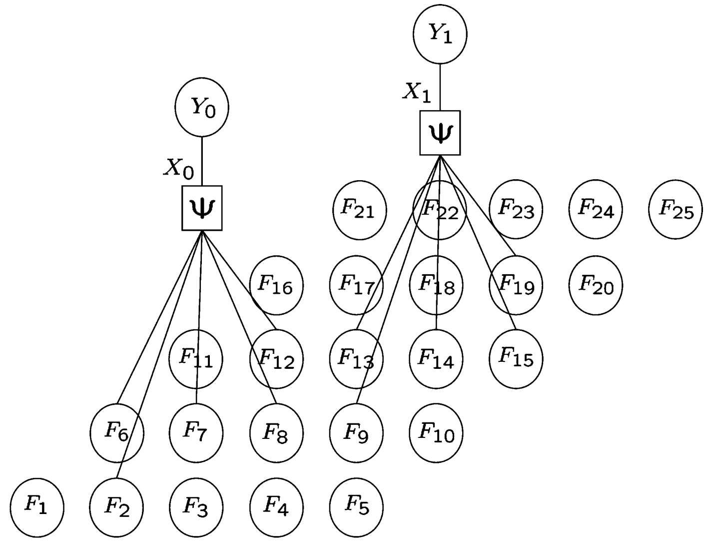

Fig. 11. Sensor network model with $k = 5, n = 2, c = 1$.

Euclidean radius c with equal probability among the set of possible regions of radius c. This would occur, for example, if sensors were randomly dropped on a field. All definitions from the one-dimensional contiguous model extend directly, with target vectors $\boldsymbol{v}$ replaced by fields $\boldsymbol{f}$. The rate is defined as $R = \frac{k^2}{n}$.

For a sensor located randomly in the target field, the probability of a sensor producing a value depends on the number of target patterns that correspond to that value in the sensor's range, and thus, can be written as a function of the frequency of patterns in the field. The type $\boldsymbol{\gamma}_i$ is a vector that corresponds to the normalized counts over the set of possible target configurations in the sensor's field of view in a field $\boldsymbol{f}_i$. For a sensor of range $c, \boldsymbol{\gamma}_i$ is a $2^{|\mathcal{S}_c|}$ dimensional vector, where each entry in the vector $\boldsymbol{\gamma}_i$ corresponds to the frequency of occurrence of one of the possible $|\mathcal{S}_c|$ bit patterns. The set of sensor types $\boldsymbol{\gamma}$ of a $k \times k$ field is denoted $\mathcal{P}_k^2(\{0,1\}^{|\mathcal{S}_c|})$. $\gamma_{(0)}$ and $\gamma_{(1)}$ are the number of zeros and ones, respectively, in a vector of type $\boldsymbol{\gamma}$. These quantities can be directly computed from $\boldsymbol{\gamma}$.

Next, we note that for sensor of range c the conditional probability $P_{\boldsymbol{X_i X_j}}$ depends on the joint type $\boldsymbol{\lambda}$ of the ith and jth target fields $\boldsymbol{f_i}, \boldsymbol{f_j}$. For $\boldsymbol{a}, \boldsymbol{b} \in \{0,1\}^{|\mathcal{S}_c|}, \boldsymbol{\lambda}$ is the matrix of $\lambda_{(\boldsymbol{a})(\boldsymbol{b})}$, the fraction of positions in $\boldsymbol{f_i}, \boldsymbol{f_j}$ where $\boldsymbol{f_i}$ has a target pattern $\boldsymbol{a}$ while $\boldsymbol{f_j}$ has a target pattern $\boldsymbol{b}$. We denote the set of all joint sensor types for sensors of range c observing a target field of area k^2, as $\mathcal{P}_k^2(\{0,1\}^{|\mathcal{S}_c|}, \{0,1\}^{|\mathcal{S}_c|})$. Since the output of each sensor depends only on the contiguous region of targets which it senses, $P_{\boldsymbol{X_i X_j}}$ depends only on $\boldsymbol{\lambda}$ (discussed in Section V-A). $\lambda_{(1)(0)}$ is the number of grid locations where field i has a target and field j does not, and can be computed directly from $\boldsymbol{\lambda}$. $\lambda_{(0)(1)}$ is similarly defined and computed.

Using the definitions of two dimensional types in the definitions of $P_{X_iY}^{\boldsymbol{\gamma}}$ and $Q_{X_iY}^{\boldsymbol{\lambda}}$ from the one-dimensional contiguous connections model, we can prove the following bound for sensing a field. The sensing capacity at distortion D satisfies,

$$C(D) \geq C_{LB}(D) = \min_{\substack{\boldsymbol{\lambda} \\ \lambda_{(0)(1)}+\lambda_{(1)(0)} \geq D}} \frac{D\left(P_{X_iY}^{\gamma_i} \| Q_{X_iY}^{\boldsymbol{\lambda}}\right)}{H((\gamma_{j(0)}, \gamma_{j(1)}))} \tag{29}$$

where $\boldsymbol{\gamma}_i, \boldsymbol{\gamma}_j \in \mathcal{P}^2(\{0,1\}^{|\mathcal{S}_c|}), \gamma_{i(0)} = 0.5$ and $\gamma_{i(1)} = 0.5$, and $\boldsymbol{\lambda} \in \mathcal{P}^2(\{0,1\}^{|\mathcal{S}_c|}, \{0,1\}^{|\mathcal{S}_c|})$.

Proof Outline: The proof is essentially identical to the proof of Theorem 1, with types and joint types replaced by types and joint types. For types, we bound α as follows:

$$\alpha(\boldsymbol{\gamma}_i, k) \leq 2^{k^2 H((\gamma_{i(0)}, \gamma_{i(1)}))}. \tag{30}$$

For joint types, we bound β as

$$\beta(\boldsymbol{\lambda}, k) \leq 2^{k^2 H((\gamma_{j(0)}, \gamma_{j(1)}))}. \tag{31}$$

The bounds on α and β are loose, and the authors are not aware of tighter combinatorial bounds for types. The set $S_{\boldsymbol{\gamma}}(D)$ is defined as in (28). Given these new bounds and definitions, and the substitution of 2D types for types, the proof of Theorem 1 can be applied directly to prove this result.

VI. Conclusion and Discussion

The results presented in this paper provide limits on the accuracy of sensor networks for large-scale detection applications. These results are obtained by drawing on an analogy between channel coding and sensor networks. We define the sensing capacity and lower bound it for several sensor network models. For all rates below the sensing capacity, detection to within a desired accuracy with arbitrarily small error is achievable. This threshold behavior contrasts with classical detection problems, where probability of error goes to zero as the number of sensor measurements go to infinity while the number of hypotheses remains fixed [43]. The sensing capacity captures complex sensor tradeoffs. For example, our bounds show that the efficiency of using long range, noisy sensors or shorter range, less noisy sensors depends on the desired detection accuracy. Further, our results show that the mutual information is not the correct notion of information for the large-scale detection problems considered in this paper. This has implications for the problem of sensor selection due to the popularity of "information gain" as a sensor selection metric.

An important contribution of this paper is its demonstration of a close connection between sensor networks and communication channels. It is thought-provoking to consider that one could apply insights from the large body of work available for communication channels to the sensor network setting. For example, channel coding theory contains a large number of results that are used to build practical communication systems. Can we fruitfully apply ideas from coding theory to sensor networks? To demonstrate the potential benefit of a channel coding perspective, in [48], [49] we proposed extending ideas from convolutional coding to sensor networks. We demonstrated that a version of sequential decoding (a low complexity decoding heuristic for convolutional codes) can be applied to detection in sensor networks, as an alternative to the belief propagation algorithm. Our empirical results indicate that above a certain number of sensor measurements, the sequential decoding algorithm achieves accurate decoding with bounded computations per bit (target position). This empirical result suggests the existence of a "computational cut-off rate," similar to one that exists for channel codes.

Our work on the theory of sensing points to a large set of open problems on large-scale detection. Obvious directions include strengthening the theory by considering alternative settings of the problem, tightening the sensing capacity bounds, and proving a converse to sensing capacity. For example, we presented extensions to the work presented in this paper by considering the impact of spatial [47] and temporal [50] dependence on the sensing capacity. Another direction for future work is to explore the connection between sensor networks and communication channels, including the exploitation of existing channel codes to design sensor networks.

REFERENCES

[1] S. Thrun, , G. Lakemeyer and B. Nebel, Eds., "Robotic mapping: A survey," in *Exploring Artificial Intelligence in the New Millenium*. San Mateo, CA: Morgan Kaufmann, 2002.

[2] A. Elfes, "Occupancy grids: A probabilistic framework for mobile robot perception and navigation," Ph.D. dissertation, Elect. Comput. Eng. Dep., Carnegie Mellon Univ., Pittsburgh, PA, 1989.

[3] R. Collins, A. Lipton, T. Kanade, H. Fujiyoshi, D. Duggins, Y. Tsin, D. Tolliver, N. Enomoto, and O. Hasegawa, Robot. Inst., Carnegie Mellon Univ., Pittsburgh, PA, A System for Video Surveillance and Monitoring Tech. Rep. CMU-RI-TR-00-12, May 2000.

[4] A. Hoover and B. Olsen, "A real-time occupancy map from multiple video streams," in *Proc. Int. Conf. Robot. Autom.*, 1999.

[5] M. Burl, B. Sisk, T. Vaid, and N. Lewis, "Classification performance of carbon black-polymer composite vapor detector arrays as a function of array size and detector composition," *Sens. Actuators B*, vol. 87, pp. 130–149, 2002.

[6] T. C. Pearce and M. A. Sanchez-Montanes, "Chemical sensor array optimization: Geometric and information theoretic approaches," *Handbook of Mach. Olfact.*, pp. 347–375, 2002.

[7] D. Li, K. Wong, Y. Hu, and A. Sayeed, "Detection, classification and tracking of targets in distributed sensor networks," *IEEE Signal Process. Mag.*, pp. 17–29, Mar. 2002.

[8] Y. Tian and H. Qi, "Target detection and classification using seismic signal processing in unattended ground sensor systems," in *Proc. Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, May 2002, vol. 4.

[9] C. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, 623–656, Jul., Oct. 1948.

[10] J. Moura, R. Negi, and M. Pueschel, "Distributed sensing and processing: A graphical model approach," in *Proc. DARPA ACMP Integr. Sensing and Process. Workshop*, Annapolis, MD, Sep. 2002.

[11] Y. Rachlin, R. Negi, and P. Khosla, "Sensing capacity for target detection," in *Proc. IEEE Inf. Theory Wksp.*, Oct. 24–29, 2004.

[12] J. Manyika and H. Durrant-Whyte*, Data Fusion and Sensor Management: A Decentralized Information-Theoretic Approach*. Englewood Cliffs, NJ: Prentice-Hall, 1994.

[13] I. Csiszar, "The method of types," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, 1998.

[14] E. J. Duarte-Melo and M. Liu, "Data-gathering wireless sensor networks: Organization and capacity," *Comput. Netw.: Special Issue on Wireless Sens. Netw.*, vol. 43, 2003.

[15] P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Trans. Inf. Theory*, vol. 46, no. 2, 2000.

[16] G. Barrenechea, B. Beferull-Lozano, and M. Vetterli, "Lattice sensor networks: Capacity limits, optimal routing and robustness to failures," in *Proc. 3rd Int. Symp. Inf. Process. in Sens. Netw.*, Apr. 2004.

[17] U. Mitra and A. Sabharwal, "Complexity constrained sensor networks: Achievable rates for two relay networks and generalizations," in *Proc. 3rd Int. Symp. Inf. Process. in Sens. Netw.*, Apr. 2004.

[18] Z. Hu and B. Li, , Y. Xiao and Y. Pan, Eds., "Fundamental performance limits of wireless sensor networks," in *Ad Hoc and Sensor Networks*. Commack, NY: Nova Science , 2004.

[19] D. Slepian and J. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inf. Theory*, vol. 19, pp. 471–480, 1973.

[20] A. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the receiver," *IEEE Trans. Inf. Theory*, vol. 22, pp. 1–10, 1976.

[21] S. Pradhan, J. Kusuma, and K. Ramachandran, "Distributed compression in a dense microsensor network," *IEEE Signal Process. Mag.*, vol. 19, pp. 51–60, Mar. 2002.

[22] Z. Xiong, A. Liveris, and S. Cheng, "Distributed source coding for sensor networks," *IEEE Signal Process. Mag.*, vol. 21, pp. 80–94, 2004.

[23] P. Varshney*, Distributed Detection and Data Fusion*. New York: Springer-Verlag, 1997.

[24] J. Chamberland and V. Veeravalli, "Decentralized detection in sensor networks," *IEEE Trans. Signal Process.*, vol. 51, no. 2, pp. 407–416, 2003.

[25] J. Chamberland and V. Veeravalli, "Asymptotic results for decentralized detection in power constrained wireless sensor networks," *IEEE J. Sel. Areas Commun. (Special Issue on Wireless Sens. Netw.)*, vol. 22, no. 6, pp. 1007–1015, 2004.

[26] A. D'Costa, V. Ramachandran, and A. Sayeed, "Distributed classification of Gaussian space-time sources in wireless sensor networks," *IEEE J. Sel. Areas Commun. (Special Issue on Fund. Perform. Limits of Wireless Sens. Netw.)*, pp. 1026–1036, Aug. 2004.

[27] J. Kotecha, V. Ramachandran, and A. Sayeed, "Distributed multi-target classification in wireless sensor networks," *IEEE J. Sel. Areas Commun. Special Issue on Self-Organizing Distrib. Collabor. Sens. Netw.*, 2005.

[28] K. Chakrabarty, S. S. Iyengar, H. Qi, and E. Cho, "Coding theory framework for target location in distributed sensor networks," in *Proc. Int. Conf. Inf. Technol.: Coding and Comput.*, Apr. 2001.

[29] A. Scaglione and S. D. Servetto, "On the interdependence of routing and data compression in multi-hop sensor networks," in *Proc. 8th ACM Int. Conf. Mobile Comput. Netw.*, Sept. 2002.

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

[31] D. Marco, E. Duarte-Melo, M. Liu, and D. Neuhoff, "On the many-to-one transport capacity of dense wireless sensor networks and the compressibility of its data," in *Proc. Inf. Process. Sens. Netw.*, 2003, pp. 1–16.

[32] P. Ishwar, A. Kumar, and K. Ramachandran, "Distributed sampling for dense sensor networks: A bit-conservation principle," in *Inf. Process. Sens. Netw.*, 2003.

[33] W. Bajwa, A. Sayeed, and R. Nowak, "Matched source-channel communication for field estimation in wireless sensor networks," in *Proc. 4th Int. Symp. Inf. Process. Sens. Netw.*, Apr. 25–27, 2005.

[34] A. Kumar, P. Ishwar, and K. Ramchandran, "On distributed sampling of smooth non-bandlimited fields," in *Proc. Int. Symp. Inf. Process. Sens. Netw.*, Apr. 2004.

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

[36] S. Aeron, M. Zhao, and V. Sailgrama, "On sensing capacity of sensor networks for the class of linear observation, fixed SNR models," in *Proc. Inf. Theory Workshop*, 2007.

[37] S. Aeron, M. Zhao, and V. Sailgrama, "On sensing capacity of sensor networks for a class of linear sensing models," in *Proc. Statist. Signal Process. Workshop*, 2007.

[38] J. Moura, J. Liu, and M. Kleiner, "Intelligent sensor fusion: A graphical model approach," in *Proc. IEEE Int. Conf. Signal Process.*, Apr. 2003.

[39] J. Moura, R. Negi, and M. Pueschel, "The network as the sensor, distributed sensing and processing: A graphical model approach," in *DARPA ISP Rev.*, St. Petersburg, FL, Oct. 2003.
[40] R. Gallager, *Information Theory and Reliable Communications*. New York: Wiley, 1968.
[41] R. Gallager, *Low-Density Parity-Check Codes*. Cambridge: M.I.T. Press, 1963.
[42] D. Mackay, *Information Theory, Inference, and Learning Algorithms*. Cambridge, U.K.: Cambridge Univ. Press, 2003.
[43] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley-Intersci., 1991.
[44] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Mateo, CA: Morgan Kaufmann, 1988.
[45] Y. Rachlin, R. Negi, and P. Khosla, "Sensing capacity for discrete sensor network applications," in *Proc. 4th Int. Symp. Inf. Process. Sens. Netw.*, Apr. 25–27, 2005.
[46] L. Davisson, G. Longo, and A. Sgarro, "The error exponent for the noiseless encoding of finite ergodic Markov sources," *IEEE Trans. Inf. Theory*, vol. 27, pp. 431–438, 1981.
[47] Y. Rachlin, R. Negi, and P. Khosla, "Sensing capacity for Markov random fields," in *Proc. Int. Symp. Inf. Theory*, 2005.
[48] Y. Rachlin, R. Negi, and P. Khosla, "On the interdependence of sensing and estimation complexity in sensor networks," in *Proc. 5th Int. Conf. Inf. Process. Sens. Netw.*, Apr. 19–21, 2006.
[49] Y. Rachlin, B. Narayanaswamy, R. Negi, J. Dolan, and P. Khosla, "Increasing sensor measurements to reduce detection complexity in large-scale detection applications," in *Proc. Military Commun. Conf.*, 2006.
[50] Y. Rachlin, R. Negi, and P. Khosla, "Temporal sensing capacity," in *Proc. Allerton Conf. Commun., Contr., Comput.*, 2006.

Yaron Rachlin (S'98–M'08) received the B.S. degrees in electrical engineering and mathematics from Virginia Polytechnic Institute, Blacksburg, in 2000, and the M.S. and Ph.D. degrees in electrical and computer engineering in 2002 and 2007 from Carnegie Mellon University, Pittsburgh, PA.

He is currently a Senior Member of Technical Staff at Draper Laboratory, Cambridge, MA. His research interests include the application of information and coding theories to detection and estimation problems, low-power sensor systems, and data privacy and security.

Rohit Negi (S'98–M'00) received the B.Tech. degree in electrical engineering from the Indian Institute of Technology, Bombay, in 1995. He received the M.S. and Ph.D. degrees from Stanford University, Stanford, CA, in 1996 and 2000, respectively, both in electrical engineering.

Since 2000, he has been with the Electrical and Computer Engineering Department, Carnegie Mellon University, Pittsburgh, PA, where he is a Professor. His research interests include signal processing, coding for communication systems, information theory, networking, cross-layer optimization, and sensor networks.

Dr. Negi received the President of India Gold Medal in 1995.

Pradeep K. Khosla (F'95) received the B.Tech. degree (honors) from the Indian Institute of Technology, Kharagpur, and the M.S. and Ph.D. degrees from Carnegie Mellon University, Pittsburgh, PA, in 1984 and 1986, respectively.

He is currently Dean of the College of Engineering, and the Philip and Marsha Dowd University Professor at Carnegie Mellon. His previous positions include—Founding Director, Carnegie Mellon CyLab, Head, Department of Electrical and Computer Engineering, Director, Information Networking Institute, Founding Director, Institute for Complex Engineered Systems (ICES), and Program Manager at Defense Advanced Researh Projects Agency (DARPA) where he managed a portfolio of programs in real-time systems, internet enabled software infrastructure, intelligent systems, and distributed systems.

Dr. Khosla received several awards including the ASEE George Westinghouse Award for Education in 1999, Siliconindia Leadership award for Excellence in Academics and Technology in 2000, the W. Wallace McDowell award from IEEE Computer Society in 2001, and Cyber Education Award from the Business Software Alliance (2007). He was awarded the Philip and Marsha Dowd Professorship in 1998, and named University Professor in 2008. He has been elected a Fellow of the American Association of Artificial Intelligence (AAAI) in 2003, Fellow of American Association for Advancement of Science (AAAS) in 2004, and a member of the National Academy of Engineering (NAE) in 2006.