Robust detection of random variables using sparse measurements

Balakrishnan Narayanaswamy, Rohit Negi, Pradeep Khosla
Department of Electrical & Computer Engineering, Carnegie Mellon University, Pittsburgh, PA, USA
Email: {bnarayan,negi,pkk@ece.cmu.edu}

***Abstract*— We look at the problem of estimating k discrete random variables from n noisy and *sparse* measurements where $k = nR$, with a 'rate' R. The model is motivated by problems studied in diverse areas including compressed sensing, group testing, multiple access channels and sensor networks. In particular, we study uncertainty and mismatch in the measurement functions and the noise model and quantify the effect of these faults on detection performance, in the large system limit as $n \to \infty$, while R remains constant. We characterize the performance of mismatched and uncertain detectors, design and analyze robust detectors and present an illustrative example where the analysis presented can be used to guide the design of robust measurement ensembles.**

I. INTRODUCTION

We study the problem of detecting k discrete random variables $\vec{\mathbf{V}} = \left[V^1, \ldots, V^k\right]$, from n noisy and *sparse* measurements $\vec{\mathbf{Y}} = \left[Y^1, \ldots, Y^n\right]$. A measurement Y^u is sparse if it is a function of a small ($\Theta(1)$) number of V^js. Sparse measurements are attractive because they are often associated with low-complexity detection algorithms that have very good performance guarantees [1]. The problem of estimating discrete variables using sparse observations [1] has been studied from many different points of view including sensor networks [2], combinatorial group testing [3], sparsely spread Code Division Multiple Access (CDMA) systems [4] and sparse compressed sensing [5].

Rachlin et al. [2] analyzed this problem from an information theoretic perspective, in the large system limit ($n \to \infty$), with $k = nR$ for a constant rate R. Their work showed that there is a strong parallel between the problem of detecting k discrete random variables using n sparse measurements and the channel coding problem of communicating k discrete symbols across a noisy channel using a code of length n with a sparse generator matrix. Furthermore, [2] showed that the detection problem could be analyzed using Gallager's bounding method [6]. They obtained a characterization of the rate R that is sufficient to detect k discrete random variables.

In this paper we are interesting in analyzing the effect of parameter uncertainty and model mismatch in this detection problem. To do so, we generalize the result in [2] to allow for a large class of (possibly sub-optimal) decoders. We show that with this more general derivation, the parallels to channel coding can be extended even further. We begin by modifying the analysis for universal decoders from [7] and mismatched decoders from [8] to analyze the detection task. This generalization allows us to quantify the effect of mismatch or uncertainty in the noise model and measurement function and design robust measurement ensembles.

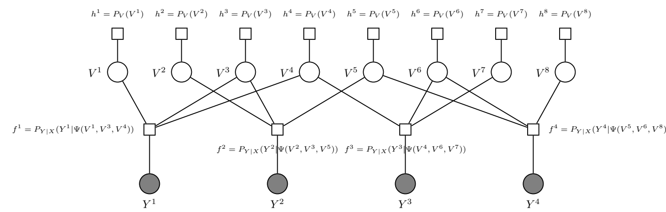

Fig. 1. Factor graph [9] of detecting variables from sparse measurements with $c = 3$, $k = 8$, $n = 4$, and the rate $R = 2$. f^u are measurement factors and h^j factors correspond to prior information.

A. Model outline

The model we study can be represented using a factor graph [9] as in Fig. 1. The state of the environment is modeled by k discrete random variables $\vec{\mathbf{V}} = \left[V^1, \ldots, V^k\right]$. Each position in the vector represents a discrete phenomenon at that 'location', which takes values in $\mathcal{V}$. Each measurement $u \in \{1, \ldots, n\}$ is *sparse* i.e., a function of c locations $l(u,1), \ldots, l(u,c)$. We call the measurements and locations measured by them a sensor network or measurement network. Each noiseless function value $X^u \in \mathcal{X}$, is a function $X^u = \Psi\left(V^{l(u,1)}, \ldots, V^{l(u,c)}\right)$. Each measurement is corrupted independently according to a probability mass function (p.m.f.) $p_{Y|X}\left(Y = y | X = x\right)$ resulting in the observations $\vec{\mathbf{Y}} \in \mathcal{Y}^n$. We use a 'decoder' $g : \mathcal{Y}^n \to \mathcal{V}^k$ to obtain an estimate $\hat{\vec{\mathbf{V}}} = g(\vec{\mathbf{Y}})$ of the true state of $\vec{\mathbf{V}}$. We quantify the performance of (possibly sub-optimal) decoders g, with uncertain or mis-specified noise model $P_{Y|X}$ or measurement function Ψ. We study the problem in the large system limit, ($n \to \infty$), while the rate R is maintained a constant.

B. Motivation for the model

We now review some applications that can be cast into the framework described above. The model is nearly identical to the one studied by [1], and so, the possible applications of our results are naturally very similar. Montanari [1] focuses on the analysis of message passing algorithms using density evolution and is not concerned with the robustness of the algorithms. However, as explained in the next section we characterize the performance of a large class of decoders, especially with model mismatch and parameter uncertainty.

- **Luby Transform (LT) codes** [10] are low-density generator matrix (LDGM) codes which are universal and rateless erasure codes that show promise in many network applications. In an LT code, each encoding symbol chooses a degree c from a degree distribution. It then chooses c message bits as neighbors. The encoded symbol is then the XOR of the chosen bits. Putting this into the model we study is very simple - the discrete V^js are the 'message bits' in the LT code and the function Ψ is XOR. The noise model $P_{Y|X}$ is the erasure channel over which we use the LT codes where $\mathcal{Y}$ is a tertiary alphabet $\mathcal{Y} = \{0, 1, *\}$, with $*$ used for erasures. We described the model above for fixed c but we show in Section IV how to extend it to the case of multiple cs.

- **Group testing** [11] is an old technique which was used to reduce the number of blood tests required to detect syphilis in soldiers during World War II. The blood from multiple people is mixed together to form a 'pool' that is tested. If any of the soldiers in that pool are 'positive' for syphilis the sample will test positive. To fit this into the model, we have one V^j for each solider which takes a value 1 if he is infected and 0 otherwise. The function Ψ is OR. Sources of errors such as human error and chemical inhibitors, are modeled using the noise model $P_{Y|X}$. Recent work [12] explores the use of sparse random pools for the group testing problem. Many different versions of this problem have been studied (as detailed in [11]) for a variety of applications such as DNA clone library screening, circuit testing or multi-access channels. These models differ in the Ψ function or the noise model, but are generally considered as group testing applications.
- **Multi-user detection** [13] is the problem of decoding the transmissions of multiple users using a shared channel resource. In general the channel can be modeled as a vector channel $\mathbf{H}$ with $\vec{\mathbf{Y}} = \mathbf{H}\vec{\mathbf{V}} + \vec{\mathbf{N}}$. In binary CDMA the alphabet of the V^j variables to be estimated is $\mathcal{V} = \{1, -1\}$. Each output Y^u is then a *linear* function of the discrete V^j. This falls very naturally into our framework when $\mathbf{H}$ is row sparse, as is the case for sparsely spread CDMA [4], in which the output at each Y^u is a linear function of a few of the binary V^js. The noise model $P_{Y|X}$ is usually assumed to be Gaussian, modeling the additive noise in the channel.

C. Focus of this paper

Prior theoretical work [1], [2] on problems similar to the one described above assume that the noise model $P_{Y|X}$ and sensing function Ψ are known perfectly at the decoder. However, in real systems this assumption can easily be violated. For example, when using LT codes [10] in networks, the erasure probability (corresponding to $P_{Y|X}$) is usually not known. In CDMA systems [4] the Ψ and $P_{Y|X}$ corresponding to the channel model and noise may only be known to be in a certain class, or the true system parameters may change over time. In applications such as group testing [11] and DNA micro-arrays [14], similar problems of uncertainty and mismatch can arise. Motivated by these examples, we analyze the effect of particular fixed measurement functions, noise and especially uncertainty and mismatch in the noise model $P_{Y|X}$ and measurement function Ψ.

In particular, we seek to answer the following questions:

- How do we design the decoder $g(\vec{\mathbf{Y}})$ if the noise model $P_{Y|X}$ or the measurement function Ψ is uncertain? What is the performance of this decoder?
- What is the performance of the decoder $g(\vec{\mathbf{Y}})$ when $P_{Y|X}$ or Ψ are mis-calibrated? That is, what is the performance of a decoder that is provided with an incorrect noise model or measurement function?
- How can we design measurement ensembles that are robust to uncertainty and model mismatch?

We approach the problem from an information theoretic point of view, following [2]. However, they were interested in the performance of optimal Maximum Likelihood (ML) or Maximum-a-posteriori (MAP) decoders and did not look at the problems of uncertainty and mismatch. In order to analyze sub-optimal decoders and address the problem of robustness, we first simplify and generalize the original derivation of sensing capacity in [2] following the 1961 Gallager-Fano bounding technique (as described [8]). We then use large deviation results from [15] to derive our main general result (Theorem 3.1). Next, we study universal decoders from [7] and analyze performance when the noise model $P_{Y|X}$ or the measurement function Ψ is uncertain. We also analyze sub-optimal decoders and decoders using mis-specified Ψ or $P_{Y|X}$ using techniques reviewed in [8]. Our analysis allows us to quantify the effect of calibration faults and uncertainty, and suggest cases where modified algorithms need to be used to compensate for them. We conclude that if a sufficient number of measurements are available (so that the sensing capacity remains positive), then even in these cases, performance guarantees can be made. We also show an example, where the analysis presented can be used to guide the design of robust measurement ensembles.

D. Comparison to prior work

In this paper we start from the ideas in [2]. However the original work did not look at mismatch and uncertainty and was for optimal Maximum Likelihood (ML) or Maximum-a-posteriori (MAP) decoders. To analyze mismatch, uncertainty and new decoders, we are forced to follow a different analysis technique. We provide a more general derivation, resulting in an expression that can be easily specialized to analyze these new problems. As pointed out in [2], the model we study is different from the set up of classical channel coding problems [6] because the randomly generated 'codewords' (or noiseless measurements) $\vec{\mathbf{X}}$ are not independent for different 'messages' (or environments) $\vec{\mathbf{v}}$. This is because, in our case, all the 'codewords' come from the same fixed generative model, resulting in constraints on the possible codewords. Also, because of the finite 'field of view' c of the measurements, each 'message bit' (location) can only affect a constant number of output measurements. For example, suppose we know the noiseless measurements for an environment $\vec{\mathbf{v}}_m$. Consider an environment $\vec{\mathbf{v}}_{m'}$ generated by changing the value at one location in $\vec{\mathbf{v}}_m$. Because of the sparse nature of the measurements, we expect that, in general, very few measurements will change i.e., the noiseless measurements $\vec{\mathbf{X}}_m$ and $\vec{\mathbf{X}}_{m'}$ will be similar. In (ideal) codes on the other hand, there is no reason for two codewords to be similar to each other. This difference causes a significant change in the techniques used.

Montanari [1] also studies a model very similar to the one we look at, but there are some fundamental differences between his approach and ours. Firstly, the case of model mismatch and uncertainty was not addressed there which is central to our work. Additionally, in that work the focus was on analyzing the performance of message passing algorithms for the detection problem. The analysis was performed through 'density evolution', which required that each measurement function is invariant to a permutation of the measured locations. The analysis results in estimates of the *symbol* error rate while we seek to provide a single letter characterization of the number of measurements (or rate R), that is sufficient to achieve a particular overall (or *word* level) distortion. Finally, in [1] the measurement networks are assumed to be actually drawn from a random distribution, whereas we use random measurements as a proof technique in the spirit of Shannon's random coding argument [6]. We are interested in the existence of good measurement networks $\mathcal{C}_n$, while [1] analyses the performance of inference algorithms on random graphs.

The rest of the paper is organized as follows. In Section II we define the notation we need for our analysis. Section III contains our main result, while its proof is relegated to the Appendix. Section IV specializes the result of Section III, and shows how our analysis is general enough to easily analyze parameter uncertainty and model mismatch. Finally, in Section V, we show a simple example highlighting how our results could be used to design measurement ensembles that are robust to model mismatch.

II. Model and Notation : Types, joint types and Conditional distributions

Notational preliminaries : In this paper, we use upper case (X) to denote a random variable and lower case (x) to denote its values. The notation $\vec{\mathbf{X}}$ denotes a random vector. We index into a vector using superscripts, and enumerate vectors using subscripts - X_m^j is the j^{th} element of the m^{th} vector $\vec{\mathbf{X}}_m$. All logs are to base e.

Model : The problem we study [2] can be modeled using the factor graph, as shown in Fig. 1 and described in Section I-A. For completeness, some terms are defined again. The 'environment' or 'target vector' is modeled as a k-dimensional discrete random vector $\vec{\mathbf{V}}$. Possible states of the environment are denoted by $\vec{\mathbf{v}}_m$, $m \in 1, 2, \ldots, |\mathcal{V}|^k$. We assume that nature generates elements of $\vec{\mathbf{V}}$ independently at random with probability mass function (p.m.f.) $p_V(V = v)$ (these correspond to the h^j factors in the factor graph). We are given n measurements and define the rate $R = \frac{k}{n}$ as the ratio of the number of random variables to be measured (k) to the number of measurements (n). Each measurement $u \in \{1, \ldots, n\}$ is *sparse* - that is, a function of c arbitrary locations $l(u,1), \ldots, l(u,c)$. We denote the resulting measurement network or sensor network as $\mathcal{C}_n$. For measurement u, we denote the (unobservable) noiseless function value $X^u \in \mathcal{X}$, which is a function $X^u = \Psi\left(V^{l(u,1)}, \ldots, V^{l(u,c)}\right)$, where Ψ is the 'measurement function' or 'sensing function'. Each X^u is then corrupted by independent identically distributed noise that follows a p.m.f. $p_{Y|X}(Y = y|X = x)$, to obtain the observed measurements $\vec{\mathbf{Y}} \in \mathcal{Y}^n$. Since the noise in the measurements is independent and identical, $P_{\vec{\mathbf{Y}}|\vec{\mathbf{X}}}\left(\vec{\mathbf{Y}} = \vec{\mathbf{y}}|\vec{\mathbf{X}} = \vec{\mathbf{x}}\right) = \prod_{u=1}^n p_{Y|X}(Y^u = y|X^u = x)$. We define a 'decoder' by a function $g : \mathcal{Y}^n \to \mathcal{V}^k$.

Error events and Sensing capacity : We define a tolerable distortion region around a particular environment vector $\vec{\mathbf{v}}_m$ as $\mathcal{D}_{\vec{\mathbf{v}}_m} = \left\{m' : \frac{1}{k} d_H(\vec{\mathbf{v}}_{m'}, \vec{\mathbf{v}}_m) < D\right\}$ for some distortion $D \in [0,1]$ where $d_H(\bullet,\bullet)$ is the Hamming distance function. This corresponds to the set of all environments within a Hamming distance of D of $\vec{\mathbf{v}}_m$. Motivated by applications, we wish to analyze under what conditions it is possible to guarantee that when $\vec{\mathbf{v}}_m$ is the true state of the environment, the decoded $g\left(\vec{\mathbf{Y}}\right) = \widehat{\vec{\mathbf{V}}} \in \mathcal{D}_{\mathbf{v}_m}$. Formally, we define an error event for a given measurement network $\mathcal{C}$ and true environment $\vec{\mathbf{v}}_m$ as the event $\left\{g\left(\vec{\mathbf{Y}}\right) \notin \mathcal{D}_{\vec{\mathbf{v}}_m}\right\}$, and the corresponding probability of error $P_{e,m,\mathcal{C}}$. The average probability of error is defined as $P_{e,\mathcal{C}} = \sum_m P_{e,m,\mathcal{C}} P_{\vec{\mathbf{V}}}\left(\vec{\mathbf{V}} = \vec{\mathbf{v}}_m\right)$. As defined in [2], the sensing capacity $C(D)$ at distortion D is the the maximum rate R such that there exists a sequence of measurement networks $\mathcal{C}_n$, such that $P_{e,\mathcal{C}_n} \to 0$ as $n \to \infty$ for fixed R. By definition, a larger sensing capacity implies that we can detect the environment using fewer measurements.

Types and joint types : Shannon used a random coding argument [6] to analyze the average probability of error for the channel coding problem. In that method the codeword for each message is drawn with independent, identically distributed symbols from a particular distribution. However in our case, as described in [2], we cannot draw symbols in this manner. The noiseless measurements $\vec{\mathbf{X}}$ (which correspond to the codewords in the channel coding problem) are constrained to be functions of the environment $\vec{\mathbf{v}}$ (which correspond to the message in the channel coding problem) through the sensing function Ψ. However, following [2] we apply a random measurement argument. That is, we assume each measurement to be a function Ψ of an independently chosen *random* subset of locations, resulting in a *random* measurement network $\mathcal{C}_n$. Because of the random $\mathcal{C}_n$, the noiseless measurements X^u are i.i.d. random variables. However, the codewords so formed ($\vec{\mathbf{X}}_m$ for $m \in 1 \ldots |\mathcal{V}|^{nR}$) are neither independent nor identically distributed and so, we use the method of types [16] to handle the correlations.

The notion of types and joint types [16] allows us to apply a version of the random coding argument, while still accounting for the correlation between codewords that arise because the measurements for different environments $\vec{\mathbf{v}}$ come from a particular measurement network $\mathcal{C}_n$. The type of a target vector $\vec{\mathbf{v}}$, denoted by $\vec{\gamma} = \left(\gamma^1, \ldots, \gamma^{|\mathcal{V}|}\right)$, is defined as the empirical p.m.f. of $\vec{\mathbf{v}}$. The joint type of two target vectors $\vec{\mathbf{v}}_m$ and $\vec{\mathbf{v}}_{m'}$, $\vec{\lambda} = \left(\lambda^{11}, \ldots, \lambda^{|\mathcal{V}||\mathcal{V}|}\right)$ is a p.m.f. of size $|\mathcal{V}|^2$ with λ^{ab} denoting the fraction of locations $\left\{i : v_m^i = a, v_{m'}^i = b\right\}$. For example, suppose $k = 4$, and consider the case of binary vectors $\vec{\mathbf{v}}_m = [0001]$ and $\vec{\mathbf{v}}_{m'} = [0101]$. Then the type of $\vec{\mathbf{v}}_m$ is $[\frac{3}{4}, \frac{1}{4}]$ and the type of $\vec{\mathbf{v}}_{m'}$ is $[\frac{1}{2}, \frac{1}{2}]$. The joint type of $\vec{\mathbf{v}}_m$ and $\vec{\mathbf{v}}_{m'}$ is $[\frac{1}{2}, \frac{1}{4}, 0, \frac{1}{4}]$.

Next, we will look at why the notion of types and joint types of the environment vector arise naturally in our analysis. As stated earlier, to parallel the random coding argument, we consider a case where each measurement (or sensor) chooses c locations to measure independently and uniformly at random, and then generates a function Ψ of those locations. In such a random drawing, the distribution of the noiseless output X^u for a measurement u is a function of only the type of the true environment, with

$$P_{X^u}\left(X^u = x\right) = \sum_{\vec{\mathbf{a}} \in \mathcal{V}^c, \Psi(\vec{\mathbf{a}})=x} \prod_{l=1}^{c} \gamma^{a^l} \doteq P_X^{\vec{\gamma}}\left(X = x\right) \quad (1)$$

Suppose that in a particular drawing of the network X_m^u is the noiseless output for measurement u when the environment is in state $\vec{\mathbf{v}}_m$. The output $X_{m'}^u$ when the environment is in state $\vec{\mathbf{v}}_{m'}$ is *not* independent of X_m^u because that measurement is taken by the same measurement network. Because of the randomly drawn measurements, the probability that the noiseless output of measurement u is X_m^u when $\vec{\mathbf{v}}_m$ occurs and $X_{m'}^u$ when $\vec{\mathbf{v}}_{m'}$ occurs depends only on the joint type of the two vectors and is given by

$$P_{X_m^u, X_{m'}^u}\left(X_m^u = x, X_{m'}^u = x'\right) = \quad (2)$$
$$\sum_{\vec{\mathbf{a}}, \vec{\mathbf{b}} \in \mathcal{V}^c, \Psi(\vec{\mathbf{a}})=x, \Psi\left(\vec{\mathbf{b}}\right)=x'} \prod_{l=1}^{c} \lambda^{a^l b^l} \doteq P^{\vec{\boldsymbol{\lambda}}}\left(X = x, X' = x'\right)$$

Naturally, the conditional $P_{X_{m'}^u|X_m^u}\left(X_{m'}^u = x' | X_m^u = x\right) = \frac{P_{X_m^u, X_{m'}^u}\left(X_m^u = x, X_{m'}^u = x'\right)}{P(X_m^u = x)}$. We also define the conditional distribution of the measurement Y^u when $\vec{\mathbf{v}}_{m'}$ of type $\vec{\gamma}$ occurs, conditioned on the fact that X_m^u is the noiseless output when a different environment $\vec{\mathbf{v}}_m$ occurs. This is also only a function of the joint type $\vec{\boldsymbol{\lambda}}$ of $\vec{\mathbf{v}}_m$ and $\vec{\mathbf{v}}_{m'}$.

$$Q_{Y^u|X_m^u}(Y^u = y | X_m^u = x) \doteq Q^{\vec{\boldsymbol{\lambda}}}(Y = y | X = x) \quad (3)$$
$$= \sum_{x' \in \mathcal{X}} P^{\vec{\boldsymbol{\lambda}}}\left(X_{m'} = x' | X_m = x\right) P_{Y|X}\left(Y = y | X = x'\right)$$

We emphasize that the fact that $X_{m'}^u$ and Y^u, when $\vec{\mathbf{v}}_{m'}$ occurs, are *not* independent of X_m^u (which corresponds to a different target $\vec{\mathbf{v}}_m$) is the core of the difference between our measurements and channel codes used for communication. We capture the dependence in a tractable manner using the notions of types and joint types.

III. Generalized derivation of the sensing capacity

A natural question that arises is, how many measurements are sufficient to guarantee that we can detect the random variables $\vec{\mathbf{V}}$ to within a specified distortion D? i.e what is the sensing capacity $C(D)$? Rachlin et al. [2] derived a lower bound on sensing capacity C_{LB} using Gallager's [6] Chernoff bounding technique to answer this question. However, they assumed that optimal ML or MAP decoders are used and furthermore, that all the problem model and parameters are known exactly.

In this paper, we are interested in situations where there is model uncertainty or mismatch. As a result a more general class of decoders needs to be considered and the analysis in [2] must be generalized. In this paper, we provide a simpler, yet more general derivation of C_{LB}, which allows us to go beyond the results in [2] and address these problems. We assume a decoder that maximizes some similarity metric $s^{\mathcal{C}}$ (which depends on the specific measurement network $\mathcal{C}$) between the received $\vec{\mathbf{Y}}$ and possible noiseless measurements $\vec{\mathbf{x}}_m$ for different states of the environment. Since there is a one-to-one correspondence between m and $\vec{\mathbf{v}}_m$, we can write such a decoder as $g\left(\vec{\mathbf{Y}}\right) = \text{argmax}_m\, s^{\mathcal{C}}\left(\vec{\mathbf{x}}_m, \vec{\mathbf{Y}}\right)$. For example, a Maximum Likelihood(ML) decoder has $s^{\mathcal{C}}\left(\vec{\mathbf{x}}_m, \vec{\mathbf{Y}}\right) = p_{\vec{\mathbf{Y}}|\vec{\mathbf{X}}}\left(\vec{\mathbf{Y}} | \vec{\mathbf{x}}_m\right)$. However, the analysis also holds for decoders maximizing other (possibly sub-optimal) metrics. We define $\mathcal{S}^{\vec{\gamma}}(D)$ be the set of joint types $\vec{\boldsymbol{\lambda}}$ at distortion D from a type $\vec{\gamma}$. $H(\bullet)$ is the entropy function. We define a tilting metric $f^{\vec{\boldsymbol{\lambda}}, w}\left(\vec{\mathbf{X}}_m, \vec{\mathbf{Y}}\right) = E_{\vec{\mathbf{X}}_{m'}|\vec{\mathbf{X}}_m}\left(s^{\mathcal{C}}\left(\vec{\mathbf{X}}_{m'}, \vec{\mathbf{Y}}\right)\right)^w$, which corresponds to the expected value of our metric for incorrect environments which have a joint type $\vec{\boldsymbol{\lambda}}$ with the true environment. Intuitively, if this is small (in a sense made rigorous in the proof) compared to the value of the metric for the true environment $\vec{\mathbf{v}}_m$ ($s^{\mathcal{C}}\left(\vec{\mathbf{x}}_m, \vec{\mathbf{Y}}\right)$) then we do not expect our decoder to output an environment $\vec{\mathbf{v}}_{m'}$ which has joint type $\vec{\boldsymbol{\lambda}}$ with $\vec{\mathbf{v}}_m$. Our theorem then concerns the large deviations of $Z_n^{\vec{\boldsymbol{\lambda}}, w}$ which is the ratio of these two quantities - the *empirical* metric for the correct environment and the *expected* metric for incorrect environments with joint type $\vec{\boldsymbol{\lambda}}$ and its log-moment generating function $\Lambda_n^{\vec{\boldsymbol{\lambda}}, w}(s)$,

$$Z_n^{\vec{\boldsymbol{\lambda}}, w} = \frac{1}{n} \log\left(\frac{\left(s^{\mathcal{C}}\left(\vec{\mathbf{X}}_m, \vec{\mathbf{Y}}\right)\right)^w}{f^{\vec{\boldsymbol{\lambda}}, w}\left(\vec{\mathbf{X}}_m, \vec{\mathbf{Y}}\right)}\right) \quad (4)$$

$$\Lambda_n^{\vec{\boldsymbol{\lambda}}, w}(s) = \log E[e^{s Z_n^{\vec{\boldsymbol{\lambda}}, w}}] \quad (5)$$

Theorem 3.1: (Sensing capacity : General statement) Suppose the limit $\Lambda^{\vec{\boldsymbol{\lambda}}, w}(s) = \lim_{n\to\infty} \frac{1}{n} \Lambda_n^{\vec{\boldsymbol{\lambda}}, w}(ns)$ exists, and $\Lambda^{*\vec{\boldsymbol{\lambda}}, w}(x) = \sup_s \left(sx - \Lambda^{\vec{\boldsymbol{\lambda}}, w}(s)\right)$ is the Fenchel-Legendre transform. If $\mathcal{D}_{\Lambda^{\vec{\lambda}, w}} = \left\{s \in \mathcal{R} : \Lambda^{\vec{\boldsymbol{\lambda}}, w}(s) < \infty\right\}$, and the origin belongs to the interior of $\mathcal{D}_{\Lambda^{\vec{\lambda}, w}}$, then, there exists a sequence of sparse measurement networks $\mathcal{C}_n$, such that the average probability of error, $P_{e, \mathcal{C}_n} \to 0$ as $n \to \infty$ of a decoder using a metric $s^{\mathcal{C}_n}$, for fixed R, *if* $R < C_{LB}$ where C_{LB} is a lower bound on sensing capacity given by,

$$C_{LB} = \min_{\vec{\boldsymbol{\lambda}} \notin \mathcal{S}^{\vec{\gamma}}(D)} \sup_{w \geq 0} \quad (6)$$
$$\frac{\sup_{T(\vec{\boldsymbol{\lambda}})} \left\{T\left(\vec{\boldsymbol{\lambda}}\right) : \inf_{x \in (-\infty, T(\vec{\boldsymbol{\lambda}})]} \Lambda^{*\vec{\boldsymbol{\lambda}}, w}(x) > 0\right\}}{\left[H(\vec{\boldsymbol{\lambda}}) - H(\vec{\gamma})\right]}$$

Proof: See Appendix. ■

This theorem states that if we have a sufficient number of measurements then there exists an appropriate measurement network and a decoder using a metric $s^{\mathcal{C}_n}$ that can guarantee detection to within a distortion D. Some discussion of the intuition behind this result may be helpful. The original large scale detection problem is a hypothesis testing problem with an exponential number of competing hypotheses. Using union bounds we group the competing hypotheses into a polynomial number of classes, with $e^{nR\left[H(\vec{\boldsymbol{\lambda}})-H(\vec{\gamma})\right]}$ hypotheses for each class, and one class for each joint type $\vec{\boldsymbol{\lambda}}$. Each hypothesis test is solved using a threshold of the form $e^{nT(\vec{\boldsymbol{\lambda}})}$ (See Eq. (A-4)). The error exponent for each class is a function of the largest threshold that still allows detection, for that class. We union bound over all the hypotheses in a class, each of which have the same error rates on average - because of the random sensor network deployment. The average probability of error is bounded by the worst exponent (or the most confusable joint type $\vec{\boldsymbol{\lambda}}$). This may be easier to visualize after we look at some special cases.

IV. Special Cases

Case 1 : We first consider the basic situation where all the measurement functions are identical and c sparse. Each measurement is a function of a subset of c locations as described earlier. We assume a Maximum Likelihood (ML) decoder. The ML decoder returns the index of the environment m that maximizes $P_{\vec{\mathbf{Y}}|\vec{\mathbf{X}}}\left(\vec{\mathbf{Y}}|\vec{\mathbf{x}}_m\right)$

Corollary 4.1: ([2] ML decoding no mismatch) There exists a sequence of sparse measurement networks $\mathcal{C}_n$, such that the average probability of error, $P_{e,\mathcal{C}_n} \to 0$ as $n \to \infty$, with ML decoding, for fixed R, *if* $R < C_{LB}^{ML}$, where

$$C_{LB}^{ML} = \min_{\vec{\boldsymbol{\lambda}} \notin \mathcal{S}^{\vec{\gamma}}(D)} \sup_{w \geq 0} \tag{7}$$

$$\frac{\sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} P^{\vec{\gamma}}(x)\, P_{Y|X}(y|x) \log\left(\frac{\left(P_{Y|X}(y|x)\right)^w}{Q^{\vec{\boldsymbol{\lambda}},w}(y|x)}\right)}{\left[H(\vec{\boldsymbol{\lambda}}) - H(\vec{\gamma})\right]}$$

Proof: We define the various terms that occur in Theorem 3.1:

$$s^{\mathcal{C}}\left(\vec{\mathbf{x}}_m, \vec{\mathbf{Y}}\right) = \prod_{u=1}^{n} P_{Y|X}\left(Y^u|x_m^u\right)$$

$$Q^{\vec{\boldsymbol{\lambda}},w}(y|x) = \sum_{x'\in\mathcal{X}} P^{\vec{\boldsymbol{\lambda}}}\left(X'=x'|X=x\right)\left(P_{Y|X}\left(Y=y|x=x'\right)\right)^w$$

$$f^{\vec{\boldsymbol{\lambda}},w}\left(\vec{\mathbf{X}}_m, \vec{\mathbf{Y}}\right) = \prod_{u=1}^{n} Q^{\vec{\boldsymbol{\lambda}},w}\left(Y^u|X_m^u\right)$$

$$Z_n^{\vec{\boldsymbol{\lambda}}} = \frac{1}{n}\sum_{u=1}^{n} \log\left(\frac{\left(P_{Y|X}\left(Y^u|X_m^u\right)\right)^w}{Q^{\vec{\boldsymbol{\lambda}},w}\left(Y^u|X_m^u\right)}\right)$$

Each of the terms in $Z_n^{\vec{\boldsymbol{\lambda}}}$, $Z^u = \log\left(\frac{\left(P_{Y|X}(Y^u|X_m^u)\right)^w}{Q^{\vec{\boldsymbol{\lambda}},w}(Y^u|X_m^u)}\right)$ are i.i.d. Thus, we are seeking a large deviations result for the average of i.i.d terms. Hence, the condition in Theorem 3.1 is met and $\Lambda^{\vec{\boldsymbol{\lambda}},w}(s) = \log E[e^{sZ^1}]$. Now, $x < E\left(Z^1\right)$, then $\Lambda^{*\vec{\boldsymbol{\lambda}}}(x) \geq 0$. So the resulting condition is that $T\left(\vec{\boldsymbol{\lambda}}\right) \leq E\left(\log\left(\frac{\left(P_{Y|X}(Y|X)\right)^w}{Q^{\vec{\boldsymbol{\lambda}},w}(Y|X)}\right)\right)$. ■

We consider simplifications of this expression. If we do not optimize over w but just set $w = 1$,

$$C_{LB} = \min_{\vec{\boldsymbol{\lambda}} \notin \mathcal{S}^{\vec{\gamma}}(D)} \frac{\mathcal{D}(P_{Y|X}||Q^{\vec{\boldsymbol{\lambda}}})}{\left[H(\vec{\boldsymbol{\lambda}}) - H(\vec{\gamma})\right]} \tag{8}$$

where $\mathcal{D}(\bullet||\bullet)$ is the KL-divergence [6]. Here the reason for the intuition given after Theorem 3.1 is even more obvious. Another specialization occurs if we assume the ideal coding case, where the codewords $\vec{\mathbf{X}}$ are actually i.i.d. Then the KL-divergence becomes a mutual information term $I(\bullet,\bullet)$, independent of the joint type $\vec{\boldsymbol{\lambda}}$ resulting in

$$C_{LB} = I(P_{Y|X}, Q) \tag{9}$$

which is Shannon's expression for channel capacity [6]. Thus, the sensing capacity in Theorem 3.1 and in [2] can be considered generalizations of the channel capacity.

Case 2 : We now derive a simple extension that will be useful for the simulations in Section V. We consider the heterogeneous case of measurement networks, $\mathcal{C}_n^{HET}$, where there are M classes of measurements, and each class t for $t = 1, \ldots, M$ has its own measurement function $\Psi^t\left(v^{l(1)}, \ldots, v^{l(c)}\right)$. Let there be a constant fraction $\alpha_t = \frac{n_t}{n}$ of measurements of class t, and we assume that the measurements are arranged such that the products below make sense.

Corollary 4.2: ([2] ML decoding, heterogeneous measurements) There exists a sequence of heterogeneous sparse measurement networks $\mathcal{C}_n^{HET}$ (as defined above), such that the average probability of error, $P_{e,\mathcal{C}_n^{HET}} \to 0$ as $n \to \infty$, with ML decoding, for fixed R, *if* $R < C_{LB}^{HET}$, where

$$C_{LB}^{HET} = \min_{\vec{\boldsymbol{\lambda}} \notin \mathcal{S}^{\vec{\gamma}}(D)} \sup_{w_t \geq 0} \sum_{t=1}^{M} \alpha_t \tag{10}$$

$$\frac{\sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} P^{\vec{\gamma}}(x)\, P_{Y|X}^t(y|x) \log\left(\frac{\left(P_{Y|X}(y|x)\right)^{w_t}}{Q^{\vec{\boldsymbol{\lambda}},w_t}(y|x)}\right)}{\left[H(\vec{\boldsymbol{\lambda}}) - H(\vec{\gamma})\right]}$$

Proof: We just outline the proof. In the derivation in Section III, we can account for the heterogeneity by assuming a single vector output measurement function $\Psi\left(v^{l(1)}, \ldots, v^{l(c)}\right) = \left[v^{l(1)}, \ldots, v^{l(c)}\right]$, and M noise functions, $p_{Y|X}^t(Y=y|X=x)$. This mapping is useful because it unifies the noise model $P_{Y|X}$ and the measurement function Ψ. Again assuming an ML decoder, that returns the index of the environment m that maximizes $P_{\vec{\mathbf{Y}}|\vec{\mathbf{X}}}\left(\vec{\mathbf{Y}}|\vec{\mathbf{x}}_m\right)$. We then define the various terms that occur in Theorem 3.1 and the result follows. ■

Case 3 : The next specialization we consider is the case where there is mismatch, which was one of the motivations for the analysis performed in this paper. Again we consider a single measurement function. Each measurement is a function of a subset of c locations as described earlier. However, instead of the true measurement function and noise, $P_{Y|X}$, we have a mismatched/mis-calibrated symbol-wise similarity metric $d(x,y)$ - possibly due to mis-calibration or model error. For example, if the noise model is mis-calibrated, then $d(x,y) = \widehat{P}_{Y|X}$. We run a decoder using this metric.

The decoder returns the index of the environment m that maximizes $d\left(\vec{\mathbf{x}}_m, \vec{\mathbf{Y}}\right) = \prod_{u=1}^{n} d\left(x_m^u, Y^u\right)$. We emphasize that using the mapping of Corollary 4.2, we can also analyze the case where the measurement function Ψ is mis-calibrated. This result also can be used in cases where we use a sub-optimal decoder that maximizes a metric that is not the ML metric.

Corollary 4.3: (ML decoding with a mismatched metric) There exists a sequence of sparse measurement networks $\mathcal{C}_n$, such that the average probability of error, $P_{e,\mathcal{C}_n} \to 0$ as $n \to \infty$, with mismatched decoding using a metric $d\left(\vec{\mathbf{x}}_m, \vec{\mathbf{Y}}\right)$, for fixed R, *if* $R < C_{LB}^{MIS}$, where

$$C_{LB}^{MIS} = \min_{\vec{\boldsymbol{\lambda}} \notin \mathcal{S}^{\vec{\gamma}}(D)} \sup_{w \geq 0} \frac{\sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} P^{\vec{\gamma}}(x) P_{Y|X}(y|x) \log\left(\frac{d^w(x,y)}{Q_d^{\vec{\boldsymbol{\lambda}},w}(y|x)}\right)}{\left[H(\vec{\boldsymbol{\lambda}}) - H(\vec{\gamma})\right]} \tag{11}$$

Proof: In this case we define ,

$$s^{\mathcal{C}}\left(\vec{\mathbf{x}}_m, \vec{\mathbf{Y}}\right) = d\left(\vec{\mathbf{x}}_m, \vec{\mathbf{Y}}\right) = \prod_{u=1}^{n} d\left(x_m^u, Y^u\right)$$

$$Q_d^{\vec{\boldsymbol{\lambda}},w}(y|x) = \sum_{x'\in\mathcal{X}} P^{\vec{\boldsymbol{\lambda}}}\left(X' = x'|X = x\right) d^w\left(x', y\right)$$

$$f^{\vec{\boldsymbol{\lambda}},w}\left(\vec{\mathbf{X}}_m, \vec{\mathbf{Y}}\right) = \prod_{u=1}^{n} Q_d^{\vec{\boldsymbol{\lambda}},w}\left(Y^u|X_m^u\right)$$

$$Z_n^{\vec{\boldsymbol{\lambda}}} = \frac{1}{n}\sum_{u=1}^{n} \log\left(\frac{\left(P\left(Y^u|X_m^u\right)\right)^w}{Q_d^{\vec{\boldsymbol{\lambda}},w}\left(Y^u|X_m^u\right)}\right)$$

Where $Z_n^{\vec{\boldsymbol{\lambda}}}$ is again an average of i.i.d terms, and so we now apply Cramer's theorem [15]. ■

This is the parallel of the Csiszar-Korner-Hui bound [8] in coding theory, modified to account for the correlations between 'codewords' that is inherent in measurement networks. This is smaller than the matched sensing capacity, indicating that we require more measurements to make a network robust to model mismatch, but at the same time, it also shows that we can still guarantee performance if a sufficient number of measurements are available.

Case 4 : Another problem we wished to study was the case where we had model uncertainty. We consider a situation where we do not know the exact value of the measurement model $P_{Y|X}$, but we know that it belongs to a finite set $\{P^\theta\}$, indexed by $\theta \in \Theta$. To simplify notation, we choose $w = 1$. Again, using the mapping of Corollary 4.2, we can also similarly analyze the case where the measurement function Ψ is uncertain.

Corollary 4.4: (ML decoding with uncertainty for a class of channels) There exists a sequence of sparse measurement networks $\mathcal{C}_n$, such that the average probability of error, $P_{e,\mathcal{C}_n} \to 0$ as $n \to \infty$, with unknown $\{P^\theta\}$, for fixed R, *if* $R < C_{LB}^{FIN}$ where C_{LB}^{FIN} is a lower bound on sensing capacity given by,

$$C_{LB}^{FIN} = \min_{\vec{\boldsymbol{\lambda}} \notin \mathcal{S}^{\vec{\gamma}}(D)} \frac{\min_{\theta_t\in\Theta} \min_{\theta'\in\Theta} E^{\theta_t}\left(\log\left(\frac{\left(P_{Y|X}^{\theta_t}(Y|X)\right)}{Q_{\theta'}^{\vec{\boldsymbol{\lambda}}}(Y|X)}\right)\right)}{\left[H(\vec{\boldsymbol{\lambda}}) - H(\vec{\gamma})\right]} \tag{12}$$

Proof: A possible metric for the decoder to maximize is

$$s^{\mathcal{C}}\left(\vec{\mathbf{x}}_i, \vec{\mathbf{Y}}\right) = \frac{1}{|\Theta|}\sum_{\theta\in\Theta} P_\theta\left(\vec{\mathbf{Y}}|\vec{\mathbf{x}}_i\right) = \frac{1}{|\Theta|}\sum_{\theta\in\Theta}\prod_{u=1}^{n} P_\theta\left(Y^u|x_i^u\right)$$

This results in,

$$f^{\vec{\boldsymbol{\lambda}}}\left(\vec{\mathbf{X}}_i, \vec{\mathbf{Y}}\right) = \frac{1}{|\Theta|}\sum_{\theta\in\Theta}\prod_{u=1}^{n} Q_\theta^{\vec{\boldsymbol{\lambda}}}\left(Y^u|X_i^u\right) \tag{13}$$

$$Z_n^{\vec{\boldsymbol{\lambda}}} = \frac{1}{n}\log\left(\frac{\sum_{\theta\in\Theta}\prod_{u=1}^{n}\left(P_\theta\left(Y^u|X_i^u\right)\right)}{\sum_{\theta'\in\Theta}\prod_{u=1}^{n} Q_{\theta'}^{\vec{\boldsymbol{\lambda}}}\left(Y^u|X_i^u\right)}\right)$$

Unfortunately, this is not the average of i.i.d terms, and so we try a bounding procedure to simplify the expression. Let θ_t be the true θ that actually occurred.

$$\begin{aligned}
Z_n^{\vec{\boldsymbol{\lambda}}} &\geq \frac{1}{n}\log\left(\frac{\prod_{u=1}^{n}\left(P_{\theta_t}\left(Y^u|X_i^u\right)\right)}{\sum_{\theta'\in\Theta}\prod_{u=1}^{n} Q_{\theta'}^{\vec{\boldsymbol{\lambda}}}\left(Y^u|X_i^u\right)}\right) && (14)\\
&\geq \frac{1}{n}\log\left(\frac{\prod_{u=1}^{n}\left(P_{\theta_t}\left(Y^u|X_i^u\right)\right)}{|\Theta|\max_{\theta'}\prod_{u=1}^{n} Q_{\theta'}^{\vec{\boldsymbol{\lambda}}}\left(Y^u|X_i^u\right)}\right) && (15)\\
&\geq -\frac{\log\left(|\Theta|\right)}{n} + \min_{\theta'\in\Theta}\frac{1}{n}\sum_{u=1}^{n}\log\left(\frac{P_{\theta_t}\left(Y^u|X_i^u\right)}{Q_{\theta'}^{\vec{\boldsymbol{\lambda}}}\left(Y^u|X_i^u\right)}\right)\\
&\geq -\epsilon + \min_{\theta'\in\Theta} S_n^{\vec{\boldsymbol{\lambda}},\theta_t,\theta'} && (16)
\end{aligned}$$

The last inequality is true for any ϵ for n large enough. Each term in $S_n^{\vec{\boldsymbol{\lambda}},\theta_t,\theta'}$ is i.i.d as before. Since $Z_n^{\vec{\boldsymbol{\lambda}}} \geq \min_{\theta'\in\Theta} S_n^{\vec{\boldsymbol{\lambda}},\theta_t,\theta'} - \epsilon$, $Pr\left(Z_n^{\vec{\boldsymbol{\lambda}}} < T\left(\vec{\boldsymbol{\lambda}}\right)\right) \leq Pr\left(\min_{\theta'\in\Theta} S_n^{\vec{\boldsymbol{\lambda}},\theta_t,\theta'} < T\left(\vec{\boldsymbol{\lambda}}\right) + \epsilon\right)$, for n large enough such that $\frac{\log(|\Theta|)}{n} \leq \epsilon$. Using a union bound over θ', this goes to zero if $T\left(\vec{\boldsymbol{\lambda}}\right) < \min_{\theta_t\in\Theta}\min_{\theta'\in\Theta} E^{\theta_t}\left(\log\left(\frac{\left(P_{Y|X}^{\theta_t}(Y|X)\right)}{Q_{\theta'}^{\vec{\boldsymbol{\lambda}}}(Y|X)}\right)\right)$, and so, we have proved the corollary. ■

Let θ^* be the minimizing θ_t in the Corollary 4.4. The Corollary also suggests a simplified decoder that just assumes $P_{Y|X}^{\theta^*}$. This decoder would achieve the same *worst-case* on performance as the suggested decoder with uncertainty, but would have a lower computational complexity. Thus, our theorem can be used to indicate and analyze a sub-optimal decoder with reasonable worst case guarantees.

V. Heterogeneous sensor networks to compensate for mismatch

We look at a simple example where mismatch can cause a sharp drop-off in the performance of a measurement/sensor network (or equivalently a large increase in the number of sensors required to achieve a specified distortion), and where

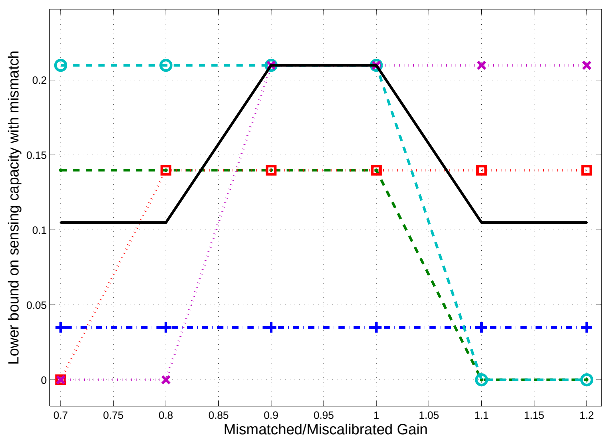

Fig. 2. The use of heterogeneous sensor networks to compensate for mismatch. The dash/dot lines are rates achievable by sensors with different thresholds, at different mismatches in the gain. The true gain is 1. The solid line represents the rates achievable using a mixture of sensors.

the bounds from the previous section can be used to guide sensor network design. The sensing model is as follows. Each sensor senses $c = 5$ random locations. The sensor transfer function takes an average of values within it's field of view and then applies a threshold. The decoder assumes that the environment takes values $\{0, 1\}$, but actually the values are $\{0, g\}$, where g is a mismatched gain. The required rates for sensors with different thresholds, to achieve a distortion of $D = .005$ are plotted with dash/dot lines in Fig. 2. These indicate that the sensors are susceptible to the gain mismatch. To ensure that the error criterion is always met despite the mismatches, we would need to use the sensor with the pessimistic 'blue dash-dot' threshold (near the bottom of the figure), which can achieve a rate of less than .035. Alternatively, we can use a heterogeneous network, composed of n_t sensors with each of the thresholds t. We use the results from Corollary 4.2 and 4.3, essentially that, $C_{LB}^{MIS,HET} = \sum_{t=1}^{r} \frac{n_t}{n} C_{LB,t}^{MIS}$. So, if we know that mismatches are possible, and the rates achievable with mismatch (using the results of the previous section), then we could use a mixture of sensors with different thresholds, which could achieve the black solid curve in the figure - a rate of more than .1.

This is a simple example where we can use the formulas derived to determine a combination of sensors so that the sensor network is robust to mismatch. Other kinds of mismatches can be analyzed similarly, and in many cases it can be seen that a mixture of heterogeneous sensors is more robust to mismatch and uncertainty than networks with a single kind of sensor.

ACKNOWLEDGMENT

This work was partially supported by NSF award CNS-0347455 and an ARO award through Cylab, CMU.

APPENDIX

Proof: (Theorem 3.1) The notation in our proof is unfortunately complicated, but is necessary to maintain rigor. So, we first provide an outline to give some intuition.

Eqs A-1 to A-3 : We use the random measurement argument outlined in Section II. We bound $P_{e,m}$, which is the expected probability of error over the random sensor networks. To do this we wish to follow the classical union bound arguments, where we union bound over all incorrect messages, as in the channel capacity proof [6]. However, because of the non-i.i.d. nature of the codewords (as explained in Section II) we first split the union bound based on the type of the incorrect environments as suggested in [2].

Eqs A-4 to A-10 : We split the error events based on two sources of error - i) the noise is so large that the measurements $\vec{\mathbf{Y}}$ are not similar to the noiseless measurements $\vec{\mathbf{X}}_m$ corresponding to the true environment $\vec{\mathbf{v}}_m$ (a 'bad noise' event) and ii) we draw a measurement network in which the $\vec{\mathbf{X}}_{m'}$ corresponding to a different $\vec{\mathbf{v}}_{m'}$ is similar to $\vec{\mathbf{Y}}$. The conditions under which the probability of both these events goes to zero as $n \to \infty$ is our main large deviation result.

Eqs A-11 to A-12 : We complete the proof by showing that if we can guarantee that the average error probability for 'typical' environments goes to zero exponentially fast in n, (using the large deviation result), then the average error probability across all environments goes to zero since there are only a polynomial number of types, and among them 'typical' ones are the most likely.

Now we start the actual proof. Let $S_m^{\vec{\lambda}}$ be the set of environments at joint type $\vec{\lambda}$ with $\vec{\mathbf{v}}_m$.

$$P_{e,m} = Pr\left(\bigcup_{m' \notin \mathcal{D}_{\vec{\mathbf{v}}_m}} \left\{ s^{\mathcal{C}}\left(\vec{\mathbf{X}}_{m'}, \vec{\mathbf{Y}}\right) > s^{\mathcal{C}}\left(\vec{\mathbf{X}}_m, \vec{\mathbf{Y}}\right) \right\} \right) \tag{A-1}$$

$$= Pr\left(\bigcup_{\vec{\lambda} \notin \mathcal{S}^{\vec{\gamma}}(D)} \bigcup_{m' \in S_m^{\vec{\lambda}}} \left\{ s^{\mathcal{C}}\left(\vec{\mathbf{X}}_{m'}, \vec{\mathbf{Y}}\right) > s^{\mathcal{C}}\left(\vec{\mathbf{X}}_m, \vec{\mathbf{Y}}\right) \right\} \right) \tag{A-2}$$

We introduce $f^{\vec{\lambda},w}\left(\vec{\mathbf{X}}_m, \vec{\mathbf{Y}}\right)$, which is a $\vec{\lambda}$ specific tilting function to be optimized later, and $w \geq 0$ is a parameter that can also be optimized.

$$\leq \sum_{\vec{\lambda} \notin \mathcal{S}^{\vec{\gamma}}(D)} \tag{A-3}$$

$$Pr\left(\bigcup_{m' \in S_m^{\vec{\lambda}}} \left\{ \frac{\left(s^{\mathcal{C}}\left(\vec{\mathbf{X}}_{m'}, \vec{\mathbf{Y}}\right)\right)^w}{f^{\vec{\lambda},w}\left(\vec{\mathbf{X}}_m, \vec{\mathbf{Y}}\right)} > \frac{\left(s^{\mathcal{C}}\left(\vec{\mathbf{X}}_m, \vec{\mathbf{Y}}\right)\right)^w}{f^{\vec{\lambda},w}\left(\vec{\mathbf{X}}_m, \vec{\mathbf{Y}}\right)} \right\} \right)$$

We define a 'bad noise' event using a $\vec{\lambda}$ specific threshold $e^{nT(\vec{\lambda})}$

$$\mathcal{Y}_{b,\lambda} = \left\{ \frac{\left(s^{\mathcal{C}}\left(\vec{\mathbf{X}}_m, \vec{\mathbf{Y}}\right)\right)^w}{f^{\vec{\lambda},w}\left(\vec{\mathbf{X}}_m, \vec{\mathbf{Y}}\right)} < e^{nT(\vec{\lambda})} \right\} \tag{A-4}$$

We now union bound in (A-3) over good and bad noise events. For the good noise condition we have reduced the problem to a tilted threshold test.

$$P_{e,m} \leq \sum_{\vec{\lambda} \notin \mathcal{S}^{\vec{\gamma}}(D)} Pr\left(\mathcal{Y}_{b,\lambda}\right) + \tag{A-5}$$

$$\sum_{\vec{\lambda} \notin \mathcal{S}^{\vec{\gamma}}(D)} \sum_{m' \in S_m^{\vec{\lambda}}} Pr\left(\left\{\frac{\left(s^{\mathcal{C}}\left(\vec{\mathbf{X}}_{m'}, \vec{\mathbf{Y}}\right)\right)^w}{f^{\vec{\lambda},w}\left(\vec{\mathbf{X}}_m, \vec{\mathbf{Y}}\right)} \geq e^{nT(\vec{\lambda})}\right\}\right)$$

Using results from [16], $|\mathcal{S}^{\vec{\gamma}}(D)| \leq n^{|\mathcal{V}|^2}$ since there are only a polynomial number of joint types, and $|S_m^{\vec{\lambda}}| \leq e^{nR\left[H(\vec{\lambda})-H(\vec{\gamma})\right]}$. Considering the second term in (A-5),

$$Term\ 2 \quad \leq \quad n^{|\mathcal{V}|^2} \max_{\vec{\lambda} \notin \mathcal{S}^{\vec{\gamma}}(D)} [e^{nR\left[H(\vec{\lambda})-H(\vec{\gamma})\right]} \tag{A-6}$$

$$Pr\left(\left\{\frac{\left(s^{\mathcal{C}}\left(\vec{\mathbf{X}}_{m'}, \vec{\mathbf{Y}}\right)\right)^w}{f^{\vec{\lambda},w}\left(\vec{\mathbf{X}}_m, \vec{\mathbf{Y}}\right)} \geq e^{nT(\vec{\lambda})}\right\}\right)]$$

Applying Markov's inequality,

$$\leq \quad n^{|\mathcal{V}|^2} \max_{\vec{\lambda} \notin \mathcal{S}^{\vec{\gamma}}(D)} e^{nR\left[H(\vec{\lambda})-H(\vec{\gamma})\right]} e^{-nT(\vec{\lambda})} \tag{A-7}$$

$$E_{\vec{\mathbf{X}}_m} E_{\vec{\mathbf{Y}}|\vec{\mathbf{X}}_m} E_{\vec{\mathbf{X}}_{m'}|\vec{\mathbf{X}}_m} \frac{\left(s^{\mathcal{C}}\left(\vec{\mathbf{X}}_{m'}, \vec{\mathbf{Y}}\right)\right)^w}{f^{\vec{\lambda},w}\left(\vec{\mathbf{X}}_m, \vec{\mathbf{Y}}\right)}$$

We now choose the tilting function $f^{\vec{\lambda},w}\left(\vec{\mathbf{X}}_m, \vec{\mathbf{Y}}\right) = E_{\vec{\mathbf{X}}_{m'}|\vec{\mathbf{X}}_m}\left(s^{\mathcal{C}}\left(\vec{\mathbf{X}}_{m'}, \vec{\mathbf{Y}}\right)\right)^w$ resulting in

$$Term\ 2 \quad \leq \quad n^{|\mathcal{V}|^2} \max_{\vec{\lambda} \notin \mathcal{S}^{\vec{\gamma}}(D)} e^{-n[T(\vec{\lambda})-R\left[H(\vec{\lambda})-H(\vec{\gamma})\right]]} \tag{A-8}$$

Because of the random sensor placement, as discussed in Section II, the distributions only depend on the type $\vec{\gamma}$ of $\vec{\mathbf{v}}_m$, and the pairwise joint types $\vec{\lambda}$ of $\vec{\mathbf{v}}_m$ and $\vec{\mathbf{v}}_{m'}$. This probability goes to zero if $R < \min_{\vec{\lambda} \notin \mathcal{S}^{\vec{\gamma}}(D)} \frac{T(\vec{\lambda})}{\left[H(\vec{\lambda})-H(\vec{\gamma})\right]}$.

For the first term in (A-5), we have

$$Term\ 1 \ \leq n^{|\mathcal{V}|^2} \min_{\vec{\lambda} \notin \mathcal{S}^{\vec{\gamma}}(D)} Pr\left(Z_n^{\vec{\lambda},w} < T\left(\vec{\lambda}\right)\right) \tag{A-9}$$

Where we define $Z_n^{\vec{\lambda},w}$ and its log-m.g.f $\Lambda_n^{\vec{\lambda},w}(s)$,

$$Z_n^{\vec{\lambda},w} = \frac{1}{n} \log\left(\frac{\left(s^{\mathcal{C}}\left(\vec{\mathbf{X}}_m, \vec{\mathbf{Y}}\right)\right)^w}{f^{\vec{\lambda},w}\left(\vec{\mathbf{X}}_m, \vec{\mathbf{Y}}\right)}\right), \Lambda_n^{\vec{\lambda},w}(s) = \log E[e^{sZ_n^{\vec{\lambda}}}] \tag{A-10}$$

If the limit $\Lambda^{\vec{\lambda},w}(s) = \lim_{n\to\infty} \frac{1}{n}\Lambda_n^{\vec{\lambda},w}(ns)$ exists, and $\Lambda^{*\vec{\lambda},w}(x) = \sup_s \left(sx - \Lambda^{\vec{\lambda},w}(s)\right)$ is the Fenchel-Legendre transform. If $\mathcal{D}_{\Lambda^{\vec{\lambda},w}} = \left\{s \in \mathcal{R} : \Lambda^{\vec{\lambda},w}(s) < \infty\right\}$, and the origin belongs to the interior of $\mathcal{D}_{\Lambda^{\vec{\lambda},w}}$, according to the Gartner-Ellis theorem [15] the probability in (A-9) goes to 0 with exponent $-\inf_{x\in(-\infty,T(\vec{\lambda})]} \Lambda^{*\vec{\lambda},w}(x)$. If this exponent is positive, then Term 2 in (A-5) also goes to 0 as $n \to \infty$.

Suppose that targets are generated independently at random with p.m.f. $p_V(v = V)$. Let this p.m.f. be $\vec{\gamma}$. As defined earlier we look at the average probability of error, P_e

$$\begin{aligned} P_e &= E_{\mathcal{C}} P_{e,\mathcal{C}} = \sum_{m=1}^{2^{nR}} P_{\vec{\mathbf{V}}}\left(\vec{\mathbf{V}} = \vec{\mathbf{v}}_m\right) E_{\mathcal{C}} P_{e,m,\mathcal{C}} \\ &= \sum_{m=1}^{2^{nR}} P_{\vec{\mathbf{V}}}\left(\vec{\mathbf{V}} = \vec{\mathbf{v}}_m\right) P_{e,m} \end{aligned} \tag{A-11}$$

We divide the environments $\vec{\mathbf{v}}_m$ based on their types $\vec{\gamma}'$. There are $e^{nRH(\vec{\gamma}')}$ target vectors of each type $\vec{\gamma}'$. According to Eq. (A-8) the error rate for an environment m is only a function of its type $\vec{\gamma'}$, which we denote by $P_{e,m} = P_{e,\vec{\gamma}'}$.

$$\begin{aligned} P_e &= \sum_{\vec{\gamma'}} e^{nRH(\vec{\gamma}')} Pr^{\vec{\gamma}}(\vec{\mathbf{v}}\ of\ type\ \vec{\gamma'}) P_{e,\vec{\gamma}'} \\ &= \sum_{\vec{\gamma'}} e^{nRH(\vec{\gamma}')} e^{-nR(H(\vec{\gamma}')+D(\vec{\gamma}||\vec{\gamma}'))} P_{e,\vec{\gamma}'} \end{aligned} \tag{A-12}$$

where $\vec{\gamma}$ is the type corresponding to the true distribution P_V. In the above sum, in terms for which $\vec{\gamma}' \neq \vec{\gamma}$, the $e^{-n\left(D(\vec{\gamma}||\vec{\gamma}')\right)}$, causes their contribution to go to zero as $n \to \infty$, using the fact that there only a polynomial number of types [16]. And so, if the $P_{e,\vec{\gamma}}$ (for typical environments m of type $\vec{\gamma}$) goes to zero exponentially, then $P_e \to 0$. The condition for this to happen was stated after (A-10). If this condition is satisfied, the average error rate across measurement networks goes to zero, and there exists a sequence of sensor measurement networks $\mathcal{C}_n$ for which the error rate goes to zero. ■

REFERENCES

[1] A. Montanari, "Estimating random variables from random sparse observations," *European Transactions on Telecommunications*, vol. 19, no. 4, pp. 385–403, 2008.

[2] Y. Rachlin, R. Negi, and P. Khosla, "The sensing capacity of sensor networks," *submitted to IEEE Transactions on Information Theory*, arXiv:0901.2094 [cs.IT].

[3] R. Berinde, A. Gilbert, P. Indyk, H. Karloff, and M. Strauss, "Combining geometry and combinatorics: A unified approach to sparse signal recovery," in *46th Annual Allerton Conference on Communication, Control, and Computing*, Sept. 2008, pp. 798–805.

[4] D. Guo and C.-C. Wang, "Multiuser detection of sparsely spread CDMA," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 3, pp. 421–431, April 2008.

[5] W. Wang, M. Wainwright, and K. Ramchandran, "Information-theoretic limits on sparse support recovery: Dense versus sparse measurements," in *IEEE International Symposium on Information Theory*, July 2008, pp. 2197–2201.

[6] R. G. Gallager, *Information Theory and Reliable Communication*. New York, NY, USA: John Wiley & Sons, Inc., 1968.

[7] A. Lapidoth and P. Narayan, "Reliable communication under channel uncertainty," *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2148–2177, Oct 1998.

[8] I. Sason and S. Shamai, "Performance analysis of linear codes under maximum-likelihood decoding: A tutorial," *Foundations and Trends in Communications and Information Theory*, vol. 3, no. 1/2, 2006.

[9] F. Kschischang, B. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Transactions on Information Theory*, vol. 47, no. 2, pp. 498–519, Feb 2001.

[10] M. Luby, "LT codes," in *The 43rd Annual IEEE Symposium on Foundations of Computer Science*, 2002, pp. 271–280. [Online]. Available: http://dx.doi.org/10.1109/SFCS.2002.1181950

[11] D. Du and F. Hwang, *Combinatorial group testing and its applications*. World Scientific, 2000.

[12] M. Mezard and C. Toninelli, "Group Testing with Random Pools: optimal two-stage algorithms," *ArXiv e-prints*, June 2007.

[13] S. Verdu, *Multiuser Detection*. Cambridge University press, 1998.

[14] M. A. Sheikh, S. Sarvotham, O. Milenkovic, and R. G. Baraniuk, "DNA array decoding from nonlinear measurements by belief propagation," in *IEEE/SP 14th Workshop on Statistical Signal Processing*, Aug. 2007, pp. 215–219.

[15] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*. Jones and Bartlett, Boston, 1993.

[16] I. Csiszar, "The method of types," *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2505–2523, Oct 1998.