Proceedings of the 2007 IEEE Symposium on
Foundations of Computational Intelligence (FOCI 2007)

Large-scale Sensor Networks as Collective and Frustrated Systems

Tatsuto Murayama
NTT Communication Science Laboratories
Nippon Telegraph and Telephone Corporation
Keihanna, Kyoto 619-0237, Japan
Email: murayama@cslab.kecl.ntt.co.jp

Peter Davis
NTT Communication Science Laboratories
Nippon Telegraph and Telephone Corporation
Keihanna, Kyoto 619-0237, Japan
Email: davis@cslab.kecl.ntt.co.jp

***Abstract*— This article presents a large-scale analysis of a distributed sensing model for systemized and networked sensors. In the system model, a data center acquires binary information from a bunch of L sensors which each independently encode their noisy observations of an original bit sequence, and transmit their encoded sequences to the data center at a combined data rate R, which is strictly limited. Supposing that the sensors use independent quantization techniques, we show that the performance can be evaluated for any given finite R when the number of sensors L goes to infinity. The analysis shows how the optimal strategy for the distributed sensing problem changes at critical values of the data rate R or the noise level p.**

I. Introduction

Some influential company laboratories world wide, include Intel Research, is actively exploring the potential of large-scale sensor networks. They typically work with the community and industry collaborators of their own, and this global trend is already demonstrating the potential of this new technology. Potential future markets include transportation and shipping, fire fighting and rescue operations, home automation and more [1]. Still, for all the promise, it is often difficult to integrate the individual components of a sensor network in a smart way. Although we see many breakthroughs in component devices, advanced software, and power managements, system-level understanding of the emerging technology is still weak. It requires a shift in our notion of "what to look for". It requires a study of collective behavior and resulting trade-offs. This is the issue that we address in this article. We demonstrate the usefulness of adopting new approaches by considering the following scenario.

Consider that a data center is interested in the data sequence, $\{X(t)\}_{t=1}^{\infty}$, which cannot be observed directly. Therefore, the data center deploys a bunch of L sensors which each independently encodes its noisy observation of the sequence, $\{Y_i(t)\}_{t=1}^{\infty}$, without sharing any information, i.e., the sensors are not permitted to communicate and decide what to send to the data center beforehand. The data center collects separate samples from all the L sensors and uses them to recover the original sequence. However, since $\{X(t)\}_{t=1}^{\infty}$ is not the only pressing matter which the data center must consider, the combined data rate R at which the sensors can communicate with it is strictly limited. A formulation of decentralized communication with estimation task, the "CEO problem", was first proposed by Berger and Zhang [2], providing a new theoretical framework for large scale sensing systems. In this outstanding work, some interesting properties of such systems have been revealed. If the sensors were permitted to communicate on the basis of their pooled observations, then they would be able to smooth out their independent observation noises entirely as L goes to infinity. Therefore, the data center can achieve an arbitrary fidelity $D(R)$, where $D(\cdot)$ denotes the distortion rate function of $\{X(t)\}$. In particular, the data center recovers almost complete information if R exceeds the entropy rate of $\{X(t)\}$. However, if the sensors are not allowed to communicate with each other, there does not exist a finite value of R for which even infinitely many sensors can make D arbitrarily small [2].

In this article, we introduce a new analytical model for a massive sensing system with a finite data rate R. More specifically, we assume that the sensors use a class of quantization methods for rate distortion coding [3], while the data center recovers the original sequence by using optimal "majority vote" estimation under the separate *decoding* condition [4]. We consider the distributed sensing problem of deciding the optimal number of sensors L given the combined data rate R. Our asymptotic analysis successfully provides the performance of the whole sensing system when L goes to infinity, where the data rate for an individual sensor information vanishes. Here, we exploit statistical methods which have recently been developed in the field of disordered statistical systems, in particular, the spin glass theory.

The article is organized as follows. In Section II, we introduce a system model for the sensor network. Section III summarizes the results of our approach, where the following section provides the outline of our analysis. Conclusions are given in the last section.

II. System Model

Let $P(x)$ be a probability distribution common to $\{X(t)\} \in \mathcal{X}$, and $W(y|x)$ be a stochastic matrix defined on $\mathcal{X} \times \mathcal{Y}$, with $\mathcal{Y}$ denotes the common alphabet of $\{Y_i(t)\}$, where $i = 1, \cdots, L$ and $t \geq 1$. In the general setup, we assume that the instantaneous joint probability distribution in the form

$$\Pr[x, y_1, \cdots, y_L] = P(x) \prod_{i=1}^{L} W(y_i|x)$$

for the temporally memoryless source $\{X(t)\}_{t=1}^{\infty}$. Here, the random variables $Y_i(t)$ are conditionally independent when $X(t)$ is given, and the conditional probabilities $W[y_i(t)|x(t)]$ are identical for all i and t. In this article, we impose the binary assumptions to the problem, i.e., the data sequence $\{X(t)\}$ and its noisy observations $\{Y_i(t)\}$ are all assumed to be binary sequences. Therefore, the stochastic matrix can be parameterized as

$$W(y|x) = \begin{cases} 1-p & (x = y) \\ p & (x \neq y) \end{cases} ,$$

where $p \in [0, 1/2]$ represents the observation noise. Note also that the alphabets have been selected as $\mathcal{X} = \mathcal{Y}$. Furthermore, for simplicity, we also assume that $P(x) = 1/2$ always holds, implying that a purely random source is observed [Fig. 1].

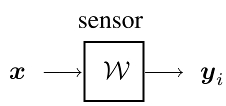

Fig. 1. Distributed sensing tasks for sensor i

At the encoding stage, a sensor i encodes a block $\boldsymbol{y}_i = [y_i(1), \cdots, y_i(n)]^T$ of length n from the noisy observation $\{y_i(t)\}_{t=1}^{\infty}$, into a block $\boldsymbol{z}_i = [z_i(1), \cdots, z_i(m)]^T$ of length m defined on $\mathcal{Z}$. Hereafter, we take the Boolean representation of the binary alphabet $\mathcal{X} = \{0, 1\}$, therefore $\mathcal{Y} = \mathcal{Z} = \{0, 1\}$ as well. Let $\hat{\boldsymbol{y}}_i$ be a reproduction sequence for the block, and we have a known integer $m < n$. Then, making use of a Boolean matrix A_i of dimensionality $n \times m$, we are to find an m bit codeword sequence $\boldsymbol{z}_i = [z_i(1), \cdots, z_i(m)]^T$ which satisfies

$$\hat{\boldsymbol{y}}_i = A_i \boldsymbol{z}_i \pmod 2 , \tag{1}$$

where the fidelity criterion

$$D = \frac{1}{n} d_{\mathrm{H}}(\boldsymbol{y}_i, \hat{\boldsymbol{y}}_i) \tag{2}$$

holds [3]. Here the Hamming distance $d_{\mathrm{H}}(\cdot, \cdot)$ is used for the distortion measure. Note that we have applied modulo-2 arithmetic for the additive operation in (1). Let A_i be characterized by K ones per row and C per column. The finite, and usually small, numbers K and C define a particular generator matrix A_i. The data center then collects the L codeword sequences, $\boldsymbol{z}_1, \cdots, \boldsymbol{z}_L$. Since all the L codewords are of the same length m, the combined data rate will be $R = L \times m/n$. Therefore, in our scenario, the data center deploys exchangeable sensors with fixed quality reproductions, $\hat{\boldsymbol{y}}_1, \cdots, \hat{\boldsymbol{y}}_L$ [Fig. 2].

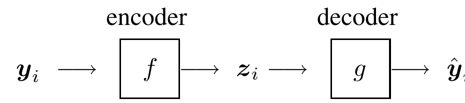

Fig. 2. Independent communications tasks for sensor i

Lastly, the tth symbol of the estimate, $\hat{\boldsymbol{x}} = [\hat{x}(1), \cdots, \hat{x}(n)]^T$, is to be calculated by optimal majority vote [5],

$$\hat{x}(t) = \begin{cases} 0 & (\hat{y}_1(t) + \cdots + \hat{y}_L(t) \leq L/2) \\ 1 & (\hat{y}_1(t) + \cdots + \hat{y}_L(t) > L/2) \end{cases} . \tag{3}$$

Therefore, overall performance of the system can be measured by the expected bit error frequency for decisions by (3), $P_{\mathrm{e}} = \Pr[x \neq \hat{x}]$ [Fig. 3].

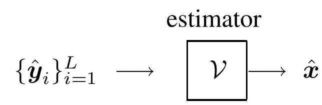

Fig. 3. Collective estimation tasks for the data center

In this article, we consider two limit cases of decentralization levels; (1) The extreme situation of $L \to \infty$, and (2) the case of $L = R$. The former case means that the data rate for an individual sensor information vanishes, while the latter case results in the transmission without quantization techniques. In general, it is difficult to determine which level is optimal for the estimation, i.e., which scenario results in the smaller value of P_{e}. Indeed, by using some quantization methods, such as the vector quantization, the data center could use as many sensors as possible for a given R. However, the quality of the individual reproduction would be less informative. The best choice seems to depend largely on R, as well as p.

III. Main Results

For simplicity, we consider the following two solvable cases; $K = 1$ for $C \geq K$ and the complicated case of $K = 2$. Let p be a given observation noise level, and R the finite real value of a given combined data rate. Letting $L \to \infty$, we find the expected bit error frequency to be

$$P_{\mathrm{e}}(p, R) = \int_{-\infty}^{-(1-2p)c_g\sqrt{R}} dr \; \mathrm{N}(0, 1) \tag{4}$$

with the constant value

$$c_g = \begin{cases} 1 & (K = 1) \\ \frac{1}{\sqrt{2}} \left[\frac{\sqrt{\alpha}}{2} + \frac{2\ln 2}{\sqrt{\alpha}} - \left(\frac{\sqrt{\alpha}}{2} - \frac{\sigma^2}{\sqrt{\alpha}} \right) \langle \tanh^2 x \rangle_{\pi(x)} \right] & (K = 2) \end{cases} \tag{5}$$

where the rescaled variance $\sigma^2 = \alpha \, \langle \hat{x}^2 \rangle_{\hat{\pi}(\hat{x})}$ and the *first step RSB* enforcement

$$-\frac{1}{2} + \frac{2}{\alpha} \ln 2 + \left(\frac{1}{2} - \frac{\sigma^2}{\alpha} \right) \langle \tanh^2 x \, (1 + 2x \; \mathrm{csch} \; x \; \mathrm{sech} \; x) \rangle_{\pi(x)} = 0$$

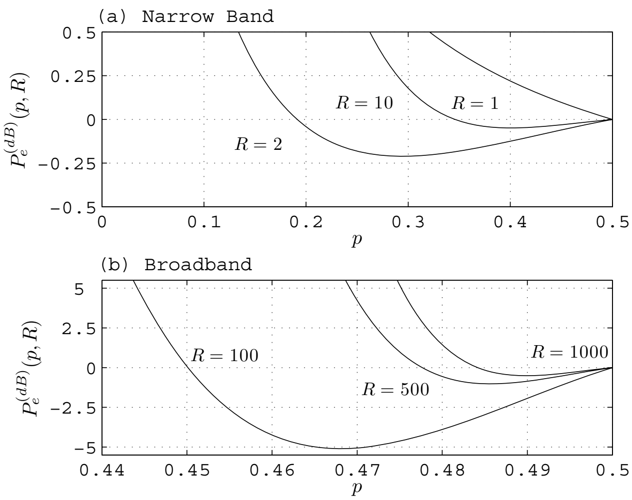

Fig. 4. $P_{\mathrm{e}}^{(\mathrm{dB})}(p,R)$ for $K = 1$. (a) Narrow band case with $R = 1, 2$, and, 10. (b) Broadband case with $R = 100, 500$, and, 1000. The critical value of p_c converges to the point 0.5.

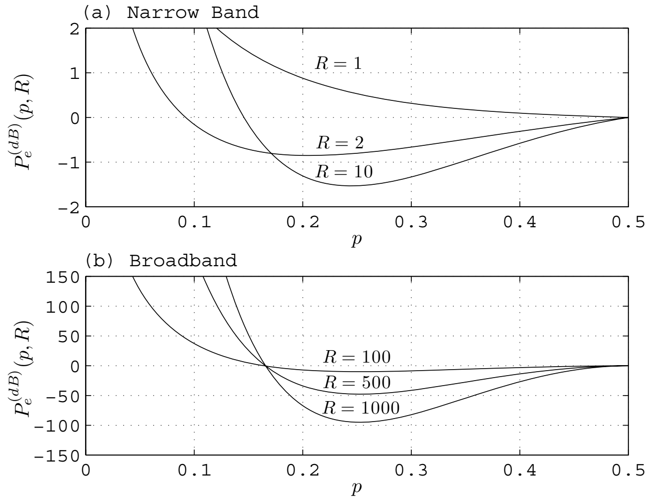

Fig. 5. $P_{\mathrm{e}}^{(\mathrm{dB})}(p,R)$ for $K = 2$. (a) Narrow band case with $R = 1, 2$, and, 10. (b) Broadband case with $R = 100, 500$, and, 1000. The critical value of p_c converges to the point 0.165.

holds. Here $\mathrm{N}(X,Y)$ denotes the normal distribution with the mean X and the variance Y. The rescaled variance σ^2 and the scale invariant parameter α is determined numerically, where we use the following notations.

$$\langle \cdot \rangle_{\pi(x)} = \int_{-\infty}^{\infty} \frac{dx}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{x^2}{2\sigma^2}\right] (\cdot) ,$$

$$\langle \cdot \rangle_{\hat{\pi}(\hat{x})} = \int_{-1}^{+1} \frac{d\hat{x}}{\sqrt{2\pi\sigma^2}} (1-\hat{x}^2)^{-1} \exp\left[-\frac{(\tanh^{-1}\hat{x})^2}{2\sigma^2}\right] (\cdot) .$$

Therefore, it is straightforward to evaluate (4) with (5) for given parameters, p and R.

For a given *finite* value of R, we see what happens to the quality of the estimate when the noise level p varies. Fig. 4 and Fig. 5 shows the typical behavior of the bit error frequency, $P_{\mathrm{e}}(p,R)$, in decibel (dB), where the reference level is chosen as

$$P_{\mathrm{e}}^{(0)}(p,R) = \begin{cases} \sum_{l=0}^{(R-1)/2} \binom{R}{l} (1-p)^l p^{R-l}, & (R \text{ is odd}) \\ \sum_{l=0}^{R/2-1} \binom{R}{l} (1-p)^l p^{R-l} \\ +\frac{1}{2}\binom{R}{R/2} (1-p)^{R/2} p^{R/2} & (R \text{ is even}) \end{cases} \tag{6}$$

for a given integer R. The reference (6) denotes P_{e} for the case of $L = R$, i.e., the case when the sensors are not allowed to compress their observations. Here, in decibel, we have

$$P_{\mathrm{e}}^{(\mathrm{dB})}(p,R) = 10 \log \frac{P_{\mathrm{e}}(p,R)}{P_{\mathrm{e}}^{(0)}(p,R)} ,$$

where the $\log$ is to base 10. Note that the zero level in decibel occurs when the measured error frequency $P_{\mathrm{e}}(p,R)$ is equal to the reference level. Therefore, it is also possible to have negative levels, which would mean an expected bit error frequency much smaller than the reference level. In the case of small combined data rate R, the narrow band case, the numerical results in Fig. 4 (a) and Fig. 5 (a) show that the quality of the estimate is sensitive to the parity of the integer R. In particular, the $R = 2$ case has the lowest threshold level p_c beyond which the $L \to \infty$ scenario outperforms the $L = R$ scenario, while the $R = 1$ case does not have such a threshold. In contrast, if the bandwidth is wide enough, the difference of the expected bit error probabilities in decibel, $P_{\mathrm{e}}^{(\mathrm{dB})}(p,R)$, is proved to have quite different qualitative characteristics as shown in Fig. 4 (b) and Fig. 5 (b). The critical value of p_c converges to the point 0.5 in the case of $K = 1$, while it seems to have nontrivial value of $p_c = 0.165$ in the case of $K = 2$.

IV. Statistics of Collective Estimation

Since the predetermined matrices $A_1, \cdots, A_L$ are selected randomly, it is quite natural to say that the instantaneous series, defined by $\hat{\boldsymbol{y}}(t) = [\hat{y}_1(t), \cdots, \hat{y}_L(t)]^T$, can be modeled using the Bernoulli trials. Here, the reproduction problem reduces to a channel model, where the stochastic matrix is defined as

$$W(\hat{y}|x) = \begin{cases} q & (x = \hat{y}) \\ 1-q & (x \neq \hat{y}) \end{cases} , \tag{7}$$

where q denotes the quality of the reproductions, i.e., $\Pr[x \neq \hat{y}_i] = 1 - q$ for $i = 1, \cdots, L$. Letting the channel model (7) for the reproduction problem be valid, the expected bit error frequency can be well captured by using the cumulative probability distributions

$$\Pr[x \neq \hat{x}] = \begin{cases} B(\frac{L-1}{2} : L, q), & (L \text{ is odd}) \\ B(\frac{L}{2} - 1 : L, q) + \frac{1}{2} b(\frac{L}{2} : L, q) & (L \text{ is even}) \end{cases} \tag{8}$$

with

$$B(L':L,q) = \sum_{l=0}^{L'} b(l:L,q) \ ,$$
$$b(l:L,q) = \binom{L}{l} q^l (1-q)^{L-l} \ ,$$

where an integer l be the total number of non-flipped elements in $\hat{\boldsymbol{y}}(t)$, and the second term $(1/2)b(L/2 : L, q)$ represents random guessing with $l = L/2$. Note that the reproduction quality q can be easily obtained by the simple algebra $q = pD + (1-p)(1-D)$, where D is the distortion with respect to coding.

Lastly, by using the cumulative probability distribution (8), we get

$$P_{\rm e} = \sum_{l=0}^{L/2} \binom{L}{l} q^l (1-q)^{L-l} \sim \int_0^{L/2} dr \ \mathrm{N}(Lq, Lq(1-q)) \ . \quad (9)$$

It is easy to see that (9) can be converted to a standard normal distribution by changing variables to [6]

$$\tilde{r} = \frac{r - Lq}{\sqrt{Lq(1-q)}} \ ,$$

so

$$d\tilde{r} = dr/\sqrt{Lq(1-q)} \ ,$$

yielding the general formula

$$P_{\rm e} \sim \int_{-\sqrt{L}}^{\tilde{r}_g} d\tilde{r} \ \mathrm{N}(0,1) \quad (10)$$

with

$$\tilde{r}_g = 2\sqrt{L}(1-2p)\left(D - \frac{1}{2}\right) \ . \quad (11)$$

Since the error probability (10) is given by a function of D, we need to derive an analytical solution for the quality D in the limit $L \to \infty$, keeping R finite. In this approach, we apply the method of statistical mechanics to evaluate the *typical* performance of the separate decoding [3].

V. Method of Statistical Physics

As a first step, we translate the Boolean alphabets $\mathcal{Z} = \{0,1\}$ to the "Ising" ones, $\mathcal{S} = \{+1,-1\}$. Consequently, we need to translate the additive operations, such as, $z_i(s) + z_i(s') \pmod 2$ into their multiplicative representations, $\sigma_i(s) \times \sigma_i(s') \in \mathcal{S}$ for $s, s' = 1, \cdots, m$. Similarly, we translate the Boolean $y_i(t)$s into the Ising $J_i(t)$s. For simplicity, we omit the subscript i, which labels the L agents, in the rest of this section. Following the prescription of Sourlas [7], we examine the *Gibbs-Boltzmann distribution*

$$\Pr[\boldsymbol{\sigma}] = \frac{\exp\left[-\beta H(\boldsymbol{\sigma}|\boldsymbol{J})\right]}{Z(\boldsymbol{J})} \quad (12)$$

with the partition function

$$Z(\boldsymbol{J}) = \sum_{\boldsymbol{\sigma}} e^{-\beta H(\boldsymbol{\sigma}|\boldsymbol{J})} \ ,$$

where the *Hamiltonian* of the Ising system is defined as

$$H(\boldsymbol{\sigma}|\boldsymbol{J}) = -\sum_{s_1 < \cdots < s_K} \mathcal{A}_{s_1 \ldots s_K} J[t(s_1, \ldots, s_K)] \sigma(s_1) \ldots \sigma(s_K) \ . \quad (13)$$

The observation index $t(s_1, \ldots, s_K)$ specifies the proper value of t given the set $s_1, \ldots, s_K$, so that it corresponds to the parity check equation (1). Here the elements of the symmetric tensor $\mathcal{A}_{s_1 \ldots s_K}$, representing dilution, is either zero or one depending on the set of indices $(s_1, \ldots, s_K)$. Since there are C non-zero elements randomly chosen for any given index s, we find $\sum_{s_2, \ldots, s_K} \mathcal{A}_{s s_2 \ldots s_K} = C$. The code rate is $R/L = K/C$ because a reproduction sequence has C bits per index s and carries K bits of the codeword. It is easy to see that the Hamiltonian (13) is counting the reproduction errors, $[1 - J[t(s_1, \ldots, s_K)] \cdot \sigma(s_1) \ldots \sigma(s_K)]/2$.

Moreover, according to the statistical mechanics, we can easily derive the *observable* quantities using the per-bit *free energy* defined as

$$f = -\frac{1}{\beta n} \left\langle \ln Z(\boldsymbol{J}) \right\rangle_{\mathcal{A}, \boldsymbol{J}} \quad (14)$$

which carries all information about the statistics of the system. Here, β denotes an *inverse temperature* for the Gibbs-Boltzmann distribution (12), and $\langle \cdot \rangle_{\mathcal{A}, \boldsymbol{J}}$ represents the configurational average. Therefore, we have to average the logarithm of the partition function $Z(\boldsymbol{J})$ over the given distribution $\langle \cdot \rangle_{\mathcal{A}, \boldsymbol{J}}$ after the calculation of the partition function.

A. Case of $K = 1$

Obviously, the simple case of $K = 1$ does not induce frustrations in the system. Indeed, the case corresponds to a naive vector quantization scheme, in which the encoding as well as decoding only requires easy calculations. In the analytical point of view, we may resort to straightforward and simple calculations for evaluating such a free energy; we do not require any strange 'tricks'.

Using the set $\mathcal{M}(s) = \{t | t \text{ s.t. } a_{ts} = 1\}$ with $A = (a_{ts})$, we can rewrite (14) as

$$f = -\frac{1}{\beta n} \left\langle \ln \left[\sum_{\boldsymbol{\sigma}} \exp\left(\beta \sum_{s=1}^{m} \sum_{t \in \mathcal{M}(s)} J(t) \cdot \sigma(s) \right) \right] \right\rangle_{\mathcal{M}, \boldsymbol{J}} \ . \quad (15)$$

Since we know that the realizations of $J(t)$s is independent with respect to the t, the function is reduced to the expression:

$$f = -\frac{L}{\beta R} \left\langle \ln \left[\sum_{\sigma = \pm 1} \exp\left(\beta \bar{J} \cdot \sigma \right) \right] \right\rangle_{\bar{J}} \ , \quad (16)$$

where the corresponding models are characterized by only the one-body interactions $\bar{J}$s. In parallel with the random walk statistics, we will impose here the 'mean-field' approximation for $\bar{J}$ such that $\bar{J} \approx \sqrt{L/R}$. This approximation enables us to write the average free energy as simple as

$$f = -\frac{L}{\beta R} \ln \left[2 \cosh \left(\beta \sqrt{\frac{L}{R}} \right) \right] \ . \tag{17}$$

Lastly, one can invoke the general relation $D = (1 + f)/2$ to find the average distortion:

$$D = \frac{1}{2} \left[1 + \frac{\tanh(\beta\sqrt{L/R})}{\sqrt{L/R}} \right] \tag{18}$$

with the optimal condition of $\beta \to \infty$. Therefore, it is easy to obtain

$$D = \frac{1}{2} + \frac{1}{2}\sqrt{\frac{R}{L}} \ . \tag{19}$$

Inserting (19) into (11) in the definite integral (10), we can easily see (4) holds.

B. Case of $K = 2$

To perform a similar program in the more complicated case of $K = 2$, the *replica trick* is now used [8]. The theory of *replica symmetry breaking* can provide the free energy resulting in the expression for the general Ks [3]

$$\begin{aligned} f = -\frac{1}{\beta n} \Bigg[& \ln\cosh\beta - K \left\langle \ln\left[1 + \tanh(\beta x)\tanh(\beta \hat{x})\right]\right\rangle_{\pi(x),\hat{\pi}(\hat{x})} \\ & + \frac{1}{2} \left\langle \sum_{J=\pm 1} \ln \left[1 + \tanh(\beta J) \prod_{l=1}^{K} \tanh(\beta x_l) \right] \right\rangle_{\pi(\boldsymbol{x})} \\ & + \frac{C}{K} \left\langle \ln \sum_{\sigma=\pm 1} \prod_{l=1}^{C} [1 + \sigma \tanh(\beta \hat{x}_l)] \right\rangle_{\hat{\pi}(\hat{\boldsymbol{x}})} \Bigg] \ , \end{aligned} \tag{20}$$

where $\langle \cdot \rangle_{\pi(\boldsymbol{x})}$ denotes the averaging over $p(x_l)$s and so on. The variation of (20) by $\pi(\boldsymbol{x})$ and $\hat{\pi}(\hat{\boldsymbol{x}})$ under the condition of normalization gives the saddle point condition

$$\begin{aligned} \pi(x) &= \left\langle \delta \left[x - \sum_{l=1}^{C-1} \hat{x}_l \right] \right\rangle_{\hat{\pi}(\hat{\boldsymbol{x}})} \ , \\ \hat{\pi}(\hat{x}) &= \left\langle \frac{1}{2} \sum_{J=\pm 1} \delta\left[\hat{x} - \mu(x_1, \ldots, x_{K-1}; J)\right] \right\rangle_{\pi(\boldsymbol{x})} \ , \end{aligned} \tag{21}$$

where

$$\mu(x_1, \ldots, x_{K-1}; J) = \frac{1}{\beta} \tanh^{-1} \left[\tanh(\beta J) \prod_{l=1}^{K-1} \tanh(\beta x_l) \right] \ .$$

The stability of the solution (21) is well investigated in [9] in the context of examining the effect of 'diluted' interactions in spin glasses [10].

We now focus on the case of $K = 2$. Applying the central limit theorem to $\pi(x)$ [6], we get

$$\pi(x) = \frac{1}{\sqrt{2\pi C \sigma^2}} e^{-\frac{x^2}{2C\sigma^2}} \ , \tag{22}$$

where σ^2 is the variance of $\hat{\pi}(\hat{x})$. Here the resulting distribution (22) is a even function. The leading contribution to μ is then given by $\mu(x; J) \sim J \cdot \tanh(\beta x)$ as β goes to zero; The expression is valid in the asymptotic region $L \gg 1$ for a fixed R. Then, the formula for the delta function yields [11]

$$\begin{aligned} \hat{\pi}(\hat{x}) &= \left\langle \delta \left[x - \frac{1}{\beta} \tanh^{-1} \hat{x} \right] \left| \rho' \left(\frac{1}{\beta} \tanh^{-1} \hat{x}; \hat{x} \right) \right|^{-1} \right\rangle_{\pi(x)} \\ &= \frac{(1-\hat{x}^2)^{-1}}{\sqrt{2\pi\beta^2 C\sigma^2}} \exp\left[-\frac{(\tanh^{-1}\hat{x})^2}{2\beta^2 C\sigma^2} \right] \ , \end{aligned} \tag{23}$$

where we have used

$$\rho(x; \hat{x}) = \hat{x} - \tanh(\beta x) \ .$$

Therefore, we have

$$\begin{aligned} \sigma^2 &= \langle \hat{x}^2 \rangle_{\hat{\pi}(\hat{x})} \\ &= \int_{-1}^{+1} \frac{d\hat{x}}{\sqrt{2\pi\beta^2 C\sigma^2}} \frac{\hat{x}^2}{1-\hat{x}^2} \exp\left[-\frac{(\tanh^{-1}\hat{x})^2}{2\beta^2 C\sigma^2} \right] \end{aligned}$$

for given $\beta^2 C$. Inserting (22), (23) into (20), we get

$$f = -\frac{\beta}{2} - \frac{R}{\beta} \ln 2 + \frac{1 - 2\sigma^2}{2} \beta \left\langle \tanh^2 \tilde{x} \right\rangle_{\tilde{\pi}(\tilde{x})}$$

with

$$\tilde{\pi}(\tilde{x}) = \frac{1}{\sqrt{2\pi\beta^2 C\sigma^2}} e^{-\frac{\tilde{x}^2}{2\beta^2 C\sigma^2}} \ ,$$

where we rewrite $\tilde{x} = \beta x$. The theory of *replica symmetry breaking* tells us that relevant value of β should not be smaller than the "freezing point" β_g, which implies the vanishing entropy condition:

$$\begin{aligned} \frac{\partial f}{\partial \beta} &= -\frac{1}{2} + \frac{2}{\beta_g^2 C} \ln 2 \\ &+ \frac{1 - 2\sigma^2}{2} \left\langle \tanh^2 \tilde{x}\ (1 + 2\tilde{x}\ \mathrm{csch}\ \tilde{x}\ \mathrm{sech}\ \tilde{x}) \right\rangle_{\tilde{\pi}(\tilde{x})} = 0 \ . \end{aligned}$$

Accordingly, it is convenient for us to define a scaling invariant parameter $\alpha = \beta_g^2 C$, and to rewrite the variance $\tilde{\sigma}^2 = \alpha\sigma^2$ for simplicity. Introducing these newly defined parameters, the above results could be summarized as follows. Given R and L, we find

$$f = \sqrt{\frac{R}{L}} \left[-\frac{1}{2}\sqrt{\frac{\alpha}{2}} - \ln 2 \sqrt{\frac{2}{\alpha}} + \sqrt{\frac{\alpha}{2}} \left(\frac{1}{2} - \frac{\tilde{\sigma}^2}{\alpha} \right) \langle \tanh^2 \tilde{x} \rangle_{\tilde{\pi}(\tilde{x})} \right]$$

with

$$\tilde{\sigma}^2 = \alpha\ \langle \hat{x}^2 \rangle_{\hat{\pi}(\hat{x})} \ ,$$

where the condition

$$\begin{aligned} & -\frac{1}{2} + \frac{2}{\alpha} \ln 2 \\ & + \left(\frac{1}{2} - \frac{\tilde{\sigma}^2}{\alpha} \right) \left\langle \tanh^2 \tilde{x}\ (1 + 2\tilde{x}\ \mathrm{csch}\ \tilde{x}\ \mathrm{sech}\ \tilde{x}) \right\rangle_{\tilde{\pi}(\tilde{x})} = 0 \end{aligned} \tag{24}$$

holds [12]. Here we denote

$$\langle \cdot \rangle_{\tilde{\pi}(\tilde{x})} = \int_{-\infty}^{\infty} \frac{d\tilde{x}}{\sqrt{2\pi\tilde{\sigma}^2}} \exp\left[-\frac{\tilde{x}^2}{2\tilde{\sigma}^2}\right] (\cdot) \ ,$$

$$\langle \cdot \rangle_{\hat{\pi}(\hat{x})} = \int_{-1}^{+1} \frac{d\hat{x}}{\sqrt{2\pi\tilde{\sigma}^2}} (1-\hat{x}^2)^{-1} \exp\left[-\frac{(\tanh^{-1}\hat{x})^2}{2\tilde{\sigma}^2}\right] (\cdot) \ .$$

Note that the relation $D = (1+f)/2$ holds at the vanishing entropy condition (24) [3]. Finally, we obtain the main result (4) in Section III in the limit $L \to \infty$, when we use proper notations for the variables and the name of the function.

VI. Conclusion

This article provides a system-level perspective for large-scale sensor networks. The decentralized sensing problem argued in this article was first addressed by Berger and his collaborators. In the present work, we imposed strict restrictions for the decoding stage to give a practical scheme to analyze a class of low-complexity quantization methods in the given finite data rate. Surprisingly, our results show the existence of threshold level of noise of which the optimal levels of decentralization changes. Future work includes the theoretical derivation of the threshold level p_c where R goes to infinity, as well as the implementation problem.

Acknowledgments: The authors thank Jun Muramatsu and Yasutada Oohama for useful discussions. This work was in part supported by the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan, under the Grant-in-Aid for Scientific Research on Priority Areas, 18079015.

References

[1] (2005) Intel@Mote. [Online]. Available: http://www.intel.com/research/exploratory/motes.htm

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

[3] T. Murayama and M. Okada, “Rate distortion function in the spin glass state: a toy model,” in *Advances in Neural Information Processing Systems 15 (NIPS’02)*, Vancouver, Canada, Dec. 2002, pp. 423–430.

[4] T. Murayama, “On the asymptotic performance of sparse matrix codes in the CEO problem,” in *Proc. 2005 IEEE International Symposium on Information Theory (ISIT’05)*, Adelaide, Australia, 2005.

[5] D. J. C. MacKay, *Information Theory, Inference and Learning Algorithms*. Cambridge, UK: Cambridge University Press, 2003.

[6] W. Hays, *Statistics (5th Edition)*. Belmont, CA: Wadsworth Publishing, 1994.

[7] N. Sourlas, “Spin-glass models as error-correcting codes,” *Nature*, vol. 339, pp. 693–695, June 1989.

[8] V. Dotsenko, *Introduction to the Replica Theory of Disordered Statistical Systems*. Cambridge, UK: Cambridge University Press, 2001.

[9] A. Montanari and F. Ricci-Tersenghi, “On the nature of the low-temperature phase in discontinuous mean-field spin glasses,” *Eur. Phys. J. B.*, vol. 33, p. 339, 2003.

[10] S. Franz, M. Leone, F. Ricci-Tersenghi, and R. Zecchina, “Exact solutions for diluted spin glasses and optimization problems,” *Phys. Rev. Lett.*, vol. 87, p. 127209, 2001.

[11] C. W. Wong, *Introduction to Mathematical Physics: Methods and Concepts*. Oxford, UK: Oxford University Press, 1991.

[12] T. Murayama and P. Davis, “Rate distortion codes in sensor networks: A system-level analysis,” in *Advances in Neural Information Processing Systems 18 (NIPS’05)*, Vancouver, Canada, Dec. 2005, pp. 931–938.