Multi-Round Computation of Type-Threshold Functions in Collocated Gaussian Networks

Chien-Yi Wang
School of Computer and Communication Sciences
EPFL
Lausanne, Switzerland
chien-yi.wang@epfl.ch

Sang-Woon Jeon
Department of Information and Communication Engineering
Andong National University
Andong, Korea
swjeon@anu.ac.kr

Michael Gastpar†
School of Computer and Communication Sciences
EPFL
Lausanne, Switzerland
michael.gastpar@epfl.ch

***Abstract*—In wireless sensor networks, various applications involve learning one or multiple functions of the measurements observed by sensors, rather than the measurements themselves. This paper focuses on the computation of type-threshold functions which include the maximum, minimum, and indicator functions as special cases. Previous work studied this problem under the collocated collision network model and showed that under many probabilistic models for the measurements, the achievable computation rates tend to zero as the number of sensors increases.**

In this paper, wireless sensor networks are modeled as fully connected Gaussian networks with equal channel gains, which are termed collocated Gaussian networks. A general multi-round coding scheme exploiting not only the broadcast property but also the superposition property of Gaussian networks is developed. Through careful scheduling of concurrent transmissions to reduce redundancy, it is shown that given any independent measurement distribution, all type-threshold functions can be computed reliably with a non-vanishing rate even if the number of sensors tends to infinity.

I. Introduction

Consider a wireless sensor network consisting of multiple sensors and one fusion center. Suppose that the fusion center only needs to learn whether one (or more) of the sensor readings exceed a certain threshold. What communication resources are needed to attain this goal?

Such a problem setting is well motivated by applications such as environmental monitoring, e.g., air/water quality monitoring and forest fire detection, in which the fusion center is only interested in acquiring an *indication* or, more generally, a *function* of the measurements, rather than the measurements themselves.

The problem of function computation in wireless sensor networks has recently received significant attention. One interesting problem formulation was developed by Giridhar and Kumar [1]. They model the wireless medium as a collision channel: Any single transmission is received by all other nodes. Concurrent transmissions by multiple nodes result in a collision. They considered the class of *symmetric* functions and the subclasses of *type-sensitive* and *type-threshold* functions. Type-threshold functions include the maximum, minimum, and indicator functions as special cases and they are the main focus of this paper.

For the collocated collision network model, Giridhar and Kumar showed that the worst-case scaling law of type-threshold functions is $\Theta(\frac{1}{\log K})$ as the number of sensors K increases. Here the worst case means the worst source (measurement) distribution for computing the desired function, which may depend on K. Later, Ma, Ishwar, and Gupta [2] followed the same model and studied the problem within the framework of interactive source coding. Still, the worst-case scaling law of type-threshold functions is $\Theta(\frac{1}{\log K})$.

On the other hand, Subramanian, Gupta, and Shakkottai [3] showed that the scaling law of $\Theta(1)$ is achievable for the type-threshold function computation assuming that the number of nodes in the vicinity, i.e., the number of nodes within a direct communication range, is upper bounded by a fixed number, independent of K. Furthermore, Kowshik and Kumar [4] showed that, if the source distribution is independent of K, then the scaling law of $\Theta(1)$ is achievable.

In this paper, instead of the collocated collision network model, we consider a simplified Gaussian network in which all nodes are connected to each other and all channel gains are equal. Our achievability follows by extending the linear computation coding [5] (and, more directly, the recently developed strategy in [6]) to *multi-round communication*. Through careful scheduling of concurrent transmissions, the amount of information which is redundant with respect to the desired function can be greatly reduced. We show that for any independent source distribution, all type-threshold functions can be computed reliably with a non-vanishing rate even as K tends to infinity. Finally, we show how to tailor the proposed multi-round coding scheme to specific functions in order to achieve even higher rates.

II. Problem Statement

Throughout the paper, all logarithms are to base 2. We denote $[1:K] := \{1, 2, \cdots, K\}$ and $\log^+(x) = \max\{\log(x), 0\}$. Let $|\cdot|$ denote the cardinality of a set and $\mathbf{1}_{(\cdot)}$ denote the indicator function of an event. Random variables are represented by uppercase letters, while their realizations are represented by lowercase letters. Given a function f and vectors $\mathbf{x} = [x_1 \ \cdots \ x_l]$, $\mathbf{y} = [y_1 \ \cdots \ y_l]$, we denote $f(\mathbf{x}, \mathbf{y}) = [f(x_1, y_1) \ \cdots \ f(x_l, y_l)]$.

†He is also with the Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA, USA.

A. Network Model

Consider a sensor network consisting of K sensors labeled nodes $1, 2, \cdots, K$ and a fusion center labeled node 0. Sensor node i observes a length-k source vector $\mathbf{s}_i = \begin{bmatrix} s_i[1] & \cdots & s_i[k] \end{bmatrix} \in [0:q-1]^k$ with $q \geq 2$. For $j \in [1:k]$, the sources $s_1[j], \cdots, s_K[j]$ are assumed to be independently drawn from a joint probability mass function $p_{S_1,\cdots,S_K}$. The fusion center itself has no observation and wishes to compute a symbol-by-symbol function of K sources, i.e., $f(\mathbf{s}_1, \cdots, \mathbf{s}_K)$.

The length-n time-extended input–output is given by

$$\mathbf{y}_i = \sum_{j \in [1:K] \setminus \{i\}} \mathbf{x}_j + \mathbf{z}_i,$$

where $i \in [0:K]$ and the elements of $\mathbf{z}_i$ are independent and identically distributed (i.i.d.) drawn from $\mathcal{N}(0,1)$. Each transmit signal should satisfy the average power constraint P, i.e., $\frac{1}{n}\|\mathbf{x}_j\|^2 \leq P$. We assume full-duplex in which each node can transmit and receive simultaneously. We term the above network model the *collocated Gaussian network*.

B. Computation Capacity

A length-n block code for function computation in the collocated Gaussian network is defined as

- (Sensor Node Encoding) At time $t \in [1:n]$, sensor node $i \in [1:K]$ broadcasts $x_i[t] = \mathcal{E}_i^{(t)}\left(\mathbf{s}_i, \mathbf{y}_i^{t-1}\right)$.
- (Fusion Center Decoding) The fusion center estimates $\hat{f}\left(\mathbf{s}_1, \cdots, \mathbf{s}_K\right) = \mathcal{D}\left(\mathbf{y}_0\right)$.

Here $\mathbf{y}_i^{t-1}$ denotes the length-$(t-1)$ vector containing the first $t-1$ elements of $\mathbf{y}_i$.

The probability of error is defined as $P_e^{(n)} := \mathsf{P}\left(\bigcup_{j=1}^{k}\left\{\hat{f}(s_1[j], \cdots, s_K[j]) \neq f(s_1[j], \cdots, s_K[j])\right\}\right)$. We say that the computation rate $R := \frac{k}{n}$ is achievable if there exists a sequence of length-n block codes such that $P_e^{(n)}$ converges to zero as n tends to infinity. Note that the computation rate is the number of reliably computed functions per channel use. The computation capacity C is the supremum over all achievable computation rates.

C. Type-Threshold Function

Before defining the type-threshold functions, we first introduce symmetric functions, which is a more general class containing all type-threshold functions. Symmetric functions $f : [0:q-1]^K \to \mathcal{R}$ are invariant under any permutation σ of its arguments, i.e., $f(s_{\sigma(1)}, \cdots, s_{\sigma(K)}) = f(s_1, \cdots, s_K)$. Thus, for symmetric functions, we simply denote $f(s_1, \cdots, s_K)$ as $f(\{s_i\})$. The type vector of $\{s_i\}_{i \in [1:K]}$ is defined as the length-q vector $\mathbf{b}$ where the $(l+1)$-th element $b_l := |\{i \in [1:K] | s_i = l\}|$ for $l \in [0:q-1]$. Then, every symmetric function can be represented as $f(\{s_i\}) = f'(b_0, \cdots, b_{q-1})$ by some function f' with domain $\mathcal{B} := \{\mathbf{b} \in [0:K]^q | \sum_{l=0}^{q-1} b_l = K\}$.

A symmetric function is said to be type-threshold if there exists a non-negative length-q vector $\boldsymbol{\theta}$, called *threshold vector*, such that for all $\mathbf{b} \in \mathcal{B}$, $f'(b_0, \cdots, b_{q-1}) = f'(\overline{b}_0, \cdots, \overline{b}_{q-1})$, where $\overline{b}_l = \min\{b_l, \theta_l\}$. We denote the clipped type vector as $\overline{\mathbf{b}} = \begin{bmatrix} \overline{b}_0 & \cdots & \overline{b}_{q-1} \end{bmatrix}$. From now on, we only consider type-threshold functions.

III. Preliminaries

A. Cut-Set Upper Bound

The computation capacity can be upper bounded from a cut-set based argument. Let $\Omega \subseteq [1:K]$ and $\Omega^c := [1:K] \setminus \Omega$. First, assume that a genie provides $\{s_i\}_{i \in \Omega^c}$ to the fusion center. Also, we allow sensor nodes in Ω to exchange each other's source information, so from the nodes in Ω to the fusion center becomes a point-to-point multiple-input single-output (MISO) channel in which the source–channel separation theorem holds and feedback does not increase the capacity since the channel is memoryless. Given $\{s_i\}_{i \in \Omega^c}$ as side information at the fusion center, Orlitsky and Roche [7] showed that the source coding rate should be at least the conditional graph entropy $H_{\mathcal{G}_\Omega}(\{S_i\}_{i \in \Omega} | \{S_i\}_{i \in \Omega^c})$, where $\mathcal{G}_\Omega$ is the characteristic graph of $(\{S_i\}_{i \in \Omega}, \{S_i\}_{i \in \Omega^c}, f)$. Therefore, we have

$$C \leq \min_{\Omega \subseteq [1:K]} \frac{\frac{1}{2}\log(1 + |\Omega|^2 P)}{H_{\mathcal{G}_\Omega}(\{S_i\}_{i \in \Omega} | \{S_i\}_{i \in \Omega^c})}.$$

B. Round-Robin Broadcast with Interactive Source Coding

The round-robin approach within the framework of interactive source coding follows from [2]. The whole communication consists of t rounds. In each round, only one sensor node is activated and broadcasts a common message to all other nodes in the network with rate $\frac{1}{2}\log(1+P)$. After t rounds, the fusion center computes the desired function based on the received t messages. Note that t can be larger than K, which means each sensor can act as a sender for multiple rounds and the communication becomes *interactive* between sensors.

Assume that the communication is initiated by sensor node i and let $R_j^{(i)}$ denote the block-coding rate to transmit the j-th message. If the sources are independent, then the minimum sum rate $\sum_{j=1}^{t} R_j^{(i)}$ such that the function can be computed reliably is characterized in [2].[1] Therefore, for independent sources, the interactive round-robin approach achieves[2]

$$R_{\mathsf{rr}}^{(t)} = \frac{\frac{1}{2}\log(1+P)}{\sum_{j=1}^{t} R_j^{(i)}}. \tag{1}$$

IV. Computation in Collocated Gaussian Networks

In general, it is unclear how to let the fusion center learn $f'(\overline{b}_0, \cdots, \overline{b}_{q-1})$ directly. One natural approach is to convey some sufficient information so that the fusion center learns $\overline{\mathbf{b}}$. Our developed coding scheme is termed *multi-round group broadcast* in which sensor nodes are partitioned into groups and in each round only members of the activated group cooperatively broadcast partial information of a type.

Theorem 1: Consider type-threshold function computation in collocated Gaussian networks with respect to the threshold vector $\boldsymbol{\theta}$. For all $l \in [0:q-1]$, let $J_l \in [1:K]$ and $\{\mathcal{A}_1^{(l)}, \cdots, \mathcal{A}_{J_l}^{(l)}\}$ be a partition of $[1:K]$, i.e.,

$$\bigcup_{m=1}^{J_l} \mathcal{A}_m^{(l)} = [1:K] \text{ and } \mathcal{A}_i^{(l)} \bigcap \mathcal{A}_j^{(l)} = \emptyset \text{ for all } i \neq j.$$

[1] We refer readers to [2] for the rate expression.

[2] The achievability can be improved through power control. Due to page limitation, we refer to the full paper in preparation for details.

The computation rate

$$R=\left(\sum_{l=0}^{q-1}\sum_{m=1}^{J_l}\frac{H\left(U_m^{(l)}\middle|U_{m-1}^{(l)},\cdots,U_1^{(l)},\{U_i\}_{i=0}^{l-1}\right)}{\frac{1}{2}\log^+\left(\frac{1}{|\mathcal{A}_m^{(l)}|}+P\right)}\right)^{-1} \quad (2)$$

is achievable, where

$$U_m^{(l)}=U_{m-1}^{(l)}+\textstyle\sum_{i\in\mathcal{A}_m^{(l)}}\mathbf{1}_{\{U_{m-1}^{(l)}<\theta_l\}\bigcap\{S_i=l\}},$$

$U_0^{(l)}=0$, and $U_l=(U_1^{(l)},\cdots,U_{J_l}^{(l)})$.

Remark 1: If we set $J_l=1$ and $\mathcal{A}_1^{(l)}=[1:K]$ for all $l\in[0:q-1]$, termed 1-partition, then Theorem 1 reduces to [6, Theorem 1] and the fusion center learns the complete type vector. On the other hand, if we set $J_l=K$ and $\mathcal{A}_m^{(l)}=\{m\}$ for all $m\in[1:J_l]$, $l\in[0:q-1]$, termed K-partition, then Theorem 1 reduces to a round-robin scheme without the sophisticated interactive source coding. Similarly, the coding scheme can be improved through power control, which will be addressed in the full paper in preparation.

Proof: Define the following recursion for all $j\in[1:k]$:

$$u_m^{(l)}[j]=u_{m-1}^{(l)}[j]+\textstyle\sum_{i\in\mathcal{A}_m^{(l)}}v_{im}^{(l)}[j],$$

where $u_0^{(l)}[j]=0$ and $v_{im}^{(l)}[j]=\mathbf{1}_{\{u_{m-1}^{(l)}[j]<\theta_l\}\bigcap\{s_i[j]=l\}}$. Also, we form $\mathbf{u}_m^{(l)}=\begin{bmatrix}u_m^{(l)}[1] & \cdots & u_m^{(l)}[k]\end{bmatrix}$ and $\mathbf{v}_{im}^{(l)}=\begin{bmatrix}v_{im}^{(l)}[1] & \cdots & v_{im}^{(l)}[k]\end{bmatrix}$.

The whole communication is divided into q stages. Consider stage $l\in[0:q-1]$ in which the communication consists of J_l rounds. In round $m\in[1:J_l]$, sensor nodes in $\mathcal{A}_m^{(l)}$ cooperate to let all nodes learn $\mathbf{u}_m^{(l)}$ using $n_m^{(l)}$ time slots, which will be specified shortly.

Since sensor node $i\in\mathcal{A}_m^{(l)}$ learns $\mathbf{u}_{m-1}^{(l)}$ in the previous round, it can compute $\mathbf{v}_{im}^{(l)}$. To let the receiving nodes learn $\sum_{i\in\mathcal{A}_m^{(l)}}\mathbf{v}_{im}^{(l)}$, we combine the linear computation coding [5] and the compute-and-forward [8] as in [6, Theorem 1]. Then, the $\sum_{i\in\mathcal{A}_m^{(l)}}\mathbf{v}_{im}^{(l)}$ can be computed reliably as k increases if

$$n_m^{(l)}\geq\frac{kH\left(\sum_{i\in\mathcal{A}_m^{(l)}}V_{im}^{(l)}\right)}{\frac{1}{2}\log^+\left(\frac{1}{|\mathcal{A}_m^{(l)}|}+P\right)},$$

where $V_{im}^{(l)}=\mathbf{1}_{\{U_{m-1}^{(l)}<\theta_l\}\bigcap\{S_i=l\}}$.

Actually, treating the messages received in previous rounds and stages as receiver side information and noting that $U_m^{(l)}=U_{m-1}^{(l)}+\sum_{i\in\mathcal{A}_m^{(l)}}V_{im}^{(l)}$, it suffices that

$$n_m^{(l)}\geq\frac{kH\left(U_m^{(l)}\middle|U_{m-1}^{(l)},\cdots,U_1^{(l)},\{U_i\}_{i=0}^{l-1}\right)}{\frac{1}{2}\log^+\left(\frac{1}{|\mathcal{A}_m^{(l)}|}+P\right)}. \quad (3)$$

Note that $\min\{u_{J_l}^{(l)}[j],\theta_l\}=\bar{b}_l[j]$. Thus, using $n=\sum_{l=0}^{q-1}\sum_{m=1}^{J_l}n_m^{(l)}$ time slots, the fusion center learns $(\mathbf{u}_{J_0}^{(0)},\mathbf{u}_{J_1}^{(1)},\cdots,\mathbf{u}_{J_{q-1}}^{(q-1)})$ and can compute $f(\{\mathbf{s}_i\})$. Finally, the theorem follows by calculating the computation rate $R=\frac{k}{n}$ using (3) and $n=\sum_{l=0}^{q-1}\sum_{m=1}^{J_l}n_m^{(l)}$. ■

Depending on the joint source distribution and the threshold vector, one can optimize the partitions $\{\mathcal{A}_m^{(l)}\}$ for all $l\in[0:q-1]$ to find the highest computation rate in Theorem 1. By lower bounding $\log^+\left(\frac{1}{|\mathcal{A}_m^{(l)}|}+P\right)$ by $\log^+(P)$, the rate expression (2) can be simplified as

$$R\geq\frac{\frac{1}{2}\log^+(P)}{H\left(U_0,\cdots,U_{q-1}\right)}. \quad (4)$$

V. Scaling Law for Number of Sensors

In this section, we study the scaling law of the computation rate as the number of sensors K increases. For this, we consider the source distribution ensembles with respect to K. A distribution ensemble is a family of probability distributions and we consider $\{p_{S_1,\cdots,S_K}\}_{K\in\mathbb{N}}$ in this paper. In this section, we assume that the sources are independent, i.e., $p_{S_1,\cdots,S_K}=\prod_{i=1}^K p_{S_i}$. First, we investigate the maximum function of binary sources, which is probably the simplest but most representative example of type-threshold functions. Second, given any distribution ensemble, we show that by carefully choosing the partitions $\{\mathcal{A}_m^{(l)}\}$ for all $l\in[0:q-1]$, Theorem 1 provides a non-vanishing computation rate even if the number of sensors K tends to infinity.

A. The Binary Maximum Problem

Assume that $S_i\in\{0,1\}$ follows Bernoulli(p) and is independent for all $i\in[1:K]$. Consider the binary maximum computation in which the desired function is given as $S_{\max}=\max\{S_1,\cdots,S_K\}$. The key feature of this binary maximum problem is that once the fusion center learns that there is at least one sensor observing one, no more communication is needed.

We now examine the achievable computation rates of the interactive round-robin approach and the proposed multi-round group broadcast. We assume that for the interactive round-robin approach the number of rounds t can be arbitrarily large. To apply Theorem 1, we need to choose a threshold vector and fix a partition. Noticing that $(\theta_0,\theta_1)=(0,1)$ is a threshold vector of the binary maximum function, we can simply convey the clipped type $\bar{b}_1=\min\{b_1,1\}$. If we choose the 1-partition (see Remark 1), then the computation rate

$$R=\frac{\frac{1}{2}\log^+\left(\frac{1}{K}+P\right)}{H\left(U_1^{(1)}\right)} \quad (5)$$

is achievable, where $U_1^{(1)}=\sum_{i=1}^K S_i$ has binomial distribution and its entropy can be approximated as [9, equation (7)]

$$H\left(U_1^{(1)}\right)=\frac{1}{2}\log\left(2\pi eKp\left(1-p\right)\right)+O\left(\frac{1}{Kp(1-p)}\right). \quad (6)$$

To clearly see the significance of selecting a good partition in Theorem 1, we consider the following three distribution ensembles.

1) i.i.d. Bernoulli$\left(\frac{1}{2}\right)$

In this ensemble, $\mathsf{P}(S_{\max}=0)$ decreases exponentially to zero as K increases. The interactive round-robin approach or simply using the K-partition in Theorem 1 achieve the scaling law of $\Theta(1)$. On the contrary, because the 1-partition results in $H\left(U_1^{(1)}\right) = \Theta(\log K)$, it only achieves the scaling law of $\Theta(\frac{1}{\log K})$.

2) i.i.d. Bernoulli$\left(\frac{1}{K}\right)$

In this ensemble, $\{S_i = 1\}$ becomes a rare event when K becomes large and $\lim_{K\to\infty} \mathsf{P}(S_{\max} = 0) = e^{-1}$. The interactive round-robin approach achieves only $\Theta(\frac{1}{\log K})$ scaling. On the contrary, the 1-partition achieves $\Theta(1)$ scaling since $H(U_1^{(1)})$ in (6) is bounded by some constant in this case.

3) i.i.d. Bernoulli$\left(\frac{1}{\sqrt{K}}\right)$

Figure 1 plots the computation rates of the proposed multi-round group broadcast in Theorem 1 and the interactive round-robin approach. For the interactive round-robin approach, we upper bound the achievable computation rate using the first bound of Theorem 3 in [2]. For the 1-partition, we approximate (5) using (6) for $K > 200$ and evaluate (5) directly otherwise. As shown in Figure 1, both the interactive round-robin and the 1-partition fail to achieve a non-vanishing rate as K increases. By contrast, if we set $\mathcal{A}_m^{(1)} = [(m-1)a_1 + 1 : ma_1]$ for all $m \in [1 : J_1 - 1]$, where $a_1 = \lceil\sqrt{K}\rceil$, $J_1 = \lceil K/a_1\rceil$, and $\mathcal{A}_{J_1}^{(1)} = [(J_1-1)a_1 + 1 : K]$ in Theorem 1, which is termed $\sqrt{K}$-partition, the proposed multi-round group broadcast can achieve a non-vanishing rate, which will be proved next.

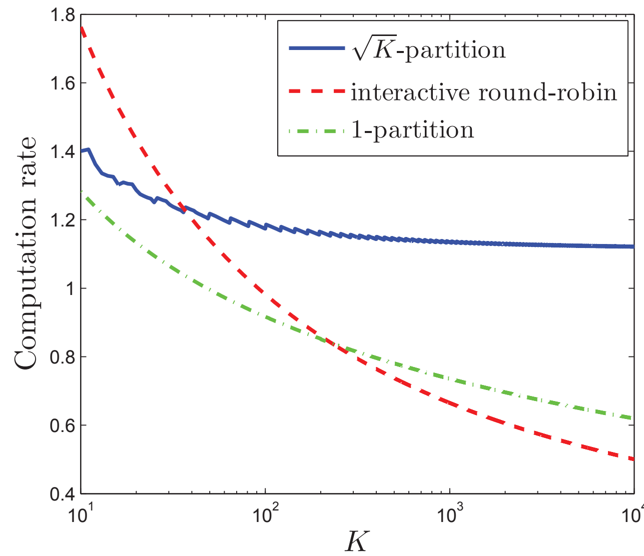

Fig. 1. Computation rates of the binary maximum function for the i.i.d. Bernoulli$\left(\frac{1}{\sqrt{K}}\right)$ ensemble and $P = 20$ dB. Both the 1-partition and the $\sqrt{K}$-partition use the multi-round group broadcast scheme in Theorem 1.

B. $\Theta(1)$ Scaling

Theorem 2: Consider the setup in Theorem 1. If the sources are independent, i.e., $p_{S_1,\cdots,S_K} = \prod_{i=1}^K p_{S_i}$, then for all $l \in [0 : q-1]$, there exists a partition $\{\mathcal{A}_m^{(l)}\}$ such that $H\left(U_1^{(l)}, \cdots, U_{J_l}^{(l)}\right) \le g(\theta_l)$ for some function g, independent of K, and the computation rate $R = \frac{\log^+(P)}{\sum_{l=0}^{q-1} g(\theta_l)}$ is achievable.

Proof: Due to page limitation, we prove here based on i.i.d. sources, i.e., $p_{S_1,\cdots,S_K} = (p_S)^K$. We refer to the full paper in preparation for general independent sources.

Since there are only K sensors, we can assume $\theta_l \le K$ without loss of generality. If $\theta_l = 0$, then $U_m^{(l)} = 0$ for all $m \in [1 : J_l]$ and thus $H\left(U_1^{(l)}, \cdots, U_{J_l}^{(l)}\right) = 0$. In the following, we consider the case $1 \le \theta_l \le K$. Since the sources are independent, for all $l \in [0 : q-1]$, $U_1^{(l)} - \cdots - U_{J_l}^{(l)}$ forms a Markov chain.

Set $\mathcal{A}_m^{(l)} = [(m-1)a_l + 1 : ma_l]$ for all $m \in [1 : J_l]$, where $a_l = \min\{\lceil\theta_l/p_S(l)\rceil, K\}$ and $J_l = \lceil K/a_l\rceil$. For convenience, set $S_{K+1} = \cdots = S_{J_l a_l} = 0$ and denote $W_m^{(l)} = \sum_{i\in\mathcal{A}_m^{(l)}} \mathbf{1}_{\{S_i = l\}}$. Then, the entropy $H\left(U_1^{(l)}, \cdots, U_{J_l}^{(l)}\right)$ can be upper bounded as follows.

$$
\begin{aligned}
&H\left(U_1^{(l)}, \cdots, U_{J_l}^{(l)}\right) \\
&= \sum_{m=1}^{J_l} H(U_m^{(l)} | U_{m-1}^{(l)}, \cdots, U_1^{(l)}) \\
&\overset{(a)}{=} H(U_1^{(l)}) + \sum_{m=2}^{J_l} H(U_m^{(l)} | U_{m-1}^{(l)}) \\
&\overset{(b)}{=} H(U_1^{(l)}) + \sum_{m=2}^{J_l} \sum_{j=0}^{\theta_l - 1} \mathsf{P}\left(U_{m-1}^{(l)} = j\right) H\left(U_m^{(l)} | U_{m-1}^{(l)} = j\right) \\
&= H\left(W_1^{(l)}\right) + \sum_{m=2}^{J_l} \sum_{j=0}^{\theta_l - 1} \mathsf{P}\left(U_{m-1}^{(l)} = j\right) H\left(W_m^{(l)}\right),
\end{aligned} \tag{7}
$$

where (a) follows since $U_1^{(l)} - \cdots - U_{J_l}^{(l)}$ forms a Markov chain and (b) follows since $U_m^{(l)}$ conditioned on $U_{m-1}^{(l)} \ge \theta_l$ is deterministic.

Since for $j \in [0 : \theta_l - 1]$, $\mathsf{P}\left(U_{m-1}^{(l)} = j\right) = \mathsf{P}(X = j)$ where $X \sim$ Binomial$((m-1)a_l, p_S(l))$, we have

$$
\begin{aligned}
&\mathsf{P}\left(U_{m-1}^{(l)} = j\right) \\
&= \binom{(m-1)a_l}{j} p_S(l)^j (1 - p_S(l))^{(m-1)a_l - j} \\
&\le ((m-1)a_l)^j p_S(l)^j \left((1 - p_S(l))^{a_l}\right)^{m-2} \\
&\overset{(a)}{\le} (m-1)^j (\theta_l + 1)^j \left((1 - p_S(l))^{\theta_l / p_S(l)}\right)^{m-2} \\
&\overset{(b)}{\le} (m-1)^j (\theta_l + 1)^j \left(e^{-\theta_l}\right)^{m-2},
\end{aligned} \tag{8}
$$

where (a) follows since $\theta_l \le a_l p_S(l) < \theta_l + 1$ and (b) follows since $(1-x)^{1/x} \le e^{-1}$ for all $x \in (0, 1]$.

Also, $W_m^{(l)} = \sum_{i\in\mathcal{A}_m^{(l)}} \mathbf{1}_{\{S_i = l\}} \sim$ Binomial$(a_l, p_S(l))$. Let $Q^{(l)}$ be a Poisson random variable with mean $a_l p_S(l)$. Then, we have

$$
\begin{aligned}
H(W_m^{(l)}) &\overset{(a)}{\le} H(Q^{(l)}) \\
&\overset{(b)}{\le} \frac{1}{2} \log\left(2\pi e\left(a_l p_S(l) + \frac{1}{12}\right)\right) \\
&\overset{(c)}{\le} \frac{1}{2} \log\left(2\pi e\left((\theta_l + 1) + \frac{1}{12}\right)\right) := g'(\theta_l)
\end{aligned} \tag{9}
$$

where (a) follows from [10, Theorems 7 and 8], (b) follows from [9, equation (1)], and (c) follows since $a_l p_S(l) \leq \theta_l + 1$.

Combining (7) to (9), we have

$$\begin{aligned} & H\left(U_1^{(l)}, \cdots, U_{J_l}^{(l)}\right) \\ \leq & g'(\theta_l) + \sum_{m=2}^{J_l} \sum_{j=0}^{\theta_l - 1} (m-1)^j (\theta_l + 1)^j \left(e^{-\theta_l}\right)^{m-2} g'(\theta_l) \\ = & g'(\theta_l) \left(1 + \sum_{j=0}^{\theta_l - 1} (\theta_l + 1)^j \sum_{m=1}^{J_l - 1} m^j \left(e^{-\theta_l}\right)^{m-1}\right). \end{aligned}$$

It can be easily shown that for any fixed $j \in \mathbb{Z}$ and $\theta_l \geq 1$, the series $\sum_{m=1}^{\infty} m^j \left(e^{-\theta_l}\right)^{m-1}$ converges (using the ratio test, for example). Thus,

$$\begin{aligned} & H\left(U_1^{(l)}, \cdots, U_{J_l}^{(l)}\right) \\ \leq\ & g(\theta_l) := g'(\theta_l) \left(1 + \sum_{j=0}^{\theta_l - 1} (\theta_l + 1)^j \sum_{m=1}^{\infty} m^j \left(e^{-\theta_l}\right)^{m-1}\right). \end{aligned} \tag{10}$$

Finally, the inequality $H(U_0, \cdots, U_{q-1}) \leq \sum_{l=0}^{q-1} H(U_l) \leq \sum_{l=0}^{q-1} g(\theta_l)$ and (4) imply $R \geq \frac{\log^+(P)}{\sum_{l=0}^{q-1} g(\theta_l)}$. ■

VI. Computation of the Maximum Function

In general, the coding scheme in Theorem 1 can be tailored to any particular type-threshold function in order to achieve a higher computation rate. As an example, consider the maximum function assuming i.i.d. sources.

As shown in [1], the all-one vector can be a threshold vector of the maximum function. A naive approach is to let the fusion center learn the whole clipped type vector $\overline{\mathbf{b}}$ and then deduce the maximum function from $\overline{\mathbf{b}}$. If we consider the scaling of the alphabet size q, Theorems 1 and 2 show that this naive approach achieves $\Theta(\frac{1}{q})$ scaling. However, computing the maximum does not require to learn all elements of $\overline{\mathbf{b}}$. The following theorem shows that $\Theta(\frac{1}{\log q})$ scaling is achievable. The achievability follows from an adaptation of Theorem 1 combined with the binary search algorithm.

Theorem 3: Consider the maximum function computation in collocated Gaussian networks. The computation rate

$$R = \frac{\frac{1}{2}\log^+(P)}{C\lceil \log q \rceil}$$

is achievable, where $C > 0$ is a constant.

Proof: The coding scheme is almost the same used in Theorem 1. The only differences are that the number of stages reduces to $\lceil \log q \rceil$ and that $\{\mathcal{A}_m^{(l)}\}$ and $\{V_{im}^{(l)}\}$ are tailored to the maximum function computation using the binary search algorithm.

The whole communication is divided into $\lceil \log q \rceil$ stages. Define c_l and $\mathcal{R}_l$ as the midpoint and the search range, respectively, of the binary search algorithm in stage $l \in [0 : \lceil \log q \rceil - 1]$. Initially, $\mathcal{R}_0 = [0 : q-1]$. The search ranges in the later stages are determined by the midpoints and the boundaries. The midpoint $c_l \in \mathcal{R}_l$ is chosen such as to minimize $|\,\mathsf{P}(S_i \geq c_l | S_i \in \mathcal{R}_l) - \frac{1}{2}|$. Note that the midpoints and the search ranges are functions of $\{s_i\}$.

Given the definitions of midpoints and search ranges, for all $l \in [0 : \lceil \log q \rceil - 1]$, set $\mathcal{A}_m^{(l)} = [(m-1)a_l + 1 : m a_l]$ for all $m \in [1 : J_l]$, where $a_l = \min\{\lceil \frac{1}{\mathsf{P}(S_i \geq c_l | S_i \in \mathcal{R}_l)} \rceil, K\}$ and $J_l = \lceil K/a_l \rceil$. For convenience, set $S_{K+1} = \cdots = S_{J_l a_l} = 0$. Note that the partitions $\{\mathcal{A}_m^{(l)}\}$ are also functions of $\{s_i\}$.

Define the following recursion for all $j \in [1 : k]$:

$$u_m^{(l)}[j] = u_{m-1}^{(l)}[j] + \textstyle\sum_{i \in \mathcal{A}_m^{(l)}} v_{im}^{(l)}[j],$$

where $u_0^{(l)}[j] = 0$ and

$$v_{im}^{(l)}[j] = \mathbf{1}_{\{u_{m-1}^{(l)}[j] = 0\} \bigcap \{s_i[j] \in \mathcal{R}_l[j], s_i[j] \geq c_l[j]\}}.$$

Also, let $w_l[j] = \mathbf{1}_{\{u_{J_l}^{(l)}[j] > 0\}}$ and $\mathbf{w}_l = \begin{bmatrix} w_l[1] & \cdots & w_l[k] \end{bmatrix}$. Since the $(\mathbf{w}_0, \mathbf{w}_1, \cdots, \mathbf{w}_{\lceil \log q \rceil - 1})$ determines all the midpoints, the binary search algorithm can be run in a distributed fashion and the maximum function can be computed after $\lceil \log q \rceil$ stages.

Following the similar lines as in the proof of Theorem 2, it can be shown that $H\left(U_1^{(l)}, \cdots, U_{J_l}^{(l)}\right) \leq g(1)$ where g is given in (10). Therefore, the achievable computation rate can be lower bounded as

$$R \geq \frac{\frac{1}{2}\log^+(P)}{\lceil \log q \rceil g(1)},$$

which establishes the theorem. ■

Acknowledgement

This work has been supported in part by the European ERC Starting Grant 259530-ComCom. The second author was also funded in part by the MSIP (Ministry of Science, ICT & Future Planning), Korea in the ICT R&D Program 2013.

References

[1] A. Giridhar and P. R. Kumar, "Computing and communicating functions over sensor networks," *IEEE J. Select. Areas Commun.*, vol. 23, pp. 755–764, Apr. 2005.

[2] N. Ma, P. Ishwar, and P. Gupta, "Interactive source coding for function computation in collocated networks," *IEEE Trans. Inf. Theory*, vol. 58, pp. 4289–4305, Jul. 2012.

[3] S. Subramanian, P. Gupta, and S. Shakkottai, "Scaling bounds for function computation over large networks," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Nice, France, Jun. 2007.

[4] H. Kowshik and P. R. Kumar, "Zero-error function computation in sensor networks." in *Proc. IEEE Conf. Decision and Control (CDC)*, Shanghai, China, Dec. 2009.

[5] B. Nazer and M. Gastpar, "Computation over multiple-access channels," *IEEE Trans. Inf. Theory*, vol. 53, pp. 3498–3516, Oct. 2007.

[6] S.-W. Jeon, C.-Y. Wang, and M. Gastpar, "Computation over Gaussian networks with orthogonal components," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Istanbul, Turkey, Jul. 2013.

[7] A. Orlitsky and J. R. Roche, "Coding for computing," *IEEE Trans. Inf. Theory*, vol. 47, pp. 903–917, Mar. 2001.

[8] B. Nazer and M. Gastpar, "Compute-and-forward: Harnessing interference through structured codes," *IEEE Trans. Inf. Theory*, vol. 57, pp. 6463–6486, Oct. 2011.

[9] J. Adell, A. Lekuona, and Y. Yu, "Sharp bounds on the entropy of the poisson law and related quantities," *IEEE Trans. Inf. Theory*, vol. 56, pp. 2299 –2306, May 2010.

[10] P. Harremoës, "Binomial and poisson distributions as maximum entropy distributions," *IEEE Trans. Inf. Theory*, vol. 47, pp. 2039 –2041, Jul. 2001.