Cooperative Sensor Networks With Misinformed Nodes

Zhiyu Yang, *Student Member, IEEE,* and Lang Tong, *Fellow, IEEE*

***Abstract*—The communication capacity of Cooperative SEnsor Networks with Mobile Access (C-SENMA) is considered when some sensors may be misinformed with erroneous messages. It is assumed that a global message is first distributed to all the nodes, each node receiving the message correctly with probability $1 - \beta$. The nodes cooperate in delivering the global message to the mobile access point. Three system configurations are discussed based on whether a polling channel and/or an energy constraint are present. The first type is C-SENMA with Polling with No Energy constraints (PNE), where the mobile access point has the ability to poll individual sensors. Without energy constraints, each sensor can transmit for an unlimited number of times. The second type is C-SENMA with No Polling with No Energy constraints (NPNE), where adaptive polling is not allowed and sensors have to transmit according to a predetermined schedule. The third type is C-SENMA with No Polling with an Energy constraint (NPE), where each node has a limit on the number of transmissions.**

The capacities of the three system configurations are analyzed. It is shown that, the capacity for C-SENMA PNE is the same as that when there are no misinformed sensors. For C-SENMA NPNE, with the absence of the polling channel, there is a loss on the achievable rate, proportional to β, the probability that a sensor is misinformed. Results are extended to multiple simultaneous transmissions with the presence of channel fading. The optimal number of simultaneous transmissions is investigated under three different fading situations.

***Index Terms*—Channel capacity, cooperative transmission and networking, multiaccess channels, sensor networks.**

I. Introduction

We consider large-scale sensor networks in which each sensor is limited in power and constrained in energy. Sensor networks are application specific, and we are interested in those cases when the network operates in two different phases: information gathering and information retrieval. The former focuses on sensing and local processing of information by the sensors whereas the latter is concerned with efficient and reliable information delivery to the outside world.

Manuscript received July 21, 2004; revised April 28, 2005. This work was supported in part by the Army Research Laboratory CTA on Communications and Networks under Grant DAAD19-01-2-0011, the Multidisciplinary University Research Initiative (MURI) under the Office of Naval Research Contract N00014-00-1-0564, and the National Science Foundation under Contract CCR-0311055. The material in this paper was presented in part at the 38th Annual Conference on Information Sciences and Systems, Princeton, NJ, March 2004; The Global Communications Conference, Dallas, TX, November/December 2004; and the International Conference on Communications, Seoul, Korea, May 2005.

The authors are with the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY 14853 USA (e-mail: zy26@cornell.edu; ltong@ece.cornell.edu).

Communicated by R. R. Müller, Associate Editor for Communications.
Digital Object Identifier 10.1109/TIT.2005.858922

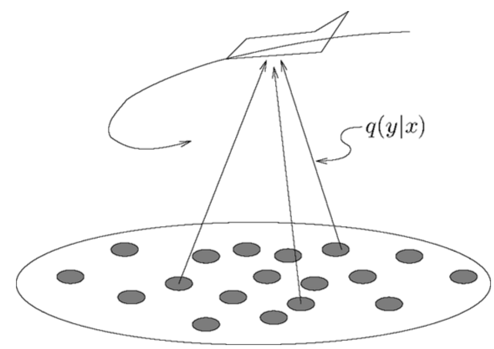

Fig. 1. SEnsor Network with Mobile Access (SENMA).

We focus in this paper on the information retrieval aspect of the sensor network, taking an informationtheoretic approach to the efficient and reliable extraction of data. To this end, we assume a special architecture referred to as SEnsor Networks with Mobile Access (SENMA) [1]. As illustrated in Fig. 1, SENMA consists of two types of nodes: a large number of geographically distributed sensors with low power, and a mobile access point in charge of collecting data from sensors. The presence of a mobile access point, presumably less constrained in energy, power, and computation complexity, considerably simplifies the design. Similar ideas include the so-called Message Ferry [2] and Data Mule [3], [4].

In SENMA, sensors can, of course, transmit data directly to the mobile access point. Such transmissions are not likely to be reliable due to the lack of transmission power and the low complexity of the sensor nodes. A better approach is to deliver the information in a cooperative fashion. By cooperative SENMA (C-SENMA) we mean that when communicating to the mobile access point, sensors may reach an agreement on a message and transmit using an appropriate coding scheme. Such a cooperative scheme, as an alternative to collaborative transmission at the signaling level, makes information retrieval robust against failures of individual sensors. A coding-across-sensors scheme to cope with packet losses is presented in [5].

The process of reaching agreement, referred to as *orientation*, is nontrivial. Orientation can be carried out in many ways. For example, nodes may exchange information via conference links among themselves and establish a global message. Alternatively, the global message may be propagated by a software agent that travels across the sensor network. When sensor networks are viewed as a form of storage devices in which one mobile access point deposits information meant to be retrieved by other mobile access points at a different time, the process of orientation is simply the broadcast of messages from a mobile access point.

We consider a discrete-time system. When only one sensor transmits at a time, the channel between each sensor and the mobile access point can be modeled as an identical discrete memoryless channel (DMC) with conditional probability $q(y\,|\,x)$. If the orientation is perfect, i.e., there is no disagreement among sensors, then the maximum achievable rate of the information retrieval is given by

$$C^{(0)} = \max_{p(x)} I(X;Y).$$

In such a setting, there is no difference between retrieving information from a single sensor or multiple sensors since all sensors have the same message.

For a large-scale sensor network, however, perfect orientation may not be possible. For example, the software agent responsible for distributing the message may not have reached all sensors, or sensors make errors because of unreliable conference links. In practice, there is always a possibility that some sensors do not have the correct message for transmission. We refer to such sensors as *misinformed* sensors, and this paper focuses on the maximum achievable rate of information retrieval with the presence of misinformed sensors.

A. Summary of Results

We study three types of C-SENMA configurations in this paper. The first type is C-SENMA with Polling with No Energy constraints (PNE). In this case, the mobile access point has the ability to poll individual sensors, making it possible to locate sensors with the correct message. Without energy constraints, each sensor can transmit for an unlimited number of times, and information can be retrieved reliably from those well-informed sensors. The second type is C-SENMA with No Polling with No Energy constraints (NPNE). In this case, polling is not allowed, and sensors must transmit according to a predetermined schedule. The problem here is more challenging because the mobile access point can no longer make choices based on the signal it receives from the sensors. The third type is C-SENMA with No Polling with an Energy constraint (NPE), a case that models the more practical situation that battery operated sensors cannot transmit indefinitely. In such a case, it is impossible to schedule a single sensor to transmit for all time. The last possible type, not considered in this paper, is C-SENMA with Polling with an Energy constraint.

The capacities of there three C-SENMA systems are the main focus of this paper. We assume that each sensor has a probability β to be misinformed.[1] The challenge for a system with a polling channel is to design strategies to adapt the polling sequence according to the previous receptions. For a system without polling channels, the activation sequence is predetermined and is an optimization parameter for the maximum achievable rate.

If only one sensor transmits at a time, and the channel between each sensor and the mobile access point is an identical DMC $q(y\,|\,x)$, we show that the capacity of C-SENMA PNE is

$$C^{\mathrm{PNE}} = C^{(0)}$$

provided that the sensor is not always misinformed, i.e., $\beta < 1$. When polling is not allowed, the capacity of C-SENMA NPNE is shown to be

$$C^{\mathrm{NPNE}} = (1-\beta)C^{(0)}.$$

We also present the capacity of C-SENMA NPE, which has a more complicated expression than the previous two system configurations.

Allowing one sensor to transmit at a time simplifies the design and the operation of the system, but it may not be optimal, depending on the channel between sensors and the mobile access point. We thus investigate multiple simultaneous transmissions and consider channels with fading. We focus on C-SENMA NPNE and consider a Gaussian channel with a network-wise total power constraint.[2] We prove an achievable rate and show that if there is no fading, the achievable rate increases to infinity as the number of simultaneous transmissions increases. If there is only phase fading, however, the achievable rate is maximized when there is only one sensor transmitting at any time slot. For Rayleigh fading, the optimal number of simultaneous transmissions varies with the misinformed probability β.

B. Related Work

For wireless networks with many power/energy-constrained nodes, the idea of cooperation among nodes for the purpose of more reliable and efficient communication has attracted much attention. Cooperation among nodes can be found in many forms according to the traffic pattern, including, *broadcasting* where information is delivered from a source to all the other nodes in the network with nodes relaying the information [7]–[9], *relaying* where information is delivered from a source to a single destination with the help of relay nodes [10], [11], and *multiple sources type* where each node in the network has its own information to send, and nodes assist each other [12]–[16]. The traffic pattern in this paper is different from the aforementioned three patterns. We assume that nodes in the network have established a common message and they cooperate to deliver the message to the mobile access point. This traffic pattern can also be found in [5], [17]. The problem of establishing the common message was addressed in the sensor broadcast problem [18]. When all the nodes are assumed to know the exact common message, there is no difference between multiple cooperating nodes and multiple transmitting antennas (in multiple-input single-output or multiple-input multiple-output channels). In this paper, however, we take into account errors in distributing the common message in the cooperation process under a specific sensor network architecture.

The communication model used in this paper is conceptually related to relay networks, where the source node distributes its information via a broadcast medium, and the relay nodes in the network receive the transmission and forward the information to the destination node through some coding strategies [10], [11], [19]. Our model differs from [10], [11], [19] in that we model the message distribution process via a packet error rate

[1]When a sensor is misinformed, it chooses its local message randomly from the possible message set with equal probability.

[2]Instead of the maximum achievable rate of C-SENMA PNE, results presented in [6] are in fact achievable rates for C-SENMA NPNE with no fading.

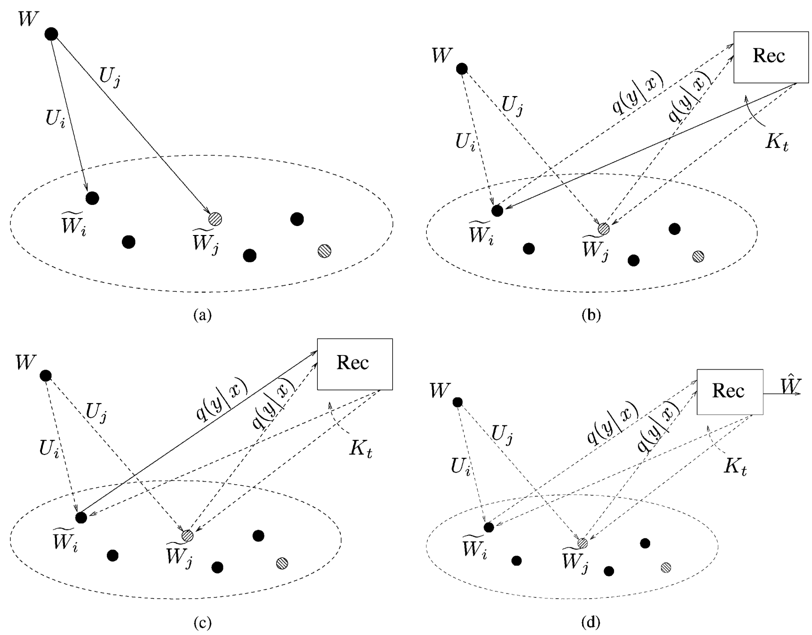

Fig. 2. Communication steps. (a) Step I: orientation. (b) Step II: polling. (c) Step III: transmission and reception. (d) Step IV: decoding.

model, in which an intermediate node received the global message correctly with a fixed probability, while [10], [11], [19] via a Gaussian channel model.

The paper is organized as follows. Section II describes the communication procedure and introduces the channel model. The capacities of the three system configurations are studied in Sections III–V, respectively. We extend the model to multiple simultaneous transmissions in Section VI and investigate the optimal number of simultaneous transmissions for a Gaussian multiple-access channel (MAC) with a network power constraint in Section VII. The paper is concluded in Section VIII.

II. Model

We first describe the model for systems with a polling channel. The communication of the global message from the network to the mobile access point is divided into four steps as shown in Fig. 2: Fig. 2 (a) presents orientation, Fig. 2 (b) polling, Fig. 2 (c) transmission and reception, and Fig. 2 (d) decoding. In the first step, nodes are informed with a global message $W \in \{1,\dots,M\}$ that is uniformly distributed. We assume that each node receives the global message independently and with probability β being correct. Specifically, the reception of node i is controlled by a binary random variable U_i, independent of W and independent and identically distributed (i.i.d.) across node index i with distribution

$$p(u_i) = \begin{cases} \beta, & \text{if } u_i = 0 \\ 1-\beta, & \text{if } u_i = 1 \end{cases} \tag{1}$$

where $\beta \in [0,1]$ is a constant. When $U_i = 1$, the received message at node i, denoted by $\widetilde{W}_i$, is equal to the global message W. When $U_i = 0$, $\widetilde{W}_i$ is uniformly distributed from 1 to M. Thus,

$$p(\tilde{w}_i \,|\, w, u_i) = \begin{cases} \delta(\tilde{w}_i, w), & \text{if } u_i = 1 \\ \frac{1}{M}, & \text{if } u_i = 0 \end{cases}$$

where $\delta(a,b)$ is equal to 1 if $a = b$, or 0 otherwise. The constant β affects the reception of the global message by individual nodes and is referred to as the orientation error probability of the network.

We consider a time-slotted system. The mobile access point comes to retrieve information from the field after the information orientation is complete. The information retrieval process consists of Step II: polling and Step III: transmission and reception. To avoid collision, a polling-based multiple access is employed: The mobile access point polls one node to transmit one symbol at each time slot. At time t, the receiver polls node K_t to transmit the tth symbol of the codeword corresponding to the $\widetilde{W}_{K_t}$th message. Since the mobile access point is not power limited, we assume that the polling channel is error free. The uplink channels from each node to the receiver are assumed to be identical and modeled by a DMC $\{\mathcal{X}, \mathcal{Y}, q(y\,|\,x)\}$, where $\mathcal{X}$ and $\mathcal{Y}$ are the input and output alphabets, respectively, and $q(y\,|\,x)$ is the transition probability of the channel. Node K_t, after receiving the polling signal, transmits the selected symbol to the uplink channel. Denote X_t and Y_t the input and output of the DMC at time t, respectively. After receiving Y_t, the mobile access point moves to the next time slot $t+1$ and starts the polling step again. It may poll a node that has or has not been polled before. Steps II and III alternate until t reaches n, the number of slots the receiver spends to retrieve information from the field.

At the last step, the receiver decodes the global message based on the channel outputs Y^n and the polling history K^n. The decoded message is denoted by $\hat{W} \in \{1,\dots,M\}$. A decoding error occurs if $\hat{W} \neq W$. We assume that the sensor network is large in the sense that there are infinite number of nodes. With this assumption, the probability of all nodes being misinformed is zero. Thus, it is possible to have a positive achievable rate.

In the above scheme, the mobile access point has a polling channel to address individual sensors. The polling channel could

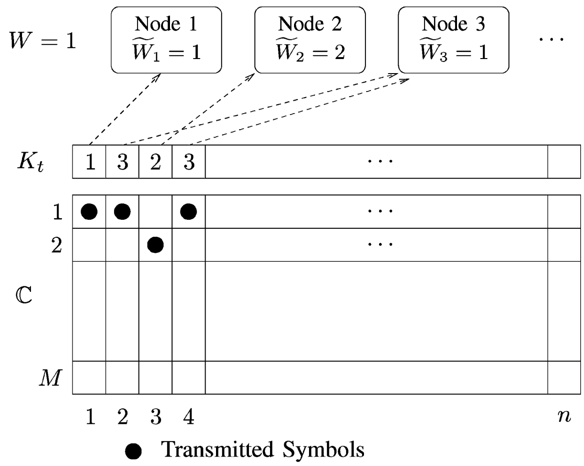

Fig. 3. Transmission example. The global message is $W = 1$. Node 2 is misinformed where the local message is $\widetilde{W}_2 = 2$. At time $t = 3$, the activated node is $K_t = 2$. Hence, the symbol in row 2 column 3 of the codebook $\mathbb{C}$ is transmitted at time $t = 3$, resulting in a transmission from a misinformed node.

be costly to implement. Therefore, it is also desirable to consider schemes with predetermined scheduling, i.e., the scheduling sequence is preset in sensors before deployment. We refer to $\{K_t\}$ as the *polling* sequence if a polling channel is implemented, or the *prescheduling* sequence if no polling channels are implemented. In systems without polling, the prescheduling sequence $\{K_t\}$ does not depend on the channel outputs.

For systems without energy constraints (PNE and NPNE), we assume that there is no limit on how many times a sensor can transmit. In reality, battery-powered sensors are energy limited, which may impose constraints on the lifetime of sensors. For systems with an energy constraint (NPE), we assume that each sensor has up to Q transmissions.

Fig. 3 illustrates a transmission example of C-SENMA, where $\mathbb{C}$ represents the shared codebook, and the actual transmitted symbols are labeled by solid dots in the codebook. In Fig. 3, the global message is $W = 1$. Node 2 is misinformed where the local message is $\widetilde{W}_2 = 2$. At time $t = 3$, the activated node is $K_t = 2$. Hence, the symbol in row 2 column 3 of the codebook $\mathbb{C}$ is transmitted at time $t = 3$, resulting in a transmission from a misinformed node.

The *rate* of a codebook is defined as $R \triangleq \log(M)/n$, where M is the number of messages in the codebook and n is the length of a codeword. The probability of error is defined as $P_e \triangleq \mathcal{P}(\hat{W} \neq W)$, where $W \in \{1, \ldots, M\}$ is uniformly distributed and $\hat{W}$ is the decoded message. A rate R is called *achievable* if for any given error $\epsilon > 0$, there exists a codebook with rate larger than $R - \epsilon$ and probability of error less than ϵ. The *capacity* of a system configuration is defined as the maximum of all achievable rates for the system configuration.

In the next section, we study the capacity of C-SENMA PNE in which a polling channel is implemented and each sensor does not have a limit on how many times it can transmit.

III. Capacity of C-SENMA PNE

When a polling channel is implemented, it is possible for the mobile access point to first locate a sensor that, with a high probability, has the correct global message, and then retrieve the global message from that sensor. If the number of time slots needed in locating such a sensor is only a function of the given probability of error, but not a function of the codeword length, then the overhead associated with the first phase can be made arbitrarily small by increasing the codeword length of the second phase. Therefore, the capacity of C-SENMA PNE is just the capacity of the DMC $(\mathcal{X}, \mathcal{Y}, q(y\,|\,x))$

$$C^{(0)} = \max_{p(x)} I(X;Y).$$

This strategy is elaborated in the proof of the following theorem, which, along with a sketch of the proof, was suggested by an anonymous reviewer for the 2004 IEEE International Symposium on Information Theory.

Theorem 1: The capacity of C-SENMA PNE is

$$C^{\mathrm{PNE}} = \begin{cases} C^{(0)}, & \text{if } \beta < 1 \\ 0, & \text{if } \beta = 1. \end{cases}$$

Theorem 1 indicates that, for C-SENMA PNE, we achieve the capacity of the DMC as long as $\beta < 1$. There is no rate loss in the C-SENMA PNE case.

Proof: From the DMC capacity definition, for all $\epsilon > 0, R_1 < C^{(0)}$, there exists an n_0 such that for all $n_1 > n_0$, there exists a code $(2^{n_1 R_1}, n_1, \epsilon)$, where (M, n, ϵ) denotes a code over the DMC $q(y\,|\,x)$ with M messages, codeword length n, and probability of error less than ϵ. Let $M = 2^{n_1 R_1}$ be the total number of messages. Let $j > 1/\epsilon$. Select l large enough such that there exists a code (j, l, ϵ) over the DMC. Divide the M messages evenly into j groups,[3] indexed from 1 to j. Let

$$k > \frac{\log(\epsilon)}{\log(1 - (1-\epsilon)^2(1-\beta)^2)}. \tag{2}$$

The mobile access point consecutively polls $2k$ sensors, each transmitting the group index of its message using the (j, l, ϵ) code. The total number of channel uses for this phase is $2kl$, which is only a function of ϵ. Let g_i denote the group index of the local message at the ith polled node and $\hat{g}_i$ denote the ith decoded group index. The mobile access point checks if there exists an $i, 1 \leq i \leq k$, such that $\hat{g}_{2i-1} = \hat{g}_{2i}$. If there is no such i, the mobile access point declares an error; otherwise, the mobile access point selects one node from one of the matching pairs. Denote by b the index of the selected node. The access point polls node b to transmit its message using the $(2^{n_1 R_1}, n_1, \epsilon)$ code. The decoded message is declared as the global message. The total channel uses are $n = 2kl + n_1$ and the overall communication rate is given by

$$R = \frac{n_1}{2kl + n_1} R_1$$

which converges to R_1 as n_1 increases.

[3]Some groups may have one more message than others. For j fixed, as n_1 increases, the ratio of the number of messages in each group over the total number of messages converges to $1/j$.

To analyze the probability of error, define the following error events.

- A_1: For all $i, 1 \le i \le k, \hat{g}_{2i-1} \ne \hat{g}_{2i}$.
- A_2: There exists an i such that $\hat{g}_{2i-1} = \hat{g}_{2i}$ and the selected node b does not have the correct global message, i.e., $\widetilde{W}_b \ne W$.
- A_3: There exists an i such that $\hat{g}_{2i-1} = \hat{g}_{2i}$ and $\widetilde{W}_b = W$, but the mobile access point does not decode the message correctly, i.e., $\hat{W} \ne W$.

Let B_1 be the event that the first pair of decoded group indices match, $B_1 = \{\hat{g}_1 = \hat{g}_2\}$, B_2 the event that the first pair of nodes have the correct global message, $B_2 = \{\widetilde{W}_1 = W, \widetilde{W}_2 = W\}$, and B_3 the event that the first pair of transmitted group indices are decoded correctly, $B_3 = \{\hat{g}_1 = g_1, \hat{g}_2 = g_2\}$. We have

$$\begin{aligned}\mathcal{P}(B_1) &\ge \mathcal{P}(B_2 \cap B_3)\\ &= \mathcal{P}(B_2)\mathcal{P}(B_3 \mid B_2)\\ &\ge (1-\beta)^2(1-\epsilon)^2.\end{aligned}$$

Since the matching outcome of any pair of decoded group indices is i.i.d., the first kind of error is upper-bounded by

$$\begin{aligned}\mathcal{P}(A_1) &= (1-\mathcal{P}(B_1))^k\\ &\le (1-(1-\beta)^2(1-\epsilon)^2)^k\\ &\le \epsilon\end{aligned}$$

where the last inequality is because of (2).

Let B_4 be the event that the first node does not have the correct global message, $B_4 = \{\widetilde{W}_1 \ne W\}$. The second kind of error is upper-bounded by

$$\begin{aligned}\mathcal{P}(A_2) &= \mathcal{P}(A_2 \cap A_1^c) &(3)\\ &= \mathcal{P}(A_2 \mid A_1^c)\,\mathcal{P}(A_1^c)\\ &\le \mathcal{P}(A_2 \mid A_1^c)\\ &= \mathcal{P}(B_4 \mid B_1) &(4)\\ &\le \mathcal{P}(B_2^c \mid B_1)\\ &= 1 - \frac{\mathcal{P}(B_1 \mid B_2)\mathcal{P}(B_2)}{\mathcal{P}(B_1)} &(5)\end{aligned}$$

where (3) holds because A_1 and A_2 are disjoint events, (4) because for any pair whose decoded group indices match, the probability that either node in the pair receives an incorrect message is the same. Let B_5 be the event that the group indices of the first pair match (before transmission), $B_5 = \{g_1 = g_2\}$. We have

$$\begin{aligned}\mathcal{P}(B_1) &= \mathcal{P}(B_1 \cap B_3^c) + \mathcal{P}(B_1 \cap B_3)\\ &\le \mathcal{P}(B_3^c) + \mathcal{P}(B_1 \cap B_3)\\ &\le 2\epsilon + \mathcal{P}(B_5)\\ &= 2\epsilon + (1-\beta)^2 + \frac{1}{j}(1-(1-\beta)^2)\\ &\le 3\epsilon + (1-\beta)^2.\end{aligned}$$

Since $\mathcal{P}(B_1 \mid B_2) \ge \mathcal{P}(B_3 \mid B_2) \ge (1-\epsilon)^2$, (5) is further bounded as

$$\begin{aligned}\mathcal{P}(A_2) &\le 1 - \frac{(1-\epsilon)^2(1-\beta)^2}{3\epsilon + (1-\beta)^2}\\ &= \frac{3\epsilon + (1-(1-\epsilon)^2)(1-\beta)^2}{3\epsilon + (1-\beta)^2}\\ &\le \frac{5\epsilon}{(1-\beta)^2}.\end{aligned}$$

Since $\mathcal{P}(A_3) \le \epsilon$, the probability of error is upper bounded by

$$\begin{aligned}\mathcal{P}_e &\le \mathcal{P}(A_1 \cup A_2 \cup A_3)\\ &= \mathcal{P}(A_1) + \mathcal{P}(A_2) + \mathcal{P}(A_3)\\ &\le 2\epsilon + \frac{5\epsilon}{(1-\beta)^2}.\end{aligned}$$

For $\beta < 1$, the probability of error $\mathcal{P}_e$ converges to zero as ϵ goes to zero and the communication rate R converges to R_1 as n_1 goes to infinity, where R_1 can be arbitrarily close to $C^{(0)}$. Therefore, $C^{(0)}$ is achievable for $\beta < 1$.

Since the achievable rate of the system cannot exceed $C^{(0)}$, we have proven the theorem. □

IV. Capacity of C-SENMA NPE

With a polling channel, C-SENMA PNE achieves the DMC capacity. For a system without polling channels, the achievable rate is expected to be lower than that for a system with a polling channel. Next, we study the capacity of C-SENMA NPE in which the prescheduling sequence does not depend on channel outputs and each sensor has up to Q transmissions.

Theorem 2: Let

$$\bar{C}_k = \frac{1}{k}\max_{p(s^k)} I(S^k; Y^k) \tag{6}$$

where $S^k \in \mathcal{X}^k, Y^k \in \mathcal{Y}^k$, and

$$p(s^k, y^k) = p(s^k)\cdot\left((1-\beta)\prod_{i=1}^{k} q(y_i \mid s_i) + \beta \sum_{s'^k \in \mathcal{X}^k} p(s'^k)\prod_{i=1}^{k} q(y_i \mid s'_i)\right). \tag{7}$$

The capacity of C-SENMA NPE is

$$C^{\mathrm{NPE}} = \max_{1 \le k \le Q} \bar{C}_k$$

where Q is the maximum number of transmissions allowed from one sensor.

For the optimal achievable rate, the prescheduling sequence is also a design issue besides the codebook. In Theorem 2, $\bar{C}_k$ is the maximum achievable rate when we fix the prescheduling to the predetermined sequence in which every k time slots are allocated to a different node. Theorem 2 indicates that we only need to consider at most Q prescheduling sequences.

From (6), $\bar{C}_k$ can be seen as the capacity of a special channel whose input output joint distribution is given by (7). The channel as defined by (7) is different from a regular DMC in that the input distribution $p(s^k)$ affects the transition probability $p(y^k \mid s^k)$. Intuitively, $p(y^k \mid s^k)$ can be viewed as the combination of the orientation process and the DMC: With probability $1-\beta+\beta p(s^k)$, the original input s^k is transmitted and goes through the DMC; with probability $\beta p(s'^k)$, s'^k is transmitted and goes through the DMC, for $s'^k \in \mathcal{X}^k$ and $s'^k \ne s^k$. The reason that s'^k ($s'^k \in \mathcal{X}^k$ and $s'^k \ne s^k$) is transmitted with probability $\beta p(s'^k)$ is that, when a node is misinformed, it randomly chooses a codeword in the codebook. For a given codebook that is randomly generated with input distribution $p(s^k)$, the probability that s'^k is selected is roughly $p(s'^k)$.

To show the achievability of $\bar{C}_k$, however, it is not sufficient to use the above equivalent channel argument since we need to show that there is a sequence of codebooks such that the probability of error goes to zero. For any given codebook, when a node is misinformed, the probability that a particular sequence of symbols is transmitted from this node depends on the statistics of the given codebook, not the input probability that generates the codebook. Additional care should be taken when using the random coding argument.

A. Direct Part of Theorem 2

We first prove the achievability of $\bar{C}_1$, from which we derive the achievability of $\bar{C}_k$. To achieve $\bar{C}_1$, a new sensor[4] is prescheduled to transmit in every time slot.

1) Codebook Generation: Generate a codebook with $M = 2^{nR}$ messages and codeword length n at random according to the distribution $p(s)$. Let all sensors have the same codebook.

2) Decoder: The typical-set decoding is employed. Define the typical set $A_\epsilon^{(n)}$ with respect to the distribution $p(s,y)$ defined in (7) for $k=1$

$$A_\epsilon^{(n)} \triangleq \Bigg\{ (s^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n : \left| -\frac{1}{n}\log_2 p(s^n) - H(S) \right| \le \epsilon, \left| -\frac{1}{n}\log_2 p(y^n) - H(Y) \right| \le \epsilon, \left| -\frac{1}{n}\log_2 p(s^n, y^n) - H(S,Y) \right| \le \epsilon \Bigg\}$$

where $p(s^n, y^n) = \prod_{i=1}^n p(s_i, y_i)$.[5] Upon receiving channel outputs y^n, the receiver declares the message $\hat{w}$ as the received message if there is one and only one $\hat{w} \in \mathcal{W}$ such that $(s^n(\hat{w}), y^n) \in A_\epsilon^{(n)}$; otherwise, the receiver declares an error.

3) Error Analysis: Let $S_t(j,n)$ denote the tth symbol of the jth message in the codebook with 2^{nR} messages. Let $Y_t(j,n)$ be the tth channel output given that the codebook has 2^{nR} messages and the jth message is the intended global message. In the definitions of $S_t(j,n)$ and $Y_t(j,n)$, we explicitly include n to indicate that $Y_t(i,n)$ and $Y_t(i,n')$ may have different distributions for $n \neq n'$.

Without loss of generality, we assume that the first message is the global message. Let $p^{(n)}(s^n, y^n \,|\, j), 1 \le j \le M$, denote the distribution of $(S^n(j,n), Y^n(1,n))$, where the superscript (n) on $p^{(n)}(\cdot)$ indicates that the codebook has 2^{nR} messages. Notice that a node receives the global message correctly with probability $1-\beta+\beta/M$ and any one of the other messages with probability β/M. It can be shown that

$$p^{(n)}(s^n, y^n \,|\, 1) = \prod_{t=1}^n p^{(n)}(s_t, y_t \,|\, 1)$$

and, for $2 \le j \le M$

$$p^{(n)}(s^n, y^n \,|\, j) = \prod_{t=1}^n p^{(n)}(s_t, y_t \,|\, j)$$

[4]By new sensors we mean sensors that have not transmitted before.

[5]The distribution $p(s^n, y^n)$ is derived from $p(s,y)$, which is given by (7) for $k=1$. It should not be confused with $p(s^k, y^k)$ in (7).

where

$$p^{(n)}(s, y \,|\, 1) = p(s)\left(\left(1-\beta+\frac{\beta}{M}\right) q(y\,|\,s) + \beta\frac{M-1}{M}\sum_{s'\in\mathcal{X}} q(y\,|\,s')p(s')\right)$$

$$p^{(n)}(s, y \,|\, j) = p(s)\left(\frac{\beta}{M} q(y\,|\,s) + \left(1-\frac{\beta}{M}\right)\sum_{s'\in\mathcal{X}} q(y\,|\,s')p(s')\right), \text{ for } j \ge 2.$$

Define the following events for $1 \le j \le M$:

$$E_j \triangleq \left\{ (S^n(j,n), Y^n(1,n)) \in A_\epsilon^{(n)} \right\}.$$

To evaluate the probability of the occurrence of each event, we need the following lemma.

Lemma 1 (Joint AEP, Extension of [20, Theorem 8.6.1]): Consider random variables $S_i^{(n)}$ and $Y_i^{(n)}$ for $n \ge 1$ and $1 \le i \le n$, and let $(S^{(n)})^n$ denote the sequence $(S_1^{(n)}, \ldots, S_n^{(n)})$. Suppose that $((S^{(n)})^n, (Y^{(n)})^n)$ has distribution

$$p^{(n)}(s^n, y^n) = \prod_{i=1}^n p^{(n)}(s_i, y_i)$$

where $p^{(n)}(s,y)$ converges to $p(s,y)$ as n goes to infinity. Let $A_\epsilon^{(n)}$ be the jointly typical set with respect to the distribution $p(s,y)$. Then

$$\lim_{n\to\infty} \mathcal{P}\left[\left(\left(S^{(n)}\right)^n, \left(Y^{(n)}\right)^n\right) \in A_\epsilon^{(n)}\right] = 1. \tag{8}$$

If $((\tilde{S}^{(n)})^n, (\tilde{Y}^{(n)})^n)$ is drawn according to

$$\tilde{p}^{(n)}(s^n, y^n) = \prod_{i=1}^n \tilde{p}^{(n)}(s_i, y_i)$$

and $\tilde{p}^{(n)}(s,y)$ converges to $p(s)p(y)$ as n goes to infinity, then for n large

$$\mathcal{P}\left[\left(\left(\tilde{S}^{(n)}\right)^n, \left(\tilde{Y}^{(n)}\right)^n\right) \in A_\epsilon^{(n)}\right] \le 2^{-n(I(S;Y)-4\epsilon)}.$$

Proof: See Appendix I. □

If we let $p^{(n)}(s,y) = p(s,y)$ and $\tilde{p}^{(n)}(s,y) = p(s)p(y)$ for all n, Lemma 1 reduces to [20, Theorem 8.6.1].

Since $p^{(n)}(s,y\,|\,1)$ converges to $p(s,y)$ and $p^{(n)}(s,y\,|\,j)$, for $j \ge 2$, converges to $p(s)p(y)$ as n goes to infinity, by the joint asymptotic equipartition property (AEP), we have, for n sufficiently large

$$\mathcal{P}(E_1^c) \le \epsilon$$
$$\mathcal{P}(E_j) \le 2^{-n(I(S;Y)-4\epsilon)}, \qquad \text{for } j \ge 2.$$

The average probability of error, averaged over all codewords and all codebooks, is given by

$$\begin{aligned}\mathcal{P}_e^{(n)} &= \mathcal{P}(\hat{W} \neq W \,|\, W=1) \\ &\le \mathcal{P}(E_1^c) + \sum_{j=2}^{2^{nR}} \mathcal{P}(E_j) \\ &\le \epsilon + 2^{nR}2^{-n(I(S;Y)-4\epsilon)} \\ &= \epsilon + 2^{-n(I(S;Y)-R-4\epsilon)}.\end{aligned}$$

If $R < I(S;Y)$, $\mathcal{P}_e^{(n)}$ can be made arbitrarily small by letting n go to infinity and ϵ go to zero. Recall that a rate R is achievable if there exists a codebook with a rate arbitrarily close to R and error probability arbitrarily small. Hence, $I(S;Y)$ is achievable. Optimized over $p(s)$, $\bar{C}_1 = \max_{p(s)} I(S;Y)$ is achievable.

To prove the achievability of $\bar{C}_k$, consider the kth extended C-SENMA where the DMC is the kth extension of the original DMC

$$q(y^k \mid x^k) = \prod_{i=1}^{k} q(y_i \mid x_i).$$

Schedule a new sensor to transmit in every time slot and let $p(s^k)$ be the input distribution in the kth extended system. From the proof of the achievability of $\bar{C}_1$, $I(S^k;Y^k)$ is achievable by the kth extended system. Now consider scheduling a new sensor to transmit in every k time slots in the original system. The operation of the original C-SENMA with a new sensor in every k slots is equivalent to that of the kth extended C-SENMA with a new sensor in every slot, except that, in the original system, it takes k times longer to transmit one codeword. Therefore, $\bar{C}_k = \frac{1}{k}\max_{p(s^k)} I(S^k;Y^k)$ is achievable by the original system.

B. Converse of Theorem 2

For a given n, suppose there is a total number of a nodes involved in the transmission of the codeword. Let $\mathcal{I}_i$ be the set of time slots allocated to node $i, 1 \le i \le a$. We have $\sum_{i=1}^{a} |\mathcal{I}_i| = n$. By the energy constraint, $|\mathcal{I}_i| \le Q$ for all $1 \le i \le a$. Let S^n denote a codeword which is uniformly distributed among all codewords in the codebook, and let Y^n denote the channel outputs. By Fano's inequality, for all achievable rate R

$$\begin{aligned} nR &= H(S^n) \\ &= H(S^n \mid Y^n) + I(S^n;Y^n) \\ &\le 1 + \mathcal{P}_e^{(n)} nR + I(S^n;Y^n). \end{aligned}$$

Let $A \backslash B$ denote the set consisting of all the elements that belong to set A but not set B. Since $\mathcal{P}_e^{(n)}$ goes to zero as n goes to infinity, for n large, the achievable rate R is upper-bounded by the C^{NPE} as follows:

$$\begin{aligned} R &\le \frac{1}{n} I(S^n;Y^n) \\ &= \frac{1}{n}(H(Y^n) - H(Y^n \mid S^n)) \\ &= \frac{1}{n}\sum_{i=1}^{a}\left(H\left(Y_{\mathcal{I}_i} \mid Y_{\cup_{j<i}\mathcal{I}_j}\right) - H\left(Y_{\mathcal{I}_i} \mid S^n, Y_{\cup_{j<i}\mathcal{I}_j}\right)\right) \\ &= \frac{1}{n}\sum_{i=1}^{a}\left(H\left(Y_{\mathcal{I}_i} \mid Y_{\cup_{j<i}\mathcal{I}_j}\right) - H\left(Y_{\mathcal{I}_i} \mid S_{\mathcal{I}_i}\right)\right) \quad (9) \\ &\le \frac{1}{n}\sum_{i=1}^{a}\left(H\left(Y_{\mathcal{I}_i}\right) - H\left(Y_{\mathcal{I}_i} \mid S_{\mathcal{I}_i}\right)\right) \\ &= \frac{1}{n}\sum_{i=1}^{a} I\left(S_{\mathcal{I}_i};Y_{\mathcal{I}_i}\right) \quad (10) \\ &\le \frac{1}{n}\sum_{i=1}^{a} |\mathcal{I}_i| C^{\mathrm{NPE}} \quad (11) \\ &= C^{\mathrm{NPE}} \end{aligned}$$

where (9) holds because $Y_{\mathcal{I}_i}$ is independent of $S^n \setminus S_{\mathcal{I}_i}$ and $Y_{\cup_{j<i}\mathcal{I}_j}$ conditioning on $S_{\mathcal{I}_i}$, and (11) holds because

$$C^{\mathrm{NPE}} = \max_{1 \le k \le Q} \bar{C}_k \ge \frac{1}{|\mathcal{I}_i|} I(S_{\mathcal{I}_i};Y_{\mathcal{I}_i}).$$

C. Properties of $\bar{C}_k$

We discuss some properties of $\bar{C}_k$ in this subsection. For the special case when $\beta = 0$, the communication channel model of C-SENMA NPE reduces to a DMC model. When $\beta = 0$, $\bar{C}_k$ is the capacity of the kth extension of the DMC. Hence, from the properties of a DMC's capacity [20], we have that, when $\beta = 0$

- $\bar{C}_1 = \bar{C}_k$ for $k \ge 1$;
- the optimal input distribution $p^*(s^k)$ for $I(S^k;Y^k)$ is the kth extension of some $p^*(s)$;
- the mutual information $I(S^k;Y^k)$ is a concave function of the input distribution $p(s^k)$.

The first and the second properties indicate that, to calculate the capacity of the DMC, there is no need to consider extensions of the DMC. Optimizing $I(S;Y)$ over $p(s)$ is enough. The third property says that any kind of hill-climbing technique can be used to find the optimal input distribution since the input distribution region is convex. From the third property it is also easy to show that the optimal input distribution for a symmetrical channel is symmetrical. However, for $\beta > 0$, these nice properties do not hold. Counterexamples and other properties are provided as follows.

For $k = ij$, where $i, j, k > 0$, we have $\bar{C}_k \ge \bar{C}_i$. To see this, let the input distribution $p(s^k)$ be the jth product extension of $p(s^i)$. Because of the independence of the input distribution, we have

$$\begin{aligned} I(S^k;Y^k) &\ge \sum_{l=0}^{j-1} I\left(S_{il+1}^{i(l+1)};Y_{il+1}^{i(l+1)}\right) \\ &= jI(S^i;Y^i). \end{aligned}$$

Therefore, $\bar{C}_k \ge \bar{C}_i$ for $k = ij$. Thus, when calculating C^{NPE}, if $k = ij \le Q$ and $i < k$, there is no need to calculate $\bar{C}_i$. Our conjecture is that $\bar{C}_k \ge \bar{C}_i$ for $k > i$.

In general, the input distribution $p(s^k)$ that maximize $I(S^k;Y^k)$ is not the kth extension of some $p(s)$. Fig. 4 shows an example that the optimal $p(s^2)$ is not a product extension of some $p(s)$.

In the definition of $\bar{C}_k$, the optimization is over $p(s^k)$. It is desirable that the mutual information $I(S^k;Y^k)$ is concave with respect to the input distribution $p(s^k)$, which is true for DMCs (when $\beta = 0$). However, $I(S^k;Y^k)$ is not a concave function of $p(s^k)$ in general. Fig. 5 shows an example.

For $Q = 1$, we have the following result concerning the concavity of the mutual information.

Proposition 1: If for all $y_0 \in \mathcal{Y}$, there exists an $x_0 \in \mathcal{X}$ such that for all $x \in \mathcal{X}$ and $x \ne x_0$, $q(y_0 \mid x) = \epsilon(y_0)$, then $I(S;Y)$ is a concave function of the input distribution $p(s)$.

Proof: See Appendix II. □

V. Capacity of C-SENMA NPNE

In this section, we investigate the capacity of C-SENMA NPNE in which no polling channels are implemented and no

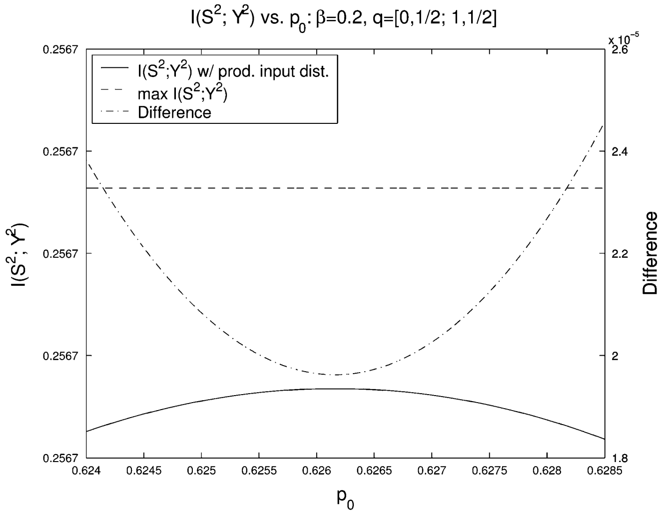

Fig. 4. $I(S^2;Y^2)$ versus p_0. In this figure, $\beta = 0.2$ and $q = [0, 1/2; 1, 1/2]$, where $q(i,j)$ means $q(i\,|\,j)$. The solid line represents $I(S^2;Y^2)$ for $p(s^2)$ being the second product extension of the distribution $[p_0, 1-p_0]$, where the maximum is achieved with the input distribution $[0.392076, 0.234084, 0.234084, 0.139756]$. The dashed line is the maximum of $I(S^2;Y^2)$ over all possible $p(s^2)$, where the optimal input distribution is $[0.396776, 0.229409, 0.229409, 0.144406]$. This figure shows that the optimal input distribution for $I(S^k;Y^k)$ may not be a product extension of some distribution.

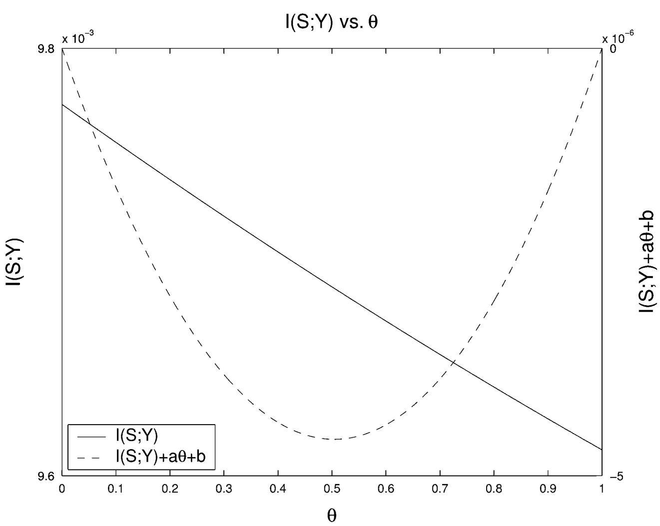

Fig. 5. $I(S;Y)$ versus θ. In this figure, $\beta = 2/3$ and $q = [1,4,5,2;4,5,2,1;5,2,1,4;2,1,4,5]/12$, where $q(i,j)$ means $q(i\,|\,j)$. The input distribution is alone a line $p_s = (1-\theta)p_{s0} + \theta p_{s1}$, where $p_{s0} = [0.18; 0; 0.17; 0.65]$ and $p_{s1} = [0; 0.16; 0.01; 0.83]$. The dashed line is $I(S;Y)$ plus some affine function of θ, which shows that $I(S;Y)$ is not a concave function of $p(s)$.

limits are imposed on how many times a sensor can transmit. Although for C-SENMA NPE, the optimal input distribution is not necessary the Qth extension of some $p(s)$, the Qth extension of the DMC capacity achieving distribution $p^*(s)$ is asymptotically optimal, which is shown in the proof of the direct part of the following theorem.

Theorem 3: The capacity of C-SENMA NPNE is

$$C^{\text{NPNE}} = (1-\beta)C^{(0)}.$$

Next we give an intuitive argument for the proof, and then we prove Theorem 3 rigorously. We show the converse first. Since the transmission scheduling is predetermined in C-SENMA NPNE, β portion of the transmission is from misinformed sensors, hence wasted. Therefore, the best rate possible is $(1-\beta)C^{(0)}$. For the direct part, as we increase the number of transmissions from one node, the difference between the rate achievable by C-SENMA NPNE and the rate achievable by a receiver supplied with extra information whether a node is misinformed is vanishing. Since the receiver supplied with the extra information achieves rate $(1-\beta)C^{(0)}$, so does C-SENMA NPNE.

A. *Direct Part of Theorem 3*

Let $p^*(s)$ be the capacity-achieving input distribution for the DMC $q(y\,|\,x)$. For a given k, schedule a new sensor to transmit in every k slots. Therefore, $\bar{C}_k = \frac{1}{k}\max_{p(s^k)} I(S^k;Y^k)$ is achievable by the direct part of Theorem 2. Let the input distribution be $p^*(s^k) = \prod_{i=1}^{k} p^*(s_i)$. The resulting rate

$$R_k^* = \frac{1}{k} I(S^k;Y^k)$$

is less than or equal to $\bar{C}_k$ and therefore achievable. We will next show that $\lim_{k\to\infty} R_k^* \geq C^{\text{NPNE}}$ and thus completing the direct part.

The calculation of $I(S^k;Y^k)$ involves the joint distribution of (S^k, Y^k) given by (7) where $p(s^k) = p^*(s^k)$. To evaluate the lower bound $I(S^k;Y^k)$, introduce a Bernoulli random variable $U \in \{0,1\}$ with mean $1-\beta$. Let U be independent of S^k. It can be verified that if we let

$$p(y^k\,|\,s^k, u) = \begin{cases} \prod_{i=1}^{k} q(y_i\,|\,s_i), & \text{if } u = 1 \\ \sum_{s'^k \in \mathcal{X}^k} p^*(s'^k) \prod_{i=1}^{k} q(y_i\,|\,s'_i), & \text{if } u = 0 \end{cases}$$

then the resulting marginal distribution $p(s^k, y^k)$ is compatible with (7). Therefore,

$$\begin{aligned} I(S^k;Y^k) &= I(S^k;Y^k,U) - I(S^k;U\,|\,Y^k) \\ &\geq I(S^k;Y^k\,|\,U) - 1 && (12) \\ &= \mathcal{P}(U=1) I(S^k;Y^k\,|\,U=1) - 1 && (13) \\ &= (1-\beta)\sum_{i=1}^{k} I(S_i;Y_i\,|\,U=1) - 1 && (14) \\ &= (1-\beta)kC^{(0)} - 1 && (15) \end{aligned}$$

where (12) holds because U is a binary random variable, (13) because S^k and Y^k are independent given $U = 0$, (14) because $S_1,\ldots,S_k$ are independent and $Y_1,\ldots,Y_k$ are independent given $(S^k, U=1)$, and (15) because $p(y_i\,|\,s_i, U=1) = q(y_i\,|\,s_i)$ and S_i has the capacity-achieving distribution for the DMC $q(y\,|\,x)$. Hence,

$$\begin{aligned} \lim_{k\to\infty} R_k^* &= \lim_{k\to\infty} \frac{1}{k} I(S^k;Y^k) \\ &\geq (1-\beta)C^{(0)} \\ &= C^{\text{NPNE}}. \end{aligned}$$

B. Converse of Theorem 3

Recall that U_i is the random variable that controls the ith node's reception of the global message during the orientation phase as in (1). Using Fano's inequality as in the proof of the converse of Theorem 2, we bound the achievable rate R from (10)

$$\begin{aligned} R &\le \frac{1}{n}\sum_{i=1}^{a} I(S_{\mathcal{I}_i}; Y_{\mathcal{I}_i}) \\ &\le \frac{1}{n}\sum_{i=1}^{a} I(S_{\mathcal{I}_i}; Y_{\mathcal{I}_i}, U_i) \\ &= \frac{1}{n}\sum_{i=1}^{a} I(S_{\mathcal{I}_i}; Y_{\mathcal{I}_i} \,|\, U_i) && (16) \\ &= \frac{1}{n}\sum_{i=1}^{a} \mathcal{P}(U_i = 1) I(S_{\mathcal{I}_i}; Y_{\mathcal{I}_i} \,|\, U_i = 1) && (17) \\ &\le \frac{1}{n}\sum_{i=1}^{a} \mathcal{P}(U_i = 1)|\mathcal{I}_i| C^{(0)} \\ &= (1-\beta)C^{(0)} && (18) \\ &= C^{\mathrm{NPNE}} \end{aligned}$$

where (16) holds because U_i is independent of $S_{\mathcal{I}_i}$, (17) because given $U_i = 0$, $S_{\mathcal{I}_i}$ is independent of $Y_{\mathcal{I}_i}$, and (18) because $\mathcal{P}(U_i = 1) = 1 - \beta$ and $\sum_{i=1}^{a} |\mathcal{I}_i| = n$.

VI. Extension to Multiple Simultaneous Transmissions

We have considered C-SENMA that only allows one sensor to transmit at a time to avoid collision. Depending on the MAC between the mobile access point and multiple sensors, one sensor transmitting at a time may not be the optimal scheme. In the following two sections, we consider multiple simultaneous transmissions and investigate the optimal number of simultaneous transmissions. We first extend the channel model to MAC with fading. We then focus on C-SENMA NPNE and consider a Gaussian MAC with fading and a network power constraint. The optimal number of simultaneous transmissions is studied for Gaussian codebooks in three fading scenarios.

A. Model Extension

The communication for C-SENMA with multiple simultaneous transmissions consists of the same steps as those introduced in Section II, except that in the third step, there are d sensors transmitting simultaneously at one time slot, either polled (for C-SENMA PNE) or prescheduled (for C-SENMA NPNE and C-SENMA NPE). Denote K_{jt} the jth node among the d nodes activated at time $t, 1 \le j \le d$. The transmission from node K_{jt} at time t depends on j, the node index among the activated group, as well as the message at this node and the time it is activated. Let $\mathcal{D} \triangleq (1, \ldots, d)$ and $Z_{\mathcal{D}}$ denote vector $(Z_1, \ldots, Z_d)$ where $Z_1, \ldots, Z_d$ are generic symbols. Denote $K_{\mathcal{D}t}$ the activating vector $(K_{1t}, \ldots, K_{dt})$. The uplink MAC with fading is modeled as follows. Denote $\widetilde{H}_{it} \in \mathcal{H}$ the channel state associate with node i at time t. Assume that the channel states associated with nodes $K_{\mathcal{D}t}$ at time $t, \widetilde{H}_{K_{\mathcal{D}t}t} \in \mathcal{H}^d$, has distribution $p(h_{\mathcal{D}})$, i.i.d. across t. The fading process $\widetilde{H}_{K_{\mathcal{D}t}t}$ is independent of the transmission from the sensors and the access point. For convenience, denote $H_{\mathcal{D}t} \triangleq \widetilde{H}_{K_{\mathcal{D}t}t}$ the channel states associated with nodes $K_{\mathcal{D}t}$ activated at time t. We assume that the realization of $H_{\mathcal{D}t}$, unknown to the sensors, is known to the mobile access point at the end of time slot t. The memoryless channel output, conditioning on the simultaneous transmissions from nodes $K_{\mathcal{D}t}$ and the associated channel states, is governed by the transition probability $q(y \,|\, x_{\mathcal{D}}, h_{\mathcal{D}})$, where $y \in \mathcal{Y}$ is the channel output to the mobile access point, $h_j \in \mathcal{H}$ the channel state associated with the jth node among the d transmitting nodes, $1 \le j \le d$, and $x_j \in \mathcal{X}$ the transmission from the jth among the d nodes.

Let $\pi_{\mathcal{D}} = [\pi_1, \pi_2, \ldots, \pi_d]^T$ be a permutation in the domain $\mathcal{D}$. From the above assumption, the channel states associated with vector $(K_{\pi_1 t}, \ldots, K_{\pi_d t})$ has the identical distribution $p(h_{\mathcal{D}})$. Therefore, a necessary condition for $p(h_{\mathcal{D}})$ is that $p(h_{\mathcal{D}})$ is symmetrical with respect to input permutations, i.e., $p(h_{\pi_1}, \ldots, h_{\pi_d})$ is identical for all permutations $\pi_{\mathcal{D}}$. Similarly, $q(y \,|\, x_{\mathcal{D}}, h_{\mathcal{D}})$ needs to be symmetrical with respect to node permutation, i.e., $q(y \,|\, x_{\pi_1}, \ldots, x_{\pi_d}, h_{\pi_1}, \ldots, h_{\pi_d})$ is identical for all permutations $\pi_{\mathcal{D}}$.

B. Achievable Rate

The results for C-SENMA PNE is extended in the following theorem.

Theorem 4: Let

$$C_d^{(0)} = \max_{p(x_{\mathcal{D}})} I(X_{\mathcal{D}}; Y \,|\, H_{\mathcal{D}}) \tag{19}$$

where $p(x_{\mathcal{D}}, h_{\mathcal{D}}, y) = p(x_{\mathcal{D}})p(h_{\mathcal{D}})q(y \,|\, x_{\mathcal{D}}, h_{\mathcal{D}})$. The capacity of C-SENMA PNE with d simultaneous transmissions is

$$C_d^{\mathrm{PNE}} = \begin{cases} C_d^{(0)}, & \text{if } \beta < 1 \\ 0, & \text{if } \beta = 1. \end{cases}$$

The proof is a straightforward extension of the proof of Theorem 1. In the first phase, we locate d nodes that, with high probability, have the correct global message. We use the same method (message group index transmission and comparison) to locate one node as in the proof of Theorem 1, and repeat it d times to locate d nodes. The only technical difference is that in the multiple simultaneous transmissions case we have a MAC. To locate one node, we select d nodes as the transmission set but fix $d-1$ nodes's transmission to a given symbol in the channel input alphabet. This way, we convert the MAC to a DMC where the only node whose transmission is not fixed is the input node to the DMC. In the second phase, we retrieve the global message from these d nodes. Since these d nodes have the same message, we can achieve the multiple-input single-output (MISO) channel capacity (19) in the second phase. The number of time slots used in the first phase is only a function of the target error probability, but not a function of the codeword length in the second phase. Hence, by increasing the codeword length in the second phase, we can make the overhead associated with the first phase arbitrarily small. Thus, the MISO channel capacity (19) is achievable by C-SENMA PNE with d simultaneous transmissions.

The next theorem gives an achievable rate for C-SENMA NPNE with multiple simultaneous transmissions.

Theorem 5: For a given d, consider random variables $X_{\mathcal{D}}$ with distribution $p(x_{\mathcal{D}})$. Denote $p_{\mathcal{I}}(x_{\mathcal{I}})$ the marginal distribution of $X_{\mathcal{I}}$ for $\mathcal{I} \subset \mathcal{D}$. Let $X_{\mathcal{D}}^{(\mathcal{I})} = (X_1^{(\mathcal{I})}, \ldots, X_d^{(\mathcal{I})})$ be a random vector derived from $X_{\mathcal{D}}$ with distribution

$$p^{(\mathcal{I})}\left(x_{\mathcal{D}}^{(\mathcal{I})}\right) = p_{\mathcal{I}}\left(x_{\mathcal{I}}^{(\mathcal{I})}\right)\prod_{j\notin\mathcal{I}} p_j\left(x_j^{(\mathcal{I})}\right).$$

Let $H_{\mathcal{D}}$, independent of $X_{\mathcal{D}}^{(\mathcal{I})}$, have distribution $p(h_{\mathcal{D}})$. And let the conditional distribution of Y be

$$p(y \mid x_{\mathcal{D}}^{(\mathcal{I})}, h_{\mathcal{D}}) = q(y \mid x_{\mathcal{D}}^{(\mathcal{I})}, h_{\mathcal{D}}).$$

The following rate is achievable for C-SENMA NPNE with d simultaneous transmissions:

$$C_d^{\text{NPNE}} = \max_{p(x_{\mathcal{D}})} \sum_{\mathcal{I}\subset\mathcal{D}} (1-\beta)^{|\mathcal{I}|}\beta^{d-|\mathcal{I}|} I\left(X_{\mathcal{I}}^{(\mathcal{I})}; Y \mid H_{\mathcal{D}}\right). \quad (20)$$

For $d = 1$, (20) is the capacity as shown in Theorem 3. The strategy to achieve (20) is similar to the proof of the direct part of Theorem 3. We schedule d new sensors to transmit simultaneously in every k consecutive slots. Let symbols of codewords in the codebook be d-dimensional vectors. The codebook is generated randomly with 2^{nR} messages and codeword length n according to some distribution $p(s_{\mathcal{D}})$ where $s_{\mathcal{D}} \in \mathcal{X}^d$. When a node is scheduled to transmit at time t as the jth node in the activated group, it transmits the jth element in the tth symbol vector of the codeword corresponding to its local message. Typical set decoding is employed. It can be shown that the following rate is achievable with the aforementioned scheduling and coding:

$$R_{dk} \triangleq \frac{1}{k} I\left(S_{\mathcal{D}}^k; Y^k \mid H_{\mathcal{D}}^k\right) \quad (21)$$

where $S_{\mathcal{D}}^k \in \mathcal{X}^{dk}, Y^k \in \mathcal{Y}^k, H_{\mathcal{D}}^k \in \mathcal{H}^{dk}$, and

$$\begin{aligned} p&\left(s_{\mathcal{D}}^k, h_{\mathcal{D}}^k, y^k\right) \\ &= \left(\prod_{i=1}^k p(s_{\mathcal{D}i})p(h_{\mathcal{D}i})\right)\left(\sum_{\mathcal{I}\subset\mathcal{D}} (1-\beta)^{|\mathcal{I}|}\beta^{|\bar{\mathcal{I}}|}\right. \\ &\quad \left.\cdot \prod_{i=1}^k \sum_{s'_{\bar{\mathcal{I}}}\in\mathcal{X}^{|\bar{\mathcal{I}}|}} q(y_i \mid s_{\mathcal{I}i}, s'_{\bar{\mathcal{I}}}, h_{\mathcal{I}i}, h_{\bar{\mathcal{I}}i}) \prod_{j\in\bar{\mathcal{I}}} p_j(s'_j)\right). \quad (22)\end{aligned}$$

In (22), $\bar{\mathcal{I}} = \mathcal{D}\setminus\mathcal{I}$, and $p_j(s)$ is the marginal distribution of the jth element in $p(s_{\mathcal{D}})$. The next step (omitted) is to show that R_{dk} converges to C_d^{NPNE} as k goes to infinity and $p(s_{\mathcal{D}})$ is optimized, the proof of which can be carried out as in the proof of the direct part of Theorem 3.

To explain (20) intuitively, view $\mathcal{I}$ in (20) as the set of indices of well-informed nodes. The probability that $\mathcal{I}$ appears is $(1-\beta)^{|\mathcal{I}|}\beta^{d-|\mathcal{I}|}$. We can assume that the access point knows which nodes are misinformed among the d transmitting nodes since the difference between the achievable rate when the access point knows the index set of misinformed nodes and the achievable rate when the access point does not is vanishingly small as k goes to infinity. Once the access point knows the index set of the misinformed nodes, we can view those misinformed nodes as noise sources and the white noises are $X_{\bar{\mathcal{I}}}^{(\mathcal{I})}$ with independent distribution $\prod_{j\in\bar{\mathcal{I}}} p_j(x_j^{(\mathcal{I})})$. The noise sources $X_{\bar{\mathcal{I}}}^{(\mathcal{I})}$ are independent because each misinformed node randomly selects a codeword in the codebook independently. Although transmissions at different time slots by the same misinformed node are from the same random codeword, the noise generated by the node is white in time because the codebook is generated with independent distribution in time. Conditioning on $\mathcal{I}$, $X_{\mathcal{I}}^{(\mathcal{I})}$ is the input, Y the output, and $H_{\mathcal{D}}$ the channel side information. Thus, (20) is achievable.

Next, we consider a Gaussian MAC with a network-wise power constraint and investigate the optimal number of simultaneous transmissions under three different fading situations.

VII. Gaussian MAC With a Network Power Constraint

Consider the following Gaussian MAC:

$$Y = V + \sum_i \widetilde{H}_i \widetilde{X}_i$$

where $V \in \mathcal{C}$ is the additive white Gaussian noise with zero mean and unit variance, $\widetilde{X}_i \in \mathcal{C}$ the input from the ith sensor, $\widetilde{H}_i \in \mathcal{C}$ the fading channel gain associated with the ith sensor, and $Y \in \mathcal{C}$ the channel output. We impose a total power constraint P on the network,[6] i.e., the total transmit power from all sensors is less than or equal to P

$$\frac{1}{n}\sum_{t=1}^n \sum_{j=1}^d |x_{jt}|^2 \le P$$

where x_{jt} is the transmission from the jth node among the d transmitting nodes at time t. We assume that the channel gains $\widetilde{H}_i$'s have distribution $p(h)$, i.i.d. across sensors and time. For convenience, denote X_j and H_j the transmission from and the channel gain associated with the jth node among the activated group, respectively, $1 \le j \le d$. Let $H_{\mathcal{D}} = (H_1, \ldots, H_d)$ be the channel gains associated with the d transmitting sensors. The realization of the channel states $H_{\mathcal{D}}$ is assumed to be known to the mobile access point. We focus on C-SENMA NPNE, and an achievable rate is given by (20) with the power constraint on the input distribution, $\sum_{j=1}^d E[|X_j|^2] \le P$. We consider three types of fading channels: nonfading, phase-fading, and Rayleigh-fading. The optimal number of simultaneous transmissions is studied for Gaussian codebooks.

A. Nonfading Case

In the nonfading case, $p(h) = \delta(h-1)$, i.e., the channel gain for each sensor is 1. If $\beta = 0$, the C-SENMA channel reduces to a Gaussian MISO channel with power constraint P. From the MISO channel capacity, we know that $C_d^{(0)} = \log_2(1+dP) = O(\log_2 d)$, which goes to infinity as d increases. Next we show that C_d^{NPNE} is still $O(\log_2 d)$ as long as $\beta < 1$.

Consider activating d sensors at a time. Let the input random vector $X_{\mathcal{D}} = (X_1, \ldots, X_d) \sim \mathcal{N}_C(\mathbf{0}, \frac{P}{d}\mathbf{1}\mathbf{1}^T)$, where $\mathbf{1} = [1, 1, \ldots, 1]^T$. For $\mathcal{I} \subset \mathcal{D}$, since the derived random

[6]If, instead, a power constraint is imposed on individual nodes, the total transmit power can reach infinity if more and more nodes transmit simultaneously.

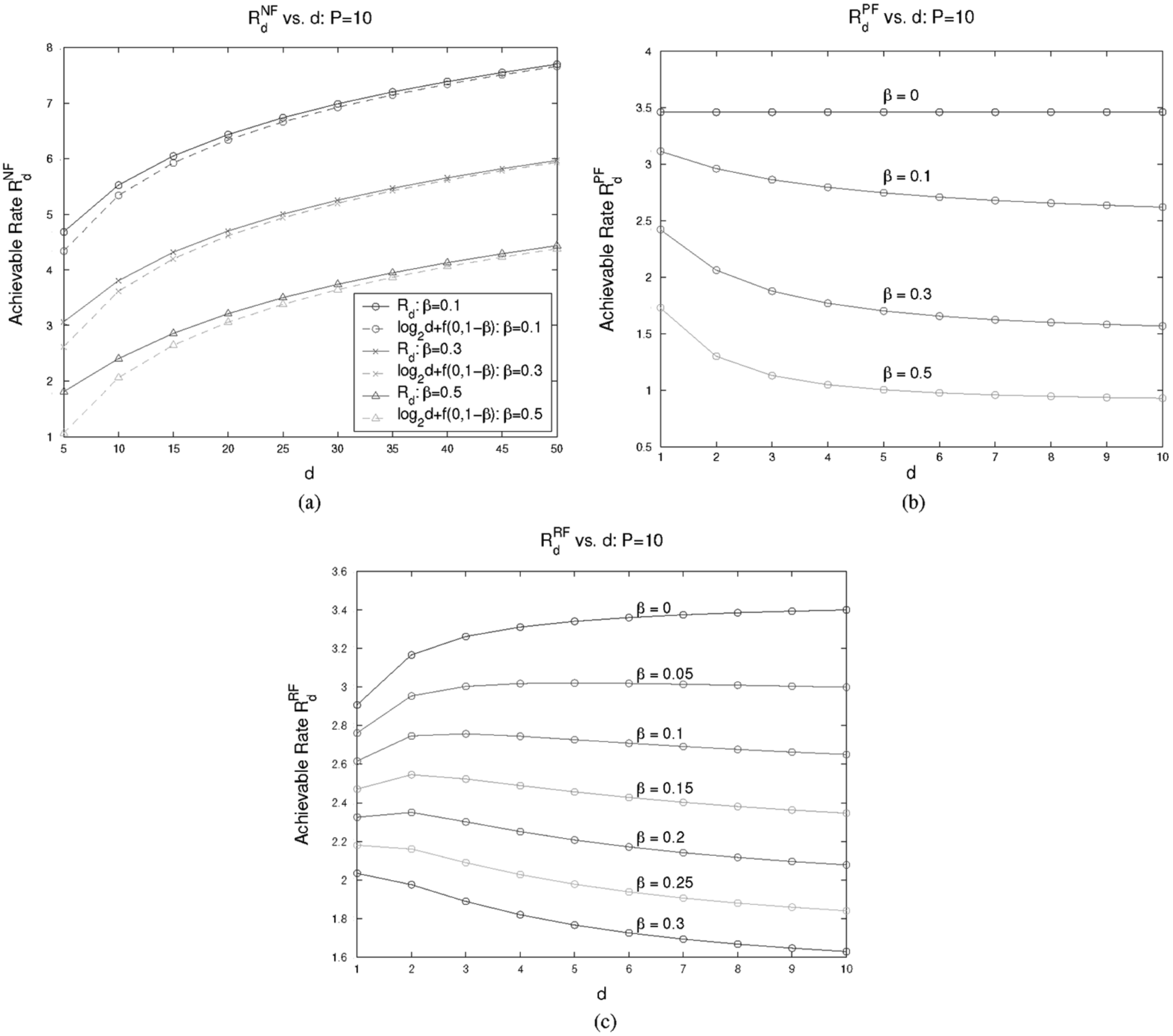

Fig. 6. Achievable rates versus d for C-SENMA NPNE with the Gaussian MAC: $P = 10$ and various β's. (a) Nonfading case. (b) Phase-fading case. (c) Rayleigh-fading case.

variables $X_{\bar{\mathcal{I}}}^{(\mathcal{I})}$, independent of each other, are independent of $X_{\mathcal{I}}^{(\mathcal{I})}, X_{\bar{\mathcal{I}}}^{(\bar{\mathcal{I}})}$ contribute $(d - |\mathcal{I}|)P/d$ power to the additive noise. Therefore,

$$I\left(X_{\mathcal{I}}^{(\mathcal{I})}; Y\right) = \log_2\left(1 + \frac{|\mathcal{I}|^2 P/d}{1 + (d - |\mathcal{I}|)P/d}\right).$$

From Theorem 5, the following rate is achievable:

$$R_d^{\mathrm{NF}} = \sum_{i=0}^{d} (1-\beta)^i \beta^{d-i} \binom{d}{i} \log_2\left(1 + \frac{i^2 P/d}{1 + (d-i)P/d}\right).$$

The next proposition shows that $R_d^{\mathrm{NF}} = O(\log_2 d)$. Since

$$R_d^{\mathrm{NF}} \leq C_d^{\mathrm{NPNE}} \leq C_d^{(0)} = O(\log_2 d)$$

we have $C_d^{\mathrm{NPNE}} = O(\log_2 d)$.

Proposition 2: For $\beta \in [0, 1)$ and $p > 0$

$$\lim_{d \to \infty} \left(R_d^{\mathrm{NF}} - \log_2 d\right) = \log_2\left(\frac{(1-\beta)^2 P}{1 + \beta P}\right).$$

Proof: See Appendix III. □

Proposition 2 implies that for the nonfading case, R_d^{NF} grows at the rate of $O(\log_2 d)$, increasing to infinity as d goes to infinity. Let $f(a, b) = \log_2(a + \frac{b^2 P}{1+(1-b)P})$. Fig. 6(a) shows the achievable rate R_d^{NF} and the approximation function $\log_2 d + f(0, 1-\beta)$ versus d when $P = 10$ and $\beta = 0.1, 0.3, 0.5$. As shown in Fig. 6(a), R_d^{NF} and the approximation function converge as d increases. As expected, the achievable rate is higher for a network with a smaller orientation error probability.

B. Phase-Fading Case

In the phase-fading case, $p(h)$ is a uniform distribution on the unit circle.

Proposition 3: Suppose d sensors are activated at a time. With Gaussian codebooks $X_{\mathcal{D}} \sim \mathcal{N}_C(\mathbf{0}, \Sigma)$ where $\mathrm{Tr}(\Sigma) \leq P$, the maximum achievable rate, optimized over Σ, is

$$R_{\max}^{\mathrm{PF}} = (1-\beta)\log_2(1+P).$$

The maximum is achieved when Σ has only one nonzero diagonal entry, which is equal to P.

Proof: See Appendix IV. □

Proposition 3 suggests that, with Gaussian codebooks, the optimal d for the phase-fading case is one, i.e., it is better to activate one sensors at a time. To see how the achievable rate depends on d, consider input distribution $X_{\mathcal{D}} \sim \mathcal{N}_C(\mathbf{0}, \frac{P}{d}\boldsymbol{I})$ for a given d. From Theorem 5, the achievable rate is given by

$$R_d^{\mathrm{PF}} = \sum_{i=1}^{d}(1-\beta)^i\beta^{d-i}\binom{d}{i}\log_2\left(1+\frac{iP/d}{1+(d-i)P/d}\right).$$

Fig. 6(b) shows the achievable rate R_d^{PF} versus d when $P=10$ and $\beta = 0, 0.1, 0.3, 0.5$. As expected from Proposition 3, R_d^{PF} achieves maximum when $d=1$.

C. Rayleigh-Fading Case

In the Rayleigh-fading case, $p(h)$ is the distribution of a complex Gaussian with zero mean and unit variance. Consider activating d sensors at a time and using a Gaussian codebook $X_{\mathcal{D}} \sim \mathcal{N}_C(\mathbf{0}, \frac{P}{d}\boldsymbol{I})$. For $\mathcal{I} \subset \mathcal{D}$, let $i = |\mathcal{I}|$. We have

$$\begin{aligned}
I&\left(X_{\mathcal{I}}^{(\mathcal{I})}; Y \mid H_{\mathcal{D}}\right)\\
&= E\log_2\left(1+\frac{\sum_{j\in\mathcal{I}}|H_j|^2P/d}{1+\sum_{j\in\bar{\mathcal{I}}}|H_j|^2P/d}\right)\\
&= \begin{cases} \int_0^\infty\int_0^\infty \log_2\left(1+\frac{z_1P/d}{1+z_2P/d}\right) \\ \quad \frac{z_1^{i-1}e^{-z_1}}{(i-1)!}\frac{z_2^{d-i-1}e^{-z_2}}{(d-i-1)!}\mathrm{d}z_1\mathrm{d}z_2, & \text{if } 0<i<d\\ \int_0^\infty \log_2\left(1+\frac{z_1P}{d}\right)\frac{z_1^{d-1}e^{-z_1}}{(d-1)!}\mathrm{d}z_1, & \text{if } i=d\\ 0, & \text{if } i=0 \end{cases} \qquad (23)\\
&\triangleq F(d,i)
\end{aligned}$$

where (23) uses the fact that $Z = \sum_{j\in\mathcal{I}}|H_j|^2$ is chi-square distributed with $2i$ degrees of freedom and distribution $\frac{z^{i-1}e^{-z}}{(i-1)!}$, [21]. From Theorem 5, the following rate is achievable:

$$R_d^{\mathrm{RF}} = \sum_{i=1}^{d}(1-\beta)^i\beta^{d-i}\binom{d}{i}F(d,i).$$

Fig. 6(c) shows the achievable rate R_d^{RF} versus d when the power constraint $P=10$ and the orientation error probability $\beta = 0, 0.05, 0.1, 0.15, 0.2, 0.25, 0.3$. As shown in Fig. 6(c), when $\beta = 0$, R_d^{RF} increases monotonically and converges to the limit $\log_2(1+P)$ as d increases, which is expected because of the diversity gain. With $\beta > 0$, R_d^{RF} is no longer monotonically increasing. R_d^{RF} first increases and then decreases as d increases. As β increases, the optimal d decreases, reaching one when $\beta \geq 0.25$ in Fig. 6(c).

The optimal number of sensors to activate at a time is infinity for the nonfading case, one for the phase-fading case. For the Rayleigh-fading case, the optimal number decreases as the orientation error probability increases. The phenomenon is related to the channel diversity gain associated with activating more than one sensors at a time. For the phase-fading case, there is no channel diversity gain. Activating more than one nodes at a time reduces the achievable rate due to the interference transmission from misinformed nodes. For the nonfading case, the diversity gain is so strong that it counteracts the rate loss due to the interference from misinformed sensors. For the Rayleigh-fading case, the diversity gain is in between. The optimal d has a tradeoff between the diversity gain and the interference loss.

VIII. Conclusion

We have considered and modeled the communication aspects of cooperative sensor networks with misinformed sensors. We have derived the capacity expressions for three system configurations: C-SENMA with Polling with No Energy constraints, C-SENMA with No Polling with No Energy constraints, and C-SENMA with No Polling with an Energy constraint. For C-SENMA PNE, there is no rate loss with the presence of misinformed sensors. For C-SENMA NPNE, the capacity is lowered by a factor of $1-\beta$ compared to systems without misinformed sensors.

We have extended the results to systems with multiple simultaneous transmissions and studied the effect of fading on the achievable rate of a Gaussian MAC with a total power constraint. The optimal number of sensors to activate at a time is infinity for the nonfading case, one for the phase-fading case. For the Rayleigh-fading case, the optimal number decreases as the orientation error probability increases.

Another possible system configuration is C-SENMA with Polling with an Energy constraint (PE). Obviously, C^{NPE}, converging to C^{NPNE} as the energy limit Q goes to infinity, is an achievable rate for C-SENMA PE. With similar strategies used in proving Theorem 1 (first probe the correctness of the nodes' local messages, then poll a node that is most likely to be correct up to its transmission limit, and then repeat the procedure of probing and polling), it can be shown that there exist an achievable rate for C-SENMA PE that converges to C^{PNE} as Q increases. However, the exact capacity expression for C-SENMA PE with finite Q is unknown due to the flexibility of its polling strategies.

Besides the investigation of the achievable rate of C-SENMA PE, future research directions on cooperate SENMA also include the effect of noise correlation and distance differences between nodes and the access point on the achievable rate of C-SENMA. As the network becomes denser and larger, noise is likely to be correlated among nearby sensors, and some nodes may have better channels due to shorter distance to the mobile access point. In this paper, we focused on the information retrieval aspect of the cooperative sensor network with misinformed nodes. Another possible future research is to investigate a framework that unifies the sensing and the information retrieval aspects of the sensor network.

APPENDIX I
PROOF OF LEMMA 1

Let (S,Y) have joint distribution $p(s,y)$. Since $p^{(n)}(s,y)$ converges to $p(s,y)$, we have the variance of $\log_2 p(S_1^{(n)})$ converges to the variance of $\log_2 p(S)$ as n goes to infinity. Therefore, the variance of $\log_2 p(S_1^{(n)})$ is bounded away from infinity for large n. Thus, by the weak law of large numbers on triangular arrays (e.g., see [22, Theorem (5.4)])

$$-\frac{1}{n}\log_2 p\left((S^{(n)})^n\right) + E\left[\log_2 p(S_1^{(n)})\right] \to 0 \text{ in prob.}$$

Since $E\left[\log_2 p(S_1^{(n)})\right]$ converges to $E[\log_2 p(S)]$, we have

$$-\frac{1}{n}\log_2 p\left((S^{(n)})^n\right) \to -E[\log_2 p(S)] = H(S) \text{ in prob.}$$

Similarly

$$-\frac{1}{n}\log_2 p\left((Y^{(n)})^n\right) \to H(Y) \text{ in prob.}$$
$$-\frac{1}{n}\log_2 p\left((S^{(n)})^n, (Y^{(n)})^n\right) \to H(S,Y) \text{ in prob.}$$

The first part of the lemma, (8), is established from the above three convergences in probability.

To prove the second part of the lemma, we first establish two properties of the typical set $A_\epsilon^{(n)}$. Since

$$\begin{aligned}1 &= \sum p(s^n,y^n)\\ &\geq \sum_{A_\epsilon^{(n)}} p(s^n,y^n)\\ &\geq |A_\epsilon^{(n)}| 2^{-n(H(S,Y)+\epsilon)}\end{aligned}$$

we have

$$\left|A_\epsilon^{(n)}\right| \leq 2^{n(H(S,Y)+\epsilon)}. \tag{24}$$

If $(s^n,y^n) \in A_\epsilon^{(n)}$, for all $i \in \{1,\ldots,n\}$, we have

$$p(s_i)p(y_i) > 0 \tag{25}$$

since otherwise $p(s^n,y^n) = 0$ and therefore,

$$\left|-\frac{1}{n}\log_2 p(s^n,y^n) - H(S,Y)\right| = \infty$$

contradicting the assumption that $(s^n,y^n) \in A_\epsilon^{(n)}$.

With the above two properties of $A_\epsilon^{(n)}$, we are ready to prove the second part of the lemma. Since S and Y have finite support and $\tilde{p}^{(n)}(s,y)$ converges to $p(s)p(y)$ as n goes to infinity, for the same $\epsilon > 0$, there exists n_0 such that for all $n \geq n_0$ and for all (s,y) where $p(s)p(y) > 0$

$$\tilde{p}^{(n)}(s,y) \leq p(s)p(y)2^\epsilon. \tag{26}$$

Therefore, for $n \geq n_0$

$$\begin{aligned}&\mathcal{P}\left[\left((\widetilde{S}^{(n)})^n, (\widetilde{Y}^{(n)})^n\right) \in A_\epsilon^{(n)}\right]\\
&= \sum_{(s^n,y^n)\in A_\epsilon^{(n)}} \tilde{p}^{(n)}(s^n,y^n)\\
&= \sum_{(s^n,y^n)\in A_\epsilon^{(n)}} \prod_{i=1}^n \tilde{p}^{(n)}(s_i,y_i)\\
&\leq \sum_{(s^n,y^n)\in A_\epsilon^{(n)}} \prod_{i=1}^n p(s_i)p(y_i)2^\epsilon \qquad (27)\\
&= \sum_{(s^n,y^n)\in A_\epsilon^{(n)}} p(s^n)p(y^n)2^{n\epsilon}\\
&\leq 2^{n(H(S,Y)+\epsilon)}2^{-n(H(S)-\epsilon)}2^{-n(H(Y)-\epsilon)}2^{n\epsilon} \qquad (28)\\
&= 2^{-n(I(S;Y)-4\epsilon)}\end{aligned}$$

where (27) is due to (25) and (26), and (28) is due to (24). □

APPENDIX II
PROOF OF PROPOSITION 1

Given $p_S(s)$ for $s \in \mathcal{X}$, we have

$$I(S;Y) = H(Y) - H(Y\,|\,S)$$

where

$$\begin{aligned}p_Y(y) &= \sum_{x\in\mathcal{X}} p_S(x)q(y\,|\,x)\\
H(Y) &= -\sum_{y\in\mathcal{Y}} p_Y(y)\log p_Y(y)\\
p_{Y|S}(y\,|\,s) &= (1-\beta)q(y\,|\,s) + \beta\sum_{x\in\mathcal{X}} p_S(x)q(y\,|\,x)\\
H(Y\,|\,S) &= -\sum_{s\in\mathcal{X}} p_S(s)\sum_{y\in\mathcal{Y}} p_{Y|S}(y\,|\,s)\log p_{Y|S}(y\,|\,s).\end{aligned}$$

The mutual information is measured in nats. Let $\Delta p_S(i)$ denote the ith component of the directional vector on the probability simplex plane $\sum_{i\in\mathcal{X}}\Delta p_S(i) = 0$. To prove that $I(S;Y)$ is concave in the input distribution $p_S(s)$, we need to show that, when $p_S(i) > 0$ for all $i \in \mathcal{X}, \sum_{i\in\mathcal{X}} p_S(i) = 1$, and $\sum_{i\in\mathcal{X}}\Delta p_S(i) = 0$, the following inequality holds:

$$\sum_{i\in\mathcal{X}}\sum_{j\in\mathcal{X}} \frac{\partial^2 I(S;Y)}{\partial p_S(i)\partial p_S(j)}\Delta p_S(i)\Delta p_S(j) \leq 0.$$

Some calculations give

$$\begin{aligned}&\sum_{i\in\mathcal{X}}\sum_{j\in\mathcal{X}} \frac{\partial^2 I(S;Y)}{\partial p_S(i)\partial p_S(j)}\Delta p_S(i)\Delta p_S(j)\\
&= \sum_{y\in\mathcal{Y}}\Bigg[-\frac{\left(\sum_{i\in\mathcal{X}} q(y\,|\,i)\Delta p_S(i)\right)^2}{\sum_{x\in\mathcal{X}} p_S(x)q(y\,|\,x)} + 2\beta\left(\sum_{i\in\mathcal{X}} q(y\,|\,i)\Delta p_S(i)\right)\\
&\quad\cdot\left(\sum_{i\in\mathcal{X}}\Delta p_S(i)\log\left((1-\beta)q(y\,|\,i) + \beta\sum_{x\in\mathcal{X}} p_S(x)q(y\,|\,x)\right)\right)\\
&\quad + \sum_{s\in\mathcal{X}} \frac{p_S(s)\beta^2\left(\sum_{i\in\mathcal{X}} q(y\,|\,i)\Delta p_S(i)\right)^2}{(1-\beta)q(y\,|\,s) + \beta\sum_{x\in\mathcal{X}} p_S(x)q(y\,|\,x)}\Bigg]. \qquad (29)\end{aligned}$$

It is then sufficient to prove that for all $y \in \mathcal{Y}$, the expression inside the brackets of (29) is nonpositive, which will be shown next. For notation simplicity, for a given y, let $q = q(y \,|\, x_0(y)), \epsilon = \epsilon(y), p = p_S(x_0(y)), \Delta p = \Delta p_S(x_0(y))$, and

$$
\begin{aligned}
a_1 &\triangleq \sum_{i\in\mathcal{X}} q(y\,|\,i)\Delta p_S(i) = (q-\epsilon)\Delta p\\
a_2 &\triangleq \sum_{x\in\mathcal{X}} p_S(x)q(y\,|\,x) = pq + \epsilon(1-p)\\
b_1 &\triangleq (1-\beta)q + \beta a_2\\
b_2 &\triangleq (1-\beta)\epsilon + \beta a_2\\
a_3 &\triangleq \sum_{i\in\mathcal{X}} \Delta p_S(i)\log\left((1-\beta)q(y\,|\,i) + \beta\sum_{x\in\mathcal{X}} p_S(x)q(y\,|\,x)\right)\\
&= \Delta p \log\frac{b_1}{b_2}\\
a_4 &\triangleq \sum_{s\in\mathcal{X}} \frac{p_S(s)}{(1-\beta)q(y\,|\,s) + \beta\sum_{x\in\mathcal{X}} p_S(x)q(y\,|\,x)}\\
&= \frac{p}{b_1} + \frac{1-p}{b_2}.
\end{aligned}
$$

We need the following lemma to prove the nonpositivity of (29).

Lemma 2: For $x \geq 1$

$$2\log(x) \leq x - \frac{1}{x}.$$

Proof: When $x = 1$, $2\log(x) = x - 1/x$. For $x \geq 1$

$$\frac{\mathrm{d}(2\log(x))}{\mathrm{d}x} = \frac{2}{x} \leq 1 + \frac{1}{x^2} = \frac{\mathrm{d}(x - 1/x)}{\mathrm{d}x}.$$

Therefore, the lemma is established. □

If $q \geq \epsilon$, applying the lemma to $2a_1a_3$, we have

$$2a_1a_3 \leq \Delta p^2(q-\epsilon)\left(\frac{b_1}{b_2} - \frac{b_2}{b_1}\right).$$

The preceding inequality also holds for $q < \epsilon$. Thus, the expression inside the brackets of (29) is upper-bounded as follows:

$$
\begin{aligned}
&-\frac{a_1^2}{a_2} + 2\beta a_1 a_3 + \beta^2 a_1^2 a_4\\
&\quad\leq -\frac{a_1^2}{a_2} + \beta\Delta p^2(q-\epsilon)\left(\frac{b1}{b2} - \frac{b2}{b1}\right) + \beta^2 a_1^2 a_4\\
&\quad= -\frac{c_1}{c_2c_3c_4}\\
&\quad\leq 0
\end{aligned}
$$

where

$$
\begin{aligned}
c_1 &\triangleq \epsilon q\Delta p^2(1-\beta)^2(q-\epsilon)^2 \geq 0\\
c_2 &\triangleq \epsilon\beta(1-p) + q(1-\beta(1-p)) \geq 0\\
c_3 &\triangleq \epsilon(1-p) + qp \geq 0\\
c_4 &\triangleq \epsilon(1-\beta p) + q\beta p \geq 0.
\end{aligned}
$$

Therefore, the expression inside the brackets of (29) is nonpositive, proving the proposition. □

APPENDIX III
PROOF OF PROPOSITION 2

Let $B_1, B_2, \ldots$, be i.i.d. Bernoulli with support $\{0,1\}$ and mean $1-\beta$. Then $T_d \triangleq \sum_{i=1}^{d} B_i$ is binomially distributed. Let

$$f(a,b) \triangleq \log_2\left(a + \frac{b^2P}{1+(1-b)P}\right).$$

We have

$$
\begin{aligned}
R_d^{\mathrm{NF}} - \log_2 d &= \sum_{i=0}^{d}(1-\beta)^i\beta^{d-i}\binom{d}{i} f(1/d, i/d)\\
&= E[f(1/d, T_d/d)].
\end{aligned}
$$

Since $f(a,b)$ is continues at $(a = 0, b = 1-\beta)$, for all $\epsilon > 0$, there exists a $\delta > 0$ such that for all $a \in [0,\delta) \triangleq \mathcal{A}$ and $b \in (1-\beta-\delta, 1-\beta+\delta) \triangleq \mathcal{B}$, $|f(a,b) - f(0,1-\beta)| \leq \epsilon/3$. For the same ϵ and δ

$$
\begin{aligned}
&|E[f(1/d,T_d/d)] - f(0,1-\beta)|\\
&\quad\leq E[|f(1/d,T_d/d) - f(0,1-\beta)|]\\
&\quad\leq E_{T_d/d\in\mathcal{B}}[|f(1/d,T_d/d) - f(0,1-\beta)|]\\
&\qquad + E_{T_d/d\notin\mathcal{B}}[|f(1/d,T_d/d)|]\\
&\qquad + \mathcal{P}(T_d/d \notin \mathcal{B})f(0,1-\beta).
\end{aligned} \tag{30}
$$

For large d, $1/d \in \mathcal{A}$. Therefore, due to the continuity of f, the first term in (30) is upper-bounded by $\epsilon/3$. The third term, by the weak law of large number, is upper-bounded by $\epsilon/3$ for large d. We bound the second term as follows. For $d \geq 1$ and $i = 0, \ldots, d$, we have

$$
\begin{aligned}
f(1/d, i/d) &\leq \log_2(1/d + P)\\
f(1/d, i/d) &\geq \log_2(1/d).
\end{aligned}
$$

For large d such that $1/d + P \leq d$, we have

$$|f(1/d, i/d)| \leq \log_2 d.$$

Hence, by Chebyshev inequality

$$
\begin{aligned}
E_{T_d/d\notin\mathcal{B}}[|f(1/d,T_d/d)|] &\leq \mathcal{P}(T_d/d \notin \mathcal{B})\log_2 d\\
&\leq \frac{\beta(1-\beta)}{\delta^2 d}\log_2 d\\
&\leq \epsilon/3, \qquad \text{for large } d.
\end{aligned}
$$

Therefore,

$$|E[f(1/d,T_d/d)] - f(0,1-\beta)| \leq \epsilon, \quad \text{for large } d.$$

Since $\epsilon > 0$ is arbitrary, we have $\lim_{d\to\infty} E[f(1/d,T_d/d)] = f(0,1-\beta)$, proving the proposition. □

APPENDIX IV
PROOF OF PROPOSITION 3

The following lemma and corollary are used in the proof of Proposition 3.

Lemma 3: For $d \geq 1, i \leq d$, and $r_1, \ldots, r_d \geq 0$, let $r_{j+d} = r_j$ for $j = 1, \ldots, d-1$. The following inequality holds:

$$\left(1 + \sum_{j=1}^{d} r_j\right)^i \leq \prod_{k=1}^{d}\left(1 + \sum_{j=k}^{k+i-1} r_j\right). \quad (31)$$

Proof: Both sides of (31) can be written as the sum of terms $\prod_{j=1}^{d} r_j^{t_j}$, where t_j is the frequency of r_j. Let $\boldsymbol{t} = [t_1, \ldots, t_d]^T$ be a d-dimensional vector with nonnegative integer entries. Rewrite the left-hand side (LHS) of (31) as

$$\left(1 + \sum_{j=1}^{d} r_j\right)^i = \sum_{\boldsymbol{t}: \sum_{j=1}^{d} t_j \leq i} \frac{i!}{(i - \sum_{k=1}^{d} t_k)! \cdot \prod_{k=1}^{d} t_k!} \prod_{j=1}^{d} r_j^{t_j}. \quad (32)$$

Rewrite the right-hand side (RHS) of (31) as

$$\prod_{k=1}^{d}\left(1 + \sum_{j=k}^{k+i-1} r_j\right) = \sum_{\boldsymbol{t}: \sum_{j=1}^{d} t_j \leq i} c_{\boldsymbol{t}} \prod_{j=1}^{d} r_j^{t_j} \quad (33)$$

where $c_{\boldsymbol{t}}$ is the count associate with the term $\prod_{j=1}^{d} r_j^{t_j}$. Next we show a lower bound on $c_{\boldsymbol{t}}$. For all $l \in \{1, \ldots, d\}$, the set

$$\mathcal{S} \triangleq \left\{1 + \sum_{j=k}^{k+i-1} r_j : k = 1, \ldots, d\right\}$$

has i elements that contain r_l in the summation.[7] To calculate $c_{\boldsymbol{t}}$, we first choose from the set $\mathcal{S}$ t_1 elements that contains r_1 in the summation, which has $\binom{i}{t_1}$ combinations. And then choose from the remaining set t_2 elements that contains r_2 in the summation. Since the remaining set has at least $i - t_1$ elements that contains r_2 in the summation, the second step has at least $\binom{i-t_1}{t_2}$ combinations. The procedure goes on until choosing elements containing r_d is finished. Therefore, we have

$$\begin{aligned} c_{\boldsymbol{t}} &\geq \binom{i}{t_1}\binom{i-t_1}{t_2}\cdots\binom{i-t_1-\cdots-t_{d-1}}{t_d} \\ &= \frac{i}{(i - \sum_{k=1}^{d} t_k)! \cdot \prod_{k=1}^{d} t_k!}. \end{aligned} \quad (34)$$

The desire result is obtained from (32), (33), and (34). □

Corollary 1: For $d \geq 1, i \leq d$, and $r_1, \ldots, r_d \geq 0$, let $\mathcal{D} \triangleq \{1, \ldots, d\}$. The following inequality holds:

$$\sum_{\mathcal{I} \subset \mathcal{D}: |\mathcal{I}| = i} \log_2\left(1 + \sum_{j \in \mathcal{I}} r_j\right) \geq \binom{d-1}{i-1} \log_2\left(1 + \sum_{j=1}^{d} r_j\right).$$

[7]Here, r_{j+d} is equal to r_j for $j = 1, \ldots, d-1$.

Proof: Consider permutation $\pi_{\mathcal{D}}$ on the domain $\mathcal{D}$. From Lemma 3 and since $\pi_{\mathcal{D}}$ is a permutation

$$\begin{aligned} \sum_{\pi_{\mathcal{D}}} \sum_{k=1}^{d} \log_2\left(1 + \sum_{j=k}^{k+i-1} r_{\pi_j}\right) &\geq \sum_{\pi_{\mathcal{D}}} i \log_2\left(1 + \sum_{j=1}^{d} r_{\pi_j}\right) \\ &= d! i \log_2\left(1 + \sum_{j=1}^{d} r_j\right). \end{aligned} \quad (35)$$

Rewrite the LHS of (35) as follows:

$$\begin{aligned} &\sum_{\pi_{\mathcal{D}}} \sum_{k=1}^{d} \log_2\left(1 + \sum_{j=k}^{k+i-1} r_{\pi_j}\right) \\ &= \sum_{k=1}^{d} \sum_{\pi_{\mathcal{D}}} \log_2\left(1 + \sum_{j=k}^{k+i-1} r_{\pi_j}\right) \\ &= \sum_{k=1}^{d} \sum_{\pi_{\mathcal{D}}} \log_2\left(1 + \sum_{j=1}^{i} r_{\pi_j}\right) \\ &= d \sum_{\substack{\mathcal{I} \in \mathcal{D}: \\ |\mathcal{I}| = i}} \sum_{\substack{\pi_{\mathcal{D}}: \\ \{\pi_j\}_1^i = \mathcal{I}}} \log_2(1 + \sum_{j \in \mathcal{I}} r_j) \\ &= d \sum_{\substack{\mathcal{I} \in \mathcal{D}: \\ |\mathcal{I}| = i}} i!(d-i)! \log_2\left(1 + \sum_{j \in \mathcal{I}} r_j\right). \end{aligned} \quad (36)$$

Equations (35) and (36) conclude the proof. □

For $\mathcal{I} \subset \mathcal{D}$, since the derived random variables $X_{\bar{\mathcal{I}}}^{(\mathcal{I})}$ are independent of $X_{\mathcal{I}}^{(\mathcal{I})}$, $X_{\bar{\mathcal{I}}}^{(\mathcal{I})}$ contribute $\sum_{j \in \bar{\mathcal{I}}} \sigma_j^2$ power to the additive noise, where σ_j^2 is the jth diagonal entry of Σ. Since the channel gains are uniformly distributed around the unit circle and $X_{\bar{\mathcal{I}}}^{(\mathcal{I})}$ are Gaussian, we have

$$I\left(X_{\mathcal{I}}^{(\mathcal{I})}; Y \mid H_{\mathcal{D}}\right) \leq \log_2\left(1 + \frac{\sum_{k \in \mathcal{I}} \sigma_k^2}{1 + \sum_{j \in \bar{\mathcal{I}}} \sigma_j^2}\right).$$

Therefore, for $i \leq d$

$$\begin{aligned} &\sum_{\mathcal{I} \subset \mathcal{D}: |\mathcal{I}| = i} I\left(X_{\mathcal{I}}^{(\mathcal{I})}; Y \mid H_{\mathcal{D}}\right) \\ &\leq \binom{d}{i} \log_2\left(1 + \sum_{k=1}^{d} \sigma_k^2\right) - \sum_{\mathcal{I} \subset \mathcal{D}: |\mathcal{I}| = i} \log_2\left(1 + \sum_{j \in \bar{\mathcal{I}}} \sigma_j^2\right) \\ &\leq \left(\binom{d}{i} - \binom{d-1}{d-i-1}\right) \log_2\left(1 + \sum_{j=1}^{d} \sigma_j^2\right) \quad (37) \\ &\leq \binom{d-1}{i-1} \log_2(1 + P) \end{aligned}$$

where (37) is due to Corollary 1. Therefore, the achievable rate with the Gaussian codebook $\mathcal{N}_C(\mathbf{0}, \Sigma)$ is upper-bounded as follows:

$$\begin{aligned}\sum_{i=1}^{d}(1-\beta)^i\beta^{d-i}&\sum_{\mathcal{I}\subset\mathcal{D}:|\mathcal{I}|=i} I(X_{\mathcal{I}}^{(\mathcal{I})};Y\mid H_{\mathcal{D}})\\ &\leq \sum_{i=1}^{d}(1-\beta)^i\beta^{d-i}\binom{d-1}{i-1}\log_2(1+P)\\ &=(1-\beta)\log_2(1+P).\end{aligned}$$

The achievability part of Proposition 3 can be verified from Theorem 5. □

REFERENCES

[1] L. Tong, Q. Zhao, and S. Adireddy, "Sensor networks with mobile agents," in *Proc. 2003 Military Commun. Int. Symp.*, Boston, MA, Oct. 2003, pp. 688–693.

[2] W. Zhao and M. H. Ammar, "Message ferrying: Proactive routing in highly-partitioned wireless ad hoc networks," in *Proc. IEEE Workshop on Future Trends of Distributed Computing Systems*, San Juan, Puerto Rico, May 2003, pp. 308–314.

[3] R. C. Shah, S. Roy, S. Jain, and W. Brunette, "Data mules: Modeling a three-tier architecture for sparse sensor networks," in *Proc. 1st IEEE Int. Workshop on Sensor Network Protocols and Applications*, Anchorage, AK, May 2003, pp. 30–41.

[4] A. Kansal, M. Rahimi, D. Estrin, W. J. Kaiser, G. J. Pottie, and M. B. Srivastava, "Controlled mobility for sustainable wireless networks," in *Proc. IEEE Sensor and Ad Hoc Communications and Networks (SECON)*, Santa Clara, CA, Oct. 2004, pp. 1–6.

[5] P. Venkitasubramaniam, S. Adireddy, and L. Tong, "Sensor networks with mobile access: Optimal random access and coding," *IEEE J. Sel. Areas Commun. (Special Issue on Sensor Networks)*, vol. 22, no. 6, pp. 1058–1068, Aug. 2004.

[6] Z. Yang and L. Tong, "Cooperative sensor networks with misinformed sensors," in *Proc. 38th Annual Conf. Information Sciences and Systems*, Princeton, NJ, Mar. 2004, pp. 154–159.

[7] A. Scaglione and Y. Hong, "Opportunistic large arrays: Cooperative transmission in wireless multihop ad hoc networks to reach far distances," *IEEE Trans. Signal Process.*, vol. 51, no. 8, pp. 2082–2092, Aug. 2003.

[8] I. Maric and R. D. Yates, "Cooperative multihop broadcast for wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 22, no. 6, pp. 1080–1088, Aug. 2004.

[9] R. Dabora and S. Servetto, "Broadcast channels with cooperating receivers: A downlink for the sensor reachback problem," in *Proc. IEEE Int. Symp. Information Theory*, Chicago, IL, Jun./Jul. 2004, p. 176.

[10] M. Gastpar and M. Vetterli, "On the capacity of wireless networks: The relay case," in *Proc. IEEE INFOCOM'02*, vol. 3, New York, Jun. 2002, pp. 1577–1586.

[11] B. Schein and R. Gallager, "The Gaussian parallel relay network," in *Proc. IEEE Int. Symp. Information Theory*, Sorrento, Italy, Jun. 2000, p. 22.

[12] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity—Part I: System description," *IEEE Trans. Commun.*, vol. 51, no. 11, pp. 1927–1938, Nov. 2003.

[13] ——, "User cooperation diversity—Part II: Implementation aspects and performance analysis," *IEEE Trans. Commun.*, vol. 51, no. 11, pp. 1939–1948, Nov. 2003.

[14] S. Cui, A. J. Goldsmith, and A. Bahai, "Energy-efficiency of MIMO and cooperative MIMO techniques in sensor networks," *IEEE J. Sel. Areas Commun.*, vol. 22, no. 6, pp. 1089–1098, Aug. 2004.

[15] J. N. Laneman and G. W. Wornell, "Distributed space-time-coded protocols for exploiting cooperative diversity in wireless networks," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2415–2425, Oct. 2003.

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

[17] A. Hu and S. Servetto, "Optimal detection for a distributed transmission array," in *Proc. IEEE Int. Symp. Information Theory*, Yokohama, Japan, Jun./Jul. 2003, p. 200.

[18] S. Servetto, "Distributed signal processing algorithms for the sensor broadcast problem," in *Proc. 37th Annu. Conf. Information Sciences and Systems*, Baltimore, MD, Mar. 2003.

[19] M. Gastpar and M. Vetterli, "On the asymptotic capacity of Gaussian relay networks," in *Proc. IEEE Int. Symp. Information Theory*, Lausanne, Switzerland, Jun./Jul. 2002, p. 195.

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

[21] J. Proakis, *Digital Communications*, 3rd ed. New York: McGraw-Hill, 1995.

[22] R. Durrett, *Probability: Theory and Examples*, 2nd ed. Belmont, CA: Duxbury, 1995.