The value of feedback in decentralized detection

Tay, Wee Peng

2012

Tay, W. P. (2012). The value of feedback in decentralized detection. IEEE Transactions on Information Theory, 58(12), 7226-7239. doi:10.1109/TIT.2012.2211331

https://hdl.handle.net/10356/96115

https://doi.org/10.1109/TIT.2012.2211331

The Value of Feedback in Decentralized Detection

Wee Peng Tay, *Member, IEEE*

***Abstract*—We consider the decentralized binary hypothesis testing problem in networks with feedback, where some or all of the sensors have access to compressed summaries of other sensors' observations. We study certain two-message feedback architectures, in which every sensor sends two messages to a fusion center, with the second message based on full or partial knowledge of the first messages of the other sensors. We also study one-message feedback architectures, in which each sensor sends one message to a fusion center, with a group of sensors having full or partial knowledge of the messages from the sensors not in that group. Under either a Neyman-Pearson or a Bayesian formulation, we show that the asymptotically optimal (in the limit of a large number of sensors) detection performance (as quantified by error exponents) does not benefit from the feedback messages, if the fusion center remembers all sensor messages. However, feedback can improve the Bayesian detection performance in the one-message feedback architecture if the fusion center has limited memory; for that case, we determine the corresponding optimal error exponents.**

***Index Terms*—Decentralized detection, feedback, error exponent, sensor networks.**

I. Introduction

In the problem of decentralized detection, introduced by Tenney and Sandell [1], each one of several sensors makes an observation and sends a summary by first applying a quantization function to its observation and then communicating the result to a fusion center. The fusion center makes a final decision based on all of the sensor messages. The goal is to design the sensor quantization functions and the fusion rule so as to minimize a cost function, such as the probability of an incorrect final decision.

In this paper we consider sensor network architectures that are more complex than those in [1], and which involve feedback: some or all of the sensors have access to compressed summaries of other sensors' observations. We are interested in characterizing the performance under different architectures, and, in particular, to determine whether the presence of feedback can substantially enhance performance. Because an exact analysis is seemingly intractable, we focus on the asymptotic regime, involving a large number of sensors, and quantify performance in terms of error exponents. The numerical examples in [2], [3] show that feedback can improve the detection performance if the number of sensors is fixed. In the asymptotic regime however, the somewhat unexpected conclusion is that for most of the models considered in this paper, feedback does not improve performance in binary hypothesis testing.[1] The only exception we have found is Bayesian hypothesis testing in a "daisy-chain architecture" (cf. Section II) where the fusion center has limited memory. In this configuration, feedback can result in a better optimal error exponent.

A. Related Literature

The decentralized detection problem has been widely studied for various network architectures, including the above described "parallel" configuration of [1] (see [4]–[15]), tandem networks [16]–[19], and bounded height tree architectures [20]–[27]. For sensor observations not conditionally independent given the hypothesis, the problem of designing the quantization functions is known to be NP-hard [28]. For this reason, most of the literature assumes that the sensor observations are conditionally independent. Several works have considered the case of correlated observations, but under specific assumptions like observations having Gaussian distributions [29]–[33] or hierarchical Markovian models [34]. In this paper, we consider the case where observations are conditionally independent given the hypothesis, but the information available at each sensor may become correlated after feedback messages are transmitted to them.

Non-tree networks are harder to analyze because the different messages received by a sensor are not in general conditionally independent. While some structural properties of optimal decision rules are available (see, e.g., [35]), not much is known about the optimal performance. Networks with feedback face the same difficulty, and the relevant literature (discussed in the next paragraph) is limited.

A variety of feedback architectures, under a Bayesian formulation, have been studied in [2], [3]. These references show that it is person-by-person optimal for every sensor to use a likelihood ratio quantizer, with thresholds that depend on the feedback messages. However, because of the difficulty of optimizing these thresholds when the number of sensors becomes large, it is difficult to analytically compare the performance of networks with and without feedback. Numerical examples in [3] show that a system with feedback has lower probability of error, as expected. To better understand the asymptotics of the error probability, [36] studies the error probability decay rate under a Neyman-Pearson formulation for two different feedback architectures. For either case, it shows that if the fusion center also has access to the feedback messages, then feedback does not improve the optimal error exponent. References [37], [38] consider the Neyman-Pearson

Manuscript received July 17, 2011; revised March 31, 2012; accepted July 10, 2012. This work was supported by the MOE AcRF Tier 1 Grant RG25/10. Preliminary versions of parts of this paper were presented at the International Symposium on Wireless and Pervasive Computing, Feb 2011, and at the IEEE International Conference on Acoustics, Speech and Signal Processing, May 2011. The author is with the Nanyang Technological University, Singapore. E-mail: `wptay@ntu.edu.sg`

[1] Although feedback has been studied in various areas of information theory, including channel capacity, those results have no direct relationships with the topic of decentralized detection that we address in this paper.

problem in a daisy-chain architecture (see Figure 2), and obtain a similar result. However, the analogous questions under a Bayesian formulation were left open in [36]–[38].

B. Summary and Contributions

In this paper, we revisit some of the architectures studied in [36]–[38], and extend the available results. We also study certain feedback architectures that have not been studied before. In what follows, we describe briefly the architectures that we consider, and summarize our results.

1) We study a new **two-message sequential feedback** architecture. Sensors are indexed, and the second message of a sensor can take into account the first message of all sensors with lower indices. We show that under either the Neyman-Pearson or Bayesian formulation, feedback does not improve the error exponent.
2) We consider the **two-message full feedback** architecture studied in [36]. Here, each sensor gets to transmit two messages, and the second message can take into account the first messages of all sensors. We resolve an open problem for the Bayesian formulation, by showing that there is no performance gain over the non-feedback case. We also provide a variant of the result of [36] for the Neyman-Pearson case. Our model is somewhat more general than that in [36], because we do not restrict the sensors' raw observations and the sensor messages to be finite-valued. More crucially, we also remove the constraint in [36] that the feedback message alphabet can grow at most subexponentially with the number of sensors.
3) We consider the **one-message sequential feedback** architecture studied in [39], [40] (under the name of "full observation network topology"), where sensors are indexed, and each sensor knows the messages of all sensors with lower indices. Unlike [39], [40], which investigate "myopic" strategies where each sensor selfishly minimizes its local error probability, we show that if there is cooperation amongst sensors so that the last sensor makes the final decision for the whole network, there is no loss of asymptotic optimality if sensors other than the last ignore information from the other sensors, for both the Neyman-Pearson and the Bayesian formulation.
4) We consider the **daisy chain** or **one-message** architectures studied in [37], under which the sensors are divided into two groups, and sensors in the second group have full or partial knowledge of the messages sent by the first group. Reference [37] dealt with the Neyman-Pearson formulation. In this paper, we turn to the Bayesian formulation and resolve several questions that had been left open.
   a) In a **full feedback daisy chain**, sensors in the second group, as well as the fusion center, have access to all messages sent by sensors in the first group. Similar to the Neyman-Pearson case, we show that the Bayesian optimal error exponent is the same as for a parallel configuration with the same number of sensors; in particular, feedback offers no performance improvement.
   b) In a **restricted feedback daisy chain**, the second group of sensors, as well as the fusion center, have access to only a 1-bit summary of the messages sent by sensors in the first group. For the Neyman-Pearson formulation, [38] shows that feedback does not improve the error exponent. In contrast, for the Bayesian formulation, we show that in general, feeding this 1-bit summary to the second group of sensors can improve the detection performance. We provide sufficient conditions for feedback to result in no performance gain. Furthermore, we show that this architecture is strictly inferior to the full feedback daisy chain and the parallel configuration. We also provide a characterization of the optimal error exponent.

The study of feedback mechanisms in parallel configurations or daisy chain architectures provides insights into the performance of more complex networks in which groups of sensors may have access to the information at other sensors. The results in this paper serve as a first step to a better understanding of the performance of complex networks.

Feedback messages can complicate the design of optimal sensor quantization functions and fusion rules [3], and may improve the detection performance when the number of sensors is limited. However, the results in this paper suggest that for binary hypothesis testing, and in most message architectures, feedback does not significantly improve the detection performance when the number of sensors is large. Therefore, it is better to adopt simple sensor quantization functions and fusion rules and optimize other aspects of the network when designing a decentralized detection network.

The remainder of the paper is organized as follows. In Section II we define the model, formulate the problems that we will be studying, and provide some background material. In Section III, we study two-message feedback architectures (sequential and full feedback). In Section IV, we study one-message feedback architectures. We offer concluding remarks and discuss open problems in Section V. Some mathematical results that we use frequently are presented in the Appendix.

II. Problem Formulation

In this section, we describe the feedback architectures of interest, define our model, and present some preliminary results. We consider a decentralized binary detection problem involving n sensors and a fusion center. Sensor k observes a random variable X_k taking values in some measurable space $(\mathcal{X}, \mathcal{F})$, and is distributed according to a measure $\mathbb{P}_j$ under hypothesis H_j, for $j = 0, 1$. Under either hypothesis H_j, $j = 0, 1$, the random variables X_k are assumed to be independent and identically distributed. We use $\mathbb{E}_j$ to denote the expectation operator with respect to (w.r.t.) $\mathbb{P}_j$, and X_1^n to denote the vector $(X_1, \ldots, X_n)$. A similar notation, e.g., Y_1^n will be used for other vectors of random variables as well.

Let $\mathcal{T}$ be the set from which messages take their values. In most engineering applications, $\mathcal{T}$ is assumed to be a finite alphabet, although we do not require this restriction. This allows us to model the received messages at the fusion center over noisy channels. Furthermore, we use Γ to denote the set of allowed quantization functions, that is, functions $\gamma : \mathcal{X} \mapsto \mathcal{T}$, that can be used to map observations to messages. One possible choice is to let Γ consist of all measurable functions. Alternatively, for the problems considered in this paper, it is known that for $\mathcal{T}$ finite, there is no loss of optimality if we let Γ be the set of likelihood-ratio quantizers [2], [3], [35].

We consider two classes of feedback architectures: the two-message and one-message architectures.

A. Two-Message Feedback Architectures

In two-message feedback architectures (see Figure 1), each sensor k sends a message $Y_k = \gamma_k(X_k)$, with $\gamma_k \in \Gamma$, which is a "quantized" version of its observation X_k, to the fusion center.

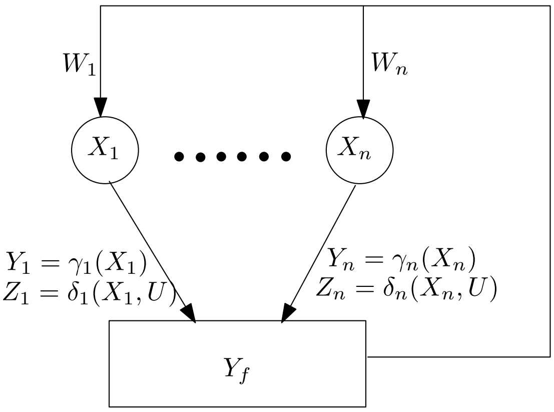

Fig. 1. A two-message architecture.

We assume that the sensors are indexed in the order that they send their messages to the fusion center. We consider three forms of feedback under the two-message architecture.

(a) **Sequential feedback.** Here, for $k = 2, \ldots, n$, the feedback message sent by the fusion center to sensor k is $W_k = (Y_1, \ldots, Y_{k-1})$, the vector of messages generated by the previous sensors.
(b) **Full feedback.** The feedback message sent by the fusion center to sensor k is the vector $W_k = (Y_1^{k-1}, Y_{k+1}^n)$ of messages generated by all of the other sensors.
(c) **Restricted feedback.** The feedback message sent by the fusion center to sensor k is a function $W_k = f_k(Y_1^{k-1}, Y_{k+1}^n)$ of the other sensors' first messages, whose alphabet does not increase with the number of sensors.

In all of the above scenarios, each sensor forms a new, second message $Z_k = \delta_k(X_k, W_k)$ based on the additional information W_k, and sends it to the fusion center.

For simplicity, we assume that Z_k takes values in the same alphabet $\mathcal{T}$ and, furthermore, that for any w, the function $\delta_k^w(\cdot) = \delta_k(\cdot, w)$ is constrained to belong to the same set Γ that applies to the first round. As alluded to earlier, when $\mathcal{T}$ is finite, it is known that there is no loss of optimality if we restrict to log-likelihood ratio quantizers of X_k, with thresholds that depend on the received messages.

Finally, the fusion center makes a decision $Y_f = \gamma_f(Y_1^n, Z_1^n)$. Here, we assume that the fusion center always remembers the first messages $Y_1, \ldots, Y_n$. The collection $(\gamma_f, \gamma_1, \ldots, \gamma_n, \delta_1, \ldots, \delta_n)$ is called a strategy. A sequence of strategies, one for each value of n, is called a strategy sequence. We wish to design strategy sequences that are asymptotically optimal (in the sense of error exponents), as n increases to infinity.

B. One-Message Feedback Architectures

In one-message architectures, every sensor sends a single message to an intermediate aggregator or the fusion center, but some of the sensors have access to the messages of some other sensors. Specifically, we consider a one-message sequential feedback architecture [39], [40], and a daisy chain architecture [37], [38]. As before, we let Γ be the set of allowed quantization functions.

(a) **One-message sequential feedback.** Here, sensor k has access to the messages $Y_1, \ldots, Y_{k-1}$ of all sensors with lower indices. Sensor k forms a message $Y_k = \gamma_k(X_k, Y_1^{k-1})$, and broadcasts it to all sensors with higher indices. The last sensor, n, makes a final decision and plays the role of a fusion center. We assume that for any Y_1^{k-1}, the mapping from X_k to Y_k belongs to Γ.
(b) **Daisy chain.** This architecture consists of two stages (see Figure 2) with the first stage involving m sensors and the second $n - m$. Each sensor k in the first stage sends a message $Y_k = \gamma_k(X_k)$ to an aggregator, with $\gamma_k \in \Gamma$. The aggregator forms a message U that is broadcast to all sensors in the second stage and to the fusion center. Each sensor l in the second stage forms a message $Z_l = \delta_l^U(X_l) = \delta_l(X_l, U)$, which depends on its own observation and the message U. Again, we assume that $\delta_i^u \in \Gamma$, for every possible value u of U. The fusion center makes a final decision using a fusion rule $Y_f = \gamma_f(U, Z_{m+1}, \ldots, Z_n)$. We can view the daisy chain as a parallel configuration, in which the fusion center feeds sensors $m+1, \ldots, n$ with a message based on information from sensors $1, \ldots, m$.

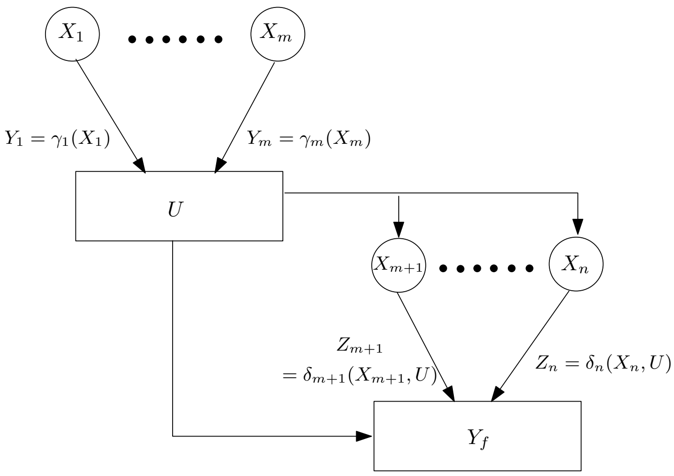

Fig. 2. The daisy chain architecture.

We consider two cases for how U is formed.

(i) **Full feedback daisy chain.** Here, we let $U = (Y_1, \ldots, Y_m)$, i.e., the second stage sensors and fusion center have the full information available at the first stage aggregator.

(ii) **Restricted feedback daisy chain.** Here, we let $U = \gamma_u(Y_1, \ldots, Y_m) \in \{0,1\}$. This architecture can be viewed as a parallel configuration in which the fusion center makes a preliminary decision based on the messages from the first m sensors, broadcasts the preliminary decision, and forgets (e.g., due to memory or security constraints) the messages sent by the first m sensors.

C. Assumptions and Preliminaries

In this section, we list the basic assumptions that we will be making throughout this paper, and note a useful consequence that will be used in our subsequent proofs. The first assumption we make results in no loss of generality (see [41]).

Assumption 1: The measures $\mathbb{P}_0$ and $\mathbb{P}_1$ are absolutely continuous w.r.t. each other.

Let $\mathbb{P}_i^X$ be the distribution of a random variable X under hypothesis H_i. Consider the Radon-Nikodym derivative $\mathrm{d}\mathbb{P}_i^X/\mathrm{d}\mathbb{P}_j^X$ of the measure $\mathbb{P}_i^X$ with respect to the measure $\mathbb{P}_j^X$. Informally, this is the likelihood ratio associated with an observation of X, and is a random variable whose value is determined by X; accordingly, its value should be denoted by a notation such as $\ell_{ij}^X(X)$, where ℓ_{ij}^X is a function from $\mathcal{X}$ into $[0,\infty)$ determined by the distributions of X under the two hypotheses. However, in order to avoid cluttered expressions, we will abuse notation and just write $\ell_{ij}(X)$. Furthermore, to simplify notation, we use $\ell_{ij}(X,Y)$ in place of $\ell_{ij}((X,Y))$, and similarly for random vectors of arbitrary length. We also use $\ell_{ij}(\gamma(X))$ to denote the Radon-Nikodym derivative of the random variable $Z = \gamma(X)$. Throughout the paper, we deal with various conditional distributions. Abusing notation as before, we let $\ell_{ij}(X|Y)$ be the Radon-Nikodym derivative of the conditional distribution of X given Y. Other notations like $\ell_{ij}(\gamma(X)|Y)$ will also be used.

In this paper, we are interested in the decay rates of the detection error probabilities. As such, we make extensive use of quantities like the Kullback-Leibler divergence $\mathbb{E}_i\left[\log \ell_{ij}(X_1)\right]$ in the subsequent discussions. Note however, that in some places it is more convenient to use $\mathbb{E}_i\left[\log \ell_{ji}(X_1)\right] = -\mathbb{E}_i\left[\log \ell_{ij}(X_1)\right]$, which is the negative of the Kullback-Leibler divergence.

The following assumption is made to simplify the exposition, and can often be relaxed. See [42] for a discussion.

Assumption 2: We have $\mathbb{E}_i\left[\log^2 \ell_{ji}(X_1)\right] < \infty$ for $i,j = 0,1$.[2]

Assumption 2 implies the following lemma, which follows from Proposition A.1 in Appendix A. This result was first proved in [42].

Lemma 1: There exists some finite constant a, such that for all $\gamma \in \Gamma$, and $i,j = 0,1$,

$$\mathbb{E}_i\left[\log^2 \ell_{ji}(\gamma(X_1))\right] \leq \mathbb{E}_i\left[\log^2 \ell_{ji}(X_1)\right] + 1 < a,$$
$$\mathbb{E}_i\left[|\log \ell_{ji}(\gamma(X_1))|\right] < a.$$

Under both the Neyman-Pearson and Bayesian formulations, the optimal fusion policy at the fusion center is a likelihood ratio test [43]. For $i,j \in \{0,1\}$, we consider the likelihood ratio of the information at the fusion center under H_i to that under H_j. Let the logarithm of this likelihood ratio be denoted by $\mathcal{L}_{ij}^{(n)}$. In the two-message architectures, $\mathcal{L}_{ij}^{(n)} = \log \ell_{ij}(Y_1^n, Z_1^n)$, and in the one-message architectures, we have $\mathcal{L}_{ij}^{(n)} = \log \ell_{ij}(Y_1^n)$ for the sequential feedback configuration, and $\mathcal{L}_{ij}^{(n)} = \log \ell_{ij}(U, Z_1^n)$ for the daisy chain network.

For the convenience of the reader, we end this section by summarizing some frequently encountered notations in the following table.

Notation	Definition
$\ell_{ij}(X)$	likelihood ratio of distribution of X under H_i to that under H_j
$\ell_{ij}(X\|Y)$	likelihood ratio of the conditional distribution of X given Y, under H_i to that under H_j
$\mathcal{L}_{ij}^{(n)}$	log likelihood ratio of the total information at the fusion center under H_i to that under H_j, when there are n sensors
$\psi_n(s)$	the log moment generating function of $\mathcal{L}_{10}^{(n)}$ under H_0, i.e., $\log \mathbb{E}_0[\exp(s\mathcal{L}_{10}^{(n)})]$
$\varphi_\xi(s)$	the log moment generating function $\log \mathbb{E}_0\left[(\ell_{10}(\xi(X_1))^s\right]$, where ξ is a measurable function.
$g_{2p}^*, g_{sf}^*, g_{rf}^*, g_f^*$	optimal Neyman-Pearson error exponent for the two-message parallel, and sequential, restricted, and full feedback architectures respectively
$\mathcal{E}_{2p}^*, \mathcal{E}_{sf}^*, \mathcal{E}_{rf}^*, \mathcal{E}_f^*$	optimal Bayesian error exponent for the two-message parallel, and sequential, restricted, and full feedback architectures respectively
$\mathcal{E}_{dc}^*, \mathcal{E}_t^*$	optimal Bayesian error exponent for the one-message daisy chain and tree architectures respectively

TABLE I
SUMMARY OF FREQUENTLY USED NOTATIONS.

III. Two-Message Architectures

In this section, we study the Neyman-Pearson and Bayesian formulations of the decentralized detection problem in two-message architectures. The log-likelihood ratio at the fusion center is given by

$$\begin{aligned}
\mathcal{L}_{10}^{(n)} &= \log \ell_{10}(Y_1^n, Z_1^n) \\
&= \sum_{k=1}^n \log \ell_{10}(Y_k) + \log \ell_{10}(Z_1^n \mid Y_1^n) \\
&= \sum_{k=1}^n \log \ell_{10}(Y_k) + \sum_{k=1}^n \log \ell_{10}(Z_k \mid Y_1^n) \\
&= \sum_{k=1}^n \log \ell_{10}(Y_k) + \sum_{k=1}^n \log \ell_{10}(Z_k \mid Y_k, W_k).
\end{aligned}$$

The third equality above holds because, under either hypothesis, and given Y_1^n, the random variables Z_k are functions of

[2] For the Neyman-Pearson formulation, we will only require that this assumption and Lemma 1 hold for $i = 0, j = 1$.

the respective X_k; thus, the Z_k are conditionally independent, given Y_1^n. The last equality holds because Z_k depends on Y_1^n through Y_k and W_k.

To simplify notation, we define, for every possible value w of W_k, a random variable $\mathcal{L}_{10}^k(w)$, according to

$$\begin{aligned}\mathcal{L}_{10}^k(w) &= \log \ell_{10}(Y_k) + \log \ell_{10}(Z_k \mid Y_k, W_k = w)\\ &= \log \ell_{10}(\gamma_k(X_k), \delta_k^w(X_k)).\end{aligned}$$

Note that $\mathcal{L}_{10}^k(w)$ is a random variable which is a function of a non-random argument w and the random variable X_k. Note also that

$$\mathcal{L}_{10}^{(n)} = \sum_{k=1}^{n} \mathcal{L}_{10}^k(W_k).$$

A. Neyman-Pearson Formulation

Let $\alpha \in (0,1)$ be a given constant. A strategy is *admissible* if its Type I error probability satisfies $\mathbb{P}_0(Y_f = 1) < \alpha$. Let $\beta_n^* = \inf \mathbb{P}_1(Y_f = 0)$, where the infimum is taken over all admissible strategies for the n-sensor problem. Our objective is to characterize the optimal error exponent

$$\limsup_{n\to\infty} \frac{1}{n} \log \beta_n^*, \tag{1}$$

under different feedback architectures. In this paper, we define the error exponent to be the worst case limiting bound for the error decay rate. One can also define the error exponent with an infimum limit in (1) in place of the supremum limit. In general, these two error exponents are not equal [44]. However, for a parallel configuration without feedback, these two error exponents coincide and the limit $\lim_{n\to\infty}(1/n)\log\beta_n^*$ exists [42]. It will be clear from our subsequent proofs that the limit also exists for the feedback architectures that we consider in this paper.

Let g_{2p}^* be the optimal error exponent for the two-message parallel configuration, in which there is no feedback from the fusion center, i.e., when each sensor k sends two messages, $(\gamma_k(X_k), \delta_k(X_k))$, to the fusion center. From [42], the optimal error exponent is

$$g_{2p}^* = \inf_{(\gamma,\delta)\in\Gamma^2} \mathbb{E}_0\left[\log \ell_{10}(\gamma(X_1), \delta(X_1))\right].$$

Let g_{sf}^*, g_{rf}^*, and g_f^* be the optimal error exponents for the sequential, restricted, and full feedback architectures respectively. Since the sensors can ignore some or all of the feedback messages from the fusion center, we have

$$g_f^* \le g_{sf}^* \le g_{2p}^*, \tag{2}$$

$$g_f^* \le g_{rf}^* \le g_{2p}^*, \tag{3}$$

(Note that error exponents are nonpositive and that smaller error exponents correspond to better performance.)

We will show that under appropriate but mild assumptions, the inequalities in (2) and (3) are equalities. Hence, from an asymptotic viewpoint, feedback results in no gain in detection performance. From [42], this implies that there is no loss in optimality if the sensors ignore the feedback messages from the fusion center and are constrained to using the same quantization function. We first show a useful result that underlies a key step in our proofs.

Lemma 2: Consider a sequence of strategies, indexed by n, the number of sensors. Let β_n be the associated Type II error probabilities. Suppose that for every strategy in the sequence, the Type I error probability $\mathbb{P}_0(Y_f = 1) \le \alpha$, where $\alpha \in (0,1)$. If there exists a nonnegative constant R such that

$$\limsup_{n\to\infty} \mathbb{P}_0(\mathcal{L}_{10}^{(n)} < -nR) < 1 - \alpha,$$

then

$$\liminf_{n\to\infty} \frac{1}{n} \log \beta_n \ge -R.$$

Proof: We have

$$\begin{aligned}\beta_n &= \mathbb{P}_1(Y_f = 0)\\ &= \mathbb{E}_0\left[\exp(\mathcal{L}_{10}^{(n)})\mathbf{1}_{\{Y_f=0\}}\right]\\ &\ge \mathbb{E}_0\left[\exp(\mathcal{L}_{10}^{(n)})\mathbf{1}_{\{Y_f=0,\mathcal{L}_{10}^{(n)}\ge -nR\}}\right]\\ &\ge e^{-nR}\mathbb{P}_0(Y_f = 0, \mathcal{L}_{10}^{(n)} \ge -nR).\end{aligned}$$

Therefore,

$$\mathbb{P}_0(Y_f = 0, \mathcal{L}_{10}^{(n)} \ge -nR) \le \beta_n e^{nR}.$$

This upper bound yields

$$\begin{aligned}1 - \alpha &\le \mathbb{P}_0(Y_f = 0)\\ &= \mathbb{P}_0(Y_f = 0, \mathcal{L}_{10}^{(n)} \ge -nR)\\ &\quad + \mathbb{P}_0(Y_f = 0, \mathcal{L}_{10}^{(n)} < -nR)\\ &\le \beta_n e^{nR} + \mathbb{P}_0(\mathcal{L}_{10}^{(n)} < -nR),\end{aligned}$$

and we have

$$\frac{1}{n}\log\beta_n + R \ge \frac{1}{n}\log(1 - \alpha - \mathbb{P}_0(\mathcal{L}_{10}^{(n)} < -nR)).$$

The lemma follows by taking the limit as $n \to \infty$, and the proof is complete. ∎

B. Neyman-Pearson Formulation — Sequential Feedback

For the case of sequential feedback, the proof that feedback yields no performance improvement is relatively simple. The core of the proof is an inequality on the (conditional) expectation of the log-likelihood ratio at the fusion center. We use this inequality together with a variance bound to obtain a bound on the tail probabilities associated with the log-likelihood ratio, and finally use Lemma 2.

Theorem 1: Suppose that Assumptions 1-2 hold. Then, the optimal error exponent for the sequential feedback architecture is $g_{sf}^* = g_{2p}^*$, i.e., there is no loss in optimality if the sensors ignore the feedback messages from the fusion center and are constrained to using the same quantization function.

Proof: From (2), we have $g_{sf}^* \le g_{2p}^*$. To show the reverse inequality, we first bound $\mathbb{E}_0[\mathcal{L}_{10}^k(W_k) \mid W_k]$ from below by g_{2p}^*. We have, for any w,

$$\begin{aligned}&\mathbb{E}_0\left[\mathcal{L}_{10}^k(W_k) \mid W_k = w\right]\\ &= \mathbb{E}_0\left[\log \ell_{10}(\gamma_k(X_k), \delta_k^w(X_k)) \mid W_k = w\right]\\ &\ge \inf_{(\gamma,\delta)\in\Gamma^2} \mathbb{E}_0\left[\log \ell_{10}(\gamma(X_1), \delta(X_1))\right]\\ &= g_{2p}^*.\end{aligned} \tag{4}$$

In particular, $\mathbb{E}_0\left[\mathcal{L}_{10}^k(W_k)\right] \geq g_{2p}^*$ and $\mathbb{E}_0\left[\mathcal{L}_{10}^{(n)}\right] \geq n g_{2p}^*$.

We next obtain a suitable variance bound. Let $Q_k = \mathcal{L}_{10}^k(W_k) - \mathbb{E}_0[\mathcal{L}_{10}^k(W_k) \mid W_k]$. From Lemma 1, there exists some constant $a > 0$ such that

$$\begin{aligned}\mathrm{var}_0(Q_k) &\leq \mathbb{E}_0\left[\mathbb{E}_0\left[\left(\mathcal{L}_{10}^k(W_k)\right)^2 \mid W_k\right]\right]\\ &\leq a. \end{aligned} \tag{5}$$

Recall that $W_k = Y_1^{k-1}$. We have, for $m < k$,

$$\begin{aligned}&\mathbb{E}_0\left[Q_m \cdot Q_k\right]\\ &= \mathbb{E}_0\left[Q_m \mathbb{E}_0\left[Q_k \mid W_k\right]\right]\\ &= 0. \end{aligned} \tag{6}$$

Let $\epsilon > 0$. Inequality (4), together with the bounds (5) and (6), and Chebyshev's inequality, yield

$$\begin{aligned}\mathbb{P}_0\left(\mathcal{L}_{10}^{(n)} < n(1+\epsilon)g_{2p}^*\right) &\leq \mathbb{P}_0\left(\sum_{k=1}^n Q_k < n\epsilon g_{2p}^*\right)\\ &\leq \frac{a}{n\epsilon^2 (g_{2p}^*)^2}. \end{aligned}$$

Letting $n \to \infty$, we get

$$\lim_{n\to\infty} \mathbb{P}_0\left(\mathcal{L}_{10}^{(n)} < n(1+\epsilon)g_{2p}^*\right) = 0 < 1-\alpha.$$

Therefore, applying Lemma 2, we have $g_{sf}^* \geq (1+\epsilon)g_{2p}^*$. Since ϵ was chosen arbitrarily, we obtain $g_{sf}^* \geq g_{2p}^*$, and the proof is complete. ■

C. Neyman-Pearson Formulation — Full Feedback

Next, we consider the full feedback architecture. The same architecture has been studied in [36], using the method of types, and under a more restrictive set of assumptions. In the following, we show a result similar to the one in [36], i.e., that there is no gain from the feedback messages asymptotically. For a comparison, we note that [36] involved a constraint that feedback messages take values in an alphabet that grows at most subexponentially. This constraint excludes the full feedback case, in which the feeback messages W_k take values in an exponentially growing alphabet.

The following result subsumes, in some sense Theorem 1; indeed, if full feedback cannot improve performance, then sequential feedback cannot either. On the other hand, for this more general result we will need a stronger assumption. In Theorem 1, we used the property that the "innovations" $\mathcal{L}_{10}^m(W_m) - \mathbb{E}_0[\mathcal{L}_{10}^m(W_m) \mid W_m]$ were uncorrelated, which allowed us to use Chebyshev's inequality. Such a property is no longer true in the full feedback case. Instead, we impose an exponential tail bound on the original log-likelihood ratios; equivalently, we make a finiteness assumption on the log moment generating function of the original log-likelihood ratios about a neighborhood of the origin, which is standard in the theory of large deviations [45]. We then proceed to derive related bounds that refer to the log-likelihood ratios associated with various messages. This step is somewhat tedious but unsurprising.

Let $\varphi_\xi(s) = \log \mathbb{E}_0\left[(\ell_{10}(\xi(X_1))^s\right]$ be the log moment generating function of the log-likelihood ratio of the distribution of $\xi(X_1)$ under H_1 versus that under H_0, where ξ is a measurable function. If $\xi = \mathrm{Id}$ the identity function, we have $\varphi_{\mathrm{Id}}(s) = \log \mathbb{E}_0[(\ell_{10}(X_1))^s]$, which is the log moment generating function of the log-likelihood ratio $\log \ell_{10}(X_1))$. We make the following assumption about $\varphi_{\mathrm{Id}}(s)$.

Assumption 3: There exists some $\bar{s} < 0$ such that $\varphi_{\mathrm{Id}}(\bar{s}) < \infty$.

Since $\varphi_{\mathrm{Id}}(\cdot)$ is nonincreasing on $[\bar{s}, 0]$ (cf. Lemma 2.2.5 of [45]), Assumption 3 implies that $\varphi_{\mathrm{Id}}(s) < \infty$ for all $s \in [\bar{s}, 0]$. Furthermore, the second moment of $\log \ell_{10}(X_1)$ under H_0 exists and is finite. Therefore, Assumption 3 implies Assumption 2 for $i = 0$ and $j = 1$.

Now consider a pair $\xi = (\gamma, \delta) \in \Gamma^2$ of quantization functions. We have $\varphi_\xi(s)$ is the log moment generating function of the log-likelihood ratio of the distribution of $\xi(X_1) = (\gamma(X_1), \delta(X_1))$ under H_1 versus that under H_0. Suppose that a strategy sequence has been fixed. Based on Assumption 3, we will show some properties of φ_ξ and ψ_n. (Recall that $\psi_n(s) = \log \mathbb{E}_0[\exp(s\mathcal{L}_{10}^{(n)})]$ is the log moment generating function of $\mathcal{L}_{10}^{(n)}$.) We will then use these properties to obtain tail bounds on $\mathcal{L}_{10}^{(n)}$, which will play the same role as the Chebyshev bound in the proof of Theorem 1. [3] A proof of the following lemma is provided in Appendix B.

Lemma 3: Suppose Assumption 1-3 holds, and let $\bar{s}$ be as in Assumption 3.

(i) There exists a positive constant c such that for all $s \in [\bar{s}/2, 0]$, and for all $\xi \in \Gamma^2$, we have $0 \leq \varphi_\xi''(s) \leq c$.

(ii) Let $\xi^* \in \Gamma^2$ be such that $\varphi_{\xi^*}'(0) \leq g_{2p}^* + \epsilon$, where ϵ is a small positive constant so that $h = \sqrt{\epsilon/(2c)} < \min\{|\bar{s}|/2, 1/4\}$. Then, for all $s \in [-h, 0]$, and for all $\xi \in \Gamma^2$, we have $\varphi_\xi(s) \leq \varphi_{\xi^*}(s) + \epsilon/2$.

(iii) For all $n \geq 1$ and $s \in [-h, 0]$, we have $\psi_n(s)/n \leq \varphi_{\xi^*}(s) + \epsilon$.

Finally, we show that for both the full and restricted feedback architectures, feedback does not improve the optimal error exponent.

Theorem 2: Suppose that Assumptions 1-3 hold. Then, in both the full and restricted feedback architectures, there is no loss in optimality if sensors ignore the feedback messages from the fusion center, i.e., $g_f^* = g_{rf}^* = g_{2p}^*$.

Proof: From (3), it suffices to show $g_f^* \geq g_{2p}^*$. Choose a sufficiently small $\epsilon > 0$. Let ξ^* and h be chosen as in Lemma 3(ii), and let $t_\epsilon = -(\varphi_{\xi^*}(-h) + \epsilon)/h$. From the Chernoff bound and Lemma 3(iii), we have

$$\begin{aligned}&\limsup_{n\to\infty} \frac{1}{n}\log \mathbb{P}_0\left(\mathcal{L}_{10}^{(n)} < n(t_\epsilon - \epsilon)\right)\\ &\leq \varphi_{\xi^*}(-h) + \epsilon + h(t_\epsilon - \epsilon)\\ &= -h\epsilon < 0. \end{aligned}$$

Applying Lemma 2, we have $g_f^* \geq t_\epsilon - \epsilon$. The Taylor series expansion of φ_{ξ^*} yields

$$t_\epsilon = \varphi_{\xi^*}'(0) - \varphi_{\xi^*}''(\theta)\frac{1}{2}\sqrt{\frac{\epsilon}{2c}} - \sqrt{2c\epsilon},$$

[3]Throughout the paper, we use $f'(s)$ and $f''(s)$ to denote the first and second derivatives of f w.r.t. s.

where $\theta \in [-h, 0]$, and c is the same constant as in Lemma 3(i). Since $0 \le \varphi''_{\xi^*}(\theta) \le c$, $t_\epsilon \to g^*_{2p}$ as ϵ decreases to 0. Letting $\epsilon \to 0$, we obtain the theorem. ■

D. Bayesian Formulation

In this section, we show that feedback does not improve the optimal error exponent for the binary Bayesian decentralized detection problem in the sequential, full, and restricted feedback architectures. Let the prior probability of hypothesis H_j be $\pi_j > 0$, $j = 0, 1$. Given a strategy, the probability of error at the fusion center is $P_e(n) = \pi_0 \mathbb{P}_0(Y_f = 1) + \pi_1 \mathbb{P}_1(Y_f = 0)$. Let $P^*_e(n)$ be the minimum probability of error, over all strategies, for the n-sensor problem. We seek to characterize the optimal error exponent

$$\limsup_{n\to\infty} \frac{1}{n} \log P^*_e(n).$$

From [42], the optimal error exponent for the parallel configuration without any feedback is given by

$$\mathcal{E}^*_{2p} = \inf_{(\gamma,\delta)\in\Gamma^2} \min_{s\in[0,1]} \log \mathbb{E}_0 \left[(\ell_{10}(\gamma(X_1), \delta(X_1)))^s \right]. \tag{7}$$

Similar to the Neyman-Pearson formulation, we let $\mathcal{E}^*_{sf}$, $\mathcal{E}^*_{rf}$ and $\mathcal{E}^*_f$ denote the optimal error exponents for the sequential, restricted, and full feedback architectures respectively. Note that the counterparts of inequalities (2) and (3) also hold for the Bayesian error exponents. Therefore, to show that feedback does not improve the asymptotic performance, it suffices to show a lower bound for the full feedback architecture. Recall that $\psi_n(s) = \log \mathbb{E}_0 \left[\exp(s\mathcal{L}^{(n)}_{10}) \right]$ is the log moment generating function of $\mathcal{L}^{(n)}_{10}$. The following lemma, whose proof is in Appendix C, provides uniform bounds for ψ_n and its derivatives, over all strategies.

Lemma 4: Suppose that Assumptions 1 and 2 hold.

(i) For all $s \in [0, 1]$, we have $\mathbb{E}_0 [\log \ell_{10}(X_1)] \le \psi'_n(s)/n \le \mathbb{E}_1 [\log \ell_{10}(X_1)]$.

(ii) For any bounded sequence (t_n) and for any given strategy such that there exists $s_n \in (0, 1)$ with $\psi'_n(s_n) = t_n$ for each n,[4] we have $\psi''_n(s_n) \le nC$, where C is a constant independent of the strategy.

(iii) For all $s \in [0, 1]$, we have $\psi_n(s) \ge n\mathcal{E}^*_{2p}$.

The following result shows that feedback does not improve Bayesian detection performance in the full feedback architecture.

Theorem 3: Suppose that Assumptions 1 and 2 hold. Then $\mathcal{E}^*_f = \mathcal{E}^*_{2p}$, i.e., there is no loss in optimality if sensors are constrained to using the same quantization function that ignores the feedback messages from the fusion center.

Proof: It is clear that $\mathcal{E}^*_f \le \mathcal{E}^*_{2p}$. To show the reverse bound, we make use of Proposition A.2. Let the conditional probability of error under H_j be $P_{n,j}$ for $j = 0, 1$. Let $s^*_n = \arg\min_{s\in(0,1)} \psi_n(s)$ so that $\psi'_n(s^*_n) = 0$. From Proposition A.2, we have

$$\begin{aligned}\max_{j=0,1} P_{n,j} &\ge \frac{1}{4} \exp\left(\psi_n(s^*_n) - \sqrt{2\psi''_n(s^*_n)} \right) \\ &\ge \exp(\psi_n(s^*_n) - C\sqrt{n}) \\ &\ge \exp(n\mathcal{E}^*_{2p} - C\sqrt{n})\end{aligned}$$

where C is some constant. The penultimate inequality follows from Lemma 4(ii), and the last inequality from Lemma 4(iii). Letting $n \to \infty$, we have

$$\begin{aligned}\liminf_{n\to\infty} \frac{1}{n} \log P_e(n) &= \liminf_{n\to\infty} \frac{1}{n} \log \max_{j=0,1} P_{n,j} \\ &\ge \mathcal{E}^*_{2p}.\end{aligned}$$

This implies that $\mathcal{E}^*_f \ge \mathcal{E}^*_{2p}$, and the proof is complete. ■

Since the sequential and restricted feedback configurations can perform no better than the full feedback architecture, and no worse than the parallel configuration, we have the following result.

Theorem 4: Suppose that Assumptions 1 and 2 hold. Then $\mathcal{E}^*_{sf} = \mathcal{E}^*_{rf} = \mathcal{E}^*_{2p}$. Moreover, there is no loss in optimality if sensors are constrained to using the same quantization function, which ignore the feedback messages from the fusion center.

IV. One-Message Architectures

In this section, we consider the one-message architecture. We study both the Neyman-Pearson and Bayesian formulations for the binary hypothesis testing problem. Similar to the two-message architecture, feedback in general does not improve the asymptotic detection performance, except for the case of Bayesian detection with restricted feedback in the daisy chain architecture. In the case where there is no feedback [42], the optimal Neyman-Pearson error exponent is

$$g^*_{1p} = \inf_{\gamma\in\Gamma} \mathbb{E}_0 \left[\log \ell_{10}(\gamma(X_1)) \right],$$

while the optimal Bayesian error exponent is

$$\mathcal{E}^*_{1p} = \inf_{\gamma\in\Gamma} \min_{s\in[0,1]} \log \mathbb{E}_0 \left[(\ell_{10}(\gamma(X_1)))^s \right].$$

A. Full Information at Fusion Center

We consider the case where the fusion center has access to all sensor messages. This is the case for the sequential feedback architecture in which the fusion center is the last sensor. The same applies for the full feedback daisy chain architecture. By ignoring all feedback messages except at the fusion center, these architectures are equivalent to the parallel configuration with the same number of sensors. Therefore, the optimal error exponents under both the Neyman-Pearson and Bayesian formulations are at least as negative as those for the parallel configuration. The proof of the reverse direction involves the same steps as in the proofs for the two-message architectures in Section III. Specifically, the proof for the one-message sequential feedback architecture is similar to that of Theorem 1, with suitable modifications (remove all references to the first messages γ_k and replace Y_k by Z_k). The proof for the daisy-chain architecture corresponds to that of Theorems

[4]Note that the sequence (s_n) depends on the strategy used.

2 and 3. The result for the daisy-chain architecture under the Neyman-Pearson formulation is also provided in [37]. The above discussion is summarized in the following result, whose proof is omitted.

Theorem 5: Suppose that Assumptions 1 and 2 hold.

1) Under either the Neyman-Pearson or Bayesian formulation, the optimal error exponents for the one-message sequential feedback are the same as that of the parallel configuration under either corresponding formulations.
2) Under the Bayesian formulation, the optimal error exponent for the full feedback daisy chain is the same as that of the parallel configuration. In addition, if Assumption 3 holds, the Neyman-Pearson error exponent is the same as that of the parallel configuration.

B. Restricted Feedback Daisy Chain

In this section, we consider the restricted feedback daisy chain (RFDC) architecture. References [37], [38] have shown that under the Neyman-Pearson formulation, feedback again does not improve the optimal error exponent. In this section, we consider the Bayesian formulation, and show that unlike the Neyman-Pearson formulation, feedback may improve the detection performance. We provide a characterization of the optimal error exponent in this case.

Recall that m is the number of sensors in the first stage of the RFDC architecture. We assume that $\lim_{n\to\infty} m/n = r \in (0,1)$, otherwise the architecture is equivalent to a parallel configuration. Let $\mathcal{E}_{dc}^*$ be the optimal error exponent. For $\gamma \in \Gamma$, and $j = 0, 1$, let the Fenchel-Legendre transform of the log moment generating functions be

$$\Lambda_j^*(\gamma, t) = \sup_{s\in\mathbb{R}} \left\{ st - \log \mathbb{E}_j \left[e^{s \log \ell_{10}(\gamma(X_1))} \right] \right\}.$$

These are also known as rate functions [45] for the log likelihood ratio $\log \ell_{10}(\gamma(X_1))$. For $i, j \in \{0, 1\}$, and for any given sequence of strategies for the first m sensors, let the rate of decay of the conditional probabilities be

$$e_{ij} = -\liminf_{n\to\infty} \frac{1}{m} \log \mathbb{P}_i(U = j).$$

We collect the decay rates into a vector

$$\vec{e} = [e_{01}, e_{10}, e_{00}, e_{11}]. \tag{8}$$

Suppose that the quantization functions for the first stage sensors have been fixed. We characterize the optimal error exponent of the second stage in terms of $\vec{e}$ in the following lemma. The proof can be found in Appendix D.

Lemma 5: Suppose Assumptions 1 and 2 hold. Suppose that the quantization functions for sensors $1, \ldots, m$ in a RFDC have been fixed. Then, we have

$$\lim_{n\to\infty} \frac{1}{n} \log P_e(n) = -h(\vec{e})$$

where $\vec{e}$ is as defined in (8),

$$\begin{aligned} &h(\vec{e}) \\ &= \min \Big\{ (1-r) \sup_{\delta^0\in\Gamma} \Lambda_0^* \left(\delta^0, \frac{r}{1-r}(e_{10} - e_{00}) \right) + re_{00}, \\ &(1-r) \sup_{\delta^1\in\Gamma} \Lambda_1^* \left(\delta^1, -\frac{r}{1-r}(e_{01} - e_{11}) \right) + re_{11} \Big\}, \end{aligned} \tag{9}$$

and $P_e(n)$ is the optimal probability of error under the given quantization functions for sensors $1, \ldots, m$.

Reference [37] shows that if $e_{01} > 0$ and $e_{10} > 0$, then there is no loss in optimality if sensors within each stage are constrained to using the same quantization function. In the following, we show that it is optimal to require that $e_{01} > 0$ and $e_{10} > 0$. We also provide a characterization of the optimal error exponent.

Theorem 6: Suppose that Assumptions 1 and 2 hold. Then, the following statements hold for a RFDC architecture.

(i) There is no loss in optimality if e_{01} and e_{10} are constrained to be strictly positive.
(ii) There is no loss in optimality if sensors in the first stage are constrained to using the same quantization function.
(iii) There is no loss in optimality if sensors in the second stage are constrained to using the same quantization function (which may depend on the feedback message).
(iv) The optimal error exponent for the RFDC is

$$\begin{aligned} \mathcal{E}_{dc}^* = -(1-r) \sup_{\substack{\gamma,\delta^0,\delta^1\in\Gamma \\ t\in\mathbb{R}}} \min \Big\{ &\Lambda_0^* \left(\delta^0, \frac{r}{1-r} \Lambda_1^*(\gamma, t) \right), \\ &\Lambda_1^* \left(\delta^1, -\frac{r}{1-r} \Lambda_0^*(\gamma, t) \right) \Big\}. \end{aligned} \tag{10}$$

Proof: We first show claim (i). Note that only one of e_{01} and e_{00} can be strictly positive. The same applies to e_{10} and e_{11}. If $e_{01} > 0$ and $e_{10} > 0$, we have $e_{00} = e_{11} = 0$, and (9) yields

$$\begin{aligned} h(\vec{e}) &= (1-r) \min \Big\{ \sup_{\delta\in\Gamma} \Lambda_0^* \left(\delta, \frac{r}{1-r} e_{10} \right), \\ &\qquad \sup_{\delta\in\Gamma} \Lambda_1^* \left(\delta, -\frac{r}{1-r} e_{01} \right) \Big\} \\ &\geq (1-r) \min \Big\{ \sup_{\delta\in\Gamma} \Lambda_0^*(\delta, 0), \sup_{\delta\in\Gamma} \Lambda_1^*(\delta, 0) \Big\} \\ &= (1-r) \sup_{\delta\in\Gamma} \Lambda_1^*(\delta, 0). \end{aligned}$$

On the other hand, if $e_{00} > 0$ and $e_{10} > 0$. Then, $e_{01} = e_{11} = 0$, and from (9), we have $h(\vec{e}) \leq (1-r)\sup_{\delta\in\Gamma} \Lambda_1^*(\delta, 0)$. The same argument applies for the case where $e_{01} > 0$ and $e_{11} > 0$, and the case where all the decay rates are zero. Therefore, there is no loss in optimality if e_{01} and e_{10} are constrained to be strictly positive.

Claims (ii) and (iii) follow from either an application of Cramèr's Theorem (cf. [46]) and (9), or from [37].

Finally, we prove claim (iv). Since there is no loss in optimality if all first stage sensors are restricted to some same quantization function $\gamma \in \Gamma$, the first stage Type I and II error decay rates are $e_{01} = \Lambda_0^*(\gamma, t)$ and $e_{10} = \Lambda_1^*(\gamma, t)$ respectively, for some t (cf. [27]). Applying Lemma 5, and optimizing over γ and t, we have shown that the optimal error exponent is lower bounded by the right hand side of (10). This bound is achievable, hence the claim follows. The proof is now complete. ∎

Let $\mathcal{E}_t^*$ be the optimal error exponent of the daisy-chain if the second stage sensors ignore the feedback message. This is equivalent to a tree architecture with a height of two [27]. Using the same arguments as above, it can be shown that

$$\mathcal{E}_t^* = -(1-r) \sup_{\substack{\gamma,\delta\in\Gamma \\ t\in\mathbb{R}}} \min\left\{ \Lambda_0^*\left(\delta, \frac{r}{1-r}\Lambda_1^*(\gamma,t)\right), \Lambda_1^*\left(\delta, -\frac{r}{1-r}\Lambda_0^*(\gamma,t)\right)\right\}. \quad (11)$$

Comparing (10) and (11), we have $\mathcal{E}_{dc}^* \leq \mathcal{E}_t^*$, i.e., the optimal error exponent for the RFDC is in general better than the tree configuration where feedback is absent. In the following, we provide a sufficient condition for no loss in performance when feedback is ignored, i.e., $\mathcal{E}_{dc}^* = \mathcal{E}_t^*$. We also provide a numerical example in which $\mathcal{E}_{dc}^* < \mathcal{E}_t^*$, i.e., feedback can strictly improve the asymptotic performance in some cases.

Proposition 1: Suppose that there exists $\delta \in \Gamma$ such that

$$\mathcal{E}_t^* = -(1-r) \sup_{\substack{\gamma\in\Gamma \\ t\in\mathbb{R}}} \min\left\{ \Lambda_0^*\left(\delta, \frac{r}{1-r}\Lambda_1^*(\gamma,t)\right), \Lambda_1^*\left(\delta, -\frac{r}{1-r}\Lambda_0^*(\gamma,t)\right)\right\}, \quad (12)$$

and $\Lambda_1^*(\delta,t) = \Lambda_0^*(\delta,-t)$ for all t. Then,

$$\mathcal{E}_{dc}^* = \mathcal{E}_t^* = -(1-r) \sup_{\gamma,\delta\in\Gamma} \Lambda_0^*\left(\delta, \frac{r}{1-r}\Lambda_1^*(\gamma,0)\right).$$

Therefore, there is no loss in optimality if the RFDC second stage sensors ignore the feedback message.

Proof: To simplify the proof, we assume that $\gamma \in \Gamma$ and $t \in \mathbb{R}$ can be chosen so that the supremum in (12) is achieved. To find the optimal threshold t, we set

$$\Lambda_0^*\left(\delta, \frac{r}{1-r}\Lambda_1^*(\gamma,t)\right) = \Lambda_1^*\left(\delta, -\frac{r}{1-r}\Lambda_0^*(\gamma,t)\right).$$

From the proposition hypothesis, we obtain $\Lambda_1^*(\gamma,t) = \Lambda_0^*(\gamma,t)$, which implies that $t = 0$. Therefore, from (11), δ satisfies

$$\Lambda_0^*\left(\delta, \frac{r}{1-r}\Lambda_1^*(\gamma,0)\right) = \sup_{\delta'\in\Gamma} \Lambda_0^*\left(\delta', \frac{r}{1-r}\Lambda_1^*(\gamma,0)\right). \quad (13)$$

Suppose that there exists $\delta^0 \neq \delta^1$, and $v \neq 0$ such that

$$\min\left\{ \Lambda_0^*\left(\delta^0, \frac{r}{1-r}\Lambda_1^*(\gamma,v)\right), \Lambda_1^*\left(\delta^1, -\frac{r}{1-r}\Lambda_0^*(\gamma,v)\right)\right\} > \Lambda_0^*\left(\delta, \frac{r}{1-r}\Lambda_1^*(\gamma,0)\right). \quad (14)$$

If $v > 0$, we have $\Lambda_1^*(\gamma,v) \leq \Lambda_1^*(\gamma,0)$ since $\Lambda_1^*(\gamma,\cdot)$ is a decreasing function. Therefore from (14), we obtain

$$\Lambda_0^*\left(\delta^0, \frac{r}{1-r}\Lambda_1^*(\gamma,0)\right) \geq \Lambda_0^*\left(\delta^0, \frac{r}{1-r}\Lambda_1^*(\gamma,v)\right) > \Lambda_0^*\left(\delta, \frac{r}{1-r}\Lambda_1^*(\gamma,0)\right),$$

a contradiction to (13). A similar argument produces a contradiction if $v < 0$. Therefore, we must have $v = 0$. But this implies that (14) cannot hold as it again contradicts (13). Hence, $\mathcal{E}_{dc}^* = \mathcal{E}_t^*$, and the proposition is proved. ■

The following example shows that in some cases, the RFDC performs strictly better in the presence of feedback.

Example 1: Let X_k take values in the set $\{1,2,3\}$, and suppose that sensor messages are restricted to a single bit. Assume that the probability mass functions under the two hypotheses are as shown in Table II. We also let $m = n/2$, i.e., $r = 1 - r = 1/2$.

	1	2	3
H_0	4/5	3/20	1/20
H_1	1/20	3/20	4/5

TABLE II
PROBABILITY MASS FUNCTIONS FOR EXAMPLE 1.

Since $\ell_{10}(X_k)$ is increasing with X_k, the two possible 1-bit quantizers are $\gamma_1(X_k) = 0$ iff $X_k = 1$, and $\gamma_2(X_k) = 0$ iff $X_k \in \{1,2\}$. We optimize (10) over these two quantizers and the threshold t. The results are shown in Figure 3. The optimal error exponent is found to be $-0.5 \cdot 0.73 = -0.365$, and is achieved by having all second stage sensors use γ_2 if the feedback message is 0, and γ_1 if the feedback message is 1. On the other hand, if feedback is ignored, the optimal quantizer is γ_2, and the optimal error exponent is -0.356, which is strictly worse than that with feedback.

It is interesting to note that unlike the daisy chain with full information at the fusion center (cf. Section IV-A), feedback in the RFDC may improve the detection performance in some scenarios. The fusion center in the RFDC architecture receives only a compressed summary of the information available at the first stage. We can think of the message from the first stage as a preliminary decision about the true hypothesis. At the fusion center, a significant weight is given to the first stage preliminary decision as compared to individual messages from sensors in the second stage. If sensors in the second stage ignore the feedback message, any errors in the preliminary decision cannot be controlled at the second stage. Errors in the preliminary decision therefore stay large. The first stage fusion should then try to balance the Type I and II errors by choosing a zero threshold for the likelihood ratio test. However, if the rate functions $\Lambda_1^*(\delta,t)$ and $\Lambda_0^*(\delta,t)$ for δ as in Proposition 1, are not symmetrical about the origin, then it is possible to choose a threshold for the first stage so that there is a bias towards one error probability type on average, and utilize feedback to allow the second stage sensors to control this bias. It turns out that since error exponents are not additive over stages, this is a better strategy. On the other hand, feedback is not required for the daisy chain with full information as sensor messages from both stages can be equally weighted and any errors in the preliminary decision can be averaged out.

Under the Neyman-Pearson formulation, there is no loss in optimality if feedback is ignored in the RFDC architecture [38], in contrast to the conclusion in Example 1. This is because in the Neyman-Pearson formulation, only the Type II error exponent is considered. One can design a strategy so that the first stage preliminary decision is biased in such a way that its Type II error is exponentially smaller than that for the second stage (note that the Type I error constraint applies only to the final decision, i.e., the second stage decision making), thus achieving the same optimal Type II error exponent even if feedback is ignored.

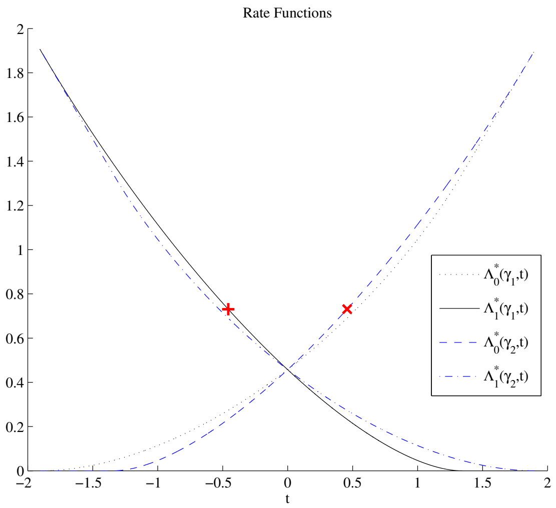

Fig. 3. Plot of the rate functions for γ_1 and γ_2. The mark 'x' indicates the optimal error decay rate (up to a constant $1/2$) when the feedback message $U=0$, while '+' indicates the optimal error decay rate (up to a constant $1/2$) when the feedback message $U=1$. The optimal quantizers are achieved on rate functions belonging to different quantizers.

In the following, we show that the RFDC performs strictly worse than a parallel configuration, and hence it has performance strictly inferior to a full feedback daisy-chain architecture.

Proposition 2: Suppose that the supremum in (10) is achieved. Then, the RFDC performs strictly worse than the parallel configuration with the same total number of sensors, i.e., $\mathcal{E}_t^* \geq \mathcal{E}_{dc}^* > \mathcal{E}_{1p}^* = -\sup_{\delta\in\Gamma}\Lambda_0^*(\delta,0)$.

Proof: Let $\gamma,\delta^0,\delta^1,t$ achieve the supremum in (10). If $\Lambda_1^*(\gamma,t)=0$, then from (10), we have

$$\begin{aligned}\mathcal{E}_{dc}^* &\geq -(1-r)\Lambda_0^*\left(\delta^0,0\right)\\ &> -\sup_{\delta\in\Gamma}\Lambda_0^*(\delta,0),\end{aligned}$$

since $r>0$. A similar argument shows that $\mathcal{E}_{dc}^* > \mathcal{E}_{1p}^*$ if $\Lambda_0^*(\gamma,t)=0$. Therefore, in the following, we assume that $\Lambda_j^*(\gamma,t)>0$ for $j=0,1$.

We have

$$\begin{aligned}&(1-r)\Lambda_0^*\left(\delta^0,\frac{r}{1-r}\Lambda_1^*(\gamma,t)\right)\\ &=(1-r)\Lambda_1^*\left(\delta^0,\frac{r}{1-r}\Lambda_1^*(\gamma,t)\right)+r\Lambda_1^*(\gamma,t)\\ &<(1-r)\Lambda_1^*\left(\delta^0,0\right)+r\Lambda_1^*(\gamma,t)\\ &\leq(1-r)\sup_{\delta\in\Gamma}\Lambda_1^*(\delta,0)+r\Lambda_1^*(\gamma,t),\end{aligned}\tag{15}$$

where the penultimate inequality follows from $\Lambda_1^*\left(\delta^0,\cdot\right)$ being a decreasing function, and $\Lambda_1^*(\gamma,t)>0$. Similarly,

$$\begin{aligned}&(1-r)\Lambda_1^*\left(\delta^1,-\frac{r}{1-r}\Lambda_0^*(\gamma,t)\right)\\ &<(1-r)\sup_{\delta\in\Gamma}\Lambda_0^*(\delta,0)+r\Lambda_0^*(\gamma,t).\end{aligned}\tag{16}$$

Combining (15) and (16), and since $\Lambda_0^*(\delta,0)=\Lambda_1^*(\delta,0)$ for all $\delta\in\Gamma$, we obtain

$$\begin{aligned}\mathcal{E}_{dc}^* &> -(1-r)\sup_{\delta\in\Gamma}\Lambda_0^*(\delta,0)-r\min\left\{\Lambda_1^*(\gamma,t),\Lambda_0^*(\gamma,t)\right\}\\ &\geq -(1-r)\sup_{\delta\in\Gamma}\Lambda_0^*(\delta,0)-r\Lambda_0^*(\gamma,0)\\ &\geq -\sup_{\delta\in\Gamma}\Lambda_0^*(\delta,0)=\mathcal{E}_{1p}^*.\end{aligned}$$

The proof is now complete. ∎

V. Conclusion

We have studied two-message feedback architectures, in which each sensor has access to compressed summaries of some or all other sensors' first messages to the fusion center. In the sequential feedback architecture, each sensor has access to the first messages of those sensors that communicate with the fusion center before it. In the restricted and full feedback architectures, each sensor has partial and full information respectively, about the first messages of every other sensor. Under both the Neyman-Pearson and Bayesian formulations, we show that the optimal error exponent is not improved by the feedback messages. We have also studied the one-message feedback architectures in which a group of sensors have access to information from sensors in a first group. We show that if the fusion center has knowledge of all the messages from the sensors in the first group, then feedback does not improve the optimal error exponent, which is the same as the parallel configuration. In the case where the fusion center has only limited knowledge (a 1-bit summary) of the messages, feedback can improve the optimal error exponent, but the optimal error exponent is strictly worse than that of the parallel configuration. Our results suggest that in the regime of a large number of sensors, and where the fusion center has sufficient memory, the performance gain in binary hypothesis testing due to feedback does not justify the increase in communication and computation costs incurred in a feedback architecture.

In the two-message feedback architecture, we assumed that the fusion center has unlimited memory and remembers all the first messages. The case where the fusion center retains only a finite-valued summary of the first messages has been studied in [36], but under various assumptions including finite-valued observation spaces, sensors all using the same quantization functions and constraints on the feedback messages. Reference [36] shows that feedback does not improve the error exponent. The same problem in the general setting that we have considered in this paper remains open.

In the case of Bayesian M-ary hypothesis testing, where $M>2$, we conjecture that feedback improves the optimal error exponent. Characterizing the optimal feedback strategy and error exponent is part of future work. This research is also part of our ongoing efforts to quantify the performance of various network architectures. Future research directions include studying network architectures with more general loop structures.

VI. Acknowledgements

The author wishes to thank Professor John Tsitsiklis for fruitful discussions and numerous suggestions that led to simplifications of some proofs, and significant improvements in the organization and presentation of this paper. He also wishes to thank the associate editor and the anonymous reviewers for their careful reading of the manuscript, and their valuable comments that have greatly improved the presentation of this paper.

Appendix A
Mathematical Preliminaries

In this appendix, we collect two well known results that are useful in our proofs. The first result is an elementary fact, which is an application of Jensen's inequality. A proof can be found in [42], and is omitted here.

Proposition A.1: Suppose $\phi : (0,\infty) \mapsto \mathbb{R}$ is a convex function. Then for any function γ, we have

$$\mathbb{E}_j\left[\phi\left(\ell_{ij}(\gamma(X))\right)\right] \leq \mathbb{E}_j\left[\phi\left(\ell_{ij}(X)\right)\right].$$

The following lower bound for the maximum of the Type I and II error probabilities was first proved in [47] for the case of discrete observation spaces. The following proposition generalizes the result to a general observation space. The proof is identical to that in [47], with some notation changes, and is provided for completeness.

Proposition A.2: Consider a hypothesis testing problem based on an observation X with distribution $\mathbb{P}_j$ under hypothesis H_j, $j = 0, 1$. Suppose that the measures $\mathbb{P}_0$ and $\mathbb{P}_1$ are absolutely continuous w.r.t. each other. Let $P_{e,j}$ be the probability of error when H_j is true. Let $Z = \log \frac{\mathrm{d}\mathbb{P}_1}{\mathrm{d}\mathbb{P}_0}(X)$ be the log Radon-Nikodym derivative. For any $s \in \mathbb{R}$, let $\Lambda(s) = \log \mathbb{E}_0[\exp(sZ)]$ be the log-moment generating function of Z. Then, for $s^* \in [0,1]$ such that $\Lambda'(s^*) = 0$, we have

$$\max(P_{e,0}, P_{e,1}) \geq \frac{1}{4}\exp\left(\Lambda(s^*) - \sqrt{2\Lambda''(s^*)}\right).$$

Proof: The proof steps are identical to that of Theorem 5 in [47]. Let P_j be the probability measure of Z under hypothesis H_j, $j = 0, 1$. For $s \in (0,1)$, define the probability measure Q such that

$$\frac{\mathrm{d}Q}{\mathrm{d}P_0}(z) = e^{sz-\Lambda(s)},$$

and let $\mathbb{E}_Q$ and var_Q be the mathematical expectation and variance w.r.t. Q, respectively. Let Y be a random variable with distribution Q. Then, it is easy to check that $\mathbb{E}_Q[Y] = \Lambda'(s)$ and $\mathrm{var}_Q(Y) = \Lambda''(s)$. Let $A_s = \{y : |y - \Lambda'(s)| \leq \sqrt{2\Lambda''(s)}\}$. From Chebychev's inequality, we have

$$Q(A_s) > \frac{1}{2}. \tag{17}$$

For any measurable set A, we have

$$\begin{aligned}P_0(A) &= \mathbb{E}_Q[\exp(-sZ + \Lambda(s))\mathbf{1}_{\{Z\in A\}}]\\ &\geq \mathbb{E}_Q[\exp(-sZ + \Lambda(s))\mathbf{1}_{\{Z\in A\cap A_s\}}]\\ &\geq \exp\left(\Lambda(s) - s\Lambda'(s) - s\sqrt{2\Lambda''(s)}\right)\, Q(A\cap A_s).\end{aligned}$$

Similarly, we have

$$\begin{aligned}P_1(A^c) \geq{}& \exp\left(\Lambda(s) + (1-s)\Lambda'(s) - (1-s)\sqrt{2\Lambda''(s)}\right)\\ &\cdot Q(A^c\cap A_s).\end{aligned}$$

From (17), either $Q(A\cap A_s) > 1/4$ or $Q(A^c\cap A_s) > 1/4$. Therefore, we have either

$$P_0(A) \geq \frac{1}{4}\exp\left(\Lambda(s) - s\Lambda'(s) - s\sqrt{2\Lambda''(s)}\right), \tag{18}$$

or

$$P_1(A^c) \geq \frac{1}{4}\exp\left(\Lambda(s) + (1-s)\Lambda'(s) - (1-s)\sqrt{2\Lambda''(s)}\right). \tag{19}$$

Since $\Lambda(s)$ is convex with $\Lambda(0) = \Lambda(1) = 0$, there exists $s^* \in (0,1)$ such that $\Lambda(s^*) = 0$. Substituting this into (18) and (19), we obtain

$$\max(P_0(A), P_1(A^c)) \geq \frac{1}{4}\exp\left(\Lambda(s^*) - \sqrt{2\Lambda''(s^*)}\right).$$

The proof is now complete. ■

Appendix B
Proof of Lemma 3

We first prove claim (i). From Lemma 2.2.5 of [45], φ_ξ is a convex function with nonnegative second derivatives. We next show that its second derivative is uniformly upper bounded for all $\xi \in \Gamma^2$. From Lemma A.1, we have

$$\mathbb{E}_0\left[(\ell_{10}(\xi(X_1)))^s\right] \leq \mathbb{E}_0\left[(\ell_{10}(X_1))^s\right] \leq e^{\varphi_{\mathrm{Id}}(\bar{s})}. \tag{20}$$

Let $f(s) = \mathbb{E}_0\left[(\ell_{10}(\xi(X_1)))^s\right]$, and $\eta = \min\{|\bar{s}|/2, 1\}$. There exists a positive constant M such that for all $|x| > M$, we have $x^2 \leq \exp(\eta x) + \exp(-\eta x)$. Making use of this bound, we obtain

$$\begin{aligned}\varphi_\xi''(s) &= \frac{\mathbb{E}_0\left[(\ell_{10}(\xi(X_1)))^s \log^2 \ell_{10}(\xi(X_1))\right]}{\mathbb{E}_0\left[(\ell_{10}(\xi(X_1)))^s\right]} - (\varphi_\xi'(s))^2\\ &\leq \mathbb{E}_0\left[(\ell_{10}(\xi(X_1)))^s \log^2 \ell_{10}(\xi(X_1))\right]\\ &\leq M^2 f(s) + f(s+\eta) + f(s-\eta)\\ &\leq (M^2+2) f(\bar{s})\\ &\leq (M^2+2)e^{\varphi_{\mathrm{Id}}(\bar{s})}.\end{aligned}$$

The third inequality follows from the bounds $\bar{s} < s + \eta < 1$ and $\bar{s} < s - \eta < 0$, and the facts that $f(x)$ is nonincreasing over $[\bar{s}, 0]$, while $f(x) \leq 1 \leq f(\bar{s})$ for $x \in [0,1]$. The final inequality follows from (20). Claim (i) is now proved.

We now use a Taylor series expansion to prove claim (ii). Since $\varphi_\xi(0) = 0$ for any $\xi \in \Gamma^2$, we have for $s \in [-h, 0]$,

$$\begin{aligned}&\varphi_\xi(s) - \varphi_{\xi^*}(s)\\ &= (\varphi_\xi'(0) - \varphi_{\xi^*}'(0))s + (\varphi_\xi''(s_1) - \varphi_{\xi^*}''(s_2))\frac{s^2}{2}\\ &\leq \epsilon|s| + c\frac{s^2}{2}\\ &\leq \epsilon/2,\end{aligned}$$

where s_1 and s_2 are between s and 0, and the first inequality follows from $\varphi_\xi'(0) \geq g_{2p}^*$, $\varphi_{\xi^*}'(0) \leq g_{2p}^* + \epsilon$, $\varphi_\xi''(s_1) \leq c$, and $\varphi_{\xi^*}''(s_2) \geq 0$.

Finally, we turn to the proof of claim (iii). Recall that $Z_k = \delta_k^{W_k}(X_k)$. For $s \in [-h, 0]$, we have

$$\begin{aligned}
&\mathbb{E}_0\left[\prod_{k=1}^n (\ell_{10}(Z_k|Y_k, W_k))^s \;\Big|\; Y_1^n\right] \\
&= \mathbb{E}_0\left[\prod_{k=1}^n (\ell_{10}(\delta_k^{W_k}(X_k)|Y_k))^s \;\Big|\; Y_1^n\right] \\
&= \prod_{k=1}^n \mathbb{E}_0\left[(\ell_{10}(\delta_k^{W_k}(X_k)|Y_k))^s \;\Big|\; Y_1^n\right] \\
&\leq \prod_{k=1}^n \left(\mathbb{E}_0\left[(\ell_{10}(\delta^{Y_k}(X_k)|Y_k))^s \;\Big|\; Y_k\right] + \epsilon_1\right),
\end{aligned} \tag{21}$$

where $\epsilon_1 = \epsilon \exp(-\varphi_{\mathrm{Id}}(\bar{s}))/2$, and $\delta^{Y_k} \in \Gamma$ is a function depending on the value of Y_k, and is such that

$$\begin{aligned}
&\mathbb{E}_0\left[(\ell_{10}(\delta^{Y_k}(X_k)|Y_k))^s \;\Big|\; Y_k\right] \\
&\geq \sup_{\delta\in\Gamma} \mathbb{E}_0\left[(\ell_{10}(\delta(X_k)|Y_k))^s \;\Big|\; Y_k\right] - \epsilon_1.
\end{aligned}$$

From (21), we have

$$\begin{aligned}
&\frac{\psi_n(s)}{n} \\
&= \frac{1}{n}\log \mathbb{E}_0\left[(\ell_{10}(Y_1^n))^s \cdot \mathbb{E}_0\left[\prod_{k=1}^n (\ell_{10}(Z_k|Y_k, W_k))^s \;\Big|\; Y_1^n\right]\right] \\
&\leq \frac{1}{n}\log \mathbb{E}_0\left[(\ell_{10}(Y_1^n))^s \cdot \prod_{k=1}^n \left(\mathbb{E}_0\left[(\ell_{10}(\delta^{Y_k}(X_k)|Y_k))^s \;\Big|\; Y_k\right] + \epsilon_1\right)\right] \\
&= \frac{1}{n}\log \prod_{k=1}^n \left(\mathbb{E}_0\left[(\ell_{10}(Y_k, \delta^{Y_k}(X_k)))^s\right] + \epsilon_1 \mathbb{E}_0\left[(\ell_{10}(Y_k))^s\right]\right) \\
&\leq \frac{1}{n}\sum_{k=1}^n \log \mathbb{E}_0\left[(\ell_{10}(\gamma_k(X_k), \delta^{Y_k}(X_k)))^s\right] + \frac{\epsilon}{2},
\end{aligned}$$

where the last inequality follows from (20), and the inequality $\log(x+\epsilon) \leq \log x + \epsilon$ for $x \geq 1$. Let $\xi_k \in \Gamma^2$ such that $\xi_k(X_k) = (\gamma_k(X_k), \delta_k(X_k))$, where $\delta_k(X_k) = \delta^u(X_k)$ iff $\gamma_k(X_k) = u \in \mathcal{T}$. We therefore have

$$\frac{\psi_n(s)}{n} \leq \frac{1}{n}\sum_{k=1}^n \varphi_{\xi_k}(s) + \frac{\epsilon}{2} \leq \varphi_{\xi^*}(s) + \epsilon,$$

where the second inequality follows from claim (ii). The proof is now complete.

Appendix C
Proof of Lemma 4

We first show claim (i). To show the bounds on $\psi_n'(s)$, we note that ψ_n is convex, so $\psi_n'(0) \leq \psi_n'(s) \leq \psi_n'(1)$ for all $s \in [0, 1]$. Using Proposition A.1, it is then easy to check that $\psi_n'(0) \geq n\mathbb{E}_0[\log \ell_{10}(X_1)]$ and $\psi_n'(1) \leq n\mathbb{E}_1[\log \ell_{10}(X_1)]$.

Next, we prove claim (ii). We have

$$\begin{aligned}
\psi_n''(s_n) &= \frac{\mathbb{E}_0[(\mathcal{L}_{10}^{(n)})^2 \exp(s_n \mathcal{L}_{10}^{(n)})]}{\mathbb{E}_0[\exp(s_n \mathcal{L}_{10}^{(n)})]} - (\psi_n'(s_n))^2 \\
&\leq C_1 \mathbb{E}_0[(\mathcal{L}_{10}^{(n)})^2 \exp(s_n \mathcal{L}_{10}^{(n)})],
\end{aligned} \tag{22}$$

where the inequality follows from the bound $\mathbb{E}_0[\exp(s_n \mathcal{L}_{10}^{(n)})] \geq 1/C_1$, for some constant C_1 independent of the strategy. (This fact is proved in Proposition 3 of [42].) The right-hand side of (22) can be upper bounded by observing that

$$\begin{aligned}
&\mathbb{E}_0[(\mathcal{L}_{10}^{(n)})^2 \exp(s_n \mathcal{L}_{10}^{(n)})] \\
&= \mathbb{E}_0\left[(\mathcal{L}_{10}^{(n)})^2 e^{s_n \mathcal{L}_{10}^{(n)}} \mathbf{1}_{\{\mathcal{L}_{10}^{(n)} \leq 0\}}\right] \\
&\quad + \mathbb{E}_1\left[(\mathcal{L}_{10}^{(n)})^2 e^{-(1-s_n)\mathcal{L}_{10}^{(n)}} \mathbf{1}_{\{\mathcal{L}_{10}^{(n)} > 0\}}\right] \\
&\leq 4\left(\frac{1}{s_n^2} + \frac{1}{(1-s_n)^2}\right),
\end{aligned} \tag{23}$$

where in the inequality, we use the result that the function $f_1(x) = x^2 \exp(s_n x)\mathbf{1}_{\{x \leq 0\}}$ is maximized at $-2/s_n$, and the function $f_2(x) = x^2 \exp(-(1-s_n)x)\mathbf{1}_{\{x>0\}}$ is maximized at $2/(1-s_n)$. It now suffices to show that both s_n and $1 - s_n$ are at least $C_2/\sqrt{n}$ for some positive constant C_2 independent of the particular strategy chosen. To simplify the notation, let $\ell_n = \exp(\mathcal{L}_{10}^{(n)})$. Suppose that $|t_n| \leq t$ for all n. Using the inequalities $x^s \leq sx + 1$ for $0 < s < 1$, and $x^s \geq x$ for $x \leq 1$, we obtain from the equation $\psi_n'(s_n) = t_n$,[5]

$$\begin{aligned}
&t_n \mathbb{E}_0\left[\exp(s_n \mathcal{L}_{10}^{(n)})\right] \\
&= \mathbb{E}_0\left[(\ell_n)^{s_n} \log \ell_n\right] \\
&= \mathbb{E}_0\left[(\ell_n)^{s_n} (\log \ell_n)^+\right] - \mathbb{E}_0\left[(\ell_n)^{s_n} (\log \ell_n)^-\right] \\
&\leq s_n\left(\mathbb{E}_0\left[\ell_n (\log \ell_n)^+\right] + \mathbb{E}_0\left[(\log \ell_n)^+\right]\right) - \mathbb{E}_0\left[\ell_n (\log \ell_n)^-\right],
\end{aligned}$$

which yields

$$s_n \geq \frac{\mathbb{E}_1\left[(\log \ell_n)^-\right] - t}{\mathbb{E}_1\left[(\log \ell_n)^+\right] + \mathbb{E}_0\left[(\log \ell_n)^+\right]}, \tag{24}$$

since $0 \leq \mathbb{E}_0\left[\exp(s_n \mathcal{L}_{10}^{(n)})\right] \leq 1$ and $|t_n| \leq t$. We first bound the denominator in (24) by using $g(x) = x(\log x)^+$, which is a convex function, and Proposition A.1 to get

$$\begin{aligned}
&\mathbb{E}_1\left[(\log \ell_n)^+\right] \\
&= \mathbb{E}_0\left[g(\ell_n)\right] \\
&\leq \mathbb{E}_0\left[g(\ell_{10}(X_1^n))\right] \\
&\leq \mathbb{E}_0\left[\ell_{10}(X_1^n) |\log \ell_{10}(X_1^n)|\right] \\
&= \mathbb{E}_1\left[|\log \ell_{10}(X_1^n)|\right] \\
&\leq \sum_{k=1}^n \mathbb{E}_1\left[|\log \ell_{10}(X_k)|\right] \\
&\leq \sum_{k=1}^n \mathbb{E}_1\left[\log^2 \ell_{10}(X_k)\right] + n \\
&\leq nC_3,
\end{aligned} \tag{25}$$

where C_3 is a constant, and the last inequality follows from Lemma 1. Similarly, it can be shown that $\mathbb{E}_0\left[(\log \ell_n)^+\right] = \mathbb{E}_0[(\mathcal{L}_{01}^{(n)})^-] \leq \mathbb{E}_0[(\mathcal{L}_{01}^{(n)})^+]$ is bounded by nC_3. Next, we show a lower bound for the numerator in (24). Let

$$f(x) = \begin{cases} x(\log x)^-, & \text{if } 0 \leq x \leq 1, \\ 1 - x, & \text{if } x > 1, \end{cases}$$

[5] We use the notations $x^+ = \max(x, 0)$ and $x^- = -\min(x, 0)$.

which is a concave function not greater than $x(\log x)^-$. From Proposition A.1, we obtain

$$\begin{aligned}
&\mathbb{E}_1\left[(\log \ell_n)^-\right] \\
&= \mathbb{E}_0\left[\ell_n(\log \ell_n)^-\right] \\
&\geq \mathbb{E}_0\left[f(\ell_n)\right] \\
&\geq \mathbb{E}_0\left[f(\ell_{10}(X_1^n))\right] \\
&\geq \mathbb{E}_1\left[(\log \ell_{10}(X_1^n))^-\right] - \mathbb{P}_1(\ell_{10}(X_1^n) > 1) \\
&\geq \sqrt{n}\mathbb{E}_1\left[\left(\frac{1}{\sqrt{n}}\sum_{k=1}^n \log \ell_{10}(X_k)\right)^-\right] - 1.
\end{aligned}$$

Applying Fatou's Lemma and the Central Limit Theorem, we obtain

$$\liminf_{n\to\infty} \frac{\mathbb{E}_1\left[(\log \ell_n)^-\right]}{\sqrt{n}} \geq C_4, \tag{26}$$

where C_4 is a positive constant. Substituting the bounds (25) and (26) into (24), we finally have

$$\liminf_{n\to\infty} s_n\sqrt{n} \geq C_2,$$

for some positive constant C_2. A similar proof using

$$\begin{aligned}
&\mathbb{E}_1[\mathcal{L}_{01}^{(n)} \exp((1-s_n)\mathcal{L}_{01}^{(n)})] \\
&= -\mathbb{E}_0[\mathcal{L}_{10}^{(n)} \exp(s_n\mathcal{L}_{10}^{(n)})] \\
&= -t_n\mathbb{E}_0\left[\exp(s_n\mathcal{L}_{10}^{(n)})\right] \geq -t
\end{aligned}$$

shows that the same bound holds for $1-s_n$. Therefore, from (23), claim (ii) holds.

In the following, we establish claim (iii). Let ϵ be a positive constant. Similar to the proof of Lemma 3(iii), let $\delta^{Y_k} \in \Gamma$ be a function depending on the value of Y_k, so that

$$\begin{aligned}
&\mathbb{E}_0\left[(\ell_{10}(\delta^{Y_k}(X_k)|Y_k))^s \;\middle|\; Y_k\right] \\
&\leq \inf_{\delta\in\Gamma} \mathbb{E}_0\left[(\ell_{10}(\delta(X_k)|Y_k))^s \;\middle|\; Y_k\right] + \epsilon.
\end{aligned}$$

We have

$$\begin{aligned}
&\psi_n(s) \\
&= \log \mathbb{E}_0\left[(\ell_{10}(Y_1^n))^s \cdot \mathbb{E}_0\left[\prod_{k=1}^n (\ell_{10}(Z_k|Y_k, W_k))^s \;\middle|\; Y_1^n\right]\right] \\
&\geq \frac{1}{n}\log \mathbb{E}_0\Bigg[(\ell_{10}(Y_1^n))^s \cdot \\
&\qquad \prod_{k=1}^n \left(\mathbb{E}_0\left[(\ell_{10}(\delta^{Y_k}(X_k)|Y_k))^s \;\middle|\; Y_k\right] - \epsilon\right)\Bigg] \\
&= \sum_{k=1}^n \log \mathbb{E}_0\left[(\ell_{10}(Y_k))^s\left(\mathbb{E}_0\left[(\ell_{10}(\delta^{Y_k}(X_k)|Y_k))^s \;\middle|\; Y_k\right] - \epsilon\right)\right] \\
&\geq \sum_{k=1}^n \log\left(\mathbb{E}_0\left[(\ell_{10}(Y_k, \delta^{Y_k}(X_k)))^s\right] - \epsilon\right),
\end{aligned} \tag{27}$$

where we have used the inequality $\mathbb{E}_0\left[(\ell_{10}(Y_k))^s\right] \leq 1$ in (27). Recall that $Y_k = \gamma_k(X_k)$. We can define $\xi_k \in \Gamma^2$ such that $\xi_k(X_k) = (\gamma_k(X_k), \delta_k(X_k))$, where $\delta_k(X_k) = \delta^u(X_k)$ iff $\gamma_k(X_k) = u \in \mathcal{T}$. From (27), we obtain the bound

$$\begin{aligned}
\psi_n(s) &\geq \sum_{k=1}^n \log\left(\mathbb{E}_0\left[(\ell_{10}(\xi_k(X_k)))^s\right] - \epsilon\right) \\
&\geq n\log\left(\inf_{\xi\in\Gamma^2} \mathbb{E}_0\left[(\ell_{10}(\xi(X_1)))^s\right] - \epsilon\right).
\end{aligned} \tag{28}$$

Since ϵ is arbitrary, the lemma is proved.

Appendix D
Proof of Lemma 5

Let us fix a sequence of strategies that conform to the given quantization functions for sensors $1,\ldots,m$. Let $\hat{\alpha}_n$ and $\hat{\beta}_n$ be the Type I and II error probabilities of a strategy with the fusion rule

$$Y_f = \begin{cases} 0, & \text{if } \mathcal{L}_{10}^{(n)} \leq 0, \\ 1, & \text{if } \mathcal{L}_{10}^{(n)} > 0. \end{cases}$$

From the Neyman-Pearson Lemma [48], the optimal decision rule at the fusion center is the Neyman-Pearson test. Moreover, for any given fusion rule, either the Type I or II error probability is at least $\hat{\alpha}_n$ or $\hat{\beta}_n$. Therefore, we have

$$\liminf_{n\to\infty} \frac{1}{n}\log P_e(n) \geq \min\{\liminf_{n\to\infty}\frac{1}{n}\log\hat{\alpha}_n, \liminf_{n\to\infty}\frac{1}{n}\log\hat{\beta}_n\}. \tag{29}$$

Thus it suffices to find a lower bound for the strategy using a zero threshold log likelihood ratio test as a fusion rule. Henceforth, we will assume that such a fusion rule is employed. Conditioning on the value of U, we have

$$\begin{aligned}
P_1(Y_f = 0) =& \mathbb{P}_1(Y_f = 0 \mid U = 0)\mathbb{P}_1(U = 0) \\
&+ \mathbb{P}_1(Y_f = 0 \mid U = 1)\mathbb{P}_1(U = 1).
\end{aligned}$$

Fix an $\epsilon > 0$. Let $\delta_i(\cdot, u) = \delta_i^u(\cdot) \in \Gamma$ be a function that depends on the value of u. Let $l = n - m$. Using the lower bound in Cramèr's Theorem (cf. [46]) and Lemma 4, we obtain

$$\begin{aligned}
&\frac{1}{n}\log \mathbb{P}_1(Y_f = 0 \mid U = 0) \\
&= \frac{1}{n}\log \mathbb{P}_1(\mathcal{L}_{10}^{(n)}/l \leq 0 \mid U = 0) \\
&= \frac{1}{n}\log \mathbb{P}_1\left(\frac{1}{l}\sum_{k=1}^l \log \ell_{10}(\delta_k^0(X_k)) \leq -\frac{1}{l}\log\frac{\mathbb{P}_1(U=0)}{\mathbb{P}_0(U=0)}\right) \\
&\geq -\frac{l}{n}\cdot\frac{1}{l}\sum_{k=1}^l \Lambda_1^*\left(\delta_k^0, -\frac{1}{l}\log\frac{\mathbb{P}_1(U=0)}{\mathbb{P}_0(U=0)} - \epsilon\right) + o(1) \\
&\geq -\frac{l}{n}\sup_{\delta^0\in\Gamma} \Lambda_1^*\left(\delta^0, -\frac{1}{l}\log\frac{\mathbb{P}_1(U=0)}{\mathbb{P}_0(U=0)} - \epsilon\right) + o(1),
\end{aligned}$$

where $o(1)$ is a term that goes to zero as $n \to \infty$. Taking $n \to \infty$ and then $\epsilon \to 0$, we obtain

$$\begin{aligned}
&\liminf_{n\to\infty}\frac{1}{n}\log \mathbb{P}_1(Y_f = 0 \mid U = 0) + \liminf_{n\to\infty}\frac{1}{n}\log \mathbb{P}_1(U = 0) \\
&\geq -(1-r)\sup_{\delta^0\in\Gamma} \Lambda_1^*\left(\delta^0, \frac{r}{1-r}(e_{10} - e_{00})\right) - re_{10} \\
&= -(1-r)\sup_{\delta^0\in\Gamma} \Lambda_0^*\left(\delta^0, \frac{r}{1-r}(e_{10} - e_{00})\right) - re_{00}
\end{aligned} \tag{30}$$

In the same way, it can be checked that

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}_1(Y_f = 0 \mid U = 1) + \liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}_1(U = 1)$$
$$\geq -(1-r) \sup_{\delta^1 \in \Gamma} \Lambda_1^* \left(\delta^1, -\frac{r}{1-r}(e_{01} - e_{11}) \right) - r e_{11}, \tag{31}$$

and we obtain

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}_1(Y_f = 0) \geq -h(\vec{e}).$$

A similar proof shows that

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}_0(Y_f = 1) \geq -h(\vec{e}),$$

and that the optimal error exponent is lower bounded by $-h(\vec{e})$. We note that this lower bound can be asymptotically achieved by letting all sensors in the second stage quantize their observations using δ^0 and δ^1 if the feedback message is 0 or 1 respectively, and where δ^0 and δ^1 are chosen to asymptotically maximize their respective rate functions in (9). The proof is now complete.

REFERENCES

[1] R. R. Tenney and N. R. Sandell, "Detection with distributed sensors," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 17, no. 4, pp. 501–510, 1981.

[2] S. Alhakeem and P. K. Varshney, "A unified approach to the design of decentralized detection systems," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 31, no. 1, pp. 9–20, 1995.

[3] S. Alhakeem and P. Varshney, "Decentralized Bayesian detection with feedback," *IEEE Trans. Syst., Man, Cybern.*, vol. 26, no. 4, pp. 503–513, Jul. 1996.

[4] Z. Chair and P. K. Varshney, "Optimal data fusion in multiple sensor detection systems," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 22, no. 1, pp. 98–101, 1986.

[5] G. Polychronopoulos and J. N. Tsitsiklis, "Explicit solutions for some simple decentralized detection problems," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 26, no. 2, pp. 282–292, 1990.

[6] P. Willett and D. Warren, "The suboptimality of randomized tests in distributed and quantized detection systems," *IEEE Trans. Inf. Theory*, vol. 38, no. 2, pp. 355–361, Mar. 1992.

[7] J. N. Tsitsiklis, "Extremal properties of likelihood-ratio quantizers," *IEEE Trans. Commun.*, vol. 41, no. 4, pp. 550–558, 1993.

[8] ——, "Decentralized detection," *Advances in Statistical Signal Processing*, vol. 2, pp. 297–344, 1993.

[9] W. W. Irving and J. N. Tsitsiklis, "Some properties of optimal thresholds in decentralized detection," *IEEE Trans. Autom. Control*, vol. 39, no. 4, pp. 835–838, 1994.

[10] R. Viswanathan and P. K. Varshney, "Distributed detection with multiple sensors: part I - fundamentals," *Proc. IEEE*, vol. 85, no. 1, pp. 54–63, 1997.

[11] B. Chen and P. K. Varshney, "A Bayesian sampling approach to decision fusion using hierarchical models," *IEEE Trans. Signal Process.*, vol. 50, no. 8, pp. 1809–1818, Aug. 2002.

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

[13] B. Chen and P. K. Willett, "On the optimality of the likelihood-ratio test for local sensor decision rules in the presence of nonideal channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 2, pp. 693–699, Feb. 2005.

[14] A. Kashyap, "Comments on "On the optimality of the likelihood-ratio test for local sensor decision rules in the presence of nonideal channels"," *IEEE Trans. Inf. Theory*, vol. 52, no. 3, pp. 1274–1275, Mar. 2006.

[15] B. Liu and B. Chen, "Channel-optimized quantizers for decentralized detection in sensor networks," *IEEE Trans. Inf. Theory*, vol. 52, no. 7, pp. 3349–3358, Jul. 2006.

[16] R. Viswanathan, S. C. A. Thomopoulos, and R. Tumuluri, "Optimal serial distributed decision fusion," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 24, no. 4, pp. 366–376, 1988.

[17] Z. B. Tang, K. R. Pattipati, and D. L. Kleinman, "Optimization of detection networks: part I - tandem structures," *IEEE Trans. Syst., Man, Cybern.*, vol. 21, no. 5, pp. 1044–1059, 1991.

[18] J. D. Papastavrou and M. Athans, "Distributed detection by a large team of sensors in tandem," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 28, no. 3, pp. 639–653, 1992.

[19] W. P. Tay, J. N. Tsitsiklis, and M. Z. Win, "On the sub-exponential decay of detection error probabilities in long tandems," *IEEE Trans. Inf. Theory*, vol. 54, no. 10, pp. 4767–4771, Oct. 2008.

[20] L. K. Ekchian and R. R. Tenney, "Detection networks," in *Proc. IEEE Conf. Decision and Control*, vol. 21, Dec. 1982, pp. 686–691.

[21] A. R. Reibman and L. W. Nolte, "Design and performance comparison of distributed detection networks," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 23, no. 6, pp. 789–797, 1987.

[22] Z. B. Tang, K. R. Pattipati, and D. L. Kleinman, "Optimization of detection networks: part II - tree structures," *IEEE Trans. Syst., Man, Cybern.*, vol. 23, no. 1, pp. 211–221, 1993.

[23] A. Pete, K. Pattipati, and D. Kleinman, "Optimization of detection networks with multiple event structures," *IEEE Trans. Autom. Control*, vol. 39, no. 8, pp. 1702–1707, 1994.

[24] Y. Lin, B. Chen, and P. K. Varshney, "Decision fusion rules in multi-hop wireless sensor networks," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 41, no. 2, pp. 475–488, Apr. 2005.

[25] W. P. Tay, J. N. Tsitsiklis, and M. Z. Win, "On the impact of node failures and unreliable communications in dense sensor networks," *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2535–2546, Jun. 2008.

[26] ——, "Data fusion trees for detection: Does architecture matter?" *IEEE Trans. Inf. Theory*, vol. 54, no. 9, pp. 4155–4168, Sep. 2008.

[27] ——, "Bayesian detection in bounded height tree networks," *IEEE Trans. Signal Process.*, vol. 57, no. 10, pp. 4042–4051, Oct 2009.

[28] J. N. Tsitsiklis and M. Athans, "On the complexity of decentralized decision making and detection problems," *IEEE Trans. Autom. Control*, vol. 30, pp. 440–446, 1985.

[29] J.-F. Chamberland and V. V. Veeravalli, "How dense should a sensor network be for detection with correlated observations?" *IEEE Trans. Inf. Theory*, vol. 52, no. 11, pp. 5099–5106, Nov. 2006.

[30] S. Misra and L. Tong, "Error exponents for the detection of Gauss-Markov signals using randomly spaced sensors," *Signal Processing, IEEE Transactions on*, vol. 56, no. 8, pp. 3385 –3396, Aug. 2008.

[31] Y. Sung, H. V. Poor, and H. Yu, "How much information can one get from a wireless ad hoc sensor network over a correlated random field?" *IEEE Trans. Inf. Theory*, vol. 55, no. 6, pp. 2827–2847, Jun. 2009.

[32] P.-N. Chen and A. Papamarcou, "Error bounds for parallel distributed detection under the Neyman-Pearson criterion," *Proc. IEEE Int. Symp. Inform. Theory*, pp. 528–533, Mar. 1995.

[33] P. Willett, P. Swaszek, and R. Blum, "The good, bad and ugly: distributed detection of a known signal in dependent Gaussian noise," *Signal Processing, IEEE Transactions on*, vol. 48, no. 12, pp. 3266 – 3279, Dec. 2000.

[34] H. Chen, B. Chen, and P. Varshney, "A new framework for distributed detection with conditionally dependent observations," *IEEE Trans. Signal Process.*, vol. 60, no. 3, pp. 1409 –1419, Mar. 2012.

[35] O. Kreidl and A. Willsky, "An efficient message-passing algorithm for optimizing decentralized detection networks," *IEEE Trans. Autom. Control*, vol. 55, no. 3, pp. 563–578, Mar. 2010.

[36] H. Shalaby and A. Papamarcou, "A note on the asymptotics of distributed detection with feedback," *Information Theory, IEEE Transactions on*, vol. 39, no. 2, pp. 633–640, Mar. 1993.

[37] S. I. Zoumpoulis, "Decentralized detection in sensor network architectures with feedback," Master's thesis, Massachusetts Institute of Technology, Jun. 2009.

[38] O. P. Kreidl, J. N. Tsitsiklis, and S. I. Zoumpoulis, "On decentralized detection with partial information sharing among sensors," *IEEE Trans. Signal Process.*, vol. 59, no. 4, pp. 1759–1765, Apr. 2011.

[39] L. Smith and P. Sorensen, "Pathological outcomes of observational learning," *Econometrica*, vol. 68, pp. 371–398, 2000.

[40] D. Acemoglu, M. A. Dahleh, I. Lobel, and A. Ozdaglar, "Bayesian learning in social networks," *Review of Economic Studies*, 2011, forthcoming.

[41] W. P. Tay, J. N. Tsitsiklis, and M. Z. Win, "Asymptotic performance of a censoring sensor network," *IEEE Trans. Inf. Theory*, vol. 53, no. 11, pp. 4191–4209, Nov. 2007.

[42] J. N. Tsitsiklis, "Decentralized detection by a large number of sensors," *Math. Control, Signals, Syst.*, vol. 1, pp. 167–182, 1988.

[43] H. V. Poor, *An Introduction to Signal Detection and Estimation*, 2nd ed. New York, NY: Springer-Verlag, 1994.

[44] P.-N. Chen, “General formulas for the neyman-pearson type-II error exponent subject to fixed and exponential type-i error bounds,” *IEEE Trans. Inf. Theory*, vol. 42, no. 1, pp. 316–323, 1996.
[45] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*. New York, NY: Springer-Verlag, 1998.
[46] D. Stroock, *Probability Theory: An Analytic View*. Cambridge, UK: Cambridge University Press, 1993.
[47] C. E. Shannon, R. G. Gallager, and E. R. Berlekamp, “Lower bounds to error probability for coding on discrete memoryless channels, I,” *Information and Control*, vol. 10, pp. 65–103, 1967.
[48] E. Lehmann and J. P. Romano, *Testing Statistical Hypotheses*. New York, NY: Springer, 2005.

Wee Peng Tay (S’06–M’08) received the B.S. degree in electrical engineering and mathematics, and the M.S. degree in electrical engineering from Stanford University in 2002. He received the Ph.D. degree in electrical engineering and computer science from the Massachusetts Institute of Technology in 2008. He is currently an Assistant Professor in the School of Electrical and Electronics Engineering at Nanyang Technological University in Singapore. He received the Singapore Technologies Scholarship in 1998, the Stanford University President’s Award in 1999, and the Frederick Emmons Terman Engineering Scholastic Award in 2002. His research interests include distributed decision making, data fusion, distributed algorithms, communications in ad hoc networks, machine learning and applied probability.