This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.



# Hypothesis Testing in Feedforward Networks with Broadcast Failures

Zhenliang Zhang, *Student Member, IEEE,* Edwin K. P. Chong, *Fellow, IEEE,*
Ali Pezeshki, *Member, IEEE,* and William Moran, *Member, IEEE*

***Abstract*—Consider a large number of nodes, which sequentially make decisions between two given hypotheses. Each node takes a measurement of the underlying truth, observes the decisions from some immediate predecessors, and makes a decision between the given hypotheses. We consider two classes of broadcast failures: 1) each node broadcasts a decision to the other nodes, subject to random erasure in the form of a binary erasure channel; 2) each node broadcasts a randomly flipped decision to the other nodes in the form of a binary symmetric channel. We are interested in conditions under which there does (or does not) exist a decision strategy consisting of a sequence of likelihood ratio tests such that the node decisions converge in probability to the underlying truth, as the number of nodes goes to infinity. In both cases, we show that if each node only learns from a bounded number of immediate predecessors, then there does not exist a decision strategy such that the decisions converge in probability to the underlying truth. However, in case 1, we show that if each node learns from an unboundedly growing number of predecessors, then there exists a decision strategy such that the decisions converge in probability to the underlying truth, even when the erasure probabilities converge to $1$. We show that a locally optimal strategy, consisting of a sequence of Bayesian likelihood ratio tests, is such a strategy, and we derive the convergence rate of the error probability for this strategy. In case 2, we show that if each node learns from all of its previous predecessors, then there exists a decision strategy such that the decisions converge in probability to the underlying truth when the flipping probabilities of the binary symmetric channels are bounded away from $1/2$. Again, we show that a locally optimal strategy achieves this, and we derive the convergence rate of the error probability for it. In the case where the flipping probabilities converge to $1/2$, we derive a necessary condition on the convergence rate of the flipping probabilities such that the decisions based on the locally optimal strategy still converge to the underlying truth. We also explicitly characterize the relationship between the convergence rate of the error probability and the convergence rate of the flipping probabilities.**

***Index Terms*—Asymptotic learning, decentralized detection, erasure channel, herding, social learning, symmetric channel.**

## I. Introduction

We consider a large number of nodes, which sequentially make decisions between two hypotheses $H_0$ and $H_1$. At stage $k$, node $a_k$ takes a measurement $X_k$ (called its *private signal*), receives the decisions of $m_k < k$ immediate predecessors, and makes a binary decision $d_k = 0$ or $1$ about the prevailing hypothesis $H_0$ or $H_1$, respectively. It then broadcasts a decision to its successors. Note that $m_k$ is often referred to as the *memory size*. A typical question is this: Can these nodes asymptotically learn the underlying true hypothesis? In other words, does the decision $d_k$ converge (in probability) to the true hypothesis as $k \to \infty$? If so, what is the convergence rate of the error probability?

One application of the sequential hypothesis testing problem[1] is decentralized detection in sensor networks, in which case the set of nodes represents a set of spatially distributed sensors attempting to jointly solve the hypothesis testing problem, for example, the presence or absence of a target. Decentralized detection problems have been intensively studied in recent years; see [3] for a comprehensive introduction to this problem. Usually, a sensor network consists of a large number of low-cost sensors with limited resources for processing and transmitting data. Therefore, each sensor has to aggregate its measurement and the observed messages from the previous sensors into a much smaller message (e.g., a 1-bit decision) and then sends it to other sensors for further aggregation. These sensors are subject to random failures, (e.g., dead battery), in which case the failed sensor cannot transmit its message. Moreover, the communication channels between sensors are noisy and the 1-bit messages are subject to random erasures or random flippings. A central question is whether or not there exists a sequence of decision rules for aggregating the spatially distributed information such that the decisions converge to the underlying truth as the number of sensors increases.

Another application is social learning in multi-agent networks, in which case the set of nodes represents a set of agents trying to learn the underlying truth (also known as the state of the world). Each agent makes a decision based on its own measurement and what it learns from the actions/decisions of the previous agents. In this case, we usually assume that each agent uses a myopic decision rule to minimize a local objective function; for example, the probability of error is locally minimized using the Bayesian likelihood ratio test with a threshold given by the ratio of the prior probabilities. The

This work was supported in part by AFOSR under contract FA9550-09-1-0518, and by NSF under grants CCF-0916314 and CCF-1018472. Part of the results in this paper was presented in [1].

Z. Zhang, E. K. P. Chong, and A. Pezeshki are with the Department of Electrical and Computer Engineering, Colorado State University, Fort Collins, CO 80523-1373, USA (e-mail: zhenliang.zhang@colostate.edu; edwin.chong@colostate.edu; ali.pezeshki@colostate.edu).

W. Moran is with the Department of Electrical and Electronic Engineering, The University of Melbourne, Melbourne, VIC 3010, Australia (e-mail: wmoran@unimelb.edu.au).

[1]Our model for sequential hypothesis testing is different from the model that goes by a similar name, due to Wald [2]. In Wald's sequential hypothesis testing problem, there is a single decision maker, who tests the given hypotheses by sequentially collecting samples. The sample size is not fixed in advance. Instead, according to the pre-defined stoping rule, the decision maker stops sampling and then declares a hypothesis.

question in this setting is whether the agents in the social network can asymptotically learn the state of the world.

To illustrate the feedforward nature of the model we study, consider a customer having to decide whether or not to dine in a particular restaurant. Typically, this decision is made based on her own taste and also on the stated opinions of previous patrons. In this example, the customer is a node in the feedforward network. The private signal at this node represents the customer's own taste, while the received decisions from predecessor nodes represent the perceived opinions of previous patrons. Some previous patrons might not reveal their opinions or might expose erroneous versions of their opinions. The former is what we might call "erasure" of decisions, while the latter represents "flipping" of decisions. We will formalize these notions of erasure and flipping later. Similar examples along these lines include customers deciding whether or not to watch a particular movie and investors deciding whether or not to buy a certain asset. A comprehensive exposition of social learning can be found in [4].

### A. Related Work

The literature on hypothesis testing in decentralized networks is vast, spanning various disciplines including signal processing, game theory, information theory, economics, biology, physics, computer science, and statistics. Here we only review the relevant asymptotic learning results in the network structure relevant to this paper.

The research on our problem begins with a seminal paper by Cover [5], which considers the case where each node only observes the decision from its immediate previous node, i.e., $m_k = 1$ for all $k$. This structure is also known as a serial network or tandem network and has been studied extensively in [5]–[17]. We use $\mathbb{P}_j$ and $\pi_j$ to denote the probability measure and the prior probability associated with $H_j, j = 0, 1$, respectively. Cover [5] shows that if the (log)-likelihood ratio for each private signal $X_k$ is bounded almost surely, then using a sequence of likelihood ratio tests the (Bayesian) error probability

$$\mathbb{P}_e^k = \pi_0 \mathbb{P}_0(d_k = 1) + \pi_1 \mathbb{P}_1(d_k = 0)$$

does not converge in probability to 0 as $k \to \infty$. Conversely, if the likelihood ratio is unbounded, then the error probability converges to 0. In the case of unbounded likelihood ratios for the private signals, Veeravalli [12] shows that the error probability converges sub-exponentially with respect to the number $k$ of nodes in the case where the private signals are independent and follow identical Gaussian distribution. Tay *et al.* [14] show that the convergence of error probability is in general sub-exponential and derive a lower bound for the convergence rate of the error probability in the tandem network. Lobel *et al.* [15] derive a lower bound for the convergence rate in the case where each node learns randomly from one previous node (not necessarily its immediate predecessor). In the case of bounded likelihood ratios, Drakopoulos *et al.* [16] provide a non-Bayesian decision strategy, which leads to the convergence of the error probability.

Another extreme scenario is that each node can observe *all* the previous decisions; i.e., $m_k = k-1$ for all $k$. This scenario was first studied in the context of social learning [18], [19], where each node uses the Bayesian likelihood ratio test to make its decision. In the case of bounded likelihood ratios for the private signals, the authors of [18] and [19] show that the error probability does not converge to 0, which results in arriving at the wrong decision with positive probability. In [20], we show that in balanced binary trees, the decisions converge to the right decision even if the likelihood ratios of signals converge to 1 as the number of nodes increases. We further studied in [21] the convergence rate of the error probability in more general tree structures. In the case of unbounded likelihood ratios for the private signals, Smith and Sorensen [22] study this problem using martingales and show that the error probability converges to 0. Krishnamurthy [23], [24] studies this problem from the perspective of quickest time change detection. Acemoglu *et al.* [25] show that the nodes can asymptotically learn the underlying truth in more general network structures.

Most previous work including those reviewed above assume that the nodes and links are perfect. We study the sequential hypothesis testing problem when broadcasts are subject to random erasure or random flipping.

### B. Contributions

In this paper, we assume that each node uses a likelihood ratio test to generate its binary decision. We call the sequence of likelihood ratio tests a *decision strategy*. We want to know whether or not there exists a decision strategy such that the node decisions converge in probability to the underlying true hypothesis. We consider two classes of broadcast failures:

1) *Random erasure*: Each broadcasted decision is erased with a certain erasure probability, modeled by a binary erasure channel. If the decision broadcasted by a node is erased, then none of its successors will observe that decision.
2) *Random flipping*: Each broadcasted decision is flipped with a certain flipping probability, modeled by a binary symmetric channel. If the broadcasted decision of a node is flipped, then all the successors of that node observe that flipped decision.

For case 1, we show that if each node can only learn from a bounded number of immediate predecessors, i.e., there exists a constant $C$ such that $m_k \leq C$ for all $k$, then for any decision strategy, the error probability cannot converge to 0. We also show that if $m_k \to \infty$ as $k \to \infty$, then there exists a decision strategy such that the error probability converges to 0, even if the erasure probability converges to 1 (given that the convergence of the erasure probability is slower than a certain rate). In the case where an agent learns from all its predecessors, the convergence rate of the error probability is $\Theta(1/\sqrt{k})$. More specifically, we show that if the memory size $m_k = \Theta(k^\sigma)$, $\sigma \leq 1$, then the error probability decreases as $\Theta(1/k^{\min(\sigma,1/2)})$.

For case 2, we show that if each node can only learn from a bounded number of immediate predecessors, then for any decision strategy, the error probability cannot converge to 0. We also show that if each node can learn from *all* the previous nodes, i.e., $m_k = k-1$, then the error probability converges to 0 using the myopic decision strategy when the flipping probabilities are bounded away from $1/2$. In this case, we show that the error probability converges to 0 as $\Omega(1/k^2)$. In the case where the flipping probability converges to $1/2$, we derive a necessary condition on the convergence rate of the flipping probability (i.e., how fast it must converge) such that the error probability converges to 0. More specifically, we show that if there exists $p > 1$ such that the flipping probability converges to $1/2$ as $O(1/k(\log k)^p)$, then it is impossible that the error probability converges to 0. Therefore, only if the flipping probability converges as $\Omega(1/k(\log k)^p)$ for some $p \le 1$ can we hope for $\mathbb{P}_e^k \to 0$. Under this condition, we characterize explicitly the relationship between the convergence rate of the flipping probability and the convergence rate of the error probability.

## II. Preliminaries

We use $\mathbb{P}$ to denote the underlying probability measure. We use $\pi_j$ to denote the prior probability (assumed nonzero), $\mathbb{P}_j$ to denote the probability measure, and $\mathbb{E}_j$ to denote the conditional expectation associated with $H_j$, $j = 0, 1$. At stage $k$, node $a_k$ takes a measurement $X_k$ of the scene and makes a decision $d_k = 0$ or $d_k = 1$ about the prevailing hypothesis $H_0$ or $H_1$. It then broadcasts a potentially corrupted form $\hat{d}_k$ of that decision to its successors. Note that in case 1, if the decision is erased, it is equivalent to saying that the corrupted decision $\hat{d}_k$ is $e$, which is a message that carries no information and is not useful for decision-making. Inserting $e$ in place of erased messages allows us to unify the notation for cases 1 and 2. The decision $d_k$ of node $a_k$ is made based on the private signal $X_k$ and the sequence of corrupted decisions $\hat{D}_{m_k} = \{\hat{d}_1, \hat{d}_2, \ldots, \hat{d}_{m_k}\}$ received from the $m_k$ immediate predecessor nodes using a likelihood ratio test.

Our aim is to find a sequence of likelihood ratio tests such that the probability of making a wrong decision about the state of the world tends to 0 as $k \to \infty$; i.e.,

$$\lim_{k\to\infty} \mathbb{P}_e^k = \lim_{k\to\infty} (\pi_0 \mathbb{P}_0(d_k = 1) + \pi_1 \mathbb{P}_1(d_k = 0)) = 0.$$

Before proceeding, we introduce the following definitions and assumptions:

1) The private signal $X_k$ takes values in a set $S$, endowed with a $\sigma$-algebra $\mathcal{S}$. We assume that $X_k$ is independent of the broadcast history $\hat{D}_{m_k}$. Moreover, the $X_k$s are mutually independent and identically distributed with distribution $\mathbb{P}_j^X$, under $H_j$, $j = 0, 1$. (Note that $\mathbb{P}_j^X$ is a probability measure on the $\sigma$-algebra $\mathcal{S}$.) We assume that the underlying hypothesis, $H_0$ or $H_1$, does not change with $k$.
2) The two probability measures $\mathbb{P}_0^X$ and $\mathbb{P}_1^X$ are equivalent; i.e., they are absolutely continuous with respect to each other. In other words, if $A \in \mathcal{S}$, then $\mathbb{P}_0^X(A) = 0$ if and only if $\mathbb{P}_1^X(A) = 0$.
3) Let the likelihood ratio of a private signal $s \in S$ be
$$L_X(s) = \frac{d\mathbb{P}_1^X}{d\mathbb{P}_0^X}(s),$$
where $d\mathbb{P}_1^X / d\mathbb{P}_0^X$ denotes the Radon–Nikodym derivative (which is guaranteed to exist because of the assumption that the two measures are equivalent). We assume that the likelihood ratios for the private signals are unbounded; i.e., for any set $S' \subset S$ with probability 1 under the measure $(\mathbb{P}_0^X + \mathbb{P}_1^X)/2$, we have
$$\inf_{s\in S'} \frac{d\mathbb{P}_1^X}{d\mathbb{P}_0^X}(s) = 0$$
and
$$\sup_{s\in S'} \frac{d\mathbb{P}_1^X}{d\mathbb{P}_0^X}(s) = \infty.$$
4) Suppose that $\theta$ is the underlying truth. Let $\bar{b}_k = \mathbb{P}(\theta = H_1 | X_k)$, which we call the *private belief* of $a_k$. By Bayes' rule, we have
$$\bar{b}_k = \left(1 + \frac{\pi_0}{\pi_1} \frac{1}{L_X(X_k)}\right)^{-1}. \tag{1}$$
5) Recall that node $a_k$ observes $m_k$ decisions $\hat{D}_{m_k}$ from its immediate predecessors. Let $p_j^k$ be the conditional probability mass function of $\hat{D}_{m_k}$ under $H_j$, $j = 0, 1$. The likelihood ratio of a realization $\mathcal{D}_{m_k}$ is
$$L_D^k(\mathcal{D}_{m_k}) = \frac{p_1^k(\mathcal{D}_{m_k})}{p_0^k(\mathcal{D}_{m_k})} = \frac{\mathbb{P}_1(\hat{D}_{m_k} = \mathcal{D}_{m_k})}{\mathbb{P}_0(\hat{D}_{m_k} = \mathcal{D}_{m_k})}.$$
6) Let $b_k = \mathbb{P}(\theta = H_1 | \hat{D}_{m_k})$, which we call the *public belief* of $a_k$. We have
$$b_k = \left(1 + \frac{\pi_0}{\pi_1} \frac{1}{L_D^k(\hat{D}_{m_k})}\right)^{-1}. \tag{2}$$
7) Each node $a_k$ makes its decision using its own measurement and the observed decisions based on a likelihood ratio test with a threshold $t_k > 0$:
$$d_k = \begin{cases} 1 & \text{if } L_X(X_k) L_D^k(\hat{D}_{m_k}) > t_k, \\ 0 & \text{if } L_X(X_k) L_D^k(\hat{D}_{m_k}) \le t_k. \end{cases}$$
If $t_k = \pi_0/\pi_1$, then this test becomes the maximum a-posteriori probability (MAP) test, in which case the probability of error is locally minimized for node $a_k$. If $t_k = 1$, then the test becomes the maximum-likelihood (ML) test. If the prior probabilities are equal, then these two tests are identical. A decision strategy $\mathbb{T}$ is a sequence of likelihood ratio tests with thresholds $\{t_k\}_{k=1}^\infty$. Given a decision strategy, the decision sequence $\{d_k\}_{k=1}^\infty$ is a well-defined stochastic process.
8) We say that the system *asymptotically learns* the underlying true hypothesis with decision strategy $\mathbb{T}$ if
$$\lim_{k\to\infty} \mathbb{P}(d_k = \theta) = 1.$$
In other words, the probability of making a wrong decision goes to 0, i.e., $\lim_{k\to\infty} \mathbb{P}_e^k = 0$. The question we are interested in is this: In each of the two classes of failures, is there a decision strategy such that the system asymptotically learns the underlying true hypothesis?

## III. Random Erasure

In this section, we consider the sequential hypothesis testing problem in the presence of random erasures, modeled by binary erasure channels. Recall that the binary message $d_k$ is the input to a binary erasure channel and $\hat{d}_k$ is the output, which is either equal to $d_k$ (no erasure) or is equal to a symbol $e$ that represents the occurrence of an erasure. The erasure channel matrix at stage $k$ is given by $\mathbb{P}(\hat{d}_k = i|d_k = j)$, $j = 0, 1$ and $i = j, e$. Recall that each node $a_k$ observes $m_k$ immediate previous broadcasted decisions. We divide our analysis into two scenarios: *A*) $\{m_k\}$ is bounded above by a positive constant; *B*) $m_k$ goes to infinity as $k \to \infty$.

### A. Bounded Memory

*Theorem 1:* Suppose that there exists $C$ and $\epsilon > 0$ such that for all $k$, $m_k \le C$ and $\mathbb{P}(\hat{d}_k = e|d_k = j) \in [\epsilon, 1-\epsilon]$ for $j = 0, 1$. Then, there does not exist a decision strategy such that the error probability converges to 0.

*Proof:* We first prove this claim for the special case of the tandem network, where $m_k = 1$ for all $k$. For each node $a_k$, with a nonzero probability $\mathbb{P}(\hat{d}_k = e|d_k = j)$, the decision $d_{k-1} = j$ of the immediate predecessor is erased and $a_k$ makes a decision based only on its own private signal $X_k$. We use $\mathcal{E}_k$ to denote this event. Conditioned on $\mathcal{E}_k$, we claim that the error probability as a sequence of $k$,

$$\begin{aligned}
&\mathbb{P}(d_k \neq \theta|\mathcal{E}_k)\\
&= \pi_0 \mathbb{P}_0(d_k = 1|\mathcal{E}_k) + \pi_1 \mathbb{P}_1(d_k = 0|\mathcal{E}_k)\\
&= \pi_0 \mathbb{P}_0(L_X(X_k) > t_k) + \pi_1 \mathbb{P}_1(L_X(X_k) \le t_k),
\end{aligned}$$

is bounded away from 0. We prove the above claim by contradiction. Suppose that there exists a decision strategy with threshold sequence $\{t_k\}$ such that $\mathbb{P}(d_k \neq \theta|\mathcal{E}_k) \to 0$ as $k \to \infty$. Then, we must have $\mathbb{P}_1(L_X(X_k) \le t_k) \to 0$ because $\pi_1$ is positive. Because $\mathbb{P}_0^X$ and $\mathbb{P}_1^X$ are equivalent measures, we have $\mathbb{P}_0(L_X(X_k) \le t_k) \to 0$. Hence we have $\mathbb{P}_0(L_X(X_k) > t_k) \to 1$. Therefore, $\mathbb{P}(d_k \neq \theta|\mathcal{E}_k)$ does not converge to 0.

We use $\mathcal{E}_k^{\mathcal{C}}$ to denote the complement event of $\mathcal{E}_k$. By the Law of Total Probability, we have

$$\begin{aligned}
\mathbb{P}_e^k &= \mathbb{P}(\mathcal{E}_k)\mathbb{P}(d_k \neq \theta|\mathcal{E}_k) + \mathbb{P}(\mathcal{E}_k^{\mathcal{C}})\mathbb{P}(d_k \neq \theta|\mathcal{E}_k^{\mathcal{C}})\\
&\ge \mathbb{P}(\mathcal{E}_k)\mathbb{P}(d_k \neq \theta|\mathcal{E}_k).
\end{aligned}$$

Because $\mathbb{P}(\mathcal{E}_k) \ge \epsilon$, we conclude that the error probability does not converge to 0.

We can now generalize this proof to the case of a general bounded $m_k$ sequence. Let $\mathcal{E}_k$ be the event that $a_k$ receives $m_k$ erased symbols $e$. Then, the probability $\mathbb{P}(\mathcal{E}_k)$ is bounded below according to

$$\mathbb{P}(\mathcal{E}_k) \ge \left( \min_{\substack{j=0,1\\ m=k-1,\dots,k-m_k}} \mathbb{P}(\hat{d}_m = e|d_m = j) \right)^{m_k} \ge \epsilon^{m_k}.$$

We have already shown that given this event the error probability does not converge to 0. Using the Law of Total Probability, It is easy to see that the error probability does not converge to 0. ■

*Remark 1:* We use $\mathbb{P}(\hat{d}_k = e|d_k = j) \in [\epsilon, 1-\epsilon]$ for $j = 0, 1$ to mean that the erasure probability $\mathbb{P}(\hat{d}_k = e|d_k = j)$ is bounded away from 0 and 1.

This result is straightforward to understand. If the memory sizes are bounded for all nodes, then for each node, there exists a positive probability such that all the decisions received from its immediate predecessors are erased, in which case the node has to make a decision based on its own measurement. The error probability cannot converge to 0 because of the equivalent-measure assumption.

### B. Unbounded Memory

Suppose that each node $a_k$ observes $m_k$ immediate previous decisions. In this section, we deal with the case where $m_k$ is unbounded.[2] More specifically, we consider the case where $m_k$ goes to infinity. We first consider the case where the erasure probabilities are bounded away from 1. We have the following result.

*Theorem 2:* Suppose that $m_k$ goes to infinity as $k \to \infty$ and there exists $\epsilon > 0$ such that for all $j = 0, 1$ and for all $k$, $\mathbb{P}(\hat{d}_k = e|d_k = j) \le 1 - \epsilon$. Then, there exists a decision strategy such that the error probability converges to 0.

*Proof:* We prove this result by constructing a certain tandem network within the original network using a *backward-searching scheme*. The scheme is the following: Consider node $a_k$ in the original network. Let $n_k$ be the largest integer such that each node in the sequence $\{a_{k-n_k^2}, a_{k-n_k^2-1}, \dots, a_k\}$ of $n_k^2 + 1$ nodes has a memory size that is greater than or equal to $n_k$. Note that an $n_k$ satisfying this condition is guaranteed to exist. Moreover, because $m_k$ goes to infinity as $k \to \infty$, we have $n_k \to \infty$ as $k \to \infty$. Consider the event that $a_k$ receives at least one decision $j$, which is not erased, from $\{a_{k-n_k}, \dots, a_{k-1}\}$, its $n_k$ immediate predecessors. The probability of this event is at least

$$1 - \max_{\substack{j=0,1\\ m=k-n_k,\dots,k-1}} \mathbb{P}(\hat{d}_m = e|d_m = j)^{n_k},$$

which is bounded below by $1 - (1-\epsilon)^{n_k}$ by the assumption on the erasure probabilities. We denote the node that sends the unerased decision by $a_{k_1}$. Similarly, with a certain probability, $a_{k_1}$ receives at least one decision, which is not erased, from its $n_k$ immediate predecessors. Recursively, with a certain probability, we can construct a tandem network with length $n_k$ using nodes from among the $n_k^2+1$ nodes above within the original network. Let $\mathcal{E}_k$ be the event that such a tandem network exists. The probability $\mathbb{P}(\mathcal{E}_k)$ is at least $(1-(1-\epsilon)^{n_k})^{n_k}$. Recall that $\lim_{k\to\infty} n_k = \infty$, which implies that

$$\lim_{k\to\infty} (1-(1-\epsilon)^{n_k})^{n_k} = 1.$$

Hence we have

$$\lim_{k\to\infty} \mathbb{P}(\mathcal{E}_k) = 1.$$

[2]The assumption that $m_k$ is unbounded is not sufficiently strong to guarantee the convergence of error probability to 0. An example is that the memory size $m_k$ equals $\sqrt{k}$ if $\sqrt{k}$ is an integer and it equals 1 otherwise. In this case, we can use a similar argument as that in the proof of Theorem 1 to show that the error probability does not converge to 0.

Conditioned on $\mathcal{E}_k$, by using the strategy $\mathbb{T}$ consisting of a sequence of likelihood ratio tests with monotone thresholds described in [5], we can get the conditional convergence of the error probability, given $\mathcal{E}_k$, to 0. We can also use the equilibrium strategy described in [15]. Therefore, by the Law of Total Probability, we have

$$\begin{aligned} &\lim_{k\to\infty} \mathbb{P}(d_k \neq \theta) \\ &= \lim_{k\to\infty} \left(\mathbb{P}(d_k \neq \theta|\mathcal{E}_k)\mathbb{P}(\mathcal{E}_k) + \mathbb{P}(d_k \neq \theta|\mathcal{E}_k^{\mathcal{C}})(1-\mathbb{P}(\mathcal{E}_k))\right) \\ &\leq \lim_{k\to\infty} \left(\mathbb{P}(d_k \neq \theta|\mathcal{E}_k) + (1-\mathbb{P}(\mathcal{E}_k))\right) = 0. \end{aligned} \tag{3}$$

■

Note that given a strategy, the convergence rate for the error probability in this case depends on how fast $\mathbb{P}(\mathcal{E}_k)$ converges to 1 and how fast $\mathbb{P}(d_k \neq \theta|\mathcal{E}_k)$ converges to 0.

First let us consider the convergence rate of $\mathbb{P}(\mathcal{E}_k)$. Obviously this convergence rate depends on the convergence rate of $n_k$. Moreover, the convergence rate of $n_k$ depends on the convergence rate of $m_k$. For example, if $m_k$ goes to infinity extremely slowly, then $n_k$ grows extremely slowly with respect to $k$, which means that $\mathbb{P}(\mathcal{E}_k)$ converges to 1 extremely slowly with respect to $k$. Next we assume that $m_k$ increases as $\Theta(k^\sigma)$, where $\sigma \leq 1$. We first establish a relationship between the convergence rate of $m_k$ and the convergence rate of $n_k$ when using the backward-searching scheme.

*Proposition 1:* Suppose that $m_k = \Theta(k^\sigma)$ where $\sigma \leq 1$. Then, we have

$$n_k = \begin{cases} \Theta(\sqrt{k}) & \text{if } \sigma \geq 1/2, \\ \Theta(k^\sigma) & \text{if } \sigma < 1/2. \end{cases}$$

*Proof:* Suppose that we can form a tandem network with length $n_k$ within the original network. Recall that $n_k$ is the largest integer such that each node in the sequence $\{a_{k-n_k^2}, a_{k-n_k^2-1}, \dots, a_k\}$ of $n_k^2+1$ nodes has a memory size that is greater than or equal to $n_k$. Therefore, the memory size $m_{k-n_k^2}$ of $a_{k-n_k^2}$ must be larger than or equal to $n_k$ by assumption. Hence we have

$$m_{k-n_k^2} = (k-n_k^2)^\sigma \geq n_k.$$

Moreover, the memory size $m_{k-(n_k+1)^2}$ of $a_{k-(n_k+1)^2}$ must be strictly smaller than $n_k+1$ (otherwise we can construct a tandem network with length $n_k+1$). Hence we have

$$m_{k-(n_k+1)^2} = (k-(n_k+1)^2)^\sigma < n_k+1.$$

From the above two inequalities, we easily obtain the desired asymptotic rates for $n_k$. ■

*Remark 2:* Note that if $\sigma < 1/2$, then the scaling law of $n_k$ is identical to that of $m_k$: The faster the scaling of $m_k$, the faster the scaling of $n_k$ also. However, for $\sigma \geq 1/2$, the scaling law of $n_k$ "saturates" at $\sqrt{k}$, no matter how fast $m_k$ scales.

We have derived the convergence rate for $n_k$. Recall that $\mathbb{P}(\mathcal{E}_k)$ converges to 1 at least in the rate of $\Theta(n_k(1-\epsilon)^{n_k})$ (by expanding the term $(1-(1-\epsilon)^{n_k})^{n_k}$ and keeping the dominating term). From this fact and Proposition 1, we derive the convergence rate for $\mathbb{P}(\mathcal{E}_k)$.

*Corollary 1:* Suppose that $m_k = \Theta(k^\sigma)$ where $\sigma \leq 1$. Then, we have

$$1-\mathbb{P}(\mathcal{E}_k) = \begin{cases} O(\sqrt{k}(1-\epsilon)^{\sqrt{k}}) & \text{if } \sigma \geq 1/2, \\ O(k^\sigma(1-\epsilon)^{k^\sigma}) & \text{if } \sigma < 1/2. \end{cases}$$

Second, let us consider the convergence rate of $\mathbb{P}(d_k \neq \theta|\mathcal{E}_k)$. Recall that $\mathcal{E}_k$ denotes the event that a tandem network with length $n_k$ exists. Conditioned on $\mathcal{E}_k$, if we use the the equilibrium strategy[3] described in [15], then it has been shown that the error probability converges to 0 as $\Theta(1/n_k)$, with appropriate assumptions on the distributions of the private signal. From this fact and Proposition 1, we derive the convergence rate for $\mathbb{P}(d_k \neq \theta|\mathcal{E}_k)$.

*Corollary 2:* Suppose that $m_k = \Theta(k^\sigma)$ where $\sigma \leq 1$. Then, we have

$$\mathbb{P}(d_k \neq \theta|\mathcal{E}_k) = \begin{cases} \Theta(1/\sqrt{k}) & \text{if } \sigma \geq 1/2, \\ \Theta(1/k^\sigma) & \text{if } \sigma < 1/2. \end{cases}$$

Notice that the convergence rate of $\mathbb{P}(d_k \neq \theta|\mathcal{E}_k)$ is much smaller than that of $\mathbb{P}(\mathcal{E}_k)$. Moreover by (3), the convergence rate of $\mathbb{P}(d_k \neq \theta)$ depends on the smaller of the convergence rates of $\mathbb{P}(d_k \neq \theta|\mathcal{E}_k)$ and $\mathbb{P}(\mathcal{E}_k)$. We derive the convergence rate for the error probability as follows.

*Corollary 3:* Suppose that $m_k = \Theta(k^\sigma)$ where $\sigma \leq 1$. Then, we have

$$\mathbb{P}(d_k \neq \theta) = \begin{cases} \Theta(1/\sqrt{k}) & \text{if } \sigma \geq 1/2, \\ \Theta(1/k^\sigma) & \text{if } \sigma < 1/2. \end{cases}$$

We have considered the situation where the erasure probabilities are bounded away from 1. Now consider the case where the erasure probability $\mathbb{P}(\hat{d}_k = e|d_k = j)$ converges to 1.

*Theorem 3:* Suppose that $\mathbb{P}(\hat{d}_k = e|d_k = j) \to 1$ and there exists $\epsilon > 1$ and $c > 0$ such that $\mathbb{P}(\hat{d}_k = e|d_k = j) \leq (cn_k)^{-\epsilon/n_k}$. Then, there exists a decision strategy such that the error probability converges to 0.

*Proof:* We use the scheme described in the proof of Theorem 2. The probability that a tandem network with length $n_k$ exists is at least $(1-((cn_k)^{-\epsilon/n_k})^{n_k})^{n_k} = (1-(cn_k)^{-\epsilon})^{n_k}$, which converges to 1 as $k \to \infty$. Using the same arguments as those in the proof of Theorem 2, we can show that the error probability converges to 0. ■

As an example, we consider the situation where each node observes *all* the previous decisions; i.e, $m_k = k-1$ for all $k$. In this case, it is easy to show that using the backward-searching scheme, with a certain probability, we can form a tandem network with length $n_k = \lfloor\sqrt{k-1}\rfloor$. Suppose that the erasure probabilities are bounded away from 1. Then, the error probability converges to 0 as $\Theta(1/\sqrt{k})$. Moreover, the error probability converges to 0 even if the erasure probability converges to 1, provided that $\mathbb{P}(\hat{d}_k = e|d_k = j) \leq (cn_k)^{-\epsilon/n_k}$.

[3] Note that this equilibrium strategy is *not* the only strategy such that the error probability converges to 0 in a tandem network.

## IV. Random Flipping

We study in this section the sequential hypothesis testing problem with random flipping, modeled by a binary symmetric channel. Recall that $d_k$ is the input to a binary symmetric channel and $\hat{d}_k$ is the output, which is either equal to $d_k$ (no flipping) or is equal to its complement $1-d_k$ (flipping). The channel matrix is given by $\mathbb{P}(\hat{d}_k = i | d_k = j)$, $i, j = 0, 1$. We assume that $\mathbb{P}(\hat{d}_k = 1 | d_k = 0) = \mathbb{P}(\hat{d}_k = 0 | d_k = 1) = q_k$, where $q_k$ denotes the probability of a flip. The assumption of symmetry is for simplicity only, and all results obtained in this section can be generalized easily to a general binary communication channel with unequal flipping probabilities, i.e., $\mathbb{P}(\hat{d}_k = 1 | d_k = 0) \neq \mathbb{P}(\hat{d}_k = 0 | d_k = 1)$. We assume that each node $a_k$ knows the probabilities of flipping associated with the corrupted decisions $\hat{D}_{m_k}$ received from its predecessors.

### A. Bounded Memory

*Theorem 4:* Suppose that there exists $C$ and $\epsilon > 0$ such that for all $k$, $m_k \leq C$ and $q_k \in [\epsilon, 1-\epsilon]$. Then, there does not exist a decision strategy such that the error probability converges to 0.

*Proof:* We first prove this theorem in the case where each node observes the immediate previous node; i.e., $m_k = 1$ for all $k$. Node $a_k$ makes a decision $d_k$ based on its private signal $X_k$ and the decision $\hat{d}_{k-1}$ from its immediate predecessor. Recall that $q_k = \mathbb{P}(\hat{d}_k = 1 | d_k = 0) = \mathbb{P}(\hat{d}_k = 0 | d_k = 1)$. The likelihood ratio test at stage $k$ (with a threshold $t_k > 0$) is

$$d_k = \begin{cases} 1 & \text{if } L_X(X_k) L_D^k(\hat{d}_{k-1}) > t_k, \\ 0 & \text{if } L_X(X_k) L_D^k(\hat{d}_{k-1}) \leq t_k, \end{cases}$$

where for each $j_{k-1} = 0, 1$

$$L_D^k(j_{k-1}) = \frac{p_1^k(j_{k-1})}{p_0^k(j_{k-1})} = \frac{\mathbb{P}_1(\hat{d}_{k-1} = j_{k-1})}{\mathbb{P}_0(\hat{d}_{k-1} = j_{k-1})},$$

and $\mathbb{P}_j(\hat{d}_{k-1} = j_{k-1})$, $j = 0, 1$ is given by

$$\begin{aligned} \mathbb{P}_j(\hat{d}_{k-1} = j_{k-1}) &= q_k(1 - \mathbb{P}_j(d_{k-1} = j_{k-1})) \\ &\quad + (1-q_k)\mathbb{P}_j(d_{k-1} = j_{k-1}) \\ &= q_k + (1 - 2q_k)\mathbb{P}_j(d_{k-1} = j_{k-1}). \qquad (4) \end{aligned}$$

Let $t_k(\hat{d}_{k-1}) = t_k / L_D^k(\hat{d}_{k-1})$ be the testing threshold for $L_X(X_k)$ when $\hat{d}_{k-1}$ is received. Then, the likelihood ratio test can be rewritten as

$$d_k = \begin{cases} 1 & \text{if } L_X(X_k) > t(\hat{d}_{k-1}), \\ 0 & \text{if } L_X(X_k) \leq t(\hat{d}_{k-1}). \end{cases}$$

From (4), we notice that $\mathbb{P}_j(\hat{d}_{k-1})$ depends linearly on $\mathbb{P}_j(d_{k-1})$. Without loss of generality, henceforth we assume that $q_k \leq 1/2$.[4] It is obvious that $t_k(0) \geq t_k(1)$ because $L_D^k(j) = \mathbb{P}_1(\hat{d}_{k-1} = j)/\mathbb{P}_0(\hat{d}_{k-1} = j)$ is non-decreasing in $j$. Therefore, the likelihood ratio test becomes

$$d_k = \begin{cases} 1 & \text{if } L_X(X_k) > t_k(0), \\ 0 & \text{if } L_X(X_k) \leq t_k(1), \\ \hat{d}_{k-1} & \text{otherwise}, \end{cases}$$

and we can write the Type I and Type II error probabilities, denoted by $\mathbb{P}_0(d_k = 1)$ and $\mathbb{P}_1(d_k = 0)$, respectively, as follows:

$$\begin{aligned} \mathbb{P}_0(d_k = 1) &= \mathbb{P}_0(L_X(X_k) > t_k(0))\mathbb{P}_0(\hat{d}_{k-1} = 0) \\ &\quad + \mathbb{P}_0(L_X(X_k) > t_k(1))\mathbb{P}_0(\hat{d}_{k-1} = 1) \end{aligned}$$

and

$$\begin{aligned} \mathbb{P}_1(d_k = 0) &= \mathbb{P}_1(L_X(X_k) \leq t_k(0))\mathbb{P}_1(\hat{d}_{k-1} = 0) \\ &\quad + \mathbb{P}_1(L_X(X_k) \leq t_k(1))\mathbb{P}_1(\hat{d}_{k-1} = 1). \end{aligned}$$

The total error probability at stage $k$ is

$$\begin{aligned} \mathbb{P}_e^k &= \pi_0 \mathbb{P}_0(d_k = 1) + \pi_1 \mathbb{P}_1(d_k = 0) \\ &= \pi_0(\mathbb{P}_0(L_X(X_k) > t_k(0)) \\ &\quad + \mathbb{P}_0(t_k(1) < L_X(X_k) \leq t_k(0))\mathbb{P}_0(\hat{d}_{k-1} = 1)) \\ &\quad + \pi_1(\mathbb{P}_1(t_k(1) < L_X(X_k) \leq t_k(0))\mathbb{P}_1(\hat{d}_{k-1} = 0) \\ &\quad + \mathbb{P}_1(L_X(X_k) \leq t_k(1))). \end{aligned}$$

We prove the claim by contradiction. Suppose that there exists a strategy such that $\mathbb{P}_e^k \to 0$ as $k \to \infty$. Then, we must have $\mathbb{P}_0(L_X(X_k) > t_k(0)) \to 0$ and $\mathbb{P}_1(L_X(X_k) \leq t_k(1)) \to 0$. Recall that $\mathbb{P}_0^X$ and $\mathbb{P}_1^X$ are equivalent measures. Hence we have $\mathbb{P}_1(L_X(X_k) > t_k(0)) \to 0$ and $\mathbb{P}_0(L_X(X_k) \leq t_k(1)) \to 0$. These imply that $\mathbb{P}_j(t_k(1) < L_X(X_k) \leq t_k(0)) \to 1$ for $j = 0, 1$. But

$$\begin{aligned} \mathbb{P}_j(\hat{d}_{k-1} = 1 - j) &= q_k(1 - \mathbb{P}_j(d_{k-1} = 1 - j)) \\ &\quad + (1 - q_k)\mathbb{P}_j(d_{k-1} = 1 - j) \\ &= q_k + (1 - 2q_k)\mathbb{P}_j(d_{k-1} = 1 - j), \end{aligned}$$

which is bounded below by $q_k$. Hence $\mathbb{P}_e^k$ is also bounded below away from 0 in the asymptotic regime. This contradiction implies that $\mathbb{P}_e^k$ does not converge to 0. The proof for the general bounded memory case is similar and is given in Appendix A. ∎

### B. Unbounded Memory

In this section, we consider the case where $a_k$ can observe all its predecessors; i.e., $m_k = k - 1$. We will show that using the myopic decision strategy, the error probability converges to 0 in the presence of random flipping when the flipping probabilities are bounded away from $1/2$. In the case where the flipping probability converges to $1/2$, we derive a necessary condition on the convergence rate of the flipping probability such that the error probability converges to 0. Moreover, we precisely describe the relationship between the convergence rate of the flipping probability and the convergence rate of the error probability.

If we state the conditions on the private signal distributions in a symmetric way, then it suffices to consider the case when

[4] Note that the system is symmetric with respect to $q_k = 1/2$. For example, if the probability of flipping is 1, i.e., $q_k = 1$, then the receiver can revert the received decision back since it knows the predecessor always 'lies.'

the true hypothesis is $H_0$. In this case, our aim is to show that the Type I error probability converges to 0, i.e., $\mathbb{P}_0(d_k = 1) \to 0$. We consider the myopic decision strategy; i.e., the decision made by the $k$th node is on the basis of the MAP test. Again, the corruption from $d_k$ to $\hat{d}_k$ is in the form of a binary symmetric channel with flipping probability denoted by $q_k$. Without loss of generality, we assume that $q_k \leq 1/2$ (because of symmetry). We define the *public likelihood ratio* of $\mathcal{D}_k = (j_1, j_2, \ldots, j_k)$ to be

$$L_k(\mathcal{D}_k) = \frac{p_1^k(\mathcal{D}_k)}{p_0^k(\mathcal{D}_k)} = \frac{\mathbb{P}_1(\hat{D}_k = \mathcal{D}_k)}{\mathbb{P}_0(\hat{D}_k = \mathcal{D}_k)}.$$

We will consider two cases:

1) The flipping probabilities are bounded away from $1/2$ for all $k$; i.e., there exists $c > 0$ such that $q_k \leq 1/2 - c$ for all $k$. This ensures that the corrupted decision still contains some useful information about the true hypothesis. We call this the case of *uniformly informative nodes*.
2) The flipping probabilities $q_k$ converge to $1/2$; i.e., $q_k \to 1/2$ as $k \to \infty$. This means that the broadcasted decisions become increasingly uninformative as we move towards the latter nodes. We call this the case of *asymptotically uninformative nodes*.

*1) Uniformly informative nodes:* We first show that the error probability converges to 0. Recall that $\bar{b}_k = \mathbb{P}(\theta = H_1 | X_k)$ denotes the private belief given by signal $X_k$. Let $(\mathbb{G}_0, \mathbb{G}_1)$ be the conditional distributions of the private belief $\bar{b}_k$:

$$\mathbb{G}_j(r) = \mathbb{P}_j(\bar{b}_k \leq r).$$

Note that $\mathbb{G}_j$ does not depend on $k$ because the $X_k$s are identically distributed. These distributions exhibit two important properties:

a) *Proportionality*: This property is easy to get from Bayes' rule: for all $r \in (0, 1)$, we have
$$\frac{d\mathbb{G}_1}{d\mathbb{G}_0}(r) = \frac{r}{1-r},$$
where $d\mathbb{G}_1/d\mathbb{G}_0$ is the Radon-Nikodym derivative of their associated probability measures.
b) *Dominance*: $\mathbb{G}_1(r) < \mathbb{G}_0(r)$ for all $r \in (0, 1)$, and $\mathbb{G}_j(0) = 0$ and $\mathbb{G}_j(1) = 1$ for $j = 0, 1$. Moreover, $\mathbb{G}_1(r)/\mathbb{G}_0(r)$ is monotone non-decreasing as a function of $r$.

We note that the dominance property can be shown using Assumption 3) and the details of the proof is omitted.

We define an increasing sequence $\{\mathcal{F}_k\}$ of $\sigma$-algebras as follows:

$$\mathcal{F}_k = \sigma\langle X_1, X_2, \ldots, X_k; \hat{d}_1, \hat{d}_2, \ldots, \hat{d}_k\rangle.$$

Evidently $\hat{d}_k$ and $L_k(\hat{D}_k)$ are adapted to this sequence of $\sigma$-algebras. Moreover, given $\hat{D}_{k-1} = \{\hat{d}_1, \hat{d}_2, \ldots, \hat{d}_{k-1}\}$ and $X_k$, the decision $d_k$ is completely determined. Therefore, $d_k$ is also adapted to this sequence of $\sigma$-algebras.

*Lemma 1:* Under hypothesis $H_0$, the public likelihood ratio sequence $\{L_k(\hat{D}_k)\}$ is a martingale with respect to $\{\mathcal{F}_k\}$ and $L_k(\hat{D}_k)$ converges to a finite limit almost surely.

*Proof:* The expectation of $L_{k+1}(\hat{D}_{k+1})$ conditioned on $H_0$ and $\mathcal{F}_k$ is

$$\begin{aligned}
\mathbb{E}_0[L_{k+1}(\hat{D}_{k+1})|\mathcal{F}_k] &= \sum_{\hat{d}_{k+1}=0,1} \mathbb{P}_0(\hat{d}_{k+1}|\mathcal{F}_k) L_{k+1}(\hat{D}_{k+1}) \\
&= \sum_{\hat{d}_{k+1}=0,1} \mathbb{P}_0(\hat{d}_{k+1}|\mathcal{F}_k) L_k(\hat{D}_k) \frac{\mathbb{P}_1(\hat{d}_{k+1}|\mathcal{F}_k)}{\mathbb{P}_0(\hat{d}_{k+1}|\mathcal{F}_k)} \\
&= L_k(\hat{D}_k) \sum_{\hat{d}_{k+1}=0,1} \mathbb{P}_0(\hat{d}_{k+1}|\mathcal{F}_k) \frac{\mathbb{P}_1(\hat{d}_{k+1}|\mathcal{F}_k)}{\mathbb{P}_0(\hat{d}_{k+1}|\mathcal{F}_k)} \\
&= L_k(\hat{D}_k).
\end{aligned}$$

Moreover, note that

$$\int |L_1(\hat{D}_1)| \, d\mathbb{P}_0 = 1 < \infty.$$

Since $L_k(\hat{D}_k)$ a non-negative martingale, by Doob's martingale convergence theorem [26], it converges almost surely to a finite limit. ■

Let $L_\infty$ be the almost sure limit of $L_k(\hat{D}_k)$ conditioned on $H_0$, and note that $L_\infty < \infty$ almost surely. This claim holds for both cases 1 and 2. By (2), we know that the public belief $b_k < 1$ almost surely. The implication is that the public belief cannot go completely wrong. Moreover, for case 1, we can show that the public likelihood ratio converges to 0 almost surely.

*Lemma 2:* Suppose that the flipping probabilities are bounded away from $1/2$. Then under $H_0$, we have $L_\infty = 0$ almost surely.

*Proof:* For the public likelihood ratio, we have the following recursion:

$$\begin{aligned}
L_{k+1}(\hat{D}_{k+1}) &= \frac{\mathbb{P}_1(\hat{D}_{k+1})}{\mathbb{P}_0(\hat{D}_{k+1})} \\
&= \frac{\mathbb{P}_1(\hat{d}_{k+1}|\hat{D}_k)}{\mathbb{P}_0(\hat{d}_{k+1}|\hat{D}_k)} L_k(\hat{D}_k).
\end{aligned} \tag{5}$$

Consider the event $A = \{L_\infty > 0\}$. On $A$, we have

$$\frac{\mathbb{P}_1(\hat{d}_{k+1}|\hat{D}_k)}{\mathbb{P}_0(\hat{d}_{k+1}|\hat{D}_k)} \to 1, \tag{6}$$

almost everywhere. Now

$$\begin{aligned}
\frac{\mathbb{P}_1(\hat{d}_{k+1}|\hat{D}_k)}{\mathbb{P}_0(\hat{d}_{k+1}|\hat{D}_k)} &= \frac{\sum_{d_{k+1}} \mathbb{P}_1(d_{k+1}|\hat{D}_k)\mathbb{P}(\hat{d}_{k+1}|d_{k+1})}{\sum_{d_{k+1}} \mathbb{P}_0(d_{k+1}|\hat{D}_k)\mathbb{P}(\hat{d}_{k+1}|d_{k+1})} \\
&= \frac{\mathbb{P}_1(d_{k+1}|\hat{D}_k)(1-2q_k) + q_k}{\mathbb{P}_0(d_{k+1}|\hat{D}_k)(1-2q_k) + q_k}.
\end{aligned} \tag{7}$$

Equation (7) together with (6) implies

$$\frac{\mathbb{P}_1(d_{k+1}|\hat{D}_k)}{\mathbb{P}_0(d_{k+1}|\hat{D}_k)} \to 1,$$

or $\mathbb{P}_j(d_{k+1}|\hat{D}_k) \to 0$ for $j = 0, 1$, almost everywhere on $A$. We note that another possible situation is that there exists a subsequence of $\{\frac{\mathbb{P}_1(d_{k+1}|\hat{D}_k)}{\mathbb{P}_0(d_{k+1}|\hat{D}_k)}\}$ that converges to 1 and for its complement subsequence, we have $\mathbb{P}_j(d_{k+1}|\hat{D}_k) \to 0$ for

$j = 0, 1$, almost everywhere on $A$. However, the proof for this situation is similar with others and it is omitted.

We will show that $A$ has probability 0. Suppose that there exists $\omega \in A$ such that

$$\lim_{k\to\infty} \frac{\mathbb{P}_1(d_{k+1} = d_{k+1}(\omega)|\hat{D}_k = \hat{D}_k(\omega))}{\mathbb{P}_0(d_{k+1} = d_{k+1}(\omega)|\hat{D}_k = \hat{D}_k(\omega))} = 1.$$

Note that $d_{k+1}(\omega) = 0$ or 1. Without loss of generality, consider the case where $d_{k+1}(\omega) = 0$, we have

$$\lim_{k\to\infty} \frac{\mathbb{P}_1(d_{k+1} = 0|\hat{D}_k = \hat{D}_k(\omega))}{\mathbb{P}_0(d_{k+1} = 0|\hat{D}_k = \hat{D}_k(\omega))} = 1. \tag{8}$$

Note that the statement $d_{k+1} = 0$ is equivalent to

$$L_X(X_{k+1})L_k(\hat{D}_k) \leq \frac{\pi_0}{\pi_1}.$$

Because of the independence between $X_{k+1}$ and $\hat{D}_k$, we obtain

$$\begin{aligned}
&\mathbb{P}_j(d_{k+1} = 0|\hat{D}_k = \hat{D}_k(\omega)) = \\
&\mathbb{P}_j\left(L_X(X_{k+1})L_k(\hat{D}_k) \leq \frac{\pi_0}{\pi_1}\middle| \hat{D}_k = \hat{D}_k(\omega)\right) = \\
&\mathbb{P}_j\left(L_X(X_{k+1})L_k(\hat{D}_k(\omega)) \leq \frac{\pi_0}{\pi_1}\right).
\end{aligned}$$

Thus (8) is equivalent to

$$\lim_{k\to\infty} \frac{\mathbb{P}_1(L_X(X_{k+1})L_k(\hat{D}_k(\omega)) \leq \frac{\pi_0}{\pi_1})}{\mathbb{P}_0(L_X(X_{k+1})L_k(\hat{D}_k(\omega)) \leq \frac{\pi_0}{\pi_1})} = 1. \tag{9}$$

By (1) and the definitions of $\mathbb{G}_1$ and $\mathbb{G}_0$, (9) is equivalent to

$$\lim_{k\to\infty} \frac{\mathbb{G}_1((1 + L_k(\hat{D}_k(\omega)))^{-1})}{\mathbb{G}_0((1 + L_k(\hat{D}_k(\omega)))^{-1})} = 1.$$

Because $\mathbb{G}_1$ and $\mathbb{G}_0$ are right-continuous, we have $\mathbb{G}_1/\mathbb{G}_0$ is also right-continuous. Moreover, $\mathbb{G}_1/\mathbb{G}_0$ is monotone non-decreasing. Therefore, we have

$$\frac{\mathbb{G}_1((1 + L_\infty(\omega))^{-1})}{\mathbb{G}_0((1 + L_\infty(\omega))^{-1})} = 1.$$

However, this contradicts the dominance property (described earlier). We can use a similar argument to show that there does not exist $\omega$ such that $\mathbb{P}_j(d_{k+1} = d_{k+1}(\omega)|\hat{D}_k = \hat{D}_k(\omega)) \to 0$. Therefore, no such $\omega$ exists and this implies that $\mathbb{P}_0(A) = 0$. Hence, $\mathbb{P}_0(L_\infty = 0) = 1$. ∎

*Theorem 5:* Suppose that the flipping probabilities are bounded away from $1/2$. Then, $\mathbb{P}_e^k \to 0$ as $k \to \infty$.

*Proof:* We know that the likelihood ratio test states that $a_k$ decides 1 if and only if $\bar{b}_k > 1 - b_{k-1}$. The probability of deciding 1 given that $H_0$ is true (Type I error) is given by

$$\begin{aligned}
\mathbb{P}_0(d_k = 1) &= \mathbb{P}_0(\bar{b}_k > 1 - b_{k-1}) \\
&= \mathbb{E}_0(1 - \mathbb{G}_0(1 - b_{k-1})).
\end{aligned}$$

Since $L_\infty = 0$ almost surely, we have $b_k \to 0$ almost surely. We have

$$\lim_{k\to\infty} \mathbb{P}_0(d_k = 1) = \lim_{k\to\infty} \mathbb{E}_0(1 - \mathbb{G}_0(1 - b_{k-1})).$$

By the bounded convergence theorem, we have

$$\begin{aligned}
\lim_{k\to\infty} \mathbb{P}_0(d_k = 1) &= 1 - \mathbb{E}_0(\lim_{k\to\infty} \mathbb{G}_0(1 - b_{k-1})) \\
&= 1 - \mathbb{G}_0(1) = 0.
\end{aligned}$$

Similarly, we can prove that $\lim_{k\to\infty} \mathbb{P}_1(d_k = 0) = 0$ (i.e., Type II error probability converges to 0). Therefore, the error probability converges to 0. ∎

*Remark 3 (Additive Gaussian noise):* Note that our convergence proof easily generalizes to the additive Gaussian noise scenario: Suppose that after $a_k$ makes a decision $d_k \in \{0, 1\}$, it broadcasts the decision through a Gaussian broadcasting channel, in other words, the other nodes recieves $\hat{d}_k = F_k d_k + \mathcal{N}_k$, where $F_k \in (0, 1)$ denotes a fading coefficient and $\mathcal{N}_k$ denotes zero-mean Gaussian noise. Then, we can show that the error probability converges to 0 if $F_k$ are bounded away from 0 and the noise variances are bounded for all $k$. In other words, the signal-to-noise ratios are bounded away from 0.

Now let us consider the convergence rate of the error probability. Without loss of generality, we assume that the prior probabilities are equal; i.e., $\pi_0 = \pi_1 = 1/2$. The following analysis easily generalizes to unequal prior probabilities. Recall that $b_k = \mathbb{P}(\theta = H_1|\hat{D}_k)$ denotes the public belief. It is easy to see that the error probability converges to 0 if and only if $b_k \to 0$ almost surely given $H_0$ is true and $b_k \to 1$ almost surely given $H_1$ is true. Recall the proportionality property:

$$\frac{d\mathbb{G}_1}{d\mathbb{G}_0}(r) = \frac{r}{1 - r}.$$

Moreover, we assume $\mathbb{G}_1$ and $\mathbb{G}_0$ are continuous and therefore under each of $H_0$ and $H_1$, the density of the private belief exists. By the above property, we can write these densities as follows:

$$f^1(r) = \frac{d\mathbb{G}_1}{dr}(r) = r\rho(r),$$

and

$$f^0(r) = \frac{d\mathbb{G}_0}{dr}(r) = (1 - r)\rho(r),$$

where $\rho(r)$ is a non-negative function.

Without loss of generality, we assume that $H_0$ is the true hypothesis. Moreover, we assume that $\rho(1) > 0$ and $\rho$ is continuous near $r = 1$. This characterizes the behavior of the tail densities. We will generalize our analysis to polynomial tail densities later, where $\rho(r) \to 0$ as $r \to 1$.

The Bayesian update of the public belief when $\hat{d}_{k+1} = 0$ is given by:

$$\begin{aligned}
&b_{k+1} = \mathbb{P}(\theta = H_1|\hat{D}_{k+1}) \\
&= \frac{\mathbb{P}_1(\hat{d}_{k+1} = 0|\hat{D}_k)b_k}{\sum_{j=0,1} \mathbb{P}_j(\hat{d}_{k+1} = 0|\hat{D}_k)\mathbb{P}(\theta = H_j|\hat{D}_k)} \\
&= \frac{(q_k + (1 - 2q_k)\mathbb{P}_1(d_{k+1} = 0|\hat{D}_k))b_k}{\sum_{j=0,1}(q_k + (1 - 2q_k)\mathbb{P}_j(d_{k+1} = 0|\hat{D}_k))\mathbb{P}(H_j|\hat{D}_k)}.
\end{aligned} \tag{10}$$

It is easy to show that the public belief converges to 0 in the fastest rate if $\hat{d}_k = 0$ for all $k$. We will establish the rate in this special case to bound the converge rate of the error probability. Notice that $\mathbb{P}(\theta = H_1|\hat{D}_k) = b_k$ and $\mathbb{P}(\theta = H_0|\hat{D}_k) = 1 - b_k$.

By Lemma 2, we have $L_k(\hat{D}_k) \to 0$ almost surely, under $H_0$. This implies that $b_k \to 0$ almost surely. If $b_k$ is sufficiently small, then we have

$$\mathbb{P}_1(d_{k+1} = 0|\hat{D}_k) = 1 - \int_{1-b_k}^{1} f^1(x)dx \simeq 1 - \rho(1)(b_k - \frac{b_k^2}{2}) \tag{11}$$

and

$$\mathbb{P}_0(d_{k+1} = 0|\hat{D}_k) = 1 - \int_{1-b_k}^{1} f^0(x)dx \simeq 1 - \rho(1)\frac{b_k^2}{2}. \tag{12}$$

Note that $\simeq$ means asymptotically equal. We can also calculate the (conditional) Type I error probability:

$$\mathbb{P}_0(d_{k+1} = 1|\hat{D}_k) = 1 - \mathbb{P}_0(d_{k+1} = 1|\hat{D}_k) = \int_{1-b_k}^{1} f^0(x)dx \simeq \rho(1)\frac{b_k^2}{2}. \tag{13}$$

Note that (13) characterizes the relationship between the decay rate of Type I error probability and the decay rate of $b_k$. Next we derive the decay rate of $b_k$.

Substituting (11) and (12) into (10) and removing high order terms we obtain

$$b_{k+1} \simeq \frac{(1-q_k)b_k - (1-2q_k)\rho(1)b_k^2}{(1-q_k)}.$$

This implies that

$$b_{k+1} \simeq b_k \left(1 - \frac{1-2q_k}{1-q_k}\rho(1)b_k\right). \tag{14}$$

For any sequence that evolves according to (14), the following lemma characterizes the convergence rate of the sequence.

*Lemma 3:* Suppose that a non-negative sequence $c_k$ satisfies $c_{k+1} = c_k(1 - \delta c_k^n)$, where $n \geq 2$, $c_1 < 1$, and $\delta > 0$. Then, for sufficiently large $k$, there exists two constants $C_1$ and $C_2$ such that

$$\frac{C_1}{(\delta k)^{1/n}} \leq c_k \leq \frac{C_2}{(\delta k)^{1/n}}.$$

This implies that $c_k \to 0$ as $k \to \infty$ and $c_k = \Theta(k^{-1/n})$.

*Proof:* The proof is given in Appendix B. ■

*Theorem 6:* Suppose that the flipping probabilities are bounded away from $1/2$ and $\rho(1)$ is a non-negative constant. Then, the Type I error probability converges to 0 as $\Omega(k^{-2})$.

*Proof:* Using (14) and Lemma 3, we can get the convergence rate of the public belief conditioned on event that $\hat{d}_k = 0$ for all $k$, in which case we have $b_k = \Theta(k^{-1})$. Recall that the public belief converges to 0 the fastest in this case among all possible outcomes. Therefore, we have $b_k = \Omega(k^{-1})$ almost surely.

Recall that $d_k = 1$ if and only if $\bar{b}_k > 1 - b_{k-1}$. Therefore, the Type I error probability is given by

$$\mathbb{P}_0(d_k = 1) = \mathbb{P}_0(\bar{b}_k > 1 - b_{k-1}) = \mathbb{E}_0(1 - \mathbb{G}_0(1 - b_{k-1})). \tag{15}$$

Because $\rho$ is continuous at 1, we have if $x < 1$ is sufficiently close to 1, i.e., $1 - x$ is positive and sufficiently small, then

$$1 - \mathbb{G}_0(x) = \int_x^1 (1-x)\rho(x)dx \geq \frac{\rho(1)}{2}\int_x^1 (1-x)dx = \frac{\rho(1)(1-x)^2}{4}. \tag{16}$$

From (15) and (16) and invoking Jensen's Inequality, we obtain

$$\mathbb{P}_0(d_k = 1) \geq \frac{\rho(1)}{4}\mathbb{E}_0[b_{k-1}^2] \geq \frac{\rho(1)}{4}(\mathbb{E}_0[b_{k-1}])^2. \tag{17}$$

Because $b_k = \Omega(k^{-1})$ almost surely, we have $\mathbb{P}_0(d_k = 1) = \Omega(k^{-2})$. ■

Assume that $\rho(0) > 0$ and $\rho$ is continuous at 0. Then, we can use the same method to calculate the decay rate of the Type II error probability, which is the same as that of the Type I error probability. Note that the decay rate of the error probability depends linearly on $(1 - 2q_k)^{-2}$.

*2) Asymptotically uninformative nodes:* In this part, we consider the case where $q_k \to 1/2$ as $k \to \infty$, which means that the broadcasted decisions become asymptotically uninformative. Let

$$Q_k = \frac{1 - 2q_k}{1 - q_k}.$$

Note that $q_k \to 1/2$ implies that $Q_k \to 0$. This parameter measures how "informative" the corrupted decision is: For example, if $q_k = 0$ (where there is no flipping), then the decision is maximally informative in terms of updating the public belief. However if $q_k = 1/2$, in which case $Q_k = 0$, then the decision is completely uninformative in terms of updating the public belief.

We will derive a necessary condition on the decay rate of $Q_k$ to 0 for the public belief $b_k$ to converge to 0 under $H_0$, which gives us a necessary condition on $Q_k$ for asymptotic learning. For any sequence that evolve according to (14), the following lemma characterizes necessary and sufficient conditions such that the sequence converges to 0.

*Lemma 4:* Suppose that a non-negative sequence $\{c_k\}$ follows $c_{k+1} = c_k(1 - \delta_k c_k^n)$, where $n \geq 1$, $c_1 > 0$, and $\delta_k > 0$. Then, $c_k$ converges to 0 if and only if there exists $k_0$ such that $\sum_{k=k_0}^{\infty} \delta_k = \infty$.

*Proof:* We will use the following claim to prove the lemma: For a non-negative sequence satisfying $c_{k+1} = c_k(1 - r_k)$, where $c_1 > 0$ and $r_k \in [0, 1)$, we have $c_k \to 0$ if and only if there exists $k_0$ such that $\sum_{k=k_0}^{\infty} r_k = \infty$. To show this claim, we have

$$c_{k+1} = c_1 \prod_{i=1}^{k} (1 - r_i).$$

Applying natural logarithm, we obtain

$$\ln c_{k+1} = \ln c_1 + \sum_{i=1}^{k} \ln(1 - r_i).$$

From the above equation, we have $c_k \to 0$ if and only if $\sum_{i=1}^{\infty} \ln(1 - r_i) = -\infty$. In the case where there exists a subsequence of $\{r_k\}$ such that the subsequence is bounded away from 0, we have $\sum_{i=1}^{\infty} \ln(1 - r_i) = -\infty$. Therefore, $c_k \to 0$ as $k \to \infty$. In the case where $r_k \to 0$, there exists $k_0$ such that $r_i \le -\ln(1 - r_i) \le 2r_i$ for all $i \ge k_0$. Therefore, we have $c_k \to 0$ if and only if $\sum_{k=k_0}^{\infty} r_k = \infty$.

We now show the lemma. First we show that the condition is necessary. Suppose that $c_k \to 0$. Then, we have $\sum_{k=1}^{\infty} \delta_k c_k^n = \infty$. Since $c_k < 1$, we have $\sum_{k=1}^{\infty} \delta_k = \infty$. Second we show by contradiction that the condition is sufficient. Suppose that there exist $k_0$ such that $\sum_{k=k_0}^{\infty} \delta_k = \infty$ and $c_k$ does not converge to 0. Since $c_k$ is monotone decreasing, $c_k$ must converge to a nonzero limit $c$. Therefore, for all $k$, we have $c_k \ge c$. Then, we have $c_{k+1} \le c_k(1 - \delta_k c^n)$. We have

$$\sum_{k=k_0}^{\infty} \delta_k c^n = c^n \sum_{k=k_0}^{\infty} \delta_k = \infty.$$

Therefore, we have $c_k \to 0$. ■

*Theorem 7:* Suppose that there exists $p > 1$ such that

$$Q_k = O\left(\frac{1}{k(\log k)^p}\right).$$

Then, the public belief converges to a nonzero limit almost surely.

*Proof:* Suppose that there exists $p > 1$ such that $Q_k = O\left(1/(k(\log k)^p)\right)$. Then, we have

$$\sum_{k=2}^{\infty} Q_k < \infty.$$

Therefore, by Lemma 4, $b_k$ in (14) does not converge to 0. Recall that (14) represents the recursion of $b_k$ conditioned on the event that the node broadcast decisions are all 0. Therefore, the public belief is the smallest among all possible outcomes. Hence, the public belief converges to a nonzero limit almost surely. ■

By (17), it is evident that if $b_k$ converges to a nonzero limit almost surely, then $\mathbb{P}_0(d_k = 1)$ is bounded away from 0 and $\mathbb{P}_0(d_k = 0)$ is bounded away from 1. Therefore, the system does not asymptotically learn the underlying truth. Hence Theorem 7 provides a necessary condition for asymptotically learning.

Theorem 7 also implies that for there to be a nonzero probability that the public belief converges to zero, we must have that there exists $p \le 1$ such that $Q_k = \Omega(1/k(\log k)^p)$. If the public belief does not converge to zero, then it is impossible for there to be an eventual collective arrival at the true hypothesis. To explain this further, Let $\mathcal{H}$ denote the event that there exists a (random) $k_0$ such that the sequence of decisions $d_k = 0$ for all $k \ge k_0$. Occurrence of this event signifies that after a finite number of decisions, the agents arrive at the true underlying state. Such an outcome also means that, eventually, each agent's private signal is overpowered by the past collective true verdict, so that a false decision is never again declared. In the literature on social learning, this phenomenon is called *information cascade* (e.g., [27]) or *herding* (e.g., [22]). We use $\mathcal{L}$ to denote the event $\{b_k \to 0\}$. Notice that $\mathcal{H}$ occurs only if $\mathcal{L}$ occurs. Hence, $\mathcal{H}$ is a subset of the event that $b_k \to 0$, i.e., $\mathcal{H} \subset \mathcal{L}$. These leads to the following corollary of Theorem 7.

*Corollary 4:* If $Q_k = O(1/k(\log k)^p)$ for some $p > 1$, then $\mathbb{P}(\mathcal{H}) = 0$.

So, by the corollary above, only if $Q_k = \Omega(1/k(\log k)^p)$ for some $p \le 1$ can we hope for there to be a nonzero probability that $b_k \to 0$ and thus of information cascade to the truth. Even under the situation that $b_k \to 0$, i.e., conditioned on $\mathcal{L}$, we expect that the *rate* at which $b_k \to 0$ depends on the scaling law of $Q_k$. The following theorem relates the scaling laws of $\{Q_k\}$ with those of $\{b_k\}$ and the Type I error probability sequence $\{\mathbb{P}_0(d_k = 1)\}$.

*Theorem 8:* Conditioned on $\mathcal{L}$, we have the following:

(i) Suppose that $Q_k = \Theta(1/k^{1-p})$ where $p \in (0, 1)$. Then, $b_k = \Omega(k^{-p})$ almost surely and $\mathbb{P}_0(d_k = 1) = \Omega(k^{-2p})$.

(ii) Suppose that $Q_k = \Theta(1/k)$. Then, $b_k = \Omega(1/\log k)$ almost surely and $\mathbb{P}_0(d_k = 1) = \Omega(1/(\log k)^2)$.

(iii) Suppose that $Q_k = \Theta\left(1/(k(\log k)^p)\right)$ where $p \in (0, 1)$. Then, $b_k = \Omega(1/(\log k)^q)$ almost surely, where $1/q + 1/p = 1$, and $\mathbb{P}_0(d_k = 1) = \Omega(1/(\log k)^{2q})$.

(iv) Suppose that $Q_k = \Theta\left(1/(k \log k)\right)$. Then, $b_k = \Omega(1/\log\log k)$ almost surely and $\mathbb{P}_0(d_k = 1) = \Omega(1/(\log\log k)^2)$.

*Proof:* The proof is given in Appendix C. ■

Note that Theorem 8 provides upper bounds for the convergence rates of the public belief and error probability. However, recall that $\mathcal{H}$ is a subset of the event that $b_k \to 0$. Therefore, even if $b_k \to 0$ with certain probability, the probability of $\mathcal{H}$ is not guaranteed to be nonzero. Next we provide a necessary condition such that the probability of $\mathcal{H}$ is nonzero.

*Theorem 9:* Suppose that there exists $p \le 1$ such that

$$Q_k = O\left(\frac{(p + \log k)(\log k)^{p-1}}{(k(\log k)^p)^{1/2}}\right).$$

Then, we have $\mathbb{P}(\mathcal{H}) = 0$.

*Proof:* We first state a key lemma which is a corollary of the Borel-Cantelli lemma [26]. Consider a probability space $(S, \mathcal{S}, \mathcal{P})$ and a sequence of events $\{\mathcal{E}_k\}$ in $\mathcal{S}$. We define the limit superior of $\{\mathcal{E}_k\}$ as follows:

$$\limsup_{k\to\infty} \mathcal{E}_k \equiv \bigcap_{k=1}^{\infty} \Big( \bigcup_{n=k} \mathcal{E}_n \Big).$$

Note that this is the event that infinitely many of the $\mathcal{E}_k$ occur. We use $\mathcal{E}_k^C$ to denote the complement of $\mathcal{E}_k$.

*Lemma 5:* Suppose that

$$\sum_{k=1}^{\infty} \mathcal{P}(\mathcal{E}_k | \mathcal{E}_{k-1}^C, \mathcal{E}_{k-2}^C, \dots, \mathcal{E}_1^C) = \infty.$$

Then,

$$\mathcal{P}(\limsup_{k\to\infty} \mathcal{E}_k) = 1.$$

The proof of this lemma is omitted. Now we prove the theorem. Let $\mathcal{E}_k$ be the event that $d_k = 1$, i.e., $a_k$ makes the wrong decision given $H_0$. Notice that $\mathcal{E}_k^C$ is the event that $d_k = 0$. If

$$Q_k = O\left(\frac{(p + \log k)(\log k)^{p-1}}{(k(\log k)^p)^{1/2}}\right),$$

then using the similar analysis as those in Theorem 8, we have

$$\mathbb{P}_0(\mathcal{E}_k|\mathcal{E}_{k-1}^C, \mathcal{E}_{k-2}^C, \dots, \mathcal{E}_1^C) = \Omega\left(\frac{1}{k(\log k)^p}\right).$$

This implies that these terms are not summable, i.e., $\sum_{k=1}^{\infty} \mathbb{P}_0(\mathcal{E}_k|\mathcal{E}_{k-1}^C, \mathcal{E}_{k-2}^C, \dots, \mathcal{E}_1^C) = \infty$. Therefore we have $\mathbb{P}_0(\limsup_{k\to\infty} \mathcal{E}_k) = 1$, which means that with probability 1, $d_k = 1$ occurs for infinitely many $k$. Consequentially, we have $\mathbb{P}_0(\mathcal{H}) = 0$. By symmetry, $\mathbb{P}_1(\mathcal{H}) = 0$. This concludes the proof. ■

Suppose that the flipping probability converges to $1/2$ sufficiently fast. Then, even if the public belief converges to 0, its convergence rate is very small because the broadcasted decisions become uninformative in a fast rate. In this case, the private signals are capable to overcome the public belief infinitely often because of the slow convergence rate of the public belief.

*3) Polynomial tail density:* We now consider the case where the private belief has polynomial tail densities, that is, $\rho(r) \to 0$ as $r \to 1$ and there exist constants $\beta, \gamma > 0$ such that

$$\lim_{r\to 1} \frac{\rho(r)}{(1-r)^{\beta}} = \gamma. \tag{18}$$

Note that $\beta$ denotes the leading exponent of the Taylor expansion of the density at 1. The larger the value of $\beta$, the thiner the tail density. Note that Theorem 7 (necessary condition for $\mathbb{P}(\mathcal{L}) > 0$) which was stated under the constant density assumption is also valid in the polynomial tail density case. We can use the similar analysis as before to derive the explicit relationship between the convergence rate of $Q_k$ and the convergence rate of the public belief conditioned on $\mathcal{L}$. The following theorem establishes the scaling laws of the public belief and Type I error probability for both uniformly informative and asymptotic uninformative cases.

*Theorem 10:* Consider the polynomial tail density defined in (18).

1) Uniformly informative case: Suppose that the flipping probabilities are bounded away from $1/2$. Then, we have $b_k = \Omega(k^{-1/(\beta+1)})$ almost surely and $\mathbb{P}_0(d_k = 1) = \Omega(k^{-(\beta+2)/(\beta+1)})$.
2) Asymptotically uninformative case: Suppose that the flipping probabilities converge to $1/2$, i.e., $Q_k \to 0$. Conditioned on $\mathcal{L}$, we have
   (i) if $Q_k = \Theta(1/k^{1-p})$ where $p \in (0,1)$, then $b_k = \Omega(k^{-p/(\beta+1)})$ almost surely and $\mathbb{P}_0(d_k = 1) = \Omega(k^{-(\beta+2)p/(\beta+1)})$,
   (ii) if $Q_k = \Theta(1/k)$, then $b_k = \Omega((\log k)^{-1/(\beta+1)})$ almost surely and $\mathbb{P}_0(d_k = 1) = \Omega((\log k)^{-(\beta+2)/(\beta+1)})$,
   (iii) if $Q_k = \Theta\left(1/(k(\log k)^p)\right)$ where $p \in (0,1)$, then $b_k = \Omega((\log k)^{-q/(\beta+1)})$ almost surely, where $1/q + 1/p = 1$, and $\mathbb{P}_0(d_k = 1) = \Omega((\log k)^{-(\beta+2)q/(\beta+1)})$,
   (iv) if $Q_k = \Theta\left(1/(k \log k)\right)$, then $b_k = \Omega((\log\log k)^{-1/(\beta+1)})$ almost surely and $\mathbb{P}_0(d_k = 1) = \Omega((\log\log k)^{-(\beta+2)/(\beta+1)})$.

*Proof:* The proof is given in Appendix D. ■

Next we provide a necessary condition such that $\mathcal{H}$ has nonzero probability.

*Theorem 11:* Suppose that there exists $p \leq 1$ such that

$$Q_k = O\left(\frac{(p+\log k)(\log k)^{p-1}}{(k(\log k)^p)^{1/(\beta+2)}}\right).$$

Then, we have $\mathbb{P}(\mathcal{H}) = 0$.

*Proof:* The proof is similar with that of Theorem 9 and is omitted. ■

Note that as $\beta$ gets larger, this necessary condition states that $Q_k$ has to decay very slowly in order that it is possible for $\mathcal{H}$ to occur.

Similarly we can calculate the decay rate for the Type II error probability $\mathbb{P}_1(d_k = 0)$. Assume that the tail density is given by

$$\lim_{r\to 0} \frac{\rho(r)}{r^{\bar{\beta}}} = \bar{\gamma}$$

where $\bar{\beta}, \bar{\gamma} > 0$. Then, we can show that if the flipping probabilities are bounded away from $1/2$, then

$$\mathbb{P}_1(d_k = 0) = \Omega(k^{-(\bar{\beta}+2)/(\bar{\beta}+1)}).$$

The decay rate of the error probability is given by

$$\mathbb{P}_e^k = \Omega\left(k^{-(1+1/(\max(\beta,\bar{\beta})+1))}\right).$$

## V. Concluding Remarks

We have studied the sequential hypothesis testing problem in two types of broadcast failures: erasure and flipping. In both cases, if the memory sizes are bounded, then there does not exist a decision strategy such that the error probability converges to 0. In the case of random erasure, if the memory size goes to infinity, then there exists a decision strategy such that the error probability converges to 0, even if the erasure probability converges to 1. We also characterize explicitly the relationship between the convergence rate of the error probability and the convergence rate of the memory. In the case of random flipping, if each node observes all the previous decisions, then with the myopic decision strategy, the error probability converges to 0, when the flipping probabilities are bounded away from $1/2$. In the case where the flipping probability converges to $1/2$, we derive a necessary condition on the convergence rate of the flipping probability such that the error probability converges to 0. We also characterize explicitly the relationship between the convergence rate of the flipping probability and the convergence rate of the error probability. Finally, we have derived a necessary condition such that the event herding has nonzero probability.

Our analysis leads to several open questions. We expect that our results can be extended to multiple hypotheses testing problem, paralleling a similar extension in [10]. In the case of random flipping, we have not studied the case where the memory size goes to infinity but each node cannot observe all the previous decisions. We also want to generalize the techniques used in this paper to more general network topologies. Moreover, besides erasure and flipping failures, we expect that our techniques can be used in the additive Gaussian noise scenario. With finite signal-to-noise ratios (SNR), the martingale convergence proof in Lemma 2 easily generalizes to this scenario. However, if SNR goes to 0 (e.g., the fading coefficient goes to 0, the noise variance goes to infinity, or the broadcasting signal power goes to 0), it is obvious that the convergence of error probability is not always true. We want to derive necessary and sufficient conditions on the convergence rate of SNR such that the error probability still converges to 0.

## APPENDIX A
## PROOF OF THEOREM 3

W extend the proof to the case where each node observes $m_k \geq 1$ previous decisions. The likelihood ratio test in this case is given by

$$d_k = \begin{cases} 1 & \text{if } L_X(X_k) > t(\hat{d}_{k-1}, \dots, \hat{d}_{k-m_k}), \\ 0 & \text{if } L_X(X_k) \leq t(\hat{d}_{k-1}, \dots, \hat{d}_{k-m_k}), \end{cases}$$

where $t(\hat{d}_{k-1}, \dots, \hat{d}_{k-m_k}) = t_k / L_D^k(\hat{d}_{k-1}, \dots, \hat{d}_{k-m_k})$ denotes the testing threshold. Among all possible combinations of $\{\hat{d}_{k-1}, \dots, \hat{d}_{k-m_k}\}$, it suffices to assume that the likelihood ratio in the case where each decision equals 0 (denoted by $\mathbf{0}^{m_k}$) is the smallest and that in the case where each decision equals 1 (denoted by $\mathbf{1}^{m_k}$) is the largest. Otherwise, we can always find the smallest and largest likelihood ratio. The case where the likelihood ratios for all possible combinations are equal can be excluded because it means the decisions observed have no useful information for hypothesis testing; and the node has to make a decision based on its own measurement, in which case the error probability does not converge to 0.

From these, we can define the Type I and II error probabilities as in (19) and (20).

With the similar argument as that in the tandem network case, we have $\mathbb{P}_e^k = \pi_0 \mathbb{P}_0(d_k = 1) + \pi_1 \mathbb{P}_1(d_k = 0)$. Suppose that $\mathbb{P}_e^k \to 0$ as $k \to \infty$. Then, we must have $\mathbb{P}_0(L_X(X_k) > t_k(\mathbf{0}^{m_k})) \to 0$ and $\mathbb{P}_1(L_X(X_k) \leq t_k(\mathbf{1}^{m_k})) \to 0$. Recall that $\mathbb{P}_0^X$ and $\mathbb{P}_1^X$ are equivalent measures. Hence we have $\mathbb{P}_j(t_k(\mathbf{1}^{m_k}) < L_X(X_k) \leq t_k(\mathbf{0}^{m_k})) \to 1$ for $j = 0, 1$. We have

$$\begin{aligned}
&\mathbb{P}_j(\hat{d}_{k-1} = j_{k-1}, \hat{d}_{k-2} = j_{k-2}, \dots, \hat{d}_{k-m_k} = j_{k-m_k}) = \\
&\mathbb{P}_j(\hat{d}_{k-1} = j_{k-1} | \hat{d}_{k-2} = j_{k-2}, \dots, \hat{d}_{k-m_k} = j_{k-m_k}) \cdot \\
&\mathbb{P}_j(\hat{d}_{k-2} = j_{k-2} | \hat{d}_{k-3} = j_{k-3}, \dots, \hat{d}_{k-m_k} = j_{k-m_k}) \cdot \\
&\dots \mathbb{P}_j(\hat{d}_{k-m_k+1} = j_{k-m_k+1} | \hat{d}_{k-m_k} = j_{k-m_k}) \cdot \\
&\mathbb{P}_j(\hat{d}_{k-m_k} = j_{k-m_k}).
\end{aligned}$$

We already know that $\mathbb{P}_j(\hat{d}_{k-m_k} = j_{k-m_k})$ is bounded away from 0 by $q_k$. Similarly, we can show

$$\begin{aligned}
&\mathbb{P}_j(\hat{d}_{k-i} = j_{k-i} | \hat{d}_{k-i-1} = j_{k-i-1}, \dots, \hat{d}_{k-m_k} = j_{k-m_k}) \\
&= (1 - q_k)\mathbb{P}_j(d_{k-i} = j_{k-i} | \dots, \hat{d}_{k-m_k} = j_{k-m_k}) \\
&\quad + q_k(1 - \mathbb{P}_j(d_{k-i} = j_{k-i} | \dots, \hat{d}_{k-m_k} = j_{k-m_k})) \\
&= q_k + (1 - 2q_k)\mathbb{P}_j(d_{k-i} = j_{k-i} | \dots, \hat{d}_{k-m_k} = j_{k-m_k}).
\end{aligned}$$

Hence $\mathbb{P}_e^k$ is also bounded below by $q_k^{m_k} \geq q_k^C$. This contradiction implies that $\mathbb{P}_e^k$ does not converge to 0 with any decision strategy.

## APPENDIX B
## PROOF OF LEMMA 3

First it is easy to see that $c_k \to 0$ because it is the only fixed point of the recursion. To show the convergence rate, we treat the recursion (14) as an ordinary difference equation (ODE). Therefore, we have

$$\frac{dc_k}{dk} = -\delta c_k^{n+1}.$$

The solution to this ODE is for some $C > 0$

$$c_k = \frac{C}{(\delta k)^{1/n}}.$$

Therefore, for sufficiently large $k$, there exists two constants $C_1$ and $C_2$ such that

$$\frac{C_1}{(\delta k)^{1/n}} \leq c_k \leq \frac{C_2}{(\delta k)^{1/n}}.$$

which implies that

$$c_k = \Theta(k^{-1/n}).$$

## APPENDIX C
## PROOF OF THEOREM 8

(i). Suppose that $Q_k = \Theta(1/k^{1-p})$ where $p \in (0, 1)$. Conditioned on $\mathcal{H}$, we have recursion (14) for the public belief $b_k$. Using this recursion, we can get similar results as those in Lemma 3, that is, there exists $C_1 > 0$ and $C_2 > 0$ such that

$$\frac{C_1}{kQ_k} \leq b_k \leq \frac{C_2}{kQ_k}. \tag{21}$$

Plugging in the convergence rate of $Q_k$ in (21) establishes the claim.

(ii)-(iv). Suppose that $Q_k = \Theta(1/k(\log k)^p)$, where $p \in [0, 1]$. Then, by (14), we have

$$b_{k+1} - b_k = \frac{Cb_k^2}{k(\log k)^p}$$

for some constant $C > 0$. For $p = 0$, the solution to this ODE satisfies $b_k = \Theta(1/\log k)$, which proves (ii). When $p \in (0, 1)$, the solution satisfies $b_k = \Theta(1/(\log k)^q)$, where $1/q + 1/p = 1$. This establishes (iii). Finally, when $p = 1$, the solution satisfies $b_k = \Theta(1/\log \log k)$. Note that all these rates are derived conditioned on $\mathcal{H}$. By the fact that conditioned on $\mathcal{H}$, the decay rate is the fastest among all outcomes, we obtain the desired results. Having established the convergence rate of $b_k$, the convergence rate for the error probability in each claim follows from (17).

## APPENDIX D
## PROOF OF THEOREM 10

Proof of claim 1: If the flipping probabilities are bounded away from $1/2$, then the public belief $b_k$ converges to 0 and conditioned on $\mathcal{H}$ we have

$$\begin{aligned}
\mathbb{P}_1(d_{k+1} = 0 | \hat{D}_k) &= 1 - \int_{1-b_k}^{1} f^1(x) dx \\
&\simeq 1 - \frac{\gamma}{\beta} b_k^{\beta+1}
\end{aligned} \tag{22}$$

$$\begin{aligned}\mathbb{P}_0(d_k=1) &= \mathbb{P}_0(L_X(X_k)>t_k(\mathbf{0}_k^m))\mathbb{P}_0(\hat{d}_{k-1}=0,\hat{d}_{k-2}=0,\ldots,\hat{d}_{k-m_k}=0)\\ &\quad+\mathbb{P}_0(L_X(X_k)>t_k(1,0,0,\ldots,0))\mathbb{P}_0(\hat{d}_{k-1}=1,\hat{d}_{k-2}=0,\ldots,\hat{d}_{k-m_k}=0)+\ldots\\ &\quad+\mathbb{P}_0(L_X(X_k)>t_k(\mathbf{1}^{m_k}))\mathbb{P}_0(\hat{d}_{k-1}=1,\hat{d}_{k-2}=1,\ldots,\hat{d}_{k-m_k}=1)\\ &=\mathbb{P}_0(L_X(X_k)>t_k(\mathbf{0}^{m_k}))+\mathbb{P}_0(t_k(1,0,0,\ldots,0)<L_X(X_k)\le t_k(\mathbf{0}^{m_k}))\\ &\quad\mathbb{P}_0(\hat{d}_{k-1}=1,\hat{d}_{k-2}=0,\ldots,\hat{d}_{k-m_k}=0)+\ldots\\ &\quad+\mathbb{P}_0(t_k(\mathbf{1}^{m_k})<L_X(X_k)\le t_k(\mathbf{0}^{m_k}))\mathbb{P}_0(\hat{d}_{k-1}=1,\hat{d}_{k-2}=1,\ldots,\hat{d}_{k-m_k}=1)\end{aligned} \tag{19}$$

and

$$\begin{aligned}\mathbb{P}_1(d_k=0) &= \mathbb{P}_1(L_X(X_k)\le t_k(\mathbf{0}^{m_k}))\mathbb{P}_1(\hat{d}_{k-1}=0,\hat{d}_{k-2}=0,\ldots,\hat{d}_{k-m_k}=0)\\ &\quad+\mathbb{P}_1(L_X(X_k)\le t_k(1,0,0,\ldots,0))\mathbb{P}_1(\hat{d}_{k-1}=1,\hat{d}_{k-2}=0,\ldots,\hat{d}_{k-m_k}=0)+\ldots\\ &\quad+\mathbb{P}_1(L_X(X_k)\le t_k(\mathbf{1}^{m_k}))\mathbb{P}_1(\hat{d}_{k-1}=1,\hat{d}_{k-2}=1,\ldots,\hat{d}_{k-m_k}=1)\\ &=\mathbb{P}_1(t_k(\mathbf{1}^{m_k})<L_X(X_k)\le t_k(\mathbf{0}^{m_k}))\mathbb{P}_1(\hat{d}_{k-1}=0,\hat{d}_{k-2}=0,\ldots,\hat{d}_{k-m_k}=0)\\ &\quad+\mathbb{P}_1(t_k(\mathbf{1}^{m_k})<L_X(X_k)\le t_k(1,0,0,\ldots,0))\mathbb{P}_0(\hat{d}_{k-1}=1,\hat{d}_{k-2}=0,\ldots,\hat{d}_{k-m}=0)+\ldots\\ &\quad+\mathbb{P}_1(L_X(X_k)\le t_k(\mathbf{1}^{m_k})).\end{aligned} \tag{20}$$

---

and

$$\begin{aligned}\mathbb{P}_0(d_{k+1}=0|\hat{D}_k) &= 1-\int_{1-b_k}^{1} f^0(x)dx\\ &\simeq 1-\frac{\gamma}{\beta+1}b_k^{\beta+2}.\end{aligned} \tag{23}$$

We can also calculate the (conditional) Type I error probability in this case:

$$\begin{aligned}\mathbb{P}_0(d_{k+1}=1|\hat{D}_k) &= 1-\mathbb{P}_0(d_{k+1}=1|\hat{D}_k)\\ &=\int_{1-b_k}^{1} f^0(x)dx\\ &\simeq\frac{\gamma}{\beta+1}b_k^{\beta+2}.\end{aligned} \tag{24}$$

Note that (24) describes the relationship between the decay rate of Type I error probability and the decay rate of $b_k$. Next we derive the decay rate of $b_k$.

By (22) and (23), we can derive the recursion for the public belief as follows:

$$b_{k+1}=b_k-\frac{\gamma}{\beta}Q_k b_k^{\beta+2}. \tag{25}$$

By Lemma 3, we know that $b_k\to 0$ and the decay rate is $b_k=\Theta(k^{-1/(\beta+1)})$. Recall that conditioned on the event that $\hat{d}_k=0$ for all $k$, the convergence of $b_k$ is the fastest. Therefore, we have $b_k=\Omega(k^{-1/(\beta+1)})$ almost surely. From (24) and invoking Jensen's Inequality, we obtain

$$\begin{aligned}\mathbb{P}_0(d_k=1) &\ge \frac{\gamma}{\beta+1}\mathbb{E}_0[b_k^{\beta+2}]\\ &\ge\frac{\gamma}{\beta+1}(\mathbb{E}_0[b_k])^{\beta+2}.\end{aligned} \tag{26}$$

Because $b_k=\Omega(k^{-1/(\beta+1)})$ almost surely, we have $\mathbb{P}_0(d_k=H_1)=\Omega(k^{-(\beta+2)/(\beta+1)})$.

Proof of claim 2: Using Lemma 3, we can show that there exist two positive constants $C_1$ and $C_2$ such that

$$\frac{C_1}{(kQ_k)^{1/(\beta+1)}}\le b_k\le\frac{C_2}{(kQ_k)^{1/(\beta+1)}}. \tag{27}$$

Therefore, if $Q_k=1/k^{1-p}$, then using (27) and the fact that $b_k$ given $\mathcal{H}$ is the smallest among all possible outcomes, we have $b_k=\Omega(k^{-p/(\beta+1)})$. This establishes (i). For (ii)-(iv), we can solve the ODEs given by (25) and the solutions give rise to the convergence rates for $b_k$, which in turn characterize the convergence rates of the error probabilities.

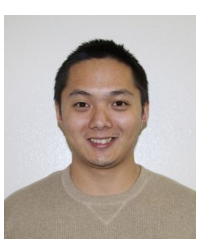
## ACKNOWLEDGMENT

The authors wish to thank the anonymous reviewers for the careful reading of the manuscript and constructive comments that have improved the presentation.

## REFERENCES

[1] Z. Zhang, E. K. P. Chong, A. Pezeshki, and W. Moran, "Asymptotic learning in feedforward networks with binary symmetric channels," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP13)*, Vancouver, Canada, to appear.
[2] A. Wald, "Sequential tests of statistical hypotheses," *Ann. Math. Statist.*, vol. 16, no. 2, pp. 117–186, 1945.
[3] P. K. Varshney, *Distributed detection and data fusion*, New York: Springer-Verlag, 1997.
[4] C. P. Chamley, *Rational herds*, New York: Cambridge, 2004.
[5] T. M. Cover, "Hypothesis testing with finite statistics," *Ann. Math. Statist.*, vol. 40, no. 3, pp. 828–835, 1969.
[6] M. E. Hellman and T. M. Cover, "Learning with finite memory," *Ann. Math. Statist.*, vol. 41, no. 3, pp. 765–782, 1970.
[7] P. Swaszek, "On the performance of serial networks in distributed detection," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 29, no. 1, pp. 254–260, Jan. 1993.
[8] Z. B. Tang, K. R. Pattipati, and D. L. Kleinman, "Optimization of detection networks: Part I—Tandem structures," *IEEE Trans. Syst., Man and Cybern.*, vol. 21, no. 5, pp. 1044–1059, Sept./Oct. 1991.
[9] 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, Jul. 1988.
[10] J. Koplowitz, "Necessary and sufficient memory size for $m$-hypothesis testing," *IEEE Trans. Inform. Theory*, vol. 21, no. 1, pp. 44–46, Jan. 1975.
[11] J. D. Papastravrou 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, Jul. 1992.
[12] V. V. Veeravalli, "Topics in decentralized detection," Ph.D. dissertation, Univ. Illinois, Urbana-Champaign, 1992.

[13] J. N. Tsitsiklis, "Decentralized detection," *Adv. Statist. Signal Process.*, vol. 2, pp. 297–344, 1993.
[14] W. P. Tay, J. N. Tsitsiklis, and M. Z. Win, "On the sub-exponential decay of detection error probabilities in long tandems," *IEEE Trans. Inform. Theory*, vol. 54, no. 10, pp. 4767–4771, Oct. 2008.
[15] I. Lobel, D. Acemoglu, M. A. Dahleh, and A. Ozdaglar, "Lower bounds on the rate of learning in social networks," in *Proc. IEEE American Control Conf.*, Hyatt Regency Riverfront, St. Louis, MO, June 10-12, 2009, pp. 2825–2830.
[16] K. Drakopoulos, A. Ozdaglar, and J. N. Tsitsiklis, "On learning with finite memory," *preprint*, available in arXiv:1209.1122.
[17] P. M. Djuric and Y. Wang, "Distributed Bayesian learning in multiagent systems: Improving our understanding of its capabilities and limitations," *IEEE Signal Process. Magazine*, vol. 29, no. 2, pp. 65–76, Mar. 2012.
[18] A. V. Banarjee, "A simple model for herd behavior," *Quart. J. Econ.*, vol. 107, no. 3, pp. 797–817, Aug. 1992.
[19] S. Bikchandani, D. Hirshleifer, and I. Welch, "A theory of fads, fashion, custom, and cultural change as information cascades," *J. Political Econ.*, vol. 100, no. 5, pp. 992–1026, Oct. 1992.
[20] Z. Zhang, A. Pezeshki, W. Moran, S. D. Howard, and E. K. P. Chong, "Error probability bounds in balanced binary relay trees," *IEEE Tran. Inform. Theory*, vol. 58, no. 6, pp. 3548–3563, Jun. 2012.
[21] Z. Zhang, E. K. P. Chong, A. Pezeshki, W. Moran, and S. D. Howard, "Learning in hierarchical social networks," *IEEE J. Sel. Topics in Signal Process.*, vol. 7, no. 2, pp. 305317, Apr. 2013.
[22] L. Smith and P. Sorensen, "Pathological outcomes of observational learning," *Econometrica*, vol. 68, no. 2, pp. 371–398, Mar. 2000.
[23] V. Krishnamurthy, "Bayesian sequential detection with phase-distributed change time and nonlinear penalty–A POMDP lattice programming approach," *IEEE Trans. Inform. Theory*, vol. 57, no. 10, pp. 7096–7124, Oct. 2011.
[24] V. Krishnamurthy, "Quickest detection POMDPs with social learning: Interaction of local and global decision makers," *IEEE Trans. Inform. Theory*, vol. 58, no. 8, pp. 5563–5587, Aug. 2012.
[25] D. Acemoglu, M. A. Dahleh, I. Lobel, and A. Ozdaglar, "Bayesian learning in social networks," *Review of Economic Studies*, vol. 78, no. 4, pp. 1201–1236, 2011.
[26] R. Durrett, *Probability: theory and examples*, Duxbury Advanced Series, 1996.
[27] J. Surowiecki, *The wisdom of crowds*, New York: Doubleday, 2005.



**Zhenliang Zhang** (S'09) received the B.S. (Physics) degree from Special Class for Gifted Young, University of Science and Technology of China, Hefei, China, in 2008.

He is currently working toward the Ph.D degree in the Department of Electrical and Computer Engineering, Colorado State University, CO. His research interests include distributed signal processing, complex network, and optimization.

**Edwin K. P. Chong** (F'04) received the B.E. degree with First Class Honors from the University of Adelaide, South Australia, in 1987; and the M.A. and Ph.D. degrees in 1989 and 1991, respectively, both from Princeton University, where he held an IBM Fellowship. He joined the School of Electrical and Computer Engineering at Purdue University in 1991, where he was named a University Faculty Scholar in 1999. Since August 2001, he has been a Professor of Electrical and Computer Engineering and Professor of Mathematics at Colorado State University. His current research interests span the areas of stochastic modeling and control, optimization methods, and communication and sensor networks. He coauthored the best-selling book, *An Introduction to Optimization* (4th Edition, Wiley-Interscience, 2013). He received the NSF CAREER Award in 1995 and the ASEE Frederick Emmons Terman Award in 1998. He was a co-recipient of the 2004 Best Paper Award for a paper in the journal *Computer Networks*. In 2010, he received the IEEE Control Systems Society Distinguished Member Award.

Prof. Chong was the founding chairman of the IEEE Control Systems Society Technical Committee on Discrete Event Systems, and served as an IEEE Control Systems Society Distinguished Lecturer. He is currently a Senior Editor of the IEEE TRANSACTIONS ON AUTOMATIC CONTROL, and also serves on the editorial boards of *Computer Networks* and the *Journal of Control Science and Engineering*. He has served a member of the IEEE Control Systems Society Board of Governors and is currently Vice President for Financial Activities. He has also served on the organizing committees of several international conferences. He has been on the program committees for the IEEE Conference on Decision and Control, the American Control Conference, the IEEE International Symposium on Intelligent Control, IEEE Symposium on Computers and Communications, and the IEEE Global Telecommunications Conference. He has also served in the executive committees for the IEEE Conference on Decision and Control, the American Control Conference, the IEEE Annual Computer Communications Workshop, the International Conference on Industrial Electronics, Technology & Automation, and the IEEE International Conference on Communications. He was the Conference (General) Chair for the Conference on Modeling and Design of Wireless Networks, part of SPIE ITCom 2001. He was the General Chair for the 2011 Joint 50th IEEE Conference on Decision and Control and European Control Conference.

**Ali Pezeshki** (S'95–M'05) received the B.Sc. and M.Sc. degrees in electrical engineering from the University of Tehran, Tehran, Iran, in 1999 and 2001, respectively. He received his Ph.D degree in electrical engineering at Colorado State University in 2004. In 2005, he was a postdoctoral research associate with the Electrical and Computer Engineering Department at Colorado State University. From January 2006 to August 2008, he was a postdoctoral research associate with The Program in Applied and Computational Mathematics at Princeton University. Since August 2008, he has been an assistant professor with the Department of Electrical and Computer Engineering, Colorado State University, Fort Collins, CO. His research interests are in statistical signal processing and coding theory and their applications to active/passive sensing.

**William Moran** (M'95) received the B.S. (Hons.) degree in mathematics from the University of Birmingham in 1965 and the Ph.D. degree in mathematics from the University of Sheffield, London, U.K., in 1968. He is currently a Professor of electrical engineering with the University of Melbourne, Australia, where he is the Research Director of Defence Science Institute and Technical Director of Melbourne Systems Laboratory. Previously, he was a Professor of mathematics with the University of Adelaide and Flinders University. He also serves as a Consultant to the Australian Department of Defence through the Defence Science and Technology Organization. His research interests are in signal processing, particularly with radar applications, waveform design and radar theory, and sensor management. He also works in various areas of mathematics, including harmonic analysis and number theory and has published widely in these areas.