Near-optimal Distributed Detection in Balanced Binary Relay Trees

Zhenliang Zhang, *Member, IEEE,* Edwin K. P. Chong, *Fellow, IEEE,* Ali Pezeshki, *Member, IEEE,* Bill Moran, *Member, IEEE,* and Stephen D. Howard, *Member, IEEE*

***Abstract*—We study the distributed detection problem in a balanced binary relay tree, where the leaves of the tree are sensors generating binary messages. The root of the tree is a fusion center that makes an overall decision. Every other node in the tree is a relay node that fuses binary messages from its two child nodes into a new binary message and sends it to the parent node at the next level. We assume that the relay nodes at the same level use identical fusion rule. The goal is to find a string of fusion rules used at all the levels in the tree that maximizes the reduction in the total error probability between the leaf nodes and the fusion center. We formulate this problem as a deterministic dynamic program and express the optimal strategy in terms of *Bellman's equation*. Moreover, we use the notion of *string-submodularity* to show that the reduction in the total error probability is a string-submodular function. Consequentially, we show that the greedy strategy, which only maximizes the level-wise reduction in the total error probability, performs at least within a factor $(1 \quad 1/e)$ of the optimal strategy in terms of reduction in the total error probability, even if the nodes and links in the trees are subject to random failures.**

***Index Terms*—Distributed detection, Bayesian social learning, submodular function, greedy strategy.**

I. Introduction

Consider a *distributed detection* network consisting of a set of sensors and fusion nodes. The objective is to solve collectively a binary hypothesis testing problem. The sensors make observations from a common event, and then communicate quantized messages to other fusion nodes, according to the network architecture. Each fusion node fuses the received messages from its child nodes into a new message and then sends it to the fusion node at the next level for further integration. A final decision is eventually made at a central node, usually known as the fusion center. A fundamental question is how to fuse messages at each fusion node such that the fusion center makes the best decision, in the sense of optimizing a global objective function. For example, under the Neyman-Pearson criterion, the objective is to minimize the probability of missed detection with an upper bound constraint on the probability of false alarm; under the Bayesian criterion, the objective is to minimize the total error probability under a given prior. The distributed detection problem is largely motivated by advances in wireless sensor networks (e.g., networks of wireless sensors deployed in smart home, smart city, connected cars, etc.), where how to aggregate the data from different sensing devices becomes a pressing issue. Most of these devices are powered by batteries and energy efficiency is of critical importance. Therefore, communication between nodes is often restricted to be one bit.

A similar hypothesis testing problem is investigated in the context of a social learning model [2]–[6], where each node represents an agent in the social network and each link represents the pairwise interaction. In most social learning literature, it is assumed that each agent owns a private observation and makes a decision by fusing it with the decisions fed by the previous agents. Moreover, it is commonly assumed that these agents in the social network are parochial and make their decisions that are only locally optimal. In contrast, most distributed detection literature takes an engineering perspective wherein each sensor makes a decision that is globally optimal, i.e., most beneficial for the overall objective. Obviously, the locally optimal strategy is in general *not* globally optimal. To connect the distributed detection and social learning problems, a nature question to ask is that how bad the locally optimal strategy can be compared to the globally optimal strategy. In this paper, we will derive the globally optimal strategy using Bellman's principle and then quantify the performance differences between the locally and globally optimal strategies for hypothesis testing problem in a network arranged as a balanced binary relay tree.

The distributed detection problem has been investigated extensively in the context of various network architectures. For the well-studied *parallel network* [7]–[23] where sensors communicate with the fusion center directly, with the assumption of independent sensor observations conditioned on either hypothesis, the optimal fusion rule at the fusion center under the Bayesian criterion is simply a likelihood-rate test with a threshold given by the ratio of the prior probabilities. For the *tandem network* [24]–[29], each fusion node combines the observation from its own sensor with the message it receives from its child node at one level down

This work was supported in part by the Foundemental Research Program of Shenzhen (JCYJ20151117161854942), the leading talents of Guangdong province Program (No. 00201510), AFOSR under Contract FA9550-09-1-0518, and by AFOSR/ASU under contract FA9550-12-1-0418, and by NSF under Grants CCF-0916314, CCF-1018472, and CCF-1422658. Part of the results in this paper was presented in *51st IEEE Conference on Decision and Control (CDC)* [1].

Z. Zhang is with Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, China 518055. He was with the Department of Electrical and Computer Engineering, Colorado State University, Fort Collins, CO 80523-1373, USA (e-mail: zzl.csu@gmail.com).

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

B. Moran is with RMIT University, Melbourne, VIC 3000, Australia (e-mail: bill.moran@rmit.edu.au).

S. D. Howard is with the Defence Science and Technology Group, P.O. Box 1500, Edinburgh, SA 5111, Australia (e-mail: sdhoward@unimelb.edu.au).

and then transmits the combined message to its parent node at the next level up. We call a collection of fusion rules at all fusion nodes a *fusion strategy*. Specifically, [25] considers the problem of finding the optimal fusion strategy in tandem network under both Neyman-Pearson and Bayesian criteria. The *bounded-height tree network* has been considered in [30]–[39], where the leaves are sensors, the root is the fusion center, and every other node is a fusion node that fuses the messages from its child nodes and sends a new message to its parent node. In general, finding a fusion strategy that minimizes the total error probability at the fusion center in bounded-height trees is computationally intractable, even for a network with a moderate number of nodes. In consequence, many recent papers focus on the asymptotic decay rate of the total error probability as the number of sensors goes to infinity. In balanced bounded-height trees where all the leaf nodes are at the same distance from the fusion center, a fusion strategy that asymptotically achieves the optimal decay exponent is studied in [31], in which all the fusion nodes at the same level use the same likelihood-ratio test as the fusion rule.

The *unbounded-height tree network* has been considered in [40]–[44]. In particular, one energy efficient and simple data aggregation solution is proposed by Gubner *et al.* [40], in which each sensor only generates one bit of data, combines it with the one bit data received from its predecessor sensor, and communicates to a nearby sensor at the next level for further integration, until a final aggregated binary decision is produced. The communication in this scheme is restricted to pairs of nearby sensors at each time, which significantly reduces the possibility of energy costly and interference generating long-range wireless communication. This scheme is in fact equivalent to the distributed detection problem in balanced binary relay trees shown in Figure 1. In this configuration, the leaf nodes, depicted as circles, are sensors generating binary messages independently and forward these binary messages to their parent nodes. Each node, depicted as a diamond, is a fusion (relay) node that fuses the two binary messages received from its child nodes and forwards the new message upward. Ultimately, the fusion center at the root makes an overall decision. This tree is *balanced* in the sense that all leaf nodes are at the same distance from the fusion center, and it is *binary* in the sense that each nonleaf node has two child nodes. This architecture is of interest because it represents the worst-case scenario in the sense that the minimum distance from the sensors to the fusion center is the largest among all possible tree architectures with the same total number of nodes. Assuming that all nonleaf nodes use the same fusion rule, the unit-threshold likelihood-ratio test (ULRT), [40] provides a proof of the convergence of detection error probabilities using a Lyapunov method. Under the same assumptions, we further show in [41] that the decay rate of the total error probability is $\sqrt{N}$, where N is the number of sensors. Under the equally likely prior probability assumption, ULRT is the locally optimal fusion rule in the sense that the total error probability of each node is minimized after each fusion. We do not, however, expect the strategy, consisting of repeated ULRT fusion rules, to be globally optimal in the sense that the total error probability at the fusion center is minimized for all possible combinations of fusion rules at all the fusion nodes. This scheme is also known as the *greedy* strategy (which we call the**ULRT strategy** throughout the paper) due to the myopic nature of decision making.

In fact, there are two main approaches for general dynamic optimization problems. One approach is to exploit Bellman's principle of optimality and derive the globally optimal solution. The main defect of this approach is its prohibitive computation complexity. The other approach is to use approximation algorithms (e.g., the greedy strategy), which are not optimal in general but computationally economic. In this paper, we explore and connect these two approaches by characterizing their performance gap. Specifically, we are interested in the following questions:

1) What is the globally optimal strategy for balanced binary relay trees?
2) What is the difference in terms of the total error probability (the performance metric for Bayesian detection) between the globally optimal strategy and the greedy strategy?

We answer the first question by formulating the problem as a deterministic dynamic program and characterizing the globally optimal strategy using Bellman's equation. We answer the second question by applying the notion of *string-submodularity* introduced in [45] and showing that the reduction in the total error probability is a string-submodular function. Subsequently, we show that the reduction in the total error probability achieved by the greedy strategy is at least a factor $(1 - e^{-1})$ of that achieved by the globally optimal strategy, even if the nodes and links in the tree network are subject to random failures.

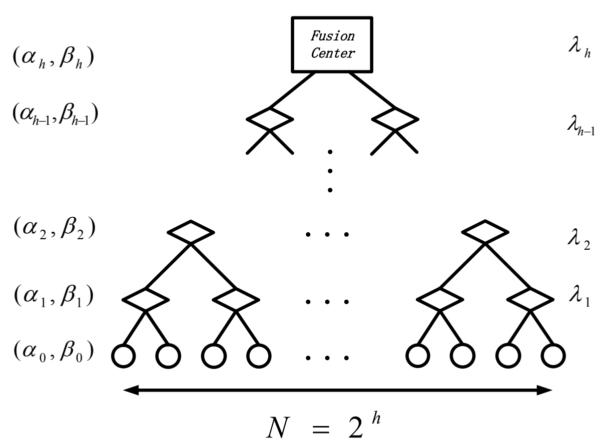


Fig. 1. A balanced binary relay tree with height h. Circles represent sensors making measurements. Diamonds represent relay nodes which fuse binary messages. The rectangle at the root represents the fusion center making the overall decision. The notation (α_k, β_k) denotes the Type I and II error probabilities for node decisions at level k and λ_k denotes the fusion rule used at level k. We will rigorously define these notation in Section II.

II. Problem Formulation

We consider the problem of testing the binary hypothesis between H_0 and H_1 in a balanced binary relay tree, with the structure shown in Figure 1. Let p be any fusion node, that is, p is a nonleaf node. We say that p is at level k if there are k hops between this node and the closest leaf node in the tree. We denote by $C(p)$ the set of child nodes of p. Suppose that p receives binary messages $x_c \in \{0,1\}$ from every $c \in C(p)$ (that is, from its child nodes), and then summarizes the two received binary messages into a new binary message $x_p \in \{0,1\}$ using a fusion rule λ^p:

$$x_p = \lambda^p(\{x_c : c \in C(p)\}).$$

The new message x_p is then transmitted to the parent node (if any) of p. Ultimately, the fusion center makes the overall decision.

We assume that all leaf nodes are independent and the binary messages associated with these leaf nodes have identical Type I error probability $\alpha_0 = \mathbb{P}(x_0 = 0|H_1)$ and identical Type II error probability $\beta_0 = \mathbb{P}(x_0 = 1|H_0)$. In addition, it is assumed that all the fusion nodes at level k ($k \in \{1, 2, \ldots, h\}$) use identical fusion rule $\lambda_k \in \mathcal{Y}$; that is, for each node p that lies at the kth level of the tree, $\lambda^p = \lambda_k$. Suppose that each fusion node fuses the two binary messages using a likelihood-ratio test. Then, the threshold for the likelihood-ratio test determines where the tie breaks. Apparently, extremely large or small threshold will cause the likelihood-ratio test to make a binary decision regardless of the two binary messages. It turns out that the only meaningful rules (among all likelihood-ratio tests with arbitrary thresholds) to aggregate two binary messages in this case are simply 'AND' and 'OR' rules defined as follows:

- AND rule (denoted by $\mathcal{A}$): a parent node decides and sends 1 if and only if both its child nodes send 1;
- OR rule (denoted by $\mathcal{O}$): a parent node decides and sends 0 if and only if both its child nodes send 0.

Henceforth, we only consider the case where each fusion node (including the fusion center) in the tree chooses a fusion rule from $\mathcal{Y} := \{\mathcal{A}, \mathcal{O}\}$.

In this case, all the output binary messages for nodes at level k have the same Type I and Type II error probabilities, denoted by α_k and β_k respectively. Given a fusion rule λ_k, it is straightforward to see that the error probabilities evolve as follows:

$$(\alpha_k, \beta_k) := \begin{cases} (1-(1-\alpha_{k-1})^2, \beta_{k-1}^2), & \text{if } \lambda_k = \mathcal{A}, \\ (\alpha_{k-1}^2, 1-(1-\beta_{k-1})^2), & \text{if } \lambda_k = \mathcal{O}. \end{cases}$$

Remark: Note that the evolution of the error probability pair (α_k, β_k) is symmetric with respect to the line $\alpha + \beta = 1$, so that it suffices to consider the case where the initial pair satisfies $\alpha_0 + \beta_0 < 1$. The case of $\alpha_0 + \beta_0 > 1$ is obtained, for instance, by only flipping the decision at the fusion center. In the case where $\alpha_0 + \beta_0 = 1$, the Type I and II error probabilities sum to 1 regardless of the fusion rule used, and this case is uninteresting.

Notice that the ULRT fusion rule is either the $\mathcal{A}$ rule or the $\mathcal{O}$ rule, depending on the values of the Type I and Type II error probabilities at a particular level of the tree. More precisely, we have

- If $\beta_k > \alpha_k$, then the ULRT fusion rule is $\mathcal{A}$;
- If $\beta_k < \alpha_k$, then the ULRT fusion rule is $\mathcal{O}$;
- If $\beta_k = \alpha_k$, then the total error probability remains unchanged using either $\mathcal{A}$ or $\mathcal{O}$. Moreover, the error probability pairs at the next level $(\alpha_{k+1}, \beta_{k+1})$ after using $\mathcal{A}$ or $\mathcal{O}$ are symmetric about the line $\beta = \alpha$.

We define a fusion strategy as a string of fusion rules $\lambda_j \in \mathcal{Y}$ used at levels $j = 1, 2, \ldots, h$, and denoted by $\pi = (\lambda_1, \lambda_2, \ldots, \lambda_h)$. Note that h denotes the height of the tree. Let the collection of all possible fusion strategies with length h be $\mathcal{Y}^h$:

$$\mathcal{Y}^h := \{\pi = (\lambda_1, \lambda_2, \ldots, \lambda_h) | \lambda_j \in \mathcal{Y} \ \text{ for } j = 1, 2, \ldots, h\}.$$

For a given initial error probability pair (α_0, β_0) at the sensor level, the pair (α_h, β_h) at the fusion center (level h) is a function of (α_0, β_0) and the specific fusion strategy π used. We consider the Bayesian criterion in this paper, under which the objective is to minimize the total error probability $\mathbb{P}(H_0)\alpha_h + \mathbb{P}(H_1)\beta_h$ at the fusion center, where $\mathbb{P}(H_0)$ and $\mathbb{P}(H_1)$ are the prior probabilities of the two hypotheses, respectively. Equivalently, we can find a strategy that maximizes the reduction of the total error probability between the sensors and the fusion center. We call this optimization problem an *h-optimal problem*. Notice that the objective function, the Bayesian probability of error, is a prior-weighted average of Types I and II error probabilities. For simplicity of the presentation, it suffices to consider the equal prior scenario; i.e., $\mathbb{P}(H_0) = \mathbb{P}(H_1) = 1/2$. The main results in later sections easily generalize to the unequal prior scenario. In this equal-prior scenario, the h-optimal problem (ignoring a factor of $1/2$) can be written as:

$$\begin{array}{l} \text{maximize } \alpha_0 + \beta_0 - (\alpha_h + \beta_h) \\ \text{subject to } \pi \in \mathcal{Y}^h. \end{array} \tag{1}$$

A fusion strategy that maximizes (1) is called the h-optimal strategy:

$$\begin{aligned} \pi^o(\alpha_0, \beta_0) &= \arg\max_{\pi \in \mathcal{Y}^h} (\alpha_0 + \beta_0 - (\alpha_h + \beta_h)) \\ &= \arg\max_{\pi \in \mathcal{Y}^h} \sum_{j=0}^{h-1} (\alpha_j + \beta_j - (\alpha_{j+1} + \beta_{j+1})). \end{aligned}$$

In contrast, the ULRT fusion rule only minimizes the step-wise reduction in the total error probability:

$$\text{ULRT} = \arg\max_{\lambda_i \in \mathcal{Y}} (\alpha_i + \beta_i - (\alpha_{i+1} + \beta_{i+1})) \quad \forall i.$$

Because of the equal prior probability assumption, a *maximum a posteriori* (MAP) fusion rule is the same as the ULRT fusion rule. In this context, we call a fusion strategy consisting of repeated ULRT fusion rules at all levels a ULRT strategy.

In the next section, we derive the h-optimal fusion strategy for balanced binary relay trees with height h using a dynamic programming approach. More specifically, we express the solution using Bellman's equation. We then show that the 2-optimal strategy is equivalent to the ULRT strategy, but this is not true for the h-optimal strategy when $h \geq 3$.

III. Optimal Strategy and Optimality of Greedy Strategy

A. Dynamic Programming Formulation

In this section, we formulate the problem of finding the optimal fusion strategy using a deterministic dynamic programming model. First we define the necessary elements of this dynamical model.

I. *Dynamical System*: We define the error probability pair at the kth level (α_k, β_k) as the system state, denoted by s_k. Notice that α_k and β_k can only take values in the interval $[0, 1]$, and the set of all possible states is $\{(\alpha, \beta) > 0 | \alpha + \beta < 1\}$. Given the fusion rule, the *state transition function* is deterministic. If we choose $\lambda_k = \mathcal{A}$, then

$$(\alpha_k, \beta_k) = (1 - (1 - \alpha_{k-1})^2, \beta_{k-1}^2).$$

On the other hand, if we choose $\lambda_k = \mathcal{O}$, then

$$(\beta_k, \alpha_k) = (1 - (1 - \beta_{k-1})^2, \alpha_{k-1}^2).$$

II. *Rewards*: At each level k, the *instantaneous reward* is defined to be the reduction of the total error probability after fusing with λ_k:

$$r(s_{k-1}, \lambda_k) = (\alpha_{k-1} + \beta_{k-1}) - (\alpha_k + \beta_k),$$

where α_k and β_k are functions of the previous state s_{k-1} and the fusion rule λ_k.

Let $v_{h-k}(s_k)$ be the cumulative reduction of the total error probability if the system is started at state s_k at level k and the strategy $(\lambda_{k+1}, \lambda_{k+2} \ldots, \lambda_h) \in \mathcal{Y}^{h-k}$ is used. Following the above definitions, we have

$$v_{h-k}(s_k) = \sum_{j=k+1}^{h} r(s_{j-1}, \lambda_j).$$

Setting $k = 0$; that is, calculating the reduction from the sensor level, reduces the cumulative reward function to the global objective function defined in Section II. For a given initial state s_0, therefore, we have to solve the following optimization problem to find the global optimal strategy over the horizon h:

$$v_h^o(s_0) = \max_{\pi \in \mathcal{Y}^h} \sum_{j=1}^{h} r(s_{j-1}, \lambda_j).$$

The globally optimal strategy π^o is

$$\pi^o(s_0) = \arg\max_{\pi \in \mathcal{Y}^h} \sum_{j=1}^{h} r(s_{j-1}, \lambda_j).$$

Notice that s_k depends on the previous state s_{k-1} and the fusion rule λ_k. We will denote the state at level k by $s_k|_{s_{k-1},\lambda_k}$. The solution of the above optimization problem can be described in terms of Bellmam's equations; namely,

$$v_h^o(s_0) = \max_{\lambda_1 \in \mathcal{Y}} \left[r(s_0, \lambda_1) + v_{h-1}^o(s_1|_{s_0,\lambda_1}) \right]$$

$$\lambda_1^o(s_0) = \arg\max_{\lambda_1 \in \mathcal{Y}} \left[r(s_0, \lambda_1) + v_{h-1}^o(s_1|_{s_0,\lambda_1}) \right],$$

where $\lambda_1^o(s_0)$ is the first element of the optimal strategy $\pi^o(s_0)$. Recursively, the solution of the optimization problem is given by

$$v_{h-(k-1)}^o(s_{k-1}) = \max_{\lambda_k \in \mathcal{Y}} \left[r(s_{k-1}, \lambda_k) + v_{h-k}^o(s_k|_{s_{k-1},\lambda_k}) \right],$$

and the kth element of the optimal strategy $\pi^o(s_0)$ is

$$\lambda_k^o(s_{k-1}) = \arg\max_{\lambda_k \in \mathcal{Y}} \left[r(s_{k-1}, \lambda_k) + v_{h-k}^o(s_k|_{s_{k-1},\lambda_k}) \right].$$

Remark: The above formulation can easily be generalized to the case where the nodes and links in the tree fail with specified probabilities [42] and even more complicated network architectures (for instance, bounded-height trees [31] and M-ary relay trees [44]) simply by changing the state transition functions and the set of all possible fusion rules. Also, we can easily generalize the above formulation to a non-equal prior probability scenario by including the prior probabilities to weight the Types I and II error probabilities.

The complexity of the explicit solution to Bellman's equation grows exponentially with respect to the horizon h, and the computation of the h-optimal strategy is intractable, especially when h is large. The ULRT strategy, in which the ULRT fusion rule is repeated at all levels, provides an alternative. We have shown in [41] that the decay rate of the total error probability with this strategy is $\sqrt{N}$. We will now consider whether the ULRT strategy is the same as the h-optimal strategy. If not, does the ULRT strategy provide a reasonable approximation of the h-optimal strategy?

B. 2-optimal Strategy

Here, we show that the 2-optimal strategy is the same as the ULRT strategy. Moreover, we give a counterexample which shows that the ULRT strategy is not 3-optimal.

Consider the 2-optimal problem in the balanced binary relay tree with height 2:

$$v_2^o(s_0) = \max_{\pi \in \mathcal{Y}^2} \sum_{j=1}^{2} r(s_{j-1}, \lambda_j),$$

where $\mathcal{Y}^2 = \{(\mathcal{A}, \mathcal{A}), (\mathcal{A}, \mathcal{O}), (\mathcal{O}, \mathcal{O}), (\mathcal{O}, \mathcal{A})\}$. The 2-optimal strategy in this case is

$$\pi^o(s_0) = \arg\max_{\pi \in \mathcal{Y}^2} \sum_{j=1}^{2} r(s_{j-1}, \lambda_j).$$

We have the following theorem.

Theorem 1: A strategy π is 2-optimal if and only if π is the ULRT strategy.

Proof: First consider the special cases where $\beta_k = \alpha_k$ in the 2-optimal problem for $k = 0$ or $k = 1$. We know that, when $\beta_k = \alpha_k$, then neither $\mathcal{A}$ nor $\mathcal{O}$ changes the total error probability after fusion and the ensuing states after the use of these rules are symmetric with respect to $\beta = \alpha$ line. Moreover, both $\mathcal{A}$ and $\mathcal{O}$ are called the ULRT strategy. Consequently, in this case, the 2-optimal problem reduces to a 1-optimal problem. In other words, if π is 2-optimal, then we can show that π is the ULRT strategy. On the other hand, if π is the ULRT strategy and $\beta_k = \alpha_k$ for $k = 0$ or $k = 1$, then it is easy to show that π is always 2-optimal.

Now we show the theorem for the case $\beta_k \neq \alpha_k$ for $k = 0$ and $k = 1$. First note that the total error probability is strictly non-increasing after fusing with the ULRT fusion rule. However, if we apply the fusion rule other than the ULRT fusion rule in $\mathcal{Y}$, then the total error probability increases strictly after fusion. For example, if $\beta_k > \alpha_k$ and we apply the $\mathcal{O}$ fusion rule, then the total error probability increases strictly; that is,

$$\alpha_{k+1} + \beta_{k+1} = \alpha_k^2 + 1 - (1-\beta_k)^2 > \alpha_k + \beta_k;$$

in other words, the instantaneous reward in this case is negative,

$$r(s_k, \mathcal{O}) < 0.$$

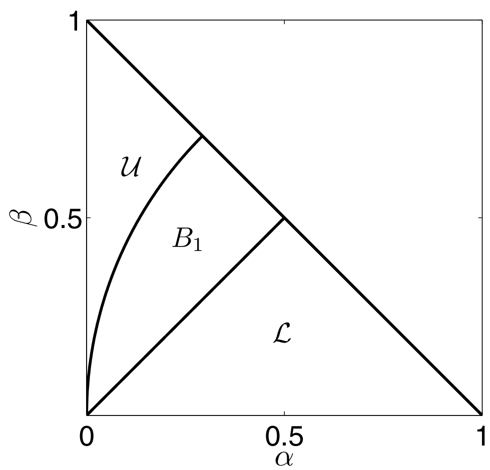


Fig. 2. Regions $\mathcal{U}$, $\mathcal{L}$, and B_1 in the (α, β) plane.

Because of symmetry, it suffices to prove this theorem in the upper triangular region $\mathcal{U}$ (see Figure 2) defined as follows:

$$\mathcal{U} := \{(\alpha, \beta) \geq 0 | \alpha + \beta < 1 \text{ and } \beta > \alpha\}.$$

We define the reflection of $\mathcal{U}$ with respect to $\beta = \alpha$ line to be $\mathcal{L}$. Recall that if $(\alpha_k, \beta_k) \in B_1$, where

$$B_1 := \{(\alpha, \beta) \in \mathcal{U} | (1-\alpha)^2 + \beta^2 \leq 1\},$$

then the next state $(\alpha_{k+1}, \beta_{k+1}) \in \mathcal{L}$. This region B_1 is illustrated in Figure 2. Also recall that if (α_0, β_0) lies on the boundary of B_1, then the next state (α_1, β_1) lies on $\beta = \alpha$ line, and this boundary is not considered.

The proof is divided into two cases:

- *Case I*: $(\alpha_0, \beta_0) \in B_1$, in which case the ULRT strategy is $(\mathcal{A}, \mathcal{O})$;
- *Case II*: $(\alpha_0, \beta_0) \in \mathcal{U} \setminus B_1$, in which case the ULRT strategy is $(\mathcal{A}, \mathcal{A})$.

For Case I where $(\alpha_0, \beta_0) \in B_1$, it is easy to see that strategy $(\mathcal{A}, \mathcal{O})$ achieves a larger reduction than $(\mathcal{A}, \mathcal{A})$, because using the $\mathcal{A}$ rule for the second level increases the total error probability. Similarly, the total error probability after using $(\mathcal{O}, \mathcal{O})$ increases with respect to the initial total error probability, and so this fusion rule, also, cannot be 2-optimal. It suffices to show that the strategy $(\mathcal{A}, \mathcal{O})$ achieves a larger reduction than $(\mathcal{O}, \mathcal{A})$:

$$r(s_0, \mathcal{A}) + r(s_1, \mathcal{O}) > r(s_0, \mathcal{O}) + r(s_1, \mathcal{A}),$$

equivalent to

$$\begin{aligned} &r(s_0, \mathcal{A}) + r(s_1, \mathcal{O}) - (r(s_0, \mathcal{O}) + r(s_1, \mathcal{A})) = \\ &(1 - (1-\beta_0)^2)^2 + 1 - (1-\alpha_0^2)^2 - \\ &((1 - (1-\alpha_0)^2)^2 + 1 - (1-\beta_0^2)^2) > 0. \end{aligned}$$

This reduces to

$$\beta_0^2(1-\beta_0)^2 - \alpha_0^2(1-\alpha_0)^2 > 0,$$

which holds for all $(\alpha_0, \beta_0) \in B_1$, and so the 2-optimal fusion strategy in this case is also $(\mathcal{A}, \mathcal{O})$. We conclude that if $(\alpha_0, \beta_0) \in B_1$, then a strategy is 2-optimal if and only if it is the ULRT strategy.

For Case II where $(\alpha_0, \beta_0) \in \mathcal{U} \setminus B_1$, it is easy to see that strategy $(\mathcal{A}, \mathcal{A})$ achieves a larger reduction than $(\mathcal{A}, \mathcal{O})$. Also, the total error probability after using $(\mathcal{O}, \mathcal{O})$ increases with respect to the initial total error probability, so that this fusion rule cannot be 2-optimal. It suffices to show that the strategy $(\mathcal{A}, \mathcal{A})$ achieves a larger reduction than $(\mathcal{O}, \mathcal{A})$:

$$r(s_0, \mathcal{A}) + r(s_1, \mathcal{A}) > r(s_0, \mathcal{O}) + r(s_1, \mathcal{A}),$$

which reduces to

$$\begin{aligned} &r(s_0, \mathcal{A}) + r(s_1|_{s_0}, \mathcal{A}) - (r(s_0, \mathcal{O}) + r(s_1|_{s_0}, \mathcal{A})) = \\ &(1 - (1-\beta_0)^2)^2 + 1 - (1-\alpha_0^2)^2 - \\ &(1 - (1-\alpha_0)^4 + \beta_0^4) > 0. \end{aligned}$$

This is equivalent to

$$\beta_0(1-\beta_0)(1+\beta_0) - \alpha_0(1-\alpha_0)(1-\alpha_0) > 0,$$

which holds for all $(\alpha_0, \beta_0) \in \mathcal{U} \setminus B_1$. As a result, the 2-optimal fusion strategy in this case is also $(\mathcal{A}, \mathcal{A})$. We conclude that, if $(\alpha_0, \beta_0) \in \mathcal{U} \setminus B_1$, then a strategy is 2-optimal if and only if it is the ULRT strategy.

■

This result also applies to any sub-tree with height 2 within a balanced binary relay tree with arbitrary height $h > 2$. However, the ULRT strategy is not in general optimal for multiple levels; that is, $h > 2$, as the following counterexample for $h = 3$ shows.

Let the initial state be $(\alpha_0, \beta_0) = (0.2, 0.3)$, in which case the ULRT strategy is $(\mathcal{A}, \mathcal{O}, \mathcal{A})$. As shown in Figure 3, the solid (red) line denotes the total error probabilities at each level up to 3. However, the 3-optimal strategy in this case is $(\mathcal{O}, \mathcal{A}, \mathcal{A})$. The total error probability curve of this strategy is shown as a dashed (green) line in Figure 3. Similar counterexamples can be found for $h > 3$. In other words, the ULRT strategy is not in general h-optimal for $h \geq 3$. In the next section, we will introduce and employ the notion of string-submodularity to quantify the gap in performances between optimal and ULRT strategies for $h \geq 3$.

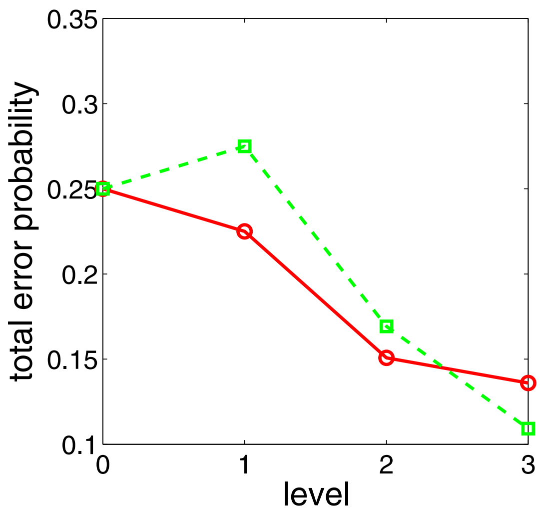


Fig. 3. Comparison of the ULRT strategy and the 3-optimal strategy. The solid (red) line represents the error probability curve using the ULRT strategy. The dashed (green) line represents the error probability curve using the 3-optimal strategy.

IV. Submodularity of Greedy Strategy

In this section, we will introduce, here, the notion of string-submodularity and use it to bound the performance of the ULRT strategy with respect to the optimal strategy.

A. *String-submodularity*

Submodularity of functions over finite sets plays an important role in combinatorial optimization. It has been shown that the greedy strategy provides at least a constant-factor approximation to the optimal strategy. For example, the celebrated result of Nemhauser *et al.* [46] states that to maximize a monotone submodular function F with$F(\emptyset) = 0$ (here $\emptyset$ denotes the empty set) over a uniform matroid, the reward of the greedy strategy is at least a factor $(1 - e^{-1})$ of the reward of the optimal strategy. Submodularity has been shown, in a sensing context, to be related to several information theoretic "rewards" such as mutual information [47] and mean square error [48][49]. However, the connection between the error probability in detection problem and submodularity has not previously been investigated. In this paper, we will establish the connection in the context of balanced binary relay trees.

Note that the submodular functions studied in most previous papers are defined on the power set of a given finite set. Many stochastic optimization problems, however, concern optimization of objective functions over a finite horizon, where an action is chosen from a given finite set at each iteration. In these cases, the objective function usually depends on the order of actions, and repetition of the same action is allowed. Unfortunately, then, the bound obtained by Nemhauser *et al.* [46] is not applicable.

For objective functions defined on strings (finite-length sequences), [50] and [51] provide some sufficient conditions for the greedy strategy to achieve a good approximation to the global optimal strategy. In this paper, we improve these results by providing sufficient conditions that are weaker than those in [50] and [51].

Next we formulate the maximization problem using submodular functions defined on strings.

I. *String*: Consider a finite set $\mathbb{A}$ of possible actions. For each step, we choose an action from $\mathbb{A}$. Let $A = (a_1, a_2, \dots, a_k)$ be a string of actions taken over k steps, where $a_i \in \mathbb{A}$ for $i = 1, 2, \dots, k$. Let the set of all strings of actions be

$$\mathbb{A}^* = \{(a_1, a_2, \dots, a_k) | k = 0, 1, \dots \text{ and } a_i \in \mathbb{A}\ \forall i\}.$$

Note that $k = 0$ corresponds to the empty string (no action taken), denoted by $\emptyset$.

II. *String length*: For a given string $A = (a_1, a_2, \dots, a_k)$, we define its *string length* as k, denoted by $|A| = k$.

III. *String concatenation*: Let $M = (a_1^m, a_2^m, \dots, a_{k_1}^m)$ and $N = (a_1^n, a_2^n, \dots, a_{k_2}^n)$ be two strings in $\mathbb{A}^*$. Their concatenation is

$$M \oplus N = (a_1^m, a_2^m, \dots, a_{k_1}^m, a_1^n, a_2^n, \dots, a_{k_2}^n).$$

IV. *String dominance*: Let M and N be two strings in $\mathbb{A}^*$. We write $M \preceq N$ if

$$N = M \oplus (a_1, a_2, \dots, a_j),$$

where $j \in \{0, 1, \dots\}$ and $a_i \in \mathbb{A}$ for all i. In other words, M is a *prefix* of N.

V. *String-submodularity*: A function from strings to real numbers, $f : \mathbb{A}^* \to \mathbb{R}$, is string-submodular if

i. f has the *monotone* property; i.e.,

$$f(M) \leq f(N), \quad \forall M \preceq N \in \mathbb{A}^*.$$

ii. f has the *diminishing-return* property; i.e.,

$$f(M \oplus (a)) - f(M) \geq f(N \oplus (a)) - f(N),$$
$$\forall M \preceq N \in \mathbb{A}^*, \forall a \in \mathbb{A}.$$

Note that the diminishing-return only requires concatenation of one additional action.

VI. *Globally optimal solution*: Consider the problem of finding a string that maximizes f under the constraint that the string length is not larger than K. Because the function f is monotone, it suffices to consider the stronger constraint with fixed length K:

$$\begin{aligned} &\text{maximize } f(M) \\ &\text{subject to } M \in \mathbb{A}^*, |M| = K. \end{aligned} \tag{2}$$

VII. *Greedy solution*: A string $G_j = (a_1^*, a_2^*, \dots, a_j^*)$ is called *greedy* if

$$\begin{aligned} a_i^* = \arg\max_{a_i \in \mathbb{A}} & f((a_1^*, a_2^*, \dots, a_{i-1}^*, a_i)) \\ & - f((a_1^*, a_2^*, \dots, a_{i-1}^*)) \quad \forall i = 1, 2, \dots, j. \end{aligned}$$

For a deterministic dynamical system, the diminishing return property can be simplified to the following.

Lemma 1: For any $M,\ N \in \mathbb{A}^*$, there exists an element $a \in \mathbb{A}$ such that

$$f(M \oplus (a)) - f(M) \geq \frac{1}{|N|}(f(M \oplus N) - f(M)).$$

Proof: Let $M, N \in \mathbb{A}^*$, write $N = (a_1, a_2, \ldots, a_n)$. We have

$$f(M \oplus N) - f(M) = \sum_{i=1}^{n}((f(M \oplus (a_1, \ldots, a_i)) - f(M \oplus (a_1, \ldots, a_{i-1})))).$$

It is easy to see that there exists an index $1 \le j \le n$ such that

$$f(M \oplus (a_1, \ldots, a_j)) - f(M \oplus (a_1, \ldots, a_{j-1})) \ge \frac{1}{|N|}(f(M \oplus N) - f(M)).$$

The diminishing-return property gives

$$\begin{aligned} &f(M \oplus (a_j)) - f(M) \ge \\ &f(M \oplus (a_1, \ldots, a_j)) - f(M \oplus (a_1, \ldots, a_{j-1})) \\ &\ge \frac{1}{|N|}(f(M \oplus N) - f(M)). \end{aligned}$$

■

Theorem 2: Consider a submodular function $f : \mathbb{A}^* \to \mathbb{R}$ such that

i. $f(\emptyset) = 0$;
ii. For any greedy strings G_i with a length less than K and optimal strings O (optimal with respect to (2)), we have $f(G_i \oplus O) \ge f(O)$, $i = 1, 2, \ldots, K-1$.

It follows that any greedy string G_K of length K satisfies

$$f(G_K) > (1 - e^{-1})f(O).$$

Proof: For the greedy strategy,

$$\begin{aligned} &f((a_1^*, a_2^*, \ldots, a_i^*)) - f((a_1^*, a_2^*, \ldots, a_{i-1}^*)) \ge \\ &f((a_1^*, a_2^*, \ldots, a)) - f((a_1^*, a_2^*, \ldots, a_{i-1}^*)) \quad \forall a \in \mathbb{A}. \end{aligned}$$

Combining this with Lemma 1, we obtain

$$\begin{aligned} &f((a_1^*, a_2^*, \ldots, a_i^*)) - f((a_1^*, a_2^*, \ldots, a_{i-1}^*)) \ge \\ &\frac{1}{K}\left(f((a_1^*, a_2^*, \ldots, a_{i-1}^*) \oplus O) - f((a_1^*, a_2^*, \ldots, a_{i-1}^*))\right) \end{aligned}$$

From the assumption, we have $f((a_1^*, a_2^*, \ldots, a_{i-1}^*) \oplus O) \ge f(O)$. Then, for $i = 1, 2, \ldots, k-1$,

$$\begin{aligned} &f((a_1^*, a_2^*, \ldots, a_i^*)) - f((a_1^*, a_2^*, \ldots, a_{i-1}^*)) \ge \\ &\frac{1}{K}\left(f(O) - f((a_1^*, a_2^*, \ldots, a_{i-1}^*))\right). \end{aligned}$$

This gives

$$f((a_1^*, a_2^*, \ldots, a_i^*)) \ge \frac{1}{K}f(O) + (1 - \frac{1}{K})f((a_1^*, a_2^*, \ldots, a_{i-1}^*)),$$

which, applied recursively, yields

$$\begin{aligned} &f((a_1^*, a_2^*, \ldots, a_K^*)) \ge \frac{1}{K}f(O) \\ &+ (1 - \frac{1}{K})\left(\frac{1}{K}f(O) + (1 - \frac{1}{K})f((a_1^*, a_2^*, \ldots, a_{K-2}^*))\right) \\ &= \frac{1}{K}f(O)\sum_{i=0}^{K-1}(1 - \frac{1}{K})^i \\ &= (1 - (1 - \frac{1}{K})^K)f(O). \end{aligned}$$

Since $\lim_{K\to\infty}(1 - 1/K)^K \nearrow e^{-1}$ we obtain

$$f((a_1^*, a_2^*, \ldots, a_K^*)) > (1 - e^{-1})f(O)$$

for all K.

■

B. Application to Distributed Detection in Balanced Binary Relay Trees

In this section, we will apply the framework of string-submodularity to the distributed detection problem in balanced binary relay trees. Note that one property of string-submodularity is monotonicity. In balanced binary relay trees, if we consider all possible fusion rules for one level, then the monotone property cannot be guaranteed; in other words, the total error probability does not necessarily decrease after one level of data integration. However, if we consider two consecutive alternative fusion rules together, namely $(\mathcal{A}, \mathcal{O})$ or $(\mathcal{O}, \mathcal{A})$, then the monotone property is satisfied and string-submodular framework can be directly applied to the distributed detection problem. Therefore, we will only consider the case of balanced binary relay trees with even heights and we assume that two fusion rules Λ of consecutive levels k (even integer) and $k+1$ are chosen from the following set $\mathcal{Z} = \{(\mathcal{A}, \mathcal{O}), (\mathcal{O}, \mathcal{A})\}$. Let $\Pi = (\Lambda_1, \Lambda_2, \ldots, \Lambda_h)$ be a fusion strategy, where $\Lambda_i \in \mathcal{Z}$ for $i = 1, \ldots, h$. Let $\mathcal{Z}^*$ be the set of all possible strategies (strings); i.e., $\mathcal{Z}^* = \{(\Lambda_1, \Lambda_2, \ldots, \Lambda_h) | h = 0, 1, \ldots \text{ and } \Lambda_i \in \mathcal{Z}\ \forall i\}$. Here we only prove the case where the prior probabilities are equally likely. The following analysis easily generalizes to non-equal prior probabilities by weighting the Types I and II error probabilities with the priors. Given the two types of error probability (α_0, β_0) at level 0, the reduction of the total error probability after applying a strategy Π is

$$u(\Pi) = \alpha_0 + \beta_0 - (\alpha_{2h}(\Pi) + \beta_{2h}(\Pi)),$$

where α_{2h} and β_{2h} represent the Type I and II error probabilities at level $2h$ after fusion with Π.

We now establish that u is string-submodular under suitable initial conditions.

Proposition 1: For sufficiently small (α_0, β_0), the function u: $\mathcal{Z}^* \to \mathbb{R}$ is string-submodular.

Proof: First we show that u is monotone. It suffices to show the following:

$$u((\Lambda_1, \ldots, \Lambda_k) \oplus (\Lambda^*)) \ge u((\Lambda_1, \ldots, \Lambda_k)),$$

for all $\Lambda_i, \Lambda^* \in \mathcal{Z}$, where $i = 1, 2, \ldots, k$. We will use (α_k, β_k) to denote the error probabilities after using $(\Lambda_1, \ldots, \Lambda_k)$. If $\Lambda^* = (\mathcal{A}, \mathcal{O})$, then we need to show that

$$\begin{aligned} &u((\Lambda_1, \ldots, \Lambda_k) \oplus (\Lambda^*)) - u((\Lambda_1, \ldots, \Lambda_k)) \\ &= \alpha_k + \beta_k - (1 - (1 - \alpha_k)^2)^2 - (1 - (1 - \beta_k^2)^2) \\ &= f_{\alpha_k} + f_{\beta_k} \ge 0, \end{aligned}$$

where $f_{\alpha_k} = \alpha_k - (1 - (1 - \alpha_k)^2)^2$ and $f_{\beta_k} = \beta_k - (1 - (1 - \beta_k^2)^2)$. It is evident that f_α and f_β are non-negative if α_k and β_k are sufficiently small. More precisely, if $\alpha_k \le 0.3$ and $\beta_k \le 0.3$, then the function u is monotone increasing.

In other words, if the initial error probabilities α_0 and β_0 are sufficiently small, u is monotone increasing. Plots of f_{α_k} and f_{β_k} against α_k and β_k, respectively, are given in Figure 4. If $\Lambda^* = (\mathcal{O}, \mathcal{A})$, then

$$u(\Lambda^*) = \alpha_k + \beta_k - (1 - (1 - \alpha_k^2)^2) - (1 - (1 - \beta_k)^2)^2 \geq 0,$$

which also holds for sufficiently small α_k and β_k. This can also be proved using the symmetry property of the problem.

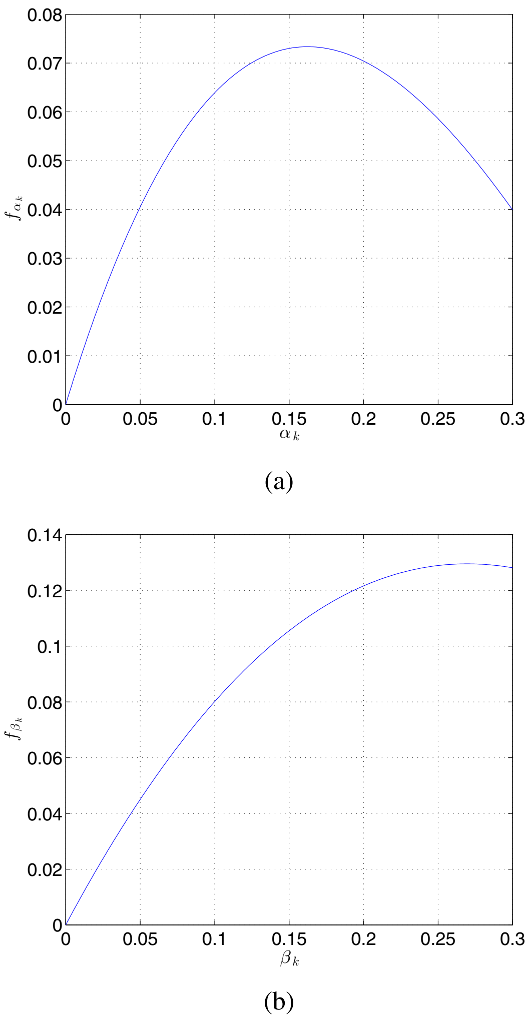


Fig. 4. (a) Values of f_{α_k} versus α_k. (b) Values of f_{β_k} versus β_k.

Now we turn to the diminishing return property of u; that is,

$$u((\Lambda_1, \Lambda_2, \ldots, \Lambda_m) \oplus (\Lambda^*)) - u((\Lambda_1, \Lambda_2, \ldots, \Lambda_m)) \geq$$
$$u((\Lambda_1, \Lambda_2, \ldots, \Lambda_n) \oplus (\Lambda^*)) - u((\Lambda_1, \Lambda_2, \ldots, \Lambda_n))$$

for all $m \leq n$, where $\Lambda_i \in \mathcal{Z}$ for all i and $\Lambda^* \in \mathcal{Z}$. First let us consider the simplest case where $m = 0$ and $n = 1$; that is,

$$u((\Lambda^*)) - u(\emptyset) \geq u((\Lambda_1, \Lambda^*)) - u((\Lambda_1)), \tag{3}$$

for all $\Lambda_1, \Lambda^* \in \mathcal{Z}$. We know that $u(\emptyset) = 0$ because the error probabilities do not change without fusion. Because of symmetry, it suffices to show the above inequality for the cases where $(\Lambda_1, \Lambda^*) = (\mathcal{A}, \mathcal{O}) \oplus (\mathcal{A}, \mathcal{O})$ and $(\Lambda_1, \Lambda^*) = (\mathcal{A}, \mathcal{O}) \oplus (\mathcal{O}, \mathcal{A})$. The error probabilities evolves as follows:

$$(\alpha_k, \beta_k) \xrightarrow{\Lambda_1} (\alpha_{k+2}, \beta_{k+2}) \xrightarrow{\Lambda^*} (\alpha_{k+4}, \beta_{k+4}).$$

Again because of symmetry, we only consider the evolution for α_k.

Case i: If $(\Lambda_1, \Lambda^*) = (\mathcal{A}, \mathcal{O}) \oplus (\mathcal{A}, \mathcal{O})$, then we can show the following

$$\begin{aligned} &\alpha_{k+4} - \alpha_{k+2} - (\alpha_{k+2} - \alpha_k) = \\ &- \alpha_k^{16} + 8\alpha_k^{14} - 24\alpha_k^{12} + 32\alpha_k^{10} \\ &- 14\alpha_k^8 - 8\alpha_k^6 + 10\alpha_k^4 - 4\alpha_k^2 + \alpha_k \geq 0, \end{aligned}$$

which holds for sufficiently small α_k. See Figure 5(a) for a plot of $\alpha_{k+4} - \alpha_{k+2} - (\alpha_{k+2} - \alpha_k)$ versus α_k. Notice that if $\alpha_k < 0.3$, then the above inequality holds. This analysis easily generalizes to the inequality for the Type II error probability by symmetry.

Case ii: If $(\Lambda_1, \Lambda^*) = (\mathcal{A}, \mathcal{O}) \oplus (\mathcal{O}, \mathcal{A})$, then

$$\begin{aligned} &\alpha_{k+4} - \alpha_{k+2} - (\bar{\alpha}_{k+2} - \alpha_k) = \\ &- \alpha_k^{16} + 2\alpha_k^8 + 2\alpha_k^4 - 4\alpha_k^2 + \alpha_k \geq 0, \end{aligned}$$

which holds for sufficiently small α_k. We note that $\bar{\alpha}_{k+2}$ denotes the Type I error probability after using Λ^*. See Figure 5(b) for a plot of $\alpha_{k+4} - \alpha_{k+2} - (\bar{\alpha}_{k+2} - \alpha_k)$ versus α_k. Notice that if $\alpha_k < 0.25$, then the above inequality holds. In other words, the inequality (3) holds for the simplest case.

From this case, it is easy to show (4). Then by recursion, we have

$$\begin{aligned} &u((\Lambda_1, \Lambda_2, \ldots, \Lambda_m) \oplus (\Lambda^*)) - u((\Lambda_1, \Lambda_2, \ldots, \Lambda_m)) \\ &\geq u((\Lambda_1, \Lambda_2, \ldots, \Lambda_n) \oplus (\Lambda^*)) - u((\Lambda_1, \Lambda_2, \ldots, \Lambda_n)) \end{aligned}$$

for all $m \leq n$, where $\Lambda_i \in \mathcal{Z}$ for all i and $\Lambda^* \in \mathcal{Z}$. ■

For a balanced binary relay tree with height $2K$, the global optimization problem is to find a strategy $\Pi \in \mathcal{Z}^*$ with length K such that the above reduction is maximized; that is

$$\begin{aligned} &\text{maximize } u(\Pi) \\ &\text{subject to } \Pi \in \mathcal{Z}^*, |\Pi| = K. \end{aligned} \tag{5}$$

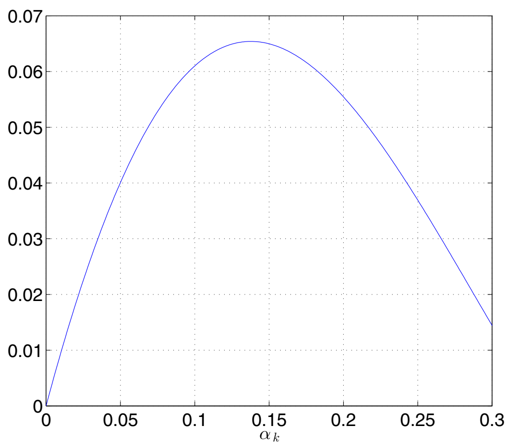
We have shown that reduction of the total error probability u is a string-submodular function. Moreover, we know that the total error probability does not change if there is no fusion; that is,

$$u(\emptyset) = 0.$$

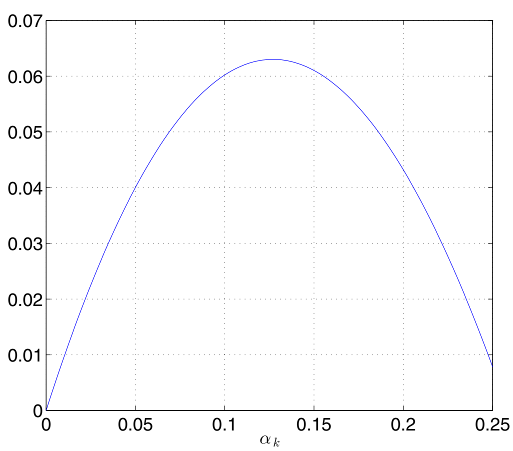
It follows that we can employ Theorem 2 to the above maximization problem (5).

Consider a balanced binary relay tree with height $2K$. We denote by $u(G_K)$ the reduction of the total error probability after using the greedy strategy G_K. We have shown that the ULRT strategy is 2-optimal. Moreover, we have also shown in [41] that the ULRT strategy only allows at most two identical consecutive fusion rules after the error probability pair enters a certain regime in the (α, β) plane. Hence, a strategy is the ULRT strategy if and only if it is the greedy strategy. We denote by $u(O)$ the reduction of the total error probability using the optimal strategy. We have the following theorem.

$$
\begin{aligned}
&u((\Lambda_1, \Lambda_2, \ldots, \Lambda_m) \oplus (\Lambda^*)) - u((\Lambda_1, \Lambda_2, \ldots, \Lambda_m)) \\
&\geq u((\Lambda_1, \Lambda_2, \ldots, \Lambda_m, \Lambda_{m+1}) \oplus (\Lambda^*)) - u((\Lambda_1, \Lambda_2, \ldots, \Lambda_m, \Lambda_{m+1})) \\
&\geq u((\Lambda_1, \Lambda_2, \ldots, \Lambda_{m+1}, \Lambda_{m+2}) \oplus (\Lambda^*)) - u((\Lambda_1, \Lambda_2, \ldots, \Lambda_{m+1}, \Lambda_{m+2})).
\end{aligned} \tag{4}
$$



(a)



(b)

Fig. 5. (a) Values of $\alpha_{k+4} \quad \alpha_{k+2} \quad (\alpha_{k+2} \quad \alpha_k)$ versus α_k. (b) Values of $\alpha_{k+4} \quad \alpha_{k+2} \quad (\bar{\alpha}_{k+2} \quad \alpha_k)$ versus α_k.

Theorem 3: For a balanced binary relay tree with height $2K$,

$$(1 - e^{-1})u(O) < u(G_K) \leq u(O).$$

Proof: The inequality on the right hand side holds because

$$u(\Pi) \leq u(O)$$

for all $\Pi \in \Pi^*$ where $|\Pi| = K$.

For the inequality on the left hand side, we have shown that u is a string-submodular function with $u(\emptyset) = 0$. For any greedy string G, we have $u(G) \geq 0$ because of the monotone property. In addition, the Type I and II error probabilities both decrease after applying the string G. In consequence, the Type I error probability after G is not larger than the initial Type I error probability.

We know that the mapping $s_\Pi : \alpha_k \to \alpha_{k+2}$ is a monotone non-decreasing function of α_k for any fusion rule $\Pi \in \mathcal{Z}$. Also the optimal string O is simply a composite of several such monotone non-decreasing functions. Hence, the Type I error probability after applying O is a monotone non-decreasing function with respect to the initial Type I error probability. With these, we conclude that $u(G \oplus O) \geq u(O)$ for any greedy string G.

We know that u is a string-submodular function from Proposition 1, and that $u(\emptyset) = 0$. With an application of Theorem 2 the proof is completed. ■

Remark: Recall that the fusion strategy is a string of fusion rules chosen from $\mathcal{Z} = \{(\mathcal{A}, \mathcal{O}), (\mathcal{O}, \mathcal{A})\}$, and so the strategies we considered in this section have at most two consecutive repeated fusion rules. For example, the strategy $(\mathcal{A}, \mathcal{A}, \mathcal{A}, \ldots)$ is not considered. Future work includes the performance analysis for the ULRT strategy with respect the optimal strategy among all possible strategies. On the other hand, it is easy to show that with repeating identical fusion rule, the total error probability goes to $1/2$, so that it is reasonable to rule out this situation.

C. Node and Link Failures

We now consider balanced binary relay trees with node and link failures, for which case the decay rate of the total error probability has been considered in [42]. A failure probability of n_k is assumed for each node at level k, and each link is modeled as a binary erasure channel; that is, with some specified probability the input message X (either 0 or 1) gets erased and the receiver does not get any data. We assume that the links between nodes at height k and height $k+1$ all have identical probability of erasure ℓ_k.

Consider a node $\mathcal{N}_k$ at level k connected to its parent node $\mathcal{N}_{k+1}$ at level $k+1$. We define several probabilities as follows:

- Local failure probability p_k: the probability that either the node $\mathcal{N}_k$ fails or the link from $\mathcal{N}_k$ to $\mathcal{N}_{k+1}$ fails.
- Silence probability q_k: the probability that $\mathcal{N}_{k+1}$ does not receive a message from $\mathcal{N}_k$.

From the above definition, we have

$$p_k = n_k + \ell_k - n_k \ell_k.$$

By the law of total probability, we have

$$q_{k+1} = q_k^2 + p_{k+1}(1 - q_k^2).$$

We can view q_k as the exogenous input of the tree network. In this case, the evolution of the Type I and II error probabilities are give in (6).

$$(\alpha_{k+1},\beta_{k+1}) := \begin{cases} \left(\frac{(1-q_k)(1-(1-\alpha_k)^2)+2q_k\alpha_k}{1+q_k}, \frac{(1-q_k)\beta_k^2+2q_k\beta_k}{1+q_k}\right), & \text{if } \lambda=\mathcal{A}, \\ \left(\frac{(1-q_k)\alpha_k^2+2q_k\alpha_k}{1+q_k}, \frac{(1-q_k)(1-(1-\beta_k)^2)+2q_k\beta_k}{1+q_k}\right), & \text{if } \lambda=\mathcal{O}. \end{cases} \quad (6)$$

Again, we consider a balanced binary relay tree with even height $2h$ with consecutive fusion rules are chosen from $\mathcal{Z}$.

Proposition 2: Suppose that the silence probability sequence is bounded above by $1/8$; that is, $q_k < 1/8$ for all k. If the initial error probabilities (α_0, β_0) are sufficiently small, then the function u: $\Pi^* \to \mathbb{R}$ is string-submodular.

Proof: We first show that u is non-decreasing. It suffices to show that $u((\mathcal{A},\mathcal{O})) \geq 0$ starting from (α_k, β_k). We can decompose $(\alpha_{k+2}, \beta_{k+2})$ into different components, for example, the expression for α_{k+2} is given in (7). Notice that in (7), the coefficient for α_k is $\frac{4q_k}{(1+q_k)(1+q_{k+1})}$. If $q_k < 1/8$ for all k, it follows that the coefficient of α_k in (7) is less than 1, which in turn implies that $\alpha_{k+2} \leq \alpha_k$ for sufficiently small α_k.

To show the diminishing return property, it suffice to consider the situation where we use the rule $(\mathcal{A},\mathcal{O}) \oplus (\mathcal{A},\mathcal{O})$. In this case, we need to show $\alpha_k + \alpha_{k+4} - 2\alpha_{k+2} \geq 0$. Again, the expression for α_{k+2} is given in (7). Notice that if we write $\alpha_k + \alpha_{k+4} - 2\alpha_{k+2}$ as a function of α_k, then the coefficient for α_k is not less than $1 - \frac{8q_k}{(1+q_k)(1+q_{k+1})}$, and hence if $q_k \leq 1/8$ the coefficient for α_k is positive. Therefore, we have $\alpha_k + \alpha_{k+4} - 2\alpha_{k+2} \geq 0$ for sufficiently small α_k. We stress that our argument relies on the fact that the initial error probabilities (α_0, β_0) are both sufficiently small. ■

With a similar analysis in the non-failure case, the following bounds, capturing the relative performance of the greedy and optimal strategies are achieved.

Corollary 1: Consider a balanced binary relay tree with height $2K$ with node and link failures. Let G_K and O denote the greedy and optimal strategies, respectively. We have

$$(1 - e^{-1})u(O) < u(G_K) \leq u(O).$$

V. Concluding Remarks

We have studied the problem of finding a fusion strategy that maximizes the reduction of the total error probability in balanced binary relay trees. This optimization problem has been reformulated using dynamic programming and its solution characterized via the Bellman equation. We have studied the notion of string submodularity, and have shown that the reduction in total error probability is a string submodular function. In consequence, the ULRT strategy, which is a string of repeated ULRT fusion rules, has a performance comparable with the globally optimal strategy measured in terms of the reduction in total error probability.

Future work includes studying the submodularity property for other architectures (for instance, bounded-height trees [31] and M-ary relay trees). We have assumed in this paper that all of the nodes at the same level use identical fusion rules. This raises the question of what will happen when this constraint is relaxed and the nodes at each level are permitted different fusion rules. Also of significant interest is the case when the trees have correlated sensor measurements. These questions are currently under consideration by us.

References

[1] Z. Zhang, E. K. P. Chong, A. Pezeshki, W. Moran, and S. D. Howard, "Submodularity and optimality of fusion rules in balanced binary relay trees," in *Proc. of 51th IEEE Conference on Decision and Control (CDC)*, Maui, Hawaii, December 10–13, 2012, pp. 3802–3807.

[2] I. Lobel, D. Acemoglu, M. A. Dahleh, and A. Ozdaglar, "Lower bounds on the rate of learning in social networks," in *Proc. of IEEE American Control Conference*, Hyatt Regency Riverfront, St. Louis, MO, June 10-12, 2009, pp. 2825–2830.

[3] K. Drakopoulos, A. Ozdaglar, and J. N. Tsitsiklis, "On learning with finite memory," *IEEE Trans. on Information Theory*, vol. 59, no. 10, pp. 6859–6872, Oct. 2013.

[4] 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.

[5] A. V. Banarjee, "A simple model for herd behavior," *Quart. J. Econ.*, vol. 107, no. 3, pp. 797–817, Aug. 1992.

[6] 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.

[7] R. R. Tenney and N. R. Sandell, "Detection with distributed sensors," *IEEE Trans. on Aerospace and Electronic Systems*, no. 4, pp. 501–510, 1981.

[8] P. Varshney *et al.*, "Optimal data fusion in multiple sensor detection systems," *IEEE Trans. on Aerospace and Electronic Systems*, no. 1, pp. 98–101, 1986.

[9] J.-F. Chamberland and V. V. Veeravalli, "Asymptotic results for decentralized detection in power constrained wireless sensor networks," *IEEE J. on Selected Areas in Communications*, vol. 22, no. 6, pp. 1007–1015, 2004.

[10] J. N. Tsitsiklis, "Decentralized detection," *Advances in Statistical Signal Processing*, vol. 2, pp. 297–344, 1993.

[11] G. Polychronopoulos and J. N. Tsitsiklis, "Explicit solutions for some simple decentralized detection problems," *IEEE Trans. on Aerospace and Electronic Systems*, vol. 26, no. 2, pp. 282–292, 1990.

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

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

[14] R. Viswanathan and P. K. Varshney, "Distributed detection with multiple sensors i. fundamentals," *Proceedings of the IEEE*, vol. 85, no. 1, pp. 54–63, 1997.

[15] R. S. Blum, S. A. Kassam, and H. V. Poor, "Distributed detection with multiple sensors ii. advanced topics," *Proceedings of the IEEE*, vol. 85, no. 1, pp. 64–79, 1997.

[16] T. M. Duman and M. Salehi, "Decentralized detection over multiple-access channels," *IEEE Trans. on Aerospace and Electronic Systems*, vol. 34, no. 2, pp. 469–476, 1998.

[17] 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. on Information Theory*, vol. 51, no. 2, pp. 693–699, 2005.

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

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

$$\alpha_{k+2} = \frac{(1-q_k)(1-q_{k+1})}{(1+q_k)(1+q_{k+1})}(1-(1-\alpha_k)^2)^2 + \frac{2q_k(1-q_{k+1})}{(1+q_k)(1+q_{k+1})}\alpha_k^2 + \frac{2q_{k+1}(1-q_k)}{(1+q_k)(1+q_{k+1})}(1-(1-\alpha_k)^2) + \frac{4q_kq_{k+1}}{(1+q_k)(1+q_{k+1})}\alpha_k. \tag{7}$$

[20] 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. on Information Theory*, vol. 52, no. 3, pp. 1274–1275, 2006.

[21] H. Chen, B. Chen, and P. K. Varshney, "Further results on the optimality of the likelihood-ratio test for local sensor decision rules in the presence of nonideal channels," *IEEE Trans. on Information Theory*, vol. 55, no. 2, pp. 828–832, 2009.

[22] G. Fellouris and G. V. Moustakides, "Decentralized sequential hypothesis testing using asynchronous communication," *IEEE Trans. on Information Theory*, vol. 57, no. 1, pp. 534–548, 2011.

[23] P. K. Varshney and C. Burrus, *Distributed Detection and Data Fusion*. Springer, 1997.

[24] T. M. Cover, "Hypothesis testing with finite statistics," *The Annals of Mathematical Statistics*, pp. 828–835, 1969.

[25] Z. Tang, K. Pattipati, and D. Kleinman, "Optimization of detection networks. i. tandem structures," *IEEE Trans. on Systems, Man and Cybernetics*, vol. 21, no. 5, pp. 1044–1059, 1991.

[26] R. Viswanathan, S. C. Thomopoulos, and R. Tumuluri, "Optimal serial distributed decision fusion," *IEEE Trans. on Aerospace and Electronic Systems*, vol. 24, no. 4, pp. 366–376, 1988.

[27] W. P. Tay, J. N. Tsitsiklis, and M. Z. Win, "On the subexponential decay of detection error probabilities in long tandems," *IEEE Trans. on Information Theory*, vol. 54, no. 10, pp. 4767–4771, 2008.

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

[29] V. Veeravalli, "Topics in decentralized detection," Ph.D. dissertation, University of Illinois, Urbrana Champiagn, 1992.

[30] Z.-B. Tang, K. R. Pattipati, and D. L. Kleinman, "Optimization of detection networks. ii. tree structures," *IEEE Trans. on Systems, Man and Cybernetics*, vol. 23, no. 1, pp. 211–221, 1993.

[31] W. P. Tay, J. N. Tsitsiklis, and M. Z. Win, "Data fusion trees for detection: Does architecture matter?" *IEEE Trans. on Information Theory*, vol. 54, no. 9, pp. 4155–4168, 2008.

[32] A. Reibman and L. Nolte, "Design and performance comparison of distributed detection networks," *IEEE Trans. on Aerospace and Electronic Systems*, no. 6, pp. 789–797, 1987.

[33] W. P. Tay and J. N. Tsitsiklis, "Error exponents for decentralized detection in tree networks," in *Networked Sensing Information and Control*. Springer, 2008, pp. 73–92.

[34] W. P. Tay, J. N. Tsitsiklis, and M. Z. Win, "Bayesian detection in bounded height tree networks," *IEEE Trans. on Signal Processing*, vol. 57, no. 10, pp. 4042–4051, 2009.

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

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

[37] S. Alhakeem and P. Varshney, "A unified approach to the design of decentralized detection systems," *IEEE Trans. on Aerospace and Electronic Systems*, vol. 31, no. 1, pp. 9–20, 1995.

[38] Y. Lin, B. Chen, and P. K. Varshney, "Decision fusion rules in multihop wireless sensor networks," *IEEE Trans. on Aerospace and Electronic Systems*, vol. 41, no. 2, pp. 475–488, 2005.

[39] B. Kailkhura, A. Vempaty, and P. K. Varshney, "Distributed Inference in Tree Networks using Coding Theory," *IEEE Trans. on Signal Processing,* vol. 63, no. 14, pp. 3715–3726, 2015.

[40] J. A. Gubner, E. K. Chong, and L. L. Scharf, "Aggregation and compression of distributed binary decisions in a wireless sensor network," in *Proceedings of the 48th IEEE Conference on Decision and Control held jointly with the 2009 28th Chinese Control Conference,* Shanghai, P. R. China, Dec. 16-18 2009, pp. 909–913.

[41] Z. Zhang, A. Pezeshki, W. Moran, S. D. Howard, and E. K. Chong, "Error probability bounds for balanced binary relay trees," *IEEE Trans. on Information Theory*, vol. 58, no. 6, pp. 3548–3563, 2012.

[42] Z. Zhang, E. Chong, A. Pezeshki, W. Moran, and S. Howard, "Detection performance in balanced binary relay trees with node and link failures," *IEEE Trans. on Signal Processing*, vol. 61, no. 9, pp. 2165–2177, May 2013.

[43] Y. Kanoria and A. Montanari, "Subexponential convergence for information aggregation on regular trees," in *2011 50th IEEE Conference on Decision and Control and European Control Conference (CDC-ECC),* . IEEE, 2011, pp. 5317–5322.

[44] Z. Zhang, E. Chong, A. Pezeshki, W. Moran, and S. Howard, "Learning in hierarchical social networks," *IEEE J. of Selected Topics in Signal Processing*, vol. 7, no. 2, pp. 305–317, Apr. 2013.

[45] Z. Zhang, E. Chong, A. Pezeshki, and W. Moran, "String sumbodular functions with curvature constraints," to appear in *IEEE Trans. on Automatic Control*.

[46] G. L. Nemhauser, L. A. Wolsey, and M. L. Fisher, "An analysis of approximations for maximizing submodular set functions—i," *Mathematical Programming*, vol. 14, no. 1, pp. 265–294, 1978.

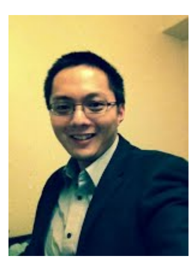
[47] A. Krause and C. Guestrin, "Near-optimal observation selection using submodular functions," in *Proceedings of the National Conference on Artificial Intelligence*, vol. 22, no. 2. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999, 2007, p. 1650.

[48] V. Cevher and A. Krause, "Greedy dictionary selection for sparse representation," *IEEE J. of Selected Topics in Signal Processing*, vol. 5, no. 5, pp. 979–988, 2011.

[49] A. Das and D. Kempe, "Submodular meets spectral: Greedy algorithms for subset selection, sparse approximation and dictionary selection," *arXiv preprint arXiv:1102.3975*, 2011.

[50] M. Streeter and D. Golovin, "An online algorithm for maximizing submodular functions," in *Neural Information Processing Systems (NIPS), 22nd Annual Conference on*, 2008.

[51] S. Alaei and A. Malekian, "Maximizing sequence-submodular functions and its application to online advertising," *arXiv preprint arXiv:1009.4153*, 2010.



Zhenliang Zhang (S'09–M'13) received the B.S. (Physics) degree from Special Class for Gifted Young, University of Science and Technology of China, Hefei, China, in 2008. He received the M.S. degree from Departments of Physics, Ph.D degree from Department of Electrical and Computer Engineering, Colorado State University in 2010 and 2013, respectively.

He is currently a visiting professor at Shenzhen Institutes of Advanced Technology. He was a senior engineer at Qualcomm Corporate R&D His research interests include statistical learning in complex network, optimization, and wireless communication.

dwin K. P. Chongdwin K. P. ChongE (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. 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 has also served on the editorial boards of *Computer Networks*, *Journal of Control Science and Engineering*, and *IEEE Expert Now*. He has served as a member of the IEEE Control Systems Society Board of Governors and as Vice President for Financial Activities until 2014. He currently serves as President-Elect. 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 BSc and MSc degrees in electrical engineering from University of Tehran, Tehran, Iran, in 1999 and 2001, respectively. He earned his PhD 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. In August 2008, he joined the faculty of Colorado State University, where he is now an Associate Professor in the Department of Electrical and Computer Engineering, and the Department of Mathematics. His research interests are in statistical signal processing, coding theory, applied harmonic analysis, and bioimaging. He is a member of the editorial board of *IEEE Access*.

William Moran (M95) 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.

Stephen D. Howard graduated in 1982 from La Trobe University, Melbourne, Australia. He received his M.S. and Ph.D. degrees in mathematics from La Trobe University in 1984 and 1990, respectively. He joined the Australian Defence Science and Technology Organisation (DSTO) in 1991, where he has been involved in research in the area electronic surveillance and radar systems. He has led the DSTO research effort into the development of algorithms in all areas of electronic surveillance, including radar pulse train deinterleaving, precision radar parameter estimation and tracking, estimation of radar intrapulse modulation, and advanced geolocation techniques. Since 2003, he has led the DSTO long-range research program in radar resource management and waveform design. He is currently the recipient of a three year DSTO research fellowship.