Decentralized Detection with State: Error Exponents for Informed and Uninformed Fusion Centers

Joni Shaska
University of Southern California
shaska@usc.edu

Urbashi Mitra
University of Southern California
ubli@usc.edu

***Abstract*—Inspired by detection in Internet of Things applications and microbial communities, we formulate the problem of state-dependent decentralized detection. In particular, we consider both the case where the fusion center knows the network state, and the case where it does not. We show that for a min-max cost structure, monotone threshold rules are optimal for the fusion center with state-knowledge, and for the Bayesian case, monotone rules are optimal for both fusion centers. We also compute the Chernoff information for both fusion centers, and show a number of interesting asymptotic properties.**

***Index Terms*—decentralized detection, multi-agent networks, hypothesis testing, state-dependent networks, error exponents**

I. Introduction

The problem of multi-agent detection has persistently been studied over the years, with initial key work [1] wherein a fully distributed system is examined. The optimal rule at each sensor is a likelihood ratio test; however, optimal thresholds are found via coupled non-linear equations. Modern applications necessitating multi-agent decision making include the Internet-of-Things [2], smart grid applications [3], cognitive radio networks, [4], millimeter-wave communications [5], and our motivating interest, decentralized decision making in microbial communities [6]–[12]. We note that multi-agent decision making includes scenarios where each agent determines their own local decision; however, we shall focus on the case where statistics are transmitted from each agent to a centralized decision-maker or fusion center as depicted in Figure 1.

The fusion center framework has been well studied for both a finite number of agents as well as asymptotically large networks [13]–[19]. For key families of cost functions, threshold-base rules, specifically likelihood ratio (LR) quantizers, are optimal [13], [14], [20], [21] and the quantization of such rules has further been examined [22], [23]. Properties of likelihood ratio functions have also been studied [24]–[26].

A classical assumption is that the agents' observations are conditionally independent. If this condition is removed, as we will do herein, the problem of multi-agent detection is, in general, NP-hard [13], [27]. Furthermore, in this case, likelihood ratio based quantizers are no longer optimal [14], [28], challenging analysis. The case of dependent observations has received recent attention [18], [29]–[32]. In particular, a hierarchical conditional independence (HCI) model was introduced in [18] for which analysis of the optimal decision making was enabled. The structure introduced in this paper allows one to solve a class of problems with correlated observations via likelihood ratio based quantizers.

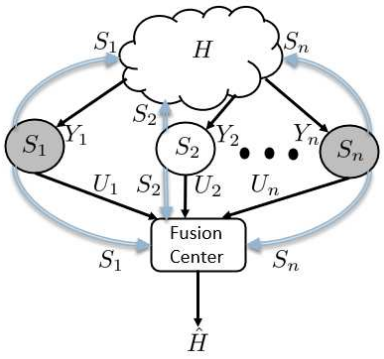


Fig. 1: Problem setup under the state-information structure

As seen in Figure 1, there is a communication link between each agent and the fusion center. Herein, we adopt the general framework of communication over channels with state [33]–[37] which enables the consideration of distorted links in wireless communications as well as directly handling gene expression (and thus different behavioral states) of bacteria [7], [9], [10]. In particular, we will examine two scenarios: one wherein the fusion center knows the network state (the *state-information structure*) and when it does not (the the *no-information structure*). As previously noted, we do not assume conditional independence amongst the observations of each agent.

Our results herein generalize over our first presentation of this work [38]. In particular, only the *state-information structure* was previously considered, whereas herein we also consider the *no-information structure*. Herein a min-max cost is considered versus the purely Bayesian approach in [38]. Furthermore, full proofs are provided.

The main contributions of this paper are as follows:

- We formulate the problem of detection over sensor networks with state. In particular, the observations seen by the agents (the signal and state of each sensor) *are not conditionally independent*.
- For binary hypotheses, we derive the optimal rule for both

This work has been funded in part by one or more of the following grants: ONR 00014-15-1-2550, NSF CCF-1817200, ARO W911NF1910269, Cisco Foundation 1980393, DOE DE-SC0021417, Swedish Research Council 2018-04359, NSF CCF-2008927, ONR 503400-78050.

This paper was presented in part at the 2021 Global Communications Conference (GLOBECOMM).

information structures, which correspond to *monotone* rules.
- For a min-max cost, we derive the optimal rule for the state-information case.
- Even though our observations are not conditionally independent, the ability to restrict ourselves to monotone rules without loss of optimality allows one to use old algorithms such as [22] with only minor adjustments.
- We derive the Chernoff information for both information structures, and show that for both information structures, one can use identical rules without loss of optimality.
- We study the effect of the hypothesis model on the asymptotics. Specifically, we show that if the hypothesis model has not adequately decayed (which corresponds to a set of bounds not being sufficiently tight) one can incur a performance loss by ignoring the hypothesis model, even for relatively large networks.

Notation: Random variables are denoted by capital letters X and realizations by lower case x. Boldfaced capital letters indicate random vectors, $\boldsymbol{V}$, and lower case, vector realizations, $\boldsymbol{v}$. The vector $\boldsymbol{v}^k$ denotes the vector $\boldsymbol{v} \setminus v_k = [v_1, v_2, ..., v_{k-1}, v_{k+1}, ..., v_N]$. The function $p(x)$ is the probability mass function or probability density function of a discrete or continuous random variable, respectively.

II. Problem Formulation

Consider the setup depicted in Fig. 1. a set of n nodes are oriented in a parallel configuration. Each agent receives an observation consisting of the random variable $Y_k \in \mathcal{Y}$, which we call the *signal*, and $S_k \in \mathcal{S}$, which we call the *the state of agent* k, $k = 1, 2, ..., n$. All agents observe the same underlying hypothesis. We denote the state vector as $\boldsymbol{S} \in \mathcal{S}^n$, which is a concatenation of all agent states, i.e., $\boldsymbol{S} = [S_1, S_2, ..., S_n]$. Moreover, the hypothesis $H \in \{0, 1, ..., m\}$ has conditional distribution $p(h|\boldsymbol{s})$, which we call *the hypothesis model*. While we do not assume conditional independence among the observations, i.e., $p(\boldsymbol{y}, \boldsymbol{s}|h) \neq \prod_{k=1}^n p(y_k, s_k|h)$, we do make the following assumption:

Assumption 1. *The signal Y_k is independent of $\boldsymbol{Y}^k$ and $\boldsymbol{S}^k$ conditioned on S_k and H, for $k = 1, 2, ..., n$. That is,*

$$p_h(\boldsymbol{y}|\boldsymbol{s}) = \prod_{k=1}^{n} p_h(y_k|s_k) \tag{1}$$

where $p_h(x)$ denotes the pmf (pdf) of the random variable X conditioned on $H = h$.

We call the conditional distribution $p_h(y_k|s_k)$ the *signal model for agent k*. Upon receiving the pair (Y_k, S_k), agent k makes a local decision $U_k = \gamma_k(Y_k, S_k) \in \mathcal{U} = \{0, 1, ..., b-1\}$. These local decisions are then sent to the fusion center, whose goal is to determine which of the possible hypotheses is true. The fusion center may take one of two information structures:

- *state-information structure*: The state network vector $\boldsymbol{S}$ is known by the fusion center, and the fusion center output is given by $U_0 = \psi(\boldsymbol{U}, \boldsymbol{S}) \in \{0, 1, ..., m-1\}$. Hence, we have the Markov chain structure $\boldsymbol{S} \to (\boldsymbol{S}, H) \to (\boldsymbol{Y}, \boldsymbol{S}) \to (\boldsymbol{U}, \boldsymbol{S}) \to U_0$.
- *no-information structure*: The fusion center only receives $\boldsymbol{U}$, and $U_0 = \psi(\boldsymbol{U}) \in \{0, 1, ..., m-1\}$. Hence, $\boldsymbol{S} \to (\boldsymbol{S}, H) \to (\boldsymbol{Y}, \boldsymbol{S}) \to \boldsymbol{U} \to U_0$.

Let the set Γ be the set of all decision rules and Ψ be the set of all fusion rules. Let $\boldsymbol{\gamma}$ denote the collection of agent rules $\{\gamma_k\}_{k=1}^n$, $\gamma_k \in \Gamma$, $k = 1, 2, ..., n$. We call a collection of agent rules $\boldsymbol{\gamma} \in \Gamma^n$ together with a fusion rule $\psi \in \Psi$ a *strategy* denoted by $\boldsymbol{\psi} \in \Gamma^n \times \Psi$. We also make the following assumption on the relationship between H, $\boldsymbol{S}$, $\boldsymbol{Y}$, $\boldsymbol{U}$, and U_0.

- $\boldsymbol{U}$ is a function only of $\boldsymbol{Y}$ and $\boldsymbol{S}$, with U_k being a function only of Y_k and S_k, i.e., the joint conditional pmf is given as

$$p(\boldsymbol{u}|\boldsymbol{y}, \boldsymbol{s}) = \prod_{k=1}^{n} p(u_k|y_k, s_k) \tag{2}$$

For the rest of the paper, we restrict ourselves to binary hypothesis, $H \in \{h_0, h_1\}$.

It is worth comparing our problem formulation and those previously considered in the literature. The formulation where the fusion center **does not** have access to the state is in part a decentralized detection problem with correlated observations, and so, could be solved under the assumption that the Hierarchical Conditional Independence model (HCI) introduced in [18] holds. However, as we will see in the sequel, the state structure we have introduced into the problem is more amenable to the development of iterative algorithms for finding optimal strategies. Specifically, the state structure allows one to use previous algorithms such as that developed in [22] with minor adjustments. Moreover, our alternative representation of the Chernoff information yields desirable properties that facilitate design in the presence of large networks. As such, the systems introduced in [18], [19] do not lend themselves to this large scale analysis. The formulation where the fusion center **does** know the network state is a solved, and well-understood problem for *a fixed network state* $\boldsymbol{s}$. It may be tempting to fix the network state and solve the problem using previously established methods for each state. However, we will see in Section V, that this approach is, in general, sub-optimal. We provide a discussion on this approach.

III. Optimum Sensor Design

In this section, we show that under a min-max cost criteria and linear combination criteria, there is no loss of optimality if we restrict ourselves to state-dependent monotone likelihood ratio (LR) partition rules. Moreover, we derive the optimal rule for agent k when the other agents rules and the fusion rule are fixed under a Bayesian cost criteria. The later result facilitates the use of iterative algorithms such as [22] with only minor adjustments to the algorithm. We first give the definition of a state-dependent monotone LR partition.

Definition 1. *A state-dependent monotone LR partition is a set of partitions on Ω indexed by s, where each member of the*

partition is denoted as $R_0^s, ..., R_v^s$, such that $\bigcup_{k=0}^{v} R_k^s = \Omega$ for any $s \in \mathcal{S}$. Moreover, for any $s \in \mathcal{S}$, $x \in R_i^s$, $x' \in R_i^s$, $y \in R_j^s$, and $y' \in R_j^s$, if $L_s(x) < L_s(y)$, then $L_s(x') < L_s(y')$, where

$$L_s(x) = \frac{p_1(x|s)}{p_0(x|s)} \tag{3}$$

is the state-dependent likelihood ratio [1].

Thus, a state-dependent monotone LR partition is simply a set of monotone LR partitions on Ω where the partition currently in use depends on s. If an agent is using a state-dependent monotone LR rule, it's code book for any state s has the following property.

Lemma 1. *Given a state-dependent monotone LR partition, for any $s \in \mathcal{S}$, assume R_i^s and R_j^s are two sets with $L_s(x) < L_s(y)$ for $x \in R_i^s$ and $y \in R_j^s$, then $L_s(U_k = i) < L_s(U_k = j)$ where*

$$L_s(U_k = i) = \frac{p_1(U_k = i|s)}{p_0(U_k = i|s)} \tag{4}$$

is the LR of the quantizer output U_k.

Proof. For a fixed $s \in \mathcal{S}$, the proof is the same as that of Lemma 2 in [22]. □

For brevity, from this point on we will simply call state-dependent monotone LR partitions monotone LR partitions.

A. No Prior Formulation

In this section, we assume no prior knowledge of the state. Typically, when one does not have prior information in standard decentralized detection, the hypothesis H is determined *a priori* by nature and knowledge of the prior distribution $p(h)$ is unknown. In our problem, we assume that the hypothesis model $p(h|\boldsymbol{s})$ is *known*, and that instead nature has determined the *state of the system* $\boldsymbol{S}$ *a priori* with the prior distribution $p(\boldsymbol{s})$ unknown. We can write the *expected state cost* under state $\boldsymbol{s}$ as

$$J(\boldsymbol{\psi}|\boldsymbol{s}) = \sum_{u_0=0}^{1} \sum_{h=0}^{1} c_{u_0,h} p(u_0, h|\boldsymbol{s}) \tag{5}$$

where $c_{u_0,h}$ be the cost of deciding $U_0 = u_0$ when $H = h$ is true. We focus on the probability of error cost ($c_{0,0} = c_{1,1} = 0$, and $c_{0,1} = c_{1,0} = 1$), so the expected state cost becomes

$$J(\boldsymbol{\psi}|\boldsymbol{s}) = p(u_0 = 1, h = 0|\boldsymbol{s}) + p(u_0 = 0, h = 1|\boldsymbol{s}) \tag{6}$$

The worst case cost under strategy $\boldsymbol{\psi}$ is

$$\max_{\boldsymbol{s} \in \mathcal{S}^n} J(\boldsymbol{\psi}|\boldsymbol{s}). \tag{7}$$

Hence, we would like to find the strategy $\boldsymbol{\psi}$ that minimizes the worst expected state cost, i.e, we wish to find the $\boldsymbol{\psi}$ that solves the optimization problem

$$\inf_{\boldsymbol{\psi} \in \Gamma^n \times \Psi} \{\max_{\boldsymbol{s} \in \mathcal{S}^n} J(\boldsymbol{\psi}|\boldsymbol{s})\}. \tag{8}$$

[1]More generally, $L_s(x) = \frac{dp_1(\bullet|s)}{dp_0(\bullet|s)}$ is the Radon-Nikodym derivative between the two conditional measures $p_1(\bullet|\boldsymbol{s})$ and $p_0(\bullet|\boldsymbol{s})$

We first wish to find the optimal fusion rule given any collection of agent rules $\boldsymbol{\gamma}$.

Lemma 2. *Given any collection of agent rules $\boldsymbol{\gamma}$, define*

$$\mathcal{A}^{\gamma} = \left\{ (\boldsymbol{u}, \boldsymbol{s}) : \frac{p_1(\boldsymbol{u}|\boldsymbol{s})}{p_0(\boldsymbol{u}|\boldsymbol{s})} \geq \frac{p(h_0|\boldsymbol{s})}{p(h_1|\boldsymbol{s})} \right\}. \tag{9}$$

Then, the optimal rule for the fusion center ψ^* for fixed $\boldsymbol{\gamma}$ is given as

$$\psi^*(\boldsymbol{u}, \boldsymbol{s}) = \begin{cases} 1, & (\boldsymbol{u}, \boldsymbol{s}) \in \mathcal{A}^{\gamma} \\ 0, & (\boldsymbol{u}, \boldsymbol{s}) \in \mathcal{A}^{\gamma^c}. \end{cases} \tag{10}$$

Proof. Note that for the state-information structure, upon observing $\boldsymbol{s}$, the fusion center is tasked with a simple binary hypothesis testing problem, of which the optimal rule is the Maximum a Posteriori (MAP) rule, which is given by (10). □

There are a few important notes regarding the optimal fusion rule.

1) ψ^* is dependent on $\boldsymbol{\gamma}$. In the proof above we dropped this dependence from the notation for brevity. Note that different $\boldsymbol{\gamma}$ will change the statistics of $\boldsymbol{u}$. That is, the set $\mathcal{A}^{\gamma}$ may change under different rules employed by the agents, in turn changing the rule ψ^*. This latter dependence is indeed captured by our notation.
2) Since the optimal rule is simply the MAP rule, it is well known that *randomization at the fusion center does not improve performance*. So we restrict ourselves to deterministic fusion rules.
3) Given a specific set of rules $\boldsymbol{\gamma} \in \Gamma^n$, the optimal fusion rule can be completely characterized by (9) and (10). That is, we treat ψ^* as a deterministic function of $\boldsymbol{\gamma}$.

Further examining the fusion rule, we see that

$$\psi^*(\boldsymbol{u}, \boldsymbol{s}) = 1 \tag{11}$$

$$\iff L_{\boldsymbol{s}}(\boldsymbol{u}) \geq \frac{p(h_0|\boldsymbol{s})}{p(h_1|\boldsymbol{s})} \tag{12}$$

$$\overset{(a)}{\iff} L_{s_1}(u_1) L_{s_2}(u_2) ... L_{s_n}(u_n) \geq \frac{p(h_0|\boldsymbol{s})}{p(h_1|\boldsymbol{s})} \tag{13}$$

$$\overset{(b)}{\iff} L_{s_k}(u_k) \geq \tau(\boldsymbol{u}^k, \boldsymbol{s}) \tag{14}$$

where (a) is due to (1), and (b) is due to the assignment

$$\tau(\boldsymbol{u}^k, \boldsymbol{s}) = \frac{p(h_0|\boldsymbol{s})}{p(h_1|\boldsymbol{s})} \frac{1}{\prod_{\substack{j=1 \\ j \neq k}}^{n} L_{s_k}(u_k)}. \tag{15}$$

Moreover, assume that for any $\gamma_k \in \Gamma$, the codewords for agent k are labelled such that for any $s \in \mathcal{S}$

$$L_s(U_k = j) > L_s(U_k = i) \quad \text{for} \quad j > i. \tag{16}$$

Since the labels of the codewords are irrelevant, if the rule γ_k does not obey (16), we can simply relabel the codewords so that (16) is satisfied without changing the performance. Hence, assumption (16) can be made without loss of generality or optimality. Then, if a rule satisfies (16), (14) implies

$$\begin{aligned} &\psi^*(\boldsymbol{u}, \boldsymbol{s}) = 1 \\ \iff &u_k \geq \eta \end{aligned} \tag{17}$$

where $\eta \in \{0, 1, ..., d-1\}$, and *is dependent on* $\boldsymbol{u}^k$ *and* $\boldsymbol{s}$. Then, we denote by f, the function that maps $\boldsymbol{u}^k$ and $\boldsymbol{s}$ to η, i.e., $f(\boldsymbol{u}^k, \boldsymbol{s}) = \eta$. Before we continue, we note that given agent k's state s, rule γ_k is a function only of the random variable Y_k (since the state S_k is fixed). We introduce, for notational simplicity, the definition of a *prescription*.

Definition 2. *A prescription* $\phi \in \Phi$ *is a function from the signal space* $\mathcal{Y}$ *to the message space* $\mathcal{U}$*. That is,*

$$\phi : \mathcal{Y} \to \mathcal{U}.$$

Notice that any rule $\gamma \in \Gamma$ can equivalently be characterized by a collection of prescriptions indexed by state. Thus, finding the optimal rule γ_k for agent k is equivalent to finding the set of prescriptions $\{\phi_{ks}\}_{s \in \mathcal{S}}$. Moreover, we can write

$$\max_{\boldsymbol{s} \in \mathcal{S}^n} J(\boldsymbol{\psi}|\boldsymbol{s}) = \max_{z \in \mathcal{S}} \max_{\boldsymbol{s}^{kz} \in \mathcal{S}^{n-1}} J(\boldsymbol{\psi}|\boldsymbol{s}^{kz}). \tag{18}$$

Thus, we can focus on finding the optimal set of prescriptions for each agent k. In order to find the optimal strategy, we first wish to find the optimal rule for agent k *when the other agents rules are fixed*. Expanding around U^k when $S_k = z$,

$$\begin{aligned}
&J(\boldsymbol{\psi}|\boldsymbol{s}^{kz}) \\
&= \sum_{\boldsymbol{u}^k} p(U_0 = 1, \boldsymbol{u}^k, h_0|\boldsymbol{s}^{kz}) + p(U_0 = 0, \boldsymbol{u}^k, h_1|\boldsymbol{s}^{kz}) \\
&= \sum_{\boldsymbol{u}^k} p_0(U_0 = 1|\boldsymbol{u}^k, \boldsymbol{s}^{kz}) p_0(\boldsymbol{u}^k|\boldsymbol{s}^{kz}) p(h_0|\boldsymbol{s}^{kz}) + \\
&\quad p_1(U_0 = 0|\boldsymbol{u}^k, \boldsymbol{s}^{kz}) p_1(\boldsymbol{u}^k|\boldsymbol{s}^{kz}) p(h_1|\boldsymbol{s}^{kz}) \\
&= \sum_{\boldsymbol{u}^k} p_0(U_k \geq f(\boldsymbol{u}^k, \boldsymbol{s}^{kz})|\boldsymbol{u}^k, \boldsymbol{s}^{kz}) p_0(\boldsymbol{u}^k|\boldsymbol{s}^k) p(h_0|\boldsymbol{s}^{kz}) + \\
&\quad p_1(U_k < f(\boldsymbol{u}^k, \boldsymbol{s}^{kz})|\boldsymbol{u}^k, \boldsymbol{s}^{kz}) p_1(\boldsymbol{u}^k|\boldsymbol{s}^k) p(h_1|\boldsymbol{s}^{kz}) \\
&= \sum_{\boldsymbol{u}^k} p_0(\phi_{kz}(Y_k) \geq f(\boldsymbol{u}^k, \boldsymbol{s}^{kz})|\boldsymbol{u}^k, \boldsymbol{s}^{kz}) p_0(\boldsymbol{u}^k|\boldsymbol{s}^k) p(h_0|\boldsymbol{s}^{kz}) \\
&\quad + p_1(\phi_{kz}(Y_k) < f(\boldsymbol{u}^k, \boldsymbol{s}^{kz})|\boldsymbol{u}^k, \boldsymbol{s}^{kz}) p_1(\boldsymbol{u}^k|\boldsymbol{s}^k) p(h_1|\boldsymbol{s}^{kz})
\end{aligned} \tag{19}$$

Define

$$\mathcal{F}^{\psi}_{\boldsymbol{s},\eta} = \{\boldsymbol{u}^k : f(\boldsymbol{u}^k, \boldsymbol{s}) = \eta\}.$$

That is, for a given fusion rule ψ and $\boldsymbol{s}$, $\mathcal{F}^{\psi}_{\boldsymbol{s},\eta}$ is the set of $\boldsymbol{u}^k$ such that $f(\boldsymbol{u}^k, \boldsymbol{s}) = \eta$. Then,

$$\begin{aligned}
&J(\boldsymbol{\psi}|\boldsymbol{s}^{kz}) = \\
&\sum_{\boldsymbol{u}^k} p_0(\phi_{kz}(Y_k) \geq f(\boldsymbol{u}^k, \boldsymbol{s}^{kz})|\boldsymbol{u}^k, \boldsymbol{s}^{kz}) p_0(\boldsymbol{u}^k|\boldsymbol{s}^k) p(h_0|\boldsymbol{s}^{kz}) + \\
&\quad p_1(\phi_{kz}(Y_k) < f(\boldsymbol{u}^k, \boldsymbol{s}^{kz})|\boldsymbol{u}^k, \boldsymbol{s}^{kz}) p_1(\boldsymbol{u}^k|\boldsymbol{s}^k) p(h_1|\boldsymbol{s}^{kz}) \\
&= \sum_{\eta} \sum_{\boldsymbol{u}^k \in \mathcal{F}^{\psi}_{\boldsymbol{s}^{kz},\eta}} p_0(\phi_{kz}(Y_k) \geq \eta)|\boldsymbol{u}^k, \boldsymbol{s}^{kz}) p_0(\boldsymbol{u}^k|\boldsymbol{s}^k) p(h_0|\boldsymbol{s}^{kz}) \\
&\quad + p_1(\phi_{kz}(Y_k) < \eta|\boldsymbol{u}^k, \boldsymbol{s}^{kz}) p_1(\boldsymbol{u}^k|\boldsymbol{s}^k) p(h_1|\boldsymbol{s}^{kz}) \\
&\overset{(a)}{=} \sum_{\eta} \sum_{\boldsymbol{u}^k \in \mathcal{F}^{\psi}_{\boldsymbol{s}^{kz},\eta}} p_0(\phi_{kz}(Y_k) \geq \eta)|z) p_0(\boldsymbol{u}^k|\boldsymbol{s}^k) p(h_0|\boldsymbol{s}^{kz}) + \\
&\quad p_1(\phi_{kz}(Y_k) < \eta|z) p_1(\boldsymbol{u}^k|\boldsymbol{s}^k) p(h_1|\boldsymbol{s}^{kz}) \\
&= \sum_{\eta} p_0(\phi_{kz}(Y_k) \geq \eta)|z) G^{\psi}_{\boldsymbol{s}^{kz}}(\eta) + \\
&\quad p_1(\phi_{kz}(Y_k) < \eta|z) Q^{\psi}_{\boldsymbol{s}^{kz}}(\eta)
\end{aligned} \tag{20}$$

with

$$G^{\psi}_{\boldsymbol{s}^{kz}}(\eta) = \sum_{\boldsymbol{u}^k \in \mathcal{F}^{\psi}_{\boldsymbol{s}^{kz},\eta}} p_0(\boldsymbol{u}^k|\boldsymbol{s}^k) p(H = 0|\boldsymbol{s}^{kz}) \tag{21}$$

$$Q^{\psi}_{\boldsymbol{s}^{kz}}(\eta) = \sum_{\boldsymbol{u}^k \in \mathcal{F}^{\psi}_{\boldsymbol{s}^{kz},\eta}} p_1(\boldsymbol{u}^k|\boldsymbol{s}^k) p(H = 1|\boldsymbol{s}^{kz}), \tag{22}$$

where (a) follows from (1), Y_k is independent of $\boldsymbol{Y}^k$ and $\boldsymbol{S}^k$ when conditioned on S_k and H. Thus, $\phi_{kz}(Y_k)$ is independent of $\boldsymbol{U}^k$ and $\boldsymbol{S}^k$ when conditioned on S_k and H. We now turn our attention the quantities $p_0(\phi_{kz}(Y_k) \geq \eta)|z)$ and $p_1(\phi_{kz}(Y_k) < \eta|z)$. Suppose we fix $p_0(\phi_{kz}(Y_k) \geq \eta)|z) = \epsilon$ for $\epsilon \in [0, 1]$. Then we seek the (not necessarily unique) prescription ϕ that solves

$$\begin{aligned}
\min_{\phi \in \Phi} \quad & p_1(\phi(Y_k) < \eta|z) \\
\text{s.t.} \quad & p_0(\phi(Y_k) \geq \eta)z) = \epsilon,
\end{aligned} \tag{23}$$

provided such a prescription exists.

Lemma 3. *For any* $s \in \mathcal{S}$*, let*

$$\mathcal{U}^{\eta}_{s,k} = \{y : L_s(y) \geq \tau^{\eta}_{s,k}\} \tag{24}$$

where $\tau^{\eta}_{s,k}$ *is chosen so that* $p_0(\mathcal{U}^{\eta}_{s,k}|s) = \epsilon$*. Then, in order for a prescription* ϕ *to be a solution to* (23)*, it is sufficient for* ϕ *to satisfy the following:*

1) $\phi(y) \geq \eta$ *if* $y \in \mathcal{U}^{\eta}_{s,k}$.
2) $\phi(y) < \eta$ *if* $y \in (\mathcal{U}^{\eta}_{s,k})^c$.

Proof. For a given $s \in \mathcal{S}$, let

$$\mathcal{Y}^i_{s,k} = \{y : \gamma_k(y, s) = \phi_{ks}(y) = i\}, \tag{25}$$

i.e., $\mathcal{Y}^i_{s,k}$ is the set of all y such that agent k sends message i when in state s. Therefore, the event $\phi_{ks}(Y_k) \geq \eta$ (given s) can be written as

$$\mathcal{D}^{\eta}_{s,k} = \bigcup_{i \geq \eta}^{d-1} \mathcal{Y}^i_{s,k} \tag{26}$$

Hence, we can rewrite (23) as

$$\begin{aligned} \min_{\mathcal{D}^{\eta}_{s,k}} \quad & p_1((\mathcal{D}^{\eta}_{s,k})^c|s) \\ \text{s.t.} \quad & p_0(\mathcal{D}^{\eta}_{s,k}|s) = \epsilon. \end{aligned} \tag{27}$$

Notice that we can think of $\mathcal{D}^{\eta}_{s,k}$ as the *acceptance region*, and thus $p_1((\mathcal{D}^{\eta}_{s,k})^c|s)$ and $p_0(\mathcal{D}^{\eta}_{s,k}|s)$ are the *missed detection* and *false alarm* probabilities, respectively. Then, by the Neyman-Pearson lemma, a solution to (27) is

$$\mathcal{U}^{\eta}_{s,k} = \{y : L_s(y) \geq \tau^{\eta}_{s,k}\}$$

where $\tau^{\eta}_{s,k}$ is chosen so that $p_0(\mathcal{U}^{\eta}_{s,k}|s) = \epsilon$. □

We observe that for fixed η, Lemma 3 *makes no assumption on how ϕ_{ks} assigns the codewords aside from conditions 1 and 2*. This result is helpful, as for a fixed fusion rule, the lemma provides a methodical approach by which to construct a monotone rule for agent k without degrading performance. Notice that in general, much like Neyman-Pearson detection, in order to satisfy $p_0(\mathcal{U}^{\eta}_{s,k}|s) = \epsilon$ one may need to design a randomized rule, our results still hold for this case.

Lemma 4. *For $s \in \mathcal{S}$ and $k = 0, 1, ..., n$, let $\mathcal{U}^{\eta}_{s,k}$ and $\mathcal{U}^{\eta'}_{s,k}$ be defined as in* (24) *with thresholds $\tau^{\eta}_{s,k}$ and $\tau^{\eta'}_{s,k}$ respectively. Moreover, let $\tau^{\eta}_{s,k}$ and $\tau^{\eta'}_{s,k}$ be such that $p_0(\mathcal{U}^{\eta}_{s,k}|s) = \epsilon$ and $p_0(\mathcal{U}^{\eta'}_{s,k}|s) = \epsilon'$ with $\epsilon' > \epsilon$. Then, $\tau^{\eta'}_{s,k} < \tau^{\eta}_{s,k}$.*

Proof. Assume $\tau^{\eta'}_{s,k} \geq \tau^{\eta}_{s,k}$. Then $\mathcal{U}^{\eta'}_{s,k} \subseteq \mathcal{U}^{\eta}_{s,k}$, and thus $p_0(\mathcal{U}^{\eta'}_{s,k}|s) \leq p_0(\mathcal{U}^{\eta}_{s,k}|s)$, which is a contradiction. Thus, we must have $\tau^{\eta'}_{s,k} < \tau^{\eta}_{s,k}$. □

Let Γ_{lr} be the set of all monotone rules. Then $\gamma \in \Gamma_{lr}$ if it uses a monotone rule for every state. Let Γ^n_{lr} be the set of all monotone strategies, i.e., strategies that have every agent use a monotone rule.

Theorem 1.

$$\inf_{\boldsymbol{\gamma} \in \Gamma^n} \{\max_{\boldsymbol{s} \in \mathcal{S}^n} J(\boldsymbol{\psi}|\boldsymbol{s})\} = \inf_{\boldsymbol{\gamma} \in \Gamma^n_{lr}} \{\max_{\boldsymbol{s} \in \mathcal{S}^n} J(\boldsymbol{\psi}|\boldsymbol{s})\}. \tag{28}$$

That is, it is sufficient to search only over strategies where each agent uses a monotone LR rule.

Notice that both infimums are taken over only the agents decision rules $\boldsymbol{\gamma}$. This is because for a fixed $\boldsymbol{\gamma}$ we can find the optimal fusion rule ψ thanks to Lemma 2.

Proof. For $k = 0, 1, ..., n$ and any $z \in \mathcal{S}$, we start with (20),

$$J(\boldsymbol{\psi}|\boldsymbol{s}^{kz})\} = \sum_{\eta} p_0(\phi^k_z(Y_k) \geq \eta)|z) G^{\psi}_{\boldsymbol{s}^{kz}}(\eta) + p_1(\phi^k_z(Y_k) < \eta|z) Q^{\psi}_{\boldsymbol{s}^{kz}}(\eta).$$

Observe that η can take on at most d distinct values. Hence, $p_0(\phi^k_z(Y_k) \geq \eta)|z)$ can take on at most d values across all $\boldsymbol{s}^{kz} \in \mathcal{S}^{n-1}$. Let $d' \in \{1, 2, .., d\}$ be the number of distinct values of $p_0(\phi^k_z(Y_k) \geq \eta)|z)$, and arranged in descending order. That is, $p_0(\phi^k_z(Y_k) \geq i)|z) = \epsilon^i_{z,k} > \epsilon^j_{z,k} = p_0(\phi^k_z(Y_k) \geq j)|z)$ for $i, j \in \{0, 1, ..., d' - 1\}$, $i < j$. Moreover, note that $p_0(\phi^k_z(Y_k) \geq 0)|z) = 1$, and is trivially satisfied by defining $\mathcal{U}^0_{z,k}$ as in (24) with the assignment $\tau^0_{z,k} = 0$. Thus, we concern ourselves only with those distinct $p_0(\phi^k_z(Y_k) \geq i|z)$ for $i \in \{1, 2, ..., d' - 1\}$. For each $p_0(\phi^k_z(Y_k) \geq i)|z)$, construct $\mathcal{U}^i_{z,k}$ according to (24) with $\tau^i_{z,k}$ such that $p_0(\phi^k_z(Y_k) \geq i)|z) = \epsilon^i_{z,k}$, for $i \in \{1, 2, ..., d' - 1\}$. Then, by Lemma 3, any rule that satisfies conditions 1 and 2 in Lemma 3 minimizes $p_1(\phi^k_z(Y_k) < i|z)$. Let the minimum value of $p_1(\phi^k_z(Y_k) < i|z)$ attained by $\mathcal{U}^i_{z,k}$ be denoted as $\alpha^i_{z,k}$. Let $\mathcal{Y}^i_{z,k} = \{y : \phi'_z(y) = i\}$ where ϕ'_z is the prescription we will construct. Given the set of $\mathcal{U}^i_{z,k}$ that solve (27) for $i \in \{0, 1, ..., d' - 1\}$, we give the assignment

$$\mathcal{Y}^i_{z,k} = \mathcal{U}^i_{z,k} \cap \bigcap_{j>i} (\mathcal{U}^j_{z,k})^c = \{y : \tau^i_{z,k} \leq L_z(y) < \tau^{i+1}_{z,k}\}, \tag{29}$$

with the understanding that $\tau^{d'}_{z,k} = \infty$. Moreover, since the $p_0(\phi^k_z(Y_k) \geq i)|z)$ are such that $p_0(\phi^k_z(Y_k) \geq i + 1)|z) < p_0(U_k \geq i)|z)$, then by Lemma 4 we have $\tau^{i+1}_{z,k} > \tau^i_{z,k}$. With the decision region for codeword i given in (29) for $i \in \{0, 1, ..., d' - 1\}$, we have completely characterized the prescription ϕ'_z. Note that with the definition of $\mathcal{Y}^i_{z,k}$ given in (29) it is clear that $\phi'_z \in \Phi_{lr}$, and thus by Lemma 1 satisfies (16). Since z is any state in $\mathcal{S}$, we can construct ϕ'_z for all $z \in \mathcal{S}$. Hence, we let γ' be the rule uses rule ϕ'_z when in state z. Since $\phi'_z \in \Phi_{lr}$ for all $z \in \mathcal{S}$, we get $\gamma' \in \Gamma_{lr}$. We have for any $\boldsymbol{s} \in \mathcal{S}^n$,

$$J(\boldsymbol{\psi}|\boldsymbol{s}) = \sum_{\eta} p_0(U_k \geq \eta)|s_k) G^{\psi}_{\boldsymbol{s}}(\eta) + p_1(U_k < \eta|s_k) Q^{\psi}_{\boldsymbol{s}}(\eta) \tag{30}$$

$$= \sum_{\eta} \epsilon^{\eta}_{s,k} G^{\psi}_{\boldsymbol{s}}(\eta) + p_1(U_k < \eta|s_k) Q^{\psi}_{\boldsymbol{s}}(\eta) \tag{31}$$

$$\geq \sum_{\eta} \epsilon^{\eta}_{s,k} G^{\psi}_{\boldsymbol{s}}(\eta) + \alpha^{\eta}_{s,k} Q^{\psi}_{\boldsymbol{s}}(\eta) \tag{32}$$

$$\overset{(a)}{=} J(\boldsymbol{\psi'}|\boldsymbol{s}), \tag{33}$$

where $\boldsymbol{\psi'} = [\psi, \gamma_1, ..., \gamma_{k-1}, \gamma', \gamma_{k+1}, ..., \gamma_n]$. To see (a), note that γ' satisfies conditions 1 and 2 in Lemma 3, and thus solves (23). Moreover, note that $\boldsymbol{\psi'}$ uses the fusion rule of strategy $\boldsymbol{\psi}$. That is, the quantities $G^{\psi}_{\boldsymbol{s}}(\eta)$ and $Q^{\psi}_{\boldsymbol{s}}(\eta)$ are unchanged under strategy $\boldsymbol{\psi'}$. Hence, if we let $\boldsymbol{\psi^*} = [\psi^*, \gamma_1, ..., \gamma_{k-1}, \gamma', \gamma_{k+1}, ..., \gamma_n]$, where ψ^* is the optimal fusion rule for $[\gamma_1, ..., \gamma_{k-1}, \gamma', \gamma_{k+1}, ..., \gamma_n]$, we have

$$J(\boldsymbol{\psi}|\boldsymbol{s}) \geq J(\boldsymbol{\psi'}|\boldsymbol{s}) \geq J(\boldsymbol{\psi^*}|\boldsymbol{s}) \tag{34}$$

for any $\boldsymbol{s} \in \mathcal{S}$. Therefore, for any $\boldsymbol{\gamma} \in \Gamma^n$, we can either improve or keep the same performance by changing agent k's rule to a monotone LR partition. Thus,

$$\inf_{\boldsymbol{\gamma} \in \Gamma^n} \{\max_{\boldsymbol{s} \in \mathcal{S}^n} J(\boldsymbol{\psi}|\boldsymbol{s})\} \geq \inf_{\boldsymbol{\gamma} \in \Gamma^n_{lr}} \{\max_{\boldsymbol{s} \in \mathcal{S}^n} J(\boldsymbol{\psi}|\boldsymbol{s})\}. \tag{35}$$

Combining the above with the obvious inequality

$$\inf_{\boldsymbol{\gamma} \in \Gamma^n} \{\max_{\boldsymbol{s} \in \mathcal{S}^n} J(\boldsymbol{\psi}|\boldsymbol{s})\} \leq \inf_{\boldsymbol{\gamma} \in \Gamma^n_{lr}} \{\max_{\boldsymbol{s} \in \mathcal{S}^n} J(\boldsymbol{\psi}|\boldsymbol{s})\}. \tag{36}$$

completes the proof. □

We direct the reader to focus on the construction of the rule γ'. Notice that the prescription ϕ'_z is constructed to minimize $p_1(U_k < i|z)$ for all $i \in \{0, 1, ..., d' - 1\}$ for a fixed $p_0(U_k \geq i|z)$. For a fixed fusion rule ψ and agent rules $\{\gamma_j\}_{\substack{j=1 \\ j\neq k}}^{n}$, the expected cost for any state $\boldsymbol{s}^{kz} \in \mathcal{S}^{n-1}$ depends on the rule for agent k only through the prescription ϕ_z. Moreover, the expected state cost only depends on ϕ_z through the probabilities $p_1(U_k < i|z)$ and $p_0(U_k \geq i|z)$, provided the fusion center's rule is fixed and the codebook for agent k while in state z satisfies (16). Hence, the prescription ϕ'_z improves the expected state cost *across all states with agent k taking state z, i.e., $\boldsymbol{s}^{kz} \in \mathcal{S}^{n-1}$*. Since we construct a prescription for every state $z \in \mathcal{S}$ for agent k, and define the rule γ' to be the collection of these prescriptions indexed by the state, i.e., γ' uses prescription ϕ'_z when in state z, we can see that the rule γ' improves the expected cost *across all states $\boldsymbol{s} \in \mathcal{S}^n$*. This observation yields the following corollary,

Corollary 1.1. *Define the optimization problem*

$$\inf_{\boldsymbol{\psi} \in \Psi \times \Gamma^n} J(\boldsymbol{\psi}) = \inf_{\boldsymbol{\psi} \in \Psi \times \Gamma^n} \sum_{\boldsymbol{s} \in \mathcal{S}^n} c_{\boldsymbol{s}} J(\boldsymbol{\psi}|\boldsymbol{s}) \tag{37}$$

where $c_{\boldsymbol{s}} \geq 0$ for all $\boldsymbol{s} \in \mathcal{S}^n$ are given constants. Then, we have the following:

a) The optimal fusion rule for a given collection of agent rules $\boldsymbol{\gamma}$ is given by (10)*. Hence, we only need to concern ourselves with the optimization over agent rules.*

b) $\inf_{\boldsymbol{\gamma} \in \Gamma^n} J(\boldsymbol{\psi}) = \inf_{\boldsymbol{\gamma} \in \Gamma^n_{lr}} J(\boldsymbol{\psi})$. That is, it is sufficient to search only over the set of monotone LR partition rules.

Proof. a) is a consequence of Lemma 2 and the fact that $c_{\boldsymbol{s}} \geq 0$ for all $\boldsymbol{s} \in \mathcal{S}^n$, since the rule given by (10) minimizes $J(\boldsymbol{\psi}|\boldsymbol{s})$ for any $\boldsymbol{s} \in \mathcal{S}$. Therefore, we can drop the optimization over the fusion rule as we did before. To prove b), note that

$$J(\boldsymbol{\psi}) = \sum_{\boldsymbol{s} \in \mathcal{S}^n} c_{\boldsymbol{s}} J(\boldsymbol{\psi}|\boldsymbol{s}) = \sum_{z \in \mathcal{S}} \sum_{\boldsymbol{s}^{kz} \in \mathcal{S}^{n-1}} c_{\boldsymbol{s}} J(\boldsymbol{\psi}|\boldsymbol{s}^{kz}) \tag{38}$$

and that for each $z \in \mathcal{S}$, we can construct a prescription ϕ'_z for agent k in the same way as in Theorem 1. Hence, the prescription ϕ'_z improves the expected state cost for all $\boldsymbol{s}^{kz} \in \mathcal{S}^{n-1}$. Letting γ' be the rule that uses prescription ϕ'_z when in state z for agent k, the rest of the proof is similiar to that of Theorem 1. □

An obvious consequence of Corollary 1.1 is that for the Bayesian case no loss of optimality occurs from restricting ourselves to monotone strategies. One must also take great care in interpreting the proof of Theorem 1. We previously stated that the rule constructed for agent k improves the performance across all network states. However, *this does not imply that said rule is optimal for all network states*. The performance improvement may only be marginal in some network states, and the overall cost may still be high given those states, resulting in sub-optimal performance for the whole system. Hence, one cannot simply "fix" the network state and design a strategy only for that state. An explicit example to show this feature is constructed in Section V.

B. Bayesian Formulation

In this section, we now consider the no-information case, and derive the optimal sensor rule under the Bayesian setting. That is, the fusion center only receives $\boldsymbol{U}$. We begin by writing the cost as

$$J(\boldsymbol{\psi}) = \mathbb{E}[J(\boldsymbol{\psi}|\boldsymbol{S})] \tag{39}$$

where the expectation is taken over $\boldsymbol{S}$. We can write $J(\boldsymbol{\psi})$ as

$$\begin{aligned} J(\boldsymbol{\psi}) = \sum_{s_k} \int_{y_k} \sum_{i=0}^{d-1} \sum_{h=0}^{1} \sum_{j=0}^{1} \sum_{\boldsymbol{u}^k, \boldsymbol{s}^k} \\ c_{j,h} p(s_k, y_k, U_k = i, H = h, U_0 = j, \boldsymbol{u}^k, \boldsymbol{s}^k) dy_k. \end{aligned} \tag{40}$$

We turn our attention to

$$p(s_k, y_k, U_k = i, H = h, U_0 = j, \boldsymbol{u}^k, \boldsymbol{s}^k) \tag{41}$$

and rewrite it as

$$\begin{aligned} p_h(U_0 = j|\boldsymbol{u}^k, U_k = i, \boldsymbol{s}^k, s_k, y_k) p_h(U_k = i, \boldsymbol{u}^k, y_k|s_k, \boldsymbol{s}^k) \\ p(H = h|s_k, \boldsymbol{s}^k) p(\boldsymbol{s}^k|s_k) p(s_k). \end{aligned} \tag{42}$$

From $\boldsymbol{S} \to (\boldsymbol{S}, H) \to (\boldsymbol{Y}, \boldsymbol{S}) \to \boldsymbol{U} \to U_0$ we get

$$\begin{aligned} p_h(U_0 = j|\boldsymbol{u}^k, U_k = i, \boldsymbol{s}^k, s_k, y_k) = \\ p(U_0 = j|\boldsymbol{u}^k, U_k = i). \end{aligned} \tag{43}$$

Moreover,

$$\begin{aligned} p_h(\boldsymbol{u}^k, u_k, y_k|\boldsymbol{s}^k, s_k) &= \sum_{\boldsymbol{y}^k} p_h(\boldsymbol{u}^k, u_k, y_k, \boldsymbol{y}^k|\boldsymbol{s}^k, s_k) \\ &= \sum_{\boldsymbol{y}^k} p_h(\boldsymbol{u}^k|\boldsymbol{y}^k, y_k, u_k, \boldsymbol{s}^k, s_k) p_h(u_k|y_k, \boldsymbol{y}^k, \boldsymbol{s}^k, s_k,) \\ &\quad p_h(y_k|\boldsymbol{s}^k, s_k, \boldsymbol{y}^k) p_h(\boldsymbol{y}^k|s_k, \boldsymbol{s}^k) \\ &\overset{(a)}{=} \sum_{\boldsymbol{y}^k} p_h(\boldsymbol{u}^k|\boldsymbol{y}^k, \boldsymbol{s}^k) p_h(u_k|y_k, s_k) \\ &\quad p_h(y_k|\boldsymbol{s}^k, s_k, \boldsymbol{y}^k) p_h(\boldsymbol{y}^k|s_k, \boldsymbol{s}^k) \\ &\overset{(b)}{=} \sum_{\boldsymbol{y}^k} p_h(\boldsymbol{u}^k|\boldsymbol{y}^k, \boldsymbol{s}^k) p(u_k|y_k, s_k) p_h(y_k|s_k) p_h(\boldsymbol{y}^k|\boldsymbol{s}^k) \\ &= p(u_k|y_k, s_k) p_h(y_k|s_k) \sum_{\boldsymbol{y}^k} p_h(\boldsymbol{u}^k|\boldsymbol{y}^k, \boldsymbol{s}^k) p_h(\boldsymbol{y}^k|\boldsymbol{s}^k) \\ &= p(u_k|y_k, s_k) p_h(y_k|s_k) p_h(\boldsymbol{u}^k|\boldsymbol{s}^k) \end{aligned} \tag{44}$$

where (a) comes from (2) and $\boldsymbol{S} \to (\boldsymbol{S}, H) \to (\boldsymbol{Y}, \boldsymbol{S}) \to \boldsymbol{U} \to U_0$ and (2), and (b) comes from (1). Therefore, (41) can be written as

$$\begin{aligned} p(U_0 = j|\boldsymbol{u}^k, U_k = i) p(u_k|y_k, s_k) p_h(y_k|s_k) \\ p_h(\boldsymbol{u}^k|\boldsymbol{s}^k) p(H = h|s_k, \boldsymbol{s}^k) p(\boldsymbol{s}^k|s_k) p(s_k). \end{aligned} \tag{45}$$

Substituting this into (40) and rearranging terms gives us

$$\begin{aligned} &J(\boldsymbol{\psi}) = \\ &\sum_{s_k} p(s_k) \int_{y_k} \left[\sum_{i=0}^{d-1} p(U_k = i|y_k, s_k) D_{ki}(y_k, s_k) \right] dy_k. \end{aligned} \tag{46}$$

where

$$D_{ki}(y_k, s_k) = \sum_{h=0}^{1} p_h(y_k|s_k) \sum_{j=0}^{1} c_{j,h} \sum_{\boldsymbol{u}^k, \boldsymbol{s}^k} \Big\{ p(U_0 = j|\boldsymbol{u}^k, U_k = i) \; p_h(\boldsymbol{u}^k|\boldsymbol{s}^k) p(H = h|s_k, \boldsymbol{s}^k) p(\boldsymbol{s}^k|s_k) \Big\}. \tag{47}$$

Thus, we see that the optimal rule is

$$i^* = \gamma_k(y_k, s_k) = \arg \min_{i \in \{0,1,\ldots,d-1\}} D_{ki}(y_k, s_k) \tag{48}$$

Let

$$g_k(i, s_k, h) = \sum_{j=0}^{1} c_{j,h} \sum_{\boldsymbol{u}^k, \boldsymbol{s}^k} \Big\{ p(U_0 = j|\boldsymbol{u}^k, U_k = i) \; p_h(\boldsymbol{u}^k|\boldsymbol{s}^k) p(H = h|s_k, \boldsymbol{s}^k) p(\boldsymbol{s}^k|s_k) \Big\} \tag{49}$$

and so

$$D_{ki}(y_k, s_k) = p_0(y_k|s_k) g_k(i, s_k, 0) + p_1(y_k|s_k) g_k(i, s_k, 1). \tag{50}$$

Note that this rule is for any arbitrary cost. Then, we have the following,

Lemma 5. *The optimal rule in* (48) *for the kth sensor amounts is a monotone rule, provided i^* in* (48) *is unique.*

Proof. For a fixed s, the proof is the same as Lemma 1 in [22]. □

Hence, the rule given by (48) corresponds to partitioning the likelihood space into b (the cardinality of the message space $\mathcal{U}$) of intervals. The sent message u_k then corresponds to the interval containing the observed likelihood ratio. We note that even in the event where the minimizing index i is not unique, we can still restrict ourselves to monotone rules by assigning the decision regions for the codewords in such a way as to satisfy Definition 1. If we restrict ourselves to the probability of error, (50) becomes

$$D_{ki}(y_k, s_k) = p_0(y_k|s_k) A_{ki}^{s_k} + p_1(y_k|s_k) B_{ki}^{s_k} \tag{51}$$

with

$$A_{ki}^{s_k} = \sum_{\boldsymbol{u}^k, \boldsymbol{s}^k} p(U_0 = 1|\boldsymbol{u}^k, \boldsymbol{s}^k, U_k = i, s_k) \; p_0(\boldsymbol{u}^k|\boldsymbol{s}^k) p(H = 0|\boldsymbol{s}^k, s_k) p(\boldsymbol{s}^k|s_k), \tag{52}$$

and

$$B_{ki}^{s_k} = \sum_{\boldsymbol{u}^k, \boldsymbol{s}^k} p(U_0 = 0|\boldsymbol{u}^k, \boldsymbol{s}^k, U_k = i, s_k) \; p_1(\boldsymbol{u}^k|\boldsymbol{s}^k) p(H = 1|\boldsymbol{s}^k, s_k) p(\boldsymbol{s}^k|s_k). \tag{53}$$

These expressions are more amenable to the devolpment of iterative algorithms, such as the algorithm given in Appendix I.

IV. Asymptotic Results

The iterative algorithm developed in Appendix I is useful for finding the optimal rules for a finite number of agents. However, as the number of agents grows, so too does the complexity of the algorithm. Moreover, since we are primarily inspired by microbial applications, it is not uncommon for us to encounter systems with millions or even tens of millions of agents. While it may be tempting to reduce complexity by restricting oneself to search only over strategies where identical agents use the same rules, we provide an example to show that, in general, this approach is sub-optimal for finite n. First, we give our definition of weakly identical agents.

Definition 3. *Given a collection of n agents, these agents are weakly identical if $p_h(Y_k = y|S_k = s) = p_h(Y_j = y|S_j = s)$ for all $k, j \in \{1, 2, ..., n\}$, $h \in \{h_0, h_1\}$, $y \in \mathcal{Y}$, $s \in \mathcal{S}$.*

Now, consider the following example with binary messages ($b = 2$), two agents ($n = 2$), two states (0, 1), and state-information. We assume that the agent states are i.i.d with $p(S = 1) = .25$. The hypothesis model is given as

$$p(h_0|s_1, s_2) = p(h_0|s_1) \begin{cases} .52 & s_1 = 0 \\ .48 & s_1 = 1. \end{cases}$$

The observations y_1 and y_2 are independent, conditioned on the hypothesis and state, take values in $\{0, 1, 2\}$, and have the following common distribution: $(p_0(y = 0|s = 0), p_0(y = 1|s = 0), p_0(y = 2|s = 0)) = (.8, .2, 0)$, $p_1(y = i|s = 0) = \frac{1}{3}$, and

$$p_h(y = i|s = 1) = \begin{cases} 1 & i = h \\ 0 & i \neq h \end{cases}$$

for $i = 0, 1, 2$ and $h = 0, 1$. Notice that even though the hypothesis model does not depend on agent 2 when conditioned on agent 1, the two agents are still identical by our definition. In a given state, each agent computes a likelihood ratio test. If we enumerate through all of the cases for this discrete observations example, we see that, in a given state each agent can choose from one of two following rules:

A) $u_i = 1$ if and only if $y_i = 0$.
B) $u_i = 1$ if and only if $y_i \in \{0, 1\}$.

Hence, there are four candidate rules for each agent. Recall that the fusion center always employs the MAP rule conditioned on the strategies used by the agents. An optimal strategy is found by exhaustive enumeration. Computing the probability of error for every strategy, we find that the optimal strategy is for agent one to use rule B when in state 0 and rule A when in state 1, and for agent two to use rule A regardless of its state. This strategy results in a probability of error of .1185. Clearly, the optimal rule is not the same for each agent. Moreover, in this example, the agents' states are i.i.d. Indeed, this case is of importance, and so we give the definition of *strongly identical agents*.

Definition 4. *A collection of n weakly identical agents are further strongly identical if the agents' states are mutually independent and identically distributed a priori.*

A. State-Information Error Exponent

We first devote our attention to the state-information case, since this case is not handled by any prior results in the literature. Let $\mathcal{A}_0$ and $\mathcal{A}_1$ be the sets where the fusion center decides 0 and 1, respectively. Recall that these sets are defined by (9), and thus depend on $\boldsymbol{\gamma}$. We drop the dependence on $\boldsymbol{\gamma}$ from the notation for simplicity. The probability of error can then be written as

$$J_n(\boldsymbol{\psi}) = \sum_{\boldsymbol{s},\boldsymbol{u}\in\mathcal{A}_1} p_0(\boldsymbol{u}|\boldsymbol{s})p(h_0|\boldsymbol{s})p(\boldsymbol{s}) + \sum_{\boldsymbol{s},\boldsymbol{u}\in\mathcal{A}_0} p_1(\boldsymbol{u}|\boldsymbol{s})p(h_1|\boldsymbol{s})p(\boldsymbol{s}). \tag{54}$$

We define the following key sequence:

$$\alpha_n \quad = \quad \min_{\boldsymbol{s},h\in\{h_0,h_1\}} p(h|\boldsymbol{s}), \tag{55}$$

that is, α_n is the smallest $p(h|\boldsymbol{s})$ appearing in either summation. Moreover, for the rest of the section, we make the following assumption:

Assumption 2. *for all n,*

a) Both hypotheses are possible under all network states, i.e., we do not condition on events with measure zero.

b) p_0 (on $\mathcal{Y}^n \times \mathcal{S}^n$) is absolutely continuous, with respect to p_1, i.e., $p_0 \ll p_1$.[2] *and $\mathbb{E}_1[\log^2 L_{\boldsymbol{s}}(\boldsymbol{Y})] < \infty$ for all $\boldsymbol{s}$, where the expectation is taken with respect to $p_1(\bullet|\boldsymbol{s})$.*

c) $\lim_{n\to\infty}(\log \alpha_n)/n = 0$.

Notice that since the fusion center is implementing the MAP rule, we can rewrite (54) as

$$J_n(\boldsymbol{\psi}) = \sum_{\boldsymbol{s},\boldsymbol{u}} \min\{p_0(\boldsymbol{u}|\boldsymbol{s})p(h_0|\boldsymbol{s}), p_1(\boldsymbol{u}|\boldsymbol{s})p(h_1|\boldsymbol{s})\}p(\boldsymbol{s}) \tag{56}$$

Since we are concerned with large n, we focus on the error exponent defined as

$$\lim_{n\to\infty} \frac{1}{n} \log J_n(\boldsymbol{\psi}). \tag{57}$$

Then, if we let $r_n(\boldsymbol{\gamma}) = \frac{1}{n}\log J_n(\boldsymbol{\psi})$ and $R_n = \inf_{\boldsymbol{\gamma}\in\Gamma^n} r_n(\boldsymbol{\gamma})$ for all n, we analyze the limiting behavior of R_n. Notice that we define r_n as a function of $\boldsymbol{\gamma}$. This is because given $\boldsymbol{\gamma}$, the optimal fusion rule is known. Similar to [39], we would like to derive upper and lower bounds on $r_n(\boldsymbol{\gamma})$. Unfortunately, we cannot use the bounds derived in [39] and [16] since our observations are not conditionally independent. Thus, we need the following.

Lemma 6. *For any n and $\boldsymbol{\gamma} \in \Gamma^n$, we have*

$$\frac{\log \alpha_n}{n} - \frac{\log 2}{n} + \frac{1}{n}\mu(\boldsymbol{\gamma},\epsilon^*) - \frac{\sqrt{2\mu''(\boldsymbol{\gamma},\epsilon^*)}}{n} \le r_n(\boldsymbol{\gamma}) \le \frac{1}{n}\mu(\boldsymbol{\gamma},\epsilon^*), \tag{58}$$

where for $\epsilon \in (0,1)$ we define [3]

$$\mu(\boldsymbol{\gamma},\epsilon) = \log\left[\sum_{\boldsymbol{s}}\sum_{\boldsymbol{u}} (p_0(\boldsymbol{u}|\boldsymbol{s}))^{1-\epsilon}(p_1(\boldsymbol{u}|\boldsymbol{s}))^{\epsilon}p(\boldsymbol{s})\right], \tag{59}$$

α_n is defined in (55)*, $\mu''(\boldsymbol{\gamma},\epsilon)$ is the second derivative of $\mu(\boldsymbol{\gamma},\epsilon)$ with respect to ϵ, and $\epsilon^* = \arg\min_{\epsilon\in[0,1]} \mu(\boldsymbol{\gamma},\epsilon)$.*

Proof. Assuming the fusion center is implementing the MAP rule, we have

$$\begin{aligned} &J_n(\boldsymbol{\gamma}) = \\ &\sum_{\boldsymbol{s},\boldsymbol{u}} \min\{p_0(\boldsymbol{u}|\boldsymbol{s})p(h=0|\boldsymbol{s}), p_1(\boldsymbol{u}|\boldsymbol{s})p(h=1|\boldsymbol{s})\}p(\boldsymbol{s}) \\ &\overset{(a)}{\le} \sum_{\boldsymbol{s},\boldsymbol{u}} (p_0(\boldsymbol{u}|\boldsymbol{s})p(h=0|\boldsymbol{s}))^{1-\epsilon}(p_1(\boldsymbol{u}|\boldsymbol{s})p(h=1|\boldsymbol{s}))^{\epsilon}p(\boldsymbol{s}) \\ &\le \sum_{\boldsymbol{s},\boldsymbol{u}} p_0(\boldsymbol{u}|\boldsymbol{s})^{1-\epsilon}p_1(\boldsymbol{u}|\boldsymbol{s})^{\epsilon}p(\boldsymbol{s}) \end{aligned} \tag{60}$$

where (a) is due to the fact that for any two positive numbers a and b,

$$\min\{a,b\} \le a^{\epsilon}b^{1-\epsilon} \quad \text{for all } \epsilon \in [0,1]. \tag{61}$$

Hence,

$$\frac{1}{n}\log J_n(\boldsymbol{\gamma}) \le \frac{1}{n}\log\left[\sum_{\boldsymbol{s},\boldsymbol{u}} (p_0(\boldsymbol{u}|\boldsymbol{s}))^{1-\epsilon}(p_1(\boldsymbol{u}|\boldsymbol{s}))^{\epsilon}p(\boldsymbol{s})\right]. \tag{62}$$

Since this is true for all ϵ, simply take the minimum over $0 \le \epsilon \le 1$. To prove the lower bound, notice that

$$\mu'(\boldsymbol{\gamma},\epsilon) = \sum_{\boldsymbol{u},\boldsymbol{s}} Q_{\epsilon}(\boldsymbol{u},\boldsymbol{s}) \log L_{\boldsymbol{s}}(\boldsymbol{u}) \tag{63}$$

$$\mu''(\boldsymbol{\gamma},\epsilon) = \left\{\sum_{\boldsymbol{u},\boldsymbol{s}} Q_{\epsilon}(\boldsymbol{u},\boldsymbol{s}) \log^2 L_{\boldsymbol{s}}(\boldsymbol{u})\right\} - (\mu'(\boldsymbol{\gamma},\epsilon))^2 \tag{64}$$

where all derivatives are with respect to ϵ. For $\epsilon \in (0,1)$, define

$$Q_{\epsilon}(\boldsymbol{u},\boldsymbol{s}) = \frac{p_0(\boldsymbol{u}|\boldsymbol{s})^{1-\epsilon}p_1(\boldsymbol{u}|\boldsymbol{s})^{\epsilon}p(\boldsymbol{s})}{\sum_{\boldsymbol{u}',\boldsymbol{s}'}(p_0(\boldsymbol{u}'|\boldsymbol{s}'))^{1-\epsilon}p_1(\boldsymbol{u}'|\boldsymbol{s}')^{\epsilon}p(\boldsymbol{s}')}. \tag{65}$$

Observe that we can think of $Q_{\epsilon}(\boldsymbol{u},\boldsymbol{s})$ as a probability distribution on $L_{\boldsymbol{s}}(\boldsymbol{u})$. Hence, $\mu'(\boldsymbol{\gamma},\epsilon)$ and $\mu''(\boldsymbol{\gamma},\epsilon)$ are the mean and variance of $\log L_{\boldsymbol{s}}(\boldsymbol{u})$, respectively, according to $Q_{\epsilon}(\boldsymbol{u},\boldsymbol{s})$. Since $\mu''(\boldsymbol{\gamma},\epsilon)$ is a variance, it is non-negative, so for a fixed $\boldsymbol{\gamma}$ $\mu(\boldsymbol{\gamma},\epsilon)$ is convex in epsilon. Furthermore, from the definition of $\mu(\boldsymbol{\gamma},\epsilon)$ we see that it is non-positive, and with convexity is zero for all $\epsilon \in [0,1]$ only in the uninteresting case where the fusion center is unable to distinguish between the two hypotheses, i.e., under all states, $p_0(\boldsymbol{u}|\boldsymbol{s}) = p_1(\boldsymbol{u}|\boldsymbol{s})$ for all $\boldsymbol{u}$. While these facts are not explicitly used in any of our proofs, they are important for two reasons. First, they show that the bounds we derive are both non-trivial and meaningful. Second, we will show that the optimal asymptotic rule can

[2] Let μ and ν be two measures on $(\mathcal{X},\mathcal{M})$. Then we say ν is *absolutely continuous* with respect to μ, written as $\nu \ll \mu$ if $\nu(\mathcal{E}) = 0$ for every $\mathcal{E} \in \mathcal{M}$, for which $\mu(\mathcal{E}) = 0$

[3] We also extend the definition to include the cases $\epsilon = 0$ and $\epsilon = 1$, with $\mu(\boldsymbol{\gamma},0) = \lim_{\epsilon\to 0^+}\mu(\boldsymbol{\gamma},\epsilon)$; $\mu(\boldsymbol{\gamma},1) = \lim_{\epsilon\to 1^-}\mu(\boldsymbol{\gamma},\epsilon)$.

be found by optimizing $\mu(\boldsymbol{\gamma}, \epsilon)$. Hence, convexity in ϵ is a desirable property. Moreover, it is not difficult to show

$$p_0(\boldsymbol{u}|\boldsymbol{s})p(\boldsymbol{s}) = \{\exp[\mu(\boldsymbol{\gamma}, \epsilon) - \epsilon \log L_{\boldsymbol{s}}(\boldsymbol{u})]\}Q_\epsilon(\boldsymbol{u}, \boldsymbol{s}) \quad (66a)$$

$$p_1(\boldsymbol{u}|\boldsymbol{s})p(\boldsymbol{s}) = \{\exp[\mu(\boldsymbol{\gamma}, \epsilon) + (1-\epsilon) \log L_{\boldsymbol{s}}(\boldsymbol{u})]\}Q_\epsilon(\boldsymbol{u}, \boldsymbol{s}). \quad (66b)$$

Define A_ϵ to be the set of pairs $(\boldsymbol{u}, \boldsymbol{s})$ for which $\log L_{\boldsymbol{s}}(\boldsymbol{u})$ is within $\sqrt{2}$ standard deviations of its mean according to $Q_\epsilon(\boldsymbol{u}, \boldsymbol{s})$,

$$A_\epsilon = \{(\boldsymbol{u}, \boldsymbol{s}) : |\log L_{\boldsymbol{s}}(\boldsymbol{u}) - \mu'(\boldsymbol{\gamma}, \epsilon)| \leq \sqrt{2\mu''(\boldsymbol{\gamma}, \epsilon)}. \quad (67)$$

For any $(\boldsymbol{u}, \boldsymbol{s}) \in A_\epsilon$ we have

$$\mu'(\boldsymbol{\gamma}, \epsilon) - \sqrt{2\mu''(\boldsymbol{\gamma}, \epsilon)} \leq \log L_{\boldsymbol{s}}(\boldsymbol{u}) \leq \mu'(\boldsymbol{\gamma}, \epsilon) + \sqrt{2\mu''(\boldsymbol{\gamma}, \epsilon)} \quad (68)$$

Then,

$$\begin{aligned}
&J_n(\boldsymbol{\gamma}) = \\
&\sum_{\boldsymbol{s},\boldsymbol{u}} \min\{p_0(\boldsymbol{u}|\boldsymbol{s})p(h=0|\boldsymbol{s}), p_1(\boldsymbol{u}|\boldsymbol{s})p(h=1|\boldsymbol{s})\}p(\boldsymbol{s}) \\
&\geq \alpha_n \sum_{\boldsymbol{s},\boldsymbol{u}\in A_\epsilon} \min\{p_0(\boldsymbol{u}|\boldsymbol{s})p(\boldsymbol{s}), p_1(\boldsymbol{u}|\boldsymbol{s})\}p(\boldsymbol{s}) \\
&= \alpha_n \exp[\mu(\boldsymbol{\gamma}, \epsilon)] \sum_{\boldsymbol{s},\boldsymbol{u}\in A_\epsilon} \min\Big\{ \\
&\exp[-\epsilon \log L_{\boldsymbol{s}}(\boldsymbol{u})], \exp[(1-\epsilon) \log L_{\boldsymbol{s}}(\boldsymbol{u})]\Big\}Q_\epsilon(\boldsymbol{u}, \boldsymbol{s}) \\
&\geq \alpha_n \exp[\mu(\boldsymbol{\gamma}, \epsilon)] \sum_{\boldsymbol{s},\boldsymbol{u}\in A_\epsilon} \min\Big\{ \\
&\exp[-\epsilon\mu'(\boldsymbol{\gamma}, \epsilon) - \epsilon\sqrt{2\mu''(\boldsymbol{\gamma}, \epsilon)}], \exp[(1-\epsilon)\mu'(\boldsymbol{\gamma}, \epsilon) - \\
&(1-\epsilon)\sqrt{2\mu''(\boldsymbol{\gamma}, \epsilon)}]\Big\}Q_\epsilon(\boldsymbol{u}, \boldsymbol{s}) \\
&\overset{(a)}{\geq} \alpha_n \exp[\mu(\boldsymbol{\gamma}, \epsilon) - \sqrt{2\mu''(\boldsymbol{\gamma}, \epsilon)}] \sum_{\boldsymbol{s},\boldsymbol{u}\in A_\epsilon} \min\Big\{ \\
&\exp[-\epsilon\mu'(\boldsymbol{\gamma}, \epsilon)], \exp[(1-\epsilon)\mu'(\boldsymbol{\gamma}, \epsilon)]\Big\}Q_\epsilon(\boldsymbol{u}, \boldsymbol{s}) \\
&= \alpha_n \exp[\mu(\boldsymbol{\gamma}, \epsilon) - \sqrt{2\mu''(\boldsymbol{\gamma}, \epsilon)}] \min\Big\{ \exp[-\epsilon\mu'(\boldsymbol{\gamma}, \epsilon)], \\
&\exp[(1-\epsilon)\mu'(\boldsymbol{\gamma}, \epsilon)]\Big\} \sum_{\boldsymbol{s},\boldsymbol{u}\in A_\epsilon} Q_\epsilon(\boldsymbol{u}, \boldsymbol{s}) \\
&\overset{(b)}{\geq} \alpha_n \exp[\mu(\boldsymbol{\gamma}, \epsilon) - \sqrt{2\mu''(\boldsymbol{\gamma}, \epsilon)}] \\
&\min\Big\{ \exp[-\epsilon\mu'(\boldsymbol{\gamma}, \epsilon)], \exp[(1-\epsilon)\mu'(\boldsymbol{\gamma}, \epsilon)]\Big\}\frac{1}{2}.
\end{aligned}$$

Where (a) comes from the fact that $0 \leq \epsilon \leq 1$, and (b) is due to the Chebyshev bound. The above is true for all $\epsilon \in [0, 1]$, thus, we can take the ϵ^* that minimizes $\mu(\boldsymbol{\gamma}, \epsilon)$. If $\epsilon^* \in (0, 1)$, then $\mu'(\boldsymbol{\gamma}, \epsilon^*) = 0$. Otherwise, if $\epsilon^* = 0$, then $\mu'(\boldsymbol{\gamma}, 0^+) \geq 0$, so

$$\begin{aligned}
\lim_{\epsilon\to 0^+} -\epsilon\mu'(\boldsymbol{\gamma}, \epsilon) &= 0 \\
\lim_{\epsilon\to 0^+} (1-\epsilon)\mu'(\gamma, \epsilon) &\geq 0.
\end{aligned} \quad (69)$$

Likewise, if $\epsilon = 1$, then $\mu'(\boldsymbol{\gamma}, 1^-) \leq 0$, and

$$\begin{aligned}
\lim_{\epsilon\to 1^-} -\epsilon\mu'(\boldsymbol{\gamma}, \epsilon) &\geq 0 \\
\lim_{\epsilon\to 1^-} (1-\epsilon)\mu'(\boldsymbol{\gamma}, \epsilon) &= 0.
\end{aligned} \quad (70)$$

Thus, $\min\Big\{ \exp[-\epsilon^*\mu'(\boldsymbol{\gamma}, \epsilon^*)], \exp[(1-\epsilon^*)\mu'(\boldsymbol{\gamma}, \epsilon^*)]\Big\} = 1$, so

$$J_n(\boldsymbol{\gamma}) \geq \frac{\alpha_n}{2} \exp[\mu(\boldsymbol{\gamma}, \epsilon^*) - \sqrt{2\mu''(\boldsymbol{\gamma}, \epsilon^*)}]. \quad (71)$$

Taking the log and dividing by n on both sides completes the proof. □

Note that if $\frac{1}{n}\log\alpha_n$ and $\mu''(\boldsymbol{\gamma}, \epsilon)$ are not properly controlled, then the bounds given in (93) could be far apart even for large n. Because of this, we elaborate on Assumption 2. The next lemma controls $\mu''(\boldsymbol{\gamma}, \epsilon)$.

Lemma 7. *Subject to Assumption 2, for all n, $\boldsymbol{\gamma} \in \Gamma^n$, and $\epsilon \in [0, 1]$, There exists a finite constant θ such that $|\mu''(\boldsymbol{\gamma}, \epsilon)| \leq n\theta$.*

Proof. See Appendix. □

Notice that Assumption 2(a) *does not imply* Assumption 2(b). Since, for a given $\boldsymbol{s}$, only one set of signals $\boldsymbol{y}$ may be possible under each hypothesis [4]. Then, for fixed n, suppose there exists some $\boldsymbol{\gamma}$ and $\epsilon \in [0, 1]$ such that $|\mu(\boldsymbol{\gamma}, \epsilon)|$ is not finite. Then, we must have that

$$\sum_{\boldsymbol{s}}\sum_{\boldsymbol{u}} p_0(\boldsymbol{u}|\boldsymbol{s})^{1-\epsilon}p_1(\boldsymbol{u}|\boldsymbol{s})^\epsilon p(\boldsymbol{s}) = 0$$

Since each term in the summation is non-negative, this implies $p_0(\boldsymbol{u}|\boldsymbol{s}) = 0$ whenever $p_1(\boldsymbol{u}|\boldsymbol{s}) \geq 0$, and vice-versa. Then, let Ω be the set of $(\boldsymbol{y}, \boldsymbol{s})$ such that the strategy $\boldsymbol{\gamma}$ maps $(\boldsymbol{y}, \boldsymbol{s})$ to $(\boldsymbol{u}, \boldsymbol{s})$ such that $p_1(\boldsymbol{u}|\boldsymbol{s}) \geq 0$. Then we have $p_0(\boldsymbol{u}|\boldsymbol{s})p_0(\boldsymbol{s}) = p_0(\boldsymbol{u}, \boldsymbol{s}) = p_0(\boldsymbol{\gamma}(\boldsymbol{y}, \boldsymbol{s}) = (\boldsymbol{u}, \boldsymbol{s})) = p_0((\boldsymbol{y}, \boldsymbol{s}) = \boldsymbol{\gamma}^{-1}(\boldsymbol{u}, \boldsymbol{s})) = 0$ on Ω. Similarly, $p_1(\boldsymbol{y}, \boldsymbol{s}) = 0$ and $p_0(\boldsymbol{y}, \boldsymbol{s}) \geq 0$ on Ω^c. Thus, the two measures are mutually singular[5], violating Assumption 2(b). Hence, if Assumption 2(b) is satisfied, then $|\mu(\boldsymbol{\gamma}, \epsilon)|$ must be finite for all $\boldsymbol{\gamma} \in \Gamma^n$, $\epsilon \in [0, 1]$. Then, we have

$$\sum_{\boldsymbol{u}} p_0(\boldsymbol{u}|\boldsymbol{s})^{1-\epsilon}p_1(\boldsymbol{u}|\boldsymbol{s})^\epsilon > 0$$

For at least one $\boldsymbol{s} \in \mathcal{S}^n$ with $p(\boldsymbol{s}) > 0$. Using (1),

$$\begin{aligned}
&\sum_{\boldsymbol{u}} p_0(\boldsymbol{u}|\boldsymbol{s})^{1-\epsilon}p_1(\boldsymbol{u}|\boldsymbol{s})^\epsilon > 0 \iff \\
&\sum_{\boldsymbol{u}}\prod_{k=1}^n p_0(u_k|s_k)^{1-\epsilon}p_1(u_k|s_k)^\epsilon > 0 \iff \\
&\prod_{k=1}^n\sum_{u_k} p_0(u_k|s_k)^{1-\epsilon}p_1(u_k|s_k)^\epsilon > 0 \iff
\end{aligned}$$

[4]Take $\mathcal{Y} = [0, 1]$ with the Borel σ-algebra. Let $p_0(y|s) = \delta_s(y)$, $s \in \{0, 1\}$, *i.e.*, a point measure at s, and let $p_1(\bullet|s)$ be Lebesgue measure.

[5]two measures μ and ν on $(\mathcal{X}, \mathcal{M})$ are *mutually singular* if there exists $\mathcal{E} \in \mathcal{M}$ such that $\nu(\mathcal{E}) = 0$ and $\mu(\mathcal{E}^c) = 0$

$$\sum_{u_k} p_0(u_k|s_k)^{1-\epsilon} p_1(u_k|s_k)^{\epsilon} > 0 \quad \text{for all } k \in \{1,2,...,n\}.$$

Therefore, if Assumption 2(a) holds, for every $\boldsymbol{\gamma} \in \Gamma^n$ and $\epsilon \in [0,1]$, each agent must posses at least one state $s \in \mathcal{S}$ with $p(s) > 0$ such that $\sum_{u_k} p_0(u_k|s_k)^{1-\epsilon} p_1(u_k|s_k)^{\epsilon} > 0$. This property will prove useful in the proof of Theorem 3.

Observe that $\mu(\boldsymbol{\gamma}, \epsilon)$ does not depend on the hypothesis model. Thus, the term $\frac{1}{n}\log \alpha_n$ can be thought of as the "loss" accrued due to removing the information the network state provides about the true hypothesis. We offer a few comments on this assumption. Recall that in the state-information structure, the fusion center has perfect knowledge of the network state. If the desired condition does not hold, the informed fusion center could could drive the probability of error to zero exponentially fast regardless of the rules used by the agents, thereby creating a pathological case that is not of interest here. Our focus herein is the design of the agents and their rules. Second, in the no-information structure, the uninformed fusion center requires messages from the agents as it does not have access to the state. In this scenario, the hypothesis model must be incorporated into the design of the strategy. However, the uninformed fusion center is still at a disadvantage relative to the informed fusion center. Thus, this assumption provides for a more fair comparison between the two information structures. The sequence α_n will play an important role later when studying the asymptotic relation between the two information structures, as well as asymptotic properties of the no-information structure.

Under Assumption 2, the bounds given in (93) will be tight for sufficiently large n, and so we define [6]

$$\Lambda^{(n)} = \inf_{\boldsymbol{\gamma} \in \Gamma^n} \min_{\epsilon \in [0,1]} \frac{1}{n} \mu(\boldsymbol{\gamma}, \epsilon). \tag{72}$$

Then, under Assumption 2, we have the following.

Theorem 2. *For the state-information structure, the optimal error exponent defined in Equation* (57) *is given by*

$$\Lambda = \lim_{n\to\infty} \inf_{\boldsymbol{\gamma} \in \Gamma^n} \min_{\epsilon \in [0,1]} \frac{1}{n} \mu(\boldsymbol{\gamma}, \epsilon) \tag{73}$$

if the limit exists.

Proof. Assume $\lim_{n\to\infty} \Lambda^{(n)}$ exists and is equal to μ^*. Then, for all $\boldsymbol{\gamma} \in \Gamma^n$, the upper bound in (93) gives us

$$R_n \leq r_n(\boldsymbol{\gamma}) \leq \min_{\epsilon \in [0,1]} \frac{1}{n} \mu(\boldsymbol{\gamma}, \epsilon) \tag{74}$$

for all n. Then, $R_n \leq \Lambda^{(n)}$. The lower bound in (93), together with assumption (b) gives us

$$\frac{\log \alpha_n}{n} - \frac{\log 2}{n} + \Lambda^{(n)} - \frac{\sqrt{2n\theta}}{n} \leq r_n(\boldsymbol{\gamma}) \tag{75}$$

for all $\boldsymbol{\gamma} \in \Gamma^n$. Therefore,

$$\frac{\log \alpha_n}{n} - \frac{\log 2}{n} + \Lambda^{(n)} - \frac{\sqrt{2n\theta}}{n} \leq R_n. \tag{76}$$

[6]We take the minimum over all ϵ since for any $\boldsymbol{\gamma} \in \Gamma^n$, $\mu(\boldsymbol{\gamma}, \epsilon)$ is continuous in ϵ and defined over a compact set.

Under assumption (c) and the assumption that $\lim_{n\to\infty} \Lambda^{(n)} = \mu^*$, we get

$$\lim_{n\to\infty} \left\{ \frac{\log \alpha_n}{n} - \frac{\log 2}{n} + \Lambda^{(n)} - \frac{\sqrt{2n\theta}}{n} \right\} = \lim_{n\to\infty} \Lambda^{(n)} = \mu^*. \tag{77}$$

Hence, $\lim_{n\to\infty} R_n = \mu^*$. □

A few remarks are in order. First, the exponent makes no assumption on the correlation between the states. Second, the hypothesis model has no affect on the asymptotics (provided assumption (c) is satisfied).

To provide a further analysis, we assume the agents' states are mutually independent *a priori*. We can then write,

$$\begin{aligned}
\mu(\boldsymbol{\gamma}, \epsilon) &= \log \left[\sum_{\boldsymbol{s}} \sum_{\boldsymbol{u}} (p_0(\boldsymbol{u}|\boldsymbol{s}))^{1-\epsilon} (p_1(\boldsymbol{u}|\boldsymbol{s}))^{\epsilon} p(\boldsymbol{s}) \right] \\
&= \log \left[\sum_{\boldsymbol{s}} \sum_{\boldsymbol{u}} \prod_{k=1}^{n} (p_0(u_k|s_k))^{1-\epsilon} (p_1(u_k|s_k))^{\epsilon} p(s_k) \right] \\
&= \log \left[\left\{ \sum_{s_1,u_1} (p_0(u_1|s_1))^{1-\epsilon} (p_1(u_1|s_1))^{\epsilon} p(s_1) \right\} ... \right. \\
&\qquad \left. ... \left\{ \sum_{s_n,u_n} (p_0(u_n|s_n))^{1-\epsilon} (p_1(u_n|s_n))^{\epsilon} p(s_n) \right\} \right] \\
&= \sum_{k=1}^{n} \log \left[\sum_{s_k,u_k} (p_0(u_k|s_k))^{1-\epsilon} (p_1(u_k|s_k))^{\epsilon} p(s_k) \right] \\
&= \sum_{k=1}^{n} \mu_k(\gamma_k, \epsilon).
\end{aligned} \tag{78}$$

Thus, $\mu(\boldsymbol{\gamma}, \epsilon)$ is decomposable, that is, it is the sum of the $\mu_k(\gamma_k, \epsilon)$s, where agent k is using rule $\gamma_k \in \Gamma$, $k = 1, 2, ..., n$. Notice that the exponent loses this property if one of the previous assumptions is removed. It can be shown that $\mu(\boldsymbol{\gamma}, \epsilon)$ is convex in ϵ and non-positive, and is zero for all ϵ only in the uninteresting case where the fusion center is unable to distinguish between the two hypotheses, i.e., under all states, $p_0(\boldsymbol{u}|\boldsymbol{s}) = p_1(\boldsymbol{u}|\boldsymbol{s})$ for all $\boldsymbol{u}$. The same result can be shown for $\mu(\gamma_k, \epsilon)$, $k = 1, 2, .., n$. Thus, $\mu(\boldsymbol{\gamma}, \epsilon)$ is non-increasing in n.

B. Identical Agents

We now turn our attention the case where the agents are identical. Define

$$\Lambda_r = \inf_{\gamma \in \Gamma} \min_{\epsilon \in [0,1]} \frac{1}{n} \sum_{k=1}^{n} \mu_k(\gamma, \epsilon). \tag{79}$$

Note that Λ_r is defined only over those strategies that have all agents use the same rule. The main result of this section is that if the agents are weakly identical, then, as $n \longrightarrow \infty$, we can restrict ourselves to strategies where all agents use the same rule without loss of asymptotic optimality.

Theorem 3. *Assume the agents are weakly identical, the agents' states are mutually independent a priori, and Assumption 2 holds. Then,* $\Lambda^{(n)} = \Lambda_r$ *for all* $n \in \mathbb{N}$.

Proof. Since having all agents use the same rule is a valid strategy, we have $\Lambda^{(n)} \leq \Lambda_r$. For any $\boldsymbol{\gamma} \in \Gamma^n$, let Φ_0 be the set of prescriptions used by the agents, i.e., $\Phi_0 \subseteq \Phi$. Now, for any $k = 1, 2, ..., n$, $\gamma \in \Gamma$, and $\epsilon \in [0, 1]$,

$$\begin{aligned}\mu_k(\gamma_k, \epsilon) &= \log \sum_{s,u} (p_0(u|s))^{1-\epsilon} (p_1(u|s))^{\epsilon} p(S_k = s) \\ &= \log \sum_{s,u} p_0(\phi_s(Y) = u|s)^{1-\epsilon} p_1(\phi_s(Y) = u|s)^{\epsilon} p(S_k = s) \qquad (80) \\ &\geq \log \sum_{s,u} p_0(\phi_s^*(Y) = u|s)^{1-\epsilon} p_1(\phi_s^*(Y) = u|s)^{\epsilon} p(S_k = s) \\ &= \mu_k(\gamma^*, \epsilon), \qquad (81)\end{aligned}$$

where

$$\phi_s^* = \arg \min_{\phi \in \Phi_0} \{ \sum_u (p_0(u|s))^{1-\epsilon} (p_1(u|s))^{\epsilon} \} \qquad (82)$$

and γ^* is the rule that uses ϕ_s^* when in state s. Thanks to Assumption 2(a) the terms inside all logarithms are strictly greater that zero, and so all logarithms are well-defined. Notice that since the agents are weakly identical, the rule γ^* can be chosen to be the same for all agents, and thus does not depend on k. For any $\boldsymbol{\gamma} \in \Gamma^n$ and $\epsilon \in [0, 1]$,

$$\frac{1}{n} \sum_{k=1}^{n} \mu_k(\gamma_k, \epsilon) \geq \frac{1}{n} \sum_{k=1}^{n} \mu(\gamma^*, \epsilon) \geq \Lambda_r. \qquad (83)$$

Hence, $\Lambda^{(n)} \geq \Lambda_r$. We then conclude $\Lambda^{(n)} = \Lambda_r$. □

Recall that our definition of weakly identical does not make any statement about the states of the agents. To provide more analysis, we turn our attention to strongly identical agents.

If the agents are strongly identical, $\mu_k(\gamma, \epsilon)$ does not depend on k. We then drop the subscript when we are discussing strongly identical sensors. We see that for any $\gamma \in \Gamma$ and $\epsilon \in [0, 1]$,

$$\frac{1}{n} \sum_{k=1}^{n} \mu_k(\gamma, \epsilon) = \frac{1}{n} \sum_{k=1}^{n} \mu(\gamma, \epsilon) = \mu(\gamma, \epsilon). \qquad (84)$$

We define

$$\Lambda_0 = \inf_{\gamma \in \Gamma} \min_{\epsilon \in [0,1]} \mu(\gamma, \epsilon). \qquad (85)$$

Corollary 3.1. *Assume the agents are strongly identical and that both hypotheses are possible under all states. Then,* $\Lambda = \Lambda_0$.

Proof. We have from Theorem 3 that $\Lambda^{(n)} = \Lambda_r$ for all $n \in \mathbb{N}$. Moreover, (84) gives us that

$$\Lambda_r = \inf_{\gamma \in \Gamma} \min_{\epsilon \in [0,1]} \frac{1}{n} \sum_{k=1}^{n} \mu_k(\gamma, \epsilon) = \inf_{\gamma \in \Gamma} \min_{\epsilon \in [0,1]} \mu(\gamma, \epsilon) = \Lambda_0.$$

Hence,

$$\Lambda_0 = \lim_{n \to \infty} \Lambda^{(n)} = \lim_{n \to \infty} \Lambda_0 = \Lambda_0.$$

□

Theorem 3 states that there is no loss of asymptotic optimality if all agents use the same rule. Moreover, if the agents are strongly identical, one only needs to solve the optimization problem in (85).

C. No-Information Error Exponent

Following classic results in the literature, [19], [39], it is readily shown that for the no-information case, the error exponent Λ_I is given as

$$\Lambda_I = \lim_n \inf_{\boldsymbol{\gamma} \in \Gamma^n} \min_{\epsilon \in [0,1]} \frac{1}{n} \log \sum_{\boldsymbol{u}} p_0(\boldsymbol{u})^{1-\epsilon} p_1(\boldsymbol{u})^{\epsilon}. \qquad (86)$$

There is however, an additional subtlety introduced by our problem. Namely, in order for (86) to hold, one needs to ensure that $\frac{1}{n} \log \min\{\pi_0, \pi_1\} \to 0$, since the priors are allowed to vary with n. Indeed,

$$\frac{1}{n} \log \alpha_n \leq \frac{1}{n} \log \sum_{\boldsymbol{s}} p(h|\boldsymbol{s}) p(\boldsymbol{s}) = \frac{1}{n} \log \pi_h < 0,$$

where $\frac{1}{n} \log \alpha_n \to 0$ by assumption. Let

$$\Lambda_I^{(n)} = \inf_{\boldsymbol{\gamma} \in \Gamma^n} \min_{\epsilon \in [0,1]} \frac{1}{n} \log \sum_{\boldsymbol{u}} p_0(\boldsymbol{u})^{1-\epsilon} p_1(\boldsymbol{u})^{\epsilon} \qquad (87)$$

To further motivate the study of α_n, consider the following.

$$\begin{aligned}&\sum_{\boldsymbol{u}} p_0(\boldsymbol{u})^{1-\epsilon} p_1(\boldsymbol{u})^{\epsilon} \\ &\geq \sum_{\boldsymbol{u}} [p_0(\boldsymbol{u})\pi_0]^{1-\epsilon} [p_1(\boldsymbol{u})\pi_1]^{\epsilon} \\ &\overset{(a)}{\geq} \sum_{\boldsymbol{u},\boldsymbol{s}} p(\boldsymbol{u}, \boldsymbol{s}, h_0)^{1-\epsilon} p(\boldsymbol{u}, \boldsymbol{s}, h_1)^{\epsilon} \\ &\overset{(b)}{\geq} \alpha_n \sum_{\boldsymbol{u},\boldsymbol{s}} p_0(\boldsymbol{u}|\boldsymbol{s})^{1-\epsilon} p_1(\boldsymbol{u}|\boldsymbol{s})^{\epsilon} p(\boldsymbol{s}).\end{aligned} \qquad (88)$$

where (a) is due to Hölder's inequality and (b) is from the definition of α_n (Equation 55). It then follows that

$$\Lambda_I^{(n)} - \Lambda^{(n)} \geq \frac{1}{n} \log \alpha_n \qquad (89)$$

Notice that $\frac{1}{n} \log \alpha_n$ is a sequence of strictly negative numbers. Hence, the bound given above potentially allows for the possibility that $\Lambda_I^{(n)} < \Lambda^{(n)}$ for a small enough n. In the numerical results section, we explicitly construct an example such that this property holds for relatively large networks. Assumption 2c characterizes the case when $\Lambda_I \geq \Lambda$, and in this case, the no-information structure cannot do better than the state-information structure, asymptotically. Moreover, the proof of Theorem 3 relies on the fact that $\mu(\boldsymbol{\gamma}, \epsilon)$ is decomposable. In contrast, $\frac{1}{n} \log \sum_{\boldsymbol{u}} p_0(\boldsymbol{u})^{1-\epsilon} p_1(\boldsymbol{u})^{\epsilon}$ does not possess the same property. This follows because even if the agents' states are independent *a priori*, they are still correlated through the hypothesis model. This motivates us to define the following quantity, which we call the *untethered exponent*.

$$\nu(\boldsymbol{\gamma}, \epsilon) \quad = \quad \log \sum_{\boldsymbol{u}} q_0(\boldsymbol{u})^{1-\epsilon} q_1(\boldsymbol{u})^{\epsilon} \qquad (90)$$

$$\text{with} \quad q_h(\boldsymbol{u}) \;=\; \sum_{\boldsymbol{s}} p_h(\boldsymbol{u}|\boldsymbol{s})p(\boldsymbol{s}), \;\; \text{for} \;\; h=0,1, \tag{91}$$

The quantity $\nu(\boldsymbol{\gamma},\epsilon)$ shares several important and useful properties with $\mu(\boldsymbol{\gamma},\epsilon)$. First, $\log\sum_{\boldsymbol{u}} q_0(\boldsymbol{u})^{1-\epsilon}q_1(\boldsymbol{u})^{\epsilon}$ is a continuous, convex function of ϵ and we have $\nu(\boldsymbol{\gamma},\epsilon)\leq 0$. From convexity, $\nu(\boldsymbol{\gamma})=0 \;\; \forall\, \epsilon\in[0,1]$ only if $q_0(\boldsymbol{u})=q_1(\boldsymbol{u})$, $\forall \;\; \boldsymbol{u}$. Moreover, if the agent states are independent *a priori* and both hypotheses are possible under all $\boldsymbol{s}$, then

$$\begin{aligned}&\log\sum_{\boldsymbol{u}}\left[\sum_{\boldsymbol{s}} p_0(\boldsymbol{u}|\boldsymbol{s})p(\boldsymbol{s})\right]^{1-\epsilon}\left[\sum_{\boldsymbol{s}} p_1(\boldsymbol{u}|\boldsymbol{s})p(\boldsymbol{s})\right]^{\epsilon}\\ &=\sum_{k}\log\sum_{u_k} q_0(u_k)^{1-\epsilon}q_1(u_k)^{\epsilon}\end{aligned} \tag{92}$$

The above computation clearly shows that $\nu(\boldsymbol{\gamma},\epsilon)$ is decomposable, a property we exploit and a property not observed in the classical exponent representation.

Proposition 1. *Let* $\nu(\boldsymbol{\gamma})=\min_{\epsilon\in[0,1]}\nu(\boldsymbol{\gamma},\epsilon)$. *Then, for all* $n\in\mathbb{N}$,

$$\begin{aligned}&\frac{1}{n}\log\alpha_n+\frac{1}{n}\inf_{\boldsymbol{\gamma}\in\Gamma^n}\nu(\boldsymbol{\gamma})\leq\Lambda_I^{(n)}\\ &\leq-\frac{1}{n}\log\min\{\pi_0,\pi_1\}+\frac{1}{n}\inf_{\boldsymbol{\gamma}\in\Gamma^n}\nu(\boldsymbol{\gamma}).\end{aligned} \tag{93}$$

For any $\epsilon\in[0,1]$, we have

$$\begin{aligned}&\sum_{\boldsymbol{u}} p_0(\boldsymbol{u})^{1-\epsilon}p_1(\boldsymbol{u})^{\epsilon}\\ &=\frac{1}{\pi_0}^{1-\epsilon}\frac{1}{\pi_1}^{\epsilon}\sum_{\boldsymbol{u}}\left[\sum_{\boldsymbol{s}} p_0(\boldsymbol{u}|\boldsymbol{s})p(h_0|\boldsymbol{s})p(\boldsymbol{s})\right]^{1-\epsilon}\\ &\qquad\left[\sum_{\boldsymbol{s}} p_1(\boldsymbol{u}|\boldsymbol{s})p(h_1|\boldsymbol{s})p(\boldsymbol{s})\right]^{\epsilon}\\ &\leq\max\{\frac{1}{\pi_0},\frac{1}{\pi_1}\}\sum_{\boldsymbol{u}}\left[\sum_{\boldsymbol{s}} p_0(\boldsymbol{u}|\boldsymbol{s})p(\boldsymbol{s})\right]^{1-\epsilon}\\ &\qquad\left[\sum_{\boldsymbol{s}} p_1(\boldsymbol{u}|\boldsymbol{s})p(\boldsymbol{s})\right]^{\epsilon}\end{aligned}$$

where we use the fact that $a^{\epsilon}b^{1-\epsilon}\leq\max\{a,b\} \quad \forall\epsilon\in[0,1]$. Combining the above with

$$\sum_{\boldsymbol{u}} p_0(\boldsymbol{u})^{1-\epsilon}p_1(\boldsymbol{u})^{\epsilon}\geq\alpha_n\sum_{\boldsymbol{u}} q_0(\boldsymbol{u})^{1-\epsilon}q_1(\boldsymbol{u})^{\epsilon},$$

taking logarithms, and dividing by n gives us Proposition 1.

Notice that thanks to Proposition 1,

$$\Lambda_I=\lim_n\inf_{\boldsymbol{\gamma}\in\Gamma^n}\frac{1}{n}\nu(\boldsymbol{\gamma}), \tag{94}$$

where equality denotes that either both limits exist, or neither does. Hence, for sufficiently large n, one can, *without loss of optimality*, restrict attention to $\nu(\boldsymbol{\gamma})$. Given that this expression does not depend on the hypothesis model, it is easier to compute for large network sizes than (87). Moreover, the untethered exponent gives us the following.

Theorem 4. *Under Assumption 2, and assuming all agents are strongly identical, we have that for all* $n\in\mathbb{N}$, $\inf_{\boldsymbol{\gamma}\in\Gamma^n}\frac{1}{n}\nu(\boldsymbol{\gamma})=\inf_{\gamma\in\Gamma}\nu(\gamma)$. *Hence, identical rules are optimal.*

Proof. For any fixed n, using identical rules is a valid strategy. Therefore, the optimal strategy can do no worse than the optimal identical strategy, and so we have $\inf_{\boldsymbol{\gamma}\in\Gamma^n}\frac{1}{n}\nu(\boldsymbol{\gamma})\leq\inf_{\gamma\in\Gamma}\nu(\gamma)$. It then suffices to show that the reverse inequality holds. First, we can show, similar to $\mu(\boldsymbol{\gamma},\epsilon)$, that under Assumption 2 $|\nu(\boldsymbol{\gamma})|$ is finite. Hence, all logarithms are well-defined. Since the agents are strongly identical, we have

$$\nu(\boldsymbol{\gamma})=\sum_k\nu_k(\gamma_k). \tag{95}$$

Since this sum is finite for any $\boldsymbol{\gamma}$, we have that $\nu_k(\gamma)$ is also finite for any rule $\gamma\in\Gamma$, $k=1,2,..,n$. Notice that if the agents are identical, $\nu_k(\gamma)$ depends only on the rule being used by agent k, and so ν_k will be the same for different agents as they all employ the same rule. Then, for a given strategy $\boldsymbol{\gamma}$, if we denote the set of distinct rules used by the agents as $\mathcal{G}$, and the number of agents using rule γ as N_γ, we get

$$\begin{aligned}\nu(\boldsymbol{\gamma})&=\sum_k\nu_k(\gamma_k)=\sum_{\gamma\in\mathcal{G}}N_\gamma\nu(\gamma)\\ &\geq\sum_{\gamma\in\mathcal{G}}N_\gamma\nu(\gamma*)=n\nu(\gamma*)\end{aligned} \tag{96}$$

where $\gamma*=\arg\min_{\gamma\in\mathcal{G}}\nu(\gamma)$. This shows that for any $\boldsymbol{\gamma}$, we can improve the system's performance by simply selecting the best $\nu_k(\gamma)$, and having all agents use agent k's rule. Taking the infimum of both sides and dividing by n yields the desired inequality. Invoking Proposition 1 we see that

$$\Lambda_I=\lim_n\inf_{\gamma\in\Gamma}\nu(\boldsymbol{\gamma}) \tag{97}$$

provided both limits exist. Hence, optimizing one agent is sufficient for asymptotic network optimality. □

It is important to note some diffirences between the proofs of Theorems 3 and 4. First, Theorem 3 holds for *weakly identical* agents whereas Theorem 4 holds only for *strongly identical* agents. The reason for this can be seen in the structural differences of μ and ν. In μ, the message and the state are in some sense "separable". This is seen by the double summation over u and s. Hence, when working with μ, it is possible, by interchanging sums, to fix the state s and design the prescription that optimizes that state. This same procedure cannot be done with ν, since one cannot interchange the summations over u and s. Hence, when working with ν, one needs the agents to be strongly identical to invoke these simplifying results.

Notice that our definition of identicality makes no assumption on the hypothesis model. Some agents' states may be more correlated with the hypothesis, and therefore with each other. Therefore, when designing a system according to Λ_I^n, changing one agent's rule may affect the other agents through $p(h|\boldsymbol{s})$. However, our result shows that for sufficiently large n this cannot happen, since Λ_I^n and ν are asymptotically equivalent. This further highlights the fact that although ν

and $\Lambda_I^{(n)}$ are asymptotically equivalent, the additional structure of ν provides insight into the structure and simplicity of the optimal solution.

V. Numerical Results

In this section we present some numerical results to illustrate several important concepts from the preceding sections.

- In the state-information case, for a fixed network state, the problem is the canonical decentralized detection problem with parallel configuration and conditionally independent observations. Hence, one might be tempted to fix the network state and use previously established methods such as [22]. We show that this approach is, in general, sub-optimal.
- In the no-information case, the only true correlation between the agents when conditioned on the hypothesis is in the agents' state. We then derive an approach which assumes the observations (Y_k, S_k), $k = 1, 2, ..., n$ are indeed conditionally independent, and use previous results. Again, we show this approach is sub-optimal, and so in both cases, this state structure should not be ignored.
- We construct a system where $\Lambda_I^{(n)} < \Lambda^{(n)}$ for a regime of network sizes n. That is, for a regime of network sizes, the Chernoff information of the no-information case ($\Lambda_I^{(n)}$) *is higher than the Chernoff Information* ($-\Lambda^{(n)}$) *of the state-information case*.

For the rest of the section, we assume each agent k takes one of two states, $s_k = 0$ or $s_k = 1$, makes a binary decision, $u_k = 0$ or $u_k = 1$, and has the following signal model,

$$Y_k = H[1 + S_k(\beta - 1)] + N_k \tag{98}$$

where $\beta \in (0, 1]$ is a constant, and $\{N_k\}_1^n$ i.i.d with $N_k \sim \mathcal{N}(0, 1)$. Notice that when $s_k = 0$, $Y_k = H + N_k$, and when $s_k = 0$, $Y_k = \beta H + N_k$. Hence, β may be thought of as a jamming constant, which is only present when $s_k = 1$.

A. Small Networks

In this section, we present numerical results to support the first two points above. We consider a simple network with two agents, and hypothesis model given by (98)

$$p(h_1|s_1, s_2) = \frac{1}{4} + \frac{1}{4}(s_1 + s_2). \tag{99}$$

Moreover, we assume the joint pmf of the states is given as $p(s_1 = i, s_2 = j) = q_{ij}$ with

$$\begin{bmatrix} q_{0,0} & q_{0,1} \\ q_{1,0} & q_{1,1} \end{bmatrix} = \begin{bmatrix} .2 & .1 \\ .1 & .6 \end{bmatrix}. \tag{100}$$

We now outline each strategy used:

A) For both the state-information and no-information structures, we use the iterative algorithm developed in the appendix for each respective information structure. This algorithm is developed based on the PBPO criteria developed in Section III. The state-information case is denoted as Strategy A1, while the no-information structure is denoted as Strategy A2.

B) For the state-information structure, fix the network state $\boldsymbol{s}$. For a fixed state, use the algorithm developed in [22]. In general, this may produce different thresholds for each agent in a given state. For example, if agent 1 is in state 0, there is no guarantee the thresholds τ_0, τ_1, corresponding to the network states $[0, 0]$ and $[0, 1]$ respectively, are the same. Hence, for a fixed s_k, we use threshold τ_0 with probability $p(s_j = 0|s_k)$ and τ_1 with probability $p(s_j = 1|s_k)$, so

$$p_h(u_k = 1|s_k) = Q(\tau_0 - h[1 + s_k(\beta - 1)])p(s_j = 0|s_k) + Q(\tau_1 - h[1 + s_k(\beta - 1)])p(s_j = 1|s_k). \tag{101}$$

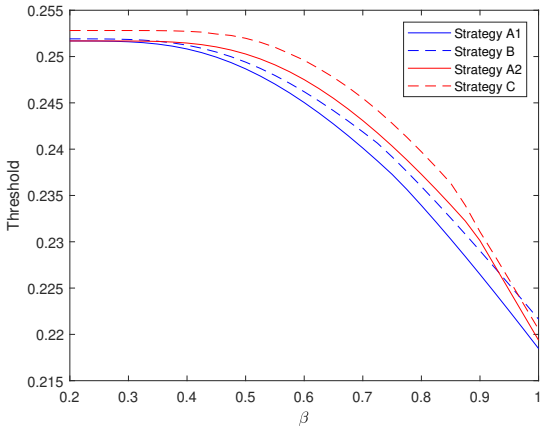
C) For the no-information structure, assume the observations $\{(Y_k, S_k)\}_1^2$ are conditionally independent and use the algorithm given in [22]. Moreover,

$$p_h(y_k, s_k) = \frac{p_h(y_k|s_k)p(h|s_k)p(s_k)}{p(h)}. \tag{102}$$

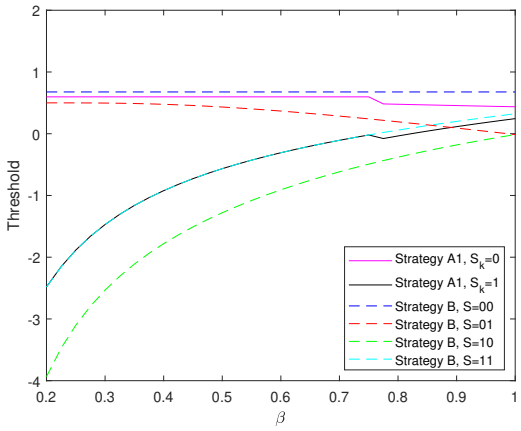
where $p(h|s_k)$ and $p(h)$ are induced by the given hypothesis model.

Also, all fusion centers use the appropriate MAP rule for each strategy. We note some interesting phenomena for each information structure.

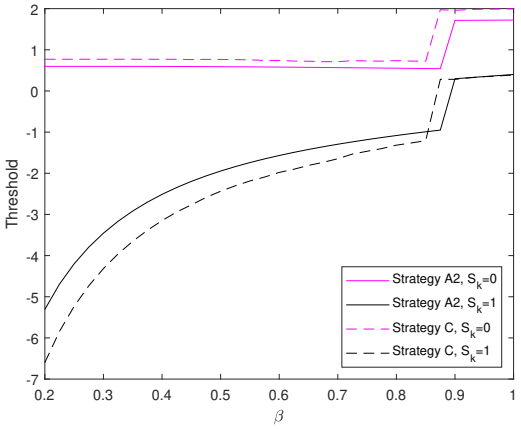
- *State-information structure*: As stated before, and illustrated in Fig. 2a, fixing the network state and using existing methods is sub-optimal. There are a few reasons for this. First, finding the optimal strategy for a fixed state neglects the performance of the other states. The strategy that optimizes the performance in one network state may only be able to do so at the expense of another state. For example, in our model, when both agents are in state 0, their signal models are identical, and both have a "true" look at the underlying hypothesis. However, if one optimizes this state, one neglects the state $[0, 1]$, in which case the second agent does not have a clear look at the hypothesis. It is not unreasonable to expect that having agent 2 use the $[0, 0]$ strategy would be sub-optimal in state $[0, 1]$. Indeed, this is the case. Second, upon fixing the network state, one removes the effect of each agent state on the hypothesis. If the network state is fixed, so too is $p(h|s_1, s_2)$, and so each agent's state has no information regarding the hypothesis. Compare this to the case where the network state is not fixed. Then, upon taking state s_k, agent k receives some knowledge of the hypothesis, namely of $p(h|s_1, s_2)$, through its state. This is captured when $\beta = 1$. Notice that in this case, the signal models under both states are identical, but the thresholds themselves are not. This results from the fact that the states themselves carry asymmetric information about the hypothesis. In this example, for each agent in a given state, we randomized between the two rules obtained for each network state, according to $p(s_j|s_k)$. Hence, one may be tempted to say that finding the optimal strategy might amount to finding the optimal randomization between optimal network rules. However, it is not



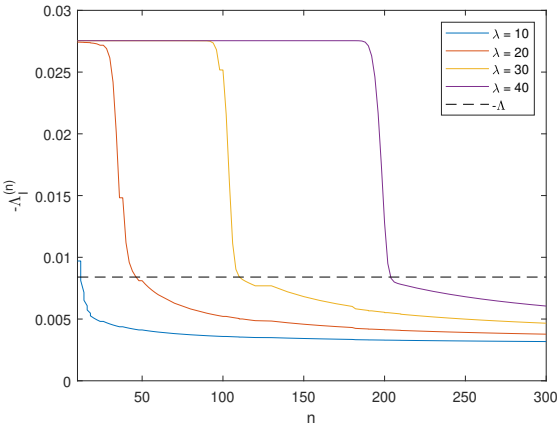
(a) Probability of error for the various strategies.



(b) Thresholds for the state-information strategies.



(c) Thresholds for the no-information strategies.



(d) Comparison of the Chernoff information for the two information structures for varying λ.

Fig. 2: Numerical results

clear that this problem has any desirable properties, i.e., convexity, quasi-convexity, etc., and so may be difficult for certain models. Moreover, this approach does not scale well in either the network size or network state-space.

- *No-information structure*: When assuming conditional independence between the observations, the thresholds themselves are different from the optimal thresholds, resulting in sub-optimal performance. Hence, one should not ignore the correlation between agents states and the hypothesis.

B. Large Networks

We now turn our attention to large networks, and focus on the Chernoff information. We explicitly construct a system with the desired behavior as described earlier. Consider a network with n agents. Each agent takes binary state, makes a binary decision, and has the same signal model given by (98). The hypothesis model is given as

$$p(h_0|\boldsymbol{s}) = \exp(-\lambda\sqrt{z+1}) \tag{103}$$

where $\lambda > 0$, and $z = \sum_k s_k$. For sufficiently large n, $\alpha_n = \exp(-\lambda\sqrt{n+1})$, so $\frac{1}{n}\log\alpha_n = -\lambda\frac{\sqrt{n+1}}{n} \to 0$. Moreover, the agent states are i.i.d with $p(S_k = 1) = p$. Hence, the agents are strongly identical, and so we can restrict ourselves to identical rules to obtain $\Lambda^{(n)}$. Moreover, we can optimize the Chernoff information for only a single agent. Then, to find the optimal thresholds for the fusion center with state information, we find the thresholds that minimize

$$\min_{[0,1]} \log \Big[Q(\tau_0)^{1-\epsilon}Q(\tau_0-1)^{\epsilon}+ (1-Q(\tau_0))^{1-\epsilon}(1-Q(\tau_0-1))^{\epsilon}\Big](1-p) + \Big[Q(\tau_1)^{1-\epsilon} Q(\tau_1-\beta)^{\epsilon} + (1-Q(\tau_1))^{1-\epsilon}(1-Q(\tau_1-\beta))^{\epsilon}\Big]p. \tag{104}$$

We now turn our attention to $\Lambda_I^{(n)}$, and begin by writing

$$\begin{aligned}
&\log \sum_{\boldsymbol{u}} p_0(\boldsymbol{u})^{1-\epsilon} p_1(\boldsymbol{u})^{\epsilon} = \\
&\log \sum_{\boldsymbol{u}} \left[\frac{1}{p(h_0)} \sum_{\boldsymbol{s}} p_0(\boldsymbol{u}|\boldsymbol{s}) p(h_0|\boldsymbol{s}) p(\boldsymbol{s})\right]^{1-\epsilon} \\
&\left[\frac{1}{p(h_1)} \sum_{\boldsymbol{s}} p_1(\boldsymbol{u}|\boldsymbol{s}) p(h_1|\boldsymbol{s}) p(\boldsymbol{s})\right]^{\epsilon} = \\
&\log \sum_{\boldsymbol{u}} \frac{1}{p(h_0)}^{1-\epsilon} \frac{1}{p(h_1)}^{\epsilon} \Bigg[\sum_{\boldsymbol{s}} \prod_k p_0(u_k|s_k) \\
&p(h_0|\boldsymbol{s}) p(\boldsymbol{s})\Bigg]^{1-\epsilon} \Bigg[\sum_{\boldsymbol{s}} \prod_k p_1(u_k|s_k) p(h_1|\boldsymbol{s}) p(\boldsymbol{s})\Bigg]^{\epsilon}.
\end{aligned} \tag{105}$$

First, notice that $p(h_0|\boldsymbol{s})$ depends only on the type of s, and so

$$p(h_0) = \sum_{z=0}^{n} \exp(-\lambda\sqrt{z+1}) \binom{n}{z} p^z (1-p)^{(n-z)} \tag{106}$$

Similarly, the summations in (105) can be written with respect to ω, z, z_0, and z_1, where ω is the number of agents that decide 1, z_0 is the number of agents that decide 0 in state 1, and z_1 is the number of agents that decide 1 in state 1. Then, we can rewrite (105) as

$$\log \frac{1}{p(h_0)}^{1-\epsilon} \frac{1}{p(h_1)}^{\epsilon} \sum_{\omega} f_0(\omega)^{1-\epsilon} f_1(\omega)^{\epsilon} \tag{107}$$

where $f_h(\omega)$ is given as

$$\begin{aligned}
&f_h(\omega) = \\
&\sum_{z} \sum_{\max\{\omega-n+z\},0\} \le z_1 \le \min\{\omega,z\}} \binom{\omega}{z_1} \binom{n-\omega}{z_0} \\
&(q_{h,0,0})^{n-\omega-z_0} (q^{h,0,1})^{z_0} (1-q_{h,0,0})^{\omega-z_1} (1-q_{h,0,1})^{z_1} \\
&\exp(-\lambda\frac{z}{n}) p^z (1-p)^{(n-z)},
\end{aligned} \tag{108}$$

with $q_{h,u,s} = p_h(u|s)$. Our results are presented in Fig. 2d. As stated before, we find the thresholds that optimize the state-information structure, and use these thresholds in the no-information structure as well. Thus, if the no-information structure outperforms the state-information structure with these thresholds, our claim is validated (since the optimal thresholds for the no-information structure must do at least as well). Indeed, this is the case as illustrated in Fig. 2d. As λ increases, the size of the network required for the Chernoff information of the state-information case to outperform the no-information case increases as well. Hence, it is in fact possible for the fusion center without state information to outperform the fusion center with state information in large networks. While this result seems highly counter-intuitive, this phenomena become more clear upon examining the expressions for the Chernoff information between the two structures. Notice that the Chernoff information for the state-information structure in no way depends upon the hypothesis model, whereas the Chernoff information for the no information structure *implicitly depends on the hypothesis model*. This however, does not result in a loss of optimality thanks to Assumption 2c, which states the information contained in the hypothesis model goes to zero in the network size. Moreover, it is because of Assumption 2c that we have $\Lambda_I \le \Lambda$. However, Assumption 2c makes no statement about the speed of decay. That is, the hypothesis model may contain substantial information about the underlying hypothesis even for large networks, but eventually, this information must decay to zero as the network grows. Indeed, this is what happens in our example. As λ increases, the rate of decay slows, and so the hypothesis model carries more information even for larger networks. This is seen in Fig. 2d from the fact that for increasing λ the no-information structure outperforms the state-information structure for larger and larger network sizes.

VI. Conclusion

In this paper, we have formulated the problem of multi-agent hypothesis testing over networks with state, and shown a number of interesting properties. We have found the optimal design rule for the state-information structure under a min-max criterion, and the optimal design rule for both structures under the Bayesian setting. We have also found the optimal error decay rate for both information structures and used them to prove a number of interesting results. Namely that identical rules are optimal for weakly identical agents, and that it is indeed possible for the no-information structure to outperform the state-information structure, even for relatively large networks.

Appendix

A. Proof of Lemma 7

Proof. For any strategy $\boldsymbol{\gamma}$, let $\mathcal{M}^{\boldsymbol{\gamma}}$ be the smallest sub σ-algebra of $\mathcal{M}_{\boldsymbol{y},\boldsymbol{s}}$ such that $\boldsymbol{\gamma}$ is measurable. Let $p_0^{\boldsymbol{\gamma}}$ and $p_1^{\boldsymbol{\gamma}}$ denote the restrictions of p_0 and p_1 to $\mathcal{M}^{\boldsymbol{\gamma}}$. Moreover, from absolute continuity, we have that $p_0^{\boldsymbol{\gamma}} \ll p_1^{\boldsymbol{\gamma}}$. Let

$$L_{\boldsymbol{s}}^{\boldsymbol{\gamma}}(\boldsymbol{y}) = \frac{dp_1^{\boldsymbol{\gamma}}(\bullet|\boldsymbol{s})}{dp_0^{\boldsymbol{\gamma}}(\bullet|\boldsymbol{s})}. \tag{109}$$

Then, we make two important observations. First,

$$L_{\boldsymbol{s}}^{\boldsymbol{\gamma}} = \mathbb{E}_0[L_{\boldsymbol{s}}(\boldsymbol{Y})|\mathcal{F}^{\boldsymbol{\gamma}}] \tag{110}$$

a.s., for all $\boldsymbol{s}$, and

$$L_{\boldsymbol{s}}(\boldsymbol{u}) = L_{\boldsymbol{s}}^{\boldsymbol{\gamma}}(\boldsymbol{y}) \tag{111}$$

for every $\boldsymbol{y}$ such that $\boldsymbol{\gamma}(\boldsymbol{y},\boldsymbol{s}) = \boldsymbol{u}$, almost surely. Using (110) together with convexity of $\zeta(x) = x\log^2 x$ yields, for fixed $\boldsymbol{s}$,

$$\begin{aligned}
&\mathbb{E}_1[\log^2 L_{\boldsymbol{s}}^{\boldsymbol{\gamma}}] = \mathbb{E}_0[L_{\boldsymbol{s}}^{\boldsymbol{\gamma}} \log^2 L_{\boldsymbol{s}}^{\boldsymbol{\gamma}}] = \mathbb{E}_0[\zeta(L_{\boldsymbol{s}}^{\boldsymbol{\gamma}})] \\
=&\mathbb{E}_0[\zeta(\mathbb{E}_0[L_{\boldsymbol{s}}|\mathcal{F}^{\boldsymbol{\gamma}}])] \le \mathbb{E}_0[\mathbb{E}_0[\zeta(L_{\boldsymbol{s}})|\mathcal{F}^{\boldsymbol{\gamma}}]] \\
=&\mathbb{E}_0[L_{\boldsymbol{s}} \log^2 L_{\boldsymbol{s}}] = \mathbb{E}_1[\log^2 L_{\boldsymbol{s}}].
\end{aligned} \tag{112}$$

Then, there exists some constant $c_l < \infty$ such that $\mathbb{E}_1[\log^2 L_{\boldsymbol{s}}^{\boldsymbol{\gamma}}] < c_l$. Moreover, using $\mathbb{E}[|X|] \le 1+\mathbb{E}[X^2]$, we get

$\mathbb{E}_1[|\log L_{\boldsymbol{s}}^{\boldsymbol{\gamma}}|] < c_l$. Next, using (111), (63), and the inequality $(L_{\boldsymbol{s}}^{\boldsymbol{\gamma}})^\epsilon \leq 1 + L_{\boldsymbol{s}}^{\boldsymbol{\gamma}}$, we obtain

$$\begin{aligned} |\mu'(\boldsymbol{\gamma},\epsilon)| &= \frac{|\mathbb{E}_{\boldsymbol{s}}[\mathbb{E}_0[(L_{\boldsymbol{s}}^{\boldsymbol{\gamma}})^\epsilon \log L_{\boldsymbol{s}}^{\boldsymbol{\gamma}}]]|}{|\mathbb{E}_{\boldsymbol{s}}[\mathbb{E}_0[(L_{\boldsymbol{s}}^{\boldsymbol{\gamma}})^\epsilon]]|} \\ &\leq \frac{|\mathbb{E}_{\boldsymbol{s}}[\mathbb{E}_1[\log L_{\boldsymbol{s}}^{\boldsymbol{\gamma}}]]| + |\mathbb{E}_{\boldsymbol{s}}[\mathbb{E}_0[\log L_{\boldsymbol{s}}^{\boldsymbol{\gamma}}]]|}{|\mathbb{E}_{\boldsymbol{s}}[\mathbb{E}_0[(L_{\boldsymbol{s}}^{\boldsymbol{\gamma}})^\epsilon]]|}. \end{aligned} \tag{113}$$

We have already bounded the numerator. The denominator is bounded by noticing that $\mathbb{E}_0[L_{\boldsymbol{s}}] = 1$ for all $\boldsymbol{s}$. Hence, there must exist some $\mathcal{M}_{\boldsymbol{y},\boldsymbol{s}}$ measurable set $\mathcal{E}$, such that for all $\boldsymbol{s}$, there exist $\delta_1^{\boldsymbol{s}} > 0$, and $\delta_2^{\boldsymbol{s}} > 0$ with $p_0(\mathcal{E}|\boldsymbol{s}) > \delta_1^{\boldsymbol{s}}$ and $L_{\boldsymbol{s}} > \delta_2^{\boldsymbol{s}}$. Then, using that $x^\epsilon \geq \min\{1, x\}$ for all $\epsilon \in [0,1]$, we get $\mathbb{E}_0[(L_{\boldsymbol{s}})^\epsilon] \geq \delta_1^{\boldsymbol{s}} \min\{1, \delta_2^{\boldsymbol{s}}\}$ for all $\epsilon \in [0,1]$. Using the fact that x^ϵ is concave for fixed $\epsilon \in [0,1]$ yields

$$\begin{aligned} &\mathbb{E}_0[(L_{\boldsymbol{s}}^{\boldsymbol{\gamma}})^\epsilon] = \mathbb{E}_0[\mathbb{E}_0[L_{\boldsymbol{s}}|\mathcal{F}^{\boldsymbol{\gamma}}]^\epsilon] \\ \geq &\mathbb{E}_0[\mathbb{E}_0[(L_{\boldsymbol{s}})^\epsilon|\mathcal{F}^{\boldsymbol{\gamma}}]] = \mathbb{E}_0[(L_{\boldsymbol{s}})^\epsilon] \geq \delta_1^{\boldsymbol{s}} \min\{1, \delta_2^{\boldsymbol{s}}\}. \end{aligned} \tag{114}$$

taking the minimum over all $\boldsymbol{s}$ of $\delta_1^{\boldsymbol{s}} \min\{1, \delta_2^{\boldsymbol{s}}\}$ completes the bound on $|\mu'(\boldsymbol{\gamma},\epsilon)|$. For $|\mu''(\boldsymbol{\gamma},\epsilon)|$, we have

$$\mu''(\boldsymbol{\gamma},\epsilon) = \frac{\mathbb{E}_{\boldsymbol{s}}[\mathbb{E}_0[(L_{\boldsymbol{s}}^{\boldsymbol{\gamma}})^\epsilon \log^2 L_{\boldsymbol{s}}^{\boldsymbol{\gamma}}]]}{\mathbb{E}_{\boldsymbol{s}}[\mathbb{E}_0[(L_{\boldsymbol{s}}^{\boldsymbol{\gamma}})^\epsilon]]} - (\mu'(\boldsymbol{\gamma},\epsilon))^2.$$

Then, the proof that $\mu''(\boldsymbol{\gamma},\epsilon)$ is bounded is identical, and is thus omitted. □

B. Iterative Algorithm for Optimal Quantizers

In this section we develop an iterative algorithm to find the local quantizers for each agent following a person-by-person optimal approach for the Bayesian setting. The iterative algorithm we present is a slightly modified version of that given in [22]. Thus, we skip some of the details in the derivation, and refer the reader to [22] for a thorough treatment. We start by exploiting the structure of the rule given in (48) with the following. Define

$$Z_{kij}^s = \{y_k : D_{ki}(y_k, s) < D_{kj}(y_k, s)\} \tag{115}$$

That is, Z_{kij}^s is the set of all y such that assigning y to codeword i results in a lower probability of error than if assigned to codeword j when in state s. Moreover,

$$\begin{aligned} &D_{ki}(y_k, s) < D_{kj}(y_k, s) \iff \\ &p_0(y_k|s)(A_{ki}^s - A_{kj}^s) + \\ &\quad p_1(y_k|s)(B_{ki}^s - B_{kj}^s) < 0 \iff \\ &(A_{ki}^s - A_{kj}^s) + L_s(y_k)(B_{ki}^s - B_{kj}^s) < 0 \end{aligned} \tag{116}$$

Hence, Z_{kij}^s is equivalent to

$$Z_{kij}^s(L) = \{L_s(y_k) : D_{ki}(y_k, s) < D_{kj}(y_k, s)\}. \tag{117}$$

That is, $Z_{kij}^s(L)$ is the set of LR values such that assigning observations with an LR value in the set to codeword i rather than j results in a smaller probability of error. Let

$$\tau_{kij}^s = -\frac{\alpha_{kij}^s}{\beta_{kij}^s}. \tag{118}$$

where

$$\alpha_{kij}^s = A_{ki}^s - A_{kj}^s \tag{119}$$

$$\beta_{kij}^s = B_{ki}^s - B_{kj}^s. \tag{120}$$

Then, we have several possibilities for $Z_{kij}^s(L)$ depending on the values of α_{kij}^s and β_{kij}^s.

$$Z_{kij}^s(L) = \begin{cases} [0, \tau_{kij}^s), & \beta_{kij}^s > 0 \\ [\tau_{kij}^s, \infty), & \beta_{kij}^s < 0 \\ [0, \infty), & \beta_{kij}^s = 0, \alpha_{kij}^s < 0 \\ \phi, & \beta_{kij}^s = 0, \alpha_{kij}^s \geq 0. \end{cases} \tag{121}$$

A few remarks are in order. First, if $Z_{kij}^s(L) = [0, \infty)$, it is always better to map the observation to codeword i instead of j. Likewise, if $Z_{kij}^s(L) = \phi$, it is never better to map the observation to codeword i instead of j. Hence, if $Z_{kij}^s(L) = [0, \infty)$ ($Z_{kij}^s(L) = \phi$), codeword j (i) should be deleted. Second, if we define

$$\tau_{ki}^{su} = \min_{j:\beta_{kij}^s > 0} \{\tau_{kij}^s\} \tag{122}$$

and

$$\tau_{ki}^{sl} = \max_{j:\beta_{kij}^s < 0} \{\tau_{kij}^s\} \tag{123}$$

then the set

$$Z_{si}^k(L) = \bigcap_{\substack{j=0 \\ j \neq i}}^{d-1} Z_{kij}^s(L)$$

which is the set of $L(y_k)$ that should be mapped to index i, can be expressed as

$$Z_{si}^k(L) = \begin{cases} \phi, & \beta_{kij}^s = 0, \alpha_{kij}^s \geq 0 \quad \text{for some } j \\ [0, \infty), & \beta_{kij}^s = 0, \alpha_{kij}^s < 0 \quad \text{for all } j \\ [\tau_{ki}^{sl}, \tau_{ki}^{su}), & \text{otherwise} \end{cases}$$

with the understanding that $[a, b) = \phi$ for $a > b$. Thus, if $\tau_{ki}^{sl} > \tau_{ki}^{su}$, $Z_{si}^k(L) = \phi$. Finally, it can be shown that for $j > i$, where neither code word i nor j has been deleted, we have $\beta_{kij}^s > 0$ and $\alpha_{kij}^s < 0$. Similarly, for $j < i$ we have $\beta_{kij}^s < 0$ and $\alpha_{kij}^s > 0$. We present the iterative algorithm below, and present some numerical results of this algorithm in Section VI.

REFERENCES

[1] R. R. Tenney and N. R. Sandell, "Detection with distributed sensors," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-17, no. 4, pp. 501–510, Jul. 1981.

[2] U. S. Shanthamallu, A. Spanias, C. Tepedelenlioglu, and M. Stanley, "A brief survey of machine learning methods and their sensor and iot applications," in *Proc. International Conference on Information, Intelligence, Systems & Applications (IISA)*, Aug. 2017, pp. 1–8.

[3] A. Tajer, S. Kar, H. V. Poor, and S. Cui, "Distributed joint cyber attack detection and state recovery in smart grids," in *2011 IEEE International Conference on Smart Grid Communications (SmartGridComm)*, Oct. 2011, pp. 202–207.

[4] A. Patel, H. Ram, A. K. Jagannatham, and P. K. Varshney, "Robust cooperative spectrum sensing for mimo cognitive radio networks under csi uncertainty," *IEEE Transactions on Signal Processing*, vol. 66, no. 1, pp. 18–33, 2018.

[5] A. Chawla, R. K. Singh, A. Patel, A. K. Jagannatham, and L. Hanzo, "Distributed detection for centralized and decentralized millimeter wave massive mimo sensor networks," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 8, pp. 7665–7680, 2021.

Algorithm 1: Iterative algorithm to find the optimal thresholds

1) Initialize the local rule for each sensor $\gamma_k^{(0)}$, $k = 1, 2, ...n$ and set $r = 1$ (iteration index)
2) Obtain the optimal fusion rule $\psi^{(r)}$ for fixed local decision rules $\boldsymbol{\gamma}^{(r-1)}$.
3) For each k, obtain the person-by-person optimal local decision rule $\gamma_k^{(r)}$ given the fixed rules for the other sensors using the following steps
   a) Initialize the state of sensor k, $s = 0$.
   b) For the state s, denote the current set of thresholds for sensor k as ν'_{ks}.
   c) Set $i = 0$ and $\tau_{k1}^{sl} = 0$.
   d) Compute τ_{kij}^{s} for $j = i + 1, ..., d - 1$ using (118).
   e) Compute τ_{ki}^{su} as (122) and obtain $j(i)$ such that $\tau_{ki}^{su} = \tau_{kij(i)}^{s}$.
   f) Delete the codewords $i + 1, i + 2, ..., j(i) - 1$ and set $\tau_{kj(i)}^{sl} = \tau_{ki}^{su}$.
   g) Set $i = j(i)$. If $i < d - 1$, go to step d). Otherwise, set $\tau_{ki}^{u} = \infty$ and set $s = s + 1$.
   h) if $s \leq b - 1$, go to step b). Otherwise, set $\gamma_k^{(r)} = \left\{ \nu_{k0}, \nu_{k1}, ..., \nu_{k(b-1)} \right\}$.
4) Check the difference (in Euclidean norm) between $\boldsymbol{\gamma}^{(r-1)}$ and $\boldsymbol{\gamma}^{(r)} = \left\{ \gamma_1^{(r)}, \gamma_2^{(r)}, ..., \gamma_n^{(r)} \right\}$. If it is less than a prescribed value, stop. Otherwise, set $r = r + 1$ and go to step 2.

[6] J. Engebrecht, K. Nealson, and M. Silverman, "Bacterial bioluminescence: isolation and genetic analysis of functions from vibrio fischeri," *Cell*, vol. 32, no. 3, pp. 773–781, Mar. 1983.

[7] J. Boedicker and K. Nealson, "Microbial communication via quorum sensing," *IEEE Transactions on Molecular, Biological, and Multi-Scale Communications*, vol. 1, no. 4, pp. 310–320, Dec. 2015.

[8] M. M. Vasconcelos, U. Mitra, O. Camara, K. P. Silva, and J. Boedicker, "Bacterial quorum sensing as a networked decision system," in *Proc. IEEE International Conference on Communications (ICC)*, May. 2018, pp. 1–6.

[9] N. Michelusi and U. Mitra, "Capacity of electron-based communication over bacterial cables: the full-csi case," *IEEE Transactions on Molecular, Biological and Multi-Scale Communications*, vol. 1, no. 1, pp. 62–75, 2015.

[10] N. Michelusi, J. Boedicker, M. Y. El-Naggar, and U. Mitra, "Queuing models for abstracting interactions in bacterial communities," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 3, pp. 584–599, 2016.

[11] M. M. Vasconcelos, U. Mitra, O. Câmara, M. S. Gangan, and J. Boedicker, "A continuous-time decision-making model for bacterial growth via quorum sensing: theory and evidence," in *Proc. ACM International Conference on Nanoscale Computing and Communication*, Sep. 2019, pp. 1–6.

[12] M. S. Gangan, M. M. Vasconcelos, U. Mitra, O. Câmara, and J. Boedicker, "Optimal strategy for public good production is set by balancing intertemporal trade-off between population growth rate and carrying capacity," *bioRxiv*, 2021. [Online]. Available: https://www.biorxiv.org/content/early/2021/07/22/2021.07.22.453373

[13] J. N. Tsitsiklis, "Decentralized detection," in *Advances in Statistical Signal Processing*. JAI Press, 1993, pp. 297–344.

[14] D. J. Warren and P. K. Willett, "Optimum quantization for detector fusion: some proofs, examples, and pathology," *Journal of The Franklin Institute-engineering and Applied Mathematics*, vol. 336, pp. 323–359, 1999.

[15] I. Y. Hoballah and P. K. Varshney, "Distributed bayesian signal detection," *IEEE Transactions on Information Theory*, vol. 35, no. 5, pp. 995–1000, Sep. 1989.

[16] J. N. Tsitsiklis, "Decentralized detection by a large number of sensors," *Mathematics of Control, Signals and Systems*, vol. 1, no. 2, pp. 167–182, Jun. 1988.

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

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

[19] Y. Lin and H. Chen, "Performance limit for distributed detection under correlated observations in wireless sensor networks," in *2012 19th IEEE Symposium on Communications and Vehicular Technology in the Benelux (SCVT)*, Nov. 2012, pp. 1–6.

[20] Q. Yan and R. Blum, "Distributed signal detection under the neyman-pearson criterion," *IEEE Transactions on Information Theory*, vol. 47, no. 4, pp. 1368–1377, May 2001.

[21] P.-N. Chen and A. Papamarcou, "Likelihood ratio partitions for distributed signal detection in correlated gaussian noise," in *Proceedings of 1995 IEEE International Symposium on Information Theory*, Sept. 1995, pp. 118–.

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

[23] A. Vempaty, H. He, B. Chen, and P. K. Varshney, "On quantizer design for distributed bayesian estimation in sensor networks," *IEEE Transactions on Signal Processing*, vol. 62, no. 20, pp. 5359–5369, Aug. 2014.

[24] B. Chen and P. Willett, "On the optimality of the likelihood-ratio test for local sensor decision rules in the presence of nonideal channels," *IEEE Transactions on Information Theory*, vol. 51, no. 2, pp. 693–699, Jan. 2005.

[25] J. Tsitsiklis, "Extremal properties of likelihood-ratio quantizers," *IEEE Transactions on Communications*, vol. 41, no. 4, pp. 550–558, 1993.

[26] H. Akaike, "Information theory and an extension of the maximum likelihood principle," in *Selected papers of hirotugu akaike*. Springer, 1998, pp. 199–213.

[27] J. Tsitsiklis and M. Athans, "On the complexity of decentralized decision making and detection problems," *IEEE Transactions on Automatic Control*, vol. 30, no. 5, pp. 440–446, May. 1985.

[28] V. Aalo and R. Viswanathou, "On distributed detection with correlated sensors: two examples," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 25, no. 3, pp. 414–421, 1989.

[29] E. Drakopoulos and C. C. Lee, "Optimum fusion of correlated local decisions," in *Proc. IEEE Conference on Decision and Control (CDC)*, Dec. 1988, pp. 2489–2494 vol.3.

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

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

[32] M. C. Vuran, Özgür B. Akan, and I. F. Akyildiz, "Spatio-temporal correlation: theory and applications for wireless sensor networks," *Computer Networks*, vol. 45, no. 3, pp. 245–259, 2004, in Memory of Olga Casals. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1389128604000519

[33] C. E. Shannon, "Channels with side information at the transmitter," *IBM journal of Research and Development*, vol. 2, no. 4, pp. 289–293, 1958.

[34] S. I. Gelfand and M. S. Pinsker, "Coding for channel with random parameters," *Probl. Contr. and Inf. Theory*, vol. 9, no. 1, pp. 19–31, 1980.

[35] C. Choudhuri, Y.-H. Kim, and U. Mitra, "Causal state communication," *IEEE transactions on information theory*, vol. 59, no. 6, pp. 3709–3719, 2013.

[36] S. I. Bross, "Causal asymmetric state communication over the degraded broadcast channel," in *2018 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2018, pp. 911–915.

[37] L. Miretti, M. Kobayashi, D. Gesbert, and P. De Kerret, "Cooperative multiple-access channels with distributed state information," *IEEE Transactions on Information Theory*, vol. 67, no. 8, pp. 5185–5199, 2021.

[38] J. Shaska and U. Mitra, "Decentralized decision making in multi-agent networks: the state-dependent case," in *2021 IEEE Global Communcations Conference*, Dec 2021.

[39] C. E. Shannon, R. G. Gallager, and E. R. Berlekamp, "Lower bounds to error probability for coding on discrete memoryless channels. I," *Information and Control*, vol. 10, no. 1, pp. 65–103, Jan. 1967.