Syracuse University
SUrface

Electrical Engineering and Computer Science | L.C. Smith College of Engineering and Computer Science

1-1-2012

A New Framework For Distributed Detection with Conditionally Dependent Observations

Hao Chen

Biao Chen
Syracuse University

Pramod Varshney
Syracuse University, varshney@syr.edu

Follow this and additional works at: http://surface.syr.edu/eecs

Part of the Electrical and Computer Engineering Commons

Recommended Citation
Chen, Hao; Chen, Biao; and Varshney, Pramod, "A New Framework For Distributed Detection with Conditionally Dependent Observations" (2012). *Electrical Engineering and Computer Science.* Paper 225.
http://surface.syr.edu/eecs/225

This Article is brought to you for free and open access by the L.C. Smith College of Engineering and Computer Science at SUrface. It has been accepted for inclusion in Electrical Engineering and Computer Science by an authorized administrator of SUrface. For more information, please contact surface@syr.edu.

A New Framework For Distributed Detection with Conditionally Dependent Observations

Hao Chen, *Member, IEEE*, Biao Chen, *Senior Member, IEEE* and Pramod K. Varshney, *Fellow, IEEE*

Abstract

Distributed detection with conditionally dependent observations is known to be a challenging problem in decentralized inference. This paper attempts to make progress on this problem by proposing a new framework for distributed detection that builds on a hierarchical conditional independence model. Through the introduction of a hidden variable that induces conditional independence among the sensor observations, the proposed model unifies distributed detection with dependent or independent observations. This new framework allows us to identify several classes of distributed detection problems with dependent observations whose optimal decision rules resemble the ones for the independent case. The new framework induces a decoupling effect on the forms of the optimal local decision rules for these problems, much in the same way as the conditionally independent case. This is in sharp contrast to the general dependent case where the coupling of the forms of local sensor decision rules often renders the problem intractable. Such decoupling enables the use of, for example, the person-by-person optimization approach to find optimal local decision rules. Two classical examples in distributed detection with dependent observations are reexamined under this new framework: detection of a deterministic signal in dependent noises and detection of a random signal in independent noises.

Copyright (c) 2011 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org. Hao Chen was with Syracuse University, Department of Electrical Engineering and Computer Science, Syracuse, NY 13244. He is now with Boise State University, Department of Electrical and Computer Engineering, Boise, ID 83725. Email: haochen@boisestate.edu. Biao Chen and Pramod K. Varshney are with Syracuse University, Department of Electrical Engineering and Computer Science, Syracuse, NY, 13244. Email: bichen{varshney}@syr.edu.

This material is based upon work supported by the Air Force Office of Scientific Research under award FA9550-10-1-0458 and FA9550-10-1-0263 and by the National Science Foundation under award 0925854. The material in this paper was presented in part at the IEEE International Symposium on Information Theory, Seoul, Korea, July 2009 and the IEEE International Conference on Acoustic, Speech, and Signal Processing, Dallas, March 2010.

Index Terms – Distributed Detection, Dependent Observations, Likelihood Quantizer

I. Introduction

Distributed inference refers to the decision making process involving multiple decentralized agents or sensors [1]. Development of the theory and methodologies for distributed inference was largely motivated by military surveillance applications in the early days [2]. Tremendous progress has been made in this area during the past few decades thanks to the collective effort of many researchers (see [3]–[6] and references therein).

Of particular interest in this paper is distributed detection, or distributed hypothesis testing. Figure 1 is an illustration of a canonical distributed detection problem where the objective is to determine at the fusion center the underlying hypothesis that drives the sensor observations. Different from a centralized system, the sensor observations are compressed prior to being used by the fusion center in determining the underlying hypothesis. This is typically a result of various system constraints, e.g., the communication between the sensors and the fusion center may be severely bandlimited. The design of a distributed detection system thus involves the design of the fusion rule $\gamma_0(\cdot)$ as well as the decision rules at local sensors $\gamma_1(\cdot), \cdots, \gamma_n(\cdot)$. While the optimal fusion rule is known to be a likelihood-ratio test (LRT) at the fusion center [7]–[9], designing decision rules at decentralized nodes is much more complicated because of their distributed nature. For the most general problem, the optimal sensor decision rule design problem has been shown to be an NP-complete problem [10]. On the other hand, if the local sensor observations are conditionally independent given the hypothesis, the design of local decision rules simplifies substantially: likelihood quantizers have been shown to be optimal for such cases under various inference regimes [4], [11]. Therefore, the decision rule design problem reduces to that of finding the quantizer thresholds for which the person-by-person optimization approach can be used to efficiently search for those thresholds [3]. In the simple case of a binary hypotheses testing where each sensor sends a single bit to the fusion center, the optimal decision rule at each sensor is simply an LRT and the remaining task is to find the LRT threshold for each sensor. The optimality of likelihood quantizers for the distributed hypothesis testing problem has since been extended to more complicated cases where the sensor outputs are to be communicated through noisy, possibly coupled channels to the fusion center [12]–[14]. These extensions are largely motivated by the emerging wireless sensor networks [15].

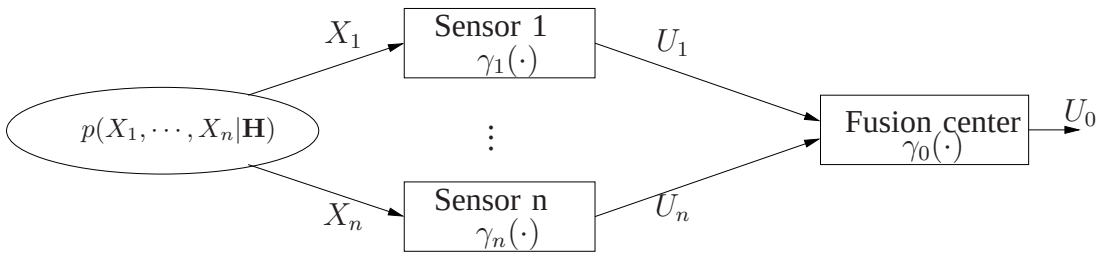

Fig. 1. A canonical distributed inference system. The kth local sensor observes X_k and sends its output $U_k = \gamma_k(X_k)$ to the fusion center. The fusion center makes a final decision U_0 regarding θ using the fusion rule $\gamma_0(U_1, \cdots, U_n)$.

In the absence of the conditional independence (CI) assumption, however, the problem of designing the optimal local decision rules becomes much more challenging. In such a case, the *form* of the optimal decision rule at a sensor is often unknown and is coupled with other sensor decision rules and the fusion rule. Even for the binary hypotheses testing problem with binary sensor output, LRTs at local sensors are often not optimal [16], [17]. The significant complexity of the dependent observation case is most clearly demonstrated in a binary detection problem with two sensors observing a shift in the mean of correlated Gaussian random variables [18], [19]. For this relatively simple problem, while the optimality of LRT can be established for certain parameter regions, the problem becomes largely intractable in other regions. In general, existing results for the dependent case are somewhat fragmented; only in some isolated cases do we have a good understanding of the optimal decision structure.

The difficulty of dealing with dependent observations can be somewhat alleviated in the large sample size regime in which the goal is to characterize or optimize error exponents of detection performance instead of the actual error probabilities. This line of work was first proposed by Berger [20] who formulated the multiterminal hypothesis testing problem under rate constraints. Significant progress has since been made on this problem (see [21] and references therein); however, most concrete results are often derived for some specialized settings, e.g., distributed test against independence under one-sided rate constraint [22]. There have also been recent works in the use of large deviation theory for distributed detection with correlated observations (see, e.g., [23]). There, the emphasis is typically not on local decision rule design but instead on the analysis of asymptotic performance as a function of sensor correlation or network size/topology. Such analysis is often feasible with simplifying assumption about local sensor decision rules, e.g., the data are often assumed to be propagated without any compression.

In this paper, we attempt to make progress in our understanding of the dependent cases in the finite sample regime. Toward this end, we propose a unifying model under the Bayesian inference framework that includes both conditionally independent and dependent observations as its special cases. Specifically, we expand the physical hierarchical structure of the distributed detection problem by introducing a "hidden" random variable. This hidden variable induces conditional independence of the sensor observations with respect to this new random variable even if the original observations are conditionally dependent given the underlying hypothesis. This new model allows for intuitive explanation of some of the known results for distributed detection with dependent observations. More importantly, it provides a powerful framework to identify broader classes of problems with dependent observations whose optimal sensor decision rules can be drastically simplified. Here, the use of optimality is rather generic and most discussions about optimality are equally applicable to both the Bayesian and the Neyman-Pearson frameworks. As this paper focuses on the Bayesian framework, the optimality here refers specifically to the minimization of the Bayesian cost at the fusion center. Preliminary results have been reported in [24], [25]. In addition to expanding on the technical details of the proposed framework, the current paper include treatment of two classical problems in distributed detection with dependent observations: the detection of a deterministic signal in dependent noises and the detection of a random signal in independent noises. The ability to deal with both problems demonstrate the power of the proposed framework.

Analogous approaches have been used in resolving some of the classical problems in multi-terminal data compression involving dependent observations. Ozarow's work in finding the rate-distortion region of the multiple description problem for a bivariate Gaussian source relies on the use of an 'artificial' random variable that induces a conditional independence of the two Gaussian random variables that are otherwise correlated [26]. Similarly, the rate distortion region for the two terminal Gaussian source coding problem [27] hinges on the introduction of an auxiliary Gaussian variable that induces a conditional independence structure, thereby allowing the coupling of the two terminal Gaussian source coding problem with the quadratic Gaussian CEO problem [28] whose rate distortion region was known [29].

The rest of this paper is organized as follows. In the next section, we review the formulation and basic results for the M-ary distributed detection problem. The unifying system model for distributed detection is proposed in Section III. In Sections IV and V, we identify two classes of distributed detection

problems with dependent observations whose optimal local decision rules are reminiscent in structure of the conditional independent case. In particular, we illustrate through two examples in Section V that the proposed framework provides a meaningful approach to tackle two classical distributed detection problems with dependent observations: 1) detection of deterministic signals in dependent noises and 2) detection of random signals in independent noises. We conclude in Section VI.

Notation: Throughout this paper, we use $p(x)$ to denote either point mass function (pmf) of a discrete random variable X or probability density function (pdf) of a continuous random variable X. Similarly, $p(x, y)$ and $p(x|y)$ denote respectively the joint or conditional pmf and pdf of random variables X and Y. Boldface capital letters (e.g., $\mathbf{X}$, $\mathbf{U}$) denote vectors of random variables while boldface lower case letters denote realizations of a random vector. Additionally, $\mathbf{X}^k$ denotes the vector $\mathbf{X}\backslash X_k = [X_1, X_2, \cdots, X_{k-1}, X_{k+1}, \cdots X_K]$, i.e., the $\mathbf{X}$ vector except for the kth term.

II. Multiple Hypotheses Testing in Distributed Multi-Sensor Systems

Consider a canonical parallel distributed hypothesis testing system with K sensors, as illustrated in Figure 1 and described below.

- M-ary hypothesis testing: $H \in \{0, 1, \cdots, M-1\}$ with prior probability π_H.
- Local sensor observations X_k, $k = 1, 2, \cdots, K$.
- Local sensor output $U_k = \gamma_k(X_k) \in \{0, \cdots, L-1\}$, $k = 1, 2, \cdots, K$.
- Fusion center output $U_0 \in \{0, 1, \cdots, M-1\}$.

Let $c_{u_0,h}$ be the Bayesian cost of deciding $U_0 = u_0$ when $H = h$ is true. For the special case of minimizing the probability of error, $c_{u_0,h}$ takes the form of $0-1$ cost, i.e., $c_{u_0,h} = 0$ when $u_0 = h$ and 1 otherwise. The expected Bayesian cost C that needs to be minimized for this M-ary hypothesis testing problem is given as

$$\begin{aligned} C &= \sum_{u_0=0}^{M-1} \sum_{h=0}^{M-1} c_{u_0,h} p(u_0, h) \\ &= \sum_{u_0=0}^{M-1} \sum_{h=0}^{M-1} c_{u_0,h} \pi_h p(u_0|h) \end{aligned}$$

From the hierarchical structure of the canonical fusion model, $H \rightarrow \mathbf{X} \rightarrow \mathbf{U} \rightarrow U_0$ forms a Markov chain. Therefore, expanding C with respect to sensor k, we have

$$\begin{aligned} C &= \int_{\mathbf{X}} \sum_{\mathbf{u}} \sum_{u_0=0}^{M-1} \sum_{h=0}^{M-1} c_{u_0,h} \pi_h p(u_0|\mathbf{u}) p(\mathbf{u}|\mathbf{x}) p(\mathbf{x}|h) d\mathbf{x} \\ &= \int_{X_k} \sum_{u_k} p(u_k|x_k) f_k(u_k, x_k) dx_k \end{aligned} \tag{1}$$

where

$$\begin{aligned} f_k(u_k, x_k) &\triangleq \sum_{\mathbf{u}^k} \sum_{u_0=0}^{M-1} \sum_{h=0}^{M-1} c_{u_0,h} p(x_k|h) \pi_h p(u_0|\mathbf{u}^k, u_k) \int_{\mathbf{X}^k} p(\mathbf{u}^k|\mathbf{x}^k) p(\mathbf{x}^k|h, x_k) d\mathbf{x}^k \\ &= \sum_{u_0=0}^{M-1} \sum_{h=0}^{M-1} c_{u_0,h} \pi_h p(u_0|u_k, x_k, h) \end{aligned} \tag{2}$$

is defined as the Bayesian cost density function (BCDF) for the kth sensor making decision u_k while observing x_k and

$$p(u_0|u_k, x_k, h) = \sum_{\mathbf{u}^k} p(u_0|\mathbf{u}^k, u_k) \int_{\mathbf{X}^k} p(\mathbf{u}^k|\mathbf{x}^k) p(\mathbf{x}^k|h, x_k) d\mathbf{x}^k. \tag{3}$$

While we use integration in Equation (1) which implicitly assumes that the sensor observations are continuous random variables, the expected Bayesian cost is similarly defined for discrete $\mathbf{X}$ where integration is replaced with summation. From Equation (1), to minimize the expected Bayesian cost C, the optimal kth sensor decision rule given fixed decision rules at all other sensors and the fusion center is to make a decision u_k such that $f_k(u_k, x_k)$ is minimized, that is

$$U_k = \gamma_k(X_k) = \arg\min_{u_k} f_k(u_k, X_k), \tag{4}$$

for all X_k except for the set D_k with $P(X_k \in D_k) = 0$.

The BCDF $f_k(U_k, X_k)$ in Equations (2), and consequently the optimal decision rule at the kth sensor, is coupled with the fusion rule $\gamma_0(\cdot)$ and other sensor decision rules $\gamma_i(\cdot)$, $i \neq k$. This coupling is what makes the problem of finding the optimal set of $\gamma_k(X_k)$ difficult since changes in other sensors' decision rule and/or the fusion rule may result in a change both in the *form* and associated parameters of $f_k(U_k, X_k)$.

A. The Conditional Independence Case

The complexity of the optimization problem reduces significantly when the sensor observations are independent conditioned on the underlying hypothesis H, i.e.,

$$p(x_1, x_2, \cdots, x_K|H) = \prod_{k=1}^{K} p(x_k|H). \tag{5}$$

For this conditional independence model, the BCDF $f_k(U_k, X_k)$ in Equation (2) becomes

$$f_k^I(u_k, x_k) = \sum_{h=0}^{M-1} \alpha_k(u_k, h) p(x_k|h), \tag{6}$$

where

$$\begin{aligned} \alpha_k(u_k, h) &\triangleq \sum_{u_0=0}^{M-1} c_{u_0,h} \pi_h p(u_0|u_k, h) \\ &= \sum_{\mathbf{u}^k} \sum_{u_0=0}^{M-1} c_{u_0,h} \pi_h p(u_0|\mathbf{u}^k, u_k) \int_{\mathbf{X}^k} p(\mathbf{u}^k|\mathbf{x}^k) p(\mathbf{x}^k|h) d\mathbf{x}^k \end{aligned} \tag{7}$$

is a scalar function of the sensor decision $U_k = u_k$ and the underlying hypothesis $H = h$. Here the superscript "I" indicates that the BCDF is obtained under the conditional independence assumption. The optimal kth sensor decision rule $\gamma_k^I(X_k)$ becomes

$$U_k = \gamma_k^I(X_k) = \arg\min_{u_k} \sum_{h=0}^{M-1} \alpha_k(u_k, h) p(X_k|H = h), \tag{8}$$

for all X_k except for the set D_k with $P(X_k \in D_k) = 0$. Equation (8) is in essence of the same form as the M-ary hypotheses Bayesian detection problem [30], i.e., γ_k^I is an optimal multiple hypotheses Bayesian test with M hypotheses and L possible decisions with Bayesian cost coefficients $\alpha_k(U_k, H)$. Compared with the BCDF given in Equation (2) where even the forms of the optimal sensor decision rules are unknown, the optimization problem reduces to determining suitable values of the scalars $\alpha_k(U_k, H)$ under the CI assumption. For binary hypothesis testing ($M = 2$ and $H \in \{0, 1\}$) with binary output at sensors ($L = 2$ and $U_k \in \{0, 1\}$), $\gamma_k(X_k)$ can be further simplified as

$$\gamma_k^I(X_k) = \begin{cases} 0, & \text{if } \tilde{\alpha}_k(1,1) \frac{p(X_k|H=1)}{p(X_k|H=0)} + \tilde{\alpha}_k(1,0) > 0 \\ 1, & \text{otherwise,} \end{cases}, \tag{9}$$

where $\tilde{\alpha}_k(u_k, h) = \alpha_k(u_k, h) - \alpha_k(0, h)$, i.e., it is a local LRT with a suitable threshold.

III. A Hierarchical Conditional Independence Model for Distributed Detection

Consider again the distributed detection system as shown in Fig. 1, where the variables involved follow the following Markov chain

$$H \to \mathbf{X} \to \mathbf{U} \to U_0. \tag{10}$$

Whether or not conditional independence holds depends on how the joint distribution $p(\mathbf{X}|H)$ factorizes, i.e., whether or not Equation (5) is satisfied. The proposed framework hinges on the introduction of a "hidden" random variable $\mathbf{Y}$ into this Markov chain, such that

1) the following Markov chain holds

$$H \to \mathbf{Y} \to \mathbf{X} \to \mathbf{U} \to U_0. \tag{11}$$

2) $X_1, X_2, \cdots, X_K$ are independent conditioned on $\mathbf{Y}$, i.e.,

$$p(x_1, x_2, \cdots, x_K|\mathbf{y}) = \prod_{i=1}^{K} p(x_k|\mathbf{y}). \tag{12}$$

The injection of the hidden variable $\mathbf{Y}$ induces conditional independence of the sensor observations with respect to this new variable regardless of the dependence structure of the original model. We refer to this new model as the hierarchical (or hidden) conditional independence (HCI) model. Although it may appear that the proposed HCI model (11) is less general than the traditional model (10), they are in fact equivalent.

Lemma 1: Any general distributed inference model in Fig. 1 and Equation (10) can be represented as a HCI model and vice versa.

Proof: Any HCI model is naturally a general distributed detection model with

$$p(x_1, x_2, \cdots, x_K|H) = \sum_{Y} p(x_1, x_2, \cdots, x_K|\mathbf{y})p(\mathbf{y}|H)$$

To prove the other direction, consider two cases. If $X_1, X_2, \cdots, X_K$ are conditionally independent given H, set $\mathbf{Y} = H$. Otherwise, let $\mathbf{Y} = \mathbf{X}$. □

Compared with the traditional model, the HCI model (11) is more flexible and provides a unified framework for analyzing distributed detection problems under various dependence assumptions.

Depending on the support of $\mathbf{Y}$, we further classify the HCI model into three categories:

1) "Discrete" HCI (DHCI) model. In this case, $\mathbf{Y}$ is a discrete random variable or vector.

2) "Continuous" HCI (CHCI) model. In this case, $\mathbf{Y}$ is either a continuous random variable or vector.
3) "Hybrid" HCI (HHCI) model. In this case, $\mathbf{Y}$ is composed of both discrete and continuous random variables or vectors.

Notice that the discrete HCI model includes cases where $\mathbf{Y}$ can take finite or infinite values. It will become apparent that the result presented in the next section requires $\mathbf{Y}$ to be of finite alphabet. For ease of notation, we still refer to the case where Y is finite as simply discrete HCI.

One subtle difference between the CI and HCI models lies in the asymptotic detection performance when the number of sensors increases. For the CI model, each sensor obtains a conditionally independent "noisy" observation of H. Therefore, as long as the Chernoff distances between any two distributions corresponding two hypotheses are not equal to 0, the probability of making a wrong decision decays exponentially as the number of sensors K increases [31]. For the HCI model, however, each sensor obtains a conditionally independent "noisy" observations of $\mathbf{Y}$ instead of H. As a result, although the knowledge of $\mathbf{Y}$ can be improved as K increases, the distributed detection performance of the entire system is limited by the clairvoyant case where $\mathbf{Y}$ is directly observable (c.f. the Markov chain (11)). That is, the probability of making a wrong decision is always lower bounded by the probability of error assuming direct knowledge of $\mathbf{Y}$, regardless of the size of the sensor network K. This, of course, does not preclude the case where the error probability may still decrease exponentially with K, e.g., in the case of conditional independence for which one can set $\mathbf{Y} = H$.

In the next two sections, we study two classes of the HCI model: Section IV deals with the DHCI model where we further assume a finite alphabet support for $\mathbf{Y}$ while Section V considers the CHCI model under additional conditions on the distributions of observations as well as the fusion rule. As it turns out, both these models admit local sensor decision rules that are reminiscent of those for the conditional independence case.

IV. Optimum Sensor Decision Rule Design for the DHCI Model

Without loss of generality, let $Y \in \{0, 1, \cdots, N-1\}$ where N is the size of its support. Note that we use Y instead of $\mathbf{Y}$ in this section since $\mathbf{Y}$ is now a scalar random variable. Under this setting, we have

$$p(x_k|H) \quad = \quad \sum_{y=0}^{N-1} p(x_k, y|H)$$

$$= \sum_{y=0}^{N-1} p(y|H)p(x_k|y), \tag{13}$$

and

$$\begin{aligned} p(\mathbf{x}|H) &= \sum_{y=0}^{N-1} p(\mathbf{x}, y|H) \\ &= \sum_{y=0}^{N-1} p(y|H) \prod_{k=1}^{K} p(x_k|y). \end{aligned} \tag{14}$$

For the general DHCI case, $p(\mathbf{x}|H)$ and $p(x_k|H)$ defined above do not necessarily satisfy Equation (5). Substitute Equation (14) into (2), the BCDF $f_k(u_k, x_k)$ can be simplified as follows.

$$f_k^D(u_k, x_k) = \sum_{y=0}^{N-1} \beta_k(u_k, y)p(x_k|y), \tag{15}$$

where

$$\begin{aligned} \beta_k(u_k, y) &\equiv \sum_{u_0=0}^{M-1}\sum_{h=0}^{M-1} c_{u_0,h} p(u_0|u_k, y)p(y|h) \\ &= \sum_{\mathbf{u}^k}\sum_{u_0=0}^{M-1}\sum_{h=0}^{M-1} c_{u_0,h}\pi_h p(y|h)p(u_0|\mathbf{u}^k, u_k) \int_{\mathbf{X}^k} p(\mathbf{u}^k|\mathbf{x}^k)p(\mathbf{x}^k|y)d\mathbf{x}^k \end{aligned} \tag{16}$$

is a scalar function of u_k and y. The superscript "D" indicates that the BCDF is obtained under the DHCI model.

To minimize the expected Bayesian cost

$$C = \int_{X_k} \sum_{u_k=0}^{M-1} P(U_k = u_k|X_k = x_k) f_k^D(u_k, x_k)dx_k,$$

the optimal kth sensor rule γ_k^D becomes

$$U_k = \gamma_k^D(X_k) = \arg\min_{u_k} \sum_{y=0}^{N-1} \beta_k(u_k, y)p(X_k|Y = y), \tag{17}$$

for all X_k except the set D_k where $P(X_k \in D_k) = 0$.

Comparing Equations (17) and (8), one can observe the similarity between the DHCI case and the CI case, i.e., $\gamma_k^D(X_k)$ under the DHCI model is an optimum multiple hypotheses Bayesian test of N hypotheses and L decisions with the cost coefficients $\beta_k(u_k, y)$ where $k = 1, 2, \cdots, K$, $u_k = 0, \cdots, L-1$ and $y = 0, 1, \cdots, N-1$. Thus, $f_k^D(u_k, x_k)$ has a similar form as $f_k^I(u_k, x_k)$ if we replace H with Y (with respective cardinalities M and N).

For example, when $M = 2$ and $L = 2$, dividing $\tilde{f}_k^I(\cdot, X_k)$ by $p(X_k|Y = 0)$, $\gamma_k^D(X_k)$ can be further simplified as

$$\gamma_k^D(X_k) = \begin{cases} 0, & \text{if } \sum_{y=0}^{N-1} \tilde{\beta}_k(1, y) \frac{p(X_k|Y=y)}{p(X_k|Y=0)} > 0, \\ 1, & \text{otherwise}, \end{cases} \tag{18}$$

where $\tilde{\beta}_k(U_k, y) = \beta_k(U_k, y) - \beta_k(0, y)$, i.e., $\gamma_k^D(X_k)$ is a local LR quantizer except that the likelihood function is defined with respect to Y instead of the original hypothesis H. In the case of $N = 2$, Equation (18) reduces to a LRT, again, with the likelihood function defined with respect to the hidden variable Y.

A. Comparison between CI and DHCI

We now examine more closely the similarities and differences between the CI and DHCI models and the resulting decision rules. First of all, for the DHCI model, let us determine the cases for which the optimal decision rules at local sensors are indeed likelihood ratio quantizers where the likelihood functions are defined with respect to the hypothesis under test.

Under the DHCI model, replacing $p(x_k|H)$ in Equation (6) with (13), the BCDF $f_k^I(u_k, x_k)$ obtained in the CI model becomes

$$f_k^I(u_k, x_k) \quad = \quad \sum_{y=0}^{N-1} p(x_k|y) \left(\sum_{h=0}^{M-1} \alpha_k(u_k, h) p(y|h) \right). \tag{19}$$

Let $P_Y^H = [P(y|h)]_{0 \le y \le N-1, 0 \le h \le M-1}$ be the $N \times M$ probability transition matrix between H and Y,

$$B_k = [\beta(U_k, 0), \beta(U_k, 1), \cdots, \beta(U_k, N-1)]^T,$$

and

$$A_k = [\alpha(U_k, 0), \alpha(U_k, 1), \cdots, \alpha(U_k, M-1)]^T.$$

For any $f_k^I(u_k, x_k)$ with coefficients $\alpha(u_k, 0), \alpha(u_k, 1), \cdots, \alpha(u_k, M-1)$, we can always find, for a given Y, a corresponding $f_k^D(u_k, x_k)$ with coefficients $\beta(u_k, y)$ where

$$B_k = P_Y^H A_k, \tag{20}$$

such that they are equivalent. The converse, however, is not always true as we may not be able to find a set A_k satisfying Equation (20) for any given B_k.

However, if P_Y^H is a full row rank matrix, i.e., Rank$(P_Y^H) = N \leq M$, then, for any arbitrary B_k, there exists at least one A_k such that Equation (20) holds [32]. That is, under this condition, for any $f_k^D(u_k, x_k)$ with coefficients B_k, we can find a corresponding $f_k^I(u_k, x_k)$ with coefficients A_k such that the resulting decision functions are the same. Therefore, when Rank$(P_Y^H) = N \leq M$, the optimal detection performance can also be achieved by fixing the form of local sensor decision rules as Equation (8) and selecting an optimal set of parameters. In other words, the optimal decision rules take the form of likelihood quantizers where the likelihood function is defined with respect to the original hypothesis under test.

B. An Example

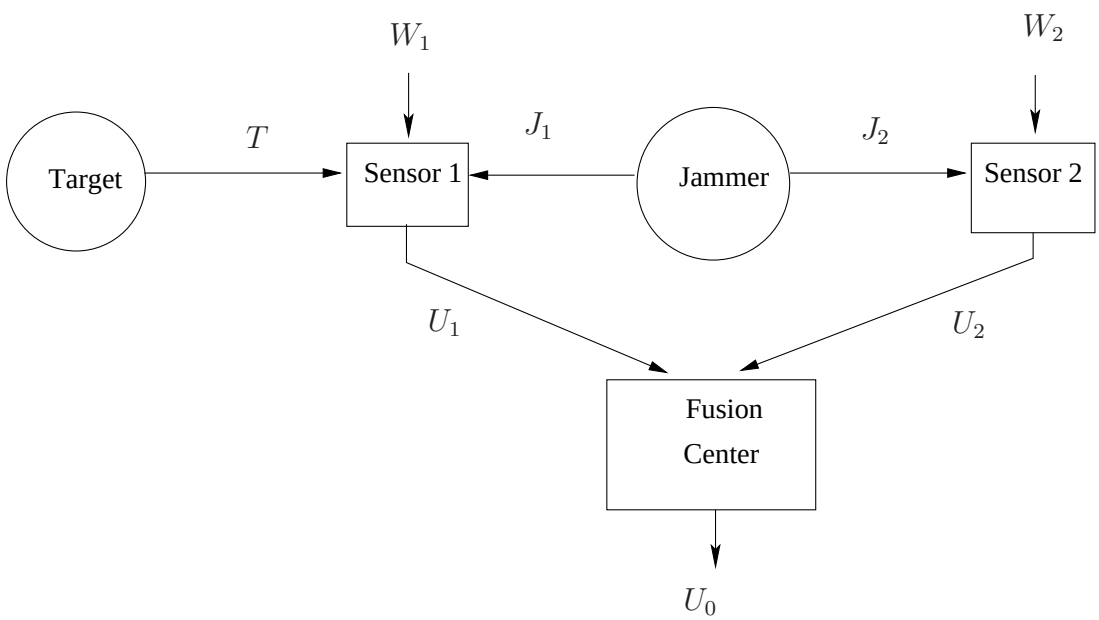

Fig. 2. A DHCI example with two sensors, one target and one jammer.

Consider the problem of target detection in the presence of a possible jammer using two sensors. Sensor 1 is placed between the jammer and the target, while sensor 2 is placed far away from the target but close to the possible jammer. Assume that depending on whether the target and/or the jammer are present, the

received signals at the two sensors are respectively:

$$X_1 = \begin{cases} T + J_1 + W_1 & \text{both target and jammer are present;} \\ T + W_1 & \text{only target is present;} \\ J_1 + W_1 & \text{only jammer is present;} \\ W_1 & \text{neither target nor jammer is present;} \end{cases}$$

$$X_2 = \begin{cases} J_2 + W_2 & \text{jammer is present;} \\ W_2 & \text{jammer is absent.} \end{cases}$$

where X_1 and X_2 are the respective received signals at sensors 1 and 2, T is the received target signal at sensor 1, J_1 and J_2 are jammer signals observed at the two sensors, and W_1 and W_2 are the noises at the two sensors and are independent of each other and of the target and jammer signals. Notice that in the above model, the received signal of sensor 2 is independent of whether the target is present or not as it is far away from the target location. Each sensor makes a binary decision independently and sends it to a fusion center where the final decision is made. This is illustrated in Figure 2.

Let $H \in \{0, 1\}$ represent the hypotheses of target absent and present respectively. We also denote by $J \in \{0, 1\}$ the random variable that represents the absence or presence of the jammer. Assume further that both H and J have equally likely prior probabilities. Notice that due to the possible presence of the jammer, the signals at the two sensors are no longer independent conditioned on the hypothesis under test. However, given both H and J, X_1 and X_2 are independent of each other due to the assumption of the independence of observation noises.

To illustrate the DHCI approach and to compare it with the one that assumes the CI model, let us consider the high signal to noise ratio (SNR) case with the additional simplifying assumption that T, J_1, and J_2 are all of equal power, say, P. Thus, the received signal power at sensors 1 and 2 are

$$P_1 \approx \begin{cases} 2P & \text{both target and jammer are present;} \\ P & \text{either target or jammer is present;} \\ 0 & \text{neither target nor jammer is present;} \end{cases}$$

$$P_2 \approx \begin{cases} P & \text{jammer is present;} \\ 0 & \text{jammer is absent.} \end{cases}$$

Notice that in the absence of any additional information regarding target and jamming signals, any meaningful detection schemes will be based solely on the received signal power. If we assume a conditional independence model, then since sensor 2's signal is independent of H, the fusion output should rely on sensor 1 output only. For sensor 1, the optimal binary decision rule that minimizes the error probability is to declare $U_1 = 1$ if $P_1 = 2P$, $U_1 = 0$ if $P_1 = 0$. If $P_1 = P$, U_1 can be set at either 1 or 0 without affecting the error probability. The optimal fusion rule would be simply to set $U_0 = U_1$, achieving an error probability of 0.25.

However, it is easy to conceive a simple scheme that achieves perfect detection: Sensor 2 first makes a binary decision informing the fusion center about the state of the jammer, while sensor 1 implements a non-monotone quantizer: $U_1 = 1$ if $P_1 = 0$ or $2P$, and $U_1 = 0$ if $P_1 = P$. This simple scheme results in a vanishing error probability as SNR grows.

Not surprisingly, such decision rules are exactly what would result if we follow the underlying DHCI model with the hidden variable $\mathbf{Y} = \{H, J\}$. Specifically, by setting $Y = \{H, J\}$ which induces conditional independence of the sensor observations, it is straightforward, *albeit* tedious to verify from (17) that for certain signal parameters, sensor 1 should implement a two threshold quantizer rule whereas for sensor 2 the decision rule is always a single threshold quantizer. As the signal to noise ratio tends to infinity, the decision rule for sensor 1 converges precisely to the non-monotone quantizer described above.

V. Optimum Sensor Rule Design for the CHCI Model

In this section, we consider the optimum design problem for the CHCI model where $\mathbf{Y}$ is a continuous scalar random variable. Analogous to the DHCI model, by replacing summation with integration in Equation (21), the BCDF under the CHCI model is given by

$$f_k^C(u_k, x_k) = \int_y \beta_k(u_k, y) p(x_k|y) dy, \tag{21}$$

where $\beta_k(u_k, y)$ is similarly defined as in (16) except that $p(\cdot)$ now denotes pdf instead of pmf. The superscript "C" indicates that the BCDF is obtained under the CHCI model.

To minimize the expected Bayesian cost C, given that all other rules are fixed, the optimal kth local sensor rule becomes

$$U_k = \gamma_k^C(X_k) = \arg\min_{u_k} f_k^C(U_k, X_k) \tag{22}$$

for all X_k except for the set D_k where $P(X_k \in D_k) = 0$.

Unlike the BCDF $f^D(u_k, x_k)$ for the DHCI model, $f_k^C(u_k, x_k)$ *cannot* be described completely by a set of finite parameters. Thus, unlike the optimal design problem under the CI or DHCI model, it is often not possible to solve the optimum sensor rule design problem solely by determining proper values of a set of parameters. However, compared to the BCDF form of the traditional model given by Equation (2), $f_k^C(u_k, x_k)$ has a much simpler structure based on the hidden random variable Y and provides a better insight for the design problem. In fact, by imposing some additional constraints on the distributions $p(x_k|y)$ and $p(y|H)$, the optimal local sensor decision rules can be determined by exploring certain property of $f_k^C(u_k, x_k)$. In the next few subsections, we present some concrete results for the CHCI model. The power of the proposed framework is highlighted by reexamining the binary detection problem with two sensors observing a shift in the mean of correlated Gaussian random variables [18], [19]. While the optimal decision rules are already known for some given parameter regimes, the new approach of solving this problem is more potent and is broadly applicable to other cases.

A. Distributed Binary Hypotheses Testing with Monotone Likelihood Ratios

Recall that a family of densities, parameterized by θ, $p_\theta(x)$, is said to have *monotone likelihood ratios* in $T(x)$ if there exists a real-valued function $T(x)$ such that for any $\theta < \theta'$ the distributions p_θ and $p_{\theta'}$ are distinct and $p_{\theta'}(T(x))/p_\theta(T(x))$ is a nondecreasing function of $T(x)$ [33]. Moreover, the distribution is said to have monotone likelihood ratios in its observation if $T(x) = x$ satisfies the above condition. The monotonicity utilized in this section can be considered to be a variation of the above classical definition. To ease our notation and presentation, we focus in this subsection on binary hypothesis testing with binary sensor outputs, i.e., $M = 2$, $L = 2$ and $U_k \in \{0, 1\}$. Both sensor observations as well as the "hidden" Y are scalar random variables. Also, the fusion rule is assumed to be monotonic such that

$$P(U_0 = 1|U_k = 1, y) \geq P(U_0 = 1|U_k = 0, y), \tag{23}$$

for all $k = 1, 2, \cdots, K$ and all y. Notice that

$$P(U_0 = 1|U_k = u_k, y) = \sum_{\mathbf{u}^k} P(U_0 = 1|U_k = u_k, \mathbf{u}^k) p(\mathbf{u}^k|y), \tag{24}$$

and this monotonicity definition is a slight deviation from the traditional definition of a monotonic fusion rule where $P(U_0 = 1|U_k = 1, \mathbf{u}^k) \geq P(U_0 = 1|U_k = 0, \mathbf{u}^k)$ is required for all possible $\mathbf{U}^k$ [3].

Without loss of generality, for this binary hypothesis testing problem, we assume the 0-1 Bayesian cost, resulting in the expected Bayesian cost to be the error probability P_e. The following proposition establishes the optimality of single threshold quantizers at the local sensors under suitable conditions.

Proposition 1: Consider a distributed binary hypothesis testing system with scalar sensor observations and binary sensor outputs. Suppose that the distributed hypothesis testing problem can be described equivalently by the CHCI model where the "hidden" random variable Y is a scalar random variable. Furthermore,

1) The fusion center implements a monotone fusion rule that satisfies (23);
2) The ratio $g(y) \triangleq \frac{p(y|H=1)}{p(y|H=0)}$ is a nondecreasing function of y;
3) The ratio $h(y; x_k, x'_k) \triangleq \frac{p(x_k|y)}{p(x'_k|y)}$ is also a nondecreasing function of y for any $x_k > x'_k$.

Then there exists a single threshold quantizer at sensor k, i.e.,

$$U_k = \begin{cases} 1, & \text{if } X_k \geq \tau_k, \\ 0, & \text{if } X_k < \tau_k, \end{cases} \tag{25}$$

for some suitable τ_k, that minimizes the error probability P_e.

Proof: As the expected Bayesian cost C is the error probability P_e, we have $c_{11} = c_{00} = 0$ and $c_{10} = c_{01} = 1$. The coefficient $\beta(U_k, Y)$ becomes

$$\begin{aligned} \beta(u_k, y) &= \pi_0 P(U_0 = 1|u_k, y) p(y|H = 0) + \pi_1 P(U_0 = 0|u_k, y) p(y|H = 1) \\ &= (\pi_0 p(y|H = 0) - \pi_1 p(y|H = 1)) P(U_0 = 1|u_k, y) + \pi_1 p(y|H = 1) \end{aligned} \tag{26}$$

Thus,

$$\begin{aligned} \tilde{f}_k^C(1, x_k) &\triangleq f_k^C(1, x_k) - f_k^C(0, x_k) \\ &= \int_Y (\beta(1, y) - \beta(0, y)) p(x_k|y) dy \\ &= \int_Y p(x_k|y)(P(U_0 = 1|U_k = 1, y) - P(U_0 = 1|U_k = 0, y)) \end{aligned} \tag{27}$$

$$
\begin{aligned}
& (\pi_0 p(y|H=0) - \pi_1 p(y|H=1))dy \\
= & -\int_Y p(x_k|y)\phi(y)dy \qquad (28)
\end{aligned}
$$

where

$$
\begin{aligned}
\phi(y) &\triangleq (P(U_0=1|U_k=1,y) - P(U_0=1|U_k=0,y))\,(\pi_1 p(y|H=1) - \pi_0 p(y|H=0)) \\
&= (P(U_0=1|U_k=1,y) - P(U_0=1|U_k=0,y))\pi_1 p(y|H=0)\left(\frac{p(y|H=1)}{p(y|H=0)} - \frac{\pi_0}{\pi_1}\right). \qquad (29)
\end{aligned}
$$

Since $p(y|H=1)/p(y|H=0)$ is nondecreasing in y and $P(U_0=1|U_k=1,Y=y) \geq P(U_0=1|U_k=0,Y=y)$, $\phi(y)$ is a function with a single change of sign. In other words there exists a value $-\infty \leq \tau_{k,y} \leq +\infty$ such that $\phi(y) \geq 0$ when $y \geq \tau_{k,y}$ and $\phi(y) \leq 0$ when $y < \tau_{k,y}$. The exact value of $\tau_{k,y}$ can be obtained by solving the likelihood ratio equation $p(\tau_{k,y}|H=1)/p(\tau_{k,y}|H=0) = \pi_0/\pi_1$.

From (22) and (28), the optimal kth sensor rule γ_k is

$$
U_k = \begin{cases} 1 & \text{if } \int_y p(x_k|y)\phi(y)dy > 0, \\ 0 & \text{otherwise.} \end{cases}
$$

To establish the sufficiency of a single threshold quantizer as defined in (25), it suffices to show that for $x_k > x'_k$,

$$
\int_Y p(x'_k|y)\phi(y)dy > 0
$$

implies

$$
\int_Y p(x_k|y)\phi(y)dy > 0.
$$

To show this, let

$$
c = \frac{p(x_k|\tau_{k,y})}{p(x'_k|\tau_{k,y})}.
$$

Hence, by the monotone property specified in condition 3), i.e., the ratio $\frac{p(x_k|y)}{p(x'_k|y)}$ is also a nondecreasing function of y for any $x_k > x'_k$, we have

$$
\begin{cases} \frac{p(x_k|y)}{p(x'_k|y)} \geq c, & \text{for} \quad y \geq \tau_{k,y}, \\ \frac{p(x_k|y)}{p(x'_k|y)} \leq c, & \text{for} \quad y \leq \tau_{k,y}. \end{cases}
$$

Thus,

$$\begin{aligned}\int_Y p(x_k|y)\phi(y)dy &= \int_Y p(x'_k|y)\frac{p(x_k|y)}{p(x'_k|y)}\phi(y)dy \\ &= \int_{-\infty}^{\tau_{k,y}} p(x'_k|y)\frac{p(x_k|y)}{p(x'_k|y)}\phi(y)dy + \int_{\tau_{k,y}}^{\infty} p(x'_k|y)\frac{p(x_k|y)}{p(x'_k|y)}\phi(y)dy \\ &\geq \int_{-\infty}^{\tau_{k,y}} cp(x'_k|y)\phi(y)dy + \int_{\tau_{k,y}}^{\infty} cp(x'_k|y)\phi(y)dy \\ &= c\int_Y p(x'_k|y)\phi(y)dy > 0\end{aligned}$$

■

Notice that the proof of this proposition is similar in spirit to that of Lemma 2 (iii) of [33, Section 3.3]. For the general case when the number of possible sensor outputs $L > 2$, we have the following result whose proof can be constructed analogously.

Theorem 1: Consider a distributed binary hypothesis testing system with scalar sensor observations and L-level sensor outputs. Suppose that the distributed hypothesis testing problem can be described equivalently by the CHCI model where the "hidden" random variable Y is a scalar random variable. Furthermore,

1) The fusion center implements a monotone fusion rule: $P(U_0 = 1|U_k = u, Y = y) \geq P(U_0 = 1|U_k = u', Y = y)$ for $u \geq u'$ and all possible y;
2) The ratio $g(y) \triangleq \frac{p(y|H=1)}{p(y|H=0)}$ is a nondecreasing function of y;
3) The ratio $h(y; x_k, x'_k) \triangleq \frac{p(x_k|y)}{p(x'_k|y)}$ is also a nondecreasing function of y for any $x_k > x'_k$,

Then there exists a $L-1$-threshold quantizer at sensor k such that

$$U_k = \begin{cases} U_k = L-1, & X_k \geq \tau_{k,L-1} \\ U_k = L-2, & \tau_{k,L-2} \leq X_k < \tau_{k,L-1} \\ \cdots, & \cdots \\ U_k = 0, & X_k < \tau_{k,1} \end{cases}, \tag{30}$$

that minimizes the error probability P_e.

We now examine two examples, the first one being the detection of a deterministic signal in dependent noises while the second one is the detection of a random signal in independent noise. These are the two classical cases of distributed detection with dependent observations. In both cases, the proposed approach leads to an optimal detection system, demonstrating the power of the proposed framework.

B. Detection of a Deterministic Signal in Correlated Gaussian Noise

We now revisit the binary hypothesis testing problem with two sensors observing correlated Gaussian data with different mean values under the two hypotheses. This problem was first considered in [18] and further explored in [19]. Specifically, the sensor observations are:

$$\begin{aligned} H = 0 : X_1, X_2 &\sim N(0, 0, 1, 1, \rho) \\ H = 1 : X_1, X_2 &\sim N(s_1, s_2, 1, 1, \rho) \end{aligned} \tag{31}$$

where $N(s_1, s_2, \sigma_1^2, \sigma_2^2, \rho)$ is the usual bivariate Gaussian density function with means $s_1 \geq 0$, $s_2 \geq 0$, variances σ_1^2, σ_2^2, and covariance $\rho\sigma_1\sigma_2$. Binary sensor output is assumed such that $L = 2$, with $U_k \in \{0, 1\}$. Without loss of generality, we assume $\rho \geq 0$. Otherwise, one can always multiply X_1 by -1. It was shown in both papers that, by restricting to the AND fusion rule, there exists the so-called "good " region defined by the set of parameters satisfying

$$(s_1 - \rho s_2)(s_2 - \rho s_1) \geq 0, \tag{32}$$

for which the optimal local sensor decision rules are single threshold quantizers with suitably chosen thresholds. Such local decision rules have exactly the same form as that when $\rho = 0$, i.e., when the sensor observations are conditionally independent of each other. For parameter sets that are outside of the "good" region, the optimal form of sensor decision rules remains largely unknown. The proof used in establishing the condition for the good region [19] relies on some particular properties of the Gaussian density functions.

We now provide a much more intuitive proof of the so-called good region using Proposition 1. Assume without loss of generality that $s_1 > 0$ and $s_2 > 0$. We note that condition (32) requires that either $s_1 > 0$ and $s_2 > 0$ or $s_1 < 0$ and $s_2 < 0$ and the latter case is equivalent to the former through multiplying both X_1 and X_2 with -1. Recall that any bivariate Gaussian random variables $(Z_1, Z_2) \sim \mathcal{N}(0, 0, 1, 1, \rho)$ admits a decomposition

$$\begin{aligned} Z_1 &= \sqrt{\rho}Y' + \sqrt{1-\rho}W_1' \\ Z_2 &= \sqrt{\rho}Y' + \sqrt{1-\rho}W_2' \end{aligned}$$

where (Y', W_1', W_2') are independent Gaussian random variables with zero mean and unit variance. Following similar approach, we can show that the sensor observations X_1, X_2 for the hypothesis testing

problem admit the following decomposition

$$\begin{aligned} X_1 &= s_1 Y + W_1 \\ X_2 &= s_2 Y + W_2, \end{aligned} \tag{33}$$

where

$$\begin{aligned} W_1 &\sim \mathcal{N}(0, 1 - \rho\tfrac{s_1}{s_2}), \\ W_2 &\sim \mathcal{N}(0, 1 - \rho\tfrac{s_2}{s_1}), \\ Y &\sim \mathcal{N}(H, \tfrac{\rho}{s_1 s_2}). \end{aligned} \tag{34}$$

Note that the dependence of Y on H is in its mean value: Y is zero mean under the $H = 0$ hypothesis and unit mean under the $H = 1$ hypothesis. Clearly, such a Y satisfies the Markov chain condition for the HCI model (11) and induces conditional independence of X_1 and X_2 given Y. From Equations (33) and (34), as W_1, W_2 and Y are all Gaussian random variables, one can easily verify the monotone properties 2) and 3) in Proposition 1. Thus, given that the AND fusion rule is monotonic, optimality of single threshold quantizers for both sensors follows from Proposition 1 directly.

Now we turn to the condition for the good region (32). For the above decomposition to hold, the variances for both W_1 and W_2 need to be non-negative, therefore

$$\left(1 - \rho\frac{s_1}{s_2}\right)\left(1 - \rho\frac{s_2}{s_1}\right) \geq 0.$$

For $s_1 > 0$ and $s_2 > 0$, this condition is easily seen to be equivalent to (32). Therefore, the good region condition specified in Equation (32) is precisely what is needed for the desired decomposition so that Proposition 1 can be applied.

For the above binary hypothesis testing problem with Gaussian observations, one can generalize the result to that of a multiple-sensor system. To illustrate this, consider the following hypothesis testing problem

$$\begin{aligned} H = 0 &: \mathbf{X} \sim N(\mathbf{0}, \Sigma_X) \\ H = 1 &: \mathbf{X} \sim N(\mathbf{s}, \Sigma_X), \end{aligned} \tag{35}$$

where $\mathbf{0} = [0, 0, \cdots, 0]^T$ and $\mathbf{s} = [s, s, \cdots, s]^T$ are the mean vectors of $\mathbf{X}$ under $H = 0$ and $H = 1$

hypotheses, respectively. If Σ_X has the following structure:

$$\Sigma_X = \begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{bmatrix}$$

then the sensor observations admit the following decomposition:

$$X_k = sY + W_k, \text{ for } k = 1, \cdots, K$$

where $Y \sim \mathcal{N}(H, \frac{\rho}{s^2})$ and $W_k \sim \mathcal{N}(0, 1-\rho)$, $k = 1, \cdots, K$ are independent of each other. It is then straightforward to verify that the conditions set forth in Proposition 1 are thus satisfied thus a single threshold quantizer is optimal for each sensor. The above result can also be extended to multi-sensor Gaussian hypothesis testing problem under more general conditions.

We now turn our attention to the other case, namely the detection of a random signal in independent noises.

C. Distributed Detection of a Random Signal

Consider the detection of a random signal S using K sensors. The random signal follows two different distributions under the two hypotheses: $S \sim p_0(s)$ if $H = 0$ and $S \sim p_1(s)$ if $H = 1$. The observation at sensor k is given by

$$X_k = a_k S + W_k, \tag{36}$$

where a_k is the attenuation factor that is determined by the distance between the source emitter and the sensor and W_k is the observation noise at the kth sensor. Each sensor makes a binary decision and sends it to a fusion center which makes a final decision regarding the hypotheses under test.

Consider first a simple Gaussian model where $p_0 \sim \mathcal{N}(0, \sigma_0^2)$ and $p_1 \sim \mathcal{N}(0, \sigma_1^2)$ with

$$0 \leq \sigma_0^2 < \sigma_1^2. \tag{37}$$

Furthermore, let us assume that the sensor noise is also zero mean Gaussian, i.e., $W_k \sim \mathcal{N}(0, \sigma^2)$. For such a Gaussian model, if there is only a single sensor (i.e., centralized detection), the hypotheses testing problem reduces to that of the detection of zero mean Gaussian signals with different variances for which

the simple energy detector is optimal. Or, equivalently, the optimal detector is a threshold test of the statistics $|X|$.

For the multiple sensor case, let the hidden variable be defined as $Y = S$. Apparently, given the signal model, $H - Y - \mathbf{X}$ forms a Markov chain and that $\mathbf{X}$ is conditionally independent given Y. Assume also that a monotone fusion rule, e.g., an L out of K majority rule, that satisfies (23) is used at the fusion center. From the proof of Prop. 1, we have

$$\tilde{f}_k^C(1, x_k) = -\int_y p(x_k|y)\phi(y)dy. \tag{38}$$

where $\tilde{f}_k^C(1, x_k)$ is defined in (28) and $\phi(y) = P(H=1)\frac{p(y|H=1)}{p(y|H=0)} - P(H=0)$. However, $\frac{p(y|H=1)}{p(y|H=0)}$ is *not* a monotonic function of $y \in \mathcal{R}$ where $\mathcal{R}$ is the entire real line. Thus we can not directly use Proposition 1 in obtaining the optimal sensor decision rules.

Nevertheless, the symmetry in the signal model allows us to still derive the optimal sensor decision rules under the proposed framework. Specifically, since $\phi(y)$ is symmetric at $Y = 0$, equation (38) can be rewritten as

$$\tilde{f}_k^C(1, x_k) = -\int_0^\infty (p(x_k|y) + p(x_k|-y))\phi(y)dy. \tag{39}$$

We first verify the monotonicity of $\phi(y)$ for $y > 0$ which is straightforward under the Gaussian model. Now comparing (39) and (38) and from the proof of Proposition 1, we need to verify that,

$$r(y) = \frac{p(x_k|y) + p(x_k|-y)}{p(x'_k|y) + p(x'_k|-y)} = \frac{e^{a_k x_k y} + e^{-a_k x_k y}}{e^{a_k x'_k y} + e^{-a_k x'_k y}} e^{x'^2_k - x_k^2},$$

is a monotone function of y for $x_k > x'_k \geq 0$. Differentiate $\ln r(y)$ with respect to y, we have,

$$\begin{aligned}
\frac{d}{dy}\ln r(y) &= a_k x_k \frac{e^{a_k x_k y} - e^{-a_k x_k y}}{e^{a_k x_k y} + e^{-a_k x_k y}} - a_k x'_k \frac{e^{a_k x'_k y} - e^{-a_k x'_k y}}{e^{a_k x'_k y} + e^{-a_k x'_k y}} \\
&> a_k x'_k \left\{ \frac{1 - e^{-2a_k x_k y}}{1 + e^{-2a_k x_k y}} - \frac{1 - e^{-2a_k x'_k y}}{1 + e^{-2a_k x'_k y}} \right\} \\
&> a_k x'_k \left\{ \frac{1 - e^{-2a_k x'_k y}}{1 + e^{-2a_k x'_k y}} - \frac{1 - e^{-2a_k x'_k y}}{1 + e^{-2a_k x'_k y}} \right\} \\
&= 0
\end{aligned}$$

for $x_k > x'_k \geq 0$ and $y > 0$. We have thus established that, for $X_k > 0$, the optimal local decision rule is

$$U_k = \begin{cases} 1 & \text{if } X_k \geq \tau_k, \\ 0 & \text{otherwise.} \end{cases} \tag{40}$$

where $\tau_k \geq 0$ is a suitable parameter. Furthermore, due to the symmetry of the observation model, $\tilde{f}_k^C$ is a symmetric function of X_k, i.e., $\tilde{f}_k^C(1, x_k) = \tilde{f}_k^C(1, -x_k)$. Therefore, the decision rule for the kth sensor is symmetric around 0 such that $\gamma_k(x_k) = \gamma_k(-x_k)$, or, equivalent, it is a function of $|X_k|$ only. Thus, the optimal local decision rule for all X_k is

$$U_k = \begin{cases} 1 & \text{if } |X_k| \geq \tau_k, \\ 0 & \text{otherwise.} \end{cases} \tag{41}$$

Perhaps it does not come across as surprising that the optimal local decision rule bears the same form as the optimal decision rule when only a single sensor is used. However, this is not true in general: absent of conditional independence among sensor observation, it is known that the optimal local sensor decision rules often differ from that of the optimal detector using a single sensor, i.e., an LRT may not be optimal in general for distributed detection with dependent observations.

Furthermore, it is not imperative to have a Gaussian model for the above detector structure to be optimal. The monotone properties can be also verified by assuming other uni-modal distribution symmetric at $S = 0$. Consider, for example, that S is a Laplacian distributed random signal with pdf

$$\frac{1}{2\sigma_i} e^{-\frac{|s|}{\sigma_i}}$$

with $0 < \sigma_0 < \sigma_1$ where σ_0 and σ_1 are the respective parameters under the two hypotheses. Again, by setting $Y = S$, Y satisfies the Markov chain conditions of a HCI model and also induces conditional independence. Compared with the Gaussian case and from the proof of Proposition 1, all we need to verify is the monotonicity of $\frac{p(y|H=1)}{p(y|H=0)}$ for $y \geq 0$. Straightforward calculation leads to

$$\frac{p(y|H=1)}{p(y|H=0)} = \frac{\sigma_0}{\sigma_1} e^{y\left(\frac{1}{\sigma_0} - \frac{1}{\sigma_1}\right)}, \quad y \geq 0.$$

Thus, it is indeed monotone increasing in y for $y \geq 0$. As such, simple thresholding of $|X_k|$ is also optimal when S is a Laplacian distributed random signal.

VI. Conclusion

We proposed a hierarchical conditional independence model that encompasses distributed detection problems with conditionally dependent and independent observations as its special cases. This model allowed us to develop a novel framework in dealing with distributed detection with dependent observations.

We have identified classes of distributed detection problems with dependent observations whose optimal sensor decision rules are uncoupled in their *form*. This decoupling effect is reminiscent of that of the conditional independence case thereby allowing the use of efficient algorithms to find these optimal sensor decision rules. The unifying power is most easily illustrated by providing a satisfying explanation of why single threshold quantizers are optimal for distributed detection of a constant signal in correlated Gaussian noises.

This general model will enable us to identify and solve new classes of distributed inference problems with dependent observations and the proposed framework can be adapted for distributed detection with other topological structures, e.g., the serial network, and under different inference regimes, e.g., the Neyman-Pearson problem.

REFERENCES

[1] R. Radner, "Team decision problems," *Annals of Mathematical Statistics*, vol. 33, pp. 857–881, 1962.

[2] R.R. Tenney and N.R. Sandell Jr., "Detection with distributed sensors," *IEEE Trans. Aerospace Elec. Syst.*, vol. 17, pp. 501–510, July 1981.

[3] P.K. Varshney, *Distributed Detection and Data Fusion*, Springer, New York, 1997.

[4] J.N. Tsitsiklis, "Decentralized detection," in *Advances in Statistical Signal Processing*, H.V. Poor and J.B. Thomas, Eds. JAI Press, Greenwich, CT, 1993.

[5] R. Viswanathan and P.K. Varshney, "Distributed detection with multiple sensors: Part I — Fundamentals," *Proceedings of the IEEE*, vol. 85, no. 1, pp. 54–63, Jan. 1997.

[6] R.S. Blum, S.A. Kassam, and H.V. Poor, "Distributed detection with multiple sensors: Part II — Advanced topics," *Proceedings of the IEEE*, vol. 85, pp. 64–79, Jan. 1997.

[7] Z. Chair and P.K. Varshney, "Optimal data fusion in multiple sensor detection systems," *IEEE Trans. Aerospace Elect. Sys.*, vol. 22, pp. 98–101, Jan. 1986.

[8] E. Drakopoulos and C.C. Lee, "Optimum multisensor fusion of correlated local decisions," *IEEE Trans. on Aerospace Elect. Syst.*, vol. 27, no. 4, pp. 593–605, July 1991.

[9] M. Kam, Q. Zhu, and W.S. Gray, "Optimal data fusion of correlated local decisions in multiple sensor detection systems," *IEEE Trans. Aerospace Elect. Syst.*, vol. 28, pp. 916–920, July 1992.

[10] J.N. Tsitsiklis and M. Athans, "On the complexity of decentralized decision making and detection problems," *IEEE Transactions on Automatic Control*, vol. 30, pp. 440–446, May 1985.

[11] D.J. Warren and P.K. Willett, "Optimum Quantization for Detector Fusion: Some Proofs, Examples, and Pathology," *Journal of the Franklin Institute*, vol. 336, pp. 323–359, Mar. 1999.

[12] B. Chen and P.K. Willett, "On the optimality of likelihood ratio test for local sensor decisions in the presence of non-ideal channels," *IEEE Trans. Inf. Theory*, vol. 51, pp. 693–699, Feb. 2005.

[13] A. Kashyap, "Comments on 'on the optimality of the likelihood-ratio test for local sensor decision rules in the presence of nonideal channels'," *IEEE Trans. on Inf. Theory*, vol. 52, pp. 1274–1275, Mar. 2006.

[14] H. Chen, B. Chen, and P.K. Varshney, "Further results on the optimality of likelihood ratio quantizer for distributed detection in non-ideal channels," *IEEE Trans. Inf. Theory*, vol. 55, pp. 828–832, Feb. 2009.

[15] B. Chen, L. Tong, and P.K. Varshney, "Channel aware distributed detection in wireless sensor networks," *IEEE Signal Processing Magazine*, vol. 23, no. 4, pp. 16–26, July 2006.

[16] Z.-B. Tang, K.R. Pattipati, and D.L. Kleinman, "A distributed M-ary hypothesis testing problem with correlated observations," *IEEE Trans. Automatic Control*, vol. 37, pp. 1042–1046, July 1992.

[17] Q. Yan and R. Blum, "Distributed signal detection under the Neyman-Pearson criterion," *IEEE Trans. Inf. Theory*, vol. 47, pp. 1368–1377, May 2001.

[18] P.-N. Chen and A. Papamarcou, "Likelihood ratio partitions for distributed signal detection in correlated gaussian noise," in *Proc. IEEE Int. Symp. Inform. Theory*, Oct. 1995, p. 118.

[19] P.K. Willett, P.F. Swaszek, and R.S. Blum, "The good, bad, and ugly: distributed detection of a known signal in dependent Gaussian noise," *IEEE Trans. Signal Processing*, vol. 48, pp. 3266–3279, Dec. 2000.

[20] T. Berger, "Decentralized estimation and decision theory," in *presented at the IEEE 7th Spring Workshop on Information Theory*, Mt. Kisco, NY, 1979.

[21] T.S. Han and S. Amari, "Statistical inference under multiterminal data compression," *IEEE Trans. Inf. Theory*, vol. 44, pp. 2300–2324, Oct. 1998.

[22] R. Ahlswede and I. Csiszar, "Hypothesis testing with communication constraints," *IEEE Trans. Inform. Theory*, vol. IT-32, pp. 533–542, July 1986.

[23] J. Chamberland and V.V. Veeravalli, "How dense should a sensor network be for detection with correlated observations," *IEEE Trans. Inform. Theory*, vol. IT-52, pp. 5099–5106, Nov. 2006.

[24] H. Chen, P.K. Varshney, and B. Chen, "Conditional dependence in distributed detection: How far can we go?," in *Proc. IEEE International Symposium on Information Theory*, Seoul, Korea, July 2009.

[25] H. Chen, P.K. Varshney, and B. Chen, "A novel framework for distributed detection with dependent observations," in *Proc. IEEE International Conference on Acoustic, Speech, and Signal Processing*, Dallas, TX, March 2010.

[26] L. Ozarow, "On a source-coding problem with two channels and three receivers," *Bell Syst. Tech. J.*, vol. 59, no. 10, pp. 1909–1921, 1980.

[27] A. Wagner, S. Tavildar, and P. Viswanath, "Rate region of the quadratic gaussian two-encoder source-coding problem," *IEEE Trans. Inf. Theory*, vol. 54, pp. 1938–1961, May 2008.

[28] H. Viswanathan and T. Berger, "The quadratic Gaussian CEO problem," *IEEE Trans. Inf. Theory*, vol. 43, pp. 1549–1559, Sep. 1997.

[29] Y. Oohama, "Rate-distortion theory for Gaussian multiterminal source coding systems with several side information at the decoder," *IEEE Trans. Inf. Theory*, vol. 51, pp. 2577–2593, July 2005.

[30] S. Kay, *Fundamentals of Statistical Signal Processing II: Detection Theory*, Prentice Hall, Englewood Cliffs, NJ 07632, 1998.

[31] T.M. Cover and J.A. Thomas, *Elements of Information Theory*, Wiley, New York, 1991.

[32] G.E. Shilov, *Linear Algebra*, Dover Publications, New York, 1977.

[33] E.L. Lehmann, *Testing Statistical Hypotheses*, Wiley, New York, second edition, 1986.

Hao Chen received the Ph.D. degree in Electrical Engineering from the Syracuse University in 2007. In 2007-2010, he was a postdoctoral research associate and then a research assistant professor at Syracuse University. Since Aug 2010, he has been an assistant professor at the Department of Electrical and Computer Engineering and Computer Science at the Boise State University. Dr. Chen's research interests include statistical signal and image processing, and wireless communications.

Biao Chen received the B.E. and M.E. degrees in electrical engineering from Tsinghua University, Beijing, China, in 1992 and 1994, respectively. He joined the University of Connecticut, Storrs, CT, in 1995, where he received the M.S. degree in statistics and the Ph.D. degree in electrical engineering, in 1998 and 1999, respectively. From 1994 to 1995, he worked at AT&T (China) Inc., Beijing, China. From 1999 to 2000, he was with Cornell University, Ithaca, NY, as a Postdoctoral Associate. Since 2000, he has been with Syracuse University, Syracuse, NY, where he is currently a Professor in the Department of Electrical Engineering and Computer Science. His area of interest mainly focuses on signal processing and information theory for wireless sensor and ad hoc networks and in multiuser multiple-input multiple-output (MIMO) systems.

Prof. Chen has served as Associate Editor for the IEEE COMMUNICATIONS LETTERS, the IEEE TRANSACTION ON SIGNAL PROCESSING, and the EURASIP Journal onWireless Communications and Networking (JWCN). He is the recipient of an NSF CAREER Award in 2006.

Pramod K. Varshney was born in Allahabad, India, on July 1, 1952. He received the B.S. degree in electrical engineering and computer science (with highest honors), and the M.S. and Ph.D. degrees in electrical engineering from the University of Illinois at Urbana-Champaign in 1972, 1974, and 1976 respectively.

During 1972-76, he held teaching and research assistantships at the University of Illinois. Since 1976 he has been with Syracuse University, Syracuse, NY where he is currently a Distinguished Professor of Electrical Engineering and Computer Science and the Director of CASE: Center for Advanced Systems and Engineering. He served as the Associate Chair of the department during 1993-96. He is also an Adjunct Professor of Radiology at Upstate Medical University in Syracuse, NY. His current research interests are in distributed sensor networks and data fusion, detection and estimation theory, wireless communications, image processing, radar signal processing and remote sensing. He has published extensively. He is the

author of Distributed Detection and Data Fusion, published by Springer-Verlag in 1997. He has served as a consultant to several major companies.

While at the University of Illinois, Dr. Varshney was a James Scholar, a Bronze Tablet Senior, and a Fellow. He is a member of Tau Beta Pi and is the recipient of the 1981 ASEE Dow Outstanding Young Faculty Award. He was elected to the grade of Fellow of the IEEE in 1997 for his contributions in the area of distributed detection and data fusion. He was the guest editor of the special issue on data fusion of the Proceedings of the IEEE, January 1997. In 2000, he received the Third Millennium Medal from the IEEE and Chancellor's Citation for exceptional academic achievement at Syracuse University. He is the recipient of the IEEE 2012 Judith A. Resnik Award. He serves as a distinguished lecturer for the AES society of the IEEE. He is on the editorial board of Journal on Advances in Information Fusion. He was the President of International Society of Information Fusion during 2001.