

Robust Binary Quantizers for Distributed Detection

Ying Lin, Biao Chen, and Bruce Suter

Abstract

We consider robust signal processing techniques for inference-centric distributed sensor networks operating in the presence of possible sensor failures and/or communication failures. Motivated by the multiple description (MD) principle, we develop robust *distributed* quantization schemes for a decentralized detection system. Specifically, focusing on a two-sensor system, our design criterion mirrors that of MD principle: if one of the two transmissions fails, we can guarantee an acceptable performance, while enhanced performance can be achieved if both transmissions are successful. Different from the conventional MD problem is the distributed nature of the problem as well as the use of error probability as the performance measure. Two different optimization criteria are used in the distributed quantizer design, the first a constrained optimization problem, and the second using an erasure channel model. We demonstrate that these two formulations are intrinsically related to each other. Further, using a person-by-person optimization approach, we propose an iterative algorithm to find the optimal local quantization thresholds. A design example is provided to illustrate the validity of the iterative algorithm and the improved robustness compared to the classical distributed detection approach that disregards the possible transmission losses.

Index Terms Sensor networks, distributed detection, erasure channel, multiple description.

Y. Lin and B. Chen (corresponding author) are with the Department of Electrical Engineering and Computer Science, Syracuse University, 335 Link Hall, Syracuse, NY 13244, USA. Email: {ylin20, bichen}@ecs.syr.edu. Bruce Suter is with AFRL/IFGC, 525 Brooks Rd, Rome, NY 13442, USA. Email: suterb@r1.af.mil. This work was supported in part by the National Science Foundation under grant ECS-0501534 and by the Air Force Research Laboratory under agreement FA8750-05-2-0120.

I. INTRODUCTION

For the emerging wireless sensor networks (WSN), distributed signal processing design has to deal with various physical limitations imposed by severe resource constraints. For example, the power and bandwidth constraints, coupled with the interference and channel fading, may result in transmission loss due to channel outage. In addition, low-cost sensor nodes deployed in harsh environments may be subject to sensor failure, making them unavailable for sensing/communication.

A conventional approach to combat transmission loss is to exploit channel diversity through the use of multiple description (MD) design [1] such as the MD codes [2] or MD quantizers [3]. This MD idea is illustrated in Fig. 1(a) with two encoders and three decoders [2]. The encoders are so designed that in the case of loss of one of the two transmissions, the side decoders (Decoder 1 or Decoder 2) are guaranteed with certain acceptable performance; if both transmissions are successful, the central decoder output (corresponding to Decoder 0) will have enhanced performance. As sensor failure can be dealt with in an identical fashion under the MD framework, we will no longer distinguish the two types of losses, one due to channel outage and the other due to sensor failure.

To carry over the MD principle to sensor network applications, care must be taken in considering the distinct features for distributed sensor networks. Two of the critical differences are listed below and are what motivate the current work.

- Distributed nature of WSN.

In the conventional MD framework, two encoders operate on a common source. In WSN, each encoder resides in a sensor and operates only on its own observations *without* access to the other sensor's observations. This is illustrated in Fig. 1.

- Inference-centric nature of WSN.

In WSN applications, all the sensor nodes are typically engaged in a collective inference task.

The ultimate goal may be the evaluation of some underlying state instead of recovering the sensor observations. In reference to Fig. 1(b), the goal may be inferring about the unknown parameter θ instead of recovering X_1 and X_2 . This is in comparison with the conventional MD problem where the goal is to recover the original source data. A direct consequence is that, instead of using the conventional distortion measures in the traditional MD quantizer design, other performance metrics that cater toward the inference task may be more relevant.

In this paper, we study how the MD principle can be adapted to inference-centric applications with distributed quantizer design. By focusing on a binary decentralized hypothesis testing problem (i.e., θ is binary in Fig. 1(b)), we investigate distributed binary quantizer design using the MD principle. We term this new framework distributed multiple description quantizer (DMDQ) design. The DMDQ approach achieves robust inference performance in the presence of channel outage or sensor failure as it strikes a better balance/tradeoff between the detection performance at the fusion center and that of local sensors.

The proposed scalar quantizer design is closely related to the classical distributed detection problem [4], [5] as it involves the design of multiple sensor decision rules that are coupled with each other. Major differences exist, and the most significant is that we no longer deal with a single objective function (minimum error probability at the fusion center). Instead, multiple design objectives need to be considered, each corresponding to the end-to-end inference performance for a particular channel outage or sensor failure state.

To explain the significance of the proposed approach, and in particular, to understand its improved robustness compared with the classical distributed detection design, consider the following simple example. Assume a binary hypothesis testing problem with a two-sensor parallel fusion system where each sensor employs a binary quantizer. The two hypotheses under test, H_0 and H_1 , are *a priori* equally likely. The local sensor observations at the two sensors, X_1 and

X_2 , are conditionally independent and identically distributed ternary random variables with

$$\begin{cases} P(X_k = 0|H_0) = 0.95 \\ P(X_k = 1|H_0) = 0.05 \\ P(X_k = 2|H_0) = 0 \end{cases} \quad \begin{cases} P(X_k = 0|H_1) = 0.05 \\ P(X_k = 1|H_1) = 0.9 \\ P(X_k = 2|H_1) = 0.05 \end{cases}$$

for $k = 1, 2$. By the monotonicity of the likelihood ratio (LR) in the sensor observations (i.e., the local sensor LR values are monotone in X_k), we need to consider only the two binary local decision rules at each sensor [6]:

$$\text{Rule A} \quad U_k = \begin{cases} 0 & X_k = 0 \\ 1 & X_k = 1 \text{ or } 2 \end{cases} \quad \text{Rule B} \quad U_k = \begin{cases} 0 & X_k = 0 \text{ or } 1 \\ 1 & X_k = 2 \end{cases}$$

Adopting the classical distributed detection approach, it is straightforward to show that the two sensors should employ different decision rules to achieve a minimum error probability of 0.04875 at the fusion center. Assume that, without loss of generality, sensor 1 uses Rule A while sensor 2 uses Rule B. If sensor 1's decision does not reach the fusion center, either due to a channel outage or a sensor failure, the actual minimum error probability by using the decision from sensor 2 alone becomes 0.475, which is a significant degradation from the case when both sensor outputs are available. This error probability essentially renders the detection system essentially useless as it is close to 0.5. A more robust design is to use decision rule Rule A at both sensors. In this case, both the fusion center and each local sensor have identical error probability 0.05 thus there is no degradation in the event of a lost transmission¹. Compared with the classical distributed detection approach (whose error probability pair are 0.04875 and 0.475), the alternative approach provides a more robust performance in the presence of a transmission loss.

The proposed DMDQ also provides an alternative approach to the channel aware design for a decentralized detection problem [7]–[9] in dealing with imperfect channels. The channel-aware quantization schemes require that the channel state information (CSI) be available to attain

¹This simple example also indicates that, *depending on the local decision rules used and the observation distributions*, having more sensors in the system may not always improve the overall performance.

optimum performance. Acquiring CSI, however, may be too costly in systems with stringent resource constraints. It is, therefore, imperative to consider quantizer design that is robust to potential channel outages *without* the knowledge of CSI. The proposed DMDQ framework is an initial attempt toward robust and proactive signaling for distributed sensor networks in the absence of CSI.

The rest of the paper is organized as follows. In the next section, we describe the problem formulation and introduce the two-sensor fusion network with possible transmission losses. In Section III, we apply the Lagrangian method to solve the constrained minimization and to obtain necessary conditions for optimum binary quantizers in the form of LR test (LRT) thresholds. In Section IV, we impose the discrete memoryless erasure channel model and obtain the corresponding optimum local decision rules using the channel-aware quantizer design methodology described in [7], [8]. Numerical results are presented in Section V to demonstrate how the proposed quantizer design can be implemented and the improved robustness over the classical distributed detection approach. We conclude in Section VI.

II. PROBLEM FORMULATION

Fig. 2 depicts a two-sensor parallel fusion network tasked with a hypothesis testing problem. Each sensor collects data that are generated according to one of the two hypotheses (H_0 and H_1) under test. We assume in the present work that the local observations X_1 and X_2 are conditionally independent given the underlying hypothesis, i.e., for $i = 0, 1$,

$$f(X_1, X_2|H_i) = f(X_1|H_i)f(X_2|H_i).$$

It is easy to establish that with this conditional independence assumption, the LR pair of the local sensor observations

$$\left(L(X_1) = \frac{f(X_1|H_1)}{f(X_1|H_0)}, L(X_2) = \frac{f(X_2|H_1)}{f(X_2|H_0)} \right)$$

form a sufficient statistic for the detection problem.

Based on its local observation X_k , the k th local sensor implements a binary quantizer whose output $U_k \in \{1, 0\}$, for $k = 1, 2$, will be sent to the fusion center. The transmission, however, is subject to channel outage or sensor failure. When both transmissions are successful, Decoder 0 will perform as a fusion center and make a final decision on which hypothesis is true using both U_1 and U_2 . Otherwise, if only one of the two transmissions is successful, either Decoder 1 or Decoder 2 will make a final decision based on the successfully received U_k . In our current work, as U_k is binary, Decoders 1 and 2 will simply take U_1 and U_2 as their respective output, as illustrated in Fig. 2.

Adopting a Bayesian framework, we use error probability as the performance measure. Define P_{ek} the probability of error at Decoder k :

$$P_{ek} = \pi_0 P(U_k = 1|H_0) + \pi_1 P(U_k = 0|H_1), \quad k = 0, 1, 2 \quad (1)$$

where $\pi_j = P(H_j)$ is the prior probability for hypothesis H_j , and U_0 denotes the decision output for Decoder 0. Thus P_{e0} corresponds to the error probability of the fusion center when both U_1 and U_2 are available while P_{e1} and P_{e2} are respectively the error probabilities at individual sensors. Classical distributed detection theory aim to minimize P_{e0} while our present work strives for a balance in performance between P_{e0} and P_{ek} for $k = 1, 2$.

Our approach is derived from the MD principle [1]: we aim to design sensor decision (quantization) rules such that if one of the two transmissions is lost, an acceptable performance (in terms of error probability) is guaranteed; if both transmissions are successful, a better performance can be achieved. Catering toward the hypothesis testing problem, we can succinctly summarize the design criterion using the following constrained minimization problem

$$\begin{aligned} \min \quad & P_{e0} \\ \text{subject to} \quad & P_{e1} \leq \varepsilon_1 \quad \text{and} \quad P_{e2} \leq \varepsilon_2. \end{aligned} \quad (2)$$

where ε_1 and ε_2 are the pre-specified error probabilities that are guaranteed if only U_1 or U_2 is successfully received. This design criterion is reminiscent of the MD scalar quantizer design [3]

where a general distortion measure is used.

III. NECESSARY CONDITIONS FOR OPTIMALITY AND A DESIGN ALGORITHM

The constrained optimization problem readily admits a Lagrangian formulation which is used to solve the minimization problem below. The Lagrangian function is given by

$$L(\tau_1, \tau_2, \lambda_1, \lambda_2) = P_{e0} + \lambda_1(P_{e1} - \varepsilon_1) + \lambda_2(P_{e2} - \varepsilon_2) \quad (3)$$

where τ_k is the local sensor LRT threshold, λ_k is the Lagrangian multipliers, for $k = 1, 2$.

Using the Kuhn-Tucker theorem [10], the set of optimum solution of the constrained minimization problem must satisfy the following necessary conditions, for $k = 1, 2$,

$$\frac{\partial P_{e0}}{\partial \tau_k} + \lambda_1 \frac{\partial P_{e1}}{\partial \tau_k} + \lambda_2 \frac{\partial P_{e2}}{\partial \tau_k} = 0 \quad (4)$$

$$\lambda_k \geq 0 \quad (5)$$

$$P_{ek} - \varepsilon_k \leq 0 \quad (6)$$

$$\lambda_k(P_{ek} - \varepsilon_k) = 0 \quad (7)$$

Given the above necessary conditions, the optimum solutions for the local decision rules are described in the following theorem.

Theorem 1: Assume that the two local observations, X_k 's, are conditionally independent.

Further, if the fusion rule and the k th local sensor decision rule satisfy, for $k = 1, 2$

$$\begin{cases} P(U_0 = 1|U_k = 1, U_{\bar{k}}) - P(U_0 = 1|U_k = 0, U_{\bar{k}}) \geq 0 \\ P(U_0 = 0|U_k = 0, U_{\bar{k}}) - P(U_0 = 0|U_k = 1, U_{\bar{k}}) \geq 0 \end{cases}$$

where $\bar{k} \triangleq 3 - k$, thus $\bar{1} = 2$ and $\bar{2} = 1$. Then the optimum solution of the constrained minimization problem in Eq. (2) is given by the following LRT, for $k = 1, 2$

$$P(U_k = 1|X_k) = \begin{cases} 1, & \text{if } \frac{p(X_k|H_1)}{p(X_k|H_0)} \geq \tau_k \\ 0, & \text{otherwise} \end{cases} \quad (8)$$

where τ_k , the optimal LRT threshold for the k th local sensor, is determined as follows:

- When $\lambda_k = 0$ (inactive constraint),

$$\tau_k = \frac{\pi_0 A_k}{\pi_1 B_k} \quad (9)$$

- When $\lambda_k > 0$ (active constraint), τ_k is obtained by solving

$$P_{ek} - \epsilon_k = 0 \quad (10)$$

The associated λ_k can be obtained by

$$\lambda_k = \frac{\pi_0 A_k - \pi_1 B_k \tau_k}{\pi_1 \tau_k - \pi_0} \quad (11)$$

from which we get,

$$\tau_k = \frac{\pi_0 (A_k + \lambda_k)}{\pi_1 (B_k + \lambda_k)} \quad (12)$$

The quantities A_k and B_k in Eqs. (9-12) are defined respectively as

$$A_k = \sum_{U_{\bar{k}}} (P(U_0 = 1|U_k = 1, U_{\bar{k}}) - P(U_0 = 1|U_k = 0, U_{\bar{k}})) P(U_{\bar{k}}|H_0) \quad (13)$$

$$B_k = \sum_{U_{\bar{k}}} (P(U_0 = 0|U_k = 0, U_{\bar{k}}) - P(U_0 = 0|U_k = 1, U_{\bar{k}})) P(U_{\bar{k}}|H_1) \quad (14)$$

Theorem 1 is proved in Appendix A.

Remarks:

- Note that the forms of τ_k , A_k , and B_k indicate that the threshold for the k th sensor is a function of the decision rule at the other sensor. Thus, as expected, the optimal thresholds at sensor 1 and 2 are coupled with each other.
- In order for the constrained optimization to have feasible solutions, ϵ_1 and ϵ_2 can not be chosen to be too small. Specifically, ϵ_k needs to be no smaller than the minimum achievable error probability at sensor k . More discussions about this can be found in Section V after we introduce an alternative design approach (Approach 3 in Section V).
- P_{e0} is the achievable error probability using both U_1 and U_2 . On the other hand, P_{e1} and P_{e2} are respectively the error probabilities at local sensors, each associated with U_1 or U_2 . Thus

$P_{e0} \leq \min\{P_{e1}, P_{e2}\}$ where the inequality is due to the fact that one can simply ignore one of $\{U_1, U_2\}$ and the error probability should thus be no worse than either P_{e1} or P_{e2} . Due to the constraints $P_{e1} \leq \epsilon_1$ and $P_{e2} \leq \epsilon_2$, we have

$$P_{e0} \leq \min\{P_{e1}, P_{e2}\} \leq \min\{\epsilon_1, \epsilon_2\}$$

That is, the error probability achieved when both transmissions are successful is upper bounded by the error probability constraints at local sensors.

- If $P_{e1} < \epsilon_1$ and $P_{e2} < \epsilon_2$, i.e., $\lambda_k = 0$ for $k = 1, 2$, P_{e0} is the minimum error probability that can be achieved at the fusion center. The constrained optimization approach yields the same result as the unconstrained approach that minimizes the error probability at the fusion center. This happens when the constraints ϵ_k are large enough.
- Eq. (12) is a unifying expression of the optimal local LR threshold for the two cases of $\lambda_k > 0$ and $\lambda_k = 0$.
- The conditions described in Theorem 1 do not admit closed-form solutions. Simultaneously optimizing τ_1 and τ_2 is intractable due to the distributed nature - it typically involves some exhaustive search over a two dimension space for the (τ_1, τ_2) pair. However, the necessary conditions established in Theorem 1 allows us to adopt a person-by-person optimization (PBPO) approach where each threshold is optimized assuming fixed threshold at the other sensor. The PBPO approach has been widely used in optimizing decentralized systems, and in particular, in the classical distributed detection (see, e.g., [11], [12]) when joint optimization is typically intractable.
- Theorem 1 describes necessary conditions for the optimum LRT thresholds; thus multiple initializations are needed to find the globally optimum thresholds.

The following iterative algorithm describes this PBPO procedure.

Iterative Algorithm

- Step 1. Initialize τ_k , for $k = 1, 2$.

- Step 2. Obtain the optimum fusion rule for fixed τ_1 and τ_2 .
- Step 3. For fixed fusion rule and τ_2 , calculate τ_1 using (9).
- Step 4. Check to see if τ_1 satisfies $P_{e1} - \epsilon_1 \leq 0$.
 - If yes, go to Step 5.
 - If no, calculate τ_1 using (10).
- Step 5. For fixed fusion rule and τ_1 , calculate τ_2 in a similar fashion.
- Step 6. Check convergence, i.e, if the obtained τ_1 and τ_2 are identical (up to a prescribed tolerance) to that from the previous iteration.
 - If yes, stop.
 - Otherwise, go to Step 2.

At each iteration, τ_k is optimized for a given fusion rule and the other threshold $\tau_{\bar{k}}$, hence the error probability is monotone decreasing until a stationary point is reached.

IV. OPTIMAL LOCAL DECISION RULE DESIGN UNDER AN ERASURE CHANNEL MODEL

The constrained minimization approach provides a proactive design methodology that avoids severe performance degradation in the absence of CSI. We propose in this section an alternative approach by imposing a certain parametric model on the channel/sensor failures. This allows us to adopt existing channel aware approach [8] to design the local quantizers. Similar to [13], we model the potential transmission loss using *erasure* channels where the erasure accounts for possible sensor failures/channel outages. This channel model is illustrated in Fig. 3 where $\delta_k = P(X_k = E|U_k)$ is the erasure probability corresponding to sensor k . Our alternative optimization criterion is to minimize the average error probability P_e , defined as

$$P_e = (1 - \delta_1)(1 - \delta_2)P_{e0} + \delta_2(1 - \delta_1)P_{e1} + \delta_1(1 - \delta_2)P_{e2} + \min\{\pi_0, \pi_1\}\delta_1\delta_2 \quad (15)$$

where the last term corresponds to the error probability when both transmissions are lost. This constant term has no effect on the quantizer design, hence can be dropped in the design problem.

The following theorem provides the solution for the sensor decision rules that minimize P_e .

Theorem 2: Assume that the two local observations, X_k 's, are conditionally independent and channels are independent discrete memoryless erasure channels. Further, if the fusion rule and the k th local sensor decision rule satisfy, for $k = 1, 2$

$$\begin{cases} P(U_0 = 1|U_k = 1, U_{\bar{k}}) - P(U_0 = 1|U_k = 0, U_{\bar{k}}) \geq 0 \\ P(U_0 = 0|U_k = 0, U_{\bar{k}}) - P(U_0 = 0|U_k = 1, U_{\bar{k}}) \geq 0 \end{cases}$$

where \bar{k} is defined similar as in Theorem 1. Then the optimum local rule for the k th sensor amounts to the following LRT, for $k = 1, 2$

$$P(U_k = 1|X_k) = \begin{cases} 1 & \text{if } \frac{p(X_k|H_1)}{p(X_k|H_0)} \geq \frac{\pi_0(A_k + \alpha_k)}{\pi_1(B_k + \alpha_k)} \\ 0 & \text{otherwise} \end{cases} \quad (16)$$

where

$$\alpha_1 = \frac{\delta_2}{1 - \delta_2} \text{ and } \alpha_2 = \frac{\delta_1}{1 - \delta_1}, \quad (17)$$

and A_k and B_k are defined in Eqs. (13) and (14).

A proof is given in Appendix B. Following the same spirit of the iterative algorithm in Section III, we can devise a similar procedure to find the optimal thresholds using Theorem 2.

Comparing Eqs. (12) and (16), we have some interesting observations that suggest intrinsic connections between the erasure channel model and the constrained minimization formulation. From Eq. (15), if we drop the last term and divide the average probability by $(1 - \delta_1)(1 - \delta_2)$, the new function to be minimized becomes

$$Q \triangleq P_{e0} + \alpha_1 P_{e1} + \alpha_2 P_{e2}.$$

with α_1 and α_2 defined as in Eq. (17). The design problem reduces to a problem of minimizing Q subject to $\alpha_1 \geq 0$ and $\alpha_2 \geq 0$. Compare this with Eq. (3), we see that α_k plays a similar role as the Lagrangian multiplier λ_k .

Further more, the first-order necessary conditions for minimizing Q are given by:

$$\frac{\partial P_{e0}}{\partial \tau_k} + \alpha_1 \frac{\partial P_{e1}}{\partial \tau_k} + \alpha_2 \frac{\partial P_{e2}}{\partial \tau_k} = 0 \quad (18)$$

$$\alpha_k \geq 0 \quad (19)$$

Comparing Eq. (18) and (19) to Eq. (4-7), we notice that these two formulations are similar except that the constrained optimization approach has more restrictive constraints (Eq. (6) and (7)). Next we elaborate when these two formulations will have identical optimal solutions.

Consider the first case: when $\lambda_k = 0$, i.e., the constraints $P_{ek} \leq \epsilon_k$ are satisfied. In this case, set $\alpha_k = \lambda_k = 0$, and the two formulations have the same optimal thresholds. The case of $\lambda_k > 0$ is more complicated. With $\lambda_k > 0$, we have $P_{ek} = \epsilon_k$, $k = 1, 2$. Assume the erasure channel model yields L local minima, with the corresponding threshold pair (τ_1^l, τ_2^l) , $l = 1, 2, \dots, L$. Denote by P_{ej}^l , $j = 0, 1, 2$, the error probabilities associated with (τ_1^l, τ_2^l) . By virtue of the problem formulation, there must exist one (τ_1^m, τ_2^m) whose local error probabilities satisfy $P_{ek}^m = \epsilon_k$, $k = 1, 2$. If

$$Q^m \triangleq P_{e0}^m + \alpha_1 \epsilon_1 + \alpha_2 \epsilon_2 \leq P_{e0}^j + \alpha_1 P_{e1}^j + \alpha_2 P_{e2}^j \triangleq Q^j \quad \text{for } j \neq m, j = 1, 2, \dots, L \quad (20)$$

then (τ_1^m, τ_2^m) is the optimal solution for both constrained minimization formulation and the erasure channel formulation. We will further illustrate these connections using some numerical examples in the next section.

V. A NUMERICAL EXAMPLE

In this section, we use several numerical examples to highlight the robust performance of the proposed local quantizer design compared with the classical distributed detection approach. Consider the detection of a known signal in independent Gaussian noises using two sensors:

$$H_0 : \quad X_k = n_k$$

$$H_1 : \quad X_k = s + n_k$$

where s is a known signal and n_k is zero mean Gaussian noise with variance σ^2 , for $k = 1, 2$. Without loss of generality, we assume $s = 1$ and $\sigma^2 = 1$. Each local sensor makes a binary decision using its observation X_k and a decision rule γ_k , i.e., $U_k = \gamma_k(X_k) \in \{1, 0\}$. The transmission of U_k , however, is subject to channel losses. If both U_1 and U_2 are successfully received, Decoder 0 will implement the maximum *a posteriori* probability decoding (detection) rule, i.e.,

$$P(U_0 = 1|U_1, U_2) = \begin{cases} 1, & \text{if } \frac{P(U_1, U_2|H_1)}{P(U_1, U_2|H_0)} \geq \frac{\pi_0}{\pi_1} \\ 0, & \text{otherwise} \end{cases} \quad (21)$$

For simplicity, we consider a symmetric setting where we use identical error probability constraints (i.e., $\epsilon_1 = \epsilon_2$) for the constrained minimization approach, and identical erasure probabilities (i.e., $\delta_1 = \delta_2$) for the erasure channel model approach.

In addition to the proposed approaches, we also present results using alternative approaches to highlight the robustness of the proposed MD principle based framework. The complete list of approaches used in the simulations is as follows.

Approach 1 Constrained minimization described in Section III (Theorem 1).

Approach 2 Erasure channel model approach described in Section IV (Theorem 2).

Approach 3 Minimizing the local error probabilities P_{e1} and P_{e2} . We denote by $P_{ek,3}$ ($k = 1, 2$) the minimum achievable local error probabilities, and $P_{e0,3}$ the corresponding error probability at Decoder 0, respectively. Note that $P_{ek,3}$ provides the lower bound for the local error probability constraint ϵ_k , i.e., one must have $\epsilon_k \geq P_{ek,3}$ for the constrained minimization formulation to have a solution.

Approach 4 Minimizing the error probability at the fusion center. We denote by $P_{e0,4}$ the minimum achievable error probability at Decoder 0, and $P_{ek,4}$ ($k = 1, 2$) the corresponding local error probabilities, respectively. *This approach corresponds to the classical distributed detection with a single objective function.* An interesting observation is that this approach can be considered as a special case of the erasure channel model with

$\delta_k = 0$, for $k = 1, 2$. As such, one only need to minimize P_{e0} as both transmissions are always assumed successful.

Notice that Approaches 3 and 4 are conflicting with each other: one can show that optimizing P_{e0} and P_{ek} for $k = 1, 2$ can not be simultaneously achieved [14]. Otherwise, the entire distributed MD framework will become trivial as one can simultaneously optimize the local error probability and that of Decoder 0 (the fusion center).

As we are considering a Gaussian problem, the obtained LRT thresholds at the sensors can be directly translated into thresholds for the original observations. Thus in the following presentation, we will use thresholds for the original observations, denoted by η_k for $k = 1, 2$.

The numerical results are summarized in Tables I-III as well as in Figs. 4-6. Specifically,

- Tables I and II enumerate respectively the parameters and the obtained thresholds and error probabilities of the two proposed approaches (Approaches 1 and 2).
- Tables III gives the obtained thresholds and error probabilities of the two alternative approaches (Approaches 3 and 4).
- Figs. 4 and 5 give the analytically calculated error probabilities (both of the fusion center and local sensors) versus threshold plots with two different priors, $\pi_0 = 0.6$ and $\pi_0 = 0.8$, respectively. In each plot, (b) is a zoom-in of (a) for better visualization.
- Fig. 6 is the error probability versus erasure probability plot.

Our observations from the numerical results are summarized below.

- For Approach 1, the iterative algorithm indeed yields thresholds that are solutions to the constrained optimization problem. For example, with $\pi_0 = 0.6$ and error probability constraint $\epsilon_k = 0.3$, the threshold obtained using Approach 1 is $\eta_k = 1.1812$ with corresponding error probabilities $P_{e0} = 0.2649$ and $P_{e1} = P_{e2} = 0.30$ (the left half of the last row of Table I). This is consistent with Fig. 4 (the corresponding values are marked on Fig. 4(b)). Similarly, with $\pi_0 = 0.8$ and $\epsilon_k = 0.2$, the minimum achievable $P_{e0} = 0.1775$ with the corresponding

threshold $\eta_k = 2.0311$ and local error probability $P_{ek} = 0.1866$. These values are marked on Fig. 5 and are consistent with those listed in Table I (right half of the last row).

- From Table I, it can be seen that, by comparing columns corresponding to P_{ek} and P_{e0} , smaller local sensor error probabilities typically result in larger error probability at the fusion center. In general, having a generous constraint on local sensor error probabilities (large ϵ_k) imposes less restriction on the admissible threshold pairs, which typically gives rise to smaller P_{e0} . In the extreme case, for example, when $\epsilon_k = 0.5$, the obtained thresholds will always coincide with that of Approach 4.
- The classical distributed detection (Approach 4) that minimizes error probability at the fusion center suffers significant performance loss in the event of a lost transmission. This can be illustrated using Fig. 4(b) along with Table III. At $\pi_0 = 0.6$, Approach 4 yields a globally minimum error probability $P_{e0} = 0.2574$ at the fusion center. However, if one of the transmission is lost, the error probability suffers a significant degradation to $P_{ek} = 0.3315$ (marked on the dash-dotted curve). Clearly, the constrained optimization approach is much more robust (a degradation from $P_{e0} = 0.2649$ to $P_{ek} = 0.30$). This effect is even more pronounced for the case of $\pi_0 = 0.8$. Approach 4 yields a fusion center error probability $P_{e0} = 0.1686$ (corresponding to the minimum point of the solid curve in Fig. 5(b)). However, if only one transmission reach the fusion center, the error probability becomes $P_{ek} = 0.2466$ which essentially renders this system useless – as the prior probability is $\pi_0 = 0.8$, the error probability should be capped at 0.2. This seemingly pathological result is due to the fact that the threshold design at local sensors for the classical distributed detection *always* assumes successful transmissions from other collaborating sensors.
- Approach 3 which optimizes local sensor performance does not have significant improvement when both transmissions are successful. From Table III and Fig. 4, for $\pi_0 = 0.6$, the minimum local sensor error probability is $P_{e1} = P_{e2} = 0.2945$. When both transmissions

are successful, the fusion center will have an error probability $P_{e0} = 0.2846$, which is only marginally better than the individual sensor's performance. This improvement is much smaller than that achieved by the proposed constrained minimization approach.

- For the erasure channel model approach, as the erasure probability δ_k approaches one, the obtained optimal local thresholds converge to that obtained using Approach 3 (minimizing the local error probabilities). This can be seen by comparing Tables II and III: the thresholds obtained using Approach 2 will approach that of Approach 3 as δ_k increases. This is expected since large δ_k implies that the channel is likely to break down, thus the local error probability will dominate the system performance. On the other hand, as the erasure probabilities approach zero, the obtained optimal local thresholds converge to those that minimize the error probability at Decoder 0 (corresponding to Approach 4). Intuitively, small δ_k indicates a high probability of successful transmissions of both U_1 and U_2 . Thus, the error probability at Decoder 0 would largely determine the system performance. The same behavior can be observed from Fig. 6, plotted for $\pi_0 = 0.6$, by looking at the two extreme points corresponding to $\delta_k = 0$ and $\delta_k = 1$. The associated error probabilities coincide with that of Approach 4 and 3 respectively.
- We have explored the intrinsic connections between Approach 1 and 2 in Section IV. Now we present numerical results to further elaborate the connections. Consider the case of $\pi_0 = 0.6$.
 - With $\epsilon_k = 0.32$, the corresponding $\lambda_k = 0.1652$. Set $\alpha_k = 0.1652$, we obtain two local minima that satisfy Eq. (18), as listed in Table IV. Since we want to choose the thresholds that minimize Q , it turns out that $\eta_1 = \eta_2 = 1.237$ (with $P_{e1} = P_{e2} = 0.3023$ and $P_{e0} = 0.2636$) is the optimal solution for the erasure channel model approach. But the constrained minimization approach results in the thresholds $\eta_1 = \eta_2 = 0.3609$ (with $P_{e1} = P_{e2} = 0.32$ and $P_{e0} = 0.2519$). From Table IV, it is clear that Eq.(20) does not

hold, i.e., the Q function corresponding to $P_{ek} = 0.32$ is not the smallest among the two. Hence in this particular setup, these two approaches do not have the same optimal solution.

- Now we examine a case when the two formulations share the same solution. Consider $\epsilon_k = 0.3$, the corresponding $\lambda_k = 0.4023$. Set $\alpha_k = 0.4023$, again there are two local minima as listed in Table V obtained using the erasure channel model approach. We notice that $\eta_1 = \eta_2 = 1.1812$ is the optimal solution for both approaches and it is easy to check that Eq.(20) is satisfied.
- In general, the rate of convergence of the proposed iterative algorithm depends on the initial values of local thresholds. Our simulations indicate that the proposed iterative algorithm converges very fast. For all the scenarios we have examined, convergence happens typically after several (< 10) iterations. For instance, the results in Table I were obtained after about six iterations on the average.

VI. CONCLUSIONS

In this paper, we developed robust signal processing techniques for distributed sensor networks applications. In particular, we presented a distributed multiple description quantization (DMDQ) framework for the design of sensor signaling in the presence of sensor failures/channel outages. Two approaches are proposed to address the DMDQ design using a two-sensor distributed detection problem. The first scheme is based on a constrained minimization approach; and a solution using Lagrangian multiplier is presented. The second imposes a discrete erasure channel model; we developed the channel-aware quantizer design that minimizes the average error probability. Iterative algorithms were constructed in search of the optimal thresholds. The intrinsic connections between the two approaches were explored. A design example was used to show how the DMDQ can be implemented in a real distributed detection problem, and to demonstrate its robust performance compared with the classical distributed detection approach

in the presence of possible transmission losses.

Our future work will address the application of the MD principle to sensor networks involving more than two sensors. The problem becomes conceivably much more complex as the number of objective functions grow exponentially as the number of sensors. Thus the constrained minimization approach may not be feasible. On the other hand, the erasure channel model essentially collapses the multi-objective functions into a single error probability, making it more appealing in dealing with large sensor networks. Thoroughly understanding the connection between the constrained minimization problem and the erasure channel model will provide valuable insight in how to choose the erasure channel model parameters.

APPENDIX A

PROOF OF THEOREM 1

Without loss of generality, we expand P_{e0} with respect to U_1 , and we get

$$\begin{aligned}
P_{e0} &= \pi_0 P(U_0 = 1|H_0) + \pi_1 P(U_0 = 0|H_1) \\
&= \pi_0 \sum_{U_1} \sum_{U_2} P(U_0 = 1|U_1, U_2) P(U_1, U_2|H_0) + \pi_1 \sum_{U_1} \sum_{U_2} P(U_0 = 0|U_1, U_2) P(U_1, U_2|H_1) \\
&= \pi_0 \sum_{U_2} P(U_2|H_0) [P(U_0 = 1|U_1 = 1, U_2) P(U_1 = 1|H_0) + P(U_0 = 1|U_1 = 0, U_2) P(U_1 = 0|H_0)] \\
&\quad + \pi_1 \sum_{U_2} P(U_2|H_1) [P(U_0 = 0|U_1 = 1, U_2) P(U_1 = 1|H_1) + P(U_0 = 0|U_1 = 0, U_2) P(U_1 = 0|H_1)] \\
&= \pi_0 P(U_1 = 1|H_0) A_1 - \pi_1 P(U_1 = 1|H_1) B_1 + C_1
\end{aligned}$$

where A_1 and B_1 are defined in Eqs. (13) and (14), and

$$C_1 = \pi_0 \sum_{U_2} P(U_2|H_0) P(U_0 = 1|U_1 = 0, U_2) + \pi_1 \sum_{U_2} P(U_2|H_1) P(U_0 = 0|U_1 = 0, U_2)$$

C_2 can be similarly defined by swapping the roles of U_1 and U_2 . Without loss of generality, we can rewrite P_{e0} as:

$$P_{e0} = \pi_0 P(U_k = 1|H_0) A_k - \pi_1 P(U_k = 1|H_1) B_k + C_k \quad (22)$$

for $k = 1, 2$.

From Eq. (1), the local error probabilities can be expressed as $P_{ek} = \pi_0 P_{fk} + \pi_1(1 - P_{dk})$, where $P_{fk} = P(U_k = 1|H_0)$ and $P_{dk} = P(U_k = 0|H_1)$. Thus the left-hand side of Eq. (4) becomes

$$\begin{aligned} & \frac{\partial P_{e0}}{\partial \tau_k} + \sum_{i=1}^2 \lambda_i \frac{\partial (P_{ei} - \epsilon_i)}{\partial \tau_k} \\ &= \pi_0 A_k \frac{\partial P_{fk}}{\partial \tau_k} - \pi_1 B_k \frac{\partial P_{dk}}{\partial \tau_k} + \lambda_k (\pi_0 \frac{\partial P_{fk}}{\partial \tau_k} - \pi_1 \frac{\partial P_{dk}}{\partial \tau_k}) \\ &= \pi_0 A_k \frac{\partial P_{fk}}{\partial \tau_k} - \pi_1 B_k \tau_k \frac{\partial P_{fk}}{\partial \tau_k} + \lambda_k (\pi_0 \frac{\partial P_{fk}}{\partial \tau_k} - \pi_1 \tau_k \frac{\partial P_{fk}}{\partial \tau_k}) \end{aligned} \quad (23)$$

where we have used the fact that $\frac{dP_{dk}}{dP_{fk}} = \tau_k$, and τ_k is the LR threshold for the k th sensor.²

Set (23) equal to zero, we have $\tau_k = \frac{\pi_0(A_k + \lambda_k)}{\pi_1(B_k + \lambda_k)}$. Eqs. (9-12) follow by directly applying the Kuhn-Tucker theorem for the two cases $\lambda_k = 0$ and $\lambda_k > 0$ separately. Thus, Theorem 1 is proved.

APPENDIX B

PROOF OF THEOREM 2

Similar to the proof in Appendix A, P_{e0} can be expanded with respect to the individual decision rules, and we get, for $k = 1, 2$,

$$P_{e0} = \int_{X_k} P(U_k = 1|X_k) [\pi_0 A_k P(X_k|H_0) - \pi_1 B_k P(X_k|H_k)] dX_k + C_k \quad (24)$$

where C_k has no effect on the decision rule at sensor k .

Similarly, the error probability at the k th sensor can be expanded as

$$\begin{aligned} P_{ek} &= \pi_0 P(U_k = 1|H_0) + \pi_1 P(U_k = 0|H_1) \\ &= \int_{X_k} P(U_k = 1|X_k) [\pi_0 P(X_k|H_0) - \pi_1 P(X_k|H_1)] dX_k + \pi_1 \end{aligned} \quad (25)$$

²This is the property of the receiver operating characteristics (ROC) curve for a likelihood ratio test. The threshold corresponding to the (P_f, P_d) pair equals the slope of the ROC curve at that point.

Thus, using sensor 1 as an illustration, the average error probability P_e can be written as, from Eqs. (15), (24), and (25),

$$\begin{aligned}
P_e &= (1 - \delta_1)(1 - \delta_2)P_{e0} + \delta_2(1 - \delta_1)P_{e1} + \delta_1(1 - \delta_2)P_{e2} + \min\{\pi_0, \pi_1\}\delta_1\delta_2 \\
&= (1 - \delta_1)(1 - \delta_2) \int_{X_1} P(U_1 = 1|X_1) [\pi_0 A_1 P(X_1|H_0) - \pi_1 B_1 P(X_1|H_1)] dX_1 \\
&\quad + \delta_2(1 - \delta_1) \int_{X_1} P(U_1 = 1|X_1) [\pi_0 P(X_1|H_0) - \pi_1 P(X_1|H_1)] dX_1 + D_1 \\
&= \int_{X_1} P(U_1 = 1|X_1) F_1 dX_1 + D_1 \tag{26}
\end{aligned}$$

where

$$\begin{aligned}
D_1 &= C_1 + \delta_2(1 - \delta_1)\pi_1 + \delta_1(1 - \delta_2) \left(\int_{X_2} P(U_2 = 1|X_2) [\pi_0 P(X_2|H_0) - \pi_1 P(X_2|H_1)] dX_2 + \pi_1 \right) \\
F_1 &= (1 - \delta_1)(1 - \delta_2) [\pi_0 A_1 P(X_1|H_0) - \pi_1 B_1 P(X_1|H_1)] + \delta_2(1 - \delta_1) [\pi_0 P(X_1|H_0) - \pi_1 P(X_1|H_1)]
\end{aligned}$$

As D_1 is independent of the quantizer rule at sensor 1, we need only to minimize the first term in Eq. (26) with respect to the local decision rule for sensor 1. Thus, the optimum local decision rule for sensor 1 is as follows.

$$P(U_1 = 1|X_1) = \begin{cases} 1, & \text{if } F_1 \geq 0 \\ 0, & \text{otherwise} \end{cases} \tag{27}$$

This is equivalent to the decision rule specified by Eqs. (16-17) for $k = 1$. The optimum quantizer rule for sensor 2 can be similar established. This completes the proof of Theorem 2.

REFERENCES

- [1] V.K. Goyal, "Multiple description coding: compression meets the network," *IEEE Signal Processing Magazine*, vol. 19, pp. 74–93, Sep. 2002.
- [2] 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.
- [3] V. Vaishampayan, "Design of multiple description scalar quantizer," *IEEE Trans. on Information Theory*, vol. 39, pp. 821–834, May 1993.
- [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 II — Fundamentals," *Proceedings of the IEEE*, vol. 85, no. 1, pp. 54–63, Jan. 1997.
- [6] J.N. Tsitsiklis, "On threshold rules in decentralized detection," in *Proc. 25th IEEE Conf. on Decision and Control*, Athens, Greece, 1986, vol. 1, pp. 232–236.
- [7] B. Chen and P.K. Willett, "Channel optimized binary quantizers for distributed sensor networks," in *Proc. IEEE Intl. Conf. Acoust. Speech, Sig. Proc. (ICASSP'2004)*, Montreal, Canada, May 2004.
- [8] B. Chen and P.K. Willett, "On the optimality of likelihood ratio test for local sensor decisions in the presence of non-ideal channels," *to appear in the IEEE Trans. Information theory*, vol. 51, Feb. 2005.
- [9] B. Liu and B. Chen, "Joint source-channel coding for distributed sensor networks," in *Proc. 2004 Annual Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, Nov. 2004.
- [10] Masanao Aoki, *Introduction to Optimization Techniques*, The Macmillan Company, New York, NY, 1971.
- [11] I.Y. Hoballah and P.K. Varshney, "Distributed Bayesian signal detection," *IEEE Trans. Inform. Theory*, vol. 35, pp. 995–1000, Sept. 1989.
- [12] Pramod K. Varshney, *Distributed Detection and Data Fusion*, Springer, New York, 1997.
- [13] Yugang Zhou and Wai-Yip Chan, "Multiple description quantizer design using a channel optimized quantizer approach," in *Proc. of CISS'2004*, Princeton, NJ, March 2004.
- [14] B. Chen, "On the local sensor signaling for inference centered wireless sensor networks," *presented in the 2nd IEEE Upstate New York Workshop on Sensor Networks*, Oct. 2003.

TABLE I

THRESHOLDS AND ERROR PROBABILITIES OBTAINED USING APPROACH 1

$\pi_0 = 0.6$					$\pi_0 = 0.8$				
ε_k	η_k	P_{ek}	P_{e0}	λ_k	ε_k	η_k	P_{ek}	P_{e0}	λ_k
0.33	0.2629	0.33	0.2574	0.0177	0.25	0.8474	0.2466	0.1686	0
0.32	0.3609	0.32	0.2591	0.1652	0.21	1.1737	0.21	0.1744	0.1788
0.31	1.2895	0.3047	0.2632	0.0001	0.2	2.0292	0.1866	0.1775	0.001
0.30	1.1812	0.30	0.2649	0.4023	0.19	2.0292	0.1866	0.1775	0.001

TABLE II

THRESHOLDS AND ERROR PROBABILITIES OBTAINED USING APPROACH 2

$\pi_0 = 0.6$					$\pi_0 = 0.8$				
δ_k	η_k	P_{ek}	P_{e0}	α_k	δ_k	η_k	P_{ek}	P_{e0}	α_k
0.0174	0.2629	0.33	0.2574	0.0177	0.01	0.8667	0.2438	0.1686	0.0101
0.2869	1.1812	0.3	0.2649	0.4023	0.4	1.9757	0.1864	0.1777	0.8
0.5	1.0971	0.2973	0.2685	1.0	0.5	1.9616	0.1863	0.1778	1
0.7	1.018	0.2955	0.2738	2.3333	0.7	1.9324	0.1863	0.1780	2.3333
0.9	0.9417	0.2946	0.2807	9.0	0.9	1.902	0.1862	0.1784	9.0

TABLE III

THRESHOLDS AND ERROR PROBABILITIES OBTAINED USING APPROACHES 3 AND 4

Approach 3	η_k	$P_{ek,3}$	$P_{e0,3}$
$\pi_0 = 0.6$	0.9059	0.2945	0.2846
$\pi_0 = 0.8$	1.8863	0.1862	0.1787
Approach 4	η_k	$P_{ek,4}$	$P_{e0,4}$
$\pi_0 = 0.6$	0.2495	0.3315	0.2574
$\pi_0 = 0.8$	0.8474	0.2466	0.1686

TABLE IV

LOCAL MINIMA OBTAINED BY THE ERASURE CHANNEL MODEL APPROACH, $\epsilon_k = 0.32$, $\alpha_k = \lambda_k = 0.1652$

δ_k	η_k	P_{ek}	P_{e0}	Q
0.1418	1.237	0.3023	0.2636	0.3635
0.1418	0.3609	0.32	0.2591	0.3648

TABLE V

LOCAL MINIMA OBTAINED BY THE ERASURE CHANNEL MODEL APPROACH, $\epsilon_k = 0.3$, $\alpha_k = \lambda_k = 0.4023$

δ_k	η_k	P_{ek}	P_{e0}	Q
0.2869	0.4784	0.3101	0.2645	0.514
0.2869	1.1812	0.3	0.2649	0.5063

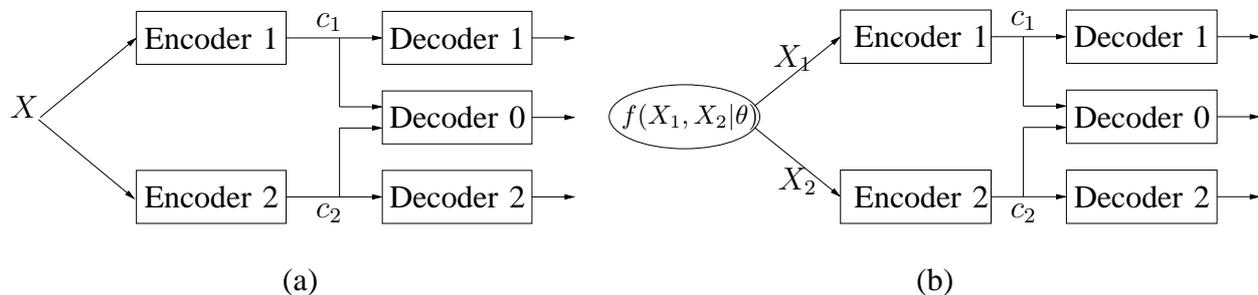


Fig. 1. Comparison between (a) conventional MD, and (b) distributed MD for sensor network applications. In (a), Encoders 1 and 2 have access to the same observation X . In (b), Encoder 1 encodes X_1 without access to X_2 while Encoder 2 encodes X_2 without access to X_1 .

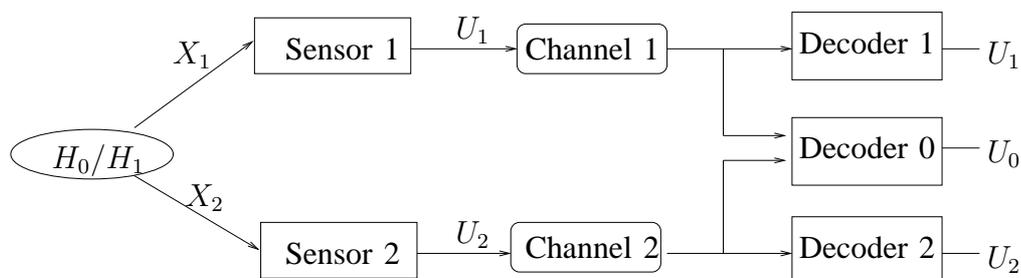


Fig. 2. A two-sensor parallel fusion network with possible transmission failures.

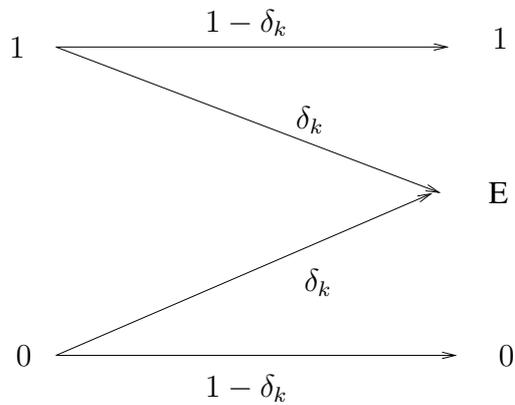


Fig. 3. A discrete memoryless erasure channel model for the channel between sensor k and the fusion center.

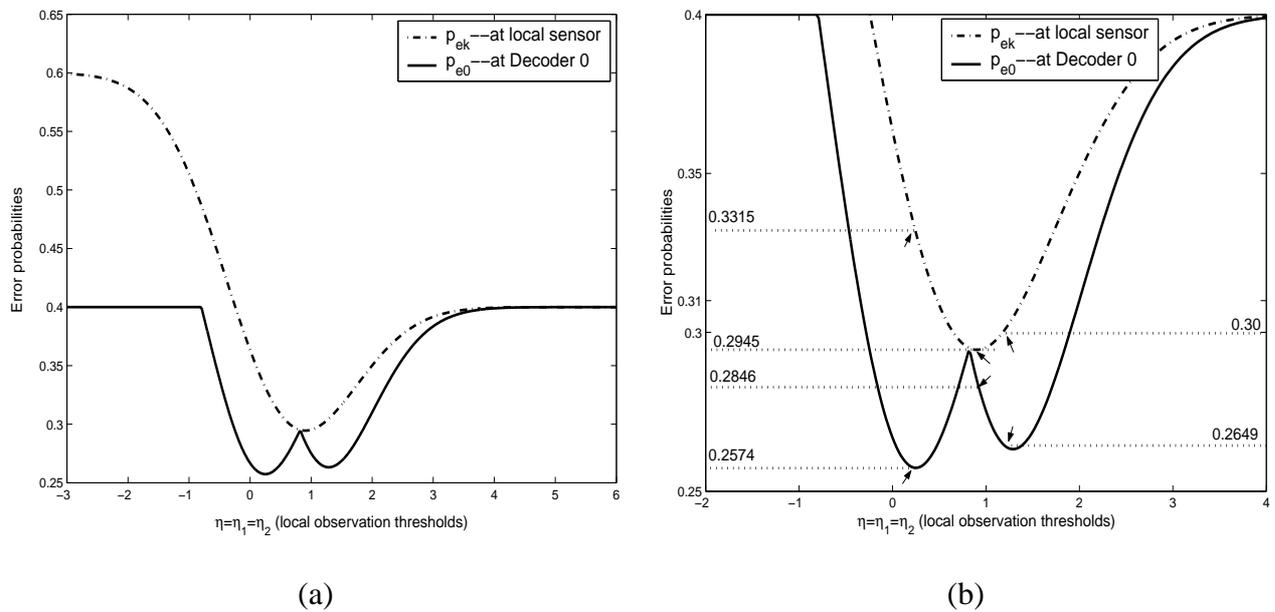


Fig. 4. Analytically calculated error probability versus threshold plot for $\pi_0 = 0.6$; (b) is a zoom-in of (a).

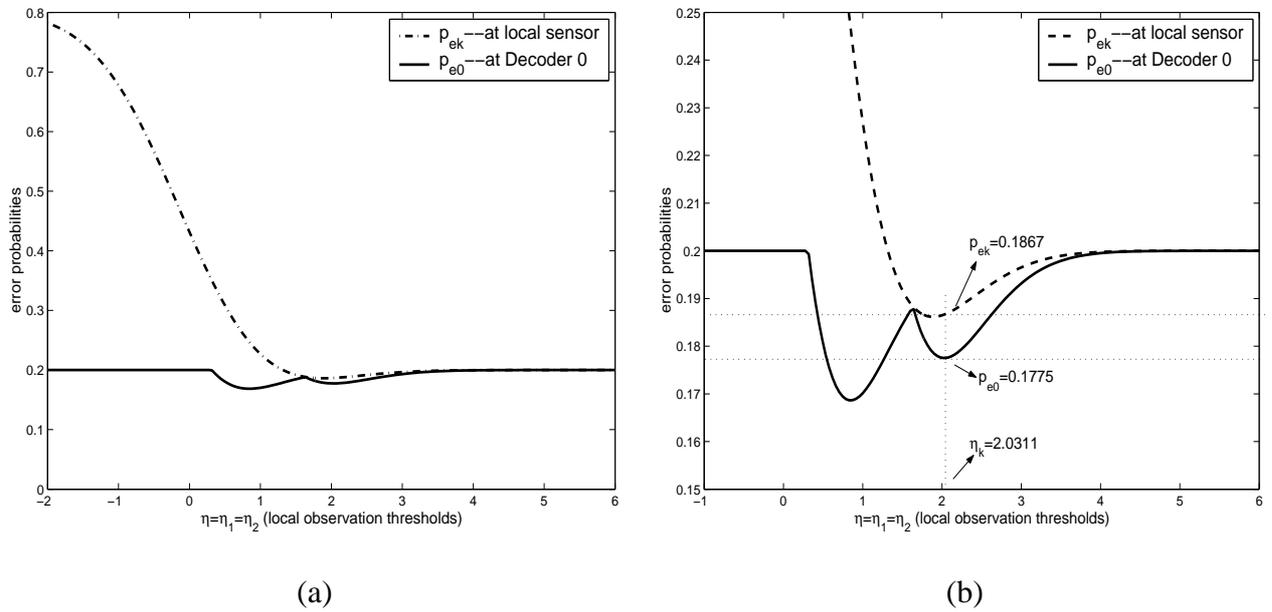


Fig. 5. Analytically calculated error probability versus threshold plot for $\pi_0 = 0.8$; (b) is a zoom-in of (a).

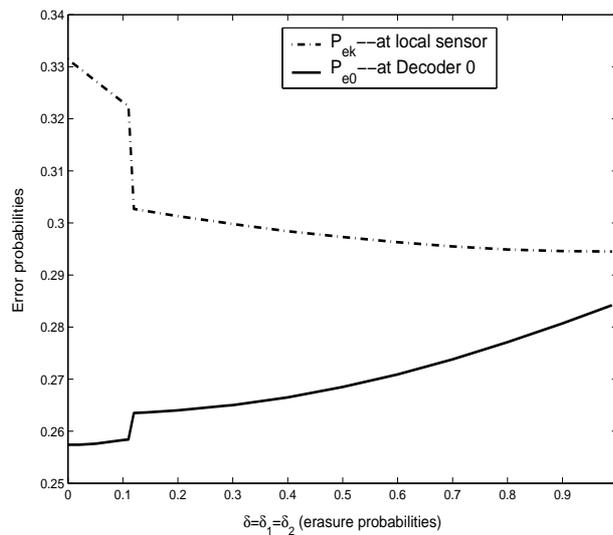


Fig. 6. Error probability versus erasure probability plot for $\pi_0 = 0.6$ obtained using the channel-aware quantization for the erasure channel model.