

Exponent Trade-off for Hypothesis Testing Over Noisy Channels

Nir Weinberger
IDSS and LIDS
MIT
Cambridge, MA, USA
nirw@mit.edu

Yuval Kochman
School of CSE
HUJI
Jerusalem, Israel
yuvalko@cs.huji.ac.il

Michèle Wigger
Comm. and Elect. Dept.
Telecom ParisTech
Paris, France
michele.wigger@telecom-paristech.fr

Abstract—The distributed hypothesis testing (DHT) problem is considered, in which the joint distribution of a pair of sequences present at separated terminals, is governed by one of two possible hypotheses. The decision needs to be made by one of the terminals (the “decoder”). The other terminal (the “encoder”) uses a noisy channel in order to help the decoder with the decision. This problem can be seen as a generalization of the side-information variant of the DHT problem, where the rate-limited link is replaced by a noisy channel. A recent work by Salehkalaibar and Wigger has derived an achievable Stein exponent for this problem, by employing concepts from the DHT scheme of Shimokawa et al., and from unequal error protection coding for a single special message. In this work we extend the view to a trade-off between the two error exponents, additionally building on multiple codebooks and two special messages with unequal error protection. As a by product, we also present an achievable exponent trade-off for a rate-limited link, which generalizes Shimokawa et al..

I. INTRODUCTION

The distributed hypothesis testing (DHT) problem enjoys a renewed interest in recent years. In this problem there are two hypotheses regarding the distribution of a pair of sequences X^n and Y^n - either independently and identically distributed (i.i.d.) according to P_{XY}^0 or P_{XY}^1 . The two sequences are observed by terminals (encoders), which communicate to a central decision function over links of rates R_X and R_Y . The decision function outputs one of the two hypotheses, and as in any hypothesis testing problem, there are two error probabilities at play. The dependence of these error probabilities on the rates, and specifically, their exponential decay rate (error exponents), is of interest.

The general DHT problem turns out to be challenging, and thus it has been mostly studied under one or more of the following simplifying assumptions:

1) Side-information (SI) setting: The terminal observing Y^n is collocated with the decision function, so $R_Y = \infty$. Then, the sequence Y^n serves as SI.

2) Stein problem: The error probability under $P_{X,Y}^0$ is not required to decay exponentially. Thus, we are only interested in the exponent under $P_{X,Y}^1$.

3) Testing against conditional independence: The side information decomposes as $Y = (\tilde{Y}, Z)$ and $P_{X,Z}^1$ coincides with $P_{X,Z}^0$ and $P_{\tilde{Y}|X,Z}^1$ coincides with $P_{\tilde{Y}|Z}^0$.

Indeed, only when all three assumptions hold, the optimal exponent is known. The state-of-the-art approach is a quantization-and-binning (Q&B) encoder [1], much like lossy compression with SI (the Wyner-Ziv problem). It was shown to be optimal when all three assumptions hold [2], and in fact when a stronger independence assumption holds a simpler quantization-only scheme where the encoder ignores the SI was already shown to be optimal in [3]. The scheme proposed in [1] is heavily tailored to the Stein problem, in the sense that any detected atypicality leads to declaring P_{XY}^1 . For example, if the encoder observes X^n which is not P_{XY}^0 -typical it sends a “special message” under which the decoder completely ignores Y^n and decides the hypothesis is that of P_{XY}^1 . Recent works aim at further generalizing the results. For example, in [4], an achievable exponent trade-off for any two hypotheses in the SI setting is derived; the resulting exponents is conjectured to be optimal for any random-coding Q&B scheme.

In this work, we consider the problem of hypothesis testing over noisy channels. It can be seen as an extension of the SI setting, where the rate-limited link (“bit-pipe”) from the encoder to the decoder is replaced by a noisy channel. Thus, it is much like the extension from a lossy source coding problem to joint source-channel coding (JSCC) problem. For this problem, the optimal Stein exponent for testing against conditional independence was derived in [5], and an achievable exponent to the Stein problem was derived in [6]. The approach of [6] is based upon mapping a Q&B codebook for typical X^n sequences to a channel codebook. The “special message” of [1] is mapped to a special channel message, with an unequal error protection (UEP) mechanism à la [7]. Quite remarkably, when this special message is ignored, the exponents of [1] with a rate matching the channel capacity are achievable; it is possible, however, that error events related to the special message will reduce that exponent.

Here, we extend the results from the Stein setting to an exponent trade-off. As a preliminary step, we propose (Section III) a refined coding scheme for the *clean* channel case. Unlike for the Stein setting, in our new scheme the encoder does not simply assign all P_0 -atypical X^n sequences to P_{XY}^1 . Instead, a different Q&B scheme with a tailored coding rate and specific detection regions is chosen for each

possible type of the observation X^n . Specializing our scheme to a degenerate decision region for all types $Q_X \neq P_X^0$ recovers the scheme in [1] for the Stein setting. This scheme is not meant to be competitive with [4], but rather reasonably amenable to an extension to the *noisy* channel case, which is the main contribution of this paper (Section IV). In this extended scheme, binning is implicit through the channel as in [8] and multiple codebooks are used as proposed by Csiszár [9] for lossy JSCC.

II. NOTATION AND PROBLEM DEFINITION

Notation conventions: Deterministic vectors and scalars will be denoted by lower case letters, e.g., $x^n = (x_1, \dots, x_n)$, and random vectors and scalars by upper case letters $X^n = (X_1, \dots, X_n)$. In general, distributions related to the source will be denoted by P, Q and types related to the channel by Φ, Γ . Joint, conditional and marginal distributions are denoted as $Q_{X,Y} = Q_X Q_{Y|X}$, and $Q_Y = [Q_{X,Y}]_Y$. The *type class* of a type Q_X at blocklength n will be denoted by $\mathcal{T}_n(Q_X)$. The set of all type classes of vectors from \mathcal{X}^n will be denoted by $\mathcal{P}_n(\mathcal{X})$, and the probability simplex will be denoted by $\overline{\mathcal{P}}(\mathcal{X})$ which is the closure of $\mathcal{P}(\mathcal{X}) \triangleq \bigcup_{n=1}^{\infty} \mathcal{P}_n(\mathcal{X})$. The *Q-shell* of x^n for $Q_{Y|X}$ will be denoted by $\mathcal{T}_n(Q_{Y|X}, x^n)$. For a given $Q_X \in \mathcal{P}_n(\mathcal{X})$, the conditional types $Q_{Y|X}$ such that $\mathcal{T}_n(Q_{Y|X}, x^n)$ is not empty when $x^n \in \mathcal{T}_n(Q_X)$ will be denoted by $\mathcal{P}_n(\mathcal{Y}, Q_X)$. The empirical type of a vector x^n will be denoted by \hat{Q}_{x^n} . Information-theoretic quantities will be denoted as a function of the distribution, e.g., $H(Q_{XY}), H(Q_{Y|X}|Q_X)$ for entropy and conditional entropy, $I(Q_{XY}), I(Q_{XY|Z}|Q_Z)$ for mutual information and conditional mutual information, and $D(Q_X||P_X), D(Q_{X|Z}||P_{X|Z}|Q_Z)$, for KL divergence and conditional KL divergence, respectively. The cardinality of a set \mathcal{A} will be denoted by $|\mathcal{A}|$. The shorthands $[t]_+ \triangleq \max\{t, 0\}$ and $[k] \triangleq \{1, 2, \dots, k\}$ will be utilized. Integer constraints on large numbers, e.g., 2^{nR} , will be ignored as they are inconsequential. Finally, the $\log(\cdot)$ function is with respect to base 2 and we use the notation $a_n \leq b_n$ to indicate that $\lim_{n \rightarrow \infty} \log(a_n/b_n) \leq 1$.

In the SI-DHT problem, the transmitter observes source sequence X^n and the receiver a source sequence Y^n . The hypotheses are given by:

$$\mathcal{H} = i: (X^n, Y^n) \text{ i.i.d. } \sim P_{XY}^i, \quad i \in \{0, 1\}, \quad (1)$$

for $P_{XY}^0, P_{XY}^1 \in \overline{\mathcal{P}}(\mathcal{X} \times \mathcal{Y})$. The transmitter can communicate with the receiver over m uses of a discrete memoryless channel $(\mathcal{W}, \mathcal{V}, P_{V|W})$ where \mathcal{W} (resp. \mathcal{V}) denotes the input (resp. output) alphabet, $|\mathcal{W}|, |\mathcal{V}| < \infty$. If we take $m = \lfloor \rho n \rfloor$ for a fixed ρ , we can index both source and channel sequences using n . A scheme for this problem comprises of a (possibly stochastic) encoding function

$$f^{(n)}: \mathcal{X}^n \rightarrow \mathcal{W}^m \quad (2)$$

and a decoding function

$$g^{(n)}: \mathcal{V}^m \times \mathcal{Y}^n \rightarrow \{0, 1\}. \quad (3)$$

Specifically, the transmitter feeds inputs $W^m = f^{(n)}(X^n)$ to the channel; the decoder observes the outputs V^m , where for a given input $W^m = w^m$,

$$V^m \sim \prod_{t=1}^m \Gamma_{V|W}(v_t|w_t). \quad (4)$$

Based on this sequence and the source sequence Y^n , the decoder produces an estimate $\hat{\mathcal{H}} = g^{(n)}(V^m, Y^n)$.

Denote for brevity $\bar{i} = 1 - i$ for $i \in \{0, 1\}$. The error probabilities at a given blocklength n are given by

$$\epsilon_n^i \triangleq \Pr[\hat{\mathcal{H}} = \bar{i} | \mathcal{H} = i], \quad i = 0, 1, \quad (5)$$

where the dependence upon the source and channel distributions, ρ , the encoder and the decoder remain implicit. An exponent pair (θ_0, θ_1) is said to be achievable, if there exists a sequence of encoding and decoding functions $\{(f^{(n)}, g^{(n)})\}$ such that the corresponding error probabilities satisfy

$$-\overline{\lim}_{n \rightarrow \infty} \frac{1}{n} \log \epsilon_n^i \geq \theta_i, \quad i = 0, 1. \quad (6)$$

The optimal exponent trade-off region is the closure of all achievable pairs; our goal is to find an achievable region, i.e., an inner bound on this region. An extreme version of this problem is when only one of the probabilities is required to decay exponentially, while the other only has to only decay sub-exponentially. The optimal exponents for this problem are known as Stein exponent, and are the corner points of the achievable region.¹

III. CLEAN CHANNEL

In this section, we propose an achievable scheme to the classic DHT scenario in which X^n must be compressed at rate R , and the decoder receives the compressed message without any noise. It can be noted that this scenario falls into the framework of Section II when the channel is a clean binary channel, i.e. $\mathcal{W} = \mathcal{V} = \{0, 1\}$, $\Gamma_{V|W}(v|w) = 1$ iff $v = w$ and $\rho = R$. Nonetheless, it will be more convenient to denote the encoding operation by $f^{(n)}: \mathcal{X}^n \rightarrow [2^{nR}]$.

The scheme is based on [1], for which the encoding is based on Q&B, and decoding is based on a two stage process. In the first decoding stage, given the bin index and the SI, the decoder decides on the specific quantized source vector within the bin. This is done by a *universal* metric, as the decoder is unaware of the hypothesis governing the source, and thus cannot use the optimal rule. In the second stage, a test decides on one of the hypotheses, based only on the joint type of the quantized source vector and the SI.

More formally, our proposed scheme is defined as follows. Let $\delta > 0$ be an arbitrary small constant.

- *Parameters:* Choose a quantization alphabet \mathcal{S} , where, w.l.o.g., $|\mathcal{S}| \geq |\mathcal{X}|$. For each $Q_X \in \mathcal{P}_n(\mathcal{X})$, choose a test channel $P_{S|X}(Q_X) \in \mathcal{P}_n(\mathcal{S}, Q_X)$, and choose detection regions $\mathcal{A}^0(Q_X) \subseteq \mathcal{P}(\mathcal{S} \times \mathcal{Y})$ and $\mathcal{A}^1(Q_X) \triangleq \mathcal{P}(\mathcal{S} \times \mathcal{Y}) \setminus \mathcal{A}^0(Q_X)$.

¹In ordinary hypothesis testing, for which strong converse exists, the Stein exponent is also the optimal exponent under the more stringent constraint that the second probability only has to remain bounded away from 1. For DHT problems, there might be a distinction.

• **Codebooks construction:** For each $Q_X \in \mathcal{P}_n(\mathcal{X})$, draw the codebook $\mathcal{C}_n(Q_X)$ uniformly at random over the set of codebooks of rate $I(Q_X P_{S|X}(Q_X)) + \delta$ that have constant type $Q_S = [Q_X P_{S|X}(Q_X)]_S$ and where for each $x^n \in \mathcal{T}_n(Q_X)$ there exists at least one codeword s^n such that $(x^n, s^n) \in \mathcal{T}_n(Q_X P_{S|X}(Q_X))$. For a given type Q_X of the source x^n , let $h_{\text{quan}} : \mathcal{T}_n(Q_X) \rightarrow \mathcal{C}_n(Q_X)$ denote the quantization operation that maps x^n to one of the codewords s^n in $\mathcal{C}_n(Q_X)$ satisfying $(x^n, s^n) \in \mathcal{T}_n(Q_X P_{S|X}(Q_X))$.² Given the source type Q_X , let further $h_{\text{bin}} : \mathcal{C}_n(Q_X) \rightarrow [2^{nR}]$ denote the random binning operation that maps each codeword of $\mathcal{C}_n(Q_X)$ uniformly at random to an index in $[2^{nR}]$, in case $I(Q_X P_{S|X}(Q_X)) \geq R - \delta$, and maps each codeword to a different index in $[2^{nR}]$, in case $I(Q_X P_{S|X}(Q_X)) < R - \delta$. The scheme operates on $(x^n, y^n) \in \mathcal{T}_n(Q_{X,Y})$ as follows:

- **Encoding:** The encoder sends an index of Q_X in an enumeration of $\mathcal{P}_n(\mathcal{X})$ and then sends $f(x^n) = h_{\text{bin}}(h_{\text{quan}}(x^n))$.
- **Decoding:** Using the received message, the decoder is aware of the type Q_X . It then performs the following steps:

- 1) Chooses a codeword \tilde{s}^n which minimizes $H(\hat{Q}_{\tilde{s}^n|y^n}|\hat{Q}_{y^n})$ over all \tilde{s}^n with $h_{\text{bin}}(\tilde{s}^n) = f(x^n)$.
- 2) If $\hat{Q}_{\tilde{s}^n,y^n} \in \mathcal{A}_n^0(Q_X)$ it outputs $\mathcal{H} = 0$, and otherwise it outputs $\mathcal{H} = 1$.

Theorem 1. *Given a choice of parameters, the exponent pair (θ^0, θ^1) is achievable, where $\theta^1 \triangleq \min(\theta^{\text{std},i}, \theta^{\text{bin},i})$ is given by:*

$$\theta^{\text{std},i} \triangleq \min_{\substack{Q_{S,X,Y}: \\ Q_{S|X}=P_{S|X}(Q_X) \\ Q_{S,Y} \in \mathcal{A}_n^i(Q_X)}} D(Q_{S,X,Y} \| P_{S,X,Y}^i), \quad (7)$$

$$\theta^{\text{bin},i} \triangleq \min_{\substack{Q_{S,X,Y}: \\ Q_{S|X}=P_{S|X}(Q_X) \\ I(Q_{S,X}) \geq R}} \left\{ D(Q_{S,X,Y} \| P_{S,X,Y}^i) + \min_{Q'_{S,Y} \in \mathcal{D}(Q_{X,Y}) \cap \mathcal{A}^i(Q_X)} [I(Q'_{S,Y}) - I(Q_{S,X}) + R]_+ \right\}, \quad (8)$$

where $P_{S,X,Y}^i = P_{S|X}(P_X^i)P_{X,Y}$ and

$$D(Q_{X,Y}) \triangleq \left\{ Q'_{S,Y} \in \mathcal{P}(\mathcal{S} \times \mathcal{Y}) : Q'_S = [Q_X P_{S|X}(Q_X)]_S, Q'_Y = Q_Y, H(Q'_{S|Y}|Q'_Y) \leq H(Q_{S|Y}|Q_Y) \right\}. \quad (9)$$

Proof outline: Since the number of types is polynomial in n , $|\mathcal{P}_n(\mathcal{X})| \leq 2^{n\delta}$ for all n sufficiently large and so the encoding rate is less than R .³ The type covering implied by the codebook construction is possible due to the type-covering lemma [10, Lemma 9.1].

We now turn to bound the error exponent assuming that $\mathcal{H} = i$. Let S^n denote the quantization codeword picked at

²For simplicity, the dependence on the type and codebook is omitted.

³For Q_X such that either $\mathcal{A}_n^0(Q_X)$ or $\mathcal{A}_n^i(Q_X)$ is empty, sending the bin index is redundant since the decoder decision is fixed. However, sending it causes no harm as the total rate does not increase.

the encoder and \tilde{S}^n the codeword picked at the decoder. We decompose the error event into the two disjoint events

$$\mathcal{E}^{\text{std},i} := \{\tilde{S}^n = S^n \text{ and } \mathcal{H} = \bar{i}\}, \quad (10)$$

and

$$\mathcal{E}^{\text{bin},i} := \{\tilde{S}^n \neq S^n \text{ and } \mathcal{H} = \bar{i}\}. \quad (11)$$

The probability of the first event can be upper bounded by a standard method-of-types argument as

$$\Pr[\mathcal{E}^{\text{std},i} | \mathcal{H} = i] \leq \sum_{\substack{Q_{S,X,Y}: \\ Q_{S|X}=P_{S|X}(Q_X) \\ Q_{S,Y} \in \mathcal{A}_n^i(Q_X)}} 2^{-nD(Q_{S,X,Y} \| P_{S,X,Y}^i)} \quad (12)$$

which leads to $\theta^{\text{std},i}$.

For the second event, we have:

$$\begin{aligned} & \Pr[\mathcal{E}^{\text{bin},i} | \mathcal{H} = i] \\ &= \sum_{Q_{S,X,Y}} \Pr[\hat{Q}_{S^n X^n Y^n} = Q_{S,X,Y} | \mathcal{H} = i] \\ & \quad \cdot \Pr[\mathcal{E}^{\text{bin},i} | \hat{Q}_{S^n X^n Y^n} = Q_{S,X,Y}, \mathcal{H} = i]. \end{aligned} \quad (13)$$

The first term in each summand is positive only if $Q_{S|X} = P_{S|X}(Q_X)$ and the second term is positive only if $I(Q_X P_{S|X}(Q_X)) > R$, because otherwise there is no binning and event $\mathcal{E}^{\text{bin},i}$ has zero conditional probability. The sum can thus be constrained to types $Q_{S,X,Y}$ satisfying these two conditions. For such types, the first conditional probability in (13) can be bounded as in (12) and the second as:

$$\begin{aligned} & \Pr[\mathcal{E}^{\text{bin},i} | \hat{Q}_{S^n X^n Y^n} = Q_{S,X,Y}, \mathcal{H} = i] \\ & \leq \sum_{S^n \neq \tilde{S}^n} \Pr[h_{\text{bin}}(S^n) \neq h_{\text{bin}}(\tilde{S}^n), \\ & \quad H(\hat{Q}_{S^n|Y^n}|\hat{Q}_{Y^n}) \leq H(\hat{Q}_{\tilde{S}^n|Y^n}|\hat{Q}_{Y^n}), \\ & \quad \hat{Q}_{S^n Y^n} \in \mathcal{A}^i(Q_X) \\ & \quad | \hat{Q}_{S^n X^n Y^n} = Q_{S,X,Y}, \mathcal{H} = i] \quad (14) \\ & \leq 2^{n(I(Q_{S,X})+\delta)} \cdot 2^{-nR} \cdot \sum_{Q'_{S,Y} \in \mathcal{D}(Q_{X,Y})} 2^{-nI(Q'_{S,Y})}, \quad (15) \end{aligned}$$

where the sum in (14) is over all codewords except the one chosen by the encoder. The asymptotic inequality (15) holds because there are $2^{n(I(Q_{S,X})+\delta)}$ codewords and each of them is uniform over the type class $\mathcal{T}_n(Q_S)$ and uniformly assigned to one of 2^{nR} bins. The proof is then concluded by clipping the upper bound by 1 and then taking $n \rightarrow \infty$ and $\delta \rightarrow 0$. ■ Theorem 1 may be particularized by special choices of $\{\mathcal{A}^i(Q_X)\}_{Q_X \in \mathcal{P}(Q_X)}$. Two natural choices are:

1) The likelihood ratio test (LRT) between $P_{X,Y}^0$ and $P_{X,Y}^1$: For some choice of a threshold $t \in \mathbb{R}$, and all $Q_X \in \mathcal{P}(Q_X)$

$$\mathcal{A}^0(Q_X) \triangleq \{D(Q_{S,Y} \| P_{S,Y}^1) - D(Q_{S,Y} \| P_{S,Y}^0) \geq t\}. \quad (16)$$

2) Typicality-based set: For all Q_X such that $D(Q_X \| P_X^0) \leq \epsilon_n$

$$\mathcal{A}^0(Q_X) = \left\{ Q_{S,Y} : D(Q_{S,Y} \| [P_{X,Y}^0 P_{S|X}]_{S,Y}) \leq \epsilon_n \right\} \quad (17)$$

where $\epsilon_n \rightarrow 0$, and $\mathcal{A}^0(Q_X) = \phi$ otherwise. In this case, the scheme reduces to the scheme of [1]. With this choice, the binning exponent dominates the standard one and the bound of [1] is recovered.

Let us elucidate the rationale behind these choices. If the word chosen in the first stage is the correct codeword, then the LRT is optimal. It should be taken into account, however, that there is a probability (of finite exponent) that the second stage test is performed on the wrong codeword. Attributing “doubtful” cases to the hypothesis that needs better “protection”, can increase the total exponent. In the extreme Stein case, the test in [1] only attributes to P^0 the typical set, i.e., types that must be included in order to have non-trivial error probability under P^0 .

Finding a test that strikes an optimal balance outside the Stein case seems to be complicated. However, a family of tests which includes the LRT and the typicality test in [1] as special cases, may be obtained by enforcing both a threshold on the difference of type-class likelihoods and on the absolute likelihood:

$$t_{\text{diff}} \leq D(Q_{S,Y} \| P_{S,Y}^1) - D(Q_{S,Y} \| P_{S,Y}^0) \quad (18a)$$

$$t_{\text{abs}} \geq D(Q_{S,Y} \| P_{S,Y}^0), \quad (18b)$$

or a similar absolute criterion with respect to P^1 .

IV. NOISY CHANNEL

In this section, we move on to consider the noisy channel case. We focus on $\rho = 1$; the extension to general ρ is straightforward. In case of a noisy channel, the codewords of the per-type source codebooks $\mathcal{C}_n(Q_X)$ are mapped into channel codewords rather than an index, where a different codebook can be used for each Q_X . However, here the rate is not explicitly constrained, and is a parameter to be designed. For our analysis it turns out that there is no loss in:

1) Choosing a channel codebook with a single codeword for all types Q_X that are attributed entirely to one of the hypotheses. Hence, we will have two special input sequences, $m^{n,0}$ and $m^{n,1}$, corresponding to each of the hypotheses.

2) Choosing the codebook rate $I(Q_X P_{S|X}(Q_X))$ for all other types.

Notice that our scheme does not use explicit binning, i.e., each of the quantizer indices is mapped to an independently chosen channel codeword. In this case, the decoder is effectively observing channel-induced “bins” and then chooses a codeword within the bin using the SI. Such strategy was coined “virtual binning” in [8] and used for JSCC for broadcast channels with side information. Another justification for this choice are the random-coding exponents of JSCC with SI [11], which also cannot be improved by binning. Notice that in [6] binning is used for the Stein case, thus we can conclude that in this special case the performance remains the same. Nonetheless, observing that a special codeword is also conceptually an extreme kind of binning - distant source sequences group together to the same channel codeword - one may wonder why they are still used in our scheme. The reason is, that

special codewords are different than zero-rate codebooks since two codewords can be chosen judiciously, and not just drawn uniformly over a type class. Then, the regular messages can be drawn in a conditional fashion, “far” from the two special ones. This is, indeed, the idea behind the UEP scheme in [7]. The scheme we propose is defined by the following:

- **Parameters:** Choose a quantization alphabet \mathcal{S} , where, w.l.o.g., $|\mathcal{S}| \geq |\mathcal{X}|$. Choose two disjoint subsets $\mathcal{M}^0, \mathcal{M}^1 \in \mathcal{P}(\mathcal{X})$ and denote $\bar{\mathcal{M}} = \mathcal{P}(\mathcal{X}) \setminus (\mathcal{M}^0 \cup \mathcal{M}^1)$. Choose a joint type for the special messages $\phi_{M^0 M^1} \in \mathcal{P}(\mathcal{W} \times \mathcal{W})$. For each $Q_X \notin \{\mathcal{M}^0 \cup \mathcal{M}^1\}$ choose a quantization test channel $P_{S|X}(Q_X) \in \mathcal{P}_n(\mathcal{S}, Q_X)$, a conditional channel input type $\phi_{W|M^0, M^1}(Q_X)$. For each Q_Y choose special-codewords biases $t^0(Q_Y)$ and $t^1(Q_Y)$. Choose detection regions $\mathcal{A}^0 \subseteq \mathcal{P}(\mathcal{S} \times \mathcal{Y}) \times \mathcal{P}(\mathcal{V} \times \mathcal{W})$ and $\mathcal{A}^1 \triangleq \mathcal{P}(\mathcal{S} \times \mathcal{Y}) \times \mathcal{P}(\mathcal{V} \times \mathcal{W}) \setminus \mathcal{A}^0$.

- **Codebooks construction:** For each $Q_X \in \mathcal{P}_n(\mathcal{X})$, draw a codebook $\mathcal{C}_n(Q_X)$ uniformly at random over the set of codebooks of rate $I(Q_X P_{S|X}(Q_X)) + \delta$ that have fixed type $[Q_X P_{S|X}(Q_X)]_{\mathcal{S}}$ and where for each $x^n \in \mathcal{T}_n(Q_X)$ there exists $s^n \in \mathcal{C}_n(Q_X)$ such that $(x^n, s^n) \in \mathcal{T}_n(Q_X P_{S|X}(Q_X))$. Let $h_{\text{quan}} : \mathcal{T}_n(Q_X) \rightarrow \mathcal{C}_n(Q_X)$ denote the quantization operation mapping x^n to a codeword s^n of joint type $Q_X P_{S|X}(Q_X)$.

Choose two special codewords $(m^{n,0}, m^{n,1}) \in \mathcal{Q}_{M^0 M^1}$. Then, for any $Q_X \in \mathcal{P}_n(\mathcal{X}) \setminus (\mathcal{M}^0 \cup \mathcal{M}^1)$ draw a channel codebook $\tilde{\mathcal{C}}_n(Q_X)$ of same size as $\mathcal{C}_n(Q_X)$, and where each codeword is picked uniformly at random over the set of vectors that have conditional type $\phi_{W|M^0, M^1}(Q_X)$ given $(m^{n,0}, m^{n,1})$. We denote the mapping from quantization to channel codewords by $h_{\text{map}} : \mathcal{C}_n(Q_X) \rightarrow \tilde{\mathcal{C}}_n(Q_X)$.

Assuming that the source has emitted $(x^n, y^n) \in \mathcal{T}_n(Q_X, Y)$. The scheme operates as follows:

- **Encoding:** If the source type is in $Q_X \in \mathcal{M}^i$, the encoder sends the corresponding $m^{n,i}$, $i \in [0, 1]$. Otherwise, it sends $w^n = h_{\text{map}}(h_{\text{quan}}(x^n))$.

- **Detection:** The detector follows the stages:

1) For each $\bar{s}^n \in \cup_{Q_X} \mathcal{C}_n(Q_X)$ computes the metric

$$-H(\hat{Q}_{\bar{s}^n | y^n} | \hat{Q}_{y^n}) + I(\hat{Q}_{\bar{w}^n, v^n}) \quad (19)$$

where $\bar{w}^n = h_{\text{map}}(s^n)$, and for $m^{n,i}$ it computes the metrics

$$-t^i(\hat{Q}_{y^n}) + I(\hat{Q}_{m^{n,i}, v^n}). \quad (20)$$

It then chooses the pair (\bar{s}^n, \bar{w}^n) or $m^{n,i}$ according to the maximal metric.

2) If the chosen codeword in the first step is $m^{n,i}$ then it outputs $\hat{\mathcal{H}} = i$. Otherwise, if $(\hat{Q}_{\bar{s}^n, y^n}, \hat{\Phi}_{\bar{w}^n, v^n}) \in \mathcal{A}^0$ it outputs $\hat{\mathcal{H}} = 0$, and otherwise it outputs $\hat{\mathcal{H}} = 1$.

The performance of the scheme is given as follows. For any given choice of parameters, let θ^i be given by the minimum over the following six exponents:

$$\begin{aligned}
\theta^{\text{std},i} &\triangleq \min D(Q_{S,X,Y} \| P_{S,X,Y}^i) \\
&\quad + D(\Phi_{V|W} \| \Gamma_{V|W} | \phi_W(Q_X)), \\
\theta^{\text{reg},i} &\triangleq \min D(Q_{S,X,Y} \| P_{S,X,Y}^i) \\
&\quad + D(\Phi_{M^i, W, V} \| \phi_{M^i, W}(Q_X) \Gamma_{V|W}), \\
\theta^{\text{reg} \rightarrow \text{opp},i} &\triangleq \min D(Q_{S,X,Y} \| P_{S,X,Y}^i) \\
&\quad + D(\Phi_{V|W} \| \Gamma_{V|W} | \phi_W(Q_X)) \\
&\quad + [I(Q'_{S,Y}) + I(\Phi'_{W,V}) - I(Q'_{S,X})]_+, \\
\theta^{\text{sp} \rightarrow \text{reg},i} &\triangleq \min D(Q_{S,X,Y} \| P_{S,X,Y}^i) \\
&\quad + D(\phi_{M^i} \Phi_{V|M^i} \| \phi_{M^i} \Gamma_{V|M^i}) \\
&\quad + [I(Q'_{S,Y}) + I(\Phi'_{W,V|M^i} | \phi_{M^i}) - I(Q'_{S,X})]_+, \\
\theta^{\text{sp} \rightarrow \text{opp},i} &\triangleq \min D(Q_{X,Y} \| P_{X,Y}^i) \\
&\quad + D(\phi_{M^i, M^i} \Phi_{V|M^i, M^i} \| \phi_{M^i, M^i} \Gamma_{V|M^i}), \\
\theta^{\text{opp},i} &\triangleq \min D(Q_X \| P_X^i), \tag{21}
\end{aligned}$$

where the minimizations are over $(Q_{S,X,Y}, \Phi_{M^i, M^i, W, V}, (Q'_{S,X,Y}, \Phi'_{M^i, M^i, W, V})$ which are consistent with $Q_X P_{S|X}(Q_X)$ and $\phi_{M^0, M^1} \phi_{W|M^0, M^1}(Q_X)$, and where each of the six minimizations is restricted to one of the following sets, respectively:

$$\begin{aligned}
&\{Q_X \in \bar{\mathcal{M}}, (Q_{S,Y}, \phi_W(Q_X) \Phi_{W|V}) \in \mathcal{A}^{\bar{i}}\}, \\
&\{Q_X \in \bar{\mathcal{M}}, \\
&\quad - H(Q_{S|Y}|Q_Y) + I(\Phi_{W,V}) \leq -t^{\bar{i}}(Q_Y) + I(\Phi_{M^i, V})\}, \\
&\{Q_X \in \bar{\mathcal{M}}, Q'_X \in \bar{\mathcal{M}}, Q'_{V,Y} = Q_{V,Y}, (Q'_{S,Y}, \Phi'_{W,V}) \in \mathcal{A}^{\bar{i}}, \\
&\quad - H(Q'_{S|Y}|Q'_Y) + I(\Phi'_{W,V}) \geq \\
&\quad - H(Q_{S|Y}|Q_Y) + I(\Phi_{W,V})\}, \\
&\{Q_X \in \mathcal{M}^i, Q'_X \in \bar{\mathcal{M}}, Q'_{V,Y} = Q_{V,Y}, (Q'_{S,Y}, \Phi'_{W,V}) \in \mathcal{A}^{\bar{i}} \\
&\quad - H(Q'_{S|Y}|Q'_Y) + I(\Phi'_{W,V}) \geq -t^i(Q_Y) + I(\Phi_{M^i, V})\}, \\
&\{Q_X \in \mathcal{M}^i, \\
&\quad - t^{\bar{i}}(Q_Y) + I(\Phi_{M^i, V}) \geq -t^i(Q_Y) + I(\Phi_{M^i, V})\}, \\
&\{Q_X \in \mathcal{M}^{\bar{i}}\}. \tag{22}
\end{aligned}$$

Theorem 2. For any given choice of parameters, the exponent pair (θ^0, θ^1) is achievable, where θ^i is given by (21)-(22).

Proof outline: The existence of quantization codebooks follows as in the clean-channel case. Assume that $\mathcal{H} = i$, and that the “true” vectors have type pair $(Q_{S,X,Y}, \Phi_{M^i, M^i, W, V})$. It can be verified that any error event belongs to one of the following cases:

1) A regular message was sent and decoded correctly, but an erroneous decision was made at stage 2.

2) A regular message was sent, but the “opposite” special message \bar{i} was decoded.

3) A regular message was sent, but another regular message was decoded, which has a type pair $(Q'_{S,X,Y}, \Phi'_{W,V, M^i, M^i})$ that leads to an erroneous decision at stage 2.

4) The “correct” special message i was sent, but a regular message was decoded, which has a type pair that leads to an erroneous decision at stage 2.

5) The “correct” special message was sent, but the “opposite” special message was decoded.

6) The “opposite” special message was sent.

The six exponents in the theorem correspond to these six events, respectively. The exponent of each event is calculated along similar lines to the clean-channel case, using the (conditionally) independent codebook construction, the encoder and decoder structure. ■

As in the clean-channel case, we have left the choice of decision regions as a design parameter. We note that also here, one natural choice is the LRT (which is based upon the source types only, as the channel likelihood is the same under both hypotheses). For the Stein case, a source- and channel-typicality region is preferable. Using a single special message, as one would in this case, we can recover the exponent of [6]. Of course, one may interpolate between the two in a manner similar to (18). We have also left the biases $t^0(Q_Y)$ and $t^1(Q_Y)$ unspecified. As these replace the source score applied to regular messages, a good choice seems to be

$$t^i(Q_Y) = \min D(Q_{X|Y} \| P^i(X|Y)|Q_Y) + t, \tag{23}$$

where the minimization is over all $Q_{X|Y}$ such that $Q_X \in \mathcal{M}_i$ and t is a constant. Detailed performance evaluation of the proposed scheme for specific instances of the noisy DHT problem, and a comparison between different choices of the decision regions and the biases is left for future work.

REFERENCES

- [1] H. Shimokawa, T. S. Han, and S. Amari. Error bound of hypothesis testing with data compression. In *IEEE International Symposium on Information Theory (ISIT)*, 1994.
- [2] M. S. Rahman and A. B. Wagner. On the optimality of binning for distributed hypothesis testing. *IEEE Trans. Info. Theory*, pages 6282 – 6303, 2012.
- [3] R. Ahlswede and I. Csiszár. Hypothesis testing with communication constraints. *IEEE Trans. Info. Theory*, 32(4):533–542, Jul 1986.
- [4] N. Weinberger and Y. Kochman. On the reliability function of distributed hypothesis testing under optimal detection. In *IEEE International Symposium on Information Theory (ISIT)*, pages 1066–1070, June 2018. Full version available on: <http://arxiv.org/pdf/1801.03687.pdf>.
- [5] S. Sreekumar and D. Gündüz. Distributed hypothesis testing over noisy channels. In *IEEE International Symposium on Information Theory*, pages 983–987, June 2017.
- [6] S. Salehkalaibar and M. Wigger. Distributed hypothesis testing based on unequal-error protection codes. *IEEE Trans. Info. Theory*, Submitted, 2018.
- [7] S. Borade, B. Nakiboglu, and L. Zheng. Unequal error protection: an information-theoretic perspective. *IEEE Trans. Info. Theory*, 55:5511–5539, 2009.
- [8] E. Tuncel. Slepian-Wolf coding over broadcast channels. *IEEE Trans. Info. Theory*, 52(4):1469–1482, 2006.
- [9] I. Csiszár. On the error exponent of source-channel transmission with a distortion threshold. *IEEE Trans. Info. Theory*, IT-28:823–838, Nov. 1982.
- [10] I. Csiszár and J. Körner. *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Cambridge University Press, 2011.
- [11] N. Merhav. Universal decoding for source-channel coding with side information. *Communications in Information and Systems*, 16:17–58, 2016.