

# Estimation with a helper who knows the interference

Yeow-Khiang Chia\*, Rajiv Soundararajan<sup>†</sup> and Tsachy Weissman<sup>‡</sup>

## Abstract

We consider the problem of estimating a signal corrupted by independent interference with the assistance of a cost-constrained helper who knows the interference causally or noncausally. When the interference is known causally, we characterize the minimum distortion incurred in estimating the desired signal. In the noncausal case, we present a general achievable scheme for discrete memoryless systems and novel lower bounds on the distortion for the binary and Gaussian settings. Our Gaussian setting coincides with that of assisted interference suppression introduced by Grover and Sahai. Our lower bound for this setting is based on the relation recently established by Verdú between divergence and minimum mean squared error. We illustrate with a few examples that this lower bound can improve on those previously developed. Our bounds also allow us to characterize the optimal distortion in several interesting regimes. Moreover, we show that causal and noncausal estimation are not equivalent for this problem. Finally, we consider the case where the desired signal is also available at the helper. We develop new lower bounds for this setting that improve on those previously developed, and characterize the optimal distortion up to a constant multiplicative factor for some regimes of interest.

## I. INTRODUCTION

Consider a joint source channel coding problem as depicted in Figure 1. We have two memoryless sources  $S_1$  (the desired signal) and  $S_2$  (the interfering signal). The decoder's aim is to estimate the source sequence  $S_1^n$  from  $Y^n$ , with the goal of minimizing the average per symbol distortion  $E(\sum_{i=1}^n d(\hat{S}_{1i}(Y^n), S_{1i}))/n$ . The encoder (helper), who knows the interfering signal  $S_2$ , aids the decoder in reconstructing the signal  $S_1$  through his choice of  $X$ , subject to a cost constraint  $\rho(X)$ .

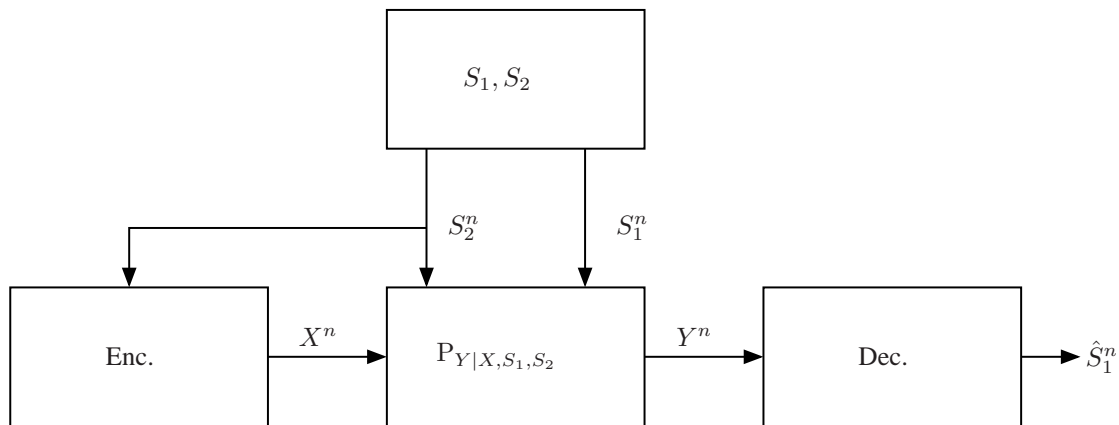


Fig. 1: Estimation with a helper who knows the interference. The interfering signal is  $S_2^n$  while the desired signal is  $S_1^n$ . The encoder (helper) tries to help the decoder in estimating  $S_1^n$  by reducing the interference due to  $S_2$ , subject to a per symbol cost constraint on its transmission  $X^n$ .

This work was partially supported by NSF under grant CCF-0916713, AFOSR under grant FA95500910063, and the Center for Science of Information (CSoI), an NSF Science and Technology Center.

\* Yeow-Khiang Chia was with Stanford University when most of this work was done. He is now with Institute for Infocomm Research, Singapore. Email: yeowkhiang@gmail.com

<sup>†</sup> Rajiv Soundararajan is with The University of Texas at Austin. Email: rajivs@utexas.edu

<sup>‡</sup> Tsachy Weissman is with Stanford University. Email: tsachy@stanford.edu

Applications may arise in sensor networks or cognitive radio systems. As a motivating example, suppose Alice is talking to Bob in his office. As a result of ongoing construction work near Bob's office, there is high interference which makes it hard for Bob to listen to Alice. Fortunately, Bob recently purchased a noise cancellation device which has a microphone placed near the construction site. The microphone measures the interfering signal from the construction site and transmits it to a noise cancellation speaker situated in Bob's office. Since electromagnetic waves travel faster than sound, the noise cancellation speaker knows the interfering signal noncausally. Due to a power constraint on the speaker, it cannot cancel the interference fully. What then, is the minimum distortion that can be achieved by Bob in trying to reconstruct Alice's speech?

Our setup is closely related to several strands of work involving communication over channels with states. In [1], the authors considered the problem of State Amplification, where a message is to be sent to the decoder and the decoder also forms a list of possible  $S_2^n$  sequences. The goal is to maximize the message transmission rate and reduce the uncertainty the decoder has regarding  $S_2$ ; i.e. reduce the list size of possible  $S_2^n$  sequences. Recently, the problem of state amplification with a distortion constraint was considered in [2], with an additional condition that the encoder only knows the state  $S_2$  causally. This setting is similar to our setting, with the main difference being that the decoder wishes to reconstruct  $S_2$  rather than  $S_1$ . When our setting is specialized to the Gaussian case with the mean squared error distortion between the reconstruction and the signal, our setting becomes equivalent to the problem of Assisted Interference Suppression considered in [3]. As detailed in [3], this problem is closely related to Witsenhausen's counterexample in Optimum Control Theory [4].

In this paper, we consider both the case when  $S_2$  is available causally at the encoder, and the case when  $S_2$  is available non causally at the encoder. Our main contributions are as follows:

- 1) When  $S_2$  is available causally at the encoder, we characterize the minimum achievable distortion in  $S_1$ . We borrow certain ideas used in the characterization of the distortion cost region for the causal state amplification problem in [2] to establish our result.
- 2) For the noncausal setting, we first give an achievable scheme for the general discrete memoryless system and then focus our attention on the case where  $S_1$  and  $S_2$  are independent Bernoulli random variables and the distortion measure is Hamming. We give two lower bounds on the achievable distortion for this binary setting. The first lower bound is based on ideas from the Assisted Interference Suppression problem [3], while the second lower bound is based on ideas from the problem of Compression with Actions [5]. Neither bound contains the other and one bound can be better than the other, depending on the regime of interest. Using our lower and upper bounds, we characterize the minimum achievable distortion in several regimes. In particular, we provide an example to show that causal and noncausal estimation of  $S_1$  are not equivalent and causal knowledge of  $S_2$  could incur a higher distortion than noncausal knowledge of  $S_2$  at the encoder. A complete characterization of the minimum achievable distortion in the noncausal case remains open.
- 3) In the Gaussian case, where  $S_1$  and  $S_2$  are independent Gaussian random variables with finite variance, the distortion measure is the mean square error and  $Y = X + S_1 + S_2$ , we note that our setting coincides with that of Assisted Interference Suppression [3]. For this setting, we give a lower bound on the minimum achievable distortion which in some places improves on that given in [3], and also its improved version given in [6]. The proof of our lower bound relies on an application of Verdu's relation between relative entropy and mismatched estimation in Gaussian noise [7]. In recent years, since the seminal paper [8] established the relationship between minimum mean square error estimation (MMSE) in Gaussian noise and the Mutual Information between the signal and the output, there has been interest in applying these information-estimation relations to problems in Information Theory (see e.g. [9] and [10]). Our lower bound, which seems difficult to obtain by traditional techniques such as the Entropy Power Inequality [11, Chapter 2], provides another application of these information-estimation relations.
- 4) In the Gaussian case, we also consider the setting when the encoder has access to  $S_1$  noncausally, in addition to  $S_2$ . This setting is a special case of a problem considered in [12]. We give a lower bound for this setting that contains the previous bounds in [12] and can be strictly better in some cases. Furthermore, we establish constant gap results between the achievable distortion and our lower bound.

We first provide the formal definitions in the next section. In Section III, we consider the causal case. In Section IV, we consider the noncausal case, present an achievable scheme for general discrete memoryless systems and analyze the binary setting in detail. Section V deals with the Gaussian version of this problem, while we

consider the Gaussian setting when  $S_1$  is also available noncausally at the encoder in Section VI. We conclude in Section VII with a summary of our findings and directions for future work.

## II. DEFINITIONS

In this section, we give formal definitions for our problem settings. We will follow the notation of [11], and assume throughout this paper that the channel in consideration is memoryless. That is,  $p(y^n|x^n, s_1^n, s_2^n) = \prod_{i=1}^n p(y_i|x_i, s_{1i}, s_{2i})$ . We also assume that  $S_1^n$  and  $S_2^n$  are independent i.i.d. sequences.

### A. Estimation with interference known at the helper

A  $(n, C)$  code for the setting shown in Figure 1 when the interference is known *noncausally* consists of

- An encoder that maps the interference  $S_2^n$  to  $X^n$ ,  $f : \mathcal{S}_2^n \rightarrow \mathcal{X}^n$ ;
- A decoder that maps the output  $Y^n$  to the reconstruction sequence  $\hat{S}_1^n$ ,  $g : \mathcal{Y}^n \rightarrow \hat{\mathcal{S}}_1^n$ ;

such that  $\mathbb{E} \sum_{i=1}^n \rho(X_i)/n \leq C$ . The expected per symbol *distortion*,  $D$ , is given by  $D = \mathbb{E} d(S_1^n, \hat{S}_1^n) = \mathbb{E} \sum_{i=1}^n d(S_{1i}, \hat{S}_{1i})/n$ .

A distortion  $D$  is said to be achievable under the cost constraint  $C$  if there exists a sequence of  $(n, C + \epsilon_n)$  codes, where  $\epsilon_n \rightarrow 0$  as  $n \rightarrow \infty$ , and

$$\limsup_{n \rightarrow \infty} \mathbb{E} d(S_1^n, \hat{S}_1^n) \leq D.$$

The *minimum achievable distortion*,  $D(C)_{\min}$ , is then defined as the infimum of the set of achievable distortions under the cost constraint  $C$ .

When the interference is only known *causally*, the definitions are mostly the same, with the difference being that the encoder is restricted to causal mapping:

$$f_i : \mathcal{S}_2^i \rightarrow \mathcal{X} \text{ for } i \in [1 : n].$$

### B. Estimation with source and interference known at the helper

This setting is shown in Figure 2. For this setting, we restrict attention to the case where  $S_1$  and  $S_2$  are independent Gaussian random variables,  $S_1 \sim \mathcal{N}(0, P_1)$  and  $S_2 \sim \mathcal{N}(0, P_2)$ . Furthermore, we assume that both  $S_1$  and  $S_2$  are known noncausally at the encoder, and the distortion measure is the mean square error between  $S_1$  and its reconstruction. That is,  $d(s_1, \hat{s}_1) = (s_1 - \hat{s}_1)^2$ . The channel is specified by  $Y = X + S_1 + S_2 + Z$ , where  $Z \sim \mathcal{N}(0, N)$  is independent of  $S_1$  and  $S_2$ . The cost constraint is the expected power constraint:  $\mathbb{E}(\sum_{i=1}^n X_i^2/n)$ . As the definitions are similar to the previous setting, we only mention the difference. That is, the encoder now maps both  $S_1^n$  and  $S_2^n$  to  $X^n$ :

$$f : \mathcal{S}_1^n \times \mathcal{S}_2^n \rightarrow \mathcal{X}^n.$$

## III. CAUSAL ESTIMATION WITH A HELPER

In this section, we give the distortion-cost tradeoff region for the setting given in II-A under the condition that the interfering signal,  $S_2$ , is causally known at the encoder. We will discuss some connections between our setting and that of the problem of Causal State Amplification discussed in [2].

**Theorem 1.** *The distortion-cost region for the problem of estimation with a helper when the interfering signal is causally known at the encoder is given by*

$$D(C)_{\min} = \min_{U, V, X, \hat{S}_1} \mathbb{E} d(S_1, \hat{S}_1(U, V, Y))$$

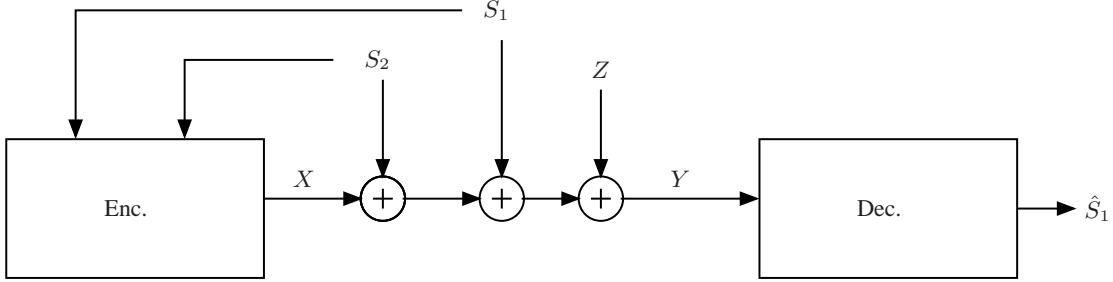


Fig. 2: Gaussian estimation with a helper that knows both the interference and the source. The random variables  $S_1$ ,  $S_2$  and  $Z$  are independent zero mean Gaussian random variables. The encoder has knowledge of  $S_1^n$  and  $S_2^n$  noncausally and the decoder tries to perform lossy reconstruction of  $S_1^n$ . The distortion criterion is the mean square error criterion and the cost constraint is the expected power constraint on the encoder output,  $X$ .

for some  $p(u)p(v|u, s_2)p(s_1)p(s_2)$  and functions  $x(u, s_2)$  and  $\hat{s}_1(u, v, y)$  such that

$$\begin{aligned} I(U; Y) &\geq I(V; S_2|U, Y), \\ E \rho(X) &\leq C. \end{aligned}$$

The cardinalities of the auxiliary random variables may be upper bounded by  $|\mathcal{U}| \leq |\mathcal{S}_2|(|\mathcal{X}| - 1) + 2$  and  $|\mathcal{V}| \leq |\mathcal{U}|(|\mathcal{S}_2| + 1)$ .

The achievability scheme in this Theorem is actually the same as that used in the problem of Causal State Amplification considered in [2], where the focus was on reconstructing  $S_2$  instead of  $S_1$ . The expressions for the distortion-cost tradeoff are also similar, with the difference being that in the Causal State Amplification setting, one is interested in minimizing the distortion between  $S_2$  and its reconstruction, rather than between  $S_1$  and its reconstruction. Of course, the optimizing choice of auxiliary random variables in the two problems are different, since in our setting, we try to minimize the interference ( $S_2$ ) as much as possible subjected to a cost constraint, whereas in the setting of Causal State Amplification, one tries to *amplify* the interfering signal. As a (trivial) example, consider the case when  $S_1, S_2, X \in \{0, 1\}$  and  $Y = X \oplus S_1 \oplus S_2$  and no cost constraint. Then, clearly, in our problem of causal estimation with a helper, we set  $X = S_2$  to cancel out the interference completely, thereby recovering  $S_1$  losslessly. In contrast, for the problem of Causal State Amplification, we will not cancel out  $S_2$ , since that is the signal we are trying to recover.

Theorem 1 gives the optimal cost-distortion tradeoff for the estimation problem when the encoder knows the interfering signal causally. A natural question to ask is whether there is any penalty incurred in this restriction? In the next section, we will give an example of a binary estimation with a helper problem under Hamming loss and show that there is indeed a penalty incurred in only knowing the interfering signal causally.

*Proof of Theorem 1:*

*Sketch of Achievability:* As the achievability scheme is similar to that in [2], we give only a sketch in Appendix I for completeness.

*Converse:* Given a  $(n, C)$  code that achieves distortion  $D$ , we have

$$\begin{aligned} 0 &\leq \frac{1}{n} \sum_{i=1}^n I(Y_{i+1}^n; Y_i) \\ &= \frac{1}{n} \sum_{i=1}^n (I(Y_{i+1}^n, S_2^{i-1}; Y_i) - I(S_2^{i-1}; Y_i | Y_{i+1}^n)) \\ &\stackrel{(a)}{=} \frac{1}{n} \sum_{i=1}^n (I(Y_{i+1}^n, S_2^{i-1}; Y_i) - I(Y_{i+1}^n; S_{2i} | S_2^{i-1})) \end{aligned}$$

$$\begin{aligned}
&\stackrel{(b)}{=} \frac{1}{n} \sum_{i=1}^n (I(Y_{i+1}^n, S_2^{i-1}; Y_i) - I(Y_{i+1}^n, S_2^{i-1}; S_{2i})) \\
&\stackrel{(c)}{=} I(Y_{Q+1}^n, S_2^{Q-1}; Y_Q | Q) - I(Y_{Q+1}^n, S_2^{Q-1}; S_{2Q} | Q) \\
&\leq I(Y_{Q+1}^n, Q, S_2^{Q-1}; Y_Q) - I(Y_{Q+1}^n, S_2^{Q-1}, Q; S_{2Q}) \\
&= I(U, V; Y) - I(U, V; S_2),
\end{aligned}$$

where in (a), we used the Csiszár sum lemma [13]; in (b), we used the fact that  $S_2$  is a memoryless source; in (c), we defined  $Q$  in the standard manner to be uniformly distributed over  $[1 : n]$  and independent of every other random variable; and in the last step, we define  $U = (U_Q, Q) = (S_2^{Q-1}, Q)$  and  $V = V_Q = Y_{Q+1}^n$ . With these definitions of auxiliary random variables, it is clear that  $U$  is independent of  $S_2$  and also, the encoder output  $X$ , is a function of both  $U$  and  $S_2$ . Further, using the relationship that  $U$  is independent of  $S_2$  and  $V = (U, S_2) - Y$ , the condition that  $I(U, V; Y) - I(U, V; S_2) \geq 0$  reduces to

$$I(U; Y) \geq I(V; S_2 | U, Y).$$

It now remains to show that the achievable distortion can be lower bounded by this choice of auxiliary random variables. To this end, we will use a technique for lower bounding distortion found in [14]. We have

$$\begin{aligned}
D + \epsilon_n &\geq \frac{1}{n} \sum_{i=1}^n \mathbb{E} d(S_{1i}, \hat{S}_{1i}(Y^n)) \\
&= \frac{1}{n} \sum_{i=1}^n \mathbb{E} d(S_{1i}, \hat{S}_{1i}(Y_i, Y_{i+1}^n, Y^{i-1})) \\
&= \frac{1}{n} \sum_{i=1}^n \mathbb{E} d(S_{1i}, \hat{S}_{1i}(Y_i, V_i, Y^{i-1})) \\
&\geq \frac{1}{n} \sum_{i=1}^n \mathbb{E} d(S_{1i}, \hat{S}'_{1i}(Y_i, V_i, Y^{i-1}, S_2^{i-1})), \tag{1}
\end{aligned}$$

where the last step follows from the observation that we can recover  $\hat{S}_{1i}$  from  $\hat{S}'_{1i}$  by simply ignoring  $S_2^{i-1}$ . Next, consider the term  $\mathbb{E} d(S_{1i}, \hat{S}'_{1i}(Y_i, V_i, Y^{i-1}, S_2^{i-1}))$ .

$$\begin{aligned}
&\mathbb{E} d(S_{1i}, \hat{S}'_{1i}(Y_i, V_i, Y^{i-1}, S_2^{i-1})) \\
&= \mathbb{E} d(S_{1i}, \hat{S}'_{1i}(Y_i, V_i, Y^{i-1}, U_i)) \\
&= \sum p(s_{1i}, u_i, v_i, y_i, y^{i-1}) d(s_{1i}, \hat{S}'_{1i}(y_i, v_i, y^{i-1}, u_i)) \\
&= \sum p(u_i, y_i, v_i) \sum p(y^{i-1}, s_{1i} | u_i, y_i, v_i) d(s_{1i}, \hat{S}'_{1i}(y_i, v_i, y^{i-1}, u_i)) \\
&\stackrel{(a)}{=} \sum p(w_i) \sum p(y^{i-1} | w_i) p(s_{1i} | w_i) d(s_{1i}, \hat{S}'_{1i}(w_i, y^{i-1})) \\
&= \sum p(w_i) \sum_{y^{i-1}} p(y^{i-1} | w_i) \sum_{s_{1i}} p(s_{1i} | w_i) d(s_{1i}, \hat{S}'_{1i}(w_i, y^{i-1})) \\
&\stackrel{(b)}{\geq} \sum p(w_i) \sum_{y^{i-1}} p(y^{i-1} | w_i) \sum_{s_{1i}} p(s_{1i} | w_i) d(s_{1i}, \hat{S}_{1i}^*(w_i)) \\
&= \sum p(w_i, s_{1i}) d(s_{1i}, \hat{S}_{1i}^*(w_i)) \\
&= \mathbb{E} d(S_{1i}, \hat{S}_{1i}^*(W_i)), \tag{2}
\end{aligned}$$

where in (a), we define  $w_i = (u_i, y_i, v_i)$  for notational convenience and the fact that  $p(y^{i-1}, s_{1i} | w_i) = p(y^{i-1} | w_i) p(s_{1i} | w_i)$  follows from the Markov Chain  $Y^{i-1} - W_i - S_{1i}$ , which in turn, follows from the fact that  $S_2$  is only causally known at the encoder. Hence, given  $S_2^{i-1}$  and also  $X^{i-1}$  since it is a function of  $S_2^{i-1}$ ,  $Y^{i-1}$  is independent of  $S_{1i}$ . (b) follows from defining  $y^{i-1*} = \arg \min_{y^{i-1}} \sum_{s_{1i}} p(s_{1i} | w_i) d(s_{1i}, \hat{S}'_{1i}(w_i, y^{i-1}))$  and  $\hat{S}_{1i}^*(w_i) = \hat{S}'_{1i}(w_i, y^{i-1*})$ .

Combining inequality (2) into inequality (1) then gives us

$$\begin{aligned} D + \epsilon_n &\geq \frac{1}{n} \sum_{i=1}^n \mathbb{E} d(S_{1i}, \hat{S}_{1i}^*(Y_i, V_i, U_i)) \\ &= \mathbb{E}_Q(\mathbb{E}(d(S_{1Q}, \hat{S}_{1Q}^*(Y_Q, V_Q, U_Q))|Q)) \\ &\geq \mathbb{E}(d(S_1, \hat{S}_1(Y, V, U))). \end{aligned}$$

The bounds on cardinality of the auxiliary random variables follow from standard arguments (see for e.g. [11, Appendix C]). This completes the proof of converse.  $\blacksquare$

#### IV. NONCAUSAL ESTIMATION WITH A HELPER

Having established the distortion-cost region for the discrete memoryless estimation with a helper problem when the interfering signal is causally known, we now turn to the noncausal setting, that is, when  $S_2$  is noncausally known at the encoder. This setting is more complicated and the distortion-cost region is still unknown. In this section, we first give an achievability scheme based on the recently proposed technique of hybrid coding [15]. We then specialize our setting to the case of binary estimation with a helper.

The problem of binary estimation with a helper is one where  $S_1 \sim \text{Bern}(p_1)$ ,  $S_2 \sim \text{Bern}(p_2)$ ,  $0 \leq p_1, p_2 \leq 1/2$ ,  $X \in \{0, 1\}$ ,  $Y = X \oplus S_1 \oplus S_2$  and  $d(S_1, \hat{S}_1) = S_1 \oplus \hat{S}_1$ , i.e., Hamming distortion. The cost is given by  $\rho(X) = 1$  if  $X = 1$  and 0 otherwise. The objective of the problem is to design a coding strategy that minimizes the Hamming distortion in  $S_1$ .

Specializing to the case of binary estimation with a helper allows us to derive a number of additional results of interest. In subsection IV-A, we give a (non-trivial) condition on the cost constraint that allows us to achieve zero expected distortion. We then show that in the binary case, there is a penalty involved if  $S_2$  is known only causally instead of noncausally. As a result, the distortion incurred in  $S_1$  is higher if  $S_2$  is only known causally as opposed to it being known noncausally. In subsection IV-B, we describe the two lower bounds for the problem of binary estimation with a helper and then compare them. In subsection IV-C, we briefly mention a non-binary setting for which we can characterize the distortion-cost tradeoff, and show that symbol by symbol encoding is optimal in that setting.

##### A. Achievable scheme

We first give an achievable scheme for the general discrete memoryless estimation with a helper problem based on hybrid coding [15]. We will extend this scheme to the Gaussian case in the next section.

**Theorem 2.** *An achievable distortion for the problem of estimation with a helper is given by*

$$D(C) \leq \inf \mathbb{E} d(S_1, \hat{S}_1(U, Y)),$$

where the minimization is over distribution  $p(u|s_2)$  and functions  $x = f(s_2, u)$  and  $\hat{s}_1(u, y)$  such that

$$\begin{aligned} I(U; Y) &> I(U; S_2), \\ \mathbb{E} \rho(X) &\leq C. \end{aligned}$$

*Sketch of Achievability:* The achievability scheme follows that of the hybrid coding scheme given in [15]. We give only a sketch here. The codebook generation consists of generating  $2^{n(I(U; S_2) + \epsilon)}$  sequences according to  $\prod_{i=1}^n p(u_i)$ . For encoding, given an  $s_2^n$  sequence, the encoder looks for a  $u^n$  sequence such that  $(u^n, s_2^n) \in \mathcal{T}_\epsilon^{(n)}$ . If there is more than one, it selects one sequence uniformly at random from the set of jointly typical sequences. It then outputs  $x^n$  according to  $f(u_i, s_{2i})$  for  $i \in [1 : n]$ . The decoder looks for the unique  $\hat{u}^n$  sequence such that  $(\hat{u}^n, y^n) \in \mathcal{T}_\epsilon^{(n)}$ . It can be shown as in [15] that the probability of decoding error goes to zero as  $n \rightarrow \infty$  if

$$I(U; Y) > I(U; S_2) + 2\epsilon.$$

The decoder then reconstructs  $S_1^n$  according to  $\hat{s}_1(\hat{u}_i, y_i)$  for  $i \in [1 : n]$ .

We now specialize the achievable distortion-cost region in Theorem 2 to the case of binary estimation with a helper. The next result shows that, in the binary case, zero expected distortion is achievable under a condition on the cost constraint.

**Proposition 1.** *For the problem of binary estimation with a helper,*

$$D(C)_{\min} = 0$$

*if  $H_2(C) > H(X \oplus S_2|Y)$ , where  $H_2(\cdot)$  is the binary entropy function,  $X \sim \text{Bern}(C)$  independent of  $S_2$  and  $Y = X \oplus S_1 \oplus S_2$ .*

*Proof:* The sufficient condition on the cost constraint follows from a particular choice of auxiliary random variable  $U$  in Theorem 2. We let  $X \sim \text{Bern}(C)$  independent of  $S_2$  and let  $U = X \oplus S_2$ . The decoder reconstructs  $S_1$  as  $\hat{S}_1 = Y \oplus U = S_1$ , incurring zero expected distortion. We now note that the cost constraint is satisfied since  $X \sim \text{Bern}(C)$ . To satisfy the mutual information condition on the choice of joint distribution, we require

$$\begin{aligned} I(U; Y) &> I(U; S_2) \\ \Rightarrow H(U|S_2) &> H(U|Y) \\ \Rightarrow H(X|S_2) &> H(X \oplus S_2|Y) \\ \Rightarrow H_2(C) &> H(X \oplus S_2|Y). \end{aligned}$$

Weakening Proposition 1 leads to the following simple sufficient condition for zero distortion. ■

**Corollary 1.** *If  $C > p_1$ ,  $D(C)_{\min} = 0$ .*

Proof of Corollary 1 follows readily from Proposition 1. Since  $0 \leq C, p_1 \leq 1/2$ , if  $C > p_1$ , then

$$\begin{aligned} H_2(C) &> H_2(p_1) \\ &= H(S_1) \\ &\geq H(S_1|Y) \\ &= H(X \oplus S_2|Y). \end{aligned}$$

**Remark IV.1.** *A trivial condition for zero distortion is when  $C \geq p_2$  in which case, the encoder just performs symbol by symbol cancellation of  $S_2$  to allow the decoder to recover  $S_1$  losslessly. Corollary 1 shows that zero expected distortion can be achieved even if  $C < p_2$  as long as  $C > p_1$ .*

By choosing  $U = (U', V')$  in Theorem 2, where  $p(u|s_2) = p(u')p(v'|u', s_2)$ , we obtain the distortion-cost region when  $S_2$  is restricted to be causally known at the encoder<sup>1</sup>. A natural question to ask is whether the achievable distortion for the same cost constraint can be lowered if  $S_2$  is noncausally known at the encoder rather than only causally known. This is indeed the case for the problem of binary estimation with a helper.

**Proposition 2.** *For the problem of binary estimation with a helper, the achievable distortion when  $S_2$  is noncausally known at the encoder can be strictly smaller than the achievable distortion when  $S_2$  is only causally known at the encoder, with the same cost constraint.*

*Proof:* To prove Proposition 2, we exhibit an example where we can achieve zero expected distortion when  $S_2$  is noncausally known at the encoder, but for which the achievable distortion is strictly greater than zero when  $S_2$  is only causally known. To this end, we let  $p_1 = 0.1$ ,  $p_2 = 0.5$  and  $C = 0.11$ . Since  $C > p_1$ , from Corollary 1, an expected distortion of 0 can be achieved when  $S_2$  is noncausally known at the encoder. That is, we have  $D(0.11)_{\min\text{-noncausal}} = 0$ . Proof of this proposition is completed using the following claim, which states that the minimum expected distortion when  $S_2$  is only causally known at the encoder,  $D(0.11)_{\min\text{-causal}}$ , is *strictly* greater than zero.

**Claim 1.**  $D(0.11)_{\min\text{-causal}} > 0$  for any choice of  $U, V$  satisfying the constraints given in Theorem 1.

Claim 1 is proven in Appendix II. ■

<sup>1</sup>The boundary case of  $I(U; Y) = I(V; S_2|U, Y)$  is treated in a similar fashion as in the causal setting.

*B. Lower bounds for binary estimation with helper*

We now turn to lower bounds for the binary estimation with a helper problem. The first lower bound that we will present uses ideas from [3] adapted from the Gaussian to the binary setting.

**Theorem 3.** *A lower bound for the achievable distortion for the problem of binary estimation with a helper is given by*

$$D(C)_{\min} \geq \min H_2^{-1}(H(S_1) + H(S_2) - H(Y)) - \mathbb{E} X,$$

where we define  $H_2^{-1} = 0$  if the argument is negative or greater than 1, and the minimization is over joint distribution  $p(x|s_2)$  such that  $\mathbb{E} X \leq C$ .

*Proof:* We first start with a simple claim.

**Claim 2.** *Let  $\hat{s}_1^n(y^n)$  be an optimal reconstruction function (with respect to Hamming distortion) for  $s_1^n$  and  $\hat{x}^n(y^n)$  be an optimal reconstruction for  $s_2^n \oplus x^n$ . Then,  $d(s_1^n, \hat{s}_1(y^n)) = d(s_2^n \oplus x^n, \hat{x}^n(y^n))$ .*

To prove this claim, observe that  $d(s_1^n, \hat{s}_1(y^n)) = \sum_{i=1}^n s_{1i} \oplus \hat{s}_{1i}(y^n)$ . Consider now the function  $\hat{x}'_i(y^n) = \hat{s}_{1i}(y^n) \oplus y_i$ . Since  $\hat{x}_i(y^n)$  is optimal for  $d(s_2^n \oplus x^n, \hat{x}^n(y^n))$ , we have

$$\begin{aligned} d(s_2^n \oplus x^n, \hat{x}^n(y^n)) &\leq \sum_{i=1}^n s_{2i} \oplus x_i \oplus \hat{x}'_i(y^n) \\ &= \sum_{i=1}^n s_{2i} \oplus x_i \oplus \hat{s}_{1i}(y^n) \oplus y_i \\ &= \sum_{i=1}^n s_{1i} \oplus \hat{s}_{1i}(y^n) \\ &= d(s_1^n, \hat{s}_1(y^n)). \end{aligned}$$

Hence, we have  $d(s_2^n \oplus x^n, \hat{x}^n(y^n)) \leq d(s_1^n, \hat{s}_1(y^n))$ . For the other direction, consider the function  $\hat{s}'_{1i} = \hat{x}_i(y^n) \oplus y_i$ . Repeating the same arguments for  $d(s_1^n, \hat{s}_1(y^n))$  instead of  $d(s_2^n \oplus x^n, \hat{x}^n(y^n))$ , it is easy to show that  $d(s_1^n, \hat{s}_1(y^n)) \leq d(s_2^n \oplus x^n, \hat{x}^n(y^n))$ , which completes the proof of claim 2.

As an aside, the proof of claim 2 shows that the optimal reconstruction functions for the respective problems are related by  $\hat{x}^n(y^n) = \hat{s}_1^n(y^n) \oplus y^n$ .

We now continue with our lower bound for the binary case. Using claim 2, we have

$$\begin{aligned} d(S_1^n, \hat{s}_1^n(Y^n)) &= d(X^n \oplus S_2^n, \hat{x}^n(Y^n)) \\ &\geq d(S_2^n, \hat{x}^n(Y^n)) - d(S_2^n, X^n \oplus S_2^n). \end{aligned} \tag{3}$$

The second line follows from the fact that the Hamming distance is a proper distance metric, and it therefore satisfies the triangular inequality. Hence,

$$\begin{aligned} \frac{1}{n} \mathbb{E} d(S_1^n, \hat{s}_1^n(Y^n)) &= \frac{1}{n} \mathbb{E} d(X^n \oplus S_2^n, \hat{x}^n(Y^n)) \\ &\geq \frac{1}{n} \mathbb{E} d(S_2^n, \hat{x}^n(Y^n)) - \frac{1}{n} \mathbb{E} d(S_2^n, X^n \oplus S_2^n). \end{aligned}$$

Let  $Q$  be uniform  $[1 : n]$ , independent of other random variables. Then,

$$\begin{aligned} \frac{1}{n} \mathbb{E} d(S_2^n, X^n \oplus S_2^n) &= \mathbb{E} \left( \frac{1}{n} \sum_{i=1}^n X_i \right) \\ &= \mathbb{E} X_Q \\ &= \mathbb{E} X \end{aligned} \tag{4}$$



This is the expected number of ones in  $X^n$ . For the term,  $\mathbb{E} d(S_2^n, \hat{x}^n(Y^n))/n$ , we lower bound it by

$$\frac{1}{n} \mathbb{E} d(S_2^n, \hat{x}^n(Y^n)) \geq \frac{1}{n} \sum_{i=1}^n \mathbb{E} d(S_{2i}, \hat{s}_{2i}(Y^n)) \quad (5)$$

where  $\hat{s}_2(Y^n)$  is an optimal reconstruction function with respect to Hamming distortion for  $S_2$ . The right hand side of inequality (5) is then further lower bounded by the following argument. From data processing inequality [16], we have

$$\begin{aligned} I(S_2^n; \hat{S}_2^n) &\leq I(S_2^n; Y^n) \\ &\leq \sum_{i=1}^n (H(Y_i) - H(Y_i | S_2^n, X^n)) \\ &= \sum_{i=1}^n H(Y_i) - nH(S_{1i}) \\ &\leq nH(Y) - nH(S_1). \end{aligned}$$

On the other hand,

$$\begin{aligned} I(S_2^n; \hat{S}_2^n) &\geq \sum_{i=1}^n (H(S_{2i}) - H(S_{2i} \oplus \hat{S}_{2i})) \\ &\geq nH_2(S_2) - nH_2 \left( \frac{1}{n} \sum_{i=1}^n \mathbb{E} d(S_{2i}, \hat{s}_{2i}(Y^n)) \right), \end{aligned}$$

where the last line follows from concavity of entropy [16]. Combining the upper and lower bounds gives us

$$\frac{1}{n} \sum_{i=1}^n \mathbb{E} d(S_{2i}, \hat{s}_{2i}(Y^n)) \geq H_2^{-1}(H(S_1) + H(S_2) - H(Y)), \quad (6)$$

where we define  $H_2^{-1}(\cdot) := 0$  if the argument is negative or greater than 1.

Substituting (5), (6) and (4) into (3), we have

$$D(C)_{\min} \geq H_2^{-1}(H(S_1) + H(S_2) - H(Y)) - \mathbb{E} X,$$

where  $\mathbb{E} X \leq C$  from the cost constraint. ■

Using the lower bound in Theorem 3, we can show that when  $p_1 = 1/2$ , symbol by symbol cancellation of  $S_2$  is optimal and hence, when  $p_1 = 1/2$ , the minimum achievable distortion for the same cost constraint is the same regardless of whether  $S_2$  is known causally or noncausally.

**Proposition 3.** *When  $p_1 = 1/2$  and  $p_2 > C$ , the distortion-cost region is given by*

$$D(C)_{\min} = p_2 - C.$$

*Proof:* When  $S_1 \sim \text{Bern}(1/2)$ ,  $Y \sim \text{Bern}(1/2)$ , regardless of the distribution of  $S_2 \oplus X$ . Hence, Theorem 3 reduces to

$$\begin{aligned} D(C)_{\min} &\geq p_2 - \mathbb{E} X \\ &\geq p_2 - C. \end{aligned}$$

Achievability of this lower bound follows from Theorem 1 by setting  $V = \emptyset$ ,  $U$  to be a random variable such that

$$X = \begin{cases} 1 & \text{w. p. } \frac{C}{p_2} & \text{if } S_2 = 1 \\ 0 & \text{w. p. } 1 - \frac{C}{p_2} & \text{if } S_2 = 1 \\ 0 & \text{otherwise} \end{cases} .$$

The existence of such a  $U$  follows from the functional representation lemma [11, Appendix B]. It is easy to verify that the expected cost constraint is satisfied with this choice of distribution  $p(x|s_2)$ . The reconstruction function in this case is simply  $\hat{S}_1 = Y$ . It also easy to verify that the distortion constraint is satisfied. ■

The optimization problem in Theorem 3 can be simplified in a number of cases.

**Corollary 2.** *Theorem 3 simplifies under the following conditions*

1) *Under the condition  $p_1 + (1 - 2p_1)(p_2 - C) \geq 1/2$ , Theorem 3 simplifies to*

$$D(C)_{\min} \geq H_2^{-1}(H(S_1) + H(S_2) - H(p_1 + (1 - 2p_1)(p_2 - C))) - C.$$

2) *Under the condition  $p_1 + (1 - 2p_1)(p_2 + C) \leq 1/2$ , Theorem 3 simplifies to*

$$D(C)_{\min} \geq H_2^{-1}(H(S_1) + H(S_2) - H(p_1 + (1 - 2p_1)(p_2 - C))) - C.$$

*Proof:* The proof follows from observing that  $p_2 - C \leq \mathbb{E}X \oplus S_2 \leq p_2 + C$ . Define  $\mathbb{E}X \oplus S_2 := p_{x \oplus s_2}$ . Then,  $Y \sim \text{Bern}(p_1 + (1 - 2p_1)p_{x \oplus s_2})$ . If condition one in the corollary is satisfied, then  $H(Y)$  is a decreasing function of  $p_{x \oplus s_2}$ . It is then easy to see from the expression in Theorem 3 that the minimizing distribution is one where  $p_{x \oplus s_2} = p_2 - C$  and  $\mathbb{E}X = C$ . A similar proof applies for the second condition, which completes proof of this corollary. ■

It appears to be quite difficult to obtain an explicit analytical solution for the general case of  $p_1 + (1 - 2p_1)(p_2 - C) < 1/2 < p_1 + (1 - 2p_1)(p_2 + C)$ . A looser bound in this case is

**Corollary 3.**

$$D(C)_{\min} \geq H_2^{-1}(H(S_1) + H(S_2) - 1) - C.$$

Proof of this corollary is omitted as it follows directly from Theorem 3.

We now present another lower bound for the binary setting, using ideas from the proof of converse for Gel'fand-Pinsker coding given in [11, Chapter 7], and also ideas from [5]. The main intuition in this lower bound comes from Claim 2 used in the proof of Theorem 3, which shows that the optimum distortion incurred in reconstructing  $X \oplus S_2$  is the same as the optimum distortion incurred in reconstructing  $S_1$ . We then try to lower bound  $D(C)_{\min}$  by lower bounding the distortion incurred in reconstructing  $X \oplus S_2$ . We will see in the sequel that in some cases, this lower bound is better than the previous lower bound given in Theorem 3.

**Theorem 4.** *A lower bound for the achievable distortion for the problem of binary estimation with a helper is given by*

$$D(C)_{\min} \geq \min H_2^{-1}(H(S_1) + H(X \oplus S_2|U) + I(U; S_2) - H(Y)),$$

where we minimize over  $p(u|s_2)$  and  $x = f(u, s_2)$  such that  $\mathbb{E}X \leq C$ . The cardinality of the auxiliary random variable  $U$  may be upper bounded by  $|\mathcal{U}| \leq |\mathcal{S}_2|(|\mathcal{X}| - 1) + 2$ . In the binary case that we are interested in,  $|\mathcal{U}| \leq 4$ .

*Proof:* For notational convenience, let  $\hat{Z}$  represent the optimal reconstruction for  $X \oplus S_2$  and  $Z = X \oplus S_2$ . From data processing inequality,

$$I(Z^n; \hat{Z}^n) \leq I(Y^n; Z^n).$$

On the one hand,

$$\begin{aligned} I(Z^n; \hat{Z}^n) &\geq H(Z^n) - H(\hat{Z}^n|Z^n) \\ &\geq \sum_{i=1}^n (H(Z_i|Z^{i-1}) - H(\hat{Z}_i \oplus Z_i)) \\ &\stackrel{(a)}{\geq} \sum_{i=1}^n (H(Z_i|Z^{i-1}, S_{2,i+1}^n) + I(S_{2,i+1}^n; Z_i|Z^{i-1})) - H_2(D(C)_{\min}) \\ &= \sum_{i=1}^n (H(Z_i|Z^{i-1}, S_{2,i+1}^n) + I(Z^{i-1}; S_{2i}|S_{2,i+1}^n)) - H_2(D(C)_{\min}) \end{aligned}$$

$$\begin{aligned}
&= \sum_{i=1}^n (H(Z_i|Z^{i-1}, S_{2,i+1}^n) + I(Z^{i-1}, S_{2,i+1}^n; S_{2i})) - H_2(D(C)_{\min}) \\
&= \sum_{i=1}^n (H(Z_i|U_i) + I(U_i; S_{2i})) - H_2(D(C)_{\min}).
\end{aligned}$$

In (a), we used concavity of entropy and Claim 2, which states that the optimum distortion for  $X \oplus S_2$  is the same as the optimum distortion for  $S_1$ . Next,

$$\begin{aligned}
I(Y^n; Z^n) &= H(Y^n) - H(Y^n|Z^n) \\
&\leq \sum_{i=1}^n H(Y_i) - H(S_1^n) \\
&= \sum_{i=1}^n (H(Y_i) - H(S_{1i})).
\end{aligned}$$

Defining the standard  $Q$  uniform random variable over  $[1 : n]$  independent of other random variables,  $U = (U_Q, Q)$ ,  $Y_Q = Y$ ,  $S_{1Q} = S_1$ ,  $S_{2Q} = S_2$  and  $Z_Q = Z$  then gives us the following lower bound

$$D \geq H_2^{-1}(H(S_1) + H(Z|U) + I(U; S_2) - H(Y)),$$

where we minimize over  $p(u|s_2)p(x|u, s_2)$  such that  $\mathbb{E} X \leq C$ . Reducing the cost constraint to this single letter expression ( $\mathbb{E} X \leq C$ ) follows the same procedure as in Theorem 3.

Next, we note that instead of minimizing over  $p(x|u, s_2)$ , it suffices to minimize over  $x = f(u, s_2)$ . To see this, note that we can always find a  $V$ , independent of  $U, S_2$ , such that  $p(x|u, s_2) = f(u, v, s_2)$ . Now, define  $\tilde{U} = (U, V)$ . Observe that since we preserve both  $p(x \oplus s_2)$  and  $p(x)$ , the cost constraint and  $H(Y) = H(Z \oplus S_1)$  remains unchanged. Now, note that

$$H(Z|\tilde{U}) \leq H(Z|U),$$

and

$$\begin{aligned}
I(\tilde{U}; S_2) &= I(U; S_2) + I(V; S_2|U) \\
&= I(U; S_2).
\end{aligned}$$

The bound on the cardinality of  $U$  follows from standard techniques and we omit it here. This completes the proof of the lower bound.  $\blacksquare$

Theorem 4 involves minimizing over joint distributions and choice of auxiliary random variable  $U$ . A looser bound that is easier to compute is given by the following corollary.

**Corollary 4.**

$$D(P)_{\min} \geq H_2^{-1}(H_2(p_1) + \min_{p_2 - c \leq \alpha \leq p_2 + c} \{H_2(\alpha) - H_2(\alpha * p_1)\}) - I(U; Z) + I(U; S_2).$$

for some joint distribution  $p(u|s_2)$  and  $x = f(u, s)$  satisfying  $\mathbb{E} X \leq C$ , and  $Z = X \oplus S_2$ .

In Corollary 4, we need to perform maximization of  $I(U; Z) - I(U; S_2)$  subjected to a cost constraint  $\mathbb{E} X \leq C$ . This is nothing but the problem of maximization of the capacity of a Gel'fand-Pinsker channel subjected to a cost constraint. There are efficient numerical algorithms for performing this maximization, cf. [17, Page 555-556] for a description of the algorithm.

*Proof:*

Starting from Theorem 4, consider the term  $H(Z|U) + I(U; S_2) - H(Y)$  in the Theorem.

$$\begin{aligned}
H(Z|U) + I(U; S_2) - H(Y) &= H(Z, U) - H(U|S_2) - H(Y) \\
&= H(Z) - H(Y) + H(U|Z) - H(U|S_2) \\
&= (H(Z) - H(Y)) - (I(U; Z) - I(U; S_2)).
\end{aligned}$$

We now minimize the terms  $(H(Z) - H(Y))$  and  $-(I(U; Z) - I(U; S_2))$  separately. We have discussed maximizing the term  $I(U; Z) - I(U; S_2)$  earlier. As for the term  $(H(Z) - H(Y))$ , using the observation  $p_2 - C \leq E Z \leq p_2 + C$ , we have

$$\min\{H(Z) - H(Y)\} = \min_{p_2 - C \leq \alpha \leq p_2 + C} \{H_2(\alpha) - H_2(\alpha * p_1)\},$$

which completes the proof. ■

### Comparison of lower bounds

As we mentioned, the expressions in Theorems 3 and 4 can be difficult to compute. For the purpose of simulations, we compare the expressions of Corollary 2 with those of Corollary 4, when the conditions given in Corollary 2 are satisfied. Note that since Corollary 4 can be weaker than Theorem 4 whereas Corollary 2 gives the same bounds as Theorem 3 when the conditions are satisfied, an advantage of this comparison is that it shows when Theorem 4 can be strictly larger than Theorem 3.

For our numerical example, we set  $p_2 = 0.1$ , vary the cost from 0.01 to 0.03 and compute plots for  $p_1 = 0.05, 0.09$ . In general, the bound in Theorem 3 is better, but we focus on small values of cost,  $p_1$  and  $p_2$  to show that there are regimes in which the expression in Theorem 4 is better. The plots are shown in Figures 3 and 4. As can be seen in Figure 3, there are regions for which Theorem 4 is strictly better than Theorem 3. However, Theorem 3 does give a better bound for a wider range of values as compared to Theorem 4.

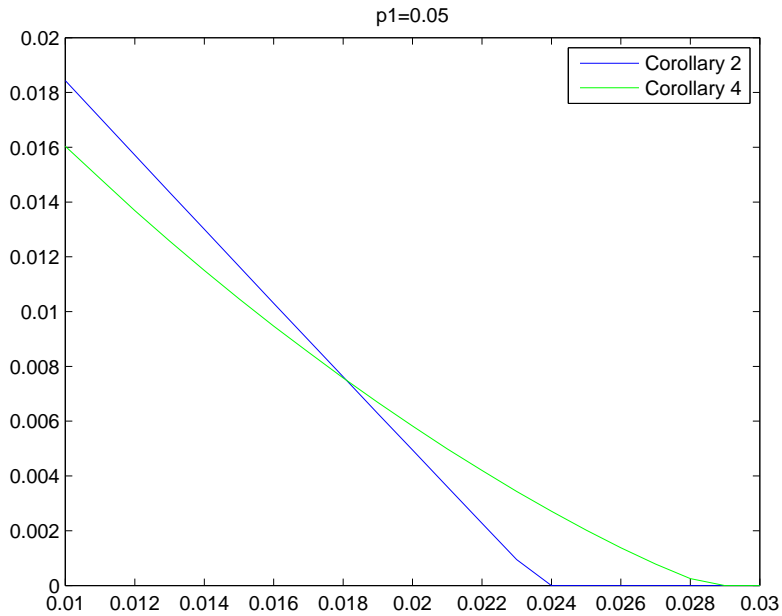


Fig. 3: Comparison of bounds for  $p_1 = 0.05$ . Y-axis represents the distortion level while X-axis represents the cost.

### C. Erasure estimation with helper

For most of this section, we have focused on the binary estimation with helper setup. In this subsection, we briefly mention a setting, erasure estimation with helper, for which we can characterize the distortion-cost function and also, for which symbol by symbol cancellation of  $S_2$  is optimal.

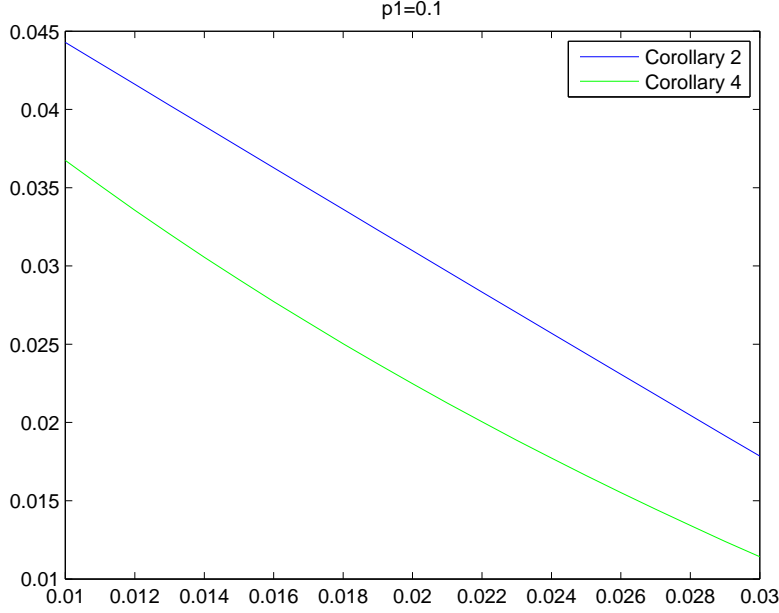


Fig. 4: Comparison of bounds for  $p_1 = 0.1$ . Y-axis represents the distortion level while X-axis represents the cost. In this case, the bound given by Corollary 2 is strictly better than that for Corollary 4.

The setting is defined by  $S_1 \sim p(s_1)$ ,  $S_2 \sim \text{Bern}(p_2)$ ,  $X \in \{0, 1\}$  and  $Y$  is defined as follows

$$Y = \begin{cases} S_1 & \text{if } X \oplus S_2 = 0 \\ e & \text{if } X \oplus S_2 = 1 \end{cases}.$$

This is a model of a channel in which when the interfering signal is large, the desired signal is erased. When the interfering signal is small, decoder receives the signal perfectly. The helper tries to help the decoder by canceling the interference. The distortion-cost region is characterized by the following proposition.

**Proposition 4.** *The distortion-cost region for the problem of erasure estimation with helper is given by*

$$D(C)_{\min} = \min \left\{ P(X \oplus S_2 = 1) \left( \min_{\hat{s}_1} E d(S_1, \hat{s}_1) \right) + P(X \oplus S_2 = 0) \left( \min_{\hat{s}_1} E d(S_1, \hat{s}_1) \right) \right\},$$

where the minimization is over  $p(x|s_2)$  satisfying  $E \rho(X) \leq C$ .

*Proof:* Achievability of the distortion-cost region uses a modified version of the achievability scheme used in Proposition 3. The modification comes in the reconstruction function where

$$\hat{s}_1(Y) = \begin{cases} \arg \min_x d(Y, x) & \text{if } Y = S_1, \\ \arg \min_x E d(S_1, x) & \text{if } Y = e \end{cases}.$$

With this choice of reconstruction function and noting that  $P(Y = S_1) = P(X \oplus S_1 = 0)$  and  $P(Y = e) = P(X \oplus S_1 = 1)$ , it is easy to see that the achievable distortion-cost region simplifies to the expression given in the Proposition.

For the converse, fixing a  $(n, C)$  code achieving distortion  $D$ , we have

$$D = \frac{1}{n} \sum_{i=1}^n E d(S_{1i}, \hat{S}_{1i}(Y^n)).$$

Consider now the term  $\mathbb{E}d(S_{1i}, \hat{S}_{1i}(Y^n))$ . We have

$$\begin{aligned}
\mathbb{E}d(S_{1i}, \hat{S}_{1i}(Y^n)) &= \sum p(s_{1i}, y^{n \setminus i}, y_i) d(s_{1i}, \hat{s}_{1i}(y^n)) \\
&= \sum \left( p(s_{1i}, y^{n \setminus i}, y_i, x \oplus s_{2i} = 0) d(s_{1i}, \hat{s}_{1i}(y^n)) \right. \\
&\quad \left. + p(s_{1i}, y^{n \setminus i}, y_i, x \oplus s_{2i} = 1) d(s_{1i}, \hat{s}_{1i}(y^n)) \right) \\
&\stackrel{(a)}{=} \sum \left( p(s_{1i}, y^{n \setminus i}, y_i, x \oplus s_{2i} = 0) d(s_{1i}, \hat{s}_{1i}(y^n)) \right. \\
&\quad \left. + p(s_{1i}, y^{n \setminus i}, x \oplus s_{2i} = 1) d(s_{1i}, \hat{s}_{1i}(y^{n \setminus i}, y_i = e)) \right) \\
&= \sum p(s_{1i}, x \oplus s_{2i} = 0) p(y^{n \setminus i}, y_i | s_{1i}, x \oplus s_{2i} = 0) d(s_{1i}, \hat{s}_{1i}(y^n)) \\
&\quad + \sum p(s_{1i}, y^{n \setminus i}, x \oplus s_{2i} = 1) d(s_{1i}, \hat{s}_{1i}(y^{n \setminus i}, y_i = e)),
\end{aligned}$$

where (a) follows from the fact that when  $x \oplus s_{2i} = 1$ ,  $y_i = e$ . Next, focusing on the first term in the sum, we note that  $\mathbb{P}(Y_i = S_{1i} | X \oplus S_2 = 0, S_{1i}) = 1$ . Hence, using  $1_{\{\cdot\}}$  to denote the indicator function, the first term simplifies to the following

$$\begin{aligned}
&\sum p(s_{1i}, x \oplus s_{2i} = 0) p(y^{n \setminus i}, y_i | s_{1i}, x \oplus s_{2i} = 0) d(s_{1i}, \hat{s}_{1i}(y^n)) \\
&= \sum p(s_{1i}, x \oplus s_{2i} = 0) p(y^{n \setminus i} | s_{1i}, x \oplus s_{2i} = 0, y_i) 1_{y_i = s_{1i}} d(s_{1i}, \hat{s}_{1i}(y^n)) \\
&\geq \sum p(s_{1i}, x \oplus s_{2i} = 0) p(y^{n \setminus i} | s_{1i}, x \oplus s_{2i} = 0, y_i) \left( \min_{x \in \hat{S}_1} d(s_{1i}, x) \right) \\
&= \sum p(s_{1i}) p(x \oplus s_{2i} = 0) \left( \min_{x \in \hat{S}_1} d(s_{1i}, x) \right) \\
&= \mathbb{P}(X_i \oplus S_{2i} = 0) \mathbb{E} \left( \min_{x \in \hat{S}_1} d(s_{1i}, x) \right).
\end{aligned}$$

Hence,  $\mathbb{E}d(S_{1i}, \hat{S}_{1i}(Y^n))$  is lower bounded by

$$\begin{aligned}
\mathbb{E}d(S_{1i}, \hat{S}_{1i}(Y^n)) &\geq \mathbb{P}(X_i \oplus S_{2i} = 0) \mathbb{E} \left( \min_{x \in \hat{S}_1} d(s_{1i}, x) \right) \\
&\quad + \sum p(s_{1i}, y^{n \setminus i}, x \oplus s_{2i} = 1) d(s_{1i}, \hat{s}_{1i}(y^{n \setminus i}, y_i = e)), \\
&= \mathbb{P}(X_i \oplus S_{2i} = 0) \mathbb{E} \left( \min_{x \in \hat{S}_1} d(s_{1i}, x) \right) \\
&\quad + \sum p(s_{1i}) p(y^{n \setminus i}, x \oplus s_{2i} = 1) d(s_{1i}, \hat{s}_{1i}(y^{n \setminus i}, y_i = e)) \\
&\geq \mathbb{P}(X_i \oplus S_{2i} = 0) \mathbb{E} \left( \min_{x \in \hat{S}_1} d(s_{1i}, x) \right) \\
&\quad + \sum p(y^{n \setminus i}, x \oplus s_{2i} = 1) \left( \min_{x \in \hat{S}_1} \sum p(s_{1i}) d(s_{1i}, x) \right) \\
&= \mathbb{P}(X_i \oplus S_{2i} = 0) \mathbb{E} \left( \min_{x \in \hat{S}_1} d(s_{1i}, x) \right) \\
&\quad + \sum \mathbb{P}(X_i \oplus S_{2i} = 1) \left( \min_{x \in \hat{S}_1} \mathbb{E}d(S_{1i}, x) \right).
\end{aligned}$$

We note now that if  $y_i = 0$  or  $1$ , then we can achieve the minimum possible distortion  $\min_{\hat{s}_1} d(s_{1i}, \hat{s}_1)$  using only knowledge of  $y_i$ , since  $s_{1i}$  is known in this case.

We therefore obtain

$$\begin{aligned}
D &= \frac{1}{n} \sum_{i=1}^n \mathbb{E} d(S_{1i}, \hat{S}_{1i}(Y^n)) \\
&\geq \frac{1}{n} \sum_{i=1}^n \left( \mathbb{P}(X_i \oplus S_{2i} = 0) \mathbb{E} \left( \min_{x \in \hat{S}_1} d(s_{1i}, x) \right) \right. \\
&\quad \left. + \sum \mathbb{P}(X_i \oplus S_{2i} = 1) \left( \min_{x \in \hat{S}_1} \mathbb{E} d(S_{1i}, x) \right) \right).
\end{aligned}$$

Defining  $Q \sim \text{Unif}[1 : n]$  independent of other random variables then give us

$$\begin{aligned}
D &\geq \mathbb{P}(X \oplus S_2 = 1) (\min_{\hat{s}_1} \mathbb{E} d(S_1, \hat{s}_1)) \\
&\quad + \mathbb{P}(X \oplus S_2 = 0) \mathbb{E} (\min_{\hat{s}_1} d(S_1, \hat{s}_1)).
\end{aligned}$$

For the cost constraint, we have

$$\begin{aligned}
\mathbb{E} \left( \frac{1}{n} \sum_{i=1}^n X_i \right) &= \mathbb{E} X_Q \\
&= \mathbb{E} X,
\end{aligned}$$

which completes the proof. ■

## V. GAUSSIAN ESTIMATION WITH HELPER

In this section, we extend our setup to the Gaussian case, where  $S_1 \sim N(0, 1)$ ,  $S_2 \sim N(0, P_2)$ ,  $Y = X + S_1 + S_2$ ,  $d(S_1, \hat{S}_1) = (S_1 - \hat{S}_1)^2$  and the cost constraint is  $\mathbb{E} X^2 \leq P$ . As we mentioned in the Introduction, the problem in the Gaussian case is equivalent to the problem of Assisted Interference Suppression considered in [3]. We present a new lower bound for this problem that can improve on that derived in [3] and [6]. The lower bound derived in [6] includes the lower bound derived in [3] as a special case and can be strictly better, but for clarity of presentation, we will first compare our lower bound to that in [3] in subsection V-C, and then compare our bound with the lower bound derived in [6] in subsection V-D. We begin with an achievability argument based on Theorem 2.

### A. Achievable distortion-cost region

We specialize Theorem 2 to the Gaussian case by choosing the auxiliary random variables as Gaussian random variables. The achievability scheme presented here is essentially the same as the scheme presented in [3], but we derive it via different means.

**Theorem 5.** *An achievable distortion for the problem of Gaussian estimation with a helper is given by*

$$D(P)_{\min} \leq \inf 1 - \frac{\mathbb{E} U^2}{\mathbb{E} Y^2 \mathbb{E} U^2 - (\mathbb{E} UY)^2},$$

where

$$\begin{aligned}
\sqrt{P'} &= \frac{-2\alpha\beta\sqrt{P} + \sqrt{4\alpha^2\beta^2P + 4(1-\alpha^2)P}}{2}, \\
\mathbb{E} U^2 &= P' + 2\gamma\beta\sqrt{P'P_2} + \gamma^2P_2, \\
\mathbb{E}(UY) &= \alpha\beta\sqrt{PP'} + P' + \alpha\gamma\sqrt{PP_2} + \gamma\beta\sqrt{P'P_2} + \beta\sqrt{P'P_2} + \gamma P_2, \\
\mathbb{E} Y^2 &= P + 1 + P_2 + 2\alpha\sqrt{PP_2} + 2\beta\sqrt{P'P_2}.
\end{aligned}$$

and the infimum is taken over  $-1 \leq \alpha \leq 1$ ,  $-1 \leq \beta \leq 1$  and  $\gamma \in \mathcal{R}$  satisfying the constraint

$$(1 - \beta^2)P' > \mathbb{E} U^2 - \frac{(\mathbb{E}(UY))^2}{\mathbb{E} Y^2}.$$

We defer the proof of Theorem 5 to Appendix III.

Similar to the binary setup, we can derive a nontrivial condition between  $P$  and the power of the source  $S_1$  (normalized to 1), such that zero expected distortion can be achieved.

**Proposition 5.** *For the problem of Gaussian estimation with a helper,  $D(C)_{\min} = 0$  if*

$$P > 1 - \frac{1}{P + P_2 + 1}.$$

*Proof:* Proof of this Proposition follows from a choice of  $\alpha$  and  $\beta$  in Theorem 5. However, we give a slightly different proof that gives more intuition to this condition and also has parallels with the problem of dirty paper coding [18] (see also [11, Chapter 7]).

Starting from Theorem 2, we let  $U = X + S_2$ , where  $X \sim N(0, P)$  independent of  $S_2$ . Note that the cost constraint is satisfied from this choice of  $U$ . If the decoder can decode  $U$ , then the distortion incurred is zero, since  $S_1 = Y - U$ . It therefore remains to satisfy the decoding condition, which is

$$I(U; Y) > I(U; S_2).$$

Since all the random variables are Gaussian, this decoding condition reduces to  $h(U|S_2) > h(U|Y)$ .

$$\begin{aligned} h(U|S_2) &= h(X|S_2) \\ &= \frac{1}{2} \log 2\pi e P. \end{aligned}$$

On the other hand,

$$\begin{aligned} h(U|Y) &= h(-S_1|Y) \\ &\stackrel{(a)}{=} h(-S_1 - \mathbb{E}(-S_1|Y)) \\ &= \frac{1}{2} \log 2\pi e \left(1 - \frac{1}{P_1 + P_2 + 1}\right), \end{aligned}$$

where (a) follows from the fact that for Gaussian random variables, the difference between  $S_1$  and its Minimum Mean Square Error Estimator is independent of the observation,  $Y$ .

We therefore derive the condition

$$P > 1 - \frac{1}{P + P_2 + 1}. \quad \blacksquare$$

Note that, similar to the binary case, the expected distortion can be made to be zero even if  $P_2$  is much larger than  $P$ .

### B. Lower bounds

We now turn to lower bounds for the problem of Gaussian Estimation with helper. We first state the following lower bound given in [3] and its improved version given in [6].

**Theorem 6.** [3] *A lower bound for the problem of Gaussian estimation with helper is given by*

$$D(P)_{\min} \geq \left( \left[ \sqrt{\frac{P_2}{P_2 2 + 2\sqrt{PP_2} + P + 1}} - \sqrt{P} \right]^+ \right)^2,$$

where  $[\cdot]^+$  denotes the positive part.

As shown in [6], the lower bound given in Theorem 6 can be improved to the following.



**Theorem 7.** [6] A lower bound for the problem of Gaussian estimation with helper is given by

$$D(P)_{\min} \geq \inf_{\sigma_{XS_2}} \sup_{\gamma > 0} \frac{1}{\gamma^2} \left( \left[ \sqrt{\frac{P_2}{1 + P_2 + P + 2\sigma_{XS_2}}} - \sqrt{(1 - \gamma)^2 P_2 + \gamma^2 P - 2\gamma(1 - \gamma)\sigma_{XS_2}} \right]^+ \right)^2,$$

where  $[\cdot]^+$  denotes the positive part and  $\sigma_{XS_2} \in [-\sqrt{P_2}\sqrt{P}, \sqrt{P_2}\sqrt{P}]$ .

From the lower bound in Theorem 6 and Proposition 5, we can show that as the power of the interfering signal goes to infinity,  $P_2 \rightarrow \infty$ , zero expected distortion is achievable if and only if  $P \geq 1$ .

**Proposition 6.**  $\lim_{P_2 \rightarrow \infty} D(P)_{\min} = 0$  if and only if  $P \geq 1$ .

*Proof:* From Proposition 5, the sufficient condition for zero expected distortion reduces to  $P > 1$  as  $P_2 \rightarrow \infty$ . From Theorem 6, we can show that this is also necessary. From Theorem 6,  $\lim_{P_2 \rightarrow \infty} D(P)_{\min} \geq \left( [1 - \sqrt{P}]^+ \right)^2$ , which is zero if and only if  $P \geq 1$ . ■

We now turn to our lower bound. For clarity, we first present a proof of a special case of our lower bound before turning to the more general expression.

**Proposition 7.** A lower bound for the problem of Gaussian estimation with a helper is given by

$$D(P)_{\min} \geq \frac{1}{(\gamma - 1)} \left[ \ln \left( \frac{1 + \gamma P_2}{1 + P_2} \right) + \frac{2\sqrt{P}}{\sqrt{P_2}(1 + \gamma P_2)} - \frac{2\sqrt{P}}{\sqrt{P_2}(1 + P_2)} - \gamma P \right],$$

for any  $\gamma \geq 1$ .

It should be noted that while finding the optimal value of  $\gamma$  that *maximizes* this lower bound is a hard optimization problem, *any*  $\gamma \geq 1$  constitutes a lower bound for  $D(P)_{\min}$ . Hence, Proposition 7 in fact gives a *family* of lower bounds.

*Proof:* This proof hinges on an application of a relationship between mismatched estimation and relative entropy given in [7, Equality (14)]. The main idea behind the proof lies in considering a decoder that performs the estimation (and reconstruction) using a wrong (or mismatched) distribution for  $P_{S_1^n|Y^n}$ . In particular, we will consider a *mismatched* decoder that attempts to estimate  $S_1^n$  assuming that  $X^n \equiv 0$ . That is, the decoder assumes that the encoder does not do anything to help the decoder. The estimation error incurred by the mismatched decoder,  $MSE_Q$ , is clearly larger than that incurred by an optimum decoder that uses the correct (true) distribution,  $D(P)_{\min}$ . We then rely on results in [7] to lower bound the difference between  $D(P)_{\min}$  and  $MSE_Q$ , thereby giving us a lower bound on  $D(P)_{\min}$ .

To derive our bound, we first consider a more general source  $S_1 \sim N(0, 1/\gamma)$  and let  $S_2 \sim N(0, P_2)$  as before. The value of  $\gamma$  that we are concerned about is  $\gamma = 1$ , which will appear later in the proof.

Define  $MSE_Q(\gamma)$  as

$$MSE_Q(\gamma) := \mathbb{E} \left\| S_1^n - \frac{\frac{1}{\gamma}}{\frac{1}{\gamma} + P_2} (X^n + S_1^n + S_2^n) \right\|^2.$$

Let  $\alpha = \frac{1}{\frac{1}{\gamma} + P_2}$  and note that  $\hat{S}_1 = \alpha Y$  is the Minimum Mean Square Error (MMSE) estimate of  $S_1$  that the decoder would employ if it assumes that  $X^n \equiv 0$ . We first give a lower bound for  $MSE_Q(\gamma)$ . Note that under the true distribution,  $\mathbb{E} \|X^n\|^2 \leq nP$ .

$$\begin{aligned} \mathbb{E} \|S_1^n - \alpha(X^n + S_1^n + S_2^n)\|^2 &= \mathbb{E} \|S_1^n - \alpha(S_1^n + S_2^n)\|^2 - 2\alpha \mathbb{E} \langle S_1^n - \alpha(S_1^n + S_2^n), X^n \rangle + \alpha^2 \mathbb{E} \|X^n\|^2 \\ &= n\alpha P_2 + 2\alpha^2 \mathbb{E} \langle S_2^n, X^n \rangle + \alpha^2 \mathbb{E} \|X^n\|^2 \\ &\stackrel{(a)}{\geq} n\alpha P_2 - 2\alpha^2 \sqrt{\mathbb{E} \|S_2^n\|^2 \mathbb{E} \|X^n\|^2} + \alpha^2 \mathbb{E} \|X^n\|^2 \\ &\geq n\alpha P_2 - 2\alpha^2 \sqrt{n^2 P_2 P} \\ &= n\alpha P_2 - 2n\alpha^2 \sqrt{P_2 P}, \end{aligned}$$

where (a) follows by Cauchy-Schwartz inequality.

Now, let  $\tilde{S}^n = S_2^n + X^n$  and let  $P_{\tilde{S}^n}$  denote the distribution of  $S_2^n + X^n$  under the optimum encoding scheme. Let  $Q_{\tilde{S}^n}$  denote the corresponding distribution under the encoding scheme of  $X^n \equiv 0$ . Note now that

$$\begin{aligned}
MSE_Q(\gamma) &= \mathbb{E} \|S_1^n - \alpha(S_1^n + \tilde{S}^n)\|^2 \\
&\stackrel{(a)}{=} \mathbb{E} \|Y^n - \tilde{S}^n - \mathbb{E}_Q(S_1^n|Y^n)\|^2 \\
&\stackrel{(b)}{=} \mathbb{E} \|Y^n - \tilde{S}^n - (Y^n - \mathbb{E}_Q(\tilde{S}^n|Y^n))\|^2 \\
&= \mathbb{E} \|\tilde{S}^n - \mathbb{E}_Q(\tilde{S}^n|Y^n)\|^2 \\
&:= MSE_{Q, \tilde{S}^n}(\gamma).
\end{aligned} \tag{7}$$

(a) follows from the fact that  $\alpha(S_1^n + \tilde{S}^n)$  is the optimum MMSE estimator for  $S_1^n$  under  $Q$ ; that is, under the assumption of  $X^n \equiv 0$ . (b) follows from  $S_1^n = Y^n - \tilde{S}^n$ .

Next, note that this analysis also holds when the decoder knows that  $\tilde{S}^n$  is distributed according to  $P_{\tilde{S}^n}$ . That is, we have

$$\begin{aligned}
MMSE(\gamma) &:= \mathbb{E} \|S_1^n - \mathbb{E}_P(S_1^n|Y^n)\|^2 \\
&= \mathbb{E} \|\tilde{S}^n - \mathbb{E}_P(\tilde{S}^n|Y^n)\|^2 \\
&:= MMSE_{P, \tilde{S}^n}(\gamma).
\end{aligned} \tag{8}$$

Note that  $nD(P)_{\min} = MMSE(1)$ .

We now relate  $MSE_Q(\gamma)$  to the optimum  $MMSE$  of  $S_1^n$  given that an optimum estimator and coding scheme were used. From (7) and (8), we see that it suffices to consider  $MSE_{Q, \tilde{S}^n}(\gamma)$  and  $MMSE_{P, \tilde{S}^n}(\gamma)$ . Using the relation between mismatched estimation and relative entropy given in [7, Equality 14], we have

$$D(P_{Y^n}^{(\gamma_0)} \| Q_{Y^n}^{(\gamma_0)}) = \frac{1}{2} \int_0^{\gamma_0} MSE_{Q, \tilde{S}^n}(\gamma) - MMSE_{P, \tilde{S}^n}(\gamma) d\gamma \tag{9}$$

Here,  $P_{Y^n}$  represents the distribution of  $Y^n$  induced by  $P_{\tilde{S}^n}$ . Similarly,  $Q_{Y^n}$  represents the distribution of  $Y^n$  induced by  $Q_{\tilde{S}^n}$ .

We first give a bound on  $D(P_{Y^n}^{(\gamma)} \| Q_{Y^n}^{(\gamma)})$ .

$$\begin{aligned}
D(P_{Y^n}^{(\gamma)} \| Q_{Y^n}^{(\gamma)}) &\leq D(P_{Y^n, S_2^n}^{(\gamma)} \| Q_{Y^n, S_2^n}^{(\gamma)}) \\
&= \mathbb{E}_{S_2^n} D(P_{Y^n|S_2^n}^{(\gamma)} \| Q_{Y^n|S_2^n}^{(\gamma)}).
\end{aligned}$$

Note that since  $X^n$  is a function of  $S_2^n$ , we have the following.

$$\text{Under } P_{Y^n|S_2^n}^{(\gamma)} : Y^n|S_2^n \sim N(S_2^n + X^n, \frac{1}{\gamma} I_{n \times n}),$$

$$\text{Under } Q_{Y^n|S_2^n}^{(\gamma)} : Y^n|S_2^n \sim N(S_2^n, \frac{1}{\gamma} I_{n \times n}).$$

Hence,  $D(P_{Y^n|S_2^n}^{(\gamma)} \| Q_{Y^n|S_2^n}^{(\gamma)})$  is given by the divergence between two multivariate Gaussian random variables with the same covariance matrix. In our case, the divergence is given by

$$D(P_{Y^n|S_2^n}^{(\gamma)} \| Q_{Y^n|S_2^n}^{(\gamma)}) = \frac{1}{2} \gamma \|X^n\|^2.$$

Hence,

$$\begin{aligned}
D(P_{Y^n}^{(\gamma)} \| Q_{Y^n}^{(\gamma)}) &\leq \mathbb{E}_{S_2^n} D(P_{Y^n|S_2^n}^{(\gamma)} \| Q_{Y^n|S_2^n}^{(\gamma)}) \\
&= \mathbb{E}_{S_2^n} \left( \frac{\gamma}{2} \|X^n\|^2 \right) \\
&\leq \frac{n\gamma P}{2}
\end{aligned}$$

From (9), we have

$$\int_{\gamma_0}^{\gamma_1} MSE_{Q, \tilde{s}^n}(\gamma) - MMSE_{P, \tilde{s}^n}(\gamma) d\gamma = 2D(P_{Y^n}^{(\gamma_1)} || Q_{Y^n}^{(\gamma_1)}) - 2D(P_{Y^n}^{(\gamma_0)} || Q_{Y^n}^{(\gamma_0)})$$

for  $\gamma_1 \geq \gamma_0$ . Hence,

$$\begin{aligned} \int_{\gamma_0}^{\gamma_1} MMSE_{P, \tilde{s}^n}(\gamma) d\gamma &\geq \int_{\gamma_0}^{\gamma_1} MSE_{Q, \tilde{s}^n}(\gamma) d\gamma - 2D(P_{Y^n}^{(\gamma_1)} || Q_{Y^n}^{(\gamma_1)}) \\ &\geq \int_{\gamma_0}^{\gamma_1} MSE_{Q, \tilde{s}^n}(\gamma) d\gamma - n\gamma_1 P \end{aligned} \quad (10)$$

Since  $MMSE_{P, \tilde{s}^n}(\gamma)$  is a non-increasing function in  $\gamma$ , we have  $\int_{\gamma_0}^{\gamma_1} MMSE_{P, \tilde{s}^n}(\gamma) d\gamma \leq (\gamma_1 - \gamma_0)MMSE_{P, \tilde{s}^n}(\gamma_0) = (\gamma_1 - \gamma_0)MMSE(\gamma_0)$ . Next, we note that  $\alpha = 1/(1 + \gamma P_2)$ , so we can write

$$\begin{aligned} MSE_{Q, \tilde{s}^n}(\gamma) &= MSE_Q(\gamma) \\ &\geq \frac{nP_2}{1 + \gamma P_2} - \frac{2n\sqrt{P_2 P}}{(1 + \gamma P_2)^2}. \end{aligned}$$

From (10) and the arguments above, we have

$$\begin{aligned} (\gamma_1 - \gamma_0)MMSE(\gamma_0) &\geq \int_{\gamma_0}^{\gamma_1} \frac{nP_2}{1 + \gamma P_2} - \frac{2n\sqrt{P_2 P}}{(1 + \gamma P_2)^2} d\gamma - n\gamma_1 P \\ &= n \ln\left(\frac{1 + \gamma_1 P_2}{1 + \gamma_0 P_2}\right) + \frac{2n\sqrt{P}}{\sqrt{P_2}(1 + \gamma_1 P_2)} - \frac{2n\sqrt{P}}{\sqrt{P_2}(1 + \gamma_0 P_2)} - n\gamma_1 P. \end{aligned}$$

Finally, using the relationship that  $D(P)_{\min} = MMSE(1)/n$ ,  $\gamma_0 = 1$  and the above completes the proof of the lower bound.  $\blacksquare$

In Proposition 7, we related the minimum mean square error that a decoder incurs when it uses the true distribution to the mean square error incurred by a decoder if it uses the possibly erroneous distribution of  $X^n \equiv 0$ . Clearly, we do not need to choose  $X^n \equiv 0$  as the erroneous distribution, but we can also choose other distributions. This is the main idea behind our generalization of Proposition 7, which we state in Theorem 8.

**Theorem 8.** *A lower bound for the problem of Gaussian estimation with helper is given by*

$$\begin{aligned} (\gamma - 1)D(P)_{\min} &\geq \log\left(\frac{1 + \gamma P_I}{1 + P_I}\right) + \frac{1}{(1 + \gamma P_I)} - \frac{1}{(1 + P_I)} \\ &\quad - \frac{P_2}{P_I(1 + \gamma P_I)} + \frac{P_2}{P_I(1 + P_I)} - \frac{c^2 \gamma}{1 + \gamma r P} P_2 \\ &\quad - \frac{1}{1 + \gamma_1 r P} + 1 + \log\left(\frac{1}{1 + \gamma_1 r P}\right) \\ &\quad + ax^{*2} - bx^*, \end{aligned}$$

where  $a = \frac{1}{P_I(1+P_I)} - \frac{1}{P_I(1+\gamma P_I)} - \frac{\gamma}{1+\gamma r P}$ ,  $b = |2(\frac{1}{P_I(1+P_I)} - \frac{1}{P_I(1+\gamma P_I)} + \frac{c\gamma}{1+\gamma r P})|\sqrt{P_2}$  and

$$x^* = \begin{cases} \sqrt{P} & \text{if } a \leq 0 \\ b/2a & \text{if } a > 0 \text{ and } b/2a < \sqrt{P} \\ \sqrt{P} & \text{otherwise} \end{cases},$$

for any  $\gamma \geq 1$ , real number  $c$  and  $r \geq 0$ .

As with Proposition 7, Theorem 8 gives a family of bounds. Any  $\gamma > 1$ , real number  $c$  and  $r \geq 0$  yields a bound on the achievable distortion. Theorem 8 is proved in Appendix IV.

### C. Comparison of bounds I

We now show some plots comparing the various bounds we derived with the lower bound proposed in [3] (Theorem 6). For the purpose of comparisons, we set  $P_2$  at a fixed level and vary the power of the encoder. We then compute the lower bounds on distortion given in Theorem 5, Proposition 7, Theorem 8 as well as the achievable distortion given in Theorem 5.

The plots for  $P_2 = 0.1$ ,  $P_2 = 1$  and  $P_2 = 10$  are shown in Figures 5, 6 and 7 respectively. As we can see from the plots, the generalized lower bound in Theorem 8 can significantly improve on the lower bound of Theorem 5 for several different levels of  $P_2$ .

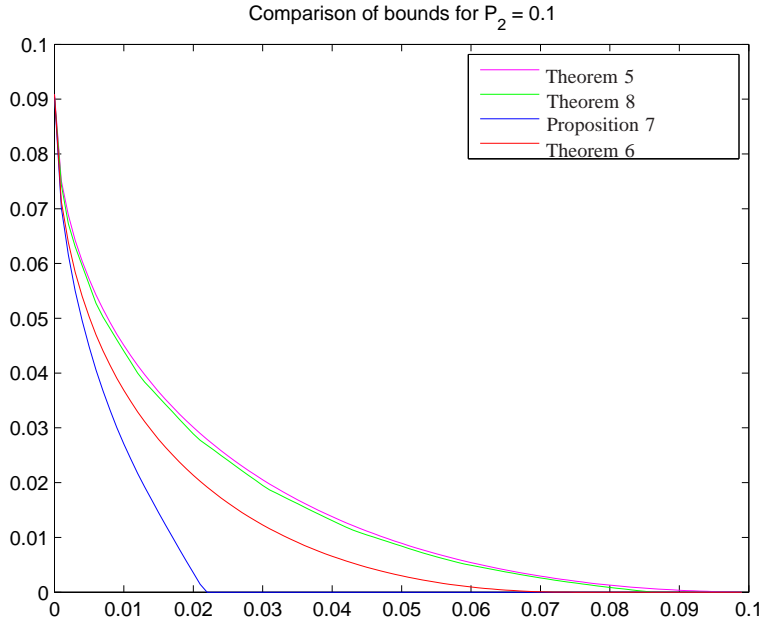


Fig. 5: Comparison of bounds for  $P_2 = 0.1$ . Y-axis represents distortion level and X-axis represents the power constraint.

### D. Comparison of bounds II

In this subsection, we compare our lower bound given in Theorem 8 to the lower bound given in [6] (Theorem 7). For ease of numerical computation, we compare our lower bound to the following *upper bound* on Theorem 7.

$$D(P)_{\min} \geq \min_{\sigma_{XS_2} \in \mathcal{A}} \sup_{\gamma > 0} \frac{1}{\gamma^2} \left( \left[ \sqrt{\frac{P_2}{1 + P_2 + P + 2\sigma_{XS_2}}} - \sqrt{(1 - \gamma)^2 P_2 + \gamma^2 P - 2\gamma(1 - \gamma)\sigma_{XS_2}} \right]^+ \right)^2, \quad (11)$$

where  $[\cdot]^+$  denotes the positive part and  $\mathcal{A}$  is a discretization of the interval  $[-\sqrt{P_2}\sqrt{P}, \sqrt{P_2}\sqrt{P}]$ . The plots showing comparisons of the lower bound proposed in Theorem 8 and the lower bound given in inequality (11) for  $P_2 = 1, 10, 100$  are given in Figures 8, 9 and 10 respectively.

As can be seen from the plots, the two bounds now cross each other. While the lower bound given [6] can be better than that given in Theorem 8 in some regimes, we can also see that Theorem 8 can be strictly better than Theorem 7 in other regimes, particularly when  $P_2$  is large and the power budget  $P$  of the encoder is small.

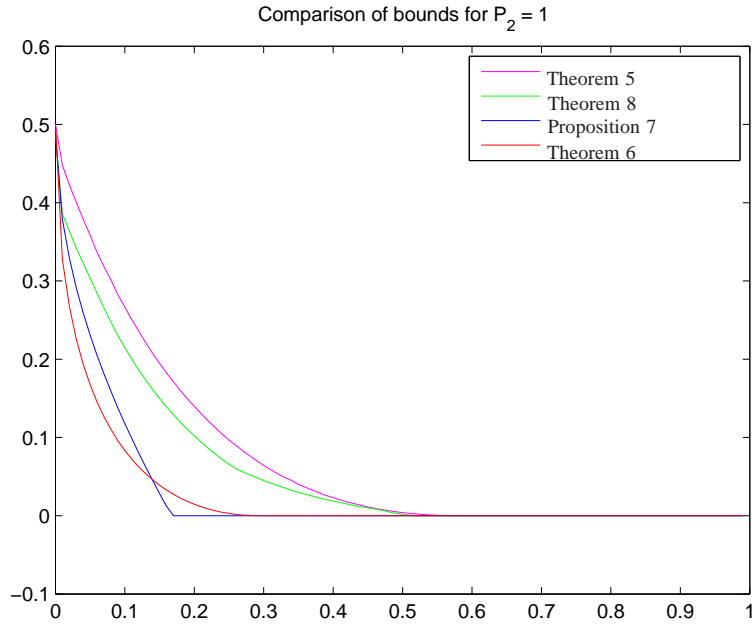


Fig. 6: Comparison of bounds for  $P_2 = 1$ . Y-axis represents distortion level and X-axis represents the power constraint.

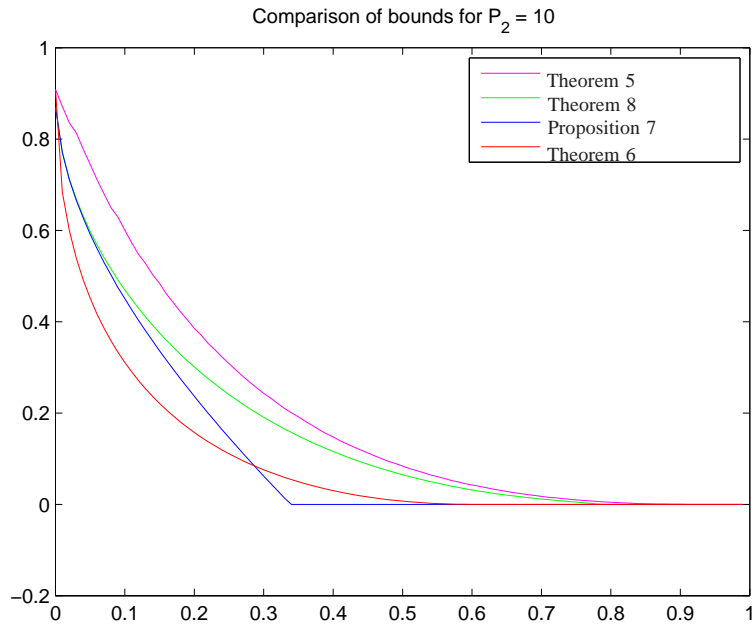


Fig. 7: Comparison of bounds for  $P_2 = 10$ . Y-axis represents distortion level and X-axis represents the power constraint.

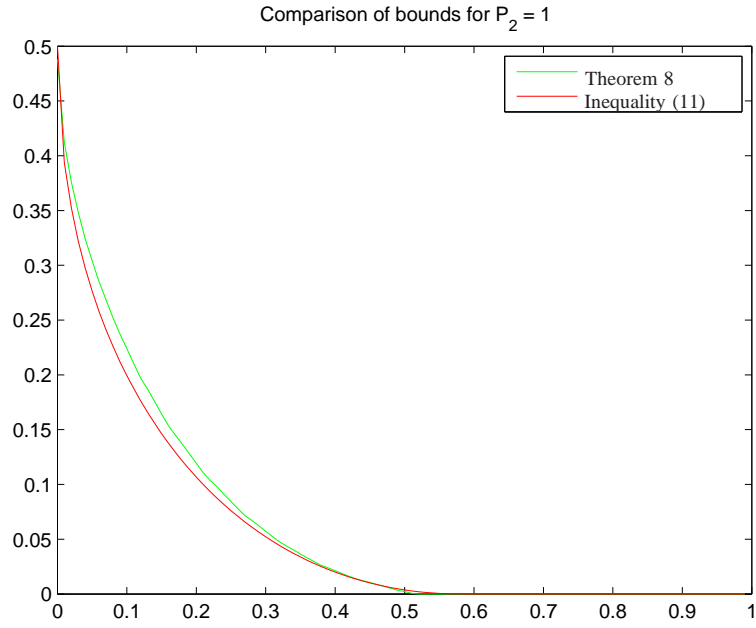


Fig. 8: Comparison of bounds for  $P_2 = 1$ . Y-axis represents distortion level and X-axis represents the power constraint.

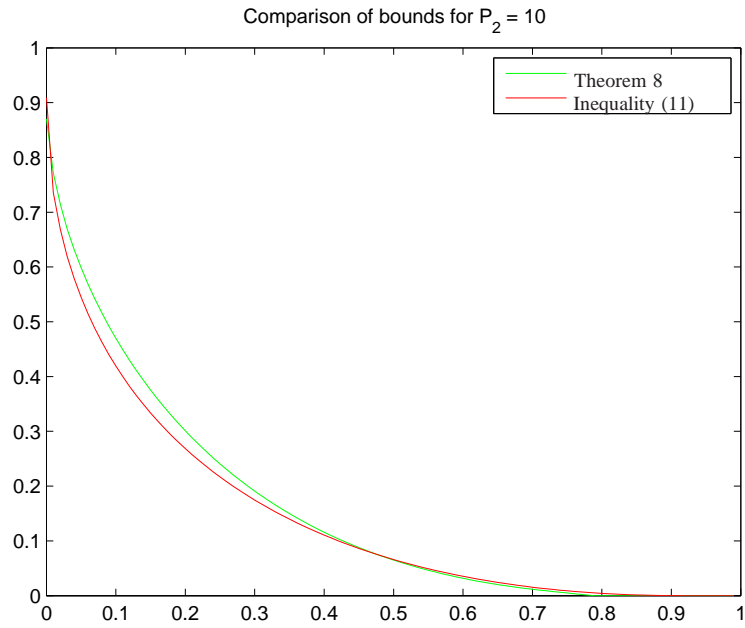


Fig. 9: Comparison of bounds for  $P_2 = 10$ . Y-axis represents distortion level and X-axis represents the power constraint.

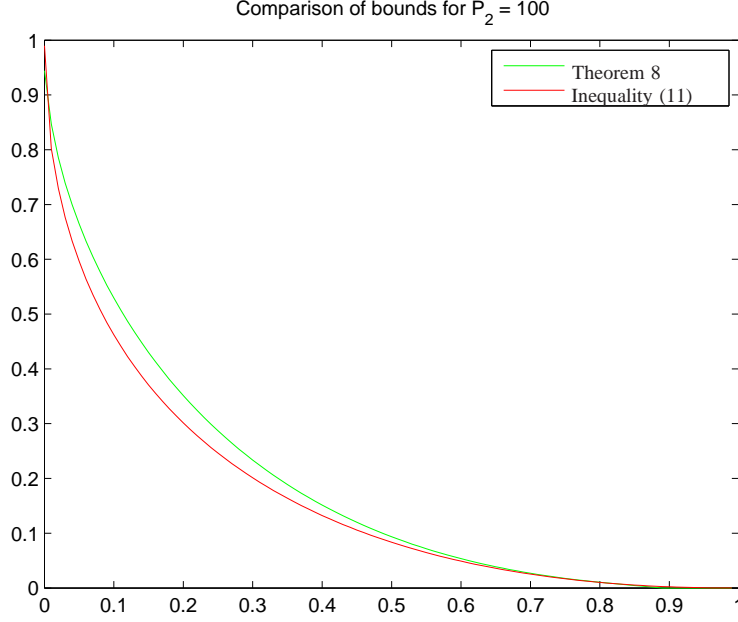


Fig. 10: Comparison of bounds for  $P_2 = 100$ . Y-axis represents distortion level and X-axis represents the power constraint.

## VI. WHEN $S_1$ IS ALSO AVAILABLE AT THE ENCODER

In this section, we turn our attention to the problem of reconstructing  $S_1$  when both  $S_1$  and  $S_2$  are available at the encoder, as defined in Section II-B. As with previous sections, the focus of this section is on lower bounds for this setup, but we also use lower and upper bounds to derive constant multiplicative gap results between the achievable distortions and lower bounds. As we mentioned in the Introduction, our setting is a special case of the setting considered in [12]. We first review some known results found in that paper specialized to our setting, and then present our results, which include a generalization of the lower bound [12] that can be strictly larger.

### A. Upper and lower bounds

We first present an achievability scheme for this setting.

**Theorem 9.** (See also [12]) *An achievable distortion-cost region for the problem of estimation with a helper who has non-causal access to both the interference and the signal is given by*

$$D(P)_{\min} \leq \frac{P_1}{\left(1 + \frac{(\alpha\sqrt{\frac{P}{P_1}}+1)^2 P_1}{(\beta\sqrt{\frac{P}{P_2}}+1)^2 P_2 + P(1-\alpha^2-\beta^2) + N}\right) \left(1 + \frac{P(1-\alpha^2-\beta^2)}{N}\right)},$$

where we minimize over  $-1 \leq \alpha \leq 1$ ,  $-1 \leq \beta \leq 1$  and  $0 \leq \alpha^2 + \beta^2 \leq 1$ .<sup>2</sup>

As the achievability scheme is largely the same as that in [12], we only give a sketch in Appendix V.

We now turn to lower bounds on the distortion-cost region. We first present without proof two lower bounds in the following two propositions. For their proofs cf. [12] or the proof of Theorem 10 below.

<sup>2</sup>In [12], the authors minimize only over  $0 \leq \alpha \leq 1$ ,  $0 \leq \beta \leq 1$  and  $0 \leq \alpha^2 + \beta^2 \leq 1$ , but it is easy to see that their proof carries over to the range stated in this Theorem.

**Proposition 8.** A lower bound for the problem of estimation with a helper who knows both the interference and the signal noncausally is given by

$$D(P)_{\min} \geq \frac{P_1}{\left(1 + \frac{P_1}{P_2}\right) \left(1 + \frac{P}{N}\right)}.$$

**Remark VI.1.** When  $P_2 \rightarrow \infty$ , we see that  $D(P)_{\min} \geq \frac{P_1}{1 + \frac{P}{N}}$ . This bound is achievable by noting that for  $\alpha = \beta = 0$ , we have  $D = \frac{P_1}{1 + \frac{P}{N}}$  in Theorem 9. Thus a separation scheme is optimal when  $P_2 \rightarrow \infty$ .

**Proposition 9.** A lower bound for the problem of estimation with a helper who knows both the interference and the signal noncausally is given by

$$D(P)_{\min} \geq \frac{P_1}{\left(1 + \frac{(\sqrt{P} + \sqrt{P_1})^2}{N}\right)}.$$

**Remark VI.2.** As  $P_2 \rightarrow 0$ , our setting reduces to that of state amplification [1]. From the results therein, the bound of Proposition 9 is optimal when  $P_2 \rightarrow 0$ .

We now present our lower bound.

**Theorem 10.** A lower bound for the problem of estimation with a helper that knows both the interference and the signal noncausally is given by

$$D(P)_{\min} \geq \frac{\left(\frac{\alpha^2 P_1 P_2}{P_1 + \alpha^2 P_2}\right) N}{MSE(\alpha)}$$

for any  $\alpha \in \mathcal{R}$ ,  $\alpha \neq 0$ , where  $MSE(\alpha)$  is given by the optimum value of the following convex (quadratic) optimization problem:

$$\max_{|\rho_{XS_1}| \leq \sqrt{PP_1}, |\rho_{XS_2}| \leq \sqrt{PP_2}} P + (1 - \alpha)^2 P_2 + 2(1 - \alpha)\rho_{XS_2} + N - \frac{((1 - \alpha)\alpha P_2 + \alpha\rho_{XS_2} + \rho_{XS_1})^2}{P_1 + \alpha^2 P_2}.$$

It can be shown that setting  $\alpha = 1$  and  $\alpha \rightarrow \infty$  recovers the bounds in Propositions 8 and 9, respectively. The cases of  $\alpha = 1$  and  $\alpha = \infty$  correspond to supplying  $S_1 + S_2$  and  $S_2$ , respectively to the decoder and then lower bounding the distortion.

Note that while finding the optimum value of  $\alpha$  may be difficult, Theorem 10 gives a lower bound for every  $\alpha$ . We note also that while computation of the lower bound requires solving an optimization problem for each  $\alpha$ , unlike the lower bounds in Propositions 8 and 9, the optimization problem is quadratic and can be efficiently solved [19], [20].

*Proof:* The idea in the proof of Theorem 10 lies in giving side information  $S_1 + \alpha S_2$  to the decoder instead of just  $S_1 + S_2$  or  $S_2$  as in Propositions 8 and 9 respectively, and then a more careful bounding of the terms appearing in the distortion calculation using Linear minimum mean square error estimation and convex optimization.

From the data processing inequality,

$$\begin{aligned} I(S_1^n; \hat{S}_1^n | S_1^n + \alpha S_2^n) &\leq I(S_1^n; Y^n | S_1^n + \alpha S_2^n) \\ &= h(Y^n | S_1^n + \alpha S_2^n) - h(Z^n) \\ &\leq \sum_{i=1}^n h(Y_i | S_{1i} + \alpha S_{2i}) - \frac{n}{2} \log 2\pi e N \\ &\stackrel{(a)}{\leq} n h(Y | S_1 + \alpha S_2, Q) - \frac{n}{2} \log 2\pi e N \\ &\leq n h(Y | S_1 + \alpha S_2) - \frac{n}{2} \log 2\pi e N. \end{aligned}$$



In (a), we defined  $Q \sim \text{Unif}[1 : n]$  independent of all other random variables and  $Y_Q = Y$ ,  $S_{1Q} = S_1$ ,  $S_{2Q} = S_2$  and  $\hat{S}_{1Q} = \hat{S}_1$ . On the other hand, we have

$$\begin{aligned}
I(S_1^n; \hat{S}_1^n | S_1^n + \alpha S_2^n) &= \sum_{i=1}^n h(S_{1i} | S_{1i} + \alpha S_{2i}) - h(S_1^n | \hat{S}_1^n, S_1^n + \alpha S_2^n) \\
&\geq \sum_{i=1}^n h(S_{1i} | S_{1i} + \alpha S_{2i}) - \sum_{i=1}^n h(S_{1i} | \hat{S}_{1i}) \\
&\geq \sum_{i=1}^n h(S_{1i} | S_{1i} + \alpha S_{2i}) - \sum_{i=1}^n h(S_{1i} - \hat{S}_{1i}) \\
&\stackrel{(a)}{\geq} nh(S_1 | S_1 + \alpha S_2) - \frac{n}{2} \log 2\pi e D(P)_{\min} \\
&= \frac{n}{2} \log \left( 2\pi e \frac{\alpha^2 P_1 P_2}{P_1 + \alpha^2 P_2} \right) - \frac{n}{2} \log 2\pi e D(P)_{\min},
\end{aligned}$$

where (a) follows from concavity of differential entropy and the property that a Gaussian distribution maximizes the differential entropy for a given second moment. Therefore,

$$\begin{aligned}
\frac{1}{2} \log \left( 2\pi e \frac{\alpha^2 P_1 P_2}{P_1 + \alpha^2 P_2} \right) - \frac{1}{2} \log 2\pi e D(P)_{\min} &\leq h(Y | S_1 + \alpha S_2) - \frac{1}{2} \log 2\pi e N \\
&= h(X + (1 - \alpha)S_2 + Z | S_1 + \alpha S_2) - \frac{1}{2} \log 2\pi e N \\
&\leq h(X + (1 - \alpha)S_2 + Z - k(S_1 + \alpha S_2)) - \frac{1}{2} \log 2\pi e N,
\end{aligned}$$

where  $k$  is defined as

$$k := \frac{(1 - \alpha)\alpha P_2 + \alpha \rho_{XS_2} + \rho_{XS_1}}{P_1 + \alpha^2 P_2},$$

with  $E X S_1 := \rho_{XS_1}$  and  $E X S_2 := \rho_{XS_2}$ . From Cauchy-Schwartz inequality and the power constraint on  $X$ ,  $|\rho_{XS_1}| \leq \sqrt{P P_{S_1}}$  and  $|\rho_{XS_2}| \leq \sqrt{P P_{S_2}}$ .

Continuing with our bound, we have

$$h(X + (1 - \alpha)S_2 + Z - k(S_1 + \alpha S_2)) \leq \frac{1}{2} \log(2\pi e (E(X + (1 - \alpha)S_2 + Z - k(S_1 + \alpha S_2))^2)).$$

In turn, we have

$$\begin{aligned}
E(X + (1 - \alpha)S_2 + Z - k(S_1 + \alpha S_2))^2 &= P + (1 - \alpha)^2 P_2 + 2(1 - \alpha)\rho_{XS_2} \\
&\quad + N - \frac{((1 - \alpha)\alpha P_2 + \alpha \rho_{XS_2} + \rho_{XS_1})^2}{P_1 + \alpha^2 P_2} \\
&:= MSE(\alpha, \rho_{XS_1}, \rho_{XS_2}).
\end{aligned}$$

Note now that for  $\alpha$  fixed,  $MSE(\alpha, \rho_{XS_1}, \rho_{XS_2})$  is a concave (quadratic) function of  $\rho_{XS_1}$  and  $\rho_{XS_2}$ , and the constraints  $|\rho_{XS_1}| \leq \sqrt{P P_{S_1}}$  and  $|\rho_{XS_2}| \leq \sqrt{P P_{S_2}}$  are linear constraints. Hence, we can find the maximum value using convex optimization. Letting  $\rho_{XS_1}^*$  and  $\rho_{XS_2}^*$  denote the optimal solutions to the optimization problem, we arrive at the lower bound for the achievable distortion:

$$D(P)_{\min} \geq \frac{(\frac{\alpha^2 P_1 P_2}{P_1 + \alpha^2 P_2})N}{MSE(\alpha, \rho_{XS_1}^*, \rho_{XS_2}^*)}.$$

■

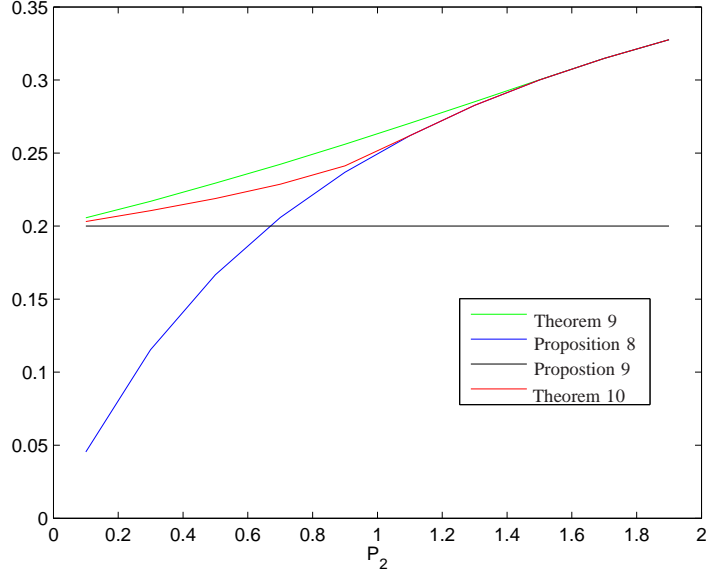


Fig. 11: Comparison of bounds for estimation with a helper that knows both the interference and the source. This figure gives a plot of the various bounds on distortion for different values of  $P_2$

### Comparison of bounds

As we mentioned earlier, Theorem 10 includes the bounds in Proposition 8 and 9. It can also be larger, as we now show in an example.

Let  $P_1 = 1$ ,  $N = 1$  and  $P = 1$ . We vary  $P_2$  and compare the bounds obtained with different values of  $P_2$ . The plots comparing the various upper and lower bounds are given in Figure 11. As can be seen from Figure 11, the lower bound given by Theorem 10 can be strictly better than that given by previous lower bounds. As we noted in the proof of Theorem 10, the improvement comes from two aspects: giving  $S_1 + \alpha S_2$  to the decoder and a more careful bounding via Linear Minimum Mean Square Error Estimation and Convex Optimization. The reader may ask whether it is necessary to use  $S_1 + \alpha S_2$  instead of just setting  $\alpha = 1$  or  $\alpha \rightarrow \infty$  and calculate the bounds more carefully using Linear Estimation and Convex Optimization. In our simulation, we noted that for some values of  $P_2$ , moderate values of  $\alpha$ , such as  $\alpha = 2, 3$  give better bounds than  $\alpha = 1$  or  $\alpha = 20$ . This shows that using  $S_1 + \alpha S_2$  does lead to better bounds than using  $S_1 + S_2$  or  $S_2$  alone.

### B. Constant gap results

In our simulations, we note that the upper bound and lower bounds appear to be quite close. This suggests that constant multiplicative gap results on the distortion may be possible, under some conditions on the input, source and interference powers. This is indeed the case as stated in our next result that when the interference power is larger than a threshold (that depends on the system parameters), the lower and upper bounds are within a constant multiplicative gap.

**Theorem 11.** *If*

$$\sqrt{P_2} \geq \frac{\sqrt{\gamma^2 P + \gamma \sqrt{P}(2 + \gamma \sqrt{P})(P(1 - \gamma^2) + N) - \gamma \sqrt{P}}}{\gamma \sqrt{P}(2 + \gamma \sqrt{P})}, \quad (12)$$

with  $\gamma = \sqrt{\frac{\epsilon(P+N)}{2P}}$ ,  $0 \leq \epsilon \leq \frac{P}{P+N}$ , then the multiplicative gap between the upper bound in Theorem 9,  $D_{\text{achievable}}$ ,

and the lower bound in Proposition 8,  $D_{\text{lb}}$ , is at most  $1/(1 - \epsilon)$ . That is,

$$\frac{D_{\text{achievable}}}{D_{\text{lb}}} \leq \frac{1}{1 - \epsilon}.$$

*Proof:* We begin the proof by evaluating the distortion achieved by Theorem 9 for  $\alpha = -\beta = \sqrt{\frac{\epsilon(P+N)}{2P}}$ . We have

$$1 + \frac{P(1 - \alpha^2 - \beta^2)}{N} = 1 + \frac{P}{N} \left(1 - \frac{\epsilon(P+N)}{P}\right) = (1 - \epsilon) \left(1 + \frac{P}{N}\right). \quad (13)$$

Now from the condition on  $P_2$  stated in the Theorem (see (12)), it follows that,

$$\begin{aligned} & P_2 \alpha \sqrt{P}(2 + \alpha \sqrt{P}) + \sqrt{P_2} 2 \alpha \sqrt{P} - P(1 - \alpha^2) - N \geq 0 \\ \Rightarrow & P_2 \alpha \sqrt{P}(2 + \alpha \sqrt{P}) + \sqrt{P_2}(2 + \alpha \sqrt{P}) \alpha \sqrt{P} - \alpha^2 P - P(1 - 2\alpha^2) - N \geq 0 \\ \Rightarrow & P_2 \left[ \alpha \sqrt{P}(2 + \alpha \sqrt{P}) - \frac{\alpha^2 P}{\sqrt{P_2}} + \frac{(2 + \alpha \sqrt{P}) \alpha \sqrt{P}}{\sqrt{P_2}} - \frac{\alpha^2 P}{P_2} \right] \geq P(1 - 2\alpha^2) + N \\ \Rightarrow & P_2 \left[ \left( \alpha \sqrt{P} + \frac{\alpha \sqrt{P}}{\sqrt{P_2}} \right) \left( 2 + \alpha \sqrt{P} - \frac{\alpha \sqrt{P}}{\sqrt{P_2}} \right) \right] \geq P(1 - 2\alpha^2) + N \\ \Rightarrow & P_2 \left[ \left( 1 + \alpha \sqrt{P} \right)^2 - \left( 1 - \alpha \sqrt{\frac{P}{P_2}} \right)^2 \right] \geq P(1 - 2\alpha^2) + N \\ \Rightarrow & \frac{(\alpha \sqrt{P} + 1)^2}{(-\alpha \sqrt{\frac{P}{P_2}} + 1)^2 P_2 + P(1 - 2\alpha^2) + N} \geq \frac{1}{P_2}. \end{aligned}$$

Therefore we have,

$$\left( 1 + \frac{(\alpha \sqrt{P} + 1)^2 P_1}{(\beta \sqrt{\frac{P}{P_2}} + 1)^2 P_2 + N + P(1 - \alpha^2 - \beta^2)} \right) \left( 1 + \frac{P(1 - \alpha^2 - \beta^2)}{N} \right) \geq \left( 1 + \frac{P_1}{P_2} \right) \left( 1 + \frac{P}{N} \right) (1 - \epsilon),$$

which implies

$$\frac{D_{\text{achievable}}}{D_{\text{lb}}} \leq \frac{1}{1 - \epsilon}.$$

## VII. CONCLUSION

In this paper, we defined and analyze the problem of estimation with a helper that knows the interference. In the discrete memoryless case when the interfering signal,  $S_2$ , is known causally at the encoder, we characterized the distortion-cost region. When  $S_2$  is known noncausally, we proposed an achievability scheme based on hybrid coding. In the binary estimation with a helper problem, we also proposed two lower bounds. Using the upper and lower bounds, we characterized the distortion-cost region when the problem parameters  $C$ ,  $p_1$  and  $p_2$  satisfy one of several nontrivial conditions.

In the Gaussian case, we derived a lower bound based on a recent result by Verdú between divergence and mismatched estimation. We showed through numerical simulations that this lower bound can be strictly better than previous lower bound derived in [3]. Similar to the binary case, we also characterized the distortion-cost region when the problem parameters  $P$ ,  $P_1$  and  $P_2$  satisfy one of several conditions.

We also extended our analysis in the Gaussian case to consider the case when the helper knows both  $S_1$  and  $S_2$  noncausally. In this case, we derived a lower bound that contains previous lower bounds proposed in [12] and can be strictly better. We also obtained constant multiplicative gap results for this setting.

In deriving our lower bound for the Gaussian case when only the interfering signal is known at the helper, we used a relationship between mismatched estimation and divergence. In the discrete case, a relationship between divergence and Hamming distortion exists too. One such relationship is Marton's inequality [21, Lemma 6.3]. An

interesting open question is whether one can use such relationships to derive a lower bound for the binary case that is strictly better than the bounds we proposed.

#### ACKNOWLEDGMENT

We thank Mr Gowtham Kumar of Stanford University for discussions that motivated this work, and Professor Sriram Vishwanath of The University of Texas at Austin for helpful discussions during the course of this work.

#### REFERENCES

- [1] Y. H. Kim, A. Sutivong, and T. Cover, "State amplification," *IEEE Trans. Inf. Theory*, vol. 54, no. 5, pp. 1850--1859, May 2008.
- [2] C. Choudhuri, Y.-H. Kim, and U. Mitra, "Causal state amplification," in *Proc. IEEE International Symposium on Information Theory*, St Petersburg, Russia, August 2011.
- [3] P. Grover and A. Sahai, "Witsenhausen's counterexample as assisted interference suppression," submitted to *International Journal of Systems, Control and Communications*. Available online at <http://www.eecs.berkeley.edu/sahai/Papers/VectorWitsenhausenJournal.pdf>.
- [4] H. Witsenhausen, "A counterexample in stochastic optimum control," *SIAM J. Control*, vol. 6, no. 1, pp. 131--147, February 1968.
- [5] L. Zhao, Y. K. Chia, and T. Weissman, "Compression with actions," in *Allerton conference on communications, control and computing*, Monticello, Illinois, September 2011.
- [6] P. Grover, A. Wagner, and A. Sahai, "Information embedding and the triple role of control," submitted to *Transactions on Information Theory*. Available online at <http://www.eecs.berkeley.edu/pulkit/papers/InfoEmbeddingITTransDec8.pdf>.
- [7] S. Verdú, "Mismatched estimation and relative entropy," *IEEE Trans. Inf. Theory*, vol. 56, no. 8, pp. 3712--3720, August 2010.
- [8] D. Guo, S. Shamai, and S. Verdú, "Mutual information and minimum mean-square error in Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1261--1282, April 2005.
- [9] S. Shamai, "From constrained signaling to network interference alignment via an information-estimation perspective," *IEEE International Symposium for Information Theory Shannon Lecture*, Saint Petersburg, Russia, 2011.
- [10] D. Guo, Y. Wu, S. Shamai, and S. Verdú, "Estimation in Gaussian noise: Properties of the minimum mean-square error," *IEEE Trans. Inf. Theory*, vol. 57, no. 4, pp. 2371--2385, April 2011.
- [11] A. El Gamal and Y. H. Kim, *Network Information Theory*, 1st ed. Cambridge University Press, 2011.
- [12] Y.-C. Huang and K. Narayanan, "Joint source-channel coding with correlated interference," in *Proc. IEEE International Symposium on Information Theory*, St. Petersburg, Russia, August 2011.
- [13] I. Csiszár and J. Körner, "Broadcast channels with confidential messages," *IEEE Trans. Inf. Theory*, vol. IT-24, pp. 339--348, May 1978.
- [14] A. H. Kaspi, "Two-way source coding with a fidelity criterion," *IEEE Trans. Inf. Theory*, vol. 31, no. 6, pp. 735--740, 1985.
- [15] S. H. Lim, P. Minero, and Y.-H. Kim, "Lossy communication of correlated sources over multiple access channels," in *Proceedings of the 48th Annual Allerton Conference on Communication*, Monticello, Illinois, September 2010, invited paper.
- [16] T. Cover and J. Thomas, *Elements of Information Theory 2nd edition*. Wiley Interscience, July 2006.
- [17] G. Keshet, Y. Steinberg, and N. Merhav, "Channel coding in the presence of side information," *IEEE Trans. Inf. Theory*, vol. 4, no. 6, June 2007.
- [18] M. Costa, "Writing on dirty paper," *IEEE Trans. Inf. Theory*, vol. 29, no. 3, pp. 439--441, 1983.
- [19] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.
- [20] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 1.21," Apr. 2011.
- [21] R. M. Gray, *Entropy and Information Theory*. Springer, 2010.
- [22] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inf. Theor.*, vol. 22, no. 1, pp. 1--10, 1976.

#### APPENDIX I

##### SKETCH OF ACHIEVABILITY FOR THEOREM 1

We use block Markov coding over  $B$  blocks. The scheme in each block is basically a separation scheme, where we use the random variable  $U$  for transmission of a message from the previous block. The message itself is a Wyner-Ziv description [22] of  $S_2^n$  from the previous block. More concretely, in each block  $j \in [1 : B]$ , the *transmission* codebook is generated as follows: Generate  $2^{n(I(U;Y)-\epsilon)}$   $U^n(l)$  sequences according to  $\prod_{i=1}^n p(u_i)$ . The *compression* codebook is generated by the following two step procedure: Generate  $2^{n(I(V;S_2,U)+\epsilon)}$   $V^n$  sequences according to  $\prod_{i=1}^n p(v_i)$ . Partition the set of  $V^n$  sequences into  $2^{n(I(V;S_2|U,Y)+2\epsilon)}$  bins,  $\mathcal{B}(M_j)$ .

For encoding, at the end of block  $j$ , assume that the codeword  $U^n(m_j)$  was sent. The encoder then finds a  $V^n(j)$  sequence that is jointly typical with  $(U^n(m_j), S_2^n(j))$ . If there is more than one such sequence, it picks from one uniformly at random from the set of jointly typical sequences. This operation succeeds with high probability as  $n \rightarrow \infty$  since there are  $2^{n(I(V;S_2,U)+\epsilon)}$   $V^n(j)$  sequences. The encoder then finds the bin index  $M_{j+1}$  such that  $V^n \in \mathcal{B}(M_{j+1})$ . It then sends out the index  $M_{j+1}$  in block  $j+1$  by selecting  $U^n(j+1)$  and sending out the  $X^n$  sequence encoded as  $x_i = f(u_i(M_{j+1}), s_{2i}(j+1))$ . For the first block, the encoder sends an arbitrary message. This encoding operation requires the condition that

$$I(U; Y) - \epsilon > I(V; S_2|U, Y) + 2\epsilon.$$

For decoding, at the end of block  $j + 1$ , the decoder first decodes the bin index  $M_{j+1}$ . From standard arguments (see for e.g. [11, Chapter 7]), this decoding operation succeeds with high probability provided

$$I(U; Y) - \epsilon > I(V; S_2|U, Y) + 2\epsilon.$$

Once the decoder recovers the bin index  $M_{j+1}$ , it then recovers the true  $V^n(j)$  codeword by looking for  $v^n(j) \in \mathcal{B}(M_{j+1})$  such that  $(u^n(m_j), y^n(j), v^n(j)) \in \mathcal{T}_\epsilon^{(n)}$ . It then reconstructs  $S_1^n(j)$  as  $\hat{s}_1(u_i(m_j), y_i(j), v_i(j))$  for  $i \in [1 : n]$ . From the rates given and standard arguments (see [11, Chapter 3 and Chapter 11]), the expected distortion for  $S_1^n(j)$  in block  $j$  is less than or equal to  $E d(S_1, \hat{S}_1(U, V, Y)) + (\epsilon)$ , where  $(\epsilon) \rightarrow 0$  as  $\epsilon \rightarrow 0$ . This decoding and reconstruction procedure applies for the first  $B - 1$  blocks and for the  $B$ th block, we simply reconstruct  $S_1^n(B)$  according to an arbitrary symbol  $\hat{s}_1 \in \hat{S}_1$ , incurring a distortion that is bounded by  $D_{\max}$ , where  $D_{\max} := \max_{\hat{s}_1} E d(S_1, \hat{s}_1)$ . The per symbol distortion over  $B$ -blocks is then upper bounded by  $D + (\epsilon)$  where  $(\epsilon) \rightarrow 0$  as  $\epsilon \rightarrow 0$ .

We now note that the above achievability scheme takes care of the case when  $I(U; Y) > I(V; S_2|U, Y)$ . The boundary case of  $I(U; Y) = I(V; S_2|U, Y)$  can be handled as follow. Assume first that  $I(U; Y) > 0$ . Define  $U' = (U, Q)$ ,  $Q \in \{1, 2\}$  independent of other random variables,  $V' = V$  when  $Q = 1$  and  $V' = \emptyset$  when  $Q = 2$ .  $X = f(U, S_2)$  regardless of  $Q$  and  $\hat{s}_1(U', V', Y') = \hat{s}_1(U, V, Y)$  if  $Q = 1$  and  $\hat{s}_1^*$  if  $Q = 2$ , where  $\hat{s}_1^*$  is an arbitrary symbol belonging to  $\hat{S}_1$ . Let  $P(Q = 1) = p_1$ . We have

$$\begin{aligned} I(U'; Y') &\geq I(U; Y), \\ I(V'; S_2|U', Y') &= p_1 I(V; S_2|U, Y), \\ E d(S_1, \hat{S}_1(U', V', Y')) &\leq p_1 E d(S_1, \hat{S}_1(U, V, Y)) + (1 - p_1) D_{\max}. \end{aligned}$$

With this choice of random variables,  $I(U'; Y') > I(V'; S_2|U', Y')$  whenever  $p_1 < 1$  and we can then apply the achievability scheme we discussed, at the expense of larger distortion. By choosing  $p_1(n) = 1 - \epsilon_n$ , where  $\epsilon_n \rightarrow 0$  as  $n \rightarrow \infty$ , we can apply our achievability scheme for blocklength  $n$  sufficiently large, with the resulting expected distortion converging to  $E d(S_1, \hat{S}_1(U, V, Y))$  as  $n \rightarrow \infty$ .

For the case of  $I(U; Y) = I(V; S_2|U, Y) = 0$ , it can be shown that in this case, the decoder can perform the reconstruction based only on  $\hat{s}_1(Y_i, U_i)$  for  $i \in [1 : n]$ . Achievability in this case requires no block Markov coding. We only need to generate one transmission codeword  $U^n$  and transmit  $X^n$  according to  $x_i = f(u_i, s_2)$ . The decoder reconstructs  $S_1^n$  as  $\hat{s}_1(u_i, y_i)$  for  $i \in [1 : n]$ .

## APPENDIX II PROOF OF CLAIM 1

The causal region in Theorem 1 is given by

$$\begin{aligned} \min \quad & E d(S_1, \hat{S}_1(U, V, Y)) \\ \text{subject to} \quad & \\ & I(U; Y) \geq I(V; S_2|U, Y) \\ & E X \leq C \end{aligned}$$

for some  $p(u)p(v|u, s_2)$  and function  $x(u, s_2)$ . We prove that  $D(0.11)_{\min\text{-causal}} > 0$  by contradiction. Suppose that there exists  $U, V$  satisfying the constraints such that  $E d(S_1, \hat{S}_1(U, V, Y)) = 0$ . This implies in particular that  $H(S_1|U, V, Y) = 0$ . Hence,

$$\begin{aligned} I(V; S_2|U, Y) &= I(V, S_1; S_2|U, Y) \\ &\geq I(S_1; S_2|U, Y) \\ &= H(S_1|U, Y) - H(S_1|S_2, U, Y) \\ &= H(S_1|U, Y) \\ &= H(S_1, Y|U) - H(Y|U) \\ &= H(S_1) + H(Y|U, S_1) - H(Y|U). \end{aligned}$$

The last step follows from  $U$  being independent of  $S_1$ . Since we require  $I(U; Y) \geq I(V; S_2|U, Y)$ , and we know that  $H(S_1) + H(Y|U, S_1) - H(Y|U) \leq I(V; S_2|U, Y)$ , a necessary condition for  $\mathbb{E}d(S_1, \hat{S}_1(U, V, Y)) = 0$  is

$$\begin{aligned} H(S_1) + H(Y|U, S_1) - H(Y|U) &\leq I(U; Y), \\ \Rightarrow H(S_1) + H(Y|U, S_1) &\leq H(Y). \end{aligned}$$

Define the subsets of  $\mathcal{U}$  as follows.  $\mathcal{U}_0 := \{u : x(u, s_2) = 0\}$ ;  $\mathcal{U}_1 := \{u : x(u, s_2) = 1\}$ ;  $\mathcal{U}_s := \{u : x(u, s_2) = s_2\}$ ; and  $\mathcal{U}_{\bar{s}} := \{u : x(u, s_2) = 1 \oplus s_2\}$ . Note the following.

- For  $u \in \mathcal{U}_0$ ,  $H(Y|U = u, S_1) = 1$  since  $S_2$  is independent of  $U, S_1$ .
- For  $u \in \mathcal{U}_1$ ,  $H(Y|U = u, S_1) = 1$  since  $S_2$  is independent of  $U, S_1$ .
- For  $u \in \mathcal{U}_s$ ,  $H(Y|U = u, S_1) = 0$  since  $S_2 \oplus X = 0$  and  $Y = S_2 \oplus X \oplus S_1$ .
- For  $u \in \mathcal{U}_{\bar{s}}$ ,  $H(Y|U = u, S_1) = 0$  since  $S_2 \oplus X = 1$  and  $Y = S_2 \oplus X \oplus S_1$ .

Further, define  $p_{u0} = \sum_{u \in \mathcal{U}_0} p(u)$ ;  $p_{u1} = \sum_{u \in \mathcal{U}_1} p(u)$ ;  $p_s = \sum_{u \in \mathcal{U}_s} p(u)$ ; and  $p_{\bar{s}} = \sum_{u \in \mathcal{U}_{\bar{s}}} p(u)$ . Then,

$$\begin{aligned} H(S_1) + H(Y|U, S_1) &= H_2(p_1) + p_{u0} + p_{u1} \\ &= H_2(p_1) + 1 - C_s, \end{aligned}$$

where  $C_s = p_s + p_{\bar{s}}$ .

The cost constraint can be expressed as

$$\begin{aligned} \mathbb{E}X &= p_1 + \frac{1}{2}(p_s + p_{\bar{s}}) \\ &= p_1 + \frac{1}{2}C_s \\ &\leq C, \end{aligned}$$

where  $C = 0.11$ . In particular, the cost constraint implies that  $C_s \leq 2C$ . Hence,

$$H(S_1) + H(Y|U, S_1) \geq 1 + H_2(q) - 2C.$$

Now, since  $p_1 = 0.1$  and  $C = 0.11$ , we see that

$$\begin{aligned} H(S_1) + H(Y|U, S_1) &> 1 \\ &\geq H(Y), \end{aligned}$$

which is a contradiction.

### APPENDIX III DERIVATION OF THEOREM 5

The derivation of Theorem 5 follows from choosing the auxiliary random variables in Theorem 2. Starting from Theorem 2, let

$$\begin{aligned} U &= X' + \gamma S_2, \\ X &= \alpha \sqrt{\frac{P}{P_2}} S_2 + X', \\ X' &\sim N(0, P'), \\ \mathbb{E}(S_2 X') &= \beta \sqrt{P_2 P'}, \end{aligned}$$

where  $P'$  is a quantity to be calculated, and  $\alpha$  and  $\beta$  are restricted to be between -1 to 1 to satisfy the power constraints. Observe that  $X$  is a function of  $U, S_2$  as required. For convenience, we will use the notation  $X|Y$  to denote Minimum Mean Square Error of  $X$  given  $Y$ . The reconstruction function is given by

$$\hat{S}_1 = \mathbb{E}(S_1|U, Y).$$

We now determine  $P'$  from other variables using  $\mathbb{E}X^2 = P$ .

$$\mathbb{E}X^2 = \alpha^2 P + P' + 2\alpha\beta\sqrt{PP'} = P.$$

Solving for  $P'$  gives

$$\sqrt{P'} = \frac{-2\alpha\beta\sqrt{P} + \sqrt{4\alpha^2\beta^2P + 4(1-\alpha^2)P}}{2}.$$

To satisfy the constraint in Theorem 2, we require

$$h(U|S_2) > h(U|Y).$$

Since  $U, S_2, Y$  are all Gaussian random variables, this condition reduces to

$$U|S_2 > U|Y.$$

Now,

$$\begin{aligned} U|S_2 &= X'|S_2 \\ &= (1-\beta^2)P'. \end{aligned}$$

As for  $U|Y$ , we have

$$U|Y = EU^2 - \frac{(E(UY))^2}{EY^2},$$

and

$$\begin{aligned} EU^2 &= P' + 2\gamma\beta\sqrt{P'P_2} + \gamma^2P_2, \\ E(UY) &= E((X' + \gamma S_2)(\alpha\sqrt{\frac{P}{P_2}}S_2 + X')) + \beta\sqrt{P'P_2} + \gamma P_2 \\ &= \alpha\beta\sqrt{PP'} + P' + \alpha\gamma\sqrt{PP_2} + \gamma\beta\sqrt{P'P_2} + \beta\sqrt{P'P_2} + \gamma P_2, \\ EY^2 &= EX^2 + ES_1^2 + ES_2^2 + 2E(S_2(\alpha\sqrt{\frac{P}{P_2}}S_2 + X')) \\ &= P + 1 + P_2 + 2\alpha\sqrt{PP_2} + 2\beta\sqrt{P'P_2}. \end{aligned}$$

The expected distortion is then given by is  $S_1|(Y, U)$ , which is

$$S_1|(U, Y) = 1 - [E(US_1) E(Y S_1)] \begin{bmatrix} EU^2 & E(UY) \\ E(UY) & EY^2 \end{bmatrix}^{-1} \begin{bmatrix} E(US_1) \\ E(Y S_1) \end{bmatrix}.$$

We note now that  $E(US_1) = 0$  and  $E(S_1Y) = 1$ . The lower bound therefore works out to

$$S_1|(U, Y) = 1 - \frac{EU^2}{EY^2 EU^2 - (EUY)^2}.$$

#### APPENDIX IV PROOF OF THEOREM 8

As Theorem 8 is a generalization of Proposition 7, the proof of this Theorem also follows closely that of Proposition 7. As such, we will only mention areas where there are differences from the proof in Proposition 7 and refer readers to Proposition 7 for the rest of the proof.

As we mentioned before, we generalize the bound by not assuming that  $X^n \equiv 0$ . Instead, let us assume that under the mismatched distribution  $Q$ ,  $X$  is distributed i.i.d according to  $X = cS_2 + Z$ , where  $Z \sim N(0, rP)$  independent of  $S_2$  and  $r \geq 0$ . Under this assumption,  $MSE_Q(\gamma)$  and  $D(P_{Y^n|S_2}^{(\gamma)} || Q_{Y^n|S_2}^{(\gamma)})$  used in the proof of Proposition 7 are now different. The bounds on  $MSE_Q(\gamma)$  and the divergence between the true distribution and the mismatched distribution are therefore different. We calculate them as follow.

Define  $\alpha$  as

$$\alpha := \frac{1}{1 + \gamma P_I},$$

where  $P_I = (1+c)^2P_2 + rP$ . Let  $E\|X^n\|_2^2 = nx^2$ , where  $x^2 \leq P$ . We now have, for  $MSE_Q(\gamma)$ ,

$$\begin{aligned} \frac{1}{n}MSE_Q(\gamma) &= \frac{1}{n}E\|S_1^n - \alpha(X^n + S_1^n + S_2^n)\|^2 \\ &= \frac{1}{n}E\|S_1^n - \alpha(S_1^n + S_2^n)\|^2 - \frac{2\alpha}{n}E\langle S_1^n - \alpha(S_1^n + S_2^n), X^n \rangle + \frac{\alpha^2}{n}E\|X^n\|^2 \\ &= \frac{(1-\alpha)^2}{\gamma} + \alpha^2P_2 + \frac{2\alpha^2}{n}E\langle S_2^n, X^n \rangle + \alpha^2x^2 \\ &= \frac{\gamma P_I^2}{(1+\gamma P_I)^2} + \frac{P_2}{(1+\gamma P_I)^2} + \frac{2\alpha^2}{n}E\langle S_2^n, X^n \rangle + \alpha^2x^2 \\ &= \frac{P_I}{1+\gamma P_I} - \frac{P_I}{(1+\gamma P_I)^2} + \frac{P_2}{(1+\gamma P_I)^2} + \frac{2}{n(1+\gamma P_I)^2}E\langle S_2^n, X^n \rangle + \frac{1}{(1+\gamma P_I)^2}x^2. \end{aligned}$$

It remains to calculate an upper bound on the divergence. As before,  $P_{Y^n|S_2^n}^{(\gamma)} \sim N(S_2^n + X^n, \frac{1}{\gamma}I)$ , but now,  $Q_{Y^n|S_2^n}^{(\gamma)} \sim N((1+c)S_2^n, (\frac{1}{\gamma} + rP)I)$ . The (conditional) divergence is now given by

$$\frac{2}{n}D(P_{Y^n|S_2^n}^{(\gamma)}\|Q_{Y^n|S_2^n}^{(\gamma)}) = \frac{1}{1+\gamma rP} - 1 - \log\left(\frac{1}{1+\gamma rP}\right) + \frac{\gamma}{n(1+\gamma rP)}\|X^n - cS_2^n\|_2^2.$$

Combining the divergence bound after taking expectation over  $S_2^n$  with the  $MSE_Q$  bound after integration gives (see (10) in the proof of Proposition 7)

$$\begin{aligned} \frac{\gamma_1 - \gamma_0}{n}MMSE(\gamma_0) &\geq \log\left(\frac{1+\gamma_1 P_I}{1+\gamma_0 P_I}\right) + \frac{1}{(1+\gamma_1 P_I)} - \frac{1}{(1+\gamma_0 P_I)} - \frac{P_2}{P_I(1+\gamma_1 P_I)} \\ &\quad + \frac{P_2}{P_I(1+\gamma_0 P_I)} - \frac{c^2\gamma_1}{1+\gamma_1 rP}P_2 - \frac{1}{1+\gamma_1 rP} + 1 + \log\left(\frac{1}{1+\gamma_1 rP}\right) \\ &\quad + \left(\frac{1}{P_I(1+\gamma_0 P_I)} - \frac{1}{P_I(1+\gamma_1 P_I)} - \frac{\gamma_1}{1+\gamma_1 rP}\right)x^2 \\ &\quad + 2\left(\frac{1}{P_I(1+\gamma_0 P_I)} - \frac{1}{P_I(1+\gamma_1 P_I)} + \frac{c\gamma_1}{1+\gamma_1 rP}\right)\frac{E\langle S_2^n, X^n \rangle}{n} \\ &\geq \log\left(\frac{1+\gamma_1 P_I}{1+\gamma_0 P_I}\right) + \frac{1}{(1+\gamma_1 P_I)} - \frac{1}{(1+\gamma_0 P_I)} - \frac{P_2}{P_I(1+\gamma_1 P_I)} \\ &\quad + \frac{P_2}{P_I(1+\gamma_0 P_I)} - \frac{c^2\gamma_1}{1+\gamma_1 rP}P_2 - \frac{1}{1+\gamma_1 rP} + 1 + \log\left(\frac{1}{1+\gamma_1 rP}\right) \\ &\quad + \left(\frac{1}{P_I(1+\gamma_0 P_I)} - \frac{1}{P_I(1+\gamma_1 P_I)} - \frac{\gamma_1}{1+\gamma_1 rP}\right)x^2 \\ &\quad - |2\left(\frac{1}{P_I(1+\gamma_0 P_I)} - \frac{1}{P_I(1+\gamma_1 P_I)} + \frac{c\gamma_1}{1+\gamma_1 rP}\right)|\sqrt{P_2}|x|. \end{aligned}$$

The final line follows from successive application of Cauchy-Schwartz on  $E\langle S_2^n, X^n \rangle$ . Minimizing the bound over  $|x| \leq \sqrt{P}$  then gives the generalized lower bound. Let  $a = \left(\frac{1}{P_I(1+\gamma_0 P_I)} - \frac{1}{P_I(1+\gamma_1 P_I)} - \frac{\gamma_1}{1+\gamma_1 rP}\right)$  and  $b = |2\left(\frac{1}{P_I(1+\gamma_0 P_I)} - \frac{1}{P_I(1+\gamma_1 P_I)} + \frac{c\gamma_1}{1+\gamma_1 rP}\right)|\sqrt{P_2}$ . We note that  $b \leq 0$  and let  $f(x) = ax^2 - b|x|$ .

We note that if  $a \leq 0$ ,  $f(x)$  is symmetric and decreasing in  $x$ . Therefore, we set  $x^* = \sqrt{P}$ . If  $a > 0$ , then  $x^* = b/(2a)$  if  $b/(2a) < \sqrt{P}$  and  $x^* = \sqrt{P}$  otherwise. The generalized lower bound is now given by

$$\begin{aligned} \frac{\gamma_1 - \gamma_0}{n}MMSE(\gamma_0) &\geq \log\left(\frac{1+\gamma_1 P_I}{1+\gamma_0 P_I}\right) + \frac{1}{(1+\gamma_1 P_I)} - \frac{1}{(1+\gamma_0 P_I)} - \frac{P_2}{P_I(1+\gamma_1 P_I)} \\ &\quad + \frac{P_2}{P_I(1+\gamma_0 P_I)} - \frac{c^2\gamma_1}{1+\gamma_1 rP}P_2 - \frac{1}{1+\gamma_1 rP} + 1 + \log\left(\frac{1}{1+\gamma_1 rP}\right) \\ &\quad + \left(\frac{1}{P_I(1+\gamma_0 P_I)} - \frac{1}{P_I(1+\gamma_1 P_I)} - \frac{\gamma_1}{1+\gamma_1 rP}\right)x^{*2} \\ &\quad - |2\left(\frac{1}{P_I(1+\gamma_0 P_I)} - \frac{1}{P_I(1+\gamma_1 P_I)} + \frac{c\gamma_1}{1+\gamma_1 rP}\right)|\sqrt{P_2}x^*, \end{aligned}$$



where we optimize over  $\gamma_1 \geq \gamma_0$ ,  $r \geq 0$  and  $c \in \mathcal{R}$ . Noting that  $MMSE(1)/n = D(P)_{\min}$  then completes the proof.

APPENDIX V  
SKETCH OF THEOREM 9

The achievability scheme in Theorem 9 closely resembles [12] and involves allocating a fraction of the power for transmitting a message (corresponding to a compressed version of the desired source  $S_1$ ) using dirty paper coding and using the remaining power for uncoded transmission of a linear combination of  $S_1$  and  $S_2$ . The compressed index is generated based on Wyner-Ziv coding and then transmitted reliably over channel using dirty paper coding as in [12]. The bin indices in Wyner-Ziv coding are transmitted at a rate equal to the capacity of the dirty paper channel. Note that the interference in this channel also includes the signal due to uncoded transmission created at the encoder. The compressed index is decoded at the receiver using the receiver side information  $Y$  and both the decoded codeword and  $Y$  are used to estimate the source  $S_1$ . Uncoded transmission helps in improving the signal to noise ratio of the desired signal  $S_1$  in  $Y$ .

Let

$$\begin{aligned} U &= X' + \left( \alpha \sqrt{\frac{P}{P_1}} + 1 \right) S_1 + \left( \beta \sqrt{\frac{P}{P_2}} + 1 \right) S_2 \\ X &= X' + \alpha \sqrt{\frac{P}{P_1}} S_1 + \beta \sqrt{\frac{P}{P_2}} S_2 \\ X' &\sim \mathcal{N}(0, P(1 - \alpha^2 - \beta^2)), \\ Y &= X + S_1 + S_2 + Z, \end{aligned}$$

where  $X'$  is independent of  $S_1$  and  $S_2$  and corresponds to the coded part of the signal. Auxiliary  $U$  is used to cancel the total interference to  $X'$  as in dirty paper coding. The total interference is equal to  $\left( \alpha \sqrt{\frac{P}{P_1}} + 1 \right) S_1 + \left( \beta \sqrt{\frac{P}{P_2}} + 1 \right) S_2$ . As a result, a clean channel (without interference) is created between  $X'$  and  $Y$ , which can be used to transmit the description of  $S_1$  at a Wyner-Ziv rate equal to  $\frac{1}{2} \log \left( 1 + \frac{P(1 - \alpha^2 - \beta^2)}{N} \right)$ . The received signal  $Y$  can also be seen as a noisy version of the desired signal  $S_1$ , and is used along with the message transmitted to reconstruct  $S_1$ .

Therefore, the resulting distortion in  $S_1$  is given by

$$D = \frac{P_1}{\left( 1 + \frac{\left( \alpha \sqrt{\frac{P}{P_1}} + 1 \right)^2 P_1}{\left( \beta \sqrt{\frac{P}{P_2}} + 1 \right)^2 P_2 + P(1 - \alpha^2 - \beta^2) + N} \right) \left( 1 + \frac{P(1 - \alpha^2 - \beta^2)}{N} \right)}.$$