

# Hypothesis Testing for Adversarial Channels: Chernoff-Stein Exponents

Eeshan Modak<sup>1</sup>, Neha Sangwan<sup>1</sup>, Mayank Bakshi<sup>2</sup>, Bikash Kumar Dey<sup>3</sup>, and Vinod M. Prabhakaran<sup>1</sup>

<sup>1</sup>Tata Institute of Fundamental Research, Mumbai, India

<sup>2</sup>Arizona State University, Tempe, AZ, USA

<sup>3</sup>Indian Institute of Technology Bombay, Mumbai, India

**Abstract**—We study the Chernoff-Stein exponent of the following binary hypothesis testing problem: Associated with each hypothesis is a set of channels. A transmitter, without knowledge of the hypothesis, chooses the vector of inputs to the channel. Given the hypothesis, from the set associated with the hypothesis, an adversary chooses channels, one for each element of the input vector. Based on the channel outputs, a detector attempts to distinguish between the hypotheses. We study the Chernoff-Stein exponent for the cases where the transmitter (i) is deterministic, (ii) may privately randomize, and (iii) shares randomness with the detector that is unavailable to the adversary. It turns out that while a memoryless transmission strategy is optimal under shared randomness, it may be strictly suboptimal when the transmitter only has private randomness.

## 1. INTRODUCTION

In binary hypothesis testing the goal is to distinguish between two distributions (sources) [1], [2]. When  $n$  independent and identically distributed (i.i.d.) observations from the source are available, the Chernoff-Stein lemma [3, Theorem 11.8.3] states that for a fixed false alarm (type-1 error) probability, the optimal missed detection (type-2 error) probability decays exponentially in  $n$  with the exponent given by the relative entropy between the distributions.

A variation on this problem is where each observation is from an arbitrarily varying source [4]. There is a set of distributions associated with each hypothesis. Given a hypothesis, the observations are independent, but each observation could be arbitrarily distributed according to any one of the distributions belonging to the set of distributions corresponding to the hypothesis. We may view the choice of distribution as being made by an adversary who is aware of the detection scheme used. Fangwei and Shiyi [5] studied this problem where the adversary's choice may be stochastic. Recently, Brandão, Harrow, Lee, and Peres [6] considered the case with an adaptive adversary who has feedback of the past observations and may use this to choose the distribution of the next observation.

In another variation on the binary hypothesis testing problem, instead of distinguishing between sources, the objective

is to distinguish between two channels with the same input and output alphabets [7], [8]. Here, a transmitter, which is unaware of the hypothesis, may choose the inputs to the channels. It can be shown that the optimal Chernoff-Stein error exponent may be attained using a deterministic transmission strategy which sends the input letter for which the relative entropy between the channel output distributions under the two hypotheses is maximized [7]. Hayashi [8] studied the adaptive case where the transmitter has feedback of the channel output when the block length is fixed and showed that feedback does not improve the optimal error exponent. Polyanskiy and Verdú [9] considered the same problem with variable-length transmissions and showed that feedback may improve the error exponent in general.

In this work we study the Chernoff-Stein exponent of the binary hypothesis testing problem for arbitrarily varying channels [10]. Associated with each hypothesis is a set of channels. All channels have the same input and output alphabets. The transmitter, without knowledge of the hypothesis, chooses the vector of inputs to the channel. Given the hypothesis, the adversary chooses a vector of channels where each element belongs to the set of channels associated with the hypothesis. The adversary is aware of the strategy of the transmitter and detector, but not necessarily the choice of channel inputs. The detector observes the outputs resulting from applying the inputs chosen by the transmitter element-wise independently to the channels selected by the adversary. We consider three different settings depending on the nature of randomness unknown to the adversary which is available to the transmitter and detector<sup>1</sup>: (i) deterministic schemes (Section 4), (ii) randomness shared between transmitter and detector (Section 3), and (iii) private randomness at the transmitter (Section 5). We also comment on the role of adaptivity both of the transmitter (under a fixed block length) and of the adversary (Section 6).

When the channels are not arbitrarily varying, randomization (and adaptivity in the fixed length case) do not change the optimal Chernoff-Stein exponent which is achieved by the deterministic transmitter strategy of repeating the input symbol for which the channel output distributions under the two hypotheses have the largest relative entropy [7], [8].

<sup>1</sup>We allow the adversary to randomize in all cases. The optimal exponent is unaffected by the availability of common randomness known also to the adversary, nor by additional private randomness at the detector.

E. Modak, N. Sangwan and V. M. Prabhakaran were supported by DAE under project no. RTI4001. N. Sangwan was additionally supported by the TCS Foundation through the TCS Research Scholar Program. The work of M. Bakshi was supported by the National Science Foundation under Grant No. CCF-2107526. The work of B. K. Dey was supported by Bharti Centre for Communication in IIT Bombay. V. M. Prabhakaran was additionally supported by SERB through project MTR/2020/000308.

	Chernoff-Stein exponent	Condition for the exponent to be non-zero
Shared randomness	$\sup_{P_X} \min_{W \in \text{conv}(\mathcal{W}), \overline{W} \in \text{conv}(\overline{\mathcal{W}})} D(W \ \overline{W}) P_X$	$\text{conv}(\mathcal{W}) \cap \text{conv}(\overline{\mathcal{W}}) = \emptyset$
Deterministic transmitter	$\max_x \min_{W_x \in \text{conv}(\mathcal{W}_x), \overline{W}_x \in \text{conv}(\overline{\mathcal{W}}_x)} D(W_x \ \overline{W}_x)$	$\text{conv}(\mathcal{W}_x) \cap \text{conv}(\overline{\mathcal{W}}_x) = \emptyset$ for some $x$
Private randomness	Open (see Theorem 6)	$\text{conv}(\mathcal{W}) \cap \text{conv}(\overline{\mathcal{W}}) = \emptyset$ and $(\mathcal{W}, \overline{\mathcal{W}})$ is not trans-symmetrizable

With arbitrarily varying channels, we see that randomization improves the exponent in general (Remark 1 and Example 2). This is analogous to the usefulness of randomization in communication over arbitrarily varying channels [11]. We also demonstrate that the optimal exponents under the three different settings are different in general. Our results also show the following interesting phenomenon: When the transmitter has private randomness which is unknown to the adversary, but shares no randomness with the detector, it turns out that a memoryless transmission strategy is strictly sub-optimal in general (Section 5). This is in contrast to the optimality of a memoryless transmission scheme when the transmitter and detector share randomness. Another related work, especially to Section 5 on the private randomness case, is [12] as we discuss there. It considered communication and testing in a similar model though error exponents for testing were not considered there.

## 2. PRELIMINARIES

**Adversarial Hypothesis Testing.** Our achievability proofs use the adversarial Chernoff-Stein lemma from [6] which we briefly describe here. Let  $\mathcal{Z}$  be a finite set. Let  $\mathcal{P}, \mathcal{Q} \subseteq \mathbb{R}^{\mathcal{Z}}$  be closed, convex sets of probability distributions. The adaptive adversary is specified by  $\hat{p}_i : \mathcal{Z}^{i-1} \rightarrow \mathcal{P}$  and  $\hat{q}_i : \mathcal{Z}^{i-1} \rightarrow \mathcal{Q}$  for  $i \in [1 : n]$ . For any  $z^n \in \mathcal{Z}^n$ , let  $\hat{p}(z^n) := \prod_{i=1}^n \hat{p}_i(z^{i-1})(z_i)$  and  $\hat{q}(z^n) := \prod_{i=1}^n \hat{q}_i(z^{i-1})(z_i)$ . Let  $A_n \subseteq \mathcal{Z}^n$  be an acceptance region for  $\mathcal{P}$ . For  $\epsilon > 0$ , the type-I and type-II errors are defined to be

$$\alpha_n \stackrel{\text{def}}{=} \sup_{(\hat{p}_i)_{i=1}^n} \hat{p}(A_n^c), \quad \beta_n^\epsilon \stackrel{\text{def}}{=} \min_{A_n : \alpha_n \leq \epsilon} \sup_{(\hat{q}_i)_{i=1}^n} \hat{q}(A_n),$$

and the adversarial Chernoff-Stein exponent is given by

$$\mathcal{E}_{\text{adv}}^\epsilon(\mathcal{P}, \mathcal{Q}) \stackrel{\text{def}}{=} \lim_{n \rightarrow \infty} -\frac{1}{n} \log \beta_n^\epsilon.$$

For any pair  $p \in \mathcal{P}, q \in \mathcal{Q}$ , since the adversary may (non-adaptively) choose  $\hat{p}_i = p$  and  $\hat{q}_i = q$  for all  $i \in [1 : n]$ , by the Chernoff-Stein lemma [3, Theorem 11.8.3] it is clear that  $\mathcal{E}_{\text{adv}}^\epsilon(\mathcal{P}, \mathcal{Q}) \leq \min_{p \in \mathcal{P}, q \in \mathcal{Q}} D(p \| q)$ . In [5] it was shown that this upper bound is achievable if the adversary is non-adaptive. The following theorem states that this remains true even when the adversary is adaptive.

**Theorem 1** (Adversarial Chernoff-Stein Lemma [6]). Let  $\mathcal{Z}$  be a finite domain. For any pair of closed, convex sets of probability distributions  $\mathcal{P}, \mathcal{Q} \subseteq \mathbb{R}^{\mathcal{Z}}$ ,

$$\mathcal{E}_{\text{adv}}^\epsilon(\mathcal{P}, \mathcal{Q}) = \min_{p \in \mathcal{P}, q \in \mathcal{Q}} D(p \| q). \quad (1)$$

**Problem Setup.** Let  $\mathcal{X}$  and  $\mathcal{Y}$  be finite sets. A discrete memoryless channel  $W(\cdot | \cdot)$  takes an input symbol  $x \in \mathcal{X}$  and

outputs a symbol  $y \in \mathcal{Y}$  with probability  $W(y|x)$ . Consider two finite sets of channels  $\mathcal{W} = \{W(\cdot | \cdot, s) : s \in \mathcal{S}\}$ ,  $\overline{\mathcal{W}} = \{\overline{W}(\cdot | \cdot, \bar{s}) : \bar{s} \in \overline{\mathcal{S}}\}$ . The goal is to distinguish between the two sets of channels. In particular, we study the asymmetric hypothesis test between the null hypothesis  $H_0 : \mathcal{W}$  and the alternative hypothesis  $H_1 : \overline{\mathcal{W}}$ . There are three entities involved: (a) the transmitter, (b) the adversary, and (c) the detector. The transmitter is unaware of which hypothesis has been realized and chooses the input symbols. The adversary, depending on which hypothesis is realized, chooses the state symbols (from  $\mathcal{S}$  under  $H_0$  and  $\overline{\mathcal{S}}$  under  $H_1$ ). The detector decides between  $H_0$  and  $H_1$  based on everything it knows. We will elaborate this in the coming sections.

## 3. SHARED RANDOMNESS

In this setting, the transmitter and detector share randomness which is unknown to the adversary. The input  $X^n$  to the channel, which is a function of this randomness, is known to the detector. For a transmitter strategy  $P_{X^n}$  and a pair of adversary strategies  $P_{S^n}$  and  $P_{\overline{S}^n}$ , the distribution induced on  $\mathcal{X}^n \times \mathcal{Y}^n$  under  $H_0$  is given by

$$Q_{\text{sh}}^n(x^n, y^n) = \sum_{s^n \in \mathcal{S}^n} P_{X^n}(x^n) P_{S^n}(s^n) \prod_{i=1}^n W(y_i | x_i, s_i). \quad (2)$$

A similar expression is obtained for  $\overline{Q}_{\text{sh}}^n$  under  $H_1$  where instead of  $P_{S^n}$  and  $W$  we have  $P_{\overline{S}^n}$  and  $\overline{W}$  respectively. The detector uses a (possibly privately randomized) decision rule  $f_{\text{sh}} : \mathcal{X}^n \times \mathcal{Y}^n \rightarrow \{0, 1\}$ . Let  $A_n$  be the (possibly random) acceptance region for  $H_0$ , i.e.,  $A_n = \{(x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n : f_{\text{sh}}(x^n, y^n) = 0\}$ . For a given transmitter and detector strategy, the type-I error is given by

$$\alpha_n^{\text{sh}} = \sup_{P_{S^n}} \mathbb{E} [Q_{\text{sh}}^n(A_n^c)],$$

where the expectation is over the random choice of  $A_n$ . For  $\epsilon > 0$ , when the type-I error  $\alpha_n^{\text{sh}}$  is at most  $\epsilon$ , the optimal type-II error is given by

$$\beta_n^{\epsilon, \text{sh}} \stackrel{\text{def}}{=} \inf_{P_{X^n}} \inf_{A_n : \alpha_n^{\text{sh}} \leq \epsilon} \sup_{P_{\overline{S}^n}} \mathbb{E} [\overline{Q}_{\text{sh}}^n(A_n)],$$

where the expectation is over the random  $A_n$  set by the inner inf. The Chernoff-Stein exponent is then defined to be

$$\mathcal{E}_{\text{sh}}^\epsilon(\mathcal{W}, \overline{\mathcal{W}}) \stackrel{\text{def}}{=} \liminf_{n \rightarrow \infty} -\frac{1}{n} \log \beta_n^{\epsilon, \text{sh}}, \quad \epsilon > 0.$$

Let  $\text{conv}(\mathcal{W})$  and  $\text{conv}(\overline{\mathcal{W}})$  be the convex hulls of the channel sets  $\mathcal{W}$  and  $\overline{\mathcal{W}}$  respectively. i.e.,

$$\text{conv}(\mathcal{W}) \stackrel{\text{def}}{=} \left\{ \sum_{s \in \mathcal{S}} P_S(s) W(\cdot | \cdot, s) : P_S \in \Delta_{\mathcal{S}} \right\},$$

where  $\Delta_{\mathcal{S}}$  is the set of all probability distributions over  $\mathcal{S}$ .  $\text{conv}(\overline{\mathcal{W}})$  is defined similarly with  $\overline{S}, \overline{W}$  instead of  $S, W$ . Let

$$D_{\text{sh}}^* \stackrel{\text{def}}{=} \sup_{P_X} \min_{\substack{W \in \text{conv}(\mathcal{W}) \\ \overline{W} \in \text{conv}(\overline{\mathcal{W}})}} D(W \parallel \overline{W} | P_X). \quad (3)$$

Since  $\text{conv}(\mathcal{W}), \text{conv}(\overline{\mathcal{W}})$  are closed, convex sets and  $D(\cdot | \cdot)$  is lower semi-continuous, the minimum exists.

**Theorem 2.** Let  $\mathcal{W}$  and  $\overline{\mathcal{W}}$  be two sets of discrete memoryless channels which map  $\mathcal{X}$  to  $\mathcal{Y}$ . For any  $\epsilon \in (0, 1)$ , we have

$$D_{\text{sh}}^* \leq \mathcal{E}_{\text{sh}}^\epsilon(\mathcal{W}, \overline{\mathcal{W}}) \leq \frac{D_{\text{sh}}^*}{1 - \epsilon}. \quad (4)$$

*Proof. Achievability* ( $\mathcal{E}_{\text{sh}}^\epsilon(\mathcal{W}, \overline{\mathcal{W}}) \geq D_{\text{sh}}^*$ ): We argue the achievability for the (stronger) adaptive adversary who has access to previous channel inputs and outputs. The transmitter transmits  $X^n$  chosen i.i.d. according to  $P_X$  using the shared randomness. This reduces the problem to the adversarial hypothesis testing problem studied in [6]. For any fixed choice of  $P_X$ , invoking Theorem 1 with  $\mathcal{P} = \{P_X W : W \in \text{conv}(\mathcal{W})\}$  and  $\mathcal{Q} = \{P_X \overline{W} : \overline{W} \in \text{conv}(\overline{\mathcal{W}})\}$ ,

$$\mathcal{E}_{\text{sh}}^\epsilon(\mathcal{W}, \overline{\mathcal{W}}) \geq \min_{\substack{W \in \text{conv}(\mathcal{W}) \\ \overline{W} \in \text{conv}(\overline{\mathcal{W}})}} D(W \parallel \overline{W} | P_X).$$

Optimizing over  $P_X$  completes the proof of achievability.

*Weak Converse* ( $\mathcal{E}_{\text{sh}}^\epsilon(\mathcal{W}, \overline{\mathcal{W}}) \leq \frac{D_{\text{sh}}^*}{1 - \epsilon}$ ): We show this converse result for an adaptive transmitter who has feedback of the outputs. Given an adaptive transmitter, we construct an adversarial strategy to show the upper bound on the exponent. Specifically, we consider a memoryless strategy (not necessarily i.i.d.) for the adversary, i.e.  $P_{S^n} = \prod_{i=1}^n P_{S_i}$  and  $P_{\overline{S}^n} = \prod_{i=1}^n P_{\overline{S}_i}$  where  $P_{S_i}$  and  $P_{\overline{S}_i}$  will be specified in course of the proof. Let  $Q^n$  and  $\overline{Q}^n$  denote the joint distributions on  $\mathcal{X}^n \times \mathcal{Y}^n$  under  $H_0$  and  $H_1$  respectively. They are given by

$$Q^n(x^n, y^n) = \prod_{i=1}^n \overline{Q}_i(x_i | x^{i-1}, y^{i-1}) \left( \sum_{s_i \in \mathcal{S}} P_{S_i}(s_i) W(y_i | x_i, s_i) \right) \quad (5)$$

and

$$\overline{Q}^n(x^n, y^n) = \prod_{i=1}^n \overline{Q}_i(x_i | x^{i-1}, y^{i-1}) \left( \sum_{\overline{s}_i \in \overline{\mathcal{S}}} P_{\overline{S}_i}(\overline{s}_i) \overline{W}(y_i | x_i, \overline{s}_i) \right). \quad (6)$$

Here,  $\overline{Q}_i(x_i | x^{i-1}, y^{i-1})$  denotes the transmitter strategy at the  $i^{\text{th}}$  timestep. We now try to get an upper bound on  $D(Q^n \parallel \overline{Q}^n)$ .

$$\begin{aligned} D(Q^n \parallel \overline{Q}^n) &= \sum_{i=1}^n D(Q_{X_i, Y_i | (X, Y)^{i-1}} \parallel \overline{Q}_{X_i, Y_i | (X, Y)^{i-1}} | Q^{i-1}) \\ &= \sum_{i=1}^n (D(\overline{Q}_i \parallel \overline{Q}_i | Q^{i-1}) \\ &\quad + D(Q_{Y_i | X^i, Y^{i-1}} \parallel \overline{Q}_{Y_i | X^i, Y^{i-1}} | Q^{i-1} \overline{Q}_i)) \end{aligned}$$

Observe that all the  $D(\overline{Q}_i \parallel \overline{Q}_i | Q^{i-1})$  terms are zero. Furthermore, from (5), (6), we can see that  $Q_{Y_i | X^i, Y^{i-1}} = Q_{Y_i | X_i}$ ,  $\overline{Q}_{Y_i | X^i, Y^{i-1}} = \overline{Q}_{Y_i | X_i}$ . Thus,

$$\begin{aligned} D(Q^n \parallel \overline{Q}^n) &= \sum_{i=1}^n D(Q_{Y_i | X_i} \parallel \overline{Q}_{Y_i | X_i} | Q^{i-1} \overline{Q}_i) \\ &= \sum_{i=1}^n D(Q_{Y_i | X_i} \parallel \overline{Q}_{Y_i | X_i} | Q_{X_i}) \end{aligned} \quad (7)$$

It is easy to see that  $(P_{S_1}, P_{\overline{S}_1})$  can be chosen such that  $D(Q_{Y_1 | X_1} \parallel \overline{Q}_{Y_1 | X_1} | \overline{Q}_1) \leq D_{\text{sh}}^*$ . We then recursively specify  $(P_{S_i}, P_{\overline{S}_i})$  such that each term in (7) is upper bounded by  $D_{\text{sh}}^*$ . Thus,

$$D(Q^n \parallel \overline{Q}^n) \leq n D_{\text{sh}}^*. \quad (8)$$

With this upper bound in place, we may follow a standard approach via the data processing inequality to complete the proof (e.g., see [8, Section VI]). See Appendix A where we complete these steps.  $\square$

A similar approach of choosing a memoryless (not necessarily i.i.d.) adversary strategy also allows us to use the proof technique of [8, Section VI], [13] to obtain the following strong converse (see Appendix B for a proof). For distributions  $\mu_{XY}, \nu_{XY}$  on  $\mathcal{X} \times \mathcal{Y}$  and  $t \in \mathbb{R}$ , let

$$\phi_t(\mu_{Y|X} \parallel \nu_{Y|X} | \mu_X) \stackrel{\text{def}}{=} \log_{X \sim \mu_X} \mathbb{E} \left[ \sum_{\mathcal{Y}} \mu_{Y|X}^{1-t} \nu_{Y|X}^t \right].$$

**Theorem 3.** If

$$\begin{aligned} \limsup_{t \rightarrow 0^-} \inf_{\substack{P_X \\ W \in \text{conv}(\mathcal{W}) \\ \overline{W} \in \text{conv}(\overline{\mathcal{W}})}} \frac{\phi_t(W \parallel \overline{W} | P_X)}{-t} \\ = \sup_{P_X} \inf_{\substack{W \in \text{conv}(\mathcal{W}) \\ \overline{W} \in \text{conv}(\overline{\mathcal{W}})}} \lim_{t \rightarrow 0^-} \frac{\phi_t(W \parallel \overline{W} | P_X)}{-t}, \end{aligned} \quad (9)$$

then

$$\mathcal{E}_{\text{sh}}^\epsilon(\mathcal{W}, \overline{\mathcal{W}}) \leq D_{\text{sh}}^*. \quad (10)$$

The following theorem characterizes the pairs of  $(\mathcal{W}, \overline{\mathcal{W}})$  for which  $\mathcal{E}_{\text{sh}}^\epsilon > 0$ .

**Theorem 4.**  $\mathcal{E}_{\text{sh}}^\epsilon(\mathcal{W}, \overline{\mathcal{W}}) > 0 \iff \text{conv}(\mathcal{W}) \cap \text{conv}(\overline{\mathcal{W}}) = \emptyset$ .

*Proof.* The if ( $\Leftarrow$ ) part follows from Theorem 2. To see the (contrapositive of the) only if ( $\Rightarrow$ ) direction, notice that under hypothesis  $H_0$  (resp.,  $H_1$ ), the adversary may induce any channel from  $\text{conv}(\mathcal{W})$  (resp.,  $\text{conv}(\overline{\mathcal{W}})$ ) from the transmitter to the detector. Hence, when the intersection is non-empty, the adversary may induce the same channel under both hypotheses so that no transmission strategy (including an adaptive one) can distinguish between the hypotheses.  $\square$

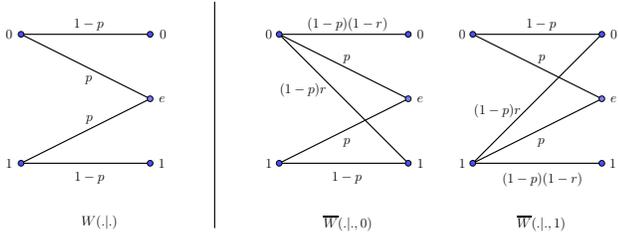


Fig. 1. An example in which, for a privately randomized transmitter, the hypotheses cannot be distinguished using memoryless transmission schemes, but a scheme with 2-step memory yields a positive Chernoff-Stein exponent.

#### 4. DETERMINISTIC TRANSMITTER

In this setting, the transmitter strategy is completely deterministic and is defined by a fixed tuple  $(x_1, x_2, \dots, x_n)$ . The distribution on  $\mathcal{Y}^n$  under  $H_0$  and  $H_1$  are similar to (2) with  $P_{X^n}$  as a point mass on  $(x_1, x_2, \dots, x_n)$ . The definitions of decision rule  $f_{\text{det}}$  and acceptance region  $A_n$  are similar to those in Section 3 except that the observation space is  $\mathcal{Y}^n$  instead of  $\mathcal{X}^n \times \mathcal{Y}^n$ . The definitions of  $\alpha_n^{\text{det}}$ ,  $\beta_n^{\epsilon, \text{det}}$  and  $\mathcal{E}_{\text{det}}^\epsilon$  are also similar except that the inf is over the input symbols in the expression for  $\beta_n^{\epsilon, \text{det}}$ .

For  $x \in \mathcal{X}$ , let  $\text{conv}(\mathcal{W}_x)$  and  $\text{conv}(\overline{\mathcal{W}}_x)$  be the convex hulls of the conditional distributions  $W(\cdot|x, s)$  and  $\overline{W}(\cdot|x, \bar{s})$ .

$$\text{conv}(\mathcal{W}_x) \stackrel{\text{def}}{=} \left\{ \sum_{s \in \mathcal{S}} P_S(s) W(\cdot|x, s) : P_S \in \Delta_{\mathcal{S}} \right\},$$

$\text{conv}(\overline{\mathcal{W}}_x)$  is defined similarly with  $\bar{S}, \overline{W}$  instead of  $S, W$ . Define  $D_{\text{det}}^*$  to be

$$D_{\text{det}}^* := \max_x \min_{\substack{W_x \in \text{conv}(\mathcal{W}_x) \\ \overline{W}_x \in \text{conv}(\overline{\mathcal{W}}_x)}} D(W_x \| \overline{W}_x) \quad (11)$$

**Theorem 5.** Let  $\mathcal{W}$  and  $\overline{\mathcal{W}}$  be two sets of discrete memoryless channels which map  $\mathcal{X}$  to  $\mathcal{Y}$ . For any  $\epsilon \in (0, 1)$ , we have

$$D_{\text{det}}^* \leq \mathcal{E}_{\text{det}}^\epsilon(\mathcal{W}, \overline{\mathcal{W}}) \leq \frac{D_{\text{det}}^*}{1 - \epsilon}. \quad (12)$$

The proof (Appendix C) is on similar lines as Theorem 2. We also show that (12) holds when both the transmitter and the adversary are adaptive (Appendix D). A strong converse and characterization theorem analogous to Theorems 3 and 4 can also be shown. We omit these in the interest of space.

#### 5. PRIVATE RANDOMNESS

We now consider the case where the transmitter may choose the channel input  $X^n$  randomly, but the realization of  $X^n$  is unavailable to the detector and the adversary. We may define the optimal Chernoff-Stein exponent  $\mathcal{E}_{\text{pvt}}^\epsilon(\mathcal{W}, \overline{\mathcal{W}})$  along the same lines as earlier sections. Note that the decision function is now a (possibly random) partition of  $\mathcal{Y}^n$ . By the discussion leading up to Theorem 1, if the transmitter adopts an i.i.d.  $P_X$

strategy, the best possible exponent (irrespective of whether the adversary is adaptive or not) is

$$D_{\text{pvt}, \text{iid}} = \sup_{P_X} \min_{\substack{Q_Y \in \mathcal{Q} \\ \bar{Q}_Y \in \bar{\mathcal{Q}}}} D(Q_Y \| \bar{Q}_Y),$$

where  $\mathcal{Q}$  (resp.  $\bar{\mathcal{Q}}$ ) is the set of (single-letter) channel output distributions that can be induced by the adversary when the input is distributed as  $P_X$  under hypothesis  $H_0$  (resp.  $H_1$ ), i.e.,  $\mathcal{Q} \stackrel{\text{def}}{=} \left\{ \sum_{x,s} P_S(s) P_X(x) W(\cdot|x, s) : P_S \in \Delta_{\mathcal{S}} \right\}$ . It turns out that in general the optimal exponent  $\mathcal{E}_{\text{pvt}}^\epsilon(\mathcal{W}, \overline{\mathcal{W}})$  could be strictly larger than  $D_{\text{pvt}, \text{iid}}$ . In the following example,  $\mathcal{E}_{\text{pvt}}^\epsilon(\mathcal{W}, \overline{\mathcal{W}}) > 0$  for all  $\epsilon > 0$  even though  $D_{\text{pvt}, \text{iid}} = 0$ .

**Example 1.**  $H_0 : \mathcal{W} = \{W(\cdot|\cdot)\}$  consists of a binary erasure channel (BEC) with parameter  $p < 1$  and  $H_1 : \overline{\mathcal{W}} = \{\overline{W}(\cdot|\cdot, 0), \overline{W}(\cdot|\cdot, 1)\}$  consists of two modified BEC( $p$ ) channels where one of the symbols flips with probability  $(1-p)r$ ,  $r > 0$  as shown in Figure 1. Here,  $\mathcal{X} = \{0, 1\}$ ,  $\mathcal{Y} = \{0, 1, e\}$ ,  $\mathcal{S} = \{0\}$ ,  $\bar{\mathcal{S}} = \{0, 1\}$ . Note that  $\mathcal{Q}$  is a singleton. It is easy to verify that, under  $H_1$ , if the adversary sets  $P_{\bar{S}}(0) = 1 - P_X(0)$ , the induced channel output distribution will be the same as the one under  $H_0$ . Hence,  $\mathcal{Q} \subset \bar{\mathcal{Q}}$  and therefore  $D_{\text{pvt}, \text{iid}} = 0$ .

Now to see that  $\mathcal{E}_{\text{pvt}}^\epsilon(\mathcal{W}, \overline{\mathcal{W}}) > 0$ , consider a transmission scheme with 2-step memory:  $n/2$  i.i.d. pairs are sent where each pair is distributed as  $P_{X_1, X_2}(0, 0) = P_{X_1, X_2}(1, 1) = 0.5$ . The effective channel is now a random map from  $\mathcal{X}^2$  to  $\mathcal{Y}^2$ . The new state space for the (non-adaptive) adversary under  $H_0$  is  $\mathcal{S}^2$  (which is still a singleton), and  $\bar{\mathcal{S}}^2$  under  $H_1$ . Let  $\mathcal{Q}_2$  (resp.  $\bar{\mathcal{Q}}_2$ ) be the set of (two-letter) channel output distributions that can be induced by the adversary when the input is distributed according to  $P_{X_1, X_2}$  under  $H_0$  (resp.  $H_1$ ). Since  $\mathcal{Q}_2$  is a singleton, let the member be denoted by  $Q_{Y_1, Y_2}$ . If we show that  $Q_{Y_1, Y_2} \notin \bar{\mathcal{Q}}_2$ , we may conclude that  $\mathcal{E}_{\text{pvt}}^\epsilon(\mathcal{W}, \overline{\mathcal{W}}) > 0$ . Assume for contradiction that this is not the case, i.e., suppose there exists  $P_{\bar{S}_1, \bar{S}_2}$  such that the resulting  $\bar{Q}_{Y_1, Y_2}$  is the same as  $Q_{Y_1, Y_2}$ . Since the marginals also have to be equal, we have  $Q_{Y_1} = \bar{Q}_{Y_1}$ . This forces  $P_{\bar{S}_1}$  to be uniform. Now, observe that  $Q_{Y_1, Y_2}(0, 1) = 0$  while, irrespective of  $P_{\bar{S}_2|S_1}$ , we have  $\bar{Q}_{Y_1, Y_2}(0, 1) > 0$  since  $r > 0$ . This is a contradiction and hence  $Q_{Y_1, Y_2} \notin \bar{\mathcal{Q}}_2$ . Therefore,  $\mathcal{E}_{\text{pvt}}^\epsilon(\mathcal{W}, \overline{\mathcal{W}}) > 0$  by Theorem 1.

The above argument does not account for an adaptive adversary. In Appendix E we show that even with an adaptive adversary the above transmission scheme leads to a positive exponent.

**Remark 1.** For the above example,  $\mathcal{E}_{\text{det}}^\epsilon(\mathcal{W}, \overline{\mathcal{W}}) < \mathcal{E}_{\text{pvt}}^\epsilon(\mathcal{W}, \overline{\mathcal{W}})$ . This follows from  $D_{\text{det}}^* \leq D_{\text{pvt}, \text{iid}}$  which is a consequence of the fact that for  $P_X$  such that  $P_X(x) = 1$  for some  $x \in \mathcal{X}$ , the corresponding  $\mathcal{Q}$  and  $\bar{\mathcal{Q}}$  are  $\text{conv}(\mathcal{W}_x)$  and  $\text{conv}(\overline{\mathcal{W}}_x)$  respectively.

In the rest of this section, we give an achievable lower bound on the error exponent  $\mathcal{E}_{\text{pvt}}^\epsilon(\mathcal{W}, \overline{\mathcal{W}})$  and characterize the pairs

$(\mathcal{W}, \overline{\mathcal{W}})$  for which it is positive<sup>2</sup>. If  $\text{conv}(\mathcal{W}) \cap \text{conv}(\overline{\mathcal{W}}) \neq \emptyset$ , then  $\mathcal{E}_{\text{pvt}}^\epsilon(\mathcal{W}, \overline{\mathcal{W}}) = 0$  (by Theorem 4). This follows from the fact that the adversary can choose  $S^n$  and  $\bar{S}^n$  i.i.d. so that a channel in the intersection may be induced which renders the hypotheses indistinguishable irrespective of the transmission scheme. It turns out that when the transmitter only has private randomness, a more carefully chosen adversary strategy which now depends on the transmission scheme may render  $\mathcal{E}_{\text{pvt}}^\epsilon(\mathcal{W}, \overline{\mathcal{W}}) = 0$  for a larger class of  $(\mathcal{W}, \overline{\mathcal{W}})$  pairs.

**Definition 1** ([12, eq. (2)]). The pair  $(\mathcal{W}, \overline{\mathcal{W}})$  is *trans-symmetrizable* if there exist conditional distributions  $P_{S|X}, P_{\bar{S}|X}$  such that, for every  $x, \tilde{x} \in \mathcal{X}$  and  $y \in \mathcal{Y}$ ,

$$\sum_{s \in \mathcal{S}} P_{S|X}(s|x) W(y|\tilde{x}, s) = \sum_{\bar{s} \in \bar{\mathcal{S}}} P_{\bar{S}|X}(\bar{s}|\tilde{x}) W(y|x, \bar{s}). \quad (13)$$

Consider a trans-symmetrizable pair  $(\mathcal{W}, \overline{\mathcal{W}})$  and a (non-adaptive<sup>3</sup>) transmission scheme  $\hat{P}$ . We will demonstrate (non-adaptive) adversary strategies under which the detector is unable to distinguish between the hypotheses. Under hypothesis  $H_1$ , the adversary, independent of the transmitter, samples a  $\tilde{X}^n$  according to  $\hat{P}$  and passes it through the (memoryless) channel  $P_{\bar{S}|X}$  of Definition 1 to produce its  $\bar{S}^n$ . This induces the following distribution on the channel output vector:

$$\begin{aligned} & \sum_{x^n, \bar{s}^n} \hat{P}(x^n) \left[ \sum_{\tilde{x}^n} \hat{P}(\tilde{x}^n) \prod_{i=1}^n (P_{\bar{S}|X}(\bar{s}_i|\tilde{x}_i)) \right] W^n(y^n|x^n, \bar{s}^n) \\ &= \sum_{x^n, \tilde{x}^n} \hat{P}(x^n) \hat{P}(\tilde{x}^n) \prod_{i=1}^n \left[ \sum_{\bar{s}_i \in \bar{\mathcal{S}}} P_{\bar{S}|X}(\bar{s}_i|\tilde{x}_i) W(y_i|x_i, \bar{s}_i) \right] \\ &\stackrel{(a)}{=} \sum_{\tilde{x}^n, x^n} \hat{P}(\tilde{x}^n) \hat{P}(x^n) \prod_{i=1}^n \left[ \sum_{s_i \in \mathcal{S}} P_{S|X}(s_i|x_i) W(y_i|\tilde{x}_i, s_i) \right] \\ &= \sum_{\tilde{x}^n, s^n} \hat{P}(\tilde{x}^n) \left[ \sum_{x^n} \hat{P}(x^n) \prod_{i=1}^n (P_{S|X}(s_i|x_i)) \right] W^n(y^n|\tilde{x}^n, s^n) \end{aligned}$$

where (a) follows from (13). This is identical to the channel output distribution under hypothesis  $H_0$  if the adversary samples from  $\hat{P}$  (independent of the transmitter) and passes through the channel  $P_{S|X}$  of Definition 1 to produce its  $S^n$ . Thus,  $\mathcal{E}_{\text{pvt}}^\epsilon(\mathcal{W}, \overline{\mathcal{W}}) = 0$  if  $(\mathcal{W}, \overline{\mathcal{W}})$  is trans-symmetrizable. The example below establishes a separation between shared and private randomness.

**Example 2** ([12, Example 1]). Let  $\mathcal{X} = \mathcal{S} = \bar{\mathcal{S}} = \{0, 1\}$  and  $\mathcal{Y} = \{0, 1\}^2$ . Suppose  $W$  deterministically outputs  $Y = (X, S)$  while  $\bar{W}$  outputs  $Y = (\bar{S}, X)$ . Clearly,  $\text{conv}(\mathcal{W}) \cap \text{conv}(\overline{\mathcal{W}}) = \emptyset$ . Hence, by Theorem 4,  $\mathcal{E}_{\text{sh}}^\epsilon > 0$ . However,  $(\mathcal{W}, \overline{\mathcal{W}})$  is trans-symmetrizable since  $P_{S|X}(x|x) = P_{\bar{S}|X}(x|x) = 1$  for all  $x \in \mathcal{X}$  satisfies (13). Hence  $\mathcal{E}_{\text{pvt}}^\epsilon(\mathcal{W}, \overline{\mathcal{W}}) = 0$ .

<sup>2</sup>This characterization is implicit in [14, Corollary 1]. Note that the ‘‘deterministic coding’’ transmitter there has access to the message which serves as a source of private randomness for the testing problem.

<sup>3</sup>This discussion can be modified to handle an adaptive transmission scheme if the adversary is also adaptive. This is omitted in the interest of space.

Our lower bound on  $\mathcal{E}_{\text{pvt}}^\epsilon(\mathcal{W}, \overline{\mathcal{W}})$  is in terms of the following quantitative measure of how far the pair  $(\mathcal{W}, \overline{\mathcal{W}})$  is from being trans-symmetrizable and/or having a non-empty intersection of their convex hulls; Lemma 1 and its proof in Appendix F makes this connection concrete.

**Definition 2.** For a distribution  $P$  over  $\mathcal{X}$ , we define  $\eta(P)$  as the set of triples  $(\eta_1, \eta_2, \eta_3)$  for which there exists  $\delta > 0$  such that there is no joint distribution  $P_{X X' \bar{S} S Y}$  with  $P_X = P_{X'} = P$  satisfying

- 1)  $I(X; \bar{S}) < \eta_1$ ,
- 2)  $I(X'; S) < \delta$ ,
- 3)  $D(P_{X \bar{S} Y} || P_{X \bar{S}} W) < \eta_2$ ,
- 4)  $D(P_{X' S Y} || P_{X' S} W) < \delta$ , and
- 5) if  $P_{X X'}(X' \neq X) > 0$ ,
  - (i)  $I(X'; X Y | \bar{S}) < \eta_3$ , and
  - (ii)  $I(X; X' Y | S) < \delta$ .

Our main theorem for this section is the following:

**Theorem 6.** Let  $\epsilon > 0$ .

$$\mathcal{E}_{\text{pvt}}^\epsilon(\mathcal{W}, \overline{\mathcal{W}}) = 0 \text{ if } (\mathcal{W}, \overline{\mathcal{W}}) \text{ is trans-symmetrizable or } \text{conv}(\mathcal{W}) \cap \text{conv}(\overline{\mathcal{W}}) \neq \emptyset$$

$$\mathcal{E}_{\text{pvt}}^\epsilon(\mathcal{W}, \overline{\mathcal{W}}) \geq \max \left\{ \max_{P, (\eta_1, \eta_2, \eta_3) \in \eta(P)} \min \left\{ \eta_1, \eta_2, \frac{\eta_3}{3} \right\}, D_{\text{det}}^* \right\}$$

**Lemma 1.** If  $(\mathcal{W}, \overline{\mathcal{W}})$  is not trans-symmetrizable and  $\text{conv}(\mathcal{W}) \cap \text{conv}(\overline{\mathcal{W}}) = \emptyset$ , there exists an input distribution  $P$  with  $(\eta_1, \eta_2, \eta_3) \in \eta(P)$  such that  $\eta_1, \eta_2, \eta_3 > 0$ .

**Corollary 1.**  $\mathcal{E}_{\text{pvt}}^\epsilon(\mathcal{W}, \overline{\mathcal{W}}) > 0$  if and only if  $(\mathcal{W}, \overline{\mathcal{W}})$  is not trans-symmetrizable and  $\text{conv}(\mathcal{W}) \cap \text{conv}(\overline{\mathcal{W}}) = \emptyset$ .

This recovers [14, Corollary 1] which gave the same characterization for  $(\mathcal{W}, \overline{\mathcal{W}})$  which allow hypothesis testing with vanishing probability of error when the transmitter has private randomness (in the form a random message). Our proof (in Appendix F) of the lower bound to  $\mathcal{E}_{\text{pvt}}^\epsilon(\mathcal{W}, \overline{\mathcal{W}})$  in Theorem 6, which is inspired by [14], entails significant careful modifications to the detector and the error analysis there.

## 6. ON THE ROLE OF ADAPTIVITY

1) *With shared randomness:* It turns out that our results hold even if the transmitter and/or adversary is adaptive. We proved the achievability part of Theorem 2 assuming that the adversary is adaptive and the converse assuming the transmitter is adaptive.

2) *Deterministic schemes:* Here the optimal exponent remains unchanged even if the adversary is adaptive (irrespective of whether the transmitter is adaptive or not). This is also the case if both the adversary and the transmitter are adaptive. These follow from our achievability proof which is shown assuming an adaptive adversary and the converse which is

shown when (a) both the transmitter and adversary are non-adaptive and (b) when both are adaptive (see Appendix D). It is also easy to see that, in general, if the transmitter is adaptive and the adversary is not, the exponent could be improved. The transmitter and detector may extract some randomness unknown to the adversary from the channel output feedback of, say, the first half of the block, and use this to implement a scheme with shared randomness during the second half. Since there are channels for which deterministic exponent is zero while the exponent under shared randomness is positive (for instance, see Example 2), these (possibly augmented by an independent random channel output component which provide additional shared randomness) serve as examples where such an improvement is feasible.

3) *With private randomness*: If the adversary is non-adaptive and the transmitter is adaptive, improved exponents are possible along the lines of the above discussion. This follows from the fact that there are channels where the exponent with shared randomness is positive, while that with private randomness is zero (specifically, trans-symmetrizable but with  $\text{conv}(\mathcal{W}) \cap \text{conv}(\overline{\mathcal{W}}) = \emptyset$ ; see Example 2). We also showed that memoryless schemes may be strictly sub-optimal even if the adversary is adaptive (Appendix E). Also, the impossibility result in Theorem 6 can be shown when both the transmitter and adversary are adaptive.

## REFERENCES

- [1] H. Chernoff, "A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations," *The Annals of Mathematical Statistics*, pp. 493–507, 1952.
- [2] W. Hoeffding, "Asymptotically optimal tests for multinomial distributions," *The Annals of Mathematical Statistics*, pp. 369–401, 1965.
- [3] T. M. Cover, *Elements of information theory*. John Wiley & Sons, 1999.
- [4] V. Strassen, "Meßfehler und information," *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, vol. 2, pp. 273–305, 1964.
- [5] F. Fangwei and S. Shiyi, "Hypothesis testing for arbitrarily varying source," *Acta Mathematica Sinica*, vol. 12, no. 1, pp. 33–39, 1996.
- [6] F. G. Brandão, A. W. Harrow, J. R. Lee, and Y. Peres, "Adversarial hypothesis testing and a quantum stein's lemma for restricted measurements," *IEEE Transactions on Information Theory*, vol. 66, no. 8, pp. 5037–5054, 2020.
- [7] R. Blahut, "Hypothesis testing and information theory," *IEEE Transactions on Information Theory*, vol. 20, no. 4, pp. 405–417, 1974.
- [8] M. Hayashi, "Discrimination of two channels by adaptive methods and its application to quantum system," *IEEE Transactions on Information Theory*, vol. 55, no. 8, pp. 3807–3820, 2009.
- [9] Y. Polyanskiy and S. Verdú, "Binary hypothesis testing with feedback," in *Information Theory and Applications Workshop (ITA)*, 2011.
- [10] D. Blackwell, L. Breiman, and A. Thomasian, "The capacities of certain channel classes under random coding," *The Annals of Mathematical Statistics*, vol. 31, no. 3, pp. 558–567, 1960.
- [11] A. Lapidoth and P. Narayan, "Reliable communication under channel uncertainty," *IEEE transactions on Information Theory*, vol. 44, no. 6, pp. 2148–2177, 1998.
- [12] S. Chaudhuri, N. Sangwan, M. Bakshi, B. K. Dey, and V. M. Prabhakaran, "Compound arbitrarily varying channels," in *2021 IEEE International Symposium on Information Theory (ISIT)*, pp. 503–508, IEEE, 2021.
- [13] H. Nagaoka, "Strong converse theorems in quantum information theory," in *Asymptotic Theory of Quantum Statistical Inference: Selected Papers*, pp. 64–65, World Scientific, 2005.
- [14] S. Chaudhuri, N. Sangwan, M. Bakshi, B. K. Dey, and V. M. Prabhakaran, "Compound arbitrarily varying channels," *arXiv preprint arXiv:2105.03420*, 2021.

- [15] T. Van Erven and P. Harremoës, "Rényi divergence and kullback-leibler divergence," *IEEE Transactions on Information Theory*, vol. 60, no. 7, pp. 3797–3820, 2014.
- [16] I. Csiszár, "The method of types [information theory]," *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2505–2523, 1998.
- [17] I. Csiszár and P. Narayan, "The capacity of the arbitrarily varying channel revisited: Positivity, constraints," *IEEE transactions on Information Theory*, vol. 34, no. 2, pp. 181–193, 1988.

## APPENDIX A PROOF OF THEOREM 2

The transmitter applies a (possibly randomized) decision rule  $f_{\text{sh}}$  on its observations  $(X^n, Y^n)$  which are distributed according to  $Q^n$  under  $H_0$  and  $\bar{Q}^n$  under  $H_1$ . The rule maps these observations to  $\{0,1\}$ . Let  $\tilde{\alpha}_n \stackrel{\text{def}}{=} \sum_{x^n, y^n} Q^n(x^n, y^n) \Pr[f_{\text{sh}}(x^n, y^n) = 1]$  and  $\tilde{\beta}_n \stackrel{\text{def}}{=} \sum_{x^n, y^n} \bar{Q}^n(x^n, y^n) \Pr[f_{\text{sh}}(x^n, y^n) = 0]$  be the type-1 and type-2 errors respectively. Observe that the distribution of the decision is  $\text{Bern}(\tilde{\alpha}_n)$  under  $H_0$  and  $\text{Bern}(1 - \tilde{\beta}_n)$  under  $H_1$ . By data processing inequality,

$$D(\text{Bern}(\tilde{\alpha}_n) \parallel \text{Bern}(1 - \tilde{\beta}_n)) \leq D(Q^n \parallel \bar{Q}^n).$$

Expanding out the L.H.S and using (8), we have

$$-h(\tilde{\alpha}_n) - (1 - \tilde{\alpha}_n) \log \tilde{\beta}_n - \tilde{\alpha}_n \log(1 - \tilde{\beta}_n) \leq nD_{\text{sh}}^*,$$

where  $h(\tilde{\alpha}_n) := -\tilde{\alpha}_n \log \tilde{\alpha}_n - (1 - \tilde{\alpha}_n) \log(1 - \tilde{\alpha}_n)$ . Since  $\tilde{\alpha}_n \log(1 - \tilde{\beta}_n) \leq 0$ , it follows that

$$-\frac{1}{n} \log \tilde{\beta}_n \leq \frac{D_{\text{sh}}^* + \frac{h(\tilde{\alpha}_n)}{n}}{1 - \tilde{\alpha}_n}.$$

Since we require  $\tilde{\alpha}_n \leq \epsilon$ , taking  $\liminf_{n \rightarrow \infty}$  on both sides,

$$\liminf_{n \rightarrow \infty} -\frac{1}{n} \log \tilde{\beta}_n \leq \frac{D_{\text{sh}}^*}{1 - \epsilon}.$$

It follows that

$$\mathcal{E}_{\text{sh}}^\epsilon(\mathcal{W}, \overline{\mathcal{W}}) \leq \frac{D_{\text{sh}}^*}{1 - \epsilon}.$$

## APPENDIX B PROOF OF THEOREM 3

Let  $\mu_{XY}, \nu_{XY}$  be distributions on  $\mathcal{X} \times \mathcal{Y}$ ,  $t \in \mathbb{R}$ .

$$\Phi_t(\mu_Y \parallel \nu_Y) \stackrel{\text{def}}{=} \sum_{\mathcal{Y}} \mu_Y^{1-t} \nu_Y^t$$

$$\Phi_t(\mu_{Y|X} \parallel \nu_{Y|X} \mid \mu_X) \stackrel{\text{def}}{=} \mathbb{E}_{X \sim \mu_X} [\Phi_t(\mu_{Y|X} \parallel \nu_{Y|X})]$$

$\phi_t$  is defined to be log of the corresponding  $\Phi_t$  quantity.

We again construct a memoryless adversary strategy. Let  $P_{S^n} = \prod_{i=1}^n P_{S_i}$ ,  $P_{\bar{S}^n} = \prod_{i=1}^n P_{\bar{S}_i}$  where  $P_{S_i}$  and  $P_{\bar{S}_i}$  will be specified in course of the proof.  $Q^n$  and  $\bar{Q}^n$  are the joint distributions on  $\mathcal{X}^n \times \mathcal{Y}^n$  as defined in (5), (6).

Define  $Q_{\text{tilt}}^{i-1}$  to be

$$Q_{\text{tilt}}^{i-1} = \frac{(Q^{i-1})^{1-t} (\bar{Q}^{i-1})^t}{\Phi_t(Q^{i-1} \parallel \bar{Q}^{i-1})}. \quad (14)$$

From the definition of  $\Phi_t(\cdot|\cdot)$ , we can see that  $Q_{\text{ult}}^{i-1}$  is a distribution on  $\mathcal{X}^{i-1} \times \mathcal{Y}^{i-1}$ . Let  $\tilde{Q}_{X_i}$  be the marginal on  $X_i$  induced by  $Q_{\text{ult}}^{i-1} \cdot \tilde{Q}_i$ ,

$$\tilde{Q}_{X_i}(x_i) = \sum_{x^{i-1}, y^{i-1}} Q_{\text{ult}}^{i-1}(x^{i-1}, y^{i-1}) \cdot \tilde{Q}_i(x_i|x^{i-1}, y^{i-1}).$$

Thus, we have

$$\begin{aligned} \Phi_t(Q^n \| \tilde{Q}^n) &= \sum_{\mathcal{X}^n \times \mathcal{Y}^n} (Q^n)^{1-t} (\tilde{Q}^n)^t \\ &\stackrel{(a)}{=} \Phi_t(Q^{n-1} \| \tilde{Q}^{n-1}) \sum_{\mathcal{X}^n \times \mathcal{Y}^n} Q_{\text{ult}}^{n-1} \tilde{Q}_n(Q_{Y_n|X_n})^{1-t} (\tilde{Q}_{Y_n|X_n})^t \\ &= \Phi_t(Q^{n-1} \| \tilde{Q}^{n-1}) \cdot \Phi_t(Q_{Y_n|X_n} \| \tilde{Q}_{Y_n|X_n} | \tilde{Q}_{X_n}), \end{aligned}$$

where (a) follows from the factorizing  $Q^n$  as  $Q^n = Q^{n-1} \cdot \tilde{Q}_n \cdot Q_{Y_n|X_n}$  and using (14). We break down the term  $\Phi_t(Q^{n-1} \| \tilde{Q}^{n-1})$  in a similar manner. Repeating this process and finally taking log on both sides, we get

$$\begin{aligned} \phi_t(Q^n \| \tilde{Q}^n) &= \log \Phi_t(Q^n \| \tilde{Q}^n) \\ &= \sum_{i=1}^n \phi_t(Q_{Y_i|X_i} \| \tilde{Q}_{Y_i|X_i} | \tilde{Q}_{X_i}) \end{aligned}$$

Define  $\phi_{\text{sh}}^*(t)$  to be

$$\phi_{\text{sh}}^*(t) \stackrel{\text{def}}{=} \sup_{P_X} \min_{\substack{W \in \text{conv}(\mathcal{W}) \\ \bar{W} \in \text{conv}(\bar{\mathcal{W}})}} \phi_t(W \| \bar{W} | P_X). \quad (15)$$

We now specify  $(P_{S_i}, P_{\bar{S}_i})$  in a manner similar to the one in the proof of the weak converse. Consider the first term in the sum. By the definition of  $\phi_{\text{sh}}^*(t)$  in (15),

$$\min_{P_{S_1}, P_{\bar{S}_1}} \phi_t(Q_{Y_1|X_1} \| \tilde{Q}_{Y_1|X_1} | \tilde{Q}_{X_1}) \leq \phi_{\text{sh}}^*(t).$$

Recall that  $\phi_t(\cdot|\cdot) = -tD_{1-t}(\cdot|\cdot)$  for  $t < 0$ , where  $D_{1-t}(\cdot|\cdot)$  is the Rényi divergence of order  $1-t$ . Since  $\mathcal{P} = \{P_X W : W \in \text{conv}(\mathcal{W})\}$ ,  $\mathcal{Q} = \{P_X \bar{W} : \bar{W} \in \text{conv}(\bar{\mathcal{W}})\}$  are closed, convex sets and  $D_{1-t}(\cdot|\cdot)$  is lower semi-continuous [15, Theorem 15], such a minimum exists. We choose  $(P_{S_1}, P_{\bar{S}_1})$  such that  $\phi_t(Q_{Y_1|X_1} \| \tilde{Q}_{Y_1|X_1} | \tilde{Q}_{X_1}) \leq \phi_{\text{sh}}^*(t)$ . We now recursively specify all the  $(P_{S_i}, P_{\bar{S}_i})$  in a similar manner. Thus, we have

$$\phi_t(Q^n \| \tilde{Q}^n) = \log \Phi_t(Q^n \| \tilde{Q}^n) \leq n\phi_{\text{sh}}^*(t). \quad (16)$$

We now follow the approach of [8, Section VI], [13]. Let  $\tilde{\alpha}_n$  and  $\tilde{\beta}_n$  be the type-1 and type-2 errors once the strategies of transmitter, detector and adversary are fixed. They are as defined in the Appendix B. Let

$$r \stackrel{\text{def}}{=} \liminf_{n \rightarrow \infty} \frac{-1}{n} \log \tilde{\beta}_n$$

Our goal is to show that if  $r > D_{\text{sh}}^*$ , then the type-1 error probability  $\tilde{\alpha}_n$  goes to 1 exponentially fast. As before the distribution of the decision is  $\text{Bern}(\tilde{\alpha}_n)$  under  $H_0$  and  $\text{Bern}(1-\tilde{\beta}_n)$  under  $H_1$ . Since data processing inequality holds for  $D_{1-t}(\cdot|\cdot)$  for  $t < 0$  [15, Theorem 9], we can apply it for  $\phi_t(\cdot|\cdot)$ .

$$\Phi_t(\text{Bern}(\tilde{\alpha}_n) \| \text{Bern}(1-\tilde{\beta}_n)) \leq \Phi_t(Q^n \| \tilde{Q}^n) = e^{\phi_t(Q^n \| \tilde{Q}^n)}$$

Expanding out the L.H.S. and using (16), we have

$$(1 - \tilde{\alpha}_n)^{1-t} (\tilde{\beta}_n)^t + (\tilde{\alpha}_n)^{1-t} (1 - \tilde{\beta}_n)^t \leq e^{n\phi_{\text{sh}}^*(t)}.$$

Since  $\tilde{\alpha}_n^{1-t} (1 - \tilde{\beta}_n)^t \geq 0$ , it can be dropped while retaining the inequality. Taking log followed by lim inf on both sides, we get

$$\begin{aligned} \liminf_{n \rightarrow \infty} -\frac{1}{n} \log(1 - \tilde{\alpha}_n) &\geq \frac{-tr - \phi_{\text{sh}}^*(t)}{1-t} \\ &\geq \sup_{t < 0} \frac{-t}{1-t} \left( r - \frac{\phi_{\text{sh}}^*(t)}{-t} \right). \end{aligned}$$

We now show that the L.H.S.  $> 0$  for some choice of  $t < 0$ .

$$\begin{aligned} \lim_{t \rightarrow 0^-} \frac{\phi_{\text{sh}}^*(t)}{-t} &\stackrel{(a)}{=} \sup_{P_X} \min_{\substack{W \in \text{conv}(\mathcal{W}) \\ \bar{W} \in \text{conv}(\bar{\mathcal{W}})}} \lim_{t \rightarrow 0^-} \frac{\phi_t(W \| \bar{W} | P_X)}{-t} \\ &\stackrel{(b)}{=} \sup_{P_X} \min_{\substack{W \in \text{conv}(\mathcal{W}) \\ \bar{W} \in \text{conv}(\bar{\mathcal{W}})}} D(W \| \bar{W} | P_X) \stackrel{(d)}{=} D_{\text{sh}}^*. \end{aligned}$$

where (a) is by the definition of  $\phi_{\text{sh}}^*$  in (15) and the assumption in (9), (b) follows from the fact that  $\frac{\phi_t(W \| \bar{W} | P_X)}{-t} = D_{1-t}(W \| \bar{W} | P_X)$  when  $t < 0$  and by the continuity  $D_{1-t}$  in  $t$  [15], (d) by the definition of  $D_{\text{sh}}^*$  (3). Since  $r > D_{\text{sh}}^*$ , we have  $r - \frac{\phi_{\text{sh}}^*(t')}{-t'} > 0$  for some  $t' < 0$ .

$$\liminf_{n \rightarrow \infty} -\frac{1}{n} \log(1 - \tilde{\alpha}_n) > 0$$

This inequality holds true for all possible transmitter and detector strategies  $(\tilde{Q}, A_n)$ . Thus, the probability of correctness under  $H_0$  decays exponentially.

## APPENDIX C PROOF OF THEOREM 5 (NO FEEDBACK)

a) *Achievability* ( $\mathcal{E}_{\text{det}}^\epsilon(\mathcal{W}, \bar{\mathcal{W}}) \geq D_{\text{det}}^*$ ): We apply the same argument given in the achievability proof of Theorem 2 for a fixed choice of  $x$ . We then optimize over  $x$  to complete the proof.

b) *Converse* ( $\mathcal{E}_{\text{det}}^\epsilon(\mathcal{W}, \bar{\mathcal{W}}) \leq \frac{D_{\text{det}}^*}{1-\epsilon}$ ): Recall that transmitter strategy is a fixed tuple  $(x_1, x_2, \dots, x_n)$ . Consider a memoryless adversary strategy. Let  $Q^n$  (resp.  $\tilde{Q}^n$ ) be the distribution induced on  $\mathcal{Y}$  under  $H_0$  (resp.  $H_1$ ). They are similar in form to (5), (6) with  $\tilde{Q}_i$  as a point mass on  $x_i$ . Under this setting,  $D(Q^n \| \tilde{Q}^n) = \sum_{i=1}^n D(Q_{Y_i} \| \tilde{Q}_{Y_i})$ . It is easy to see that each term in the sum is upper bounded by  $D_{\text{det}}^*$ . Thus,  $D(Q^n \| \tilde{Q}^n) \leq nD_{\text{det}}^*$ . The rest of the proof is similar to Theorem 2.

## APPENDIX D PROOF OF THEOREM 5 (FEEDBACK TO TRANSMITTER AND ADVERSARY)

The proof of achievability is same as Appendix C.

a) *Converse* ( $\mathcal{E}_{\text{det}}^\epsilon(\mathcal{W}, \overline{\mathcal{W}}) \leq \frac{D_{\text{det}}^*}{1-\epsilon}$ ): We restrict the adversary to choose the next state independently conditioned on the previous outputs of the channel, i.e.  $P_{S_i|S^{i-1}, Y^{i-1}} = P_{S_i|Y^{i-1}}$ ,  $P_{\bar{S}_i|\bar{S}^{i-1}, Y^{i-1}} = P_{\bar{S}_i|Y^{i-1}}$  where  $P_{S_i|Y^{i-1}}$  and  $P_{\bar{S}_i|Y^{i-1}}$  will be specified in course of the proof. The transmitter strategy is given by a set of deterministic functions  $\{g_i : \mathcal{Y}^{i-1} \rightarrow \mathcal{X}\}$ , where  $g_1$  is a constant function with value  $x_1$ . Let  $Q^n$  and  $\bar{Q}^n$  denote the joint distributions on  $\mathcal{Y}^n$  under  $H_0$  and  $H_1$  respectively.  $Q^n$  is given by

$$Q^n(y^n) = \prod_{i=1}^n \left( \sum_{s_i \in \mathcal{S}} P_{S_i|Y^{i-1}}(s_i|y^{i-1}) W(y_i|g_i(y^{i-1}), s_i) \right). \quad (17)$$

$\bar{Q}^n$  is defined similarly with  $\bar{S}, \bar{W}$ . We again try to upper bound  $D(Q^n \|\bar{Q}^n)$ .

$$D(Q^n \|\bar{Q}^n) = \sum_{i=1}^n D(Q_{Y_i|Y^{i-1}} \|\bar{Q}_{Y_i|Y^{i-1}} | Q_{Y^{i-1}}) \quad (18)$$

Consider the  $i^{\text{th}}$  term in (18). For each tuple  $(y^{i-1})$ , by the definition of  $D_{\text{det}}^*$  in (11), we have

$$\min_{\substack{P_{S_i|Y^{i-1}}(\cdot|y^{i-1}) \\ P_{\bar{S}_i|Y^{i-1}}(\cdot|y^{i-1})}} D(Q_{Y_i|Y^{i-1}}(\cdot|y^{i-1}) \|\bar{Q}_{Y_i|Y^{i-1}}(\cdot|y^{i-1})) \leq D_{\text{det}}^*. \quad (19)$$

For each tuple  $(y^{i-1})$ , we specify  $P_{S_i|Y^{i-1}}(\cdot|y^{i-1})$  and  $P_{\bar{S}_i|Y^{i-1}}(\cdot|y^{i-1})$  such that they satisfy (19). Since  $D(Q_{Y_i|Y^{i-1}} \|\bar{Q}_{Y_i|Y^{i-1}} | Q_{Y^{i-1}})$  is an averaging over  $y^{i-1}$ , it is upper bounded by  $D_{\text{det}}^*$  as well. Repeating this argument for each term in the sum (18), we get  $D(Q^n \|\bar{Q}^n) \leq nD_{\text{det}}^*$ . The rest of the proof is similar to Theorem 2.

#### APPENDIX E

##### ROLE OF MEMORY FOR A PRIVATELY RANDOMIZED TRANSMITTER: ADAPATIVE ADVERSARY CASE

Continuing the discussion from Example 1, we now allow the adversary access to feedback, i.e. its choice of state can depend on the outputs of the previous transmission. The new state spaces for the adversary are  $\mathcal{S}^2 = \{0\}$  and  $\bar{\mathcal{S}}^2 = \bar{\mathcal{S}} \times \Sigma$  where  $\Sigma = \{\sigma : \mathcal{Y} \rightarrow \{0, 1\}\}$ . Observe that  $\Sigma$  accounts for feedback. Note that  $|\bar{\mathcal{S}}^2| = 2 \times |\Sigma| = 16$ . The problem can now be thought of as a new hypothesis test between  $H_0 : \mathcal{W}^2 = \{W^2(\cdot, \cdot)\}$  where

$$W^2((y_1, y_2)|(x_1, x_2)) = W(y_1|x_1)W(y_2|x_2)$$

and  $H_1 : \bar{\mathcal{W}}^2 = \{\bar{W}^2(\cdot, \cdot, (\bar{s}, \sigma)) : (\bar{s}, \sigma) \in \bar{\mathcal{S}} \times \Sigma\}$  where

$$\begin{aligned} \bar{W}^2((y_1, y_2)|(x_1, x_2), (\bar{s}, \sigma)) \\ = \bar{W}(y_1|x_1, \bar{s})\bar{W}(y_2|x_2, \sigma(y_1)) \end{aligned}$$

Recall that the transmitter strategy was  $P_{X_1, X_2}(0, 0) = P_{X_1, X_2}(1, 1)$ . The adversary strategy is given by  $P_{\bar{S}, \sigma}$ . Let  $\mathcal{Q}$  (resp.  $\bar{\mathcal{Q}}$ ) be the set of all possible (double-letter) distributions that can be induced on  $\mathcal{Y}^2$  under  $H_0$  (resp.  $H_1$ ). Since  $\mathcal{Q}$  is a

singleton, let the member be denoted by  $Q_{Y_1, Y_2}$ . If  $\mathcal{Q} \cap \bar{\mathcal{Q}} = \emptyset$ , then by Theorem 1, we get a positive exponent. Assume for contradiction that this is not the case, i.e. there exists  $P_{\bar{S}, \sigma}$  such that the resulting  $\bar{Q}_{Y_1, Y_2}$  is same as  $Q_{Y_1, Y_2}$ . Since the marginals have to be equal, we have  $Q_{Y_1} = \bar{Q}_{Y_1}$ . This forces  $P_{\bar{S}}$  to be uniform. Now, observe that  $Q_{Y_1, Y_2}(0, 1) = 0$ . Examine the term corresponding to  $x_1 = x_2 = 1, \bar{s}_1 = 1$  in the expansion of  $\bar{Q}_{Y_1, Y_2}(0, 1)$ .

$$P_{X_1, X_2}(1, 1)P_{\bar{S}}(1) \sum_{\sigma_2 \in \Sigma} P_{\sigma|\bar{S}}(\sigma_2|1)\bar{W}(0|1, 1)\bar{W}(1|1, \sigma_2(0))$$

It cannot be zero since  $\bar{W}(0|1, 1) > 0, \bar{W}(1|1, \sigma_2(0)) > 0$  for all  $\sigma_2$  when  $0 < r < 1$ . Thus, we have a contradiction. This scheme gets us a positive exponent even when the adversary is adaptive.

#### APPENDIX F

##### PROOF OF LEMMA 1 AND THEOREM 6

We use bold faced letters to denote  $n$ -length vectors, for example,  $\mathbf{x}$  denotes a vector in  $\mathcal{X}^n$  and  $\mathbf{X}$  denotes a random vector taking values in  $\mathcal{X}^n$ . For a random variable  $X$ , we denote its distribution by  $P_X$  and use the notation  $X \sim P_X$  to indicate this. For an alphabet  $\mathcal{X}$ , let  $\mathcal{P}_{\mathcal{X}}^n$  denote the set of all empirical distributions of  $n$  length strings from  $\mathcal{X}^n$ . For a random variable  $X \sim P_X$  such that  $P_X \in \mathcal{P}_{\mathcal{X}}^n$ , let  $\mathcal{T}_X^n$  be the set of all  $n$ -length strings with empirical distribution  $P_X$ . For  $\mathbf{x} \in \mathcal{X}^n$ , the statement  $\mathbf{x} \in \mathcal{T}_X^n$  defines  $P_X$  as the empirical distribution of  $\mathbf{x}$  and a random variable  $X \sim P_X$ . For  $P_{X,Y} \in \mathcal{P}_{\mathcal{X} \times \mathcal{Y}}^n$  and  $\mathbf{y} \in \mathcal{Y}^n$ , let  $\mathcal{T}_{X|Y}^n(\mathbf{y}) = \{\mathbf{x} : (\mathbf{x}, \mathbf{y}) \in \mathcal{T}_{X,Y}^n\}$ . We denote  $2^a$  by  $\text{exp}(a)$ .

*Proof of Lemma 1.* We first prove Lemma 1 and show that show that if a pair of channels  $(\mathcal{W}, \overline{\mathcal{W}})$  is not trans-symmetrizable and  $\text{conv}(\mathcal{W}) \cap \text{conv}(\overline{\mathcal{W}}) = \emptyset$ , then for any full support input distribution  $P$ , there exist  $(\eta_1, \eta_2, \eta_3) \in \eta(P)$  such that  $\eta_1, \eta_2, \eta_3 > 0$ .

Note that if  $\text{conv}(\mathcal{W}) \cap \text{conv}(\overline{\mathcal{W}}) = \emptyset$ , there exists a constant  $\zeta_1 > 0$  such that for every  $P_{\bar{S}}$  on  $\bar{\mathcal{W}}$  and  $P_S$  on  $\mathcal{W}$ ,

$$\max_{x, y} \left| \sum_{\bar{s}} P_{\bar{S}}(\bar{s})\bar{W}(y|x, \bar{s}) - \sum_s P_S(s)W(y|x, s) \right| > \zeta_1. \quad (20)$$

Also, if  $(\mathcal{W}, \overline{\mathcal{W}})$  is not trans-symmetrizable, there exists  $\zeta_2 > 0$  such that for every  $P_{S|X}(s|x')$ ,  $s \in \mathcal{S}, x' \in \mathcal{X}$  and  $P_{\bar{S}|X}(\bar{s}|x), \bar{s} \in \bar{\mathcal{S}}, x \in \mathcal{X}$

$$\max_{x, x', y} \left| \sum_{s \in \mathcal{S}} P_{S|X}(s|x')W(y|x, s) - \sum_{\bar{s} \in \bar{\mathcal{S}}} P_{\bar{S}|X}(\bar{s}|x)W(y|x', \bar{s}) \right| > \zeta_2. \quad (21)$$

We consider a full support input distribution  $P$ . That is,  $\min_x P(x) \geq \alpha$  for some  $\alpha > 0$ . We will show that there exists  $(\eta_1, \eta_2, \eta_3) \in \eta(P)$  such that  $\eta_1, \eta_2, \eta_3 > 0$  for some  $\delta > 0$ . These choices only depend on  $\alpha, \zeta_1$  and  $\zeta_2$ .

Firstly, suppose there exists  $P_{XX'\bar{S}SY}$  such that for  $(X, X') \sim P_{XX'}$ ,  $P_{XX'}(X \neq X') > 0$  and conditions 1), 2), 3), 4) and 5) hold in Definition 2. We will show a contradiction using these conditions. We have, for  $\bar{W} = W_{Y|X, \bar{S}}$ ,

$$\begin{aligned} \eta_1 + \eta_2 + \eta_3 &> I(X; \bar{S}) + D(P_{X\bar{S}Y} \| P_{X\bar{S}}\bar{W}) + I(X'; XY|\bar{S}) \\ &= D(P_{X\bar{S}Y} \| P_X P_{X'\bar{S}} W_{Y|X, \bar{S}}) \\ &\geq D(P_{X\bar{S}Y} \| \sum_{\bar{s}} P_X P_{X'} P_{\bar{S}|X'}(\bar{s}|x') W_{Y|X, \bar{S}}(\cdot|x', \bar{s})). \end{aligned}$$

Using Pinsker's inequality, for some  $c > 0$ , this implies that

$$\begin{aligned} &\sum_{x, x', y} \left| P_{X\bar{S}Y}(x, x', y) \right. \\ &\quad \left. - \sum_{\bar{s}} P_X(x) P_{X'}(x') P_{\bar{S}|X'}(\bar{s}|x') W_{Y|X, \bar{S}}(y|x, \bar{s}) \right| \\ &\leq c\sqrt{\eta_1 + \eta_2 + \eta_3}. \end{aligned} \quad (22)$$

Similarly, from the remaining conditions in Definition 2, we can write

$$\begin{aligned} &\sum_{x, x', y} \left| P_{X\bar{S}Y}(x, x', y) - \right. \\ &\quad \left. - \sum_s P_X(x) P_{X'}(x') P_{S|X}(s|x) W_{Y|X', S}(y|x', s) \right| \leq c\sqrt{3\delta}. \end{aligned} \quad (23)$$

Combining (22) and (23), we obtain

$$\begin{aligned} &\sum_{x, x', y} P_X(x) P_{X'}(x') \left| \sum_{\bar{s}} P_{\bar{S}|X'}(\bar{s}|x') W_{Y|X, \bar{S}}(y|x, \bar{s}) - \right. \\ &\quad \left. \sum_s P_{S|X}(s|x) W_{Y|X', S}(y|x', s) \right| \leq c\sqrt{\eta_1 + \eta_2 + \eta_3} + c\sqrt{3\delta}. \end{aligned} \quad (24)$$

This implies that

$$\begin{aligned} &\max_{x, x', y} \left| \sum_{\bar{s}} P_{\bar{S}|X'}(\bar{s}|x') W_{Y|X, \bar{S}}(y|x, \bar{s}) \right. \\ &\quad \left. - \sum_s P_{S|X}(s|x) W_{Y|X', S}(y|x', s) \right| \leq \frac{c\sqrt{\eta_1 + \eta_2 + \eta_3} + c\sqrt{3\delta}}{\alpha^2} \end{aligned} \quad (25)$$

which is a contradiction to (21) for

$$\frac{c\sqrt{\eta_1 + \eta_2 + \eta_3} + c\sqrt{3\delta}}{\alpha^2} \leq \zeta_2. \quad (26)$$

Next, suppose that there exists  $P_{XX'\bar{S}SY}$  such that for  $(X, X') \sim P_{XX'}$ ,  $P_{XX'}(X = X') = 1$  such that conditions 1), 2), 3) and 4) hold in Definition 2. Setting  $X' = X$  and proceeding in a similar manner, one can show that

$$\begin{aligned} &\max_{x, y} \left| \sum_{\bar{s}} P_{\bar{S}}(\bar{s}) W_{Y|X, \bar{S}}(y|x, \bar{s}) - \sum_s P_S(s) W_{Y|X, S}(y|x, s) \right| \\ &\leq \frac{c\sqrt{\eta_1 + \eta_2} + c\sqrt{2\delta}}{\alpha} \end{aligned}$$

which is a contradiction to (20) for

$$\frac{c\sqrt{\eta_1 + \eta_2} + c\sqrt{2\delta}}{\alpha} \leq \zeta_1. \quad (27)$$

Since,  $\zeta_1$  and  $\zeta_2$  are both positive, we can choose  $\eta_1, \eta_2, \eta_3 > 0$  such that for some  $\delta > 0$ , (27) and (26) hold. Note that such a choice only depends on  $\alpha$ ,  $\zeta_1$  and  $\zeta_2$ .  $\square$

*Proof of Theorem 6.* We already discussed (after Definition 1) how trans-symmetrizability implies  $\mathcal{E}_{\text{pvt}}^\epsilon(\mathcal{W}, \bar{\mathcal{W}}) = 0$ . Here we provide the proof of the lower bound on the exponent. The proof uses the method of types (See [16]). For a distribution  $P$ ,  $(\eta_1, \eta_2, \eta_3) \in \eta(P)$  and  $R = \eta_3/3$ , we first show that we can obtain an exponent  $\gamma$  for the probability of error under Hypothesis  $H_1$ .

$$\gamma \geq \min \left\{ \min_{P_{X\bar{S}}} A_1, \eta_2 - \epsilon, \min_{P_{X\bar{S}X'SY}: I(X'; XY|\bar{S}) \geq \eta_3} A_2 \right\} \quad (28)$$

where  $A_1 = R - |R - I(X; \bar{S})|^+ - \epsilon$  and  $A_2 = \max \left\{ I(X; X'\bar{S}) - |R - I(X'; \bar{S})|^+ - \epsilon, \right.$   $(29)$

$$\left. I(Y; X'|X\bar{S}) - |R - I(X'; X\bar{S})|^+ - 2\epsilon \right\} \quad (30)$$

For  $N = \exp(nR)$ , let  $\mathcal{C}(P) = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$  be a set of sequences of type  $P$  given by Lemma 2 (proved later).

**Lemma 2.** For any  $\epsilon > 0$ , large enough  $n$ ,  $N := 2^{nR}$  for  $R \geq \epsilon$ , and type  $P$ , there exist sequences  $\mathbf{x}_1, \dots, \mathbf{x}_N \in \mathcal{X}^n$  of type  $P$ , such that for every  $\mathbf{x} \in \mathcal{X}^n$ ,  $\mathbf{s} \in \mathcal{S}^n \cup \bar{\mathcal{S}}^n$  and every joint type  $P_{XX'S}$ , we have

$$\begin{aligned} &|\{j : (\mathbf{x}, \mathbf{x}_j, \mathbf{s}) \in \mathcal{T}_{XX'S}^n\}| \\ &\leq \exp \left\{ n \left( |R - I(X'; XS)|^+ + \epsilon \right) \right\}, \end{aligned} \quad (31)$$

$$\begin{aligned} &\frac{1}{N} |\{i : (\mathbf{x}_i, \mathbf{s}) \in \mathcal{T}_{XS}^n\}| \\ &\leq \exp \left\{ n \left( |R - I(X; S)|^+ - R + \epsilon/2 \right) \right\}, \text{ and} \end{aligned} \quad (32)$$

$$\begin{aligned} &\frac{1}{N} |\{i : (\mathbf{x}_i, \mathbf{x}_j, \mathbf{s}) \in \mathcal{T}_{XX'S}^n \text{ for some } j \neq i\}| \\ &\leq \exp \left\{ n \left( |R - I(X'; S)|^+ - I(X; X'S) + \epsilon/2 \right) \right\} \end{aligned} \quad (33)$$

The transmitter sends an input sequence selected uniformly at random (using its private randomness) from  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N$ .

**Definition 3 (Detector).** Given sequences  $\{\mathbf{x}_1, \dots, \mathbf{x}_N\}$  each of type  $P$ , and for  $(\eta_1, \eta_2, \eta_3) \in \eta(P)$  and  $\delta > 0$  given by Definition 2,  $\phi(\mathbf{y}) = H_1$  if and only if  $i \in [1 : N]$  and  $\bar{\mathbf{s}} \in \bar{\mathcal{S}}^n$  exist s.t. for the joint empirical distribution  $P_{X\bar{S}Y}$  of  $(\mathbf{x}_i, \bar{\mathbf{s}}, \mathbf{y})$ ,

- 1)  $I(X; \bar{S}) < \eta_1$
- 2)  $D(P_{X\bar{S}Y} \| P_{X\bar{S}}\bar{W}) < \eta_2$ , and
- 3) for each  $j$  such that the joint empirical distribution  $P_{X\bar{S}X'SY}$  of  $(\mathbf{x}_i, \bar{\mathbf{s}}, \mathbf{x}_j, \mathbf{s}, \mathbf{y})$  for some  $\mathbf{s} \in \mathcal{S}^n$  satisfies  $I(X'; S) < \delta$  and  $D(P_{X'SY} \| P_{X'S}W) < \delta$ , we have  $I(X'; XY|\bar{S}) < \eta_3$ .

Suppose the active hypothesis is  $H_1$ . Firstly, notice that the probability of error under any randomized attack can be

written as an average over deterministic attacks and is thus maximized by a deterministic attack. So, it is sufficient to consider only deterministic attacks by the adversary. Suppose the adversary attack sequence is  $\bar{s} \in \bar{\mathcal{S}}^n$ .

Let  $P_{\mathbf{X}\mathbf{Y}}(\mathbf{x}_i, \mathbf{y}) = \frac{1}{N} W^n(\mathbf{y}|\mathbf{x}_i, \bar{s})$  for  $\mathbf{x}_i \in \mathcal{C}(P)$   $\mathbf{y} \in \mathcal{Y}^n$  and  $P_{\mathbf{X}\mathbf{Y}}(\mathbf{x}, \mathbf{y}) = 0$  for  $\mathbf{x} \notin \mathcal{C}(P)$ . Let  $(\mathbf{X}, \mathbf{Y}) \sim P_{\mathbf{X}\mathbf{Y}}$ . Define events  $\mathcal{E}_1 := \{(\mathbf{X}, \bar{s}) \in \mathcal{T}_{X\bar{s}}^n \text{ such that } I(X; \bar{S}) \geq \eta_1\}$ ,  $\mathcal{E}_2 := \{(\mathbf{X}, \bar{s}, \mathbf{Y}) \in \mathcal{T}_{X\bar{s}\mathbf{Y}}^n \text{ such that } D(P_{X\bar{s}\mathbf{Y}}||P_{X\bar{s}} \times \bar{W}) \geq \eta_2\}$ ,  $\mathcal{E}_3 := \{(\mathbf{X}, \bar{s}, \mathbf{Y}) \in \mathcal{T}_{X\bar{s}\mathbf{Y}}^n \text{ such that } I(X; \bar{S}) < \eta_1, D(P_{X\bar{s}\mathbf{Y}}||P_{X\bar{s}} \times \bar{W}) < \eta_2, \exists \mathbf{x}_j \neq \mathbf{X} \text{ such that } (\mathbf{x}_j, \mathbf{s}, \mathbf{Y}) \in \mathcal{T}_{X'\mathbf{s}\mathbf{Y}}^n \text{ for some } \mathbf{s} \in \mathcal{S}^n \text{ for which } I(X'; S) < \delta \text{ and } D(P_{X'\mathbf{s}\mathbf{Y}}||P_{X'\mathbf{s}}W) < \delta, \text{ but } I(X'; XY|\bar{S}) \geq \eta_3\}$ , and  $\mathcal{E}_4 := \{\exists \mathbf{s} \in \mathcal{S}^n \text{ such that } (\mathbf{X}, \bar{s}, \mathbf{s}, \mathbf{Y}) \in \mathcal{T}_{X\bar{s}\mathbf{s}\mathbf{Y}}^n \text{ for which } I(X; \bar{S}) < \eta_1, D(P_{X\bar{s}\mathbf{s}\mathbf{Y}}||P_{X\bar{s}} \times \bar{W}) < \eta_2, I(X; S) < \delta \text{ and } D(P_{X\mathbf{s}\mathbf{Y}}||P_{X\mathbf{s}}W) < \delta\}$ . Then,

$$\begin{aligned} P_{\mathbf{X}\mathbf{Y}}(\phi(\mathbf{Y}) = H_0) &\leq P_{\mathbf{X}\mathbf{Y}}(\mathcal{E}_1 \cup \mathcal{E}_2 \cup \mathcal{E}_3 \cup \mathcal{E}_4) \\ &\leq P_{\mathbf{X}\mathbf{Y}}(\mathcal{E}_1) + P_{\mathbf{X}\mathbf{Y}}(\mathcal{E}_2) + P_{\mathbf{X}\mathbf{Y}}(\mathcal{E}_3) + P_{\mathbf{X}\mathbf{Y}}(\mathcal{E}_4) \end{aligned}$$

We first note that  $P_{\mathbf{X}\mathbf{Y}}(\mathcal{E}_4) = 0$  because for  $(\eta_1, \eta_2, \eta_3) \in \eta(P)$  and  $\delta > 0$  given by Definition 2, there is no distribution  $\mathcal{T}_{X\bar{s}\mathbf{s}\mathbf{Y}}^n$  (with  $X' = X$ ) satisfying the conditions in  $\mathcal{E}_4$ . Next, we evaluate  $P_{\mathbf{X}\mathbf{Y}}(\mathcal{E}_1)$ ,

$$\begin{aligned} &\mathbb{P}_{\mathbf{X}\mathbf{Y}}((\mathbf{X}, \bar{s}) \in \mathcal{T}_{X\bar{s}}^n, I(X; \bar{S}) \geq \eta_1) \\ &= \frac{|i : (\mathbf{x}_i, \bar{s}) \in \mathcal{T}_{X\bar{s}}^n, I(X; \bar{S}) \geq \eta_1|}{N} \\ &= \sum_{P_{X\bar{s}} \in \mathcal{P}_{X \times \bar{s}}^n : I(X; \bar{S}) \geq \eta_1} \frac{|i : (\mathbf{x}_i, \bar{s}) \in \mathcal{T}_{X\bar{s}}^n|}{N} \\ &\stackrel{(a)}{\leq} \sum_{P_{X\bar{s}} : I(X; \bar{S}) \geq \eta_1} \exp\left\{n \left(|R - I(X; \bar{S})|^+ - R + \epsilon/2\right)\right\} \\ &\stackrel{(b)}{\leq} \max_{P_{X\bar{s}} : I(X; \bar{S}) \geq \eta_1} \exp\left\{-n \left(R - |R - I(X; \bar{S})|^+ - \epsilon\right)\right\} \end{aligned} \quad (34)$$

Here, (a) holds because of (32) and (b) holds for large  $n$  as the number of joint types is at most polynomial in  $n$ . Next, we evaluate  $P_{\mathbf{X}\mathbf{Y}}(\mathcal{E}_2)$ ,

$$\begin{aligned} &P_{\mathbf{X}\mathbf{Y}}(\{(\mathbf{X}, \bar{s}, \mathbf{Y}) \in \mathcal{T}_{X\bar{s}\mathbf{Y}}^n, D(P_{X\bar{s}\mathbf{Y}}||P_{X\bar{s}} \times \bar{W}) \geq \eta_2\}) \\ &= P_{\mathbf{X}\mathbf{Y}}\left(\bigcup_{\substack{P_{X\bar{s}\mathbf{Y}} \in \mathcal{P}_{X \times \bar{s} \times \mathcal{Y}}^n \\ D(P_{X\bar{s}\mathbf{Y}}||P_{X\bar{s}} \times \bar{W}) \geq \eta_2}} \{(\mathbf{X}, \bar{s}, \mathbf{Y}) \in \mathcal{T}_{X\bar{s}\mathbf{Y}}^n\}\right) \\ &= \sum_{\substack{P_{X\bar{s}\mathbf{Y}} \in \mathcal{P}_{X \times \bar{s} \times \mathcal{Y}}^n \\ D(P_{X\bar{s}\mathbf{Y}}||P_{X\bar{s}} \times \bar{W}) \geq \eta_2}} P_{\mathbf{X}\mathbf{Y}}((\mathbf{X}, \bar{s}, \mathbf{Y}) \in \mathcal{T}_{X\bar{s}\mathbf{Y}}^n) \end{aligned}$$

For any  $P_{X\bar{s}\mathbf{Y}} \in \mathcal{P}_{X \times \bar{s} \times \mathcal{Y}}^n$  such that  $D(P_{X\bar{s}\mathbf{Y}}||P_{X\bar{s}} \times \bar{W}) \geq \eta_2$ , we have

$$\begin{aligned} &P_{\mathbf{X}\mathbf{Y}}(\{(\mathbf{X}, \bar{s}, \mathbf{Y}) \in \mathcal{T}_{X\bar{s}\mathbf{Y}}^n\}) \\ &= \frac{1}{N} \sum_{\mathbf{x}_i \in \mathcal{T}_{X|\bar{s}}^n(\bar{s})} \sum_{\mathbf{y} \in \mathcal{T}_{Y|X\bar{s}}^n(\mathbf{x}_i, \bar{s})} W^n(\mathbf{y}|\mathbf{x}_i, \bar{s}) \end{aligned}$$

$$\begin{aligned} &\leq \frac{1}{N} \sum_{\mathbf{x}_i \in \mathcal{T}_{X|\bar{s}}^n(\bar{s})} \exp\{-nD(P_{X\bar{s}\mathbf{Y}}||P_{X\bar{s}} \times \bar{W})\} \\ &\leq \exp(-n\eta_2). \end{aligned}$$

Thus,

$$\begin{aligned} P_{\mathbf{X}\mathbf{Y}}(\mathcal{E}_2) &\leq \sum_{\substack{P_{X\bar{s}\mathbf{Y}} \in \mathcal{P}_{X \times \bar{s} \times \mathcal{Y}}^n \\ D(P_{X\bar{s}\mathbf{Y}}||P_{X\bar{s}} \times \bar{W}) \geq \eta_2}} \exp(-n\eta_2) \\ &\leq \exp(-n(\eta_2 - \epsilon)). \end{aligned} \quad (35)$$

In order to evaluate the probability of  $\mathcal{E}_3$ , let  $\mathcal{P} \subseteq \mathcal{P}_{X \times \bar{s} \times \mathcal{Y} \times X' \times \mathcal{S}}^n$  be such that for each  $P_{X\bar{s}\mathbf{Y}X'} \in \mathcal{P}$  we have  $I(X; \bar{S}) < \eta_1, D(P_{X\bar{s}\mathbf{Y}}||P_{X\bar{s}} \times \bar{W}) < \eta_2, I(X'; XY|\bar{S}) \geq \eta_3$  and for some  $S$  distributed over  $\mathcal{S}$ ,  $I(X'; S) < \delta, D(P_{X'\mathbf{s}\mathbf{Y}}||P_{X'\mathbf{s}}W) < \delta$ .

$$\begin{aligned} P_{\mathbf{X}\mathbf{Y}}(\mathcal{E}_3) &\leq \sum_{P_{X\bar{s}\mathbf{Y}X'} \in \mathcal{P}} \frac{1}{N} \sum_{i: (\mathbf{x}_i, \mathbf{x}_j, \bar{s}) \in \mathcal{T}_{X'X'\bar{s}}^n \text{ for some } j \neq i} \sum_{\mathbf{y} \in \mathcal{T}_{Y|X'X'\bar{s}}^n(\mathbf{x}_j, \mathbf{x}_i, \bar{s})} W^n(\mathbf{y}|\mathbf{x}_i, \bar{s}) \\ &\leq \sum_{P_{X\bar{s}\mathbf{Y}X'} \in \mathcal{P}} \frac{1}{N} |i : (\mathbf{x}_i, \mathbf{x}_j, \bar{s}) \in \mathcal{T}_{X'X'\bar{s}}^n \text{ for some } j \neq i| \\ &\stackrel{(a)}{\leq} \sum_{P_{X\bar{s}\mathbf{Y}X'} \in \mathcal{P}} \exp\left\{n \left(|R - I(X'; \bar{S})|^+ - I(X; X'\bar{S}) + \epsilon/2\right)\right\} \\ &\stackrel{(b)}{\leq} \exp\left\{-n \left(I(X; X'\bar{S}) - |R - I(X'; \bar{S})|^+ - \epsilon\right)\right\} \end{aligned} \quad (36)$$

where (a) follows from (33) and (b) holds for large  $n$ . (36) is also upper bounded by

$$\begin{aligned} &\sum_{P_{X\bar{s}\mathbf{Y}X'} \in \mathcal{P}} \frac{1}{N} \sum_{\mathbf{x}_i: \mathbf{x}_i \in \mathcal{T}_{X|\bar{s}}^n(\bar{s})} \sum_{\mathbf{x}_j \in \mathcal{T}_{X'|\bar{s}}^n(\mathbf{x}_i, \bar{s})} \sum_{\mathbf{y} \in \mathcal{T}_{Y|X'X'\bar{s}}^n(\mathbf{x}_j, \mathbf{x}_i, \bar{s})} W^n(\mathbf{y}|\mathbf{x}_i, \bar{s}) \\ &\stackrel{(a)}{\leq} \sum_{P_{X\bar{s}\mathbf{Y}X'} \in \mathcal{P}} \frac{1}{N} \sum_{\substack{\mathbf{x}_i: \\ \mathbf{x}_i \in \mathcal{T}_{X|\bar{s}}^n(\bar{s})}} \exp\left\{n \left(|R - I(X'; X\bar{S})|^+ + \epsilon\right)\right\} \\ &\quad \exp\{-nI(X'; Y|X\bar{S})\} \\ &\stackrel{(b)}{\leq} \exp\left\{-n \left(I(X'; Y|X\bar{S}) - |R - I(X'; X\bar{S})|^+ - 2\epsilon\right)\right\} \end{aligned} \quad (37)$$

where (a) follows from (31) and  $\sum_{\mathbf{y} \in \mathcal{T}_{Y|X'X'\bar{s}}^n(\mathbf{x}_j, \mathbf{x}_i, \bar{s})} W^n(\mathbf{y}|\mathbf{x}_i, \bar{s}) \leq \exp(-nI(X'; Y|X\bar{S}))$  and (b) holds for large  $n$ . The exponent in (28) follows from (34), (35), (37) and (38).

Next, we show the exponent in Theorem 6.

For  $R \geq I(X; \bar{S})$ ,  $A_1 = I(X; \bar{S}) - \epsilon \geq \eta_1 - \epsilon$ . When  $R < I(X; \bar{S})$ ,  $A_1 = R - \epsilon$ . Next, we evaluate  $A_2$ . When  $I(X; X'\bar{S}) - |R - I(X'; \bar{S})|^+ - \epsilon \geq t$  for some  $t$  (TBD),  $A_2 \geq t$ . Otherwise, when  $I(X; X'\bar{S}) - |R - I(X'; \bar{S})|^+ \leq$

$\epsilon + t$ , we consider two cases. When  $R \leq I(X'; \bar{S})$ , we have  $I(X; X'|\bar{S}) \leq I(X; X'\bar{S}) \leq \epsilon + t$ . Thus,

$$\begin{aligned} & I(Y; X'|X\bar{S}) - |R - I(X'; X\bar{S})|^+ - 2\epsilon \\ &= I(Y; X'|X\bar{S}) - 2\epsilon \\ &= I(YX; X'|\bar{S}) - I(X; X'|\bar{S}) - 2\epsilon \\ &\geq \eta_3 - t - 3\epsilon \text{ because } I(YX; X'|\bar{S}) > \eta_3. \end{aligned}$$

Thus,  $A_2 \geq \eta_3 - t - 3\epsilon$  in this case. When  $R > I(X'; S)$ ,

$$\begin{aligned} R &\geq I(X; X'\bar{S}) + I(X'; \bar{S}) - \epsilon - t \\ &\geq I(X'; XS) - \epsilon - t. \end{aligned}$$

This implies that  $|R - I(X'; X\bar{S})|^+ \leq R - I(X'; XS) + \epsilon + t$ . In this case,

$$\begin{aligned} & I(Y; X'|X\bar{S}) - |R - I(X'; X\bar{S})|^+ - 2\epsilon \\ &\geq I(Y; X'|X\bar{S}) - R + I(X'; XS) - \epsilon - t - 2\epsilon \\ &= I(X\bar{S}Y; X') - R - t - 3\epsilon \\ &= I(XY; X'|\bar{S}) + I(X'; \bar{S}) - R - t - 3\epsilon \\ &\geq \eta_3 - R - t - 3\epsilon. \end{aligned}$$

With this, the exponent  $\gamma$

$$\begin{aligned} \gamma &\geq \min \left\{ \min \{ \eta_1 - \epsilon, \eta_3/3 - \epsilon \}, \eta_2 - \epsilon, \right. \\ &\quad \left. \max \{ t, \min \{ \eta_3 - t - \epsilon/4, \eta_3 - R - t - 3\epsilon \} \} \right\}. \end{aligned}$$

For  $R = t = \eta_3/3$  and  $\epsilon \rightarrow 0$  (note that  $\epsilon > 0$  may be arbitrarily small as long as  $R \geq \epsilon$  as required by Lemma 2), the exponent  $\gamma$  can be made arbitrarily close to

$$\min \{ \eta_1, \eta_2, \eta_3/3 \}. \quad (39)$$

Next, we will show under Hypothesis  $H_0$  too, the probability of error is arbitrarily small. Suppose the adversary's attack is  $\mathbf{s} \in \mathcal{S}^n$ . For each  $\mathbf{x}_j \in \mathcal{C}(P)$  and  $\mathbf{y} \in \mathcal{Y}^n$ , let  $P_{\mathbf{X}'\mathbf{Y}}(\mathbf{x}_j, \mathbf{y}) = \frac{1}{N} W^n(\mathbf{y}|\mathbf{x}_j, \mathbf{s})$ . Let  $(\mathbf{X}', \mathbf{Y}) \sim P_{\mathbf{X}'\mathbf{Y}}$ . Define  $\tilde{\mathcal{E}}_1 := \{(\mathbf{X}', \mathbf{s}) \in \mathcal{T}_{X'S}^n \text{ such that } I(X'; S) \geq \delta\}$ ,  $\tilde{\mathcal{E}}_2 := \{(\mathbf{X}', \mathbf{s}, \mathbf{Y}) \in \mathcal{T}_{X'SY}^n \text{ such that } D(P_{X'SY}||P_{X'S} \times W) \geq \delta\}$ ,  $\tilde{\mathcal{E}}_3 := \{(\mathbf{X}', \mathbf{s}, \mathbf{Y}) \in \mathcal{T}_{X'SY}^n \text{ such that } I(X'; S) < \delta, D(P_{X'SY}||P_{X'S} \times W) < \delta, \exists \mathbf{x}_i \neq \mathbf{X}' \text{ such that } (\mathbf{x}_i, \bar{\mathbf{s}}, \mathbf{Y}) \in \mathcal{T}_{X\bar{S}Y}^n \text{ for some } \bar{\mathbf{s}} \in \bar{\mathcal{S}}^n \text{ for which } I(X; \bar{S}) < \eta_1 \text{ and } D(P_{X\bar{S}Y}||P_{X\bar{S}}\bar{W}) < \eta_2, \text{ but } I(X; X'Y|S) \geq \delta\}$ , and  $\tilde{\mathcal{E}}_4 := \{\exists \bar{\mathbf{s}} \in \bar{\mathcal{S}}^n \text{ such that } (\mathbf{X}', \bar{\mathbf{s}}, \mathbf{s}, \mathbf{Y}) \in \mathcal{T}_{X'\bar{S}SY}^n \text{ for which } I(X; \bar{S}) < \eta_1, D(P_{X\bar{S}Y}||P_{X\bar{S}} \times \bar{W}) < \eta_2, I(X'; S) < \delta \text{ and } D(P_{X'SY}||P_{X'S}W) < \delta\}$ .

These events are analogous to the events  $\mathcal{E}_1, \mathcal{E}_2, \mathcal{E}_3$  and  $\mathcal{E}_4$  defined under  $H_1$ , except that  $(\eta_1, \eta_2, \eta_3)$  is exchanged with  $(\delta, \delta, \delta)$ . Following a similar line of argument, one can show that  $P_{\mathbf{X}'\mathbf{Y}}(\tilde{\mathcal{E}}_1 \cup \tilde{\mathcal{E}}_2 \cup \tilde{\mathcal{E}}_3 \cup \tilde{\mathcal{E}}_4) \leq \exp(-n\delta/3)$  (see (39)).

We will next argue that conditioned on the event  $\tilde{\mathcal{E}}_1^c \cap \tilde{\mathcal{E}}_2^c \cap \tilde{\mathcal{E}}_3^c \cap \tilde{\mathcal{E}}_4^c$ , the detector will not output  $H_1$ . This is because Definition 2 ensures that for  $(\eta_1, \eta_2, \eta_3) \in \eta(P)$  and  $\delta$  given by definition 2,

- There does not exist  $\mathbf{x}_i, \bar{\mathbf{s}} \in \bar{\mathcal{S}}^n$  and such that for  $(\mathbf{x}_i, \mathbf{X}', \bar{\mathbf{s}}, \mathbf{s}, \mathbf{Y}) \in \mathcal{T}_{X\bar{S}SY}^n$ ,  $I(X; \bar{S}) < \eta_1$ ,  $D(P_{X\bar{S}Y}||P_{X\bar{S}}\bar{W}) < \eta_2$ ,  $I(X'; S) < \delta$ ,

$$D(P_{X'SY}||P_{X'S}W) < \delta, \text{ and for } X \neq X', I(X'; XY|\bar{S}) < \eta_3 \text{ and } I(X; X'Y|S) < \delta.$$

This implies that the error will happen only under  $\tilde{\mathcal{E}}_1 \cup \tilde{\mathcal{E}}_2 \cup \tilde{\mathcal{E}}_3 \cup \tilde{\mathcal{E}}_4$  which happens with probability at most  $\exp(-n\delta)$ . This can be made arbitrarily small for large  $n$ .  $\square$

*Proof of Lemma 2.* The proof of the lemma follows from the proof of [17, Lemma 3]. (31) is the same as [17, eq. (3.1)]. (32) can be obtained from the proof of [17, eq. (3.2)], specifically by replacing  $P_{X'S}$  with  $P_{XS}$  and  $\epsilon$  with  $\epsilon/2$  in [17, eq. (A8)]. Equation (33) is obtained from the proof of [17, eq. (3.3)], where for  $a = (n+1)^{|\mathcal{X}|} \exp\left\{n\left(|R - I(X'; S)|^+ - I(X; X'S) + \epsilon/4\right)\right\}$ , we choose  $t = \exp\left\{n\left(|R - I(X'; S)|^+ - I(X; X'S) + \epsilon/2\right)\right\}$ . Note that for large enough  $n$ ,  $t > a \log e$  as required by [17, eq. (A2)].  $\square$