EE 492

B.TECH. PROJECT

COMMUNICATION & COMPOUND-STATE IDENTIFICATION IN COMPOUND ARBITRARILY VARYING CHANNELS

9 May 2021

Syomantak Chaudhuri

Roll number 170070004

syomantak@iitb.ac.in

With

Neha Sangwan, TIFR

Mayank Bakshi, HONG KONG

Advised By

Prof. Bikash Dey, IIT BOMABY

Prof. Vinod Prabhakaran, TIFR

Electrical Engineering Department

Indian Institute of Technology, Bombay

Contents

Abstract

We propose a communication model, that we call compound arbitrarily varying channels (CAVC), which unifies and generalizes compound channels and arbitrarily varying channels (AVC). A CAVC can be viewed as a noisy channel with a fixed, but unknown, compound-state and an AVC-state which may vary with every channel use. The AVC-state is controlled by an adversary who is aware of the compound-state. We study four problems in this setting: 'communication', 'compound-state identification', 'communication and compound-state identification', and 'communication or compound-state identification'. For these problems, we study conditions for feasibility and capacity under deterministic coding and random coding. In addition, we analyze the best possible error exponent which can be achieved for the problem of 'compound-state identification' along with an analysis of the trade-off in error exponent for the problem of 'communication and compound-state identification'.

Keywords— compound arbitrarily varying channel, compound-state identification, adversary identification, multiple-adversaries, compound channel, arbitrarily varying channel, error exponent, hypothesis testing

1 INTRODUCTION

In communication systems modeled as discrete memoryless channels (DMC), it is assumed that the channel characteristics is fixed and known beforehand. However, the compound DMC introduced by Blackwell et al. [1] models channels with fixed but unknown characteristics due to an unknown natural state. Backwell et al. [2] also introduced arbitrarily varying channels (AVC) where the channel state may vary arbitrarily in a worst case manner for each symbol of transmission. The worst case variation of the channel state in an AVC may be viewed as the act of a malicious adversary.

The capacity of a compound DMC was characterized in [3]. For AVC, the communication capacity under random coding was obtained in [2]. The deterministic coding capacity of an AVC is zero if the channel satisfies a condition called *symmetrizability* which allows the adversary to mount an attack with a spurious message so as to confuse the decoder between this message and the sent message. When the channel is not symmetrizable, the deterministic coding capacity is the same as the random coding capacity [4].

In this work, we consider a generalization where there is an unknown compound-state as well as an AVC-state determined by an adversary (see Figure 1). The compound-state is fixed over a blocklength of transmission, whereas the AVC-state may change for every symbol of transmission. We assume that the adversary knows the compound-state. Associated with each compound-state, the adversary has a set of channels that can be be instantiated (by setting the AVC-state). We call this the Compound Arbitrarily Varying Channel (CAVC). This is a generalization of both compound channels and AVCs. For simplicity, in this paper we only consider the case of two compound-states.

We characterize the capacity of CAVCs under both random coding and deterministic coding. For non-zero rates to be achievable under deterministic coding, first, the AVC under each compound-state should be non-symmetrizable. In addition, the channel should not satisfy a new condition, called *trans-symmetrizability*, which provides the adversary with an attack strategy that can confuse the decoder between the sent message under one compound-state with another message under the other compound-state (see Fig. 3). We show that when a CAVC is not symmetrizable in either of these senses, the deterministic coding capacity is same as the random coding capacity.

Another way to view the CAVC model is to associate an adversary with each compound-state and exactly one of them being active for the entirety of the transmission. Associated with each adversary, there is a family of channels

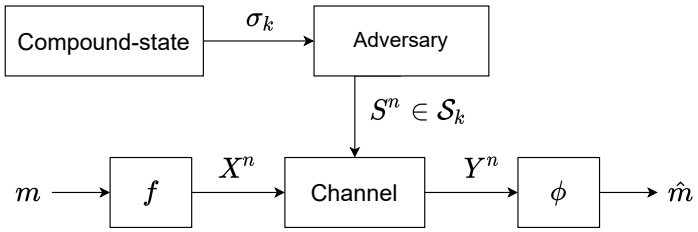

Figure 1: Compound Arbitrarily Varying Channel: The adversary knows the compound-state σ_k and for each compound-state, the adversary has a set of AVC-states $\mathcal{S}_k$. The CAVC is modeled to be discrete memoryless and the compound-state remains fixed through out the transmission of a block.

from which it can instantiate a channel for each channel use. In such a situation, it is also of interest to identify[1] the active adversary - we call this problem 'compound-stateidentification' and we give the necessary conditions to be able to identify the adversary (or equivalently, the compound-state). In addition to these two problems, we also study two other problems in the CAVC setup – joint 'communication *and* compound-state identification' and 'communication *or* compound-state identification'. In the first (resp. second) problem above, the decoder needs to decode the message and (resp. or) identify the compound-state. In both these settings, we characterize the condition for non-zero rates under deterministic codes and also the capacities under deterministic coding and random coding.

If the compound-state was known to the decoder, the CAVC model would be a special case of arbitrarily varying broadcast channels [6, 7, 8]. The trans-symmetrizability condition for non-zero rates in a CAVC arises precisely because the decoder does not know the compound-state. In [9, 10, 11], on authentication in channels which may be controlled by an adversary, a relaxed decoding requirement is considered. When there is no adversary, the decoded message must be correct; but when the adversary is active, the decoder is allowed to declare the presence of the adversary without decoding the message (however, if the decoder outputs a message instead, it must be correct). These models are close to our 'communication *or* identification' model. In fact, we recover the result in [9] as a special case (see Remark 1). The work in [12] considers communication in a Compound-Arbitrarily-Varying network where the adversary selects a subset of edges from a network which are then attacked with arbitrary transmissions.

In addition to characterizing the capacity regions for the four problem settings, we also analyze the error exponents for two related problems. The problem of obtaining the optimal error exponent for communication is a well-known problem and it remains elusive for various class for channels. [13] obtained an achievable error exponent and an upper bound for a general class of channels under random coding. [14] obtained a bound for the error exponent for AVCs. In this work, we instead focus on the error exponent for the task of 'compound-state identification' and the trade-off in error exponent of compound-state identification versus the rate of communication in the 'communication and compound-state identification' problem, i.e., we try to obtain the best error exponent for compound-state identification at any given rate of communication - it can be seen that this problem has similarity with hypothesis testing and we indeed use some of the existing theory of hypothesis Testing like Neyman-Pearson [15] lemma in our analysis. However, these problems are significantly more difficult than simple hypothesis testing. Both these problems are analyzed only under random coding and we are yet to give a tight upper and lower bound for these.

In Section 2, we formally describe the CAVC model and present the problems studied in this paper. We present our results on the four problems in Sections 3.1, 3.2, 3.3, and 3.4. Our analysis of error exponents is presented in 4. Section 6 provides brief proof sketches for the results and the full proofs are given in Section 7.

[1]Note that this is significantly different from identifying an *internal* adversary in a multiuser channel with byzantine users [5].

2 SYSTEM MODEL

Table 1: A brief summary of the problems studied and the results presented in this work.

Task	Output set $\widehat{\mathcal{M}}$	Error set $\mathcal{E}_{m,k}$	Conditions for positive deterministic capacity	Capacity expression
Communication	$\mathcal{M}$	$\{m' \in \widehat{\mathcal{M}} : m' \neq m\}$	Non-any-sym.	$\max_{P_X} \min_{W \in \overline{\mathcal{W}}_1 \cup \overline{\mathcal{W}}_2} I(X;Y)$
Compound-State Identification	$\{\sigma_1, \sigma_2\}$	$\{\sigma_{3-k}\}$	Non-trans-sym. $\overline{\mathcal{W}}_1 \cap \overline{\mathcal{W}}_2 = \emptyset$	-
Communication and Compound-State Identification	$\mathcal{M} \times \{\sigma_1, \sigma_2\}$	$\{m' \in \widehat{\mathcal{M}} : m' \neq (m, \sigma_i)\}$	Non-any-sym. $\overline{\mathcal{W}}_1 \cap \overline{\mathcal{W}}_2 = \emptyset$	$\max_{P_X} \min_{W \in \overline{\mathcal{W}}_1 \cup \overline{\mathcal{W}}_2} I(X;Y)$
Communication or Compound-State Identification	$\mathcal{M} \cup \{\sigma_1, \sigma_2\}$	$\{m' \in \widehat{\mathcal{M}} : m' \notin \{m, \sigma_i\}\}$	Non-trans-sym.	$\max_{P_X} \min_{W \in \overline{\mathcal{W}}_1 \cap \overline{\mathcal{W}}_2} I(X;Y)$

Notation: We use bold symbols like $\boldsymbol{x}$,$\boldsymbol{y}$ to denote vectors and capital letters like X,Y to denote random variables with P_X, P_Y denoting their distributions respectively. The i-th element of a vector $\boldsymbol{y}$ is denoted as y_i. For a vector $\boldsymbol{x}$, the notation $P_{\boldsymbol{x}}$ refers to its empirical distribution. For any subset $\mathcal{B}$ in a finite dimensional space $\mathbb{R}^k$, its convex closure is denoted by $\bar{\mathcal{B}}$.

A discrete-memoryless Compound Arbitrarily Varying Channel (CAVC) with a finite input alphabet $\mathcal{X}$, a finite output alphabet $\mathcal{Y}$, and two compound-states σ_1 and σ_2 is described by two families, $\mathcal{W}_1$ and $\mathcal{W}_2$, of channels with input alphabet $\mathcal{X}$ and output alphabet $\mathcal{Y}$. These families of channels correspond to the compound-states σ_1 and σ_2 respectively. In each family, the channels are indexed by a finite set $\mathcal{S}_k$ ($k = 1, 2$) called the AVC-state alphabet and, in particular, $\mathcal{W}_k$ ($k = 1, 2$) is a set of channels $\{W(\cdot|\cdot, s),\ s \in \mathcal{S}_k\}$. On input $\boldsymbol{x} \in \mathcal{X}^n$ over n uses of the channel, $n \in \{1, 2, \ldots\}$, the probability of receiving $\boldsymbol{y} \in \mathcal{Y}^n$ is given by $W^n(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{s}) = \prod_{i=1}^n W(y_i|x_i, s_i)$ for some $\boldsymbol{s} \in \mathcal{S}_1^n \cup \mathcal{S}_2^n$.

We study the CAVC under four distinct but closely-related problem settings as specified at the end of this section. In all four problems, the CAVC is analyzed under both deterministic and random (shared-randomness between encoder and decoder unknown to the adversary) coding regimes. An (M, n) deterministic code is characterized by

1. a message set $\mathcal{M} = \{1, \ldots, M\}$,
2. an encoder $f : \mathcal{M} \to \mathcal{X}^n$, and
3. a decoder $\phi : \mathcal{Y}^n \to \widehat{\mathcal{M}}$.

The set $\widehat{\mathcal{M}}$ is different for the three problems, and is described later in this section. Table 1 gives a short description of each problem and the results we present. The problems are studied under the average probability of error and it is assumed that the adversary is unaware of the message sent by the transmitter but is aware of the encoder and decoder pair (f, ϕ) used for transmission.

Let $\mathcal{E}_{m,k} \subseteq \widehat{\mathcal{M}}$ correspond to the set of erroneous outputs from the decoder when message m is sent and σ_k is the compound-state. $\mathcal{E}_{m,k}$ depends on the problem definition and we specify it at the end of this section for each problem. For $k = 1, 2$, define

$$P_{\mathsf{e}}^{\mathsf{d}}(f, \phi, k) = \max_{\boldsymbol{s} \in \mathcal{S}_k^n} \frac{1}{M} \sum_{m=1}^{M} W^n(\phi^{-1}(\mathcal{E}_{m,k})|f(m), \boldsymbol{s}). \tag{2.1}$$

The average probability of error $P_{\mathsf{e}}^{\mathsf{d}}(f, \phi)$ is given by

$$P_{\mathsf{e}}^{\mathsf{d}}(f, \phi) = \max\{P_{\mathsf{e}}^{\mathsf{d}}(f, \phi, 1), P_{\mathsf{e}}^{\mathsf{d}}(f, \phi, 2)\}. \tag{2.2}$$

A rate R is defined to be achievable under deterministic coding if there exists a sequence of $(2^{nR}, n)$ deterministic codes $\{f^{(n)}, \phi^{(n)}\}_{n=1}^{\infty}$ such that $P_{\mathsf{e}}^{\mathsf{d}}(f^{(n)}, \phi^{(n)}) \to 0$ as $n \to \infty$. The *deterministic code capacity* is defined as the supremum of all achievable rates under deterministic coding.

Let $\mathcal{F}$ be the set of all encoders $f : \mathcal{M} \to \mathcal{X}^n$ and $\mathcal{G}$ be the set of all decoders $\phi : \mathcal{Y}^n \to \widehat{\mathcal{M}}$. An (M, n) random code is given by the pair $(F, \Phi) \sim Q(f, \phi)$ where Q is a distribution on $\mathcal{F} \times \mathcal{G}$. The adversary has the knowledge of the distribution Q but does not know the realisation of (F, Φ) used during the transmission and it is unaware of the transmitted message as well. For $k = 1, 2$, define

$$P_{\mathsf{e}}^{\mathsf{r}}(Q, k) = \max_{s \in \mathcal{S}_k^n} \sum_{(f,\phi) \in \mathcal{F} \times \mathcal{G}} Q(f, \phi) \frac{1}{M} \sum_{m=1}^{M} W^n(\phi^{-1}(\mathcal{E}_{m,k}) | f(m), s). \tag{2.3}$$

The average probability of error $P_{\mathsf{e}}^{\mathsf{r}}(Q)$ for a random code is given by

$$P_{\mathsf{e}}^{\mathsf{r}}(Q) = \max\{P_{\mathsf{e}}^{\mathsf{r}}(Q, 1), P_{\mathsf{e}}^{\mathsf{r}}(Q, 2).\} \tag{2.4}$$

A rate R is defined to be achievable under random coding if there exists a sequence of $(2^{nR}, n)$ random codes $\{Q^{(n)}\}_{n=1}^{\infty}$ such that $P_{\mathsf{e}}^{\mathsf{r}}(Q^{(n)}) \to 0$ as $n \to \infty$. The *random code capacity* is defined as the supremum of all achievable rates under random coding.

We now define the four specific problems.

Communication over CAVC In this problem, the decoder needs to reconstruct the encoded message. Therefore, the decoder's reconstruction alphabet is $\hat{\mathcal{M}} = \mathcal{M}$ and the set $\mathcal{E}_{m,k}$ of erroneous decoder outputs is given by

$$\mathcal{E}_{m,k} = \{m' \in \widehat{\mathcal{M}} : m' \neq m\}.$$

Compound-State Identification over CAVC The decoder needs to identify the compound-state which was active during the transmission. Therefore, the decoder's reconstruction alphabet is $\hat{\mathcal{M}} = \{\sigma_1, \sigma_2\}$ and the set $\mathcal{E}_{m,k}$ of erroneous decoder outputs is given by

$$\mathcal{E}_{m,k} = \{\sigma_{3-k}\}.$$

Joint Communication and Compound-State Identification over CAVC Here the decoder needs to reconstruct the encoded message, and also identify the compound-state. Hence $\widehat{\mathcal{M}} = \mathcal{M} \times \{\sigma_1, \sigma_2\}$ and the set $\mathcal{E}_{m,k}$ is given by

$$\mathcal{E}_{m,k} = \{m' \in \widehat{\mathcal{M}} : m' \neq (m, \sigma_k)\}.$$

Communication or Compound-State Identification over CAVC Here the decoder needs to either reconstruct the encoded message or identify the compound-state. Hence $\widehat{\mathcal{M}} = \mathcal{M} \cup \{\sigma_1, \sigma_2\}$ and the set $\mathcal{E}_{m,k}$ is given by

$$\mathcal{E}_{m,k} = \{m' \in \widehat{\mathcal{M}} : m' \notin \{m, \sigma_k\}\}.$$

For the problem of joint communication and compound-state identification, the individual probabilities of error in communication and compound-state identification are required for analyzing the error exponent of compound-state identification separately. These can be obtained breaking the error into two terms. Formally, let $\mathcal{H}_m = \{m' \in \mathcal{M} : m' \neq m\}$, $\mathcal{I}_k = \{\sigma_{3-k}\}$, $\hat{\mathcal{H}}_m = \mathcal{H}_m \times \{\sigma_1, \sigma_2\}$, and $\hat{\mathcal{I}}_k = (\bigcup_n \mathcal{H}_n) \times \mathcal{I}_k$. Then note that $\mathcal{E}_{m,k} = \hat{\mathcal{H}}_m \cup \hat{\mathcal{I}}_k$. Here, the set $\hat{\mathcal{H}}_m$ corresponds to outputs where the message is decoded incorrectly while the compound-state may or may not be

identified correctly. Similarly, $\hat{\mathcal{I}}_k$ corresponds to the case where the compound-state is identified incorrectly. Thus, for $k = 1, 2$ under random coding, define

$$P_{\mathsf{c}}^{\mathsf{r}}(Q,k) = \max_{\boldsymbol{s}\in\mathcal{S}_k^n} \sum_{(f,\phi)\in\mathcal{F}\times\mathcal{G}} Q(f,\phi)\frac{1}{M}\sum_{m=1}^{M} W^n(\Phi^{-1}(\hat{\mathcal{H}}_m)|f(m),\boldsymbol{s}). \tag{2.5}$$

$$P_{\mathsf{i}}^{\mathsf{r}}(Q,k) = \max_{\boldsymbol{s}\in\mathcal{S}_k^n} \sum_{(f,\phi)\in\mathcal{F}\times\mathcal{G}} Q(f,\phi)\frac{1}{M}\sum_{m=1}^{M} W^n(\Phi^{-1}(\hat{\mathcal{I}}_k)|f(m),\boldsymbol{s}).. \tag{2.6}$$

and the average probabilities of error is given by

$$P_{\mathsf{c}}^{\mathsf{r}}(Q) = \max\{P_{\mathsf{c}}^{\mathsf{r}}(Q,1), P_{\mathsf{c}}^{\mathsf{r}}(Q,2)\} \tag{2.7}$$

$$P_{\mathsf{i}}^{\mathsf{r}}(Q) = \max\{P_{\mathsf{i}}^{\mathsf{r}}(Q,1), P_{\mathsf{i}}^{\mathsf{r}}(Q,2)\} \tag{2.8}$$

Note that $P_{\mathsf{i}}^{\mathsf{r}}(Q), P_{\mathsf{c}}^{\mathsf{r}}(Q) \leq P_{\mathsf{e}}^{\mathsf{r}}(Q) \leq P_{\mathsf{i}}^{\mathsf{r}}(Q) + P_{\mathsf{c}}^{\mathsf{r}}(Q)$.

In the task of joint communication and compound-state identification, the tuple (R,γ) - rate R and error exponent γ for compound-state identification - is said to be achievable if there exists a sequence of $(2^{nR}, n)$ random codes $\{Q^{(n)}\}_{n=1}^{\infty}$ such that the rate R is achievable and $\liminf_{n\to\infty} -\frac{\log(P_{\mathsf{i}}^{\mathsf{r}}(Q^{(n)}))}{n} = \gamma$. The *rate-exponent region* $\mathcal{R}$ is defined as the closure of all achievable tuples (R,γ). Note that the above definition is defined for random coding only, we do not focus on deterministic coding.

For the task of compound-state identification, we can similarly define the error exponent. In particular, an error exponent γ is said to be achievable if there exists a sequence of $(2^{nR}, n)$ random codes $\{Q^{(n)}\}_{n=1}^{\infty}$ such that $\liminf_{n\to\infty} -\frac{\log(P_{\mathsf{e}}^{\mathsf{r}}(Q^{(n)}))}{n} = \gamma$. The *optimal exponent* G is defined as the supremum of all achievable error exponents. It is immediately clear that $G = \max\{\gamma : (R,\gamma) \in \mathcal{R} \text{ for some } R\}$.

3 MAIN RESULTS

We now present the main results on the four problems in four respective subsections.

3.1 Communication over CAVC

We denote the CAVC capacity for the communication problem under deterministic coding as $C_{\mathsf{com}}^{\mathsf{d}}$ and that under randomized coding as $C_{\mathsf{com}}^{\mathsf{r}}$.

Communication over a CAVC is closely related to communication over an Arbitrarily Varying Channel (AVC). An AVC from $\mathcal{X}$ to $\mathcal{Y}$ is given by a set of channels $\{W(\cdot|\cdot,s), s\in\mathcal{S}\}$ parameterized by the state alphabet $\mathcal{S}$. The AVC-state of the channel can change arbitrarily during the transmission. A CAVC is an AVC when $\mathcal{S}_1 = \mathcal{S}_2$. Csiszar and Narayan in [4] defined the notion of a *symmetrizable* AVC and showed that the deterministic coding capacity of an AVC, $C_{\mathsf{AVC}}^{\mathsf{d}}$, is positive if and only if the channel is not symmetrizable. An AVC is symmetrizable if there exists some channel $U : \mathcal{X} \to \mathcal{S}$ such that $\forall x, x' \in \mathcal{X}, y \in \mathcal{Y}$,

$$\sum_s U(s|x')W(y|x,s) = \sum_s U(s|x)W(y|x',s). \tag{3.1}$$

Cis-symmetrizability: For a CAVC, symmetrizability can be defined under each compound-state. For $k = 1$ or 2, we define a CAVC to be $\mathcal{S}_k$-*symmetrizable* if there exists a channel $U : \mathcal{X} \to \mathcal{S}_k$ such that (3.1) holds $\forall x, x' \in \mathcal{X}, y \in \mathcal{Y}$

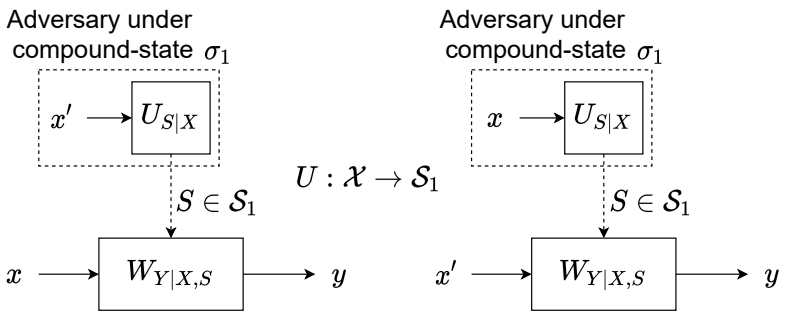

Figure 2: $\mathcal{S}_1$-symmetrizability: If there exists a channel $U : \mathcal{S}_1 \to \mathcal{X}$ such that the output distributions in the above two scenarios are the same for every pair of symbols $(x, x') \in \mathcal{X}^2$ then we call the channel $\mathcal{S}_1$-symmetrizable.

(see Figure 2). If the CAVC is $\mathcal{S}_k$-symmetrizable for $k = 1$ or $k = 2$ or both, then we say the CAVC is *cis-symmetrizable*.

If the channel is $\mathcal{S}_k$-symmetrizable and the compound-state is σ_k, then for two distinct codewords $\boldsymbol{x}_m$, $\boldsymbol{x}_{m'}$ and U satisfying (3.1), the following two situations are indistinguishable : (i) the sender sends $\boldsymbol{x}_m$ and the adversary attacks when the compound-state is σ_k with AVC-state sequence from the output of the distribution $U^n(\cdot|\boldsymbol{x}_{m'})$ and (ii) the sender sends $\boldsymbol{x}_{m'}$ and the adversary attacks when the compound-state is σ_k with the output of the distribution $U^n(\cdot|\boldsymbol{x}_m)$. Thus, this argument is formalized in Section 7 and it is possible to show that reliable decoding is not possible if a CAVC is cis-symmetrizable.

Trans-symmetrizability: The presence of two compound-states in a CAVC introduces another sufficient condition for $C_{\mathsf{com}}^{\mathsf{d}} = 0$ which we call trans-symmetrizability (see Figure 3). Define a CAVC to be *trans-symmetrizable* if there exists a pair of channels $U : \mathcal{X} \to \mathcal{S}_1$, $V : \mathcal{X} \to \mathcal{S}_2$ such that $\forall x, x' \in \mathcal{X}, y \in \mathcal{Y}$,

$$\sum_s U(s|x')W(y|x,s) = \sum_s V(s|x)W(y|x',s). \tag{3.2}$$

In a trans-symmetrizable CAVC with U, V satisfying (3.2) and $\boldsymbol{x}_m$, $\boldsymbol{x}_{m'}$ being distinct codewords, the following two situations are indistinguishable: (i) the sender sends codeword $\boldsymbol{x}_m$ and the adversary attacks when the compound-state is σ_1 with the AVC-state sequence as the output of the distribution $U^n(\cdot|\boldsymbol{x}_{m'})$ and (ii) the sender sends codeword $\boldsymbol{x}_{m'}$ and the adversary attacks when the compound-state is σ_2 with the state sequence as the output of the distribution $V^n(\cdot|\boldsymbol{x}_m)$. Note that neither of cis- and trans-symmetrizability imply the other as demonstrated by the following two examples.

Consider a CAVC where $\mathcal{W}_1$ with output alphabet $\mathcal{Y}_1$ and $\mathcal{W}_2$ with output alphabet $\mathcal{Y}_2$ are symmetrizable AVCs satisfying $\mathcal{Y}_1 \cap \mathcal{Y}_2 = \emptyset$. Clearly, the CAVC is cis-symmetrizable but not trans-symmetrizable. Example 1 below presents a CAVC which is trans-symmetrizable, but not cis-symmetrizable.

Example 1. Consider a CAVC with input alphabet $\mathcal{X}$ and output alphabet $\mathcal{Y}$. Let $\mathcal{S}_k = \mathcal{X} \times \{k\}$. For $x \in \mathcal{X}$ and $(x', k) \in \mathcal{S}_k$,

$$y = \begin{cases} (x, x') & \text{if } k = 1, \\ (x', x) & \text{if } k = 2. \end{cases} \tag{3.3}$$

This CAVC is clearly trans-symmetrizable using $U(s|x') = 1$ if $s = (x', 1)$ and $V(s|x) = 1$ if $s = (x, 2)$. To show non-cis-symmetrizability, consider the case when the compound-state is σ_1 and the input symbol is x. Since the channel reveals the input and the AVC-state completely when the compound-state is σ_k, $k = 1, 2$, it cannot be cis-symmetrizable.

We call a CAVC *any-symmetrizable* if it is cis-symmetrizable or trans-symmetrizable (or both). Note that if a CAVC is any-symmetrizable then $C_{\mathsf{com}}^{\mathsf{d}} = 0$. Further, for a CAVC with $\mathcal{W}_k$ being the family of channels corresponding to

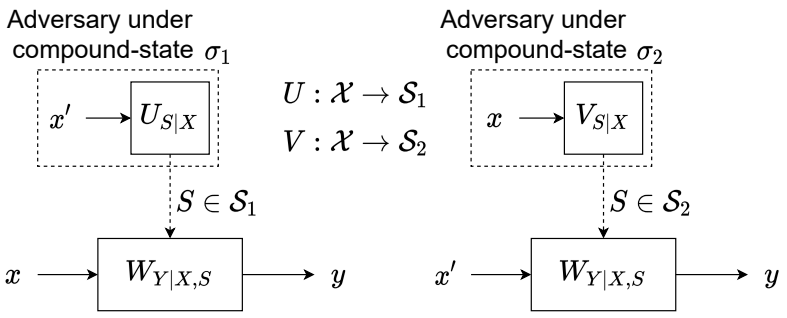

Figure 3: Trans-symmetrizability: If there exists a pair of channels $U : \mathcal{X} \to \mathcal{S}_1$, $V : \mathcal{X} \to \mathcal{S}_2$ such that the output distributions in the above two scenarios are the same for every pair of symbols $(x, x') \in \mathcal{X}^2$ then we call the channel trans-symmetrizable.

compound-state σ_k , the capacity of the AVC with the family of channels $\mathcal{W} = \mathcal{W}_1 \cup \mathcal{W}_2$ given by

$$C_{\mathsf{AVC}}^{\mathsf{d}} = \max_{P_X} \min_{W \in \overline{\mathcal{W}_1 \cup \mathcal{W}_2}} I(X;Y)$$

is a simple lower bound on $C_{\mathsf{com}}^{\mathsf{d}}$. Recall that $\overline{\mathcal{W}_1 \cup \mathcal{W}_2}$ refers to the convex closure of the family of channels $\mathcal{W}_1 \cup \mathcal{W}_2$. Using the compound nature of the channel, this bound can be improved. In particular, we show the following.

Theorem 1. (i) The random coding capacity for communication over CAVC is given by

$$C_{\mathsf{com}}^{\mathsf{r}} = \max_{P_X} \min_{W \in \overline{\mathcal{W}}_1 \cup \overline{\mathcal{W}}_2} I(X;Y). \tag{3.4}$$

(ii) The deterministic capacity $C_{\mathsf{com}}^{\mathsf{d}} > 0$ if and only if the CAVC is not any-symmetrizable. If $C_{\mathsf{com}}^{\mathsf{d}} > 0$, then

$$C_{\mathsf{com}}^{\mathsf{d}} = C_{\mathsf{com}}^{\mathsf{r}}. \tag{3.5}$$

Refer to Section 6 for proof sketches of Theorem 1.

3.2 Compound-State Identification over CAVC

For this problem, it is not meaningful to refer to the channel capacity since it will be either infinite - corresponding to the case where compound-state identification is possible - or 0 for the case when compound-state identification is not possible. Therefore, we can just consider the feasibility aspect in the problem.

Observe that if $\overline{\mathcal{W}}_1 \cap \overline{\mathcal{W}}_2 \neq \emptyset$, then there exists a channel $Z_{Y|X} \in \overline{\mathcal{W}}_1 \cap \overline{\mathcal{W}}_2$ which can be induced by either compound-state, so it is not possible to identify the compound-state in such situations - this is true even under random coding. Thus, $\overline{\mathcal{W}}_1 \cap \overline{\mathcal{W}}_2 = \emptyset$ is a necessary condition. Also, the trans-symmetrizability condition in (3.2) would hinder compound-state identification under deterministic coding as explained in Section 3.1. Thus, both $\overline{\mathcal{W}}_1 \cap \overline{\mathcal{W}}_2 = \emptyset$ and non-trans-symmetrizability is a necessary condition for compound-state identification under deterministic coding. We prove the theorem below formally in Section 7.

Theorem 2. (i) Compound-State identification is possible under random coding capacity if and only if $\overline{\mathcal{W}}_1 \cap \overline{\mathcal{W}}_2 = \emptyset$.
(ii) Compound-State identification is possible under deterministic coding capacity if and only if $\overline{\mathcal{W}}_1 \cap \overline{\mathcal{W}}_2 = \emptyset$ and the CAVC is not trans-symmetrizable.

3.3 Joint Communication and Compound-State Identification over CAVC

Let the deterministic capacity of the CAVC for the joint communication and compound-state identification be denoted by $C_{\mathsf{and}}^{\mathsf{d}}$ and let the random code capacity be denoted by $C_{\mathsf{and}}^{\mathsf{r}}$. Note that $C_{\mathsf{and}}^{\mathsf{d}} \leq C_{\mathsf{com}}^{\mathsf{d}}$ as an additional constraint has been imposed in this problem. From Theorem 1, it is clear that non-any-symmetrizability is required for joint communication and compound-state identification. Additionally, from Theorem 2, we can infer that $\overline{\mathcal{W}}_1 \cap \overline{\mathcal{W}}_2 = \emptyset$ is a necessary condition for joint communication and compound-state identification.

Note that any-symmetrizability and non-emptiness of $\overline{\mathcal{W}}_1 \cap \overline{\mathcal{W}}_2$ are not implied by each other. This can be seen by the example satisfying $\mathcal{Y}_1 \cap \mathcal{Y}_2 = \emptyset$ in Section 3.1 and the following example. Consider any non-symmetrizable AVC with state symbols in the set $\mathcal{S}$. The CAVC with $\mathcal{S}_1 = \mathcal{S}_2 = \mathcal{S}$ is not any-symmetrizable, but has $\overline{\mathcal{W}}_1 \cap \overline{\mathcal{W}}_2 \neq \emptyset$.

Theorem 3. (i) The random coding capacity for joint communication and compound-state identification over CAVC $C_{\mathsf{and}}^{\mathsf{r}} = 0$ if $\overline{\mathcal{W}}_1 \cap \overline{\mathcal{W}}_2 \neq \emptyset$. If $\overline{\mathcal{W}}_1 \cap \overline{\mathcal{W}}_2 = \emptyset$, then

$$C_{\mathsf{and}}^{\mathsf{r}} = C_{\mathsf{com}}^{\mathsf{r}}. \tag{3.6}$$

(ii) The deterministic capacity for joint communication and compound-state identification $C_{\mathsf{and}}^{\mathsf{d}} > 0$ if and only if the CAVC is not any-symmetrizable and $\overline{\mathcal{W}}_1 \cap \overline{\mathcal{W}}_2 = \emptyset$. If $C_{\mathsf{and}}^{\mathsf{d}} > 0$, then

$$C_{\mathsf{and}}^{\mathsf{d}} = C_{\mathsf{com}}^{\mathsf{r}}. \tag{3.7}$$

3.4 Communication or Compound-State Identification over CAVC

Let the deterministic code capacity for the CAVC for the 'communication or compound-state identification' problem be denoted by $C_{\mathsf{or}}^{\mathsf{d}}$ and the random code capacity as $C_{\mathsf{or}}^{\mathsf{r}}$. Observe that $C_{\mathsf{and}}^{\mathsf{d}} \leq C_{\mathsf{com}}^{\mathsf{d}} \leq C_{\mathsf{or}}^{\mathsf{d}}$. Since the decoder needs to either communicate or identify the compound-state, this is not possible if the CAVC is trans-symmetrizable as trans-symmetrizability hinders both the tasks of compound-state identification and communication. In Theorem 4, we claim that non-trans-symmetrizability is necessary and sufficient for positive capacity - a significantly more relaxed condition as compared to non-any-symmetrizability.

Remark 1. If $\overline{\mathcal{W}}_2 \subseteq \overline{\mathcal{W}}_1$, then the decoder cannot identify the compound-state σ_2 reliably, and therefore, the decoder must recover the message in this case. The model in [9] considers an AVC (with state alphabet $\mathcal{S}$) with a special no-adversary state $s_0 \in \mathcal{S}$. The decoder must decode the message correctly w.h.p. when the AVC-state sequence is $s_0^n = (s_0, \ldots, s_0)$. For any other AVC-state sequence $\boldsymbol{s} \neq s_0^n$, the decoder may declare adversarial interference. This is a special case of our model with $\mathcal{S}_2 = \{s_0\} \subseteq \mathcal{S}_1$.

For either compound-state, consider the case when the adversary samples the AVC-state symbols independently and identically distributed (i.i.d.) according to P_S such that $\sum_s P_S(s) W_{Y|X,S=s} \in \overline{\mathcal{W}}_1 \cap \overline{\mathcal{W}}_2$. Here, the decoder cannot identify the compound-state reliably, therefore the decoder must recover the message. Thus, for any channel

$W \in \overline{\mathcal{W}}_1 \cap \overline{\mathcal{W}}_2$, the capacity of W is an upper bound on C_{or}^{d}, i.e., $C_{\text{or}}^{\text{d}} \leq \max_{P_X} \min_{W \in \overline{\mathcal{W}}_1 \cap \overline{\mathcal{W}}_2} I(X;Y)$. It is also possible to show that this upper bound is achievable when the CAVC is not trans-symmetrizable as described in Section 6.

Theorem 4. (i) The random coding capacity for 'communication or compound-state identification' over CAVC is given by

$$C_{\text{or}}^{\text{r}} = \max_{P_X} \min_{W \in \overline{\mathcal{W}}_1 \cap \overline{\mathcal{W}}_2} I(X;Y). \tag{3.8}$$

In particular, if $\overline{\mathcal{W}}_1 \cap \overline{\mathcal{W}}_2 = \emptyset$, then $C_{\text{or}}^{\text{r}} = \infty$.
(ii) The deterministic capacity $C_{\text{or}}^{\text{d}} > 0$ if and only if the CAVC is not trans-symmetrizable. If $C_{\text{or}}^{\text{d}} > 0$, then

$$C_{\text{or}}^{\text{d}} = C_{\text{or}}^{\text{r}}. \tag{3.9}$$

Note that we can independently obtain Theorem 2 from this theorem - if the compound-state can be identified, then the message need not be decoded so the capacity is infinite for such a CAVC; thus, Theorem 4 implies that compound-state can be identified (i) under random coding if and only if $\overline{\mathcal{W}}_1 \cap \overline{\mathcal{W}}_2 = \emptyset$, and (ii) under deterministic coding if and only if the CAVC is not trans-symmetrizable and $\overline{\mathcal{W}}_1 \cap \overline{\mathcal{W}}_2 = \emptyset$.

Note that for a non-trans-symmetrizable, but cis-symmetrizable CAVC with $\overline{\mathcal{W}}_1 \cap \overline{\mathcal{W}}_2 \neq \emptyset$, it is impossible to just communicate and it is impossible to identify the compound-state separately; cis-symmetrizability hinders communication while $\overline{\mathcal{W}}_1 \cap \overline{\mathcal{W}}_2 \neq \emptyset$ hinders compound-state identification. However, such channels would have a positive capacity according to Theorem 4 for the problem of 'communication or compound-state identification'.

4 Error Exponent Analysis

Let the optimal error exponent in compound-state identification be denoted by G. If $\overline{\mathcal{W}}_1 \cap \overline{\mathcal{W}}_2 \neq \emptyset$, then we can not identify the adversary as explained in Section 3.2. If $\overline{\mathcal{W}}_1 \cap \overline{\mathcal{W}}_2 = \emptyset$, then suppose the sets $\overline{\mathcal{W}}_1$ and $\overline{\mathcal{W}}_2$ are singleton sets having elements Q and R respectively. In this case, if we send the channel input as i.i.d. P_X, then each input X that is sent and Y that is received, (X, Y) is distributed i.i.d. according to the distribution $P_X \times Q$ or $P_X \times R$ depending on the active adversary. Thus, we observe n realizations from either the distribution $P_X Q$ or the distribution $P_X R$. This is a classical hypothesis testing problem and the optimal decoding strategy is given by the Neyman-Pearson lemma [15]. For our problem, we are interested in the maximum of the two types of errors in hypothesis testing problem and we can use the results known for minimizing the error under Bayesian setting to get the optimal error exponent in this scheme - this gives a tight bound. However, the optimal error exponent G can be greater than this since the encoding scheme need to be i.i.d. Further, the set of channels $\overline{\mathcal{W}}_1$ and $\overline{\mathcal{W}}_2$ need not be singleton and the AVC-state can be arbitrarily varied during transmission. For $n = 1$, if an input symbol is picked using distribution P_X, then in terms of hypothesis testing, we can write $\Theta_0 = P_X \times \overline{\mathcal{W}}_1$ and $\Theta_1 = P_X \times \overline{\mathcal{W}}_2$. However, for $n > 1$, as we noted earlier, we can have a non-i.i.d. input distribution to the channel which makes it impossible to formulate our problem in terms of hypothesis testing. In particular, our problem has the following difficulties :

1. dependence among the n samples which we obtain,
2. the hypotheses being composite in nature, i.e., each hypothesis contains a set of parameters, and
3. the distributions being non-stationary, i.e., the distribution used to obtain each observation can differ.

The non-stationary composite hypothesis testing problem is significantly harder than a simple hypothesis testing problem. For our current, analysis, we contain the input symbols to be i.i.d. Note the problem still remains

non-stationary since the AVC-state can still vary arbitrarily. To get a converse under this easier problem, consider an adversarial strategy which chooses $Q \in \overline{\mathcal{W}}_1$ and $R \in \overline{\mathcal{W}}_2$ and uses it for all the transmission. The optimal error exponent can be obtained for this case and hence, for the upper bound, we can minimize this with respect to all $Q \in \overline{\mathcal{W}}_1$ and $R \in \overline{\mathcal{W}}_2$ to obtain the converse bound in Theorem 5. Additionally, for the problem of finding rate-exponent region $\mathcal{R}$, we give an achievable inner-bound in Theorem 5. The proof for this theorem is present in Section 7.6.

Let $\mathcal{U} = \mathcal{X}$ and Q_{UX} is a distribution over $\mathcal{U} \times \mathcal{X}$. Let $C(Q_{UX}) = \min_{W \in \overline{\mathcal{W}}_1 \cup \overline{\mathcal{W}}_2} I(X;Y|U)$ where $P_{UXY} = Q_{UX} \times W$ and let $C_0 = \max_{Q_{UX}} C(Q_{UX})$. Let E_k, for $k = 1, 2$, be a partition of the set of all distributions over $\mathcal{U} \times \mathcal{Y}$ ($E_1^c = E_2$), then, similar to $\overline{\mathcal{W}}_k$, define $\overline{\mathcal{V}}_k$, for $k = 1, 2$, as the closure of channels in the family $\{\sum_x Q_{U,X=x} W_{Y|X=x} : W_{Y|X} \in \mathcal{W}_k\}$. Let $\mathcal{P}^* = \{P_{UY} : \min_{Z \in \overline{\mathcal{V}}_1} D(P||Z) = \min_{Z \in \overline{\mathcal{V}}_2} D(P||Z)\}$. Define

$$\gamma(Q_{UX}) = \max_{E_k} \min_{k} \min_{\substack{P_{Y|U} \in E_k \\ Z_{Y|U} \in \overline{\mathcal{V}}_k}} D(P||Z). \tag{4.1}$$

Under the constraint of i.i.d. input symbols, let the rate-exponent region be $\mathcal{R}' \subseteq \mathcal{R}$ and let the optimal error exponent be $G' = \max\{\gamma : (R, \gamma) \in \mathcal{R}' \text{ for some } R\} \leq G$.

Theorem 5. i) Under the constraint of i.i.d. input symbols,

$$G' \leq \max_{Q_{UX}} \min_{Z \in \overline{\mathcal{V}}_k} \min_{P \in \mathcal{P}^*} D(P||Z).$$

ii) The rate-exponent pair $(C(Q_{UX}), \gamma(Q_{UX}))$ is achievable. In particular,

$$\{(C(Q_{UX}), \gamma(Q_{UX})) : Q_{UX} \text{ is any distribution over } \mathcal{U} \times \mathcal{X}\} \subseteq \mathcal{R}.$$

For a sanity check, It can be seen that $C_0 \leq C_{\text{com}}^r$ due to the Markov chain formed by $U \leftrightarrow X \leftrightarrow Y$.

5 Conclusion

We have introduced an unifying generalization of the compound channel and AVC. For the purpose of analysis, we consider the case where there are two compound-states. In this new setting, we have characterized the communication capacity in Theorem 1. Unlike compound channels or AVCs, the CAVC has a new notion of identifying the compound-state that was active during the transmission. This can also be interpreted as identifying an adversary which was active as explained in the Introduction. This naturally raises the question of trying to identify the compound-state and we give the conditions for feasibility in Theorem 2. One can also combine the problem of compound-state identification and communication in the two ways which are described in Table 1. For the problem of joint communication and compound-state identification (AND task), we present the capacity and feasibility conditions in Theorem 3. It is interesting to note that if the task is feasible, then compound-state identification can be done without any penalty on rate of communication. The capacity and feasibility condition for communication or compound-state identification (OR task) is presented in Theorem 4. It is worth noting that Theorem 4 shows that for the OR problem, we can take the 'best out of both' communication and compound-state identification. In particular, it is interesting to note that non-trans-symmetrizability and $\overline{\mathcal{W}}_1 \cap \overline{\mathcal{W}}_2 = \emptyset$ is not sufficient to ensure neither communication nor compound-state identification separately but it is sufficient for the OR task.

We also briefly consider the problem of generalizing our results to the case of more than two compound-state.

Much of the proofs that we present for the two compound-state case follows through without significant change for the generalization. Formally, for $\mathcal{K}$ adversaries, let the closures of the families of channels corresponding to each adversary be denoted by $\overline{\mathcal{W}}_k$, $k = 1, \ldots, \mathcal{K}$. The communication capacity under random coding is given by $\max_{P_X} \min_{W \in \bigcup_k \overline{\mathcal{W}}_k} I(X;Y)$. The necessary and sufficient condition for compound-state identification, under random coding, is $\bigcup_{i \neq j} \overline{\mathcal{W}}_i \cap \overline{\mathcal{W}}_j = \emptyset$. For the AND problem under random coding, the CAVC has communication capacity if compound-state identification is feasible. For the OR problem, the capacity is given by $\max_{P_X} \min_{W \in \bigcup_{i \neq j} \overline{\mathcal{W}}_i \cap \overline{\mathcal{W}}_j} I(X;Y)$. Under deterministic coding, we need to consider i, j-symmetrizability, i.e., cis- and trans-symmetrizability are replaced by the more general i, j-symmetrizability. Define non-any-symmetrizability as non-cis-symmetrizability and non-i, j-symmetrizability $\forall i, j$. Communication is feasible if the CAVC is non-any-symmetrizable. Compound-State identification is feasible if CAVC is non-i, j-symmetrizable $\forall i \neq j$ and $\overline{\mathcal{W}}_1 \cap \overline{\mathcal{W}}_2 = \emptyset$. For the AND task, we require feasibility of Compound-State identification and communication. We are still working on the feasibility of the OR task for deterministic coding.

In addition of these four problems, we also analyzed the error exponents for the AND problem and to obtain a converse for it, we tried to obtain a converse for the error exponent of compound-state identification problem. However, unsurprisingly, this problem has turned out to be much harder than what we had expected. We are currently still working on it.

6 Proof Sketches

We give a brief proof outline for the theorems. The full proofs can be found in Section 7. Let $\mathcal{P}_k$ denote the set of all distributions over $\mathcal{S}_k$, $k = 1, 2$.

6.1 Proof Sketch for Theorem 1 (i)

Both the achievability and converse parts of the proof follow along similar lines as that for standard AVCs. The achievability argument uses a randomly generated (and shared with the decoder) codebook where all code symbols are generated i.i.d. $\sim P_X$, a maximizing distribution of (3.4).

6.2 Proof Sketch for Theorem 2 (i)

If $\overline{\mathcal{W}}_1 \cap \overline{\mathcal{W}}_2 \neq \emptyset$, then the adversary under either compound-state can induce any effective channel in $\overline{\mathcal{W}}_1 \cap \overline{\mathcal{W}}_2$ using a suitable state distribution. Thus the compound-state cannot be identified reliably in this case.

For achievability, pick any vector $\boldsymbol{x}$ randomly which has an uniform empirical distribution over $\mathcal{X}$. Suppose the decoder receives the vector $\boldsymbol{y}$, then it can estimate the effective channel law from the joint distribution of $(\boldsymbol{x}, \boldsymbol{y})$. The condition $\overline{\mathcal{W}}_1 \cap \overline{\mathcal{W}}_2 = \varnothing$ ensures that the joint distribution of $(\boldsymbol{x}, \boldsymbol{y})$ is not in both $\overline{\mathcal{W}}_1$and $\overline{\mathcal{W}}_2$.

6.3 Proof Sketch for Theorem 3 (i)

The converse for the case $\overline{\mathcal{W}}_1 \cap \overline{\mathcal{W}}_2 = \emptyset$ follows from the converse of Theorem 1 (i). We now outline the achievability argument under $\overline{\mathcal{W}}_1 \cap \overline{\mathcal{W}}_2 \neq \emptyset$. The encoder constructs a vector with two parts $\boldsymbol{x} = (\hat{\boldsymbol{x}}, \tilde{\boldsymbol{x}})$. The first part is used for communication and the second part is used for compound-state identification. The vector $\boldsymbol{x}$ is randomly permuted before transmission so that the adversary cannot apply different types of attack on the two parts. The permutation

is shared with the decoder, so that it can recover $\boldsymbol{x}$. The encoding of the message in $\hat{\boldsymbol{x}}$ and its decoding is similar to that in the proof of Theorem 1 (i). The second part $\tilde{\boldsymbol{x}}$ is a fixed $|\mathcal{X}|\log(n)$ length sequence consisting of $\log(n)$ repetitions of each symbol in $\mathcal{X}$. The decoder estimates the effective channel law from this part and identifies the compound-state based on whether it is in $\overline{\mathcal{W}}_1$ or in $\overline{\mathcal{W}}_2$ (as explained in Section 6.2).

6.4 Proof Sketches for Theorem 4 (i)

For the converse proof, we first note that since the adversary under either compound-state can induce a channel from $\overline{\mathcal{W}}_1 \cap \overline{\mathcal{W}}_2$, the compound-state cannot be identified if the induced channel is in $\overline{\mathcal{W}}_1 \cap \overline{\mathcal{W}}_2$. So the decoder must decode the message reliably in such situation. However, by standard arguments, the decoder cannot decode reliably if the rate is more than $C_{\mathsf{or}}^{\mathsf{r}}$.

We now discuss the achievability argument. The same coding scheme is used as in Theorem 3 (i) using a distribution P_X that maximizes (3.8). If the effective channel induced (in both $\tilde{\boldsymbol{x}}$ and $\hat{\boldsymbol{x}}$) by the adversary is in $\overline{\mathcal{W}}_1 \cap \overline{\mathcal{W}}_2$, then the reliability in decoding follows using standard arguments since the rate is less than $\min_{W \in \overline{\mathcal{W}}_1 \cap \overline{\mathcal{W}}_2} I(X;Y)$. On the other hand, if the effective channel is outside $\overline{\mathcal{W}}_1 \cap \overline{\mathcal{W}}_2$, then the compound-state can be identified, as discussed in the proof of Theorem 3 (i).

6.5 Proof Sketches for Theorem 1 (ii), Theorem 2 (ii), Theorem 3 (ii), & Theorem 4 (ii)

It can be shown that $C_{\mathsf{com}}^{\mathsf{r}} > 0$ (resp. $C_{\mathsf{or}}^{\mathsf{d}} > 0$) when the channel is not any-symmetrizable (resp. trans-symmetrizable).

The achievability proof for deterministic coding follows along similar lines of argument as in [4]. A suitable codebook with codewords $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M$ of type P_X can be obtained using an extension of [4, Lemma 3] for all the three theorems with appropriate P_X. We only describe the decoders below, and refer the reader to Section 7 for the detailed analysis. The decoder for the task of joint 'communication and compound-state identification' (Theorem 3 (ii)) is as described below. Let

$$\mathcal{C}_\eta = \{P_{XSY} : D(P_{XSY} || P_X \times P_S \times W_{Y|X,S}) \leq \eta,\ P_S \in \mathcal{P}_1 \cup \mathcal{P}_2\}. \tag{6.1}$$

Decoder. Given codewords $\boldsymbol{x}_j$, $j = 1, \ldots, M$, set $\phi^{\mathsf{and}}(\boldsymbol{y}) = (i, \sigma_k)$, $i \in \mathcal{M}, k \in \{1,2\}$, iff an $\boldsymbol{s} \in \mathcal{S}_k^n$ exists such that:

1. the joint type $P_{\boldsymbol{x}_i, \boldsymbol{s}, \boldsymbol{y}} \in \mathcal{C}_\eta$, and
2. for each $\boldsymbol{x}_j, j \neq i$ such that there exists $\boldsymbol{s}' \in \mathcal{S}_1^n \cup \mathcal{S}_2^n$, $P_{\boldsymbol{x}_j, \boldsymbol{s}', \boldsymbol{y}} \in \mathcal{C}_\eta$, we have $I(XY; X'|S) \leq \eta$ where $P_{XX'SY} = P_{\boldsymbol{x}_i, \boldsymbol{x}_j, \boldsymbol{s}, \boldsymbol{y}}$.

Set $\phi^{\mathsf{and}}(\boldsymbol{y}) = (1, a_1)$ if no such (i, a_k) exists.

The condition $\overline{\mathcal{W}}_1 \cap \overline{\mathcal{W}}_2 = \emptyset$ ensures that if there exists $\boldsymbol{s} \in \mathcal{S}_1^n$, $P_{\boldsymbol{x}_m, \boldsymbol{s}, \boldsymbol{y}} \in \mathcal{C}_\eta$ then $\forall \boldsymbol{s}' \in \mathcal{S}_2^n$, $P_{\boldsymbol{x}_m, \boldsymbol{s}', \boldsymbol{y}} \notin \mathcal{C}_\eta$. For two distinct codewords $\boldsymbol{x}_i, \boldsymbol{x}_j$, and their corresponding $\boldsymbol{s}_i, \boldsymbol{s}_j$ respectively, (i) non-cis-symmetrizability ensures that they do not simultaneously satisfy both the decoder conditions when both $\boldsymbol{s}_i, \boldsymbol{s}_j \in \mathcal{S}_k^n$ for some $k \in \{1,2\}$, (ii) non-trans-symmetrizability ensures they do not simultaneously satisfy both the decoder conditions when $\boldsymbol{s}_i \in \mathcal{S}_k^n$, $\boldsymbol{s}_j \in \mathcal{S}_{3-k}^n$ for some $k \in \{1,2\}$ (see Section 7).

For Theorem 1 (ii), we can use a decoder similar to the above and disregard the decoder output corresponding to the compound-state identity. Similarly, for Theorem 2 (ii), we can use a similar decoder and disregard the decoder

output corresponding to the message.

For Theorem 4 (ii), we show the achievability of a non-zero rate, and then use the randomness reduction technique [16] to achieve the capacity. The following decoder is used to show positive capacity.

Decoder. Given codewords $\boldsymbol{x}_j$, $j = 1, \ldots, M$, let B_k ($k = 1, 2$) be the set of messages $m \in \mathcal{M}$ such that

1. the joint type $P_{\boldsymbol{x}_m, \boldsymbol{s}, \boldsymbol{y}} \in \mathcal{C}_\eta$, and
2. for every $m' \neq m$ such that there exists $\boldsymbol{s}' \in \mathcal{S}_{3-k}^n$, $P_{\boldsymbol{x}_{m'}, \boldsymbol{s}', \boldsymbol{y}} \in \mathcal{C}_\eta$, we have $I(XY; X'|S) \leq \eta$ where $P_{XX'SY} = P_{\boldsymbol{x}_m, \boldsymbol{x}_{m'}, \boldsymbol{s}, \boldsymbol{y}}$.

If $B_1 = B_2 = \{m\}$, then $\phi^{\mathsf{or}}(\boldsymbol{y}) = m$. If for some $k \in \{1, 2\}$, $B_k = \emptyset \neq B_{3-k}$, then the decoder outputs the compound-state $\phi^{\mathsf{or}}(\boldsymbol{y}) = \sigma_{3-k}$.

Non-trans-symmetrizability ensures that the two cases for B_k described in the decoder are the only cases which can occur (see Section 7).

The rate-converses follow from the converse for the randomized coding capacity. The zero-rate converse ideas have been discussed in Section 3 and are elaborated in Section 7.

7 Complete Proofs

For a given channel $W_{Y|X,S}$, we use the notation W_P to refer to the channel $W_P : \mathcal{X} \to \mathcal{Y}$ given by $\sum_s P(s) W_{Y|X,S=s}$. The ϵ-typical set of a random variable be denoted by $\tau_X^\epsilon = \{\boldsymbol{x} : |P_{\boldsymbol{x}}(x) - P_X(x)| \leq \epsilon \ \forall x \in \mathcal{X}\}$. In particular, τ_X denotes the typical set when $\epsilon = 0$. Let $\mathcal{P}_L^{(n)}$ denote the set all empirical distributions of length n over the set L.

7.1 Converse Proofs Under Random Coding

Lemma 1.

$$C_{\mathsf{com}}^{\mathsf{r}} \leq \max_{P_X} \min_{W \in \overline{\mathcal{W}}_1 \cup \overline{\mathcal{W}}_2} I(X; Y)$$

Proof. Consider the adversarial strategy for compound-state σ_k where the adversary chooses a distribution $P(\boldsymbol{s})$ with support over $\mathcal{S}_k^n$ and randomly samples a vector $\boldsymbol{s}$ distributed according to P. Note that the CAVC average error probability under the worst-case P is same as that under worst-case $\boldsymbol{s}$ (*c.f.* [16, Lemma 12.3, Page 210]). In other words, if $\mathcal{P}_i^{(n)}$ represents all distributions over $\mathcal{S}_k^n$, then

$$P_e^r(Q, k) = P_e^p(Q, \mathcal{P}_k^{(n)}),$$

where

$$P_e^p(Q, \mathcal{P}_k^{(n)}) = \max_{P \in \mathcal{P}_k^{(n)}} \sum_{\boldsymbol{s}} P(\boldsymbol{s}) \sum_{(f,\phi)} Q(f, \phi) \frac{1}{M} \sum_{m=1}^{M} W^n(\phi^{-1}(\mathcal{E}_{m,k}) | f(m), \boldsymbol{s}).$$

Here, $\mathcal{E}_{m,k}$ is the error event corresponding to communication error $\{m' \in \mathcal{M} : m' \neq m\}$.

Consider a particular class of adversarial strategies for compound-state σ_k where the adversary chooses the state sequence $\boldsymbol{s}$ with each bit independently from the distribution $P_k \in \mathcal{P}_k$, i.e., $P(\boldsymbol{s}) = P_k^n(\boldsymbol{s}) = \prod_{j=1}^n P_k(s_j)$. The probability of error under this adversarial strategy is given by

$$\sum_{(f,\phi)} Q(f,\phi)\frac{1}{M}\sum_{m=1}^{M} W_{P_k}^n(\phi^{-1}(\mathcal{E}_{m,k})|f(m)),$$

where $W_{P_k}(y|x) = \sum_{s\in\mathcal{S}_k} P_k(s)W(y|x,s)$. Therefore, channel distribution is given by Discrete Memoryless Channel (DMC) W_{P_k}.

Under such i.i.d. adversarial strategy, consider a sequence of codes with rate R' such that the error probability $P_e^{(n)}$ tends to 0 for large block-length. Let M be the message which is encoded into vector X^n and transmitted, and let Y^n be the vector received by the decoder. Then, $(M, Y^{i-1}) \longleftrightarrow X_i \longleftrightarrow Y_i$ form a Markov Chain under this adversarial strategy (as W_{P_k} is a DMC). Let $\hat{M}$ be the decoded message. By Data-Processing and Fano's inequalities,

$$H(M|Y^n) = H(M|\hat{M}) \leq 1 + P_e^{(n)} nR' = n\epsilon_n,$$

where ϵ_n is defined as $\frac{1}{n} + P_e^{(n)} R'$. Next, we note that

$$\begin{aligned} nR' &= H(M) && (7.1)\\ &= H(M|Y^n) + I(M;Y^n) && (7.2)\\ &\leq n\epsilon_n + I(M;Y^n). && (7.3) \end{aligned}$$

Consider the term $I(M;Y^n)$ -

$$\begin{aligned} I(M;Y^n) &= \sum_{j=1}^{n} I(M;Y_j|Y^{j-1}) && (7.4)\\ &\leq \sum_{j=1}^{n} I(M, X_j, Y^{j-1}; Y_j) && (7.5)\\ &= \sum_{j=1}^{n} I(X_j;Y_j), && (7.6)\\ & && (7.7) \end{aligned}$$

where the last equality follows from the property of Markov Chains ($(M, Y^{i-1}) \longleftrightarrow X_i \longleftrightarrow Y_i$).

Let $L \sim Uniform[1,n]$ be independent of other random variables. Note that $L \longleftrightarrow X_L \longleftrightarrow Y_L$ forms a Markov Chain. Thus, we have,

$$\begin{aligned} \frac{1}{n}\sum_{j=1}^{n} I(X_j;Y_j) &= I(X_L;Y_L|L) && (7.8)\\ &\leq I(X_L, L; Y_L) && (7.9)\\ &= I(X_L;Y_L). && (7.10) \end{aligned}$$

Since (7.3) has to hold for all such i.i.d. adversarial strategies,

$$R' \leq \epsilon_n + \min_{P\in\mathcal{P}_1\cup\mathcal{P}_2} I(X;Y),$$

where Y is related to X via the DMC W_P. Further, ϵ_n can be made arbitrarily small by choosing n large enough since

$P_e^{(n)}$ vanishes for large n. Therefore, for every achievable rate $R' < C_{\mathsf{com}}^{\mathsf{r}}$, we have,

$$C_{\mathsf{com}}^{\mathsf{r}} \leq \max_{P_X} \min_{W \in \overline{\mathcal{W}}_1 \cup \overline{\mathcal{W}}_2} I(X;Y).$$

■

Lemma 2. $\overline{\mathcal{W}}_1 \cap \overline{\mathcal{W}}_2 = \emptyset$ is necessary for compound-state identification under random coding.

Proof. Let $\overline{\mathcal{W}}_1 \cap \overline{\mathcal{W}}_2 \neq \emptyset$, then $\exists$ channel $Z : \mathcal{X} \to \mathcal{Y}$, $Z_{Y|X} \in \overline{\mathcal{W}}_1 \cap \overline{\mathcal{W}}_2$. Therefore, we can choose distribution P_k over $\mathcal{S}_k$ such that $Z_{Y|X} = \sum_s P_k(s) W_{Y|X,S=s}$ for $k = 1, 2$.
Let $T_k(\boldsymbol{s}) := \prod_{j=1}^n P_k(s_j)$. Consider an adversarial strategy where the adversary chooses the state i.i.d. from distribution P_k when the compound state is σ_k. Under this attack and compound state σ_k, we have,

$$P_e^r(Q,k) \geq \sum_{\boldsymbol{s}} T_k(\boldsymbol{s}) \sum_{(f,\phi)} Q(f,\phi) \frac{1}{M} \sum_{m=1}^{M} W^n(\phi^{-1}(\mathcal{E}m,k)|f(m),\boldsymbol{s}) \tag{7.11}$$

$$= \frac{1}{M} \sum_{m=1}^{M} \sum_{\boldsymbol{s}} \sum_{(f,\phi)} \sum_{\boldsymbol{y} \in \phi^{-1}(\mathcal{E}_{m,k})} Q(f,\phi) T_k(\boldsymbol{s}) W^n(\boldsymbol{y}|f(m),\boldsymbol{s}) \tag{7.12}$$

$$= \frac{1}{M} \sum_{m=1}^{M} \sum_{\boldsymbol{s}} \sum_{(f,\phi)} \sum_{\boldsymbol{y} \in \phi^{-1}(\mathcal{E}_{m,k})} Q(f,\phi) \prod_{j=1}^{n} T_k(s_j) W^n(y_j|f(m)_j, \boldsymbol{s}_j) \tag{7.13}$$

$$= \frac{1}{M} \sum_{m=1}^{M} \sum_{(f,\phi)} Q(f,\phi) Z^n(\phi^{-1}(\mathcal{E}_{m,k})|f(m)). \tag{7.14}$$

Hence,

$$P_e^r(Q,1) + P_e^r(Q,2) \geq \frac{1}{M} \sum_{m=1}^{M} \sum_{(f,\phi)} Q(f,\phi) Z^n(\phi^{-1}(\mathcal{E}_{m,1}) \cup \phi^{-1}(\mathcal{E}_{m,2})|f(m)) \tag{7.15}$$

$$\geq 1 \quad \forall Q, \tag{7.16}$$

where (7.16) follows as $\phi^{-1}(\mathcal{E}_{m,1} \cup \mathcal{E}_{m,2}) = \mathcal{Y}^n$. Therefore, compound-state identification is not possible if $\overline{\mathcal{W}}_1 \cap \overline{\mathcal{W}}_2 \neq \emptyset$.

■

Note that the probability of error in only compound-state identification is strictly less than or equal to the probability of error in joint compound-state identification and communication. Thus, if the error probability in compound-state identification is not vanishing for a CAVC, then the error probability in joint communication and compound-state identification cannot vanish. Lemma 2 establishes that $C_{\mathsf{and}}^{\mathsf{r}} = 0$ if $\overline{\mathcal{W}}_1 \cap \overline{\mathcal{W}}_2 \neq \emptyset$. If $\overline{\mathcal{W}}_1 \cap \overline{\mathcal{W}}_2 = \emptyset$, the fact $C_{\mathsf{and}}^{\mathsf{r}} \leq C_{\mathsf{com}}^{\mathsf{r}}$ and Lemma 1 establish that

$$C_{\mathsf{and}}^{\mathsf{r}} \leq \max_{P_X} \min_{W \in \overline{\mathcal{W}}_1 \cup \overline{\mathcal{W}}_2} I(X;Y)$$

.

Lemma 3.

$$C_{\mathsf{or}}^{\mathsf{r}} \leq \max_{P_X} \min_{W \in \overline{\mathcal{W}}_1 \cap \overline{\mathcal{W}}_2} I(X;Y) \tag{7.17}$$

Proof. When $\overline{\mathcal{W}}_1 \cap \overline{\mathcal{W}}_2 = \emptyset$, RHS of (7.17) is infinity and the relation holds trivially.

If $\overline{\mathcal{W}}_1 \cap \overline{\mathcal{W}}_2 \neq \emptyset$, then let $Z_1, \ldots, Z_n$ be any n channels $\in \overline{\mathcal{W}}_1 \cap \overline{\mathcal{W}}_2$. We represent the n-length channel as $Z^{(n)}(\boldsymbol{y}|\boldsymbol{x}) = \prod_{i=1}^{n} Z_i(y_i|x_i)$. Define $P_{i,k}(s) \in \mathcal{P}_k$ for $k = 1, 2$ such that $\sum_s P_{i,k}(s) W_{Y|X,S=s} = Z_i$. We have,

$$P_e^r(Q,k) \geq \sum_{\boldsymbol{s} \in \mathcal{S}_k^n} \prod_{i=1}^{n} P_{i,k}^n(s_i) \sum_{(f,\phi)} Q(f,\phi) \frac{1}{M} \sum_{m=1}^{M} W^n(\phi^{-1}(\mathcal{E}_{m,k})|f(m), \boldsymbol{s}) \tag{7.18}$$

$$= \frac{1}{M} \sum_{(f,\phi)} \sum_{m=1}^{M} Q(f,\phi) Z^{(n)}(\phi^{-1}(\mathcal{E}_{m,k})|f(m)); \tag{7.19}$$

$$\implies 2P_e^r(Q) \geq \frac{1}{M} \sum_{(f,\phi)} \sum_{m=1}^{M} Q(f,\phi) Z^{(n)}(\phi^{-1}(\mathcal{E}_{m,1} \cup \mathcal{E}_{m,2})|f(m)) \tag{7.20}$$

$$= \frac{1}{M} \sum_{(f,\phi)} \sum_{m=1}^{M} Q(f,\phi) Z^{(n)}(\phi^{-1}(\{\sigma_1, \sigma_2\} \cup \mathcal{M} \setminus m)|f(m)) \tag{7.21}$$

$$= \frac{1}{M} \sum_{(f,\phi)} \sum_{m=1}^{M} Q(f,\phi) Z^{(n)}(\phi^{-1}(m)^{\mathsf{C}}|f(m)). \tag{7.22}$$

In order to get $P_e^r(Q) \to 0$, we must ensure the RHS vanishes as n increases for all $Z^{(n)}$ with $Z_i \in \overline{\mathcal{W}}_1 \cap \overline{\mathcal{W}}_2$. The RHS is exactly the probability of error for communication over an AVC with the family of channels $\overline{\mathcal{W}}_1 \cap \overline{\mathcal{W}}_2$. Thus, we have,

$$C_{\mathsf{or}}^{\mathsf{r}} \leq \max_{P_X} \min_{Z \in \overline{\mathcal{W}}_1 \cap \overline{\mathcal{W}}_2} I(X;Y)$$

■

7.2 Achievability Proof of Theorem 1 (i) and Theorem 2

Lemma 4.

$$C_{\mathsf{com}}^{\mathsf{r}} \geq \max_{P_X} \min_{W \in \overline{\mathcal{W}}_1 \cup \overline{\mathcal{W}}_2} I(X;Y)$$

Proof. The proof is along the lines of [9, Lemma 5]. For any $R < C_{\mathsf{com}}^{\mathsf{r}}$, choose $\delta > 0$ such that $R + \delta < C_{\mathsf{com}}^{\mathsf{r}}$. We describe the encoder-decoder pair (F_R, Φ_R) (parameterized by the rate R) used to achieve the capacity. The codebook for a $(2^{nR}, n)$ code is obtained by uniformly and independently sampling M vectors $(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_M) \in \tau_X$ where τ_X is the typical set corresponding to some $P_X \in \mathcal{P}_{\mathcal{X}}^{(n)}$, and $F_R(i) = \boldsymbol{X}_i$. The decoder outputs $\Phi_R(\boldsymbol{y}) = i \in \mathcal{M}$ if there is a unique i for which $I(X;Y) \geq R + \delta$ where $P_{XY} = P_{\boldsymbol{X}_i, \boldsymbol{y}}$, and $\Phi_R(\boldsymbol{y}) = 1$ if no such i exists.

If message i is sent and the AVC-state sequence is $\boldsymbol{s}$ during transmission, we need to prove the following two results to show that rate R is achievable :

$$\mathbb{P}\{(\boldsymbol{X}_i, \boldsymbol{y}) \in \tau_{XY}, I(X;Y) < R + \delta\} \xrightarrow{n \to \infty} 0 \ \ \forall \boldsymbol{s} \in \mathcal{S}_1^n \cup \mathcal{S}_2^n, \text{ and} \tag{7.23}$$

$$\mathbb{P}\{(\boldsymbol{X}_j, \boldsymbol{y}) \in \tau_{XY}, I(X;Y) \geq R + \delta, \text{for some } j \neq i\} \xrightarrow{n \to \infty} 0 \ \ \forall \boldsymbol{s} \in \mathcal{S}_1^n \cup \mathcal{S}_2^n. \tag{7.24}$$

The probability expression in the LHS of (7.105) is equal to

$$\sum_{\substack{P_{XSY}: I(X;Y) < R+\delta, \\ \boldsymbol{s} \in \tau_S}} \sum_{\boldsymbol{x} \in \tau_{X|S}(\boldsymbol{s})} |\tau_X|^{-1} \sum_{\boldsymbol{y} \in \tau_{Y|XS}(\boldsymbol{x}, \boldsymbol{s})} W^n(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{s}) \tag{7.25}$$

$$\leq \sum_{\substack{P_{XSY}: I(X;Y)<R+\delta, \\ \boldsymbol{s}\in\tau_S}} \sum_{\boldsymbol{x}\in\tau_{X|S}(\boldsymbol{s})} |\tau_X|^{-1} \exp\{-nD(P_{XSY}||P_{XS}\times W)\} \tag{7.26}$$

$$= \sum_{\substack{P_{XSY}: I(X;Y)<R+\delta, \\ \boldsymbol{s}\in\tau_S}} \frac{|\tau_{X|S}(\boldsymbol{s})|}{|\tau_X|} \exp\{-nD(P_{XSY}||P_{XS}\times W)\} \tag{7.27}$$

$$\leq \sum_{\substack{P_{XSY}: I(X;Y)<R+\delta, \\ \boldsymbol{s}\in\tau_S}} \exp\{-n(D(P_{XSY}||P_{XS}\times W) + I(X;S) - \epsilon)\}. \tag{7.28}$$

Using the fact that $D(P_{XSY}||P_{XS}\times W) + I(X;S) = D(P_{XSY}||P_X \times P_S \times W)$, and taking the marginals along $\mathcal{X}\times\mathcal{Y}$, while noting that divergence does not increase with marginalization, we have

$$\mathbb{P}\{(\boldsymbol{X}_i, \boldsymbol{y}) \in \tau_{XY}, I(X;Y) < R+\delta\} \leq \sum_{\substack{P_{XSY}: I(X;Y)<R+\delta, \\ \boldsymbol{s}\in\tau_S}} \exp\{-n(D(P_{XY}||P_X \times W_{P_S}) - \epsilon)\},$$

where $W_{P_S} = \sum_s P_S(s) W_{Y|X,S=s}$. In (7.28), we can set ϵ arbitrarily small as ϵ is present to account for the $(n+1)^{|\mathcal{X}|}$ term which grows polynomially. In particular, set $\epsilon < \epsilon'$, where ϵ' is described next.
Note that if $P_{XY} = P_X \times W_{P_S}$, then $R+\delta < I(X;Y)$ (as $W_{P_S} \in \overline{\mathcal{W}}_1 \cup \overline{\mathcal{W}}_2$) by choice of R and δ as described. Since mutual information and relative entropy are continuous functions of P_{XY}, there exists $\epsilon' > 0$ such that if $I(X;Y) < R+\delta$, then

$$D(P_{XY}||P_X \times W_{P_S}) \geq \epsilon' \ \forall P_S, \text{ or equivalently, } \forall \boldsymbol{s}.$$

Since there are only polynomially many types, for sufficiently large n, (7.105) is less than $\exp\{-n(\epsilon'-\epsilon)/2\} \to 0$ as $n\to\infty$.

Next, we analyze the probability in the LHS of (7.106). The probability, for any $\boldsymbol{s}$, can be written as

$$= \sum_{\substack{P_{XX'SY}: I(X';Y)\geq R+\delta \\ \boldsymbol{s}\in\tau_S,\ P_X = P_{X'}}} \sum_{\boldsymbol{x}_i\in\tau_{X|S}(\boldsymbol{s})} |\tau_X|^{-1} \sum_{j=1, j\neq i}^{M} \sum_{\boldsymbol{x}_j\in\tau_{X'|XS}(\boldsymbol{x},\boldsymbol{s})} |\tau_X|^{-1} \sum_{\boldsymbol{y}\in\tau_{Y|XX'S}(\boldsymbol{x}_i,\boldsymbol{x}_j,\boldsymbol{s})} W^n(\boldsymbol{y}|\boldsymbol{x}_i,\boldsymbol{s}) \tag{7.29}$$

$$\leq \sum_{\substack{P_{XX'SY}: I(X';Y)\geq R+\delta \\ \boldsymbol{s}\in\tau_S,\ P_X = P_{X'}}} \exp(-n(I(X;S)-\epsilon)) \exp(nR) \exp(-n(I(X';XS)-\epsilon)) \exp(-n(I(Y;X'|XS)-\epsilon)) \tag{7.30}$$

$$\leq \sum_{\substack{P_{XX'SY}: I(X';Y)\geq R+\delta \\ \boldsymbol{s}\in\tau_S,\ P_X = P_{X'}}} \exp\{-n(I(X;S) + I(X';XSY) - R - 3\epsilon)\} \tag{7.31}$$

$$\leq \sum_{\substack{P_{XX'SY}: I(X';Y)\geq R+\delta \\ \boldsymbol{s}\in\tau_S,\ P_X = P_{X'}}} \exp\{-n(I(X';Y) - R - 3\epsilon)\} \tag{7.32}$$

$$\leq \sum_{\substack{P_{XX'SY}: I(X';Y)\geq R+\delta \\ \boldsymbol{s}\in\tau_S,\ P_X = P_{X'}}} \exp\{-n(\delta - 3\epsilon)\} \tag{7.33}$$

$$\leq \exp\{-n(\delta - 3\epsilon - \epsilon').\} \tag{7.34}$$

Note that ϵ and ϵ' can be set arbitrarily small as they are present to account for polynomially many terms. This proves the achievability of the capacity C^r_{com}. ■

We present the following 2 lemmas before describing compound-state identification.

Lemma 5. In a CAVC, let the random vector $\boldsymbol{X}$, chosen uniformly from the typical set τ_X corresponding to some distribution $P_X \in \mathcal{P}_{\mathcal{X}}^{(n)}$, be the input and the AVC-state sequence be $\boldsymbol{s} \in \mathcal{S}_k^n$. Suppose $\boldsymbol{Y}$ represents the output sequence. Then, for any $\epsilon > 0$ and sufficiently large n, the joint type $(\boldsymbol{X}, \boldsymbol{Y}) \in \tau_{XY}^{\epsilon}$ with high probability, where τ_{XY}^{ϵ} is the typical set corresponding to the distribution $P_{XY} = P_X \times \widetilde{Z}_{Y|X}$, for some $\widetilde{Z}_{Y|X} \in \overline{\mathcal{W}}_k$.

Proof. Consider the channel $\widetilde{Z}_{Y|X}$ which is the weighted average of the individual channels $W_{Y|X,S=s}$ (weighted with respect to fraction of $s \in \mathcal{S}_k$ occurrences, formalized later). We prove that the input $\boldsymbol{x}$, which is in the typical set τ_X, and the output $\boldsymbol{y}$ would be jointly typical with respect to the distribution $P_X \times \widetilde{Z}_{Y|X}$.

Without loss of generality, we analyze the problem when $\boldsymbol{s} \in \mathcal{S}_1^n$. Let $\mathcal{S}_1 = \{S_1, S_2, \ldots, S_T\}$ (where $T = |\mathcal{S}_1|$). Denote the indices of $\boldsymbol{s} \in \mathcal{S}_1^n$ where $s = S_i$ as $J_i(\boldsymbol{s})$, ie, $J_i(\boldsymbol{s}) = \{j : s_j = S_i\}$. Notice that,

$$P(\boldsymbol{y}, \boldsymbol{x}|\boldsymbol{s}) = \frac{1}{|\tau_X|} W^n(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{s}) \tag{7.35}$$

$$= \frac{1}{|\tau_X|} \prod_{i=1}^{n} W(y_i|x_i, s_i) \tag{7.36}$$

$$= \frac{1}{|\tau_X|} \prod_{i=1}^{T} \left[\prod_{j \in J_i(\boldsymbol{s})} W(y_j|x_j, S_i) \right]. \tag{7.37}$$

Fix an ϵ_1 (value described later) and from the sets $J_i(\boldsymbol{s})$, consider the sets which have $|J_i(\boldsymbol{s})| > \epsilon_1 n$, i.e., $\mathcal{G} := \{i \in \{1, 2, \ldots, T\} : |J_i(\boldsymbol{s})| > \epsilon_1 n\}$. $\mathcal{G}$ is non-empty for any value of $\epsilon_1 < 1/T$. Choose any $\epsilon_1 < \min\{1/T, 1/T'\}$ where $T' = |\mathcal{S}_2|$. Henceforth, we shall assume ϵ_1 satisfies this condition. Define the 'subset' vectors $\boldsymbol{x}_i := \{x_j : j \in J_i(\boldsymbol{s})\}$ and similarly $\boldsymbol{y}_i$. Let $\boldsymbol{S}_i$ be the vector of $|J_i(\boldsymbol{s})|$ repetitions of symbol S_i. Then, we can write (7.37) as

$$P(\boldsymbol{y}, \boldsymbol{x}|\boldsymbol{s}) = \frac{1}{|\tau_X|} \prod_{i=1}^{T} \left[W^{|J_i(\boldsymbol{s})|}(\boldsymbol{y}_i|\boldsymbol{x}_i, \boldsymbol{S}_i) \right].$$

By Lemma 10, $\boldsymbol{x}_i$, $i \in \mathcal{G}$ are of type $\tau_X^{\epsilon_2}$ with probability greater than $1 - f(\epsilon_2)$ for arbitrarily small ϵ_2 and sufficiently large n as their lengths are at least $\epsilon_1 n$ and $f(\cdot)$ satisfies $f(\epsilon_2) \to 0$ as $\epsilon_2 \to 0$. Therefore, $P\{\boldsymbol{x}_i \in \tau_X^{\epsilon_2} \ \forall i \in \mathcal{G}\} \geq 1 - f_2(\epsilon_2)$ where $f_2(\cdot) = |\mathcal{G}| f(\cdot)$ satisfies $f_2(\epsilon_2) \to 0$ as $\epsilon_2 \to 0$.

By conditional typicality lemma, if random variables X, Y_k are distributed as $P_{XY_k} = P_X \times W_{Y|X,S=S_k}$, then for any $\epsilon_4 > 0$

$$P\{(\boldsymbol{x}_i, \boldsymbol{y}_i) \in \tau_{XY_k}^{\epsilon_3} | \boldsymbol{x}_i \in \tau_X^{\epsilon_2}\} > 1 - \epsilon_4, \forall i \in \mathcal{G} \tag{7.38}$$

for any $\epsilon_3 > \epsilon_2$ and sufficiently large n. Denote the event $\{(\boldsymbol{x}_i, \boldsymbol{y}_i) \in \tau_{XY_k}^{\epsilon_3} \forall i \in \mathcal{G} | \boldsymbol{x}_i \in \tau_X^{\epsilon_2} \forall i \in \mathcal{G}\} = \mathcal{B}$. Similarly,

$$P(\mathcal{B}) > 1 - |\mathcal{G}|\epsilon_3.$$

Therefore, with high probability, the $(\boldsymbol{x}_i, \boldsymbol{y}_i)$, $i \in \mathcal{G}$ are jointly typical according to the distribution $P_X \times W_{Y|X,S=S_k}$. Denote $W_{Y|X,S=S_i}$ as $Z_{Y|X}^i$ (this is a single letter channel). We now show that $(\boldsymbol{x}, \boldsymbol{y})$ is jointly typical with $P_X \times \widetilde{Z}_{Y|X}$ with high probability, where

$$\widetilde{Z}_{Y|X}(b|a) = \frac{1}{\sum_{i \in \mathcal{G}} |J_i(\boldsymbol{s})|} \sum_{i \in \mathcal{G}} Z_{Y|X}^i(b|a) |J_i(\boldsymbol{s})|, \ (a, b) \in \mathcal{X} \times \mathcal{Y}.$$

Clearly, $\widetilde{Z}_{Y|X} \in \overline{\mathcal{W}}_1$. We need to show (w.h.p.)

$$|\pi(a,b|\boldsymbol{x},\boldsymbol{y}) - P_X(a)\widetilde{Z}_{Y|X}(b)| \leq \epsilon \; \forall (a,b) \in \mathcal{X} \times \mathcal{Y}$$

where $\pi(a,b|\boldsymbol{x},\boldsymbol{y})$ is the empirical distribution and ϵ is specified later.

Since $\mathcal{G}$ contains $J_i(\boldsymbol{s})$ which have at least cardinality of $\epsilon_1 n$, we can say that $\sum_{i\in\mathcal{G}^c} |J_i(\boldsymbol{s})| \leq (T-1)\epsilon_1 n$. Therefore, $\sum_{i\in\mathcal{G}} |J_i(\boldsymbol{s})| > n(1-(T-1)\epsilon_1)$. Hence, w.h.p.,

$$\pi(a,b|\boldsymbol{x},\boldsymbol{y}) = \frac{1}{n}\sum_{i=1}^{K} |J_i(\boldsymbol{s})|\pi(a,b|\boldsymbol{x}_i,\boldsymbol{y}_i) \tag{7.39}$$

$$= \frac{1}{n}\left(\sum_{i\in\mathcal{G}} |J_i(\boldsymbol{s})|\pi(a,b|\boldsymbol{x}_i,\boldsymbol{y}_i) + \sum_{i\in\mathcal{G}^c} |J_i(\boldsymbol{s})|\pi(a,b|\boldsymbol{x}_i,\boldsymbol{y}_i)\right). \tag{7.40}$$

Further,w.h.p.,

$$\frac{1}{n}\left(\sum_{i\in\mathcal{G}} |J_i(\boldsymbol{s})|(1-\epsilon_3)P_X(a)Z^i_{Y|X}(b|a)\right) \leq \pi(a,b|\boldsymbol{x},\boldsymbol{y}) \leq \frac{1}{n}\left(\sum_{i\in\mathcal{G}} |J_i(\boldsymbol{s})|(1+\epsilon_3)P_X(a)Z^i_{Y|X}(b|a) + \sum_{i\in\mathcal{G}^c}\epsilon_1 n\right) \tag{7.41}$$

$$(1-\epsilon_3)(1-(T-1)\epsilon_1)P_X(a)\widetilde{Z}_{Y|X}(b|a) \leq \pi(a,b|\boldsymbol{x},\boldsymbol{y}) \leq (1+\epsilon_3)P_X(a)\widetilde{Z}_{Y|X}(b|a) + (T-1)\epsilon_1. \tag{7.42}$$

Therefore (w.h.p.),

$$\begin{aligned}|\pi(a,b|\boldsymbol{x},\boldsymbol{y}) - P_X(a)\widetilde{Z}_{Y|X}(b|a)| &\leq \max\{(\epsilon_1(T-1)+\epsilon_3-\epsilon_1\epsilon_3(T-1))P_X(a)\widetilde{Z}_{Y|X}(b|a),\\ &\qquad \epsilon_3 P_X(a)\widetilde{Z}_{Y|X}(b|a) + (T-1)\epsilon_1\}\\ &\leq \epsilon_3 + (T-1)\epsilon_1. \end{aligned} \tag{7.43}$$

Pick $\epsilon \geq \max\{\epsilon_3 + (T-1)\epsilon_1, \epsilon_3 + (T'-1)\epsilon_1\}$. Since $\epsilon_1, \epsilon_2, \epsilon_3$ and ϵ_4 (with $\epsilon_2 < \epsilon_3$) can be set arbitrarily small for sufficiently large n, we can set ϵ to be arbitrarily small as well for large n. ■

Lemma 6. If $\overline{\mathcal{W}}_1 \cap \overline{\mathcal{W}}_2 = \emptyset$ then for any $Z : \mathcal{X} \to \mathcal{Y}, Z_{Y|X} \in \overline{\mathcal{W}}_1$, any $V : \mathcal{X} \to \mathcal{Y}, V_{Y|X} \in \overline{\mathcal{W}}_2$, and any distribution P over $\mathcal{X}$ such that $P(a) > 0,\ \forall a \in \mathcal{X}$, there exists some $\eta > 0$ such that

$$\sup_{(a,b)\in\mathcal{X}\times\mathcal{Y}} \{|P(a)Z_{Y|X}(b|a) - P(a)V_{Y|X}(b|a)|\} > \eta.$$

In fact, instead of just $\overline{\mathcal{W}}_1$ and $\overline{\mathcal{W}}_2$, Lemma 6 holds for any two closed and disjoint sets of channels.

Lemma 7. $\overline{\mathcal{W}}_1 \cap \overline{\mathcal{W}}_2 = \emptyset$ is sufficient for compound-state identification under random coding.

Proof. Recall the definition of probability of error in the compound-state identification task. In this setting, there is no particular need or meaning in sending any 'message' since the decoder does not even try to decode the message. However, since there is a message term used in the error probability definition, we still need to describe the encoder in terms of messages. For our achievability scheme, consider an encoder which randomly samples a vector from $\boldsymbol{F} \in \tau_X$ (for some distribution $P_X \in \mathcal{P}^{(n)}_{\mathcal{X}}$) and for each message, it outputs the same vector $\boldsymbol{F}$, i.e., for any realisation of the encoder, the output is same for all the messages (this form of degenerate encoder is sufficient for proving the

lemma). Since the decoders knows which encoder is used (shared randomness), it knows the exact vector which is transmitted by the encoder. Represent the encoder output as $F(i) = \boldsymbol{F} \in \tau_X \; \forall i \in \mathcal{M}$.

Decoder. $G(\boldsymbol{y}) = \sigma_k$ if $\exists \; Z_{Y|X} \in \overline{\mathcal{W}}_k$ such that $(\boldsymbol{F}, \boldsymbol{y}) \in \tau^{\epsilon}_{XY}$ for $P_{XY} = P_X \times Z_{Y|X}$ and there exists no such $Z_{Y|X} \in \overline{\mathcal{W}}_{3-k}$.
Else arbitrarily set $G(\boldsymbol{y}) = \sigma_1$.

We specify ϵ later in this proof.

Probability of error in identification for the encoder-decoders described is given by

$$P^r_{id}(Q,k) = \max_{\boldsymbol{s} \in \mathcal{S}^n_k} \sum_{\boldsymbol{f}} |\tau_X|^{-1} \frac{1}{M} \sum_{m=1}^{M} W^n(\Phi^{-1}(\hat{\mathcal{E}}_k)|\boldsymbol{f}, \boldsymbol{s}) \tag{7.44}$$

$$= \max_{\boldsymbol{s} \in \mathcal{S}^n_k} |\tau_X|^{-1} \sum_{\boldsymbol{f}} W^n(\Phi^{-1}(\mathcal{E}_k)|\boldsymbol{f}, \boldsymbol{s}). \tag{7.45}$$

The error event $\hat{\mathcal{E}}_k$ can be due to 2 events -
(A) When no such $Z_{Y|X} \in \overline{\mathcal{W}}_k$ such that $(\boldsymbol{f}, \boldsymbol{y})$ is in the typical set.
(B) When there is a $V_{Y|X} \in \overline{\mathcal{W}}_{3-k}$ such that $(\boldsymbol{f}, \boldsymbol{y})$ is in the typical set.

For each $\boldsymbol{s}$, we now analyze these 2 cases.
(A):
By choosing $Z_{Y|X}$ as defined in Lemma 5, for any ϵ and sufficiently large n, the probability of this event can be made arbitrarily small.

(A)$^C \cap$ (B):
The event $(A)^C \cap (B)$ implies $\exists V_{Y|X} \in \overline{\mathcal{W}}_{3-k}$ such that $(\boldsymbol{f}, \boldsymbol{y}) \in \tau^{\epsilon}_{XY}$ for $P_{XY} = U_X \times V_{Y|X}$ and $\exists Z_{Y|X} \in \overline{\mathcal{W}}_k$ such that $(\boldsymbol{f}, \boldsymbol{y}) \in \tau^{\epsilon}_{XY}$ for $P_{XY} = U_X \times Z_{Y|X}$. Therefore,

$$|U(a)Z_{Y|X}(b|a) - U(a)V_{Y|X}(b|a)| < 2\epsilon \; \forall (a,b) \in \mathcal{X} \times \mathcal{Y}.$$

We can choose sufficiently small ϵ such that $\epsilon < \eta/2$ which would violate Lemma 6, implying that this case occurs with arbitrarily low probability.
Hence, $P^r_{id}(Q,k)$ can be made arbitrarily small for large n. Thus, we can identify the compound-state under random coding as stated in the theorem when $\overline{\mathcal{W}}_1 \cap \overline{\mathcal{W}}_2 = \emptyset$.

■

7.3 Achievability Proof of Theorem 3 (i) and Theorem 4 (i)

For achievability of both Theorem 3 (i) and Theorem 4 (ii), we use a similar encoding scheme. Let $\tilde{\boldsymbol{x}}$ be an $|\mathcal{X}|\log(n)$ length sequence consisting of $\log(n)$ repetitions of each symbol in $\mathcal{X}$. For Theorem 3 (i), a $(2^{nR'}, n')$, code $(F^{\mathsf{and}}, \Phi^{\mathsf{and}})$ consists of a length-n communication part and length $n' - n$ compound-state identification part where n is such that $n' = n + |\mathcal{X}|\log(n)$. The communication part of a code is given in terms encoder of Lemma 4 F_R, $R = \frac{R'n'}{n}$, and the identification part consists of the constant vector $\tilde{\boldsymbol{x}}$ as shown in Figure 4. Let Γ be a random and uniformly chosen permutation of length $n' = n + |\mathcal{X}|\log(n)$. The encoder $F^{\mathsf{and}}(i) = \Gamma(F_R(i), \tilde{\boldsymbol{x}})$, $i \in \{1, \ldots, 2^{nR}\}$. Note that the rate $R' = \frac{Rn}{n'}$ of the code is governed by R for large block length. For Theorem 4 (i), we use the same structure of the encoder but operate at a different rate R'. The encoder of a $(2^{nR'}, n')$, code $(F^{\mathsf{or}}, \Phi^{\mathsf{or}})$ is given by

$F^{\mathsf{or}}(i) = \Gamma(F_R(i), \tilde{\boldsymbol{x}})$, $i \in \{1, \ldots, 2^{nR}\}$ (R', R is different for F^{and} and F^{or}).

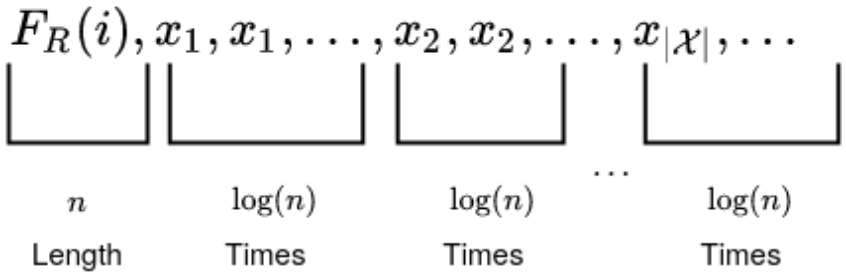

Figure 4: The vector $(F_R(i), \tilde{\boldsymbol{x}})$

Due to the shared randomness, the decoder knows the realisation of F_R and Γ. The decoder uses Γ to get back the original ordering, i.e., to get $(\hat{\boldsymbol{y}}, \tilde{\boldsymbol{y}}) = \Gamma^{-1}(\boldsymbol{y})$. Here, $\hat{\boldsymbol{y}}$ represents the vector corresponding to the first n symbols and $\tilde{\boldsymbol{y}}$ represent the vector corresponding to the last $|\mathcal{X}| \log(n)$ symbols of $\Gamma^{-1}(\boldsymbol{y})$. If the AVC-state sequence during transmission is represented as $\boldsymbol{s}$, then let $\boldsymbol{s}_a = [\Gamma^{-1}(\boldsymbol{s})]_1^n$ and $\boldsymbol{s}_b = [\Gamma^{-1}(\boldsymbol{s})]_{n+1}^{n'}$ - this notation is explained in the footnote[2].

Lemma 8. When $\overline{\mathcal{W}}_1 \cap \overline{\mathcal{W}}_2 = \emptyset$,

$$C_{\mathsf{and}}^{\mathsf{r}} \geq \max_{P_X} \min_{W \in \overline{\mathcal{W}}_1 \cup \overline{\mathcal{W}}_2} I(X;Y).$$

Proof. We use the encoding scheme described above and use Γ^{-1} at the decoder Φ^{and}, i.e., the decoder obtains $(\hat{\boldsymbol{y}}, \tilde{\boldsymbol{y}}) = \Gamma^{-1}(\boldsymbol{y})$. By the method described in Lemma 7, one can identify the compound-state as $G(\tilde{\boldsymbol{y}})$ (with $\boldsymbol{F}$ in the lemma being the vector $\tilde{\boldsymbol{x}}$) correctly w.h.p. for large block length. Note that this encoding scheme of shuffling $\tilde{\boldsymbol{x}}$ is equivalent to sending a vector from the typical set of the uniform distribution over $\mathcal{X}$ described in Lemma 7.

For any $R = \frac{n'R'}{n} < \max_{P_X} \min_{W \in \overline{\mathcal{W}}_1 \cup \overline{\mathcal{W}}_2} I(X;Y)$, we use the same decoder Φ_R used in Lemma 4 to decode the message. We obtain the message $\hat{m} = \Phi_R(\hat{\boldsymbol{y}})$ correctly w.h.p. Thus, using the $(2^{n'R'}, n')$ code, we can communicate at rate $R' = \frac{nR}{n'}$. For large block length, $R' \to R$. ∎

We now focus on proving achievability of Theorem 4. We present two lemmas before going into the main proof. The following Lemma is a well-known result and can be found in [17].

Lemma 9. An urn contains M white balls and $N - M$ black balls. If n balls are drawn uniformly without replacement and i represents the number of white balls drawn then, $\mathbb{E}[i] = n\frac{M}{N}$. Further, we can bound the deviations from the mean as shown,

$$\mathbb{P}[i \geq \mathbb{E}[i] + tn] \leq e^{-2t^2 n}, \tag{7.46}$$

$$\mathbb{P}[i \leq \mathbb{E}[i] - tn] \leq e^{-2t^2 n}, \tag{7.47}$$

$$\mathbb{P}[|i - \mathbb{E}[i]| \geq tn] \leq 2e^{-2t^2 n}. \tag{7.48}$$

Using Lemma 9, we obtain the following.

[2]For a sequence $\boldsymbol{y}$, we use the notation $[\boldsymbol{y}]_a^b$, $(b > a)$ to refer to the subsequence $(y_a, \ldots, y_b)$.

Lemma 10. Let random variable S be distributed as $P_{\boldsymbol{s}}$. Then

$$\mathbb{P}(\boldsymbol{s}_a \notin \tau_S^{\eta}) \leq 2\max\{|\mathcal{S}_1|, |\mathcal{S}_2|\} n^{-2\eta^2|\mathcal{X}|}\,, \tag{7.49}$$

$$\mathbb{P}(\boldsymbol{s}_b \notin \tau_S^{\eta}) \leq 2\max\{|\mathcal{S}_1|, |\mathcal{S}_2|\} e^{-2\eta^2 n}. \tag{7.50}$$

Proof. Since Γ shuffles randomly and uniformly, this follows directly from the definition of typicality and Lemma 9. The $\max\{.\}$ operator is present to ensure that the inequality is valid when $\boldsymbol{s}$ belongs to either of the two compound-state. ■

Lemma 10 shows that the AVC-state sequence vector corresponding to the identification part and the communication part have roughly the same type as the entire vector $\boldsymbol{s}$.

Lemma 11.

$$C_{\mathsf{or}}^{\mathsf{r}} \geq \max_{P_X} \min_{W \in \overline{\mathcal{W}}_1 \cap \overline{\mathcal{W}}_2} I(X;Y) \tag{7.51}$$

Proof. We use the encoding scheme described after Lemma 4. We specify the rate R of communication corresponding to the communication part later. Let the encoder-decoder pair for the $(2^{n'R'}, n')$, $R' = nR/n'$ code be $(F^{\mathsf{or}}, \Phi^{\mathsf{or}})$. Note that if $\overline{\mathcal{W}}_1 \cap \overline{\mathcal{W}}_2 = \emptyset$ then we can use the adversary identification scheme as described in Lemma 7 to achieve infinite capacity using $(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}})$. If $\overline{\mathcal{W}}_1 \cap \overline{\mathcal{W}}_2 \neq \emptyset$, then we first use a communication decoder $\tilde{\Phi}_R : \mathcal{Y}^n \to \mathcal{M} \cup \perp$ described below to decode the message.

The codebook for a $(2^{n'R'}, n')$ code is obtained by uniformly and independently sampling $M = 2^{n'R'}$ vectors $(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_M)$ in τ_X with some $P_X \in \mathcal{P}_{\mathcal{X}}^{(n)}$ and $F^{\mathsf{or}}(i) = (\boldsymbol{X}_i, \tilde{\boldsymbol{x}})$. The decoder outputs $\tilde{\Phi}_R(\boldsymbol{y}) = i \in \mathcal{M}$ if there is a unique i for which $I(X;Y) \geq R + \delta$ where $P_{XY} = P_{\boldsymbol{X}_i, \boldsymbol{y}}$, and $\tilde{\Phi}_R(\boldsymbol{y}) = \perp$ if no such i exists.

We show that the communication decoder correctly decodes the message w.h.p. (with high probability) for a certain class of adversarial attacks. For other attacks, we show that the decoder may output the correct message or output $\perp$ but it would not decode to a wrong message w.h.p. On receiving an error ($\perp$), a second decoder - compound-state decoder - would be used to identify the compound-state.

Suppose the compound-state is σ_k and the adversary operates with AVC-state sequence $\boldsymbol{s} \in \mathcal{S}_k^n$. Let dummy random variable $S \sim P_{\boldsymbol{s}}$. Let $||P_X||$ denote the max norm of a distribution - $\max_x P_X(x)$.
We use $\hat{\boldsymbol{y}}$ (defined in the text following Lemma 7) and $\tilde{\Phi}_R$ for decoding the message.
Define the set $\mathcal{P}_0 = \{P \in \mathcal{P}_1 \cup \mathcal{P}_2 : W_P \in \overline{\mathcal{W}}_1 \cap \overline{\mathcal{W}}_2\}$. Let $\mathcal{P}_\epsilon^+ = \{P \in \mathcal{P}_1 \cup \mathcal{P}_2 : \exists P' \in \mathcal{P}_0,\ ||P - P'|| \leq \epsilon\}$ and let $\mathcal{W}_\epsilon^- = \{W_P : P \in \mathcal{P}_\epsilon^+\}$. Also, define $\mathcal{W}_\epsilon^+ = \overline{\mathcal{W}_\epsilon^-}$. Note that $\mathcal{W}_\epsilon^+ = \overline{\mathcal{W}}_1 \cap \overline{\mathcal{W}}_2$ when $\epsilon = 0$ ($\overline{\mathcal{W}}_1 \cap \overline{\mathcal{W}}_2$ is already a closed convex set).
Let $R < \min_{W \in \mathcal{W}_\epsilon^+} I(X;Y)$ and let $\delta > 0$ be small enough such that $R + \delta < \min_{W \in \mathcal{W}_\epsilon^+} I(X;Y)$.

Case (A): $P_S \in \mathcal{P}_0$
W.h.p., $\boldsymbol{s}_a \in \tau_S^{\eta}$ by Lemma 10 for sufficiently large n - i.e., $||P_{\boldsymbol{s}_a} - P_S|| \leq \eta$ w.h.p. Set the value of $\eta < \epsilon$. Thus, it is equivalent to communication over the expanded CAVC $\mathcal{W}_\epsilon^+$ (i.e., closure of both families of channels for the CAVC is same and equal to $\mathcal{W}_\epsilon^+$) so we get arbitrarily small error in message decoding. In particular, let $\epsilon = 3\eta$.

Case (B): $P_S \notin \mathcal{P}_0$
We further divide this case into two sub-cases:
i) $P_{\boldsymbol{s}_a} \in \mathcal{P}_\epsilon^+$: Similar to Case (A), message decoding is correct and successful w.h.p.

ii) $P_{\boldsymbol{s}_a} \notin \mathcal{P}_\epsilon^+$: Note that since $\boldsymbol{s}_a \in \tau_S^\eta$ whp and $P_{\boldsymbol{s}_a} \notin \mathcal{P}_\epsilon^+$, we can see that that $P_{\boldsymbol{s}_b} \notin \mathcal{P}_0$ whp. In fact, the following is also true

$$\forall P \in \mathcal{P}_1 \cup \mathcal{P}_2,\ ||P - P_{\boldsymbol{s}_b}|| \leq \frac{\eta}{2} \implies P \notin \mathcal{P}_0.$$

Also, note that (7.106) still remains valid even if $W_{P_{\boldsymbol{s}_b}} \notin \mathcal{W}_\epsilon^+$. In other words, for any attack vector $\boldsymbol{s}_b$, we still have (7.106) as it is a very low probability event that a codeword which wasn't transmitted has high mutual information with the received vector $\hat{\boldsymbol{y}}$. Hence, w.h.p. the message decoder would not output a wrong message - it may either decode correctly or declare $\perp$. If the decoder outputs $\perp$, then we identify the adversary by $G(\tilde{\boldsymbol{y}})$- since $P_{\boldsymbol{s}_b} \notin \mathcal{P}_0$, Lemma 6 holds so the proof of achievability of Lemma 7 holds as well.

Since ϵ can be made arbitrarily small, the lemma follows. ■

7.4 Achievability Proofs Under Deterministic Coding

Let,

$$\mathcal{C}_\eta = \{P_{XSY} : D(P_{XSY}||P_X \times P_S \times W_{Y|X,S}) \leq \eta, P_S \in \mathcal{P}_1 \cup \mathcal{P}_2\}, \tag{7.52}$$

and let

$$I(P) = \min_{W \in \overline{\mathcal{W}}_1 \cup \overline{\mathcal{W}}_2, P_X = P} I(X;Y).$$

The following two lemmas establish the fact that the capacity expressions are indeed positive when the claimed necessary conditions are met.

Lemma 12. If the channel is non-any-symmetrizable, then $\min_{W \in \overline{\mathcal{W}}_1 \cup \overline{\mathcal{W}}_2} I(X;Y) > 0$ for all P_X such that $P_X(x) > 0\ \forall x \in \mathcal{X}$.

Proof. Suppose the statement is false, then there exists P_X and $P_S \in \mathcal{P}_1 \cup \mathcal{P}_2$ for which $I(X;Y) = 0$. Hence, there exists distribution $P_{XSY} \in \mathcal{C}_0$ such that X and Y are independent, i.e., $P_{Y|X}(y|x) = \sum_s W(y|x,s)P_S(s) = P_Y(y)\ \forall x, y$. The C-AVC is cis-symmetrizable in a trivial manner using $U(.|x) = P_S(.)$ in (3.1), a contradiction. ■

Lemma 13. If the channel is non-trans-symmetrizable, then $\min_{W \in \overline{\mathcal{W}}_1 \cap \overline{\mathcal{W}}_2} I(X;Y) > 0$ for all P_X such that $P_X(x) > 0\ \forall x \in \mathcal{X}$.

Proof. If $\overline{\mathcal{W}}_1 \cap \overline{\mathcal{W}}_2 = \emptyset$ then the lemma is trivially true. If $\overline{\mathcal{W}}_1 \cap \overline{\mathcal{W}}_2 \neq \emptyset$ then, let $\mathcal{P}_0 = \{P \in \mathcal{P}_1 \cup \mathcal{P}_2 : W_P \in \overline{\mathcal{W}}_1 \cap \overline{\mathcal{W}}_2\}$. Suppose the statement is false, then there exists P_X and $P_S \in \mathcal{P}_0$ for which $I(X;Y) = 0$. Hence, there exists distribution $P_{XSY} \in \mathcal{C}_0$ such that X and Y are independent, i.e., $P_{Y|X}(y|x) = \sum_s W(y|x,s)P_S(s) = P_Y(y)\ \forall x, y$. If $P_S \in \mathcal{P}_k$, then there exists $P_{S'} \in \mathcal{P}_{3-k}$ such that $W_{P_S} = W_{P_{S'}}$ as $\overline{\mathcal{W}}_1 \cap \overline{\mathcal{W}}_2 \neq \emptyset$. The C-AVC is trans-symmetrizable in a trivial manner using $U(.|x) = P_S(.)$ and $V(.|x) = P_{S'}(.)$ in (3.2), a contradiction. ■

For the achievability arguments, we describe some lemmas below. We first present a lemma based on [4, Lemma 3].

Lemma 14. For any $\epsilon > 0$, $n \geq n_0(\epsilon)$, $N \geq \exp(n\epsilon)$, and type P, there exists codewords $\boldsymbol{x}_1, \boldsymbol{x}_2, .., \boldsymbol{x}_N$ in $\mathcal{X}^n$, each of type P, such that for every $\boldsymbol{x} \in \mathcal{X}^n$, $\boldsymbol{s} \in \mathcal{S}_1^n \cup \mathcal{S}_2^n$, and every joint type $P_{XX'S}$ (with $P_S \in \mathcal{P}_1 \cup \mathcal{P}_2$), upon setting $R = \frac{1}{n} \log N$, we have:

$$\left| \left\{ j : (\boldsymbol{x}, \boldsymbol{x}_j, \boldsymbol{s}) \in \tau_{XX'S} \right\} \right| \leq \exp\left\{ n \left(|R - I(X'; XS)|^+ + \epsilon \right) \right\}; \tag{7.53}$$

$$\frac{1}{N} \left| \{ i : (\boldsymbol{x}_i, \boldsymbol{s}) \in \tau_{XS} \} \right| \leq \exp(-n\epsilon/2), \text{ if } I(X;S) > \epsilon; \tag{7.54}$$

$$\frac{1}{N} \left| \left\{ i : (\boldsymbol{x}_i, \boldsymbol{x}_j, \boldsymbol{s}) \in \tau_{XX'S} \text{ for some } j \neq i \right\} \right| \leq \exp(-n\epsilon/2), \text{ if } I(X;X'S) - |R - I(X';S)|^+ > \epsilon. \tag{7.55}$$

Proof. One can directly use [4, Lemma 3] to get the above result for a wider class of attacks by letting $\boldsymbol{s} \in (\mathcal{S}_1 \cup \mathcal{S}_2)^n$. ■

Based on [4, Lemma 4], we state the following two lemmas.

Lemma 15. If the CAVC is non-trans-symmetrizable and $\beta > 0$, then for a sufficiently small η, no quintuple of random variables X, X', S, S', Y, with $P_S \in \mathcal{P}_1$ and $P_{S'} \in \mathcal{P}_2$, can simultaneously satisfy

$$P_X = P_{X'} = P \text{ with } \min_{a \in \mathcal{X}} P(a) \geq \beta \tag{7.56}$$

$$P_{XSY} \in \mathcal{C}_\eta, P_{X'S'Y} \in \mathcal{C}_\eta \tag{7.57}$$

$$I(XY; X'|S) \leq \eta, I(X'Y; X|S') \leq \eta. \tag{7.58}$$

Proof. Suppose there exists X, X', S, S', Y which simultaneously satisfy the three conditions. Then, by definition of $\mathcal{C}_\eta$,

$$D(P_{XSY} || P_X \times P_S \times W) = \sum_{x,s,y} P_{XSY}(x,s,y) \log \frac{P_{XSY}(x,s,y)}{P_X(x) P_S(s) W(y|x,s)} \leq \eta.$$

Adding $I(XY; X'|S)$ to it,

$$D(P_{XX'SY} || P_X \times P_{X'} \times P_{S|X'} \times W) \leq 2\eta.$$

Projecting both the distributions to $\mathcal{X} \times \mathcal{X} \times \mathcal{Y}$, the divergence can not increase,

$$D(P_{XX'Y} || P_X \times P_{X'} \times V) \leq 2\eta$$

where $V(y|x,x') = \sum_s W(y|x,s) P_{S|X'}(s|x')$. By Pinsker's inequality,

$$\sum_{x,x',y} |P_{XX'Y}(x,x',y) - P(x)P(x')V(y|x,x')| \leq c\sqrt{2\eta}. \tag{7.59}$$

Similarly, starting with $P_{X'S'Y} \in \mathcal{C}_\eta$ and $I(X'Y; X|S') \leq \eta$, we get

$$\sum_{x,x',y} |P_{XX'Y}(x,x',y) - P(x)P(x')V'(y|x,x')| \leq c\sqrt{2\eta} \tag{7.60}$$

where $V'(y|x,x') = \sum_s W(y|x',s) P_{S'|X}(s|x)$. From (7.59) and (7.60),

$$\max_{x,x',y} |V(y|x,x') - V'(y|x,x')| \leq \frac{2c\sqrt{2\eta}}{\beta^2}. \tag{7.61}$$

For a non-trans-symmetrizable CAVC, there exists a ξ such that

$$\max_{x,x',y} |\sum_s W(y|x,s)U_{S|X}(s|x') - \sum_s W(y|x',s)V_{S|X}(s|x)| \geq \xi \tag{7.62}$$

for every $U_{S|X} \in \mathcal{P}_{A|X}$, $V_{S|X} \in \mathcal{P}_{B|X}$. Setting $U_{S|X'} = P_{S|X}, V_{S|X} = P_{S'|X}$ and $\eta < \frac{\xi^2\beta^4}{8c^2}$, we get a contradiction. ■

Lemma 16. If the CAVC is non-any-symmetrizable and $\beta > 0$, then for a sufficiently small η, no quintuple of random variables X, X', S, S', Y, with $P_S, P_{S'} \in \mathcal{P}_1 \cup \mathcal{P}_2$, can simultaneously satisfy

$$P_X = P_{X'} = P \text{ with } \min_{a\in\mathcal{X}} P(a) \geq \beta \tag{7.63}$$

$$P_{XSY} \in \mathcal{C}_\eta, P_{X'S'Y} \in \mathcal{C}_\eta \tag{7.64}$$

$$I(XY;X'|S) \leq \eta, I(X'Y;X|S') \leq \eta. \tag{7.65}$$

Lemma 16 can be proved in a similar manner as Lemma 15.

Lemma 17. If the CAVC is non-any-symmetrizable and $\overline{\mathcal{W}}_1 \cap \overline{\mathcal{W}}_2 = \emptyset$ then

$$C_{\mathsf{and}}^{\mathsf{d}} \geq \max_{P_X} \min_{W\in\overline{\mathcal{W}}_1\cup\overline{\mathcal{W}}_2} I(X;Y).$$

Proof. The decoder we use for achieving the capacity is described below for η described later.

Decoder. Given codewords $\boldsymbol{x}_j$, $j = 1,\ldots,M$, set $\phi^{\mathsf{and}}(\boldsymbol{y}) = (i,\sigma_k)$, $i \in \mathcal{M}, k \in \{1,2\}$, iff an $\boldsymbol{s} \in \mathcal{S}_k^n$ exists such that:

1. the joint type $P_{\boldsymbol{x}_i,\boldsymbol{s},\boldsymbol{y}} \in \mathcal{C}_\eta$, and
2. for each $\boldsymbol{x}_j, j \neq i$ such that there exists $\boldsymbol{s}' \in \mathcal{S}_1^n \cup \mathcal{S}_2^n$, $P_{\boldsymbol{x}_j,\boldsymbol{s}',\boldsymbol{y}} \in \mathcal{C}_\eta$, we have $I(XY;X'|S) \leq \eta$ where $P_{XX'SY} = P_{\boldsymbol{x}_i,\boldsymbol{x}_j,\boldsymbol{s},\boldsymbol{y}}$.

Set $\phi^{\mathsf{and}}(\boldsymbol{y}) = (1,\sigma_1)$ if no such (i,σ_k) exists.

First, we justify the consistency of the decoder - if (i,σ_k) satisfies both the conditions then $(i',\sigma_{k'}), (i',k') \neq (i,k)$ can not satisfy the conditions. Consider the following three cases

1. $i \neq i'$, $k \neq k'$, or
2. $i \neq i'$, $k = k'$, or
3. $i = i'$, $k \neq k'$.

Case (1) can not occur as by Lemma 16 (Lemma 15 can also be used), as it is impossible that first and second condition of decoder holds for both tuples (i,k) and (i',k').

Case (2) can not occur because of the same reason mentioned above.

Case (3) can not occur due to $\overline{\mathcal{W}}_1 \cap \overline{\mathcal{W}}_2 = \emptyset$. If case (3) was true then $(\boldsymbol{x},\boldsymbol{s},\boldsymbol{y}) \in \mathcal{C}_\eta$ and $(\boldsymbol{x},\boldsymbol{s}',\boldsymbol{y}) \in \mathcal{C}_\eta$. Let X,S,S',Y be random variables defined by $(\boldsymbol{x},\boldsymbol{s},\boldsymbol{s}',\boldsymbol{y}) \in \tau_{XSS'Y}$. Using Pinsker's inequality, the definition of $\mathcal{C}_\eta$ and the fact that divergence won't increase if we project P_{XSY} and $P_X \times P_S \times W$ on $\mathcal{X} \times \mathcal{Y}$,

$$\sum_{a,c} |P_{XY}(a,c) - \sum_b P_X(a)P_S(b)W(c|a,b)| \leq c\sqrt{\eta} \tag{7.66}$$

$$\sum_{a,c} |P_{XY}(a,c) - \sum_{b} P_X(a)P_{S'}(b)W(c|a,b)| \leq c\sqrt{\eta} \tag{7.67}$$

$$\sum_{a,c} |P_X(a)U(c|a) - P_X(a)V(c|a)| \leq 2c\sqrt{\eta}, \tag{7.68}$$

where $U(c|a) := \sum_b P_S(b)W(c|a,b) \in \mathcal{W}_1$ and similarly $V(c|a) \in \mathcal{W}_2$. If $\min_a P_X(a) = \beta$ then

$$\max_{a,c} |U(c|a) - V(c|a)| \leq \frac{2c\sqrt{\eta}}{\beta}. \tag{7.69}$$

However, we know that $\overline{\mathcal{W}}_1$ and $\overline{\mathcal{W}}_2$ are disjoint so (7.69) is not possible by setting η to be small enough and hence, a contradiction. Choose η sufficiently small so that (7.69) is not true and Lemma 15 and 16 are satisfied.

We need to show that the correct output indeed satisfies the decoding conditions with high probability. For this, we can show that the actual input sequence $\boldsymbol{x}$ and the AVC-state sequence $\boldsymbol{s}$ which was present in the transmission does indeed satisfy the decoder criteria. We prove this based on [4, Lemma 5].

For any arbitrarily small $\delta > 0$, choose R satisfying

$$I(P) - \delta < R < I(P) - \frac{2}{3}\delta. \tag{7.70}$$

Choose the codebook based on Lemma 14 with rate R and codewords $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M$. We analyze the error probability when the AVC-state sequence is $\boldsymbol{s} \in \mathcal{S}_t^n$ and the compound-state is σ_t, $t = 1, 2$. Since $\overline{\mathcal{W}}_1 \cap \overline{\mathcal{W}}_2 = \emptyset$, we can define the probability of error under AVC-state sequence $\boldsymbol{s}$ as shown below

$$P_e^d(f, \phi, \boldsymbol{s}) = \frac{1}{M}\sum_{i=1}^{M} W^n(\phi^{-1}(\{i, \sigma_t\})^C | \boldsymbol{x}_i, \boldsymbol{s}) \tag{7.71}$$

$$= \frac{1}{M}\sum_{i=1}^{M} \sum_{\boldsymbol{y}:\phi(\boldsymbol{y}) \neq (i,\sigma_t)} W^n(\boldsymbol{y}|\boldsymbol{x}_i, \boldsymbol{s}). \tag{7.72}$$

By (7.54),

$$\frac{1}{M}|\{i : (\boldsymbol{x}_i, \boldsymbol{s}) \in \bigcup_{I(X;S)>\epsilon} \tau_{XS}\}| \leq (\text{no. of joint types}).\exp(-n\epsilon/2) \tag{7.73}$$

$$\leq \exp(-n\epsilon/3), \tag{7.74}$$

for suitably large n, which depends on the choice of ϵ which is specified later. Therefore, it suffices to only consider codewords $\boldsymbol{x}_i$ for which $(\boldsymbol{x}_i, \boldsymbol{s}) \in \tau_{XS}$ with $I(X;S) \leq \epsilon$. If $P_{XSY} \notin \mathcal{C}_\eta$ then,

$$D(P_{XSY}||P_{XS} \times W) = D(P_{XSY}||P_X \times P_S \times W) - I(X;S) \tag{7.75}$$

$$> \eta - \epsilon. \tag{7.76}$$

Thus,

$$\sum_{\boldsymbol{y} \in \tau_{Y|XS}(\boldsymbol{x}_i, \boldsymbol{s})} W^n(\boldsymbol{y}|\boldsymbol{x}_i, \boldsymbol{s}) \leq \exp(-nD(P_{XSY}||P_{XS} \times W)$$

$$< \exp(-n(\eta - \epsilon)).$$

$$\therefore \frac{1}{M}\sum_{i=1}^{M} \sum_{\boldsymbol{y}:P_{\boldsymbol{x}_i,\boldsymbol{s},\boldsymbol{y}} \notin \mathcal{C}_\eta} W^n(\boldsymbol{y}|\boldsymbol{x}_i, \boldsymbol{s}) \leq \exp(-n(\eta - 2\epsilon)) \tag{7.77}$$

Now, if $P_{\boldsymbol{x}_i,\boldsymbol{s},\boldsymbol{y}} \in \mathcal{C}_\eta$ and yet $\phi(\boldsymbol{y}) \neq (i, \sigma_t)$, then condition (2) of the decoder must be getting violated. Let $\mathcal{D}_\eta$ be the

set of all joint distributions $P_{XX'SY}$ such that 1) $P_{XSY} \in \mathcal{C}_\eta$; 2) $P_{X'S'Y} \in \mathcal{C}_\eta$; 3) $I(XY;X'|S) > \eta$ (and $\boldsymbol{x} \neq \boldsymbol{x}'$). Then,

$$\sum_{\substack{\boldsymbol{y}: P_{\boldsymbol{x}_i,\boldsymbol{s},\boldsymbol{y}} \in \mathcal{C}_\eta; \\ \phi(\boldsymbol{y}) \neq (i,\sigma_t)}} W^n(\boldsymbol{y}|\boldsymbol{x}_i,\boldsymbol{s}) \leq \sum_{P_{XX'SY} \in \mathcal{D}_\eta} e_{XX'SY}(i,\boldsymbol{s}) \tag{7.78}$$

where

$$e_{XX'SY}(i,\boldsymbol{s}) = \sum_{\substack{\boldsymbol{y}: (\boldsymbol{x}_i,\boldsymbol{x}_j,\boldsymbol{s},\boldsymbol{y}) \in \tau_{XX'SY} \\ \text{for some } j \neq i}} W^n(\boldsymbol{y}|\boldsymbol{x}_i,\boldsymbol{s}). \tag{7.79}$$

Combining the equations so far, we have

$$P_e^d(f,\phi,\boldsymbol{s}) \leq \exp(-n\epsilon/3) + \exp(-n(\eta - 2\epsilon)) + \frac{1}{M}\sum_{i=1}^{M} \sum_{P_{XX'SY} \in \mathcal{D}_\eta} e_{XX'SY}(i,\boldsymbol{s}). \tag{7.80}$$

Notice that because of (7.55) it suffices to deal with cases when $P_{XX'SY} \in \mathcal{D}_\eta$ satisfies

$$I(X;X'S) \leq |R - I(X';S)|^+ + \epsilon. \tag{7.81}$$

From (7.79),

$$e_{XX'SY}(i,\boldsymbol{s}) \leq \sum_{j:(\boldsymbol{x}_i,\boldsymbol{x}_j,\boldsymbol{s}) \in \tau_{XX'S}} \; \sum_{\boldsymbol{y} \in \tau_{Y|XX'S}(\boldsymbol{x}_i,\boldsymbol{x}_j,\boldsymbol{s})} W^n(\boldsymbol{y}|\boldsymbol{x}_i,\boldsymbol{s}). \tag{7.82}$$

Using the fact that $W^n(\boldsymbol{y}|\boldsymbol{x}_i,\boldsymbol{s})$ is a constant upper bounded by $(|\tau_{Y|XS}(\boldsymbol{x}_i,\boldsymbol{s})|)^{-1}$, the inner sum is upper bounded by $|\tau_{Y|XX'S}(\boldsymbol{x}_i,\boldsymbol{x}_j,\boldsymbol{s})|/|\tau_{Y|XS}(\boldsymbol{x}_i,\boldsymbol{s})| \leq \exp\{-n(I(Y;X'|XS) - \epsilon)\}$. Hence, using (7.53),

$$e_{XX'SY}(i,\boldsymbol{s}) \leq \exp\{-n\left(I(Y;X'|XS) - |R - I(X';XS)|^+ - 2\epsilon\right)\}. \tag{7.83}$$

We can split the problem into two cases:

1. $R \leq I(X';S)$, or,
2. $R > I(X';S)$.

Case (1) and (7.81) yields

$$I(X;X'|S) \leq I(X;X'S) \leq \epsilon, \tag{7.84}$$

and by condition (3) in definition of $\mathcal{D}_\eta$,

$$I(Y;X'|XS) \geq \eta - \epsilon. \tag{7.85}$$

Since $R \leq I(X';S) \leq I(X';XS)$, it follows from (7.83) that

$$e_{XX'SY}(i,\boldsymbol{s}) \leq \exp(-n(\eta - 3\epsilon)). \tag{7.86}$$

For case (2), from (7.81), we get

$$\begin{aligned} R &> I(X;X'S) + I(X';S) - \epsilon \\ &= I(X';XS) + I(X;S) - \epsilon \\ &\geq I(X';XS) - \epsilon, \end{aligned} \tag{7.87}$$

and hence,

$$|R - I(X'; XS)|^+ \geq R - I(X'; XS) - \epsilon.$$

Substituting in (7.83)

$$e_{XX'SY}(i, \boldsymbol{s}) \leq \exp\{-n(I(X'; XSY) - R - 3\epsilon)\} \tag{7.88}$$
$$\leq \exp\{-n(I(X'; Y) - R - 3\epsilon)\}. \tag{7.89}$$

$P_{XX'SY} \in \mathcal{D}_\eta$ implies that $P_{X'S'Y} \in \mathcal{C}_\eta$ for some S'. Thus, by definition of $\mathcal{C}_\eta$, $P_{X'S'Y}$ is arbitrarily close to $P_{X''S''Y} \in \mathcal{C}_0$ defined by $P_{X''S''Y''} = P_X \times P_{S'} \times W$ if η is sufficiently small. This implies $I(X'; Y)$ is arbitrarily close to $I(X''; Y'')$, i.e., $I(X' : Y) \geq I(X''; Y'') - \delta/3$. By definition of $I(P)$ and assumption (7.70),

$$I(X'; Y) - R \geq I(P) - \delta/3 - R \geq \delta/3$$

if η is sufficiently small and depends only on δ (and $\overline{\mathcal{W}}_1, \overline{\mathcal{W}}_2$). Therefore, for case (2),

$$e_{XX'SY}(i, \boldsymbol{s}) \leq \exp\{-n(\frac{\delta}{3} - 3\epsilon)\}$$

Therefore,

$$P_e^d(f, \phi, \boldsymbol{s}) \leq \exp(-n\epsilon/4)$$

if $\epsilon \leq \min(\eta/4, \delta/10)$ and n sufficiently large for all $\boldsymbol{s}$. ■

Lemma 18. If the CAVC is non-any-symmetrizable then

$$C_{\mathsf{com}}^{\mathsf{d}} \geq \max_{P_X} \min_{W \in \overline{\mathcal{W}}_1 \cup \overline{\mathcal{W}}_2} I(X; Y).$$

Proof. The proof is analogous to the proof of Lemma 17. We use Lemma 14 to get a codebook with type P_X which maximizes $I(P)$ and use the following decoder to obtain the message.

Decoder. Given codewords $\boldsymbol{x}_j$, $j = 1, \ldots, M$, set $\phi(\boldsymbol{y}) = i$, $i \in \mathcal{M}$, iff an $\boldsymbol{s}$ exists such that:

1. the joint type $P_{\boldsymbol{x}_i, \boldsymbol{s}, \boldsymbol{y}} \in \mathcal{C}_\eta$, and
2. for each $\boldsymbol{x}_j$, $j \neq i$ such that there exists $\boldsymbol{s}'$, $P_{\boldsymbol{x}_j, \boldsymbol{s}', \boldsymbol{y}} \in \mathcal{C}_\eta$, we have $I(XY; X'|S) \leq \eta$ where $P_{XX'SY} = P_{\boldsymbol{x}_i, \boldsymbol{x}_j, \boldsymbol{s}, \boldsymbol{y}}$.

Set $\phi(\boldsymbol{y}) = 1$ if no such i exists.

■

Next, we show that a positive rate is attainable for 'communication or compound-state identification' if the CAVC is non-trans-symmetrizable.

Lemma 19. If CAVC is non-trans-symmetrizable then $C_{\mathsf{or}}^{\mathsf{d}} > 0$.

Proof. Use Lemma 14 to obtain a codebook at some rate $R > 0$ (described later).

Decoder. Given codewords $\boldsymbol{x}_j$, $j = 1, \ldots, M$, let B_k ($k = 1, 2$) be the set of messages $m \in \mathcal{M}$ such that

1. $\exists \boldsymbol{s} \in \mathcal{S}_k^n$ such that $P_{\boldsymbol{x}_m, \boldsymbol{s}, \boldsymbol{y}} \in \mathcal{C}_\eta$, and
2. for every $m' \neq m$ such that $\exists\ \boldsymbol{s}' \in \mathcal{S}_{3-k}^n$, $P_{\boldsymbol{x}_{m'}, \boldsymbol{s}', \boldsymbol{y}} \in \mathcal{C}_\eta$, we have $I(XY; X'|S) \leq \eta$ where $P_{XX'SY} = P_{\boldsymbol{x}_m, \boldsymbol{x}_{m'}, \boldsymbol{s}, \boldsymbol{y}}$.

If $B_1 = B_2 = \{m\}$, then $\phi^{\mathsf{or}}(\boldsymbol{y}) = m$. If for some $k \in \{1, 2\}$, $B_k = \emptyset \neq B_{3-k}$, then the decoder outputs the compound state $\phi^{\mathsf{or}}(\boldsymbol{y}) = \sigma_{3-k}$.

By Lemma 15, it is not possible to have distinct messages in the sets B_1 and B_2. Thus, the only four possibilities are listed below

1. $B_1 = B_2 = \{m\}$, $m \in \mathcal{M}$,
2. $B_1 = \emptyset$, $|B_2| \geq 1$,
3. $B_2 = \emptyset$, $|B_1| \geq 1$, and
4. $B_1 = B_2 = \emptyset$.

Suppose the AVC-state sequence during the transmission is $\boldsymbol{s} \in \mathcal{S}_t^n$, $t \in \{1, 2\}$. Using the same approach as that of the proof of Lemma 17, we can show that the correct message would be present in the set B_t w.h.p. for sufficiently large block length. To see this, refer to the proof of Lemma 17 - proof till (7.77) remains the same. The slightly different decoder changes the error event slightly and we present the new condition below.

If $P_{\boldsymbol{x}_i, \boldsymbol{s}, \boldsymbol{y}} \in \mathcal{C}_\eta$ and yet $\phi(\boldsymbol{y}) \neq i$, then condition (2) of the decoder must be getting violated. Let $\mathcal{D}'_\eta$ be the set of all joint distributions $P_{XX'SY}$ such that 1) $P_{XSY} \in \mathcal{C}_\eta$; 2) $P_{X'S'Y} \in \mathcal{C}_\eta$, $P_{S'} \in \mathcal{P}_{3-t}$; 3) $I(XY; X'|S) > \eta$ (and $\boldsymbol{x} \neq \boldsymbol{x}'$). With this modified $\mathcal{D}'_\eta$ definition, the rest of the proof remains the same till equation (7.88) where we make a slight modification as shown below,

$$
\begin{aligned}
e_{XX'SY}(i, \boldsymbol{s}) &\leq \exp\{-n(I(X'; XSY) - R - 3\epsilon)\} && (7.90)\\
&\leq \exp\{-n(I(X'; XY|S) - R - 3\epsilon)\} && (7.91)\\
&\leq \exp\{-n(\eta - R - 3\epsilon)\}, && (7.92)
\end{aligned}
$$

where (7.92) follows from definition of $\mathcal{D}'_\eta$. Choose $0 < R = \epsilon < \eta/5$. Therefore, $\mathsf{C}_{\mathsf{or}}^{\mathsf{d}} > 0$. ■

Lemma 20. $\overline{\mathcal{W}}_1 \cap \overline{\mathcal{W}}_2 = \emptyset$ and non-trans-symmetrizability of the CAVC is sufficient for compound-state identification.

Proof. The proof is similar to the proof of Lemma 19. We use Lemma 14 to get a codebook with a positive rate and use the following decoder to obtain the message. Note that although we are not even interested in decoding the messages, they still play a vital role by imparting stochasticity to the encoder.

Decoder. Given codewords $\boldsymbol{x}_j$, $j = 1, \ldots, M$, set $\phi(\boldsymbol{y}) = \sigma_k$, $k = 1, 2$, iff an $\boldsymbol{s} \in \mathcal{S}_k^n$ and a $\boldsymbol{x}$ in the codebook exists such that:

1. the joint type $P_{\boldsymbol{x}, \boldsymbol{s}, \boldsymbol{y}} \in \mathcal{C}_\eta$, and
2. for each $\boldsymbol{s}' \in \mathcal{S}_{3-k}^n$ and $\boldsymbol{x}'$ in the codebook such that $P_{\boldsymbol{x}', \boldsymbol{s}', \boldsymbol{y}} \in \mathcal{C}_\eta$, we have $I(XY; X'|S) \leq \eta$ where $P_{XX'SY} = P_{\boldsymbol{x}, \boldsymbol{x}', \boldsymbol{s}, \boldsymbol{y}}$.

Set $\phi(\boldsymbol{y}) = \sigma_1$ if no such i exists.

Note that for sufficiently small η, $\boldsymbol{x}'$ can not be equal to $\boldsymbol{x}$ in the decoder check as the occurrence of $\overline{\mathcal{W}}_1 \cap \overline{\mathcal{W}}_2 = \emptyset$ and $P_{\boldsymbol{x},\boldsymbol{s},\boldsymbol{y}}, P_{\boldsymbol{x},\boldsymbol{s}',\boldsymbol{y}} \in \mathcal{C}_\eta$ together yields a contradiction. ■

Lemma 21. If CAVC is non-trans-symmetrizable then

$$C_{\mathsf{or}}^{\mathsf{d}} \geq \max_{P_X} \min_{W \in \overline{\mathcal{W}}_1 \cap \overline{\mathcal{W}}_2} I(X;Y).$$

Proof. For some achievable rate R and block-length n under random coding, apply [16, Lemma 12.8] to show the existence of a random code distributed over $K = n^2$ encoder-decoder pairs uniformly. This small amount of shared randomness can be established using deterministic codes given by Lemma 19. Thus, we can show that $C_{\mathsf{or}}^{\mathsf{d}} = C_{\mathsf{or}}^{\mathsf{r}}$ when the CAVC is non-trans-symmetrizable. ■

7.5 Converses for Deterministic Coding

The converses of random coding results in Section 7.1 establish some of the converse results for deterministic coding.

Lemma 22. If CAVC is any-symmetrizable or $\overline{\mathcal{W}}_1 \cap \overline{\mathcal{W}}_2 \neq \emptyset$ then $C_{\mathsf{and}}^{\mathsf{d}} = 0$.

Proof. Let the codewords be $\boldsymbol{x}_1, .., \boldsymbol{x}_M$. For any distribution $R(\boldsymbol{s})$ over $\mathcal{S}_1^n$,

$$P_e^d(f, \phi, 1) \geq \sum_{\boldsymbol{s}} R(\boldsymbol{s}) P_e^d(f, \phi, \boldsymbol{s}). \tag{7.93}$$

Let $T^n(\boldsymbol{s}|\boldsymbol{x}) = \prod_i T(s_i|x_i)$ be some distribution specified later. Choose

$$R(\boldsymbol{s}) = \frac{1}{M} \sum_{i=1}^{M} T^n(\boldsymbol{s}|\boldsymbol{x}_i). \tag{7.94}$$

Then combining definition of $P_e^d(f, \phi, 1)$, (7.93), and (7.94),

$$P_e^d(f, \phi, 1) \geq \sum_{\boldsymbol{s}} \left(\frac{1}{M} \sum_{i=1}^{M} T^n(\boldsymbol{s}|\boldsymbol{x}_i) \right) \left(\frac{1}{M} \sum_{j=1}^{M} W^n(\phi^{-1}((j,\sigma_1))^C | \boldsymbol{x}_j, \boldsymbol{s}) \right) \tag{7.95}$$

$$= \frac{1}{M^2} \sum_{i=1}^{M} \sum_{j=1}^{M} \sum_{\boldsymbol{s}} T^n(\boldsymbol{s}|\boldsymbol{x}_i) W^n(\phi^{-1}((j,\sigma_1))^C | \boldsymbol{x}_j, \boldsymbol{s}) \tag{7.96}$$

$$\geq \frac{1}{M^2} \sum_{i=1}^{M} \sum_{j \neq i} \sum_{\boldsymbol{s}} T^n(\boldsymbol{s}|\boldsymbol{x}_i) W^n(\phi^{-1}((j,\sigma_1))^C | \boldsymbol{x}_j, \boldsymbol{s}). \tag{7.97}$$

We can have 3 cases:

(A) the CAVC is trans-symmetrizable, or,

(B) the CAVC is cis-symmetrizable, or,

(C) $\mathcal{W}_0 \neq \emptyset$.

For case (A), let $U(s|x)$ and $V(s|x)$ be the distributions satisfying trans-symmetrizability condition. Let $T(s|x) = U(s|x)$. By trans-symmetrizability condition on (7.97),

$$\frac{1}{M^2}\sum_{i=1}^{M}\sum_{j\neq i}\sum_{\boldsymbol{s}} U^n(\boldsymbol{s}|\boldsymbol{x}_i)W^n(\phi^{-1}((j,\sigma_1))^C|\boldsymbol{x}_j,\boldsymbol{s}) = \frac{1}{M^2}\sum_{i=1}^{M}\sum_{j\neq i}\sum_{\boldsymbol{s}} V^n(\boldsymbol{s}|\boldsymbol{x}_j)W^n(\phi^{-1}((j,\sigma_1))^C|\boldsymbol{x}_i,\boldsymbol{s}) \tag{7.98}$$

$$\geq \frac{1}{M^2}\sum_{i=1}^{M}\sum_{j\neq i}\sum_{\boldsymbol{s}} V^n(\boldsymbol{s}|\boldsymbol{x}_j)W^n(\phi^{-1}((i,\sigma_2))|\boldsymbol{x}_i,\boldsymbol{s})$$

$$= \frac{M-1}{M} - \frac{1}{M^2}\sum_{i=1}^{M}\sum_{j\neq i}\sum_{\boldsymbol{s}} V^n(\boldsymbol{s}|\boldsymbol{x}_j)W^n(\phi^{-1}((i,\sigma_2))^C|\boldsymbol{x}_i,\boldsymbol{s}) \tag{7.99}$$

$$\geq \frac{M-1}{M} - \frac{1}{M^2}\sum_{i=1}^{M}\sum_{j=1}^{M}\sum_{\boldsymbol{s}} V^n(\boldsymbol{s}|\boldsymbol{x}_j)W^n(\phi^{-1}((i,\sigma_2))^C|\boldsymbol{x}_i,\boldsymbol{s})$$

(7.100)

(note that $V^n(\boldsymbol{s}|\boldsymbol{x})$ is non-zero only over $\boldsymbol{s}\in\mathcal{S}_2^n$) (7.101)

$$= \frac{M-1}{M} - P_e^d(f,\phi,2). \tag{7.102}$$

$$\therefore\ P_e^d(f,\phi,1) + P_e^d(f,\phi,2) \geq \frac{M-1}{M}. \tag{7.103}$$

$$\implies P_e^d(f,\phi) \geq \frac{M-1}{2M}. \tag{7.104}$$

Similarly, for case (B), let $U(s|x)$ and $V(s|x)$ be the distributions satisfying cis-symmetrizability condition (without loss of generality we assume σ_1-symmetrizable). Let $T(s|x) = U(s|x)$. By performing similar steps, one can get the following inequality

$$P_e^d(f,\phi,1) \geq \frac{M-1}{2M}.$$

$$\therefore\ P_e^d(f,\phi) \geq \frac{M-1}{2M}.$$

For case (C), say $Z_{Y|X} \in \overline{\mathcal{W}}_1 \cap \overline{\mathcal{W}}_2$. Let $P_k(s)$ be a distribution over $\mathcal{S}_k$ be such that $\sum_s P_k(s)W_{Y|X,S=s} = Z_{Y|X}$. Set $T(s|x) = P_1(s)$. Simplifying (7.96), we get

$$P_e^d(f,\phi,1) \geq \frac{1}{M}\sum_{i=1}^{M} Z^n(\phi^{-1}((i,\sigma_1))^C|\boldsymbol{x}_i).$$

Similarly, setting $T(s|x) = P_2(s)$, we get,

$$P_e^d(f,\phi,2) \geq \frac{1}{M}\sum_{i=1}^{M} Z^n(\phi^{-1}((i,\sigma_2))^C|\boldsymbol{x}_i).$$

Adding both,

$$P_e^d(f,\phi,1) + P_{es}(f,\phi,2) \geq 1.$$

$$\therefore\ P_e^d(f,\phi) \geq \frac{1}{2}.$$

Therefore, non-any-symmetrizability and $\overline{\mathcal{W}}_1 \cap \overline{\mathcal{W}}_2 = \emptyset$ is necessary for non-zero rate of communication and compound-state identification. ■

Similar steps can be performed to show that any-symmetrizability implies $C_{\mathsf{com}}^{\mathsf{d}} = 0$.

Lemma 23. If CAVC is trans-symmetrizable then $C_{\mathsf{or}}^{\mathsf{d}} = 0$.

Steps similar to proof of Lemma 22 can be used to show that trans-symmetrizability leads to the condition $P_e^d(f,\phi) \geq \frac{M-1}{2M}$.

7.6 Proof of Theorem 5

Lemma 24. The rate-exponent pair $(C(Q_{UX}), \gamma(Q_{UX}))$ is achievable.

Proof. Choose R and $\delta > 0$ such that $R + \delta < C(Q_{UX})$.

Encoding: Sample an n-length sequence $\boldsymbol{u}$ uniformly from $\tau_U, U \sim Q_U$. Sample uniformly 2^{nR} codewords $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_M$ from $\tau_{X|U}(\boldsymbol{u}), UX \sim Q_{UX}$. The codebook and $\boldsymbol{u}$ is shared with the decoder and the adversaries are unaware of it. $F(i) = \boldsymbol{X}_i$.

Message Decoding: Let the received vector be $\boldsymbol{y}$. $\Phi_m(\boldsymbol{y}) = i$ if $\boldsymbol{X}_i$ is the unique codeword such that $I(X;Y|U) \geq R + \delta$ for $(\boldsymbol{u}, \boldsymbol{X}_i, \boldsymbol{y}) \in \tau_{UXY}$. If message i is sent and the state sequence is $\boldsymbol{s}$ during transmission, we need to prove the following two results to show that rate R is achievable -

$$\mathbb{P}\{(\boldsymbol{u}, \boldsymbol{X}_i, \boldsymbol{y}) \in \tau_{XY}, I(X;Y|U) < R + \delta\} \to 0 \ \ \forall \boldsymbol{s} \in \mathcal{S}_1^n \cup \mathcal{S}_2^n \tag{7.105}$$

$$\mathbb{P}\{(\boldsymbol{u}, \boldsymbol{X}_j, \boldsymbol{y}) \in \tau_{XY}, I(X;Y|U) \geq R + \delta, \text{for some } j \neq i\} \to 0 \ \ \forall \boldsymbol{s} \in \mathcal{S}_1^n \cup \mathcal{S}_2^n \tag{7.106}$$

Probability expression in (7.105) is equal to

$$\sum_{\substack{P_{UXSY}: I(X;Y|U)<R+\delta, \\ \boldsymbol{s}\in\tau_S, \\ P_{UX}=Q_{UX}}} \sum_{\boldsymbol{u}\in\tau_{U|S}(\boldsymbol{s})} |\tau_U|^{-1} \sum_{\boldsymbol{x}\in\tau_{X|US}(\boldsymbol{u},\boldsymbol{s})} |\tau_{X|U}(\boldsymbol{u})|^{-1} \sum_{\boldsymbol{y}\in\tau_{Y|UXS}(\boldsymbol{u},\boldsymbol{x},\boldsymbol{s})} W^n(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{s}) \tag{7.107}$$

$$\leq \sum_{\substack{P_{UXSY}: I(X;Y|U)<R+\delta, \\ \boldsymbol{s}\in\tau_S, \\ P_{UX}=Q_{UX}}} \sum_{\boldsymbol{u}\in\tau_{U|S}(\boldsymbol{s})} |\tau_U|^{-1} \sum_{\boldsymbol{x}\in\tau_{X|US}(\boldsymbol{u},\boldsymbol{s})} |\tau_{X|U}(\boldsymbol{u})|^{-1} \exp\{-nD(P_{UXSY}||P_{UXS} \times W_{Y|X,S})\} \tag{7.108}$$

$$\leq \sum_{\substack{P_{UXSY}: I(X;Y|U)<R+\delta, \\ \boldsymbol{s}\in\tau_S, \\ P_{UX}=Q_{UX}}} \exp\{-nD(P_{UXSY}||P_{UXS} \times W_{Y|X,S}) - nI(UX;S) + n\epsilon\} \tag{7.109}$$

$$= \sum_{\substack{P_{UXSY}: I(X;Y|U)<R+\delta, \\ \boldsymbol{s}\in\tau_S, \\ P_{UX}=Q_{UX}}} \exp\{-nD(P_{UXSY}||P_{UX}P_S \times W_{Y|X,S}) + n\epsilon\} \tag{7.110}$$

$$\leq \sum_{\substack{P_{UXSY}: I(X;Y|U)<R+\delta, \\ \boldsymbol{s}\in\tau_S, \\ P_{UX}=Q_{UX}}} \exp\{-nD(P_{UXY}||P_{UX} \times \widetilde{W}_{Y|X}) + n\epsilon\} \tag{7.111}$$

$$\tag{7.112}$$

where $\widetilde{W}_{Y|X} = \sum_s P_S(s) W_{Y|X,S=s}$. We can set ϵ arbitrarily small as ϵ is present to account for the polynomially many terms. In particular, set $\epsilon < \epsilon'$, where ϵ' is described next.
Note that if $P_{UXY} = P_{UX} \times \widetilde{W}_{Y|X}$, then $R + \delta < I(X;Y|U)$ (as $\widetilde{W}_{Y|X} \in \overline{\mathcal{W}_1} \cup \overline{\mathcal{W}_2}$) by choice of R and δ as described. Since mutual information and relative entropy are continuous functions of P_{UXY}, there exists $\epsilon' > 0$ such that if

$I(X;Y) < R + \delta$, then

$$D(P_{UXY}||P_{UX} \times \widetilde{W}_{Y|X}) \geq \epsilon' \ \forall P_S, \text{ or equivalently, } \forall \boldsymbol{s}.$$

Since there are only polynomially many types, for sufficiently large n, (7.105) is less than $\exp\{-n(\epsilon' - \epsilon)/2\} \to 0$ as $n \to \infty$.

Next, we analyze the probability in (7.106). The probability, for any $\boldsymbol{s}$, can be written as

$$= \sum_{\substack{P_{UXX'SY}: I(X';Y|U) \geq R+\delta \\ \boldsymbol{s} \in \tau_S,\ P_{UX} = P_{UX'} = Q_{UX}}} \sum_{\boldsymbol{u} \in \tau_{U|S}\boldsymbol{s}} |\tau_U|^{-1} \sum_{\boldsymbol{x}_i \in \tau_{X|US}(\boldsymbol{u},\boldsymbol{s})} |\tau_{X|U}(\boldsymbol{u})|^{-1} \sum_{j=1, j\neq i}^{M} \sum_{\boldsymbol{x}_j \in \tau_{X'|UXS}(\boldsymbol{u},\boldsymbol{x}_i,\boldsymbol{s})} |\tau_X|^{-1} \sum_{\boldsymbol{y} \in \tau_{Y|UXX'S}(\boldsymbol{u},\boldsymbol{x}_i,\boldsymbol{x}_j,\boldsymbol{s})} W^n(\boldsymbol{y}|\boldsymbol{x}_i,\boldsymbol{s}) \quad (7.113)$$

$$\leq \sum_{\substack{P_{UXX'SY}: I(X';Y|U) \geq R+\delta \\ \boldsymbol{s} \in \tau_S,\ P_{UX} = P_{UX'} = Q_{UX}}} \exp\{nR + n\epsilon - nI(UX;S) - nI(X';UXS) - nI(X';Y|UXS)\} \quad (7.114)$$

$$= \sum_{\substack{P_{UXX'SY}: I(X';Y|U) \geq R+\delta \\ \boldsymbol{s} \in \tau_S,\ P_{UX} = P_{UX'} = Q_{UX}}} \exp\{nR + n\epsilon - nI(UX;S) - nI(X';UXSY)\} \quad (7.115)$$

$$= \sum_{\substack{P_{UXX'SY}: I(X';Y|U) \geq R+\delta \\ \boldsymbol{s} \in \tau_S,\ P_{UX} = P_{UX'} = Q_{UX}}} \exp\{n(R + \epsilon - I(UX;S) - I(X';U) - I(X';Y|U) - I(X';XS|UY))\} \quad (7.116)$$

$$\leq \sum_{\substack{P_{UXX'SY}: I(X';Y|U) \geq R+\delta \\ \boldsymbol{s} \in \tau_S,\ P_{UX} = P_{UX'} = Q_{UX}}} \exp\{n(\epsilon - \delta)\} \quad (7.117)$$

Since ϵ can be set arbitrarily small, this proves the achievability of all rate upto $C(Q_{UX})$ for the described encoding-decoding scheme.

Adversary Identification : For the parition E_k, $\Phi_i(\boldsymbol{u},\boldsymbol{y}) = \sigma_{3-k}$ if $P_(\boldsymbol{u},\boldsymbol{y}) \in E_k$ For the scheme described above, (2.6) can be written as (note that random variables UX have joint distribution Q_{UX} throughout the following equations)

$$P_{\mathrm{i}}^{\mathrm{r}}(Q,k) = \sum_{\substack{P_{UXSY}: \\ P_{UX} = Q_{UX}, \\ P_{UY} \in E_k, \\ \boldsymbol{s} \in \tau_S}} \sum_{\boldsymbol{u} \in \tau_{U|S}(\boldsymbol{s})} |\tau_U|^{-1} \sum_{\boldsymbol{x} \in \tau_{X|US}(\boldsymbol{u},\boldsymbol{s})} |\tau_{X|U}(\boldsymbol{u})|^{-1} \sum_{\boldsymbol{y} \in \tau_{Y|UXS}(\boldsymbol{u},\boldsymbol{x},\boldsymbol{s})} W^n(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{s}) \quad (7.118)$$

$$\leq \sum_{\substack{P_{UXSY}: \\ P_{UX} = Q_{UX}, \\ P_{UY} \in E_k, \\ \boldsymbol{s} \in \tau_S}} \exp\{-nI(U;S) - nI(X;S|U) - nD(P_{UXSY}||P_{UXS}W_{Y|X,S}) + n\epsilon\} \quad (7.119)$$

$$= \max_{\substack{P_{UXSY}: \\ P_{UX} = Q_{UX}, \\ P_{UY} \in E_k, \\ \boldsymbol{s} \in \tau_S}} \exp\{-nD(P_{UXSY}||P_{UX}P_SW_{Y|X,S}) + n\epsilon'\} \quad (7.120)$$

Projecting distribution on $\mathcal{U} \times \mathcal{Y}$, KL divergence does not increase (7.121)

$$\leq \max_{\substack{P_{UY} \in E_k \\ P_U = Q_U, P_S \in \mathcal{P}_k}} \exp\{-nD(P_{Y|U}||\widetilde{W}_{Y|U}) + n\epsilon'\} \quad (7.122)$$

where $\widetilde{W}(Y|U) = \sum_{X,S} Q_{X|U}P_SW_{Y|X,S}$.

Note - all the above inequalities can be replaced by equality in the limit - approximation is $o(n)$. Therefore, the error

exponent is given by

$$\gamma(Q_{UX}) = \min_{k} \min_{\substack{P_{Y|U} \in E_k \\ W_{Y|U} \in \overline{\mathcal{V}}_k}} D(P_{Y|U} || W_{Y|U}). \tag{7.123}$$

■

REFERENCES

[1] D. Blackwell, L. Breiman, and A. J. Thomasian, "The capacity of a class of channels," *The Annals of Mathematical Statistics*, vol. 30, no. 4, pp. 1229–1241, 1959.

[2] D. Blackwell, L. Breiman, and A. J. Thomasian, "The capacities of certain channel classes under random coding," *The Annals of Mathematical Statistics*, vol. 31, no. 3, pp. 558–567, 1960.

[3] J. Wolfowitz, "Simultaneous channels," *Archive for Rational Mechanics and Analysis*, vol. 4, pp. 371–386, Jan 1959.

[4] I. Csiszar 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.

[5] N. Sangwan, M. Bakshi, B. K. Dey, and V. Prabhakaran, "Communication with adversary identification in byzantine multiple access channels," *IEEE International Symposium on Information Theory*, 2021.

[6] J. Jahn, "Coding of arbitrarily varying multiuser channels," *IEEE Trans. Inf. Theory*, vol. 27, pp. 212–226, 1981.

[7] U. Pereg and Y. Steinberg, "The arbitrarily varying broadcast channel with degraded message sets with causal side information at the encoder." arXiv:1709.04770, 2017.

[8] E. Hof and S. I. Bross, "On the deterministic-code capacity of the two-user discrete memoryless arbitrarily varying general broadcast channel with degraded message sets," *IEEE Transactions on Information Theory*, vol. 52, no. 11, pp. 5023–5044, 2006.

[9] O. Kosut and J. Kliewer, "Authentication capacity of adversarial channels," in *2018 IEEE Information Theory Workshop (ITW)*, pp. 1–5, 2018.

[10] A. Beemer, O. Kosut, J. Kliewer, E. Graves, and P. Yu, "Structured coding for authenticationin the presence of a malicious adversary," *IEEE International Symposium on Information Theory*, 2019.

[11] E. Graves, P. Yu, and P. Spasojevic, "Keyless authentication in the presence of asimultaneously transmitting adversary," *IEEE Information Theory Workshop (ITW)*, 2018.

[12] O. Kosut and J. Kliewer, "Network equivalence for a joint compound-arbitrarily-varying network model," *IEEE Information Theory Workshop (ITW)*, 2016.

[13] Y. Polyanskiy, H. V. Poor, and S. Verdu, "Channel coding rate in the finite blocklength regime," *IEEE Transactions on Information Theory*, vol. 56, no. 5, pp. 2307–2359, 2010.

[14] B. Hughes and T. Thomas, "On error exponents for arbitrarily varying channels," *IEEE Transactions on Information Theory*, vol. 42, no. 1, pp. 87–98, 1996.

[15] J. Neyman, E. S. Pearson, and K. Pearson, "Ix. on the problem of the most efficient tests of statistical hypotheses," *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, vol. 231, no. 694-706, pp. 289–337, 1933.

[16] I. Csiszar and J. Korner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. USA: Academic Press, Inc., 1982.

[17] M. Skala, "Hypergeometric tail inequalities: ending the insanity," 2013.