New Results on AVCs With Omniscient and Myopic Adversaries

Anuj Kumar Yadav*, Mohammadreza Alimohammadi†, Yihan Zhang‡, Amitalok J. Budkuley §, Sidharth Jaggi¶

*Department of Electrical Engineering, Indian Institute of Technology Patna
†Department of Electrical Engineering, Sharif University of Technology
‡Institute of Science and Technology (IST) Austria
§Department of Electronics and Electrical Communication Engineering, Indian Institute of Technology Kharagpur.
¶School of Mathematics, University of Bristol.

Abstract

In the classic adversarial communication problem, two parties communicate over a noisy channel in the presence of a malicious jamming adversary. The arbitrarily varying channels (AVCs) offers an elegant framework to study a wide range of interesting adversary models. The optimal throughput or *capacity* over such AVCs is intimately tied to the underlying adversary model; in some cases, capacity is unknown and the problem is known to be notoriously hard. The *omniscient* adversary, one which knows the sender's *entire* channel transmission *a priori*, is one of such classic models of interest; the capacity under such an adversary remains an exciting open problem. The *myopic* adversary is a generalization of that model where the adversary's observation may be corrupted over a noisy discrete memoryless channel. Through the adversary's *myopicity*, one can unify the slew of different adversary models, ranging from the omniscient adversary to one that is completely blind to the transmission (the latter is the well known *oblivious* model where the capacity is fully characterized).

In this work, we present new results on the capacity under both the omniscient and myopic adversary models. We completely characterize the positive capacity threshold over general AVCs with omniscient adversaries. The characterization is in terms of two key combinatorial objects: the set of *completely positive distributions* and the *CP-confusability set*. For omniscient AVCs with positive capacity, we present non-trivial lower and upper bounds on the capacity; unlike some of the previous bounds, our bounds hold under fairly general input and jamming constraints. Our lower bound improves upon the generalized Gilbert-Varshamov bound for general AVCs while the upper bound generalizes the well known Elias-Bassalygo bound (known for binary and q-ary alphabets). For the myopic AVCs, we build on prior results known for the so-called *sufficiently myopic* model, and present new results on the positive rate communication threshold over the so-called *insufficiently myopic* regime (a completely insufficient myopic adversary specializes to an omniscient adversary). We present interesting examples for the widely studied models of adversarial *bit-flip* and *bit-erasure* channels. In fact, for the bit-flip AVC with additive adversarial noise as well as random noise, we completely characterize the omniscient model capacity when the random noise is sufficiently large vis-a-vis the adversary's budget.

Index Terms

Adversarial Channels, Cloud codes, Omniscient adversary, Elias bassalygo bound.

I. Introduction

Historically, the problem of two parties, say Alice and Bob, communicating over a noisy channel has received wide attention in the information-theory community. This problem has an elegant solution when the channel noise is random and independent (cf. [Sha59]). However, the problem is notoriously hard when the channel noise maybe controlled, partially or fully, by an adversary, say jammer James. In such a case, the channel noise or jamming state is maliciously chosen and maybe arbitrarily correlated to the interactions between Alice and Bob; the nature and the capabilities of the adversary further complicate the analysis. In this work, we study the optimal communication throughput over channels with such active adversaries.

We leverage the framework of arbitrarily varying channels (AVCs) to study communication with active adversaries (cf. [BBT59], [LN98]). Given an AVC, an omniscient adversary models the most pessimistic scenario vis-a-vis adversarial communication where James knows Alice's entire transmission a priori. While the threshold for positive rate was recently characterized for general AVCs with omniscient adversary, the completely capacity remains an exciting open problem for the community. This is in stark contrast to the case of an oblivious adversary (one, that is completely unaware of Alice's channel transmissions); for such a model, Csiszar-Narayan [CN88], [CN89] characterized the capacity. The *myopic* adversary model, where James observes a noisy version of Alice's codeword, offers a framework to to unify the study of all 'adversary models'. In this work, we present new results on the capacity of channels with myopic and omniscient adversaries.

The most well studied model for adversarial communication is the binary *bit-flip AVC*, were Alice's and James' inputs as well as the channel output are binary vectors. Consider the scenario when Alice can input any binary vector while an omniscient James is allowed to flip at most a fraction p of the transmitted bits. It is well known that when James is allowed to flip at least a quarter of the transmitted bits, positive rate communication is impossible [Plo60]. More generally, when capacity is positive,

the Gilbert-Varshamov (GV) [Var57] bound and the MRRW bound [MRRW77] offer the best lower and upper bounds on the capacity. Extensions of such results are known for some other 'structured' classes of channels.

II. Our contributions

In this work we provide new results for general channels under both omniscient and myopic AVCs.

Omniscient AVCs: An AVC is said to be *omniscient* if the adversary's jamming sequence $\underline{s}$ can be arbitrarily correlated with the transmitted codeword $\underline{x}$ and codebook.

- Recent work [WBBJ19] gave a precise characterization of the rate-positivity regime for any state-deterministic AVC (where channel output is a deterministic function of the transmitted codeword and the jamming state). In Theorem 23 we extend this characterization to general non-state-deterministic AVCs, by extending both the rate-positivity result and the generalized Plotkin bound in [WBBJ19].

- We similarly generalize the classical Elias-Bassalygo bound [Bas65] to prove novel outer bounds (that are in general significantly sharper than sphere-packing converse bounds); this follows from a further generalization of this bound to myopic AVCs (see Thm 28).

- In Theorem 24 we analogously extend the achievability scheme of Csiszár and Körner [CK81] to general AVCs with both state and input constraints. Importantly, this requires us to analyze codes where codewords are selected according to a distribution $P_{\mathbf{x}|\mathbf{t}}$ with respect to a fixed time-sharing sequence $\underline{t}$ – as shown in [WBBJ19] this generalization from i.i.d. random codes is necessary for some AVCs with input constraints.

- Example of the Theorem 23 and Theorem 24 for the special case of the non state-deterministic and omniscient bit-flip AVCs is given in corollary 26.

Insufficiently myopic AVCs: We generalize some of the results above to AVCs which are "insufficiently myopic". Myopic AVCs were defined in [Sar10] as channels where the adversary does not observe the transmitted codeword $\underline{x}$ directly, but only a noisy observation $\underline{z}$, and must choose its jamming action $\underline{s}$ as a function only of $\underline{z}$ and the codebook. The followup work of [DJL19a] removed the strong assumption made in [Sar10] (that the encoder and decoder share common randomness, roughly speaking). However, [DJL19a] focused on "sufficiently myopic" channels, where results similar to the classical results of Csiszár-Narayayan [CN88] for oblivious adversaries were obtained. In this work we focus on insufficiently myopic AVCs, which are channels for which, roughly speaking (the formal definition is in Definition 18) there exist adversarial jamming actions that ensure that the mutual information on the channel from the transmitter to the receiver cannot be higher than than the mutual information on the channel from the transmitter to the jammer, and hence the channels are spiritually closer to being omniscient than to being myopic. This setting has interesting subtle differences with the truly omniscient setting. A consequence of insufficient myopicity is that whenever the receiver is able to decode the transmitter's message, then with high probability so can the jammer. Nonetheless, as shown in [BDJ$^+$20], there exist myopic AVCs for which, if the encoder uses a *stochastic code*, (i.e., for each message it randomly selects one of a potentially large set of codewords), then the rates achievable are strictly better than those obtainable for the corresponding omniscient AVC. The difference arises from the fact that the jammer does not get to observe the transmitted $\underline{x}$ directly, but only the noisy observation $\underline{z}$. In [BDJ$^+$20] and followup work in [BDJ$^+$21] it was demonstrated that if the channel from the transmitter to the jammer is a BEC(q), and the jammer can erasure up to pn bits of the transmitted codeword, then if $q > 2p - 1$ then a positive rate is achievable, whereas if $q \leq 2p - 1$ no positive rate is achievable. Note that for this channel the regime $p \geq q$ corresponds to the insufficiently myopic regime.

- Building on the myopic strong-symmetrizability condition obtained in [BDJ$^+$20] (which gives a sufficient condition for the adversary being able to jam any communication via a symmetrization strategy coupled with a generalized Plotkin-type bound), we prove in Theorem 28 a generalized Elias-Bassalygo even for myopic AVCs (in the setting where the codes used are deterministic). In particular this extension requires a novel covering lemma that produces a covering of any codebook with a set of sequences with the interesting property that most covering sequences cover a "large number" of codewords – this property is important since it enables us to use myopic symmetrization arguments even though the jammer may not know precisely what codeword was transmitted.

- In Theorem 30 we provide a rate-positivity achievability result for state-deterministic myopic AVCs (paralleling the achievability result in Theorem 23 for omniscient AVCs). This achievability result generalizes the achievability for myopic binary-erasure AVCs obtained in [BDJ$^+$20], [BDJ$^+$21] to general myopic AVCs, and in particular holds for *both*

insufficiently myopic AVCs *and* sufficiently myopic AVCs, in contrast to prior work [DJL19a] where the achievability arguments were restricted to sufficiently myopic AVCs. We note, however, that it is unclear whether for general AVCs even the rate-positivity threshold obtained via the achievability argument in this work precisely complements the zero-rate impossibility result in [BDJ$^+$20] – the challenge lies in the different notions of symmetrizability that naturally arise from our techniques here in the achievability sense, and in the converse sense in [BDJ$^+$20]. On the one hand, our achievability argument can be optimized over general classes of auxiliary input variables/auxiliary input channels (no such optimization arises in the converse argument presented in [BDJ$^+$20]) – identifying the optimizing distributions is challenging. On the other hand the converse in [BDJ$^+$20], since it holds for both sufficiently myopic and insufficiently myopic AVCs, is restricted to jamming actions where the adversary is unable to estimate the truly transmitted message. Closing the gap between these two types of results requires new ideas, and is ongoing work.

- Example of the Theorem 28 for the special case of state-deterministic and myopic bit-flip AVC is given in corollary 29.

The rest of the paper is organized as follows. In Section III, we present the notation and preliminaries used in this work. In Section V, we formally describe the communication model and state our problem. Next, we present our main results in Section VI; the proofs of our the main results appear in sections VII, VIII, IX, X and XI.

III. Notation and Preliminaries

A. Notation

We denote the random variables, their instances and their alphabet/set using boldface letters (eg. $\mathbf{x}$), lower case letters (eg. x), and calligraphic letters (eg. $\mathcal{X}$) respectively. The random vectors are denoted using boldface letters with under bar ($\underline{\mathbf{x}} = (\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n)$), their concomitant values using lower case letters with under bar $\underline{x} = (x_1, x_2, \cdots, x_n)$, resp.). The set of real numbers, non-negative real numbers, natural numbers and rational numbers are denoted by $\mathbb{R}$, $\mathbb{R}_+$, $\mathbb{N}$ and $\mathbb{Q}$. For $a \in \mathbb{N}$, let $[a] := \{1, 2, \cdots, a\}$. For $[a, b] \in \mathbb{R}$, $a \leq b$, let $[a, b]$ denote the closed interval defined between a and b. Matrices are denoted by boldface capital letters (e.g., $\mathbf{A}$, $\mathbf{B}$, etc). Deterministic and random functions will be denoted by lower case letters (eg. f) and by upper case letters (e.g., F), respectively. Unless stated otherwise, all sets are assumed to be finite.
Let $\Delta(\mathcal{X})$ denote the simplex of probability distributions on the set $\mathcal{X}$. The related notations $\Delta(\mathcal{X} \times \mathcal{Y})$ and $\Delta(\mathcal{X}|\mathcal{Y})$ are defined similarly. We denote by $P_{\mathbf{x}} \in \Delta(\mathcal{X})$, $P_{\mathbf{x}|\mathbf{y}} \in \Delta(\mathcal{X}|\mathcal{Y})$ and $P_{\mathbf{x},\mathbf{y}} \in \Delta(\mathcal{X} \times \mathcal{Y})$ the probability distribution of random variable $\mathbf{x} \in \mathcal{X}$, the conditional probability distribution induced by random variable $\mathbf{x} \in \mathcal{X}$ conditioned on events generated by random variable $\mathbf{y} \in \mathcal{Y}$ and the joint probability distribution on the pair of random variables $(\mathbf{x}, \mathbf{y}) \in \mathcal{X} \times \mathcal{Y}$ respectively. For the latter, we denote the marginal distribution on random variable $\mathbf{x}$ by $[P_{\mathbf{x},\mathbf{y}}]_{\mathbf{x}}$. For a distribution $P_{\mathbf{x},\mathbf{y}|\mathbf{u}} \in \Delta(\mathcal{X}, \mathcal{Y}|\mathcal{U})$, $[P_{\mathbf{x},\mathbf{y}|\mathbf{u}}]_{\mathbf{x}|\mathbf{u}}$ denotes the marginal distribution onto $\mathbf{x}$ given $\mathbf{u}$. Given $P_{\mathbf{x}}$, $P_{\mathbf{x}}^{(n)}$ denotes the n-fold memoryless extension of $P_{\mathbf{x}}$. We use $\Delta^{(n)}(\mathcal{X})$ to denote the set of types of n-length vectors over alphabet $\mathcal{X}$. In other words, $\Delta^{(n)}(\mathcal{X})$ consists of all distributions $P_{\mathbf{x}} \in \Delta(\mathcal{X})$ that can be induced by vectors in $\mathcal{X}^n$. Other notations such as $\Delta^{(n)}(\mathcal{X} \times \mathcal{Y})$ and $\Delta^{(n)}(\mathcal{X}|\mathcal{Y})$ are defined in similar manner. Let $\mathbb{P}(A)$ denote the probability of event A. Let $\mathbf{x} \sim \text{Bernoulli}(p)$ denote a Bernoulli random variable $\mathbf{x}$ with parameter $p \in [0, 1]$.
The Hamming distance between two vectors, say $\underline{\mathbf{x}}, \underline{\mathbf{x}}' \in \mathcal{X}^n$ is denoted by $d_H(\underline{\mathbf{x}}, \underline{\mathbf{x}}')$. Given distributions $P_{\mathbf{x}}, Q_{\mathbf{x}} \in \Delta(\mathcal{X})$, $\|P_{\mathbf{x}} - Q_{\mathbf{x}}\|_1 = \sum_{x \in \mathcal{X}} |P_{\mathbf{x}}(x) - Q_{\mathbf{x}}(x)|$ denotes the ℓ_1 distance between $P_{\mathbf{x}}$ and $Q_{\mathbf{x}}$. All logarithms are to the base 2, unless stated otherwise.

IV. Preliminaries

A. Information Measures

Let $\mathbf{x} \in \mathcal{X}$ and $\mathbf{y} \in \mathcal{Y}$ be random variables with distributions $P_{\mathbf{x}} \in \Delta(\mathcal{X})$ and $P_{\mathbf{y}} \in \Delta(\mathcal{Y})$ respectively. The Shannon entropy of $\mathbf{x}$ is defined as $H(\mathbf{x}) := \sum_{x \in \mathcal{X}} P_{\mathbf{x}}(x) \log(1/P_{\mathbf{x}}(x))$. The conditional entropy of $\mathbf{x}$ given $\mathbf{y}$ is defined as $H(\mathbf{x}|\mathbf{y}) := \sum_{x,y} P_{\mathbf{x},\mathbf{y}}(x, y) \log(1/P_{\mathbf{x},\mathbf{y}}(x, y))$. Further, the mutual information between $\mathbf{x}$ and $\mathbf{y}$ is defined as:

$$I(\mathbf{x}; \mathbf{y}) := \sum_{x,y} P_{\mathbf{x},\mathbf{y}}(x, y) \log \frac{P_{\mathbf{x},\mathbf{y}}(x, y)}{P_{\mathbf{x}}(x) P_{\mathbf{y}}(y)}$$

If $\mathbf{x}$ and $\mathbf{y}$ along with a random variable $\mathbf{z} \in \mathcal{Z}$ have joint distribution $P_{\mathbf{x},\mathbf{y},\mathbf{z}}$ then, the conditional mutual information between $\mathbf{x}$ and $\mathbf{y}$ given $\mathbf{z}$ is given by:

$$I(\mathbf{x}; \mathbf{y}|\mathbf{z}) := \sum_{z} P_Z(z) \sum_{x,y} P_{\mathbf{x},\mathbf{y}|\mathbf{z}}(x, y|z) \log \frac{P_{\mathbf{x},\mathbf{y}|\mathbf{z}}(x, y|z)}{P_{\mathbf{x}|\mathbf{z}}(x|z) P_{\mathbf{y}|\mathbf{z}}(y|z)}$$

Definition 1 (KL Divergence)**.** *Let P and Q be two discrete probability distributions defined over the same probability simplex $\Delta(\mathcal{X})$. The Kulback-Liebler (KL) divergence (Relative entropy) between P and Q is defined as:*

$$D(P\|Q) := \sum_{x\in\mathcal{X}} \log \frac{P_{\mathbf{x}}(x)}{Q_{\mathbf{x}}(x)}$$

Note that the KL divergence is defined only if P is absolutely continuous w.r.t Q i.e., $\forall x \in \mathcal{X}$ s.t. $Q_{\mathbf{x}}(x) = 0 \implies P_{\mathbf{x}}(x) = 0$. Furthermore, given $P_{\mathbf{y}|\mathbf{x}}$, $Q_{\mathbf{y}|\mathbf{x}} \in \mathcal{P}(\mathcal{Y}|\mathcal{X})$ and $x \in \mathcal{X}$, let

$$D(P_{\mathbf{y}|\mathbf{x}}(\cdot|x)||Q_{\mathbf{y}|\mathbf{x}}(\cdot|x)) := \sum_{y} P_{\mathbf{y}|\mathbf{x}}(y|x) \log\left(\frac{P_{\mathbf{y}|\mathbf{x}}(y|x)}{Q_{\mathbf{y}|\mathbf{x}}(y|x)}\right) \tag{1}$$

and given $P_{\mathbf{x},\mathbf{y}} = P_{\mathbf{x}}P_{\mathbf{y}|\mathbf{x}}$, $Q_{\mathbf{x},\mathbf{y}} = P_{\mathbf{x}}Q_{\mathbf{y}|\mathbf{x}}$, where $P_{\mathbf{x}} \in \mathcal{P}(\mathcal{X})$, let

$$D(P_{\mathbf{x},\mathbf{y}}||Q_{\mathbf{x},\mathbf{y}}) := \sum_{x} P_{\mathbf{x}}(x) D(P_{\mathbf{y}|\mathbf{x}}(\cdot|x)||Q_{\mathbf{y}|\mathbf{x}}(\cdot|x))$$

Lemma 2 (Pinsker's Inequality)**.** *Let $P_{\mathbf{x}}$ and $Q_{\mathbf{x}}$ be probability distributions defined over the probability simplex $\Delta(\mathcal{X})$. Then for any positive constant c,*

$$\|P_{\mathbf{x}} - Q_{\mathbf{x}}\| \leq c\sqrt{D(P_{\mathbf{x}}||P_{\mathbf{x}})} \tag{2}$$

Lemma 3 (Chernoff Bound [CT91])**.** *Let $\mathbf{x} := \sum_{i=1}^{n} \mathbf{x}_i$, where $\mathbf{x}_i \sim Ber(\mu_i)$, $0 \leq \mu_i \leq 1$, $\forall i$, and $(\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n)$ are independent. Let $\mu := \sum_{i=1}^{n} \mu_i$. Then,*

$$\mathbb{P}(\mathbf{x} \geq (1+\delta)\mu) \leq e^{-\frac{\delta^2}{2+\delta}\mu} \tag{3}$$

for all $\delta > 0$.

B. Method of Types

The type of a vector $\underline{x} \in \mathcal{X}^n$, denoted by $P_{\underline{\mathbf{x}}} \in \Delta(\mathcal{X})$, is defined as the probability distribution representing the relative frequency/empirical distribution of each symbol $x \in \mathcal{X}$. The set of all possible types of n-length vectors is denoted by $\Delta^{(n)}(\mathcal{X}) \subset \Delta(\mathcal{X})$. Note that the types are special probability distributions with rational entries and thus, $\bigcup_{i=1}^{\infty} \Delta^{(n)}(\mathcal{X})$ is dense in $\Delta(\mathcal{X})$. Given vectors $\underline{x} \in \mathcal{X}^n$, $\underline{y} \in \mathcal{Y}^n$, the joint type $P_{\underline{\mathbf{x}},\underline{\mathbf{y}}} \in \Delta^{(n)}(\mathcal{X}\times\mathcal{Y}) \subset \Delta(\mathcal{X}\times\mathcal{Y})$ and the conditional type $P_{\underline{\mathbf{y}}|\underline{\mathbf{x}}} \in \Delta^{(n)}(\mathcal{Y}|\mathcal{X}) \subset \Delta(\mathcal{Y}|\mathcal{X})$ are defined similarly. While $P_{\mathbf{x}}$ is an element of $\Delta(\mathcal{X})$, $P_{\underline{\mathbf{x}}}$ refers to an element in $\Delta^{(n)}(\mathcal{X})$. Given a type $T_{\underline{\mathbf{x}}}$, let $\mathcal{P}(P_{\underline{\mathbf{x}}}) = \{\underline{x} \in \mathcal{X}^n : P_{\underline{x}} = P_{\underline{\mathbf{x}}}\}$ denote the set of n-length typical $\underline{\mathbf{x}}$-sequences of type $P_{\underline{\mathbf{x}}}$. Similarly, for a given $\underline{x}$ of type $P_{\underline{x}}$, the set of conditionally typical sequences $\underline{y} \in \mathcal{Y}^n$ defined via the conditional type $P_{\mathbf{y}|\underline{x}} \in \Delta^{(n)}(\mathcal{Y}|\underline{x}) \subset \Delta^{(n)}(\mathcal{Y}|\mathcal{X})$ is denoted by $\mathcal{P}(P_{\mathbf{y}|\underline{x}})$.
Now, we state few bounds related to types and typical sets that will be used later on in our proofs:

$$|\Delta(\mathcal{X})| \leq (n+1)^{|\mathcal{X}|} \tag{4}$$

$$|\Delta(\mathcal{Y}|\mathcal{X})| \leq (n+1)^{|\mathcal{X}||\mathcal{Y}|} \tag{5}$$

$$(n+1)^{-|\mathcal{X}|}2^{nH(\mathbf{x})} \leq |\mathcal{P}(P_{\underline{\mathbf{x}}})| \leq 2^{nH(\mathbf{x})} \tag{6}$$

$$(n+1)^{-|\mathcal{X}||\mathcal{Y}|}2^{nH(\mathbf{y}|\mathbf{x})} \leq |\mathcal{P}(P_{\mathbf{y}|\underline{x}})| \leq 2^{nH(\mathbf{y}|\mathbf{x})} \tag{7}$$

Theorem 4 (Sanov's Theorem [CT91])**.** *Let $\underline{\mathbf{x}}$ be generated in an independent and identically distributed (i.i.d.) manner using the distribution $P_{\mathbf{x}} \in \mathcal{P}(\mathcal{X})$, and fix some type $Q_{\mathbf{x}} \in \mathcal{P}^{(n)}(\mathcal{X})$. Then,*

$$\mathbb{P}\left(\underline{\mathbf{x}} : P_{\underline{\mathbf{x}}} \in \mathcal{P}(Q_{\mathbf{x}})\right) \leq 2^{-nD(Q_{\mathbf{x}}||P_{\mathbf{x}})} \tag{8}$$

V. Problem Setup

A. Arbitrarily Varying Channels (AVCs)

An arbitrarily varying channel (AVC) $\mathcal{A} = (\mathcal{X}, \mathcal{S}, \mathcal{Y}, \lambda_{\mathbf{x}}, \lambda_{\mathbf{s}}, W_{\underline{\mathbf{y}}|\underline{\mathbf{x}},\underline{\mathbf{s}}})$ is defined by an input alphabet $\mathcal{X}$, state alphabet $\mathcal{S}$, output alphabet $\mathcal{Y}$, input constraints $\lambda_{\mathbf{x}}$, state constraint $\lambda_{\mathbf{s}}$, and the channel law $W_{\underline{\mathbf{y}}|\underline{\mathbf{x}},\underline{\mathbf{s}}}$. In context to this work, we make the following assumptions:

1) The alphabets $\mathcal{X}$, $\mathcal{S}$ and $\mathcal{Y}$ are finite.
2) The input constraint set $\lambda_{\mathbf{x}} \subseteq \Delta(\mathcal{X})$ and the state constraint set $\lambda_{\mathbf{s}} \subseteq \Delta(\mathcal{S})$ are convex polytopes. Mathematically, these sets are defined as:

$$\lambda_{\mathbf{x}} := \left\{P_{\mathbf{x}} \in \Delta(\mathcal{X}) : \forall j \leq k, \sum_{x\in\mathcal{X}} A_j P_{\mathbf{x}}(x) \leq \Gamma_j\right\} = \{P_{\mathbf{x}} \in \Delta(\mathcal{X}) : \mathbf{A}P_{\mathbf{x}} \leq \underline{\Gamma}\}$$

where $\mathbf{A} \in \mathbb{R}^{k\times|\mathcal{X}|}$ is a matrix with (j,x)-th entry as $A_j(x)$ and $\underline{\Gamma} \in \mathbb{R}^k$ is a vector with Γ_j as the j-th component. Further, the set of length-n input sequences satisfying input constraint $\lambda_{\mathbf{x}}$ is defined as:

$$\mathcal{P}(\lambda_{\mathbf{x}}) := \{\underline{x} \in \mathcal{X}^n : P_{\underline{x}} \in \lambda_{\mathbf{x}}\}$$

Similarly, the state constraint set $\lambda_{\mathbf{s}}$ is defined as:

$$\lambda_{\mathbf{s}} := \left\{ P_{\mathbf{s}} \in \Delta(\mathcal{S}) : \forall j \leq m, \sum_{x\in\mathcal{X}} A_j P_{\mathbf{x}}(x) \leq \Lambda_j \right\} = \{P_{\mathbf{x}} \in \Delta(\mathcal{X}) : \mathbf{A}P_{\mathbf{x}} \leq \underline{\Lambda}\}$$

where $\mathbf{A} \in \mathbb{R}^{m\times|\mathcal{X}|}$ is a matrix with (j,x)-th entry as $A_j(x)$ and $\underline{\Lambda} \in \mathbb{R}^m$ is a vector with Λ_j as the j-th component. Further, the set of length-n state sequences satisfying state constraint $\lambda_{\mathbf{s}}$ is defined as:

$$\mathcal{P}(\lambda_{\mathbf{s}}) := \{\underline{s} \in \mathcal{S}^n : P_{\underline{s}} \in \lambda_{\mathbf{s}}\}$$

3) The AVC is index-wise decomposable i.e., $W_{\underline{\mathbf{y}}|\underline{\mathbf{x}},\underline{\mathbf{s}}}(\underline{y}|\underline{x},\underline{s}) = \prod_{i=1}^n W_{\mathbf{y}|\mathbf{x},\mathbf{s}}(y_i|x_i,s_i)$.

The adversarial communication problem involves three parties, two legitimate parties- Alice and Bob, and jamming adversary- James. It is described as follows: Alice wants to transmit a message $m \in \mathcal{M}$ to Bob over the AVC in the presence of James who wish to disrupt the communication between the legitimate parties. The noise in the AVC can be partially (non-state deterministic AVCs) or fully (state deterministic AVCs) controlled by James. Alice encodes the message m into a codeword $\underline{\mathbf{x}} \in \mathcal{P}(\lambda_{\mathbf{x}})$ and transmits across the AVC, James inputs a sequence $\underline{\mathbf{s}} \in \mathcal{P}(\lambda_{\mathbf{s}})$ in order to corrupt the transmitted codeword $\underline{\mathbf{x}}$. Consequently, Bob receives a corrupted sequence $\underline{\mathbf{y}}$ and the goal is to estimate the message m with negligible probability (non-state deterministic AVC) or zero probability (state deterministic AVC) of error.

B. Non State-Deterministic and Omniscient AVCs

Definition 5. *A non-state deterministic and omniscient arbitrarily varying channel $(\mathcal{X},\mathcal{S},\mathcal{Y},\lambda_{\mathbf{x}},\lambda_{\mathbf{s}},W_{\mathbf{y}|\mathbf{x},\mathbf{s}})$ is defined via the input constraint set $\lambda_{\mathbf{x}}$, the state constraint set $\lambda_{\mathbf{s}}$, and the channel law described via the stochastic matrix $W_{\mathbf{y}|\mathbf{x},\mathbf{s}}$ such that for any time-sharing variable $\mathbf{t} \in \mathcal{T}$ and any distribution $[P_{\mathbf{t},\mathbf{x}}]_{\mathbf{x}} \in \lambda_{\mathbf{x}}$ there always exists a jamming action $V_{\mathbf{s}|\mathbf{t},\mathbf{x}} \in \Delta(\mathcal{S}|\mathcal{T}\times\mathcal{X})$ by James such that $[P_{\mathbf{t},\mathbf{x}}V_{\mathbf{s}|\mathbf{t},\mathbf{x}}]_{\mathbf{s}} \in \lambda_{\mathbf{s}}$.*

Informally, in a non state-deterministic and omniscient AVC model, the jammer 'James' is completely cognizant of the Alice's message m, transmitted codeword $\underline{x}$ (non-causally), Alice's encoder, Bob's decoder and the codebook $\mathcal{C}$ and inflicts jamming noise sequence $\underline{s}$ into the channel based on its complete knowledge of observation.

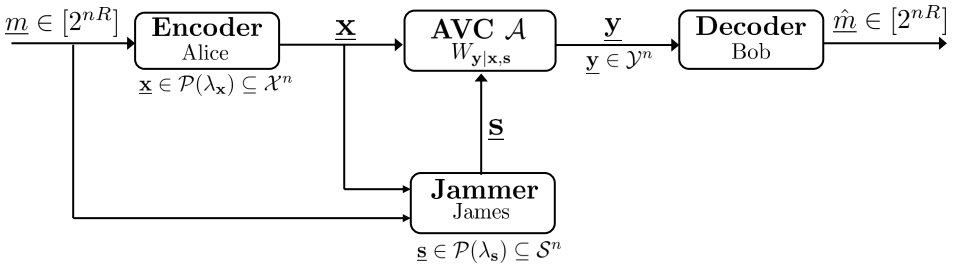


Fig. 1. Communication over a non state-deterministic and omniscient AVC $\mathcal{A}$ $(\mathcal{X},\mathcal{S},\mathcal{Y},\lambda_{\mathbf{x}},\lambda_{\mathbf{s}},W_{\mathbf{y}|\mathbf{x},\mathbf{s}})$

Definition 6 ($P_{\mathbf{x}}$-self coupling)**.** *A joint distribution $P_{\mathbf{x},\mathbf{x}'}$ over the alphabet pair $\mathcal{X}\times\mathcal{X}$ is said to be a self-coupling if its marginal distributions are identical, i.e., $[P_{\mathbf{x},\mathbf{x}'}]_{\mathbf{x}} = [P_{\mathbf{x},\mathbf{x}'}]'_{\mathbf{x}}$. In particular, given some $P_{\mathbf{x}} \in \lambda_{\mathbf{x}} \subseteq \Delta(\mathcal{X})$, a joint distribution $P_{\mathbf{x},\mathbf{x}'}$ is said to be a $P_{\mathbf{x}}$-self coupling if $[P_{\mathbf{x},\mathbf{x}'}]_{\mathbf{x}} = [P_{\mathbf{x},\mathbf{x}'}]_{\mathbf{x}'} = P_{\mathbf{x}}$.*

Definition 7 ($\lambda_{\mathbf{x}}$-self coupling)**.** *Given an input constraint set $\lambda_{\mathbf{x}} \subseteq \Delta(\mathcal{X})$, a joint distribution $P_{\mathbf{x},\mathbf{x}'}$ is said to be a $\lambda_{\mathbf{x}}$-self coupling if $[P_{\mathbf{x},\mathbf{x}'}]_{\mathbf{x}} = [P_{\mathbf{x},\mathbf{x}'}]_{\mathbf{x}'} \in \lambda_{\mathbf{x}}$.*

Definition 8 (Self-coupling space)**.** *Given a probability distribution $P_{\mathbf{x}} \in \Delta(\mathcal{X})$, the self-coupling space $\mathcal{J}(P_{\mathbf{x}})$ is the set of all $P_{\mathbf{x}}$-self couplings, i.e., joint distributions $P_{\mathbf{x},\mathbf{x}'}$ with identical marginal distribution $P_{\mathbf{x}}$.*
Similarly, we define the self coupling space $\mathcal{J}(\lambda_{\mathbf{x}})$ as the set of all $\lambda_{\mathbf{x}}$-self couplings. Alternatively, the $\mathcal{J}(\lambda_{\mathbf{x}})$ can also been

seen as the union of all self-coupling spaces $\mathcal{J}(P_{\mathbf{x}})$ where $P_{\mathbf{x}} \in \lambda_{\mathbf{x}}$ i.e.,

$$\mathcal{J}(\lambda_{\mathbf{x}}) := \bigcup_{P_{\mathbf{x}} \in \lambda_{\mathbf{x}}} \mathcal{J}(P_{\mathbf{x}})$$

Definition 9 (Confusability set)**.** *For a distribution $P_{\mathbf{x}} \in \lambda_{\mathbf{x}}$, the confusability set $\mathcal{K}(P_{\mathbf{x}})$ for an AVC $(\mathcal{A})$ is the set of all such $P_{\mathbf{x}}$-self couplings for which there exists a joint distribution $P_{\mathbf{x},\mathbf{x},\mathbf{s},\mathbf{s},\mathbf{y}} \in \Delta(\mathcal{X}^2 \times \mathcal{S}^2 \times \mathcal{Y})$ such that the following conditions hold:*

1) *(Input consistency): The marginal of the joint distribution $P_{\mathbf{x},\mathbf{x}',\mathbf{s},\mathbf{s}',\mathbf{y}}$ onto the pair of random variables is a $P_{\mathbf{x}}$-self coupling i.e., $[P_{\mathbf{x},\mathbf{x}',\mathbf{s},\mathbf{s}',\mathbf{y}}]_{\mathbf{x},\mathbf{x}'} \in \mathcal{J}(P_{\mathbf{x}})$.*
2) *(State consistency): The marginal of the joint distribution $P_{\mathbf{x},\mathbf{x}',\mathbf{s},\mathbf{s}',\mathbf{y}}$ onto random variables $\mathbf{s}$ and $\mathbf{s}'$ individually respect James' state constraints i.e., $[P_{\mathbf{x},\mathbf{x}',\mathbf{s},\mathbf{s}',\mathbf{y}}]_{\mathbf{s}}$, $[P_{\mathbf{x},\mathbf{x}',\mathbf{s},\mathbf{s}',\mathbf{y}}]_{\mathbf{s}'} \in \lambda_{\mathbf{s}}$*
3) *(Channel consistency): For any 5-tuple (x, x', s, s', y) such that $P_{\mathbf{x},\mathbf{x}',\mathbf{s},\mathbf{s}',\mathbf{y}}(x, x', s, s', y) > 0$, and the channel law is respected i.e., $[P_{\mathbf{x},\mathbf{x}',\mathbf{s},\mathbf{s}',\mathbf{y}}]_{\mathbf{y}|\mathbf{x},\mathbf{s}} = W_{\mathbf{y}|\mathbf{x},\mathbf{s}}$ and $[P_{\mathbf{x},\mathbf{x}',\mathbf{s},\mathbf{s}',\mathbf{y}}]_{\mathbf{y}|\mathbf{x}',\mathbf{s}'} = W_{\mathbf{y}|\mathbf{x},\mathbf{s}}$.*

Any distribution $P_{\mathbf{x},\mathbf{x}',\mathbf{s},\mathbf{s}',\mathbf{y}}$ satisfying the above constraints is said to be an $\mathcal{A}$-refinement of $P_{\mathbf{x},\mathbf{x}'}$. This refinement can be understood to encode the 5-tuple corresponding to the pair of possible inputs Alice inputs to the channel, the pair of states James imposes on the channel, and the channel output.

The confusability set $\mathcal{K}(\lambda_{\mathbf{x}})$ can be defined similarly. It can be represented as the union of all confusability sets $\mathcal{K}(P_{\mathbf{x}})$ where $P_{\mathbf{x}} \in \lambda_{\mathbf{x}}$ i.e.,

$$\mathcal{K}(\lambda_{\mathbf{x}}) := \bigcup_{P_{\mathbf{x}} \in \lambda_{\mathbf{x}}} \mathcal{K}(P_{\mathbf{x}})$$

Claim 10 (Properties of confusability sets)**.** *The confusability set $\mathcal{K}(\lambda_{\mathbf{x}})$ has the following properties:*

1) (Non-triviality): *For every $P_{\mathbf{x}} \in \lambda_{\mathbf{x}}$ the confusability set $\mathcal{K}(P_{\mathbf{x}})$ is a non-empty subset of $\mathcal{J}(P_{\mathbf{x}})$. In particular, Diag$(P_{\mathbf{x}})$, defined as the diagonal matrix comprising of the entries of $P_{\mathbf{x}}$, is always within the confusability set $\mathcal{K}(P_{\mathbf{x}})$. Therefore, the confusability set $\mathcal{K}(\lambda_{\mathbf{x}})$ is a non-empty subset of $\mathcal{J}(\lambda_{\mathbf{x}})$.*
2) (Convexity): *For every $P_{\mathbf{x}} \in \lambda_{\mathbf{x}}$ the confusability set $\mathcal{K}(P_{\mathbf{x}})$ is convex. Therefore, the confusability set $\mathcal{K}(\lambda_{\mathbf{x}})$ is also a convex set.*
3) (Transpositional symmetry): *A $\lambda_{\mathbf{x}}$-self coupling $P_{\mathbf{x},\mathbf{x}'}$ is in the confusability set $\mathcal{K}(\lambda_{\mathbf{x}})$ if and only if its transpose $P_{\mathbf{x}',\mathbf{x}} \triangleq (P_{\mathbf{x},\mathbf{x}'})^T$ is also in $\mathcal{K}(\lambda_{\mathbf{x}})$.*

Now, we state an alternative definition of the confusability set.

Definition 11 (Confusability set)**.** *Let $P_{\mathbf{x}} \in \Delta(\mathcal{X})$. The* confusability set *$\mathcal{K}(P_{\mathbf{x}})$ for an AVC $(\mathcal{A})$ with distribution $P_{\mathbf{x}}$ is defined as*

$$\mathcal{K}(P_{\mathbf{x}}) := \left\{ P_{\mathbf{x},\mathbf{x}'} \in \mathcal{J}(P_{\mathbf{x}}) : \begin{array}{l} \exists\ Q^1_{\mathbf{s}|\mathbf{x},\mathbf{x}'}, Q^2_{\mathbf{s}'|\mathbf{x},\mathbf{x}'} \in \Delta(\mathcal{S}|\mathcal{X} \times \mathcal{X}) \\ \quad \text{such that for } i = \{1, 2\},\ [P_{\mathbf{x},\mathbf{x}'} Q^i_{\mathbf{s}|\mathbf{x},\mathbf{x}'}]_{\mathbf{s}} \in \lambda_{\mathbf{s}}, \\ \bullet\ \forall (x, x', y) \in \mathcal{X} \times \mathcal{X} \times \mathcal{Y}, \\ \sum_{s \in \mathcal{S}} Q^1_{\mathbf{s}|\mathbf{x},\mathbf{x}'}(s|x, x') W_{\mathbf{y}|\mathbf{x},\mathbf{s}}(y|x, s) = \sum_{s' \in \mathcal{S}} Q^2_{\mathbf{s}'|\mathbf{x},\mathbf{x}'}(s'|x', x) W_{\mathbf{y}|\mathbf{x}',\mathbf{s}'}(y|x', s') \end{array} \right\}.$$

Note that the symmetrizability equation above in Definition 11, also holds for a single symmetrizing distribution defined via $V_{\mathbf{s}|\mathbf{x},\mathbf{x}'}$ such that $V_{\mathbf{s}|\mathbf{x},\mathbf{x}'} \triangleq \frac{Q^1_{\mathbf{s}|\mathbf{x},\mathbf{x}'} + Q^2_{\mathbf{s}'|\mathbf{x},\mathbf{x}'}}{2}$.

The definitions of confusability set in Definition 9 and Definition 11 can be shown to be equivalent (cf. [ZJB20] and [BDJ+21] for details).

Definition 12 (Independent self-couplings)**.** *The independent $P_{\mathbf{x}}$-self coupling $P^{ind}_{\mathbf{x},\mathbf{x}'}$ is defined as the joint distribution corresponding to two independent copies of $\mathbf{x}$ drawn with distribution $P_{\mathbf{x}} \in \lambda_{\mathbf{x}}$. Hence for each $(x, x') \in \mathcal{X} \times \mathcal{X}$, we have $P^{ind}_{\mathbf{x},\mathbf{x}'}(x, x') = P_{\mathbf{x}}(x) P_{\mathbf{x}}(x')$.*

Definition 13 (Completely positive self couplings)**.** *A $P_{\mathbf{x}}$-self-coupling that can be written (decomposed) as a convex combination of independent self-couplings is said to be a completely positive $P_{\mathbf{x}}$-self-coupling. Specifically, if $P_{\mathbf{x},\mathbf{x}'}$ is a $P_{\mathbf{x}}$-self-coupling such that*

$$P_{\mathbf{x},\mathbf{x}'}(x, x') = \sum_{t=0}^{k-1} P_{\mathbf{t}}(t) P_{\mathbf{x}|\mathbf{t}}(x|t) P_{\mathbf{x}|\mathbf{t}}(x'|t) \ \ \forall\ (x, x') \in \mathcal{X} \times \mathcal{X} \tag{9}$$

for some positive integer k and some probability distributions $P_{\mathbf{t}}$ and $P_{\mathbf{x}|\mathbf{t}}$, then $P_{\mathbf{x},\mathbf{x}'}$ is said to be a completely positive $P_{\mathbf{x}}$ self-coupling with time-sharing random variable $\mathbf{t}$. The smallest value of k (if it exists) for which $P_{\mathbf{x},\mathbf{x}'}$ can be thus decomposed is said to be the completely positive rank (CP-rank) of $P_{\mathbf{x},\mathbf{x}'}$ (following convention [], if no such k exists the CP-rank is 1;

the CP-rank for a zero matrix is 0).
Equivalently, if there exists a $k \in \mathbb{Z}^+$, a time sharing random variable $\mathbf{t}$ such that $P_{\mathbf{t}} \in \Delta(\mathcal{T})$ (where $|\mathcal{T}| := k$), and a conditional distribution $P_{\mathbf{x}|\mathbf{t}} \in \Delta(\mathcal{X}|\mathcal{T})$ such that

$$P_{\mathbf{x},\mathbf{x}'} = [P_{\mathbf{t}} P_{\mathbf{x}|\mathbf{t}}^{\otimes 2}]_{\mathbf{x},\mathbf{x}'} \tag{10}$$

The set of all completely positive $P_{\mathbf{x}}$-self-couplings is denoted by $\mathcal{CP}(P_{\mathbf{x}})$. For any $P_{\mathbf{x},\mathbf{x}'} \in \mathcal{CP}(P_{\mathbf{x}})$, the decomposition in (9) *or* (10) *is called a CP-decomposition of $P_{\mathbf{x},\mathbf{x}'}$.*

Specifically, to present our results we introduce a refined version of the confusability set, called the "CP-confusability" set $\mathcal{K}_{\mathcal{CP}}(P_{\mathbf{x}})$, which is the intersection of the confusability set $\mathcal{K}(P_{\mathbf{x}})$ and the CP set $\mathcal{CP}(P_{\mathbf{x}})$. It is the set of all such confusable self-couplings $P_{\mathbf{x},\mathbf{x}'}$ such that the joint distribution $P_{\mathbf{x},\mathbf{x}'}$ has a CP-decomposition in terms of $P_{\mathbf{t}}$ and $P_{\mathbf{x}|\mathbf{t}}$. Below, we define the refined versions of Definition 9 and Definition 11 in terms of the "CP-confusability" set.

Definition 14 (CP-confusability set)**.** *For a distribution $P_{\mathbf{x}} \in \lambda_{\mathbf{x}}$, the CP-confusability set $\mathcal{K}_{\mathcal{CP}}(P_{\mathbf{x}})$ for an AVC $(\mathcal{A})$ is the set of all such $P_{\mathbf{x}}$-self couplings for which there exists a refinement $P_{\mathbf{t},\mathbf{x},\mathbf{x},\mathbf{s},\mathbf{s},\mathbf{y}} \in \Delta(\mathcal{T} \times \mathcal{X}^2 \times \mathcal{S}^2 \times \mathcal{Y})$ such that the following conditions hold:*

1) *(Input consistency): The marginal of the joint distribution $P_{\mathbf{t},\mathbf{x},\mathbf{x}',\mathbf{s},\mathbf{s}',\mathbf{y}}$ onto the pair of random variables is a $P_{\mathbf{x}}$-self coupling i.e., $[P_{\mathbf{t},\mathbf{x},\mathbf{x}',\mathbf{s},\mathbf{s}',\mathbf{y}}]_{\mathbf{x},\mathbf{x}'} \in \mathcal{J}(P_{\mathbf{x}})$.*
2) *(State consistency): The marginal of the joint distribution $P_{\mathbf{t},\mathbf{x},\mathbf{x}',\mathbf{s},\mathbf{s}',\mathbf{y}}$ onto random variables $\mathbf{s}$ and $\mathbf{s}'$ individually respect James' state constraints i.e., $[P_{\mathbf{t},\mathbf{x},\mathbf{x}',\mathbf{s},\mathbf{s}',\mathbf{y}}]_{\mathbf{s}}$, $[P_{\mathbf{t},\mathbf{x},\mathbf{x}',\mathbf{s},\mathbf{s}',\mathbf{y}}]_{\mathbf{s}'} \in \lambda_{\mathbf{s}}$*
3) *(Channel consistency): For any 6-tuple (t, x, x', s, s', y) such that $P_{\mathbf{t},\mathbf{x},\mathbf{x}',\mathbf{s},\mathbf{s}',\mathbf{y}}(t, x, x', s, s', y) > 0$, and the channel law is respected i.e., $[P_{\mathbf{t},\mathbf{x},\mathbf{x}',\mathbf{s},\mathbf{s}',\mathbf{y}}]_{\mathbf{y}|\mathbf{t},\mathbf{x},\mathbf{s}} = W_{\mathbf{y}|\mathbf{x},\mathbf{s}}$ and $[P_{\mathbf{t},\mathbf{x},\mathbf{x}',\mathbf{s},\mathbf{s}',\mathbf{y}}]_{\mathbf{y}|\mathbf{t},\mathbf{x}',\mathbf{s}'} = W_{\mathbf{y}|\mathbf{x},\mathbf{s}}$.*

The CP-confusability set $\mathcal{K}_{\mathcal{CP}}(\lambda_{\mathbf{x}})$ can be defined similarly. It can be represented as the union of all CP-confusability sets $\mathcal{K}_{\mathcal{CP}}(P_{\mathbf{x}})$ where $P_{\mathbf{x}} \in \lambda_{\mathbf{x}}$ i.e.,

$$\mathcal{K}_{\mathcal{CP}}(\lambda_{\mathbf{x}}) := \bigcup_{P_{\mathbf{x}} \in \lambda_{\mathbf{x}}} \mathcal{K}_{\mathcal{CP}}(P_{\mathbf{x}})$$

Definition 15 (CP-confusability set)**.** *Let $P_{\mathbf{x}} \in \lambda_{\mathbf{x}} \subseteq \Delta(\mathcal{X})$. The* CP-confusability set *$\mathcal{K}_{\mathcal{CP}}(P_{\mathbf{x}})$ with input distribution $P_{\mathbf{x}}$ is defined as*

$$\mathcal{K}_{\mathcal{CP}}(P_{\mathbf{x}}) := \left\{ P_{\mathbf{x},\mathbf{x}'} \in \mathcal{CP}(P_{\mathbf{x}}) : \begin{array}{l} \forall \text{ CP-decomposition- } (P_{\mathbf{t}} P_{\mathbf{x}|\mathbf{t}}) \text{ we have,} \\ \exists\ Q^1_{\mathbf{s}|\mathbf{t},\mathbf{x},\mathbf{x}'}, Q^2_{\mathbf{s}'|\mathbf{t},\mathbf{x},\mathbf{x}'} \in \Delta(\mathcal{S}|\mathcal{T} \times \mathcal{X} \times \mathcal{X}) \\ \quad \text{such that for } i = \{1, 2\},\ [P_{\mathbf{t}} P^{\otimes 2}_{\mathbf{x}|\mathbf{t}} Q^i_{\mathbf{s}|\mathbf{t},\mathbf{x},\mathbf{x}'}]_{\mathbf{s}} \in \lambda_{\mathbf{s}}, \\ \bullet\ \forall (t, x, x', y) \in \mathcal{T} \times \mathcal{X} \times \mathcal{X} \times \mathcal{Y}, \\ \sum_{s \in \mathcal{S}} Q^1_{\mathbf{s}|\mathbf{t},\mathbf{x},\mathbf{x}'}(s|t, x, x') W_{\mathbf{y}|\mathbf{x},\mathbf{s}}(y|x, s) = \sum_{s' \in \mathcal{S}} Q^2_{\mathbf{s}'|\mathbf{t},\mathbf{x},\mathbf{x}'}(s'|t, x', x) W_{\mathbf{y}|\mathbf{x}',\mathbf{s}'}(y|x', s') \end{array} \right\}.$$

Note that the symmetrizability equation above in Definition 15, also holds for a single symmetrizing distribution defined via $V_{\mathbf{s}|\mathbf{t},\mathbf{x},\mathbf{x}'}$ such that $V_{\mathbf{s}|\mathbf{t},\mathbf{x},\mathbf{x}'} \triangleq \frac{Q^1_{\mathbf{s}|\mathbf{t},\mathbf{x},\mathbf{x}'} + Q^2_{\mathbf{s}'|\mathbf{t},\mathbf{x},\mathbf{x}'}}{2}$.

Similarly, the definitions of CP-confusability set in Definition 14 and Definition 15 can be shown to be equivalent based on the proof in [ZJB20] and [BDJ+21].

Definition 16 (Approximate constant-composition codes)**.** *A code $\mathcal{C} \subseteq \mathcal{X}^n$ is a $(\delta, P_{\mathbf{x}})$-constant composition code for some distribution $P_{\mathbf{x}} \in \lambda_{\mathbf{x}}$ if each codeword $\underline{x} \in \mathcal{C}$ has type $P_{\underline{\mathbf{x}}}$ that satisfies $d(P_{\underline{\mathbf{x}}}, P_{\mathbf{x}}) \leq \delta$.*

Definition 17 (Approximate constant-composition cloud codes)**.** *A code $\mathcal{C} \subseteq \mathcal{X}^n$ is a $(\delta, \underline{t}, P_{\mathbf{x}|\mathbf{t}})$-constant composition cloud code for some $\underline{t} \in \mathcal{T}^n$ and some distribution $P_{\mathbf{x}|\mathbf{t}} \in \Delta(\mathcal{X}|\mathcal{T})$ if each codeword $\underline{x} \in \mathcal{C}$ satisfies $d(P_{\underline{\mathbf{t}},\underline{\mathbf{x}}},\ P_{\underline{t}} P_{\mathbf{x}|\mathbf{t}}) \leq \delta$.*

A rate $R \geq 0$ is said to be achievable for an AVC if for every $\epsilon > 0$, $\delta > 0$ and every n sufficiently large, there exists a (n, R) code $\mathcal{C}$ with rate:

$$\frac{1}{n} \log |\mathcal{M}| > R - \epsilon$$

such that the maximum probability of error (P_e) given by:

$$P_e := \max_{m \in \mathcal{M}} \max_{s \in \mathcal{P}(\lambda_{\mathbf{s}})} P_e(m, s) \leq \delta$$

The supremum of all such achievable rates is defined as the capacity of AVC $(\mathcal{A})$, denoted by $C(\mathcal{A})$ under the maximum error probability criterion.

C. State-Deterministic and Insufficiently Myopic AVCs

We first define the class of insufficiently myopic AVCs, roughly speaking, as the complement of the class of sufficiently myopic AVCs considered in prior work (e.g., [DJL19a]). While our results hold for general AVCs (whether sufficiently or insufficiently myopic), our achievability arguments are particularly tailored to the insufficiently myopic regime where, for a variety of AVCs, they are able to show a separation between the rate-positivity threshold for omniscient AVCs versus insufficiently myopic AVCs.

Definition 18. *A state-deterministic and myopic AVC $(\mathcal{X}, \mathcal{Z}, \mathcal{S}, \mathcal{Y}, \lambda_{\mathbf{x}}, \lambda_{\mathbf{s}}, W_{\mathbf{z}|\mathbf{x}}, W_{\mathbf{y}|\mathbf{x},\mathbf{s}})$ is defined via the input constraint set $\lambda_{\mathbf{x}}$, the state constraint set $\lambda_{\mathbf{s}}$, Alice to James DMC with channel law $W_{\mathbf{z}|\mathbf{x}}$ and Alice to Bob channel described via the deterministic channel law $W_{\mathbf{y}|\mathbf{x},\mathbf{s}}$. Further, it is said to be insufficiently myopic if for all random variables $\mathbf{u}$ over some alphabet $\mathcal{U}$, any time-sharing variable $\mathbf{t}$ over some alphabet $\mathcal{T}$, and any distribution $P_{\mathbf{t},\mathbf{u},\mathbf{x}}$ such that $[P_{\mathbf{t},\mathbf{u},\mathbf{x}}]_{\mathbf{x}} \in \lambda_{\mathbf{x}}$, there exist a jamming action $V_{\mathbf{s}|\mathbf{z},\mathbf{t}} \in \Delta(\mathcal{S}|\mathcal{T} \times \mathcal{Z})$ such that $I(\mathbf{x};\mathbf{y}|\mathbf{t}) \leq I(\mathbf{x};\mathbf{z}|\mathbf{t})$.*

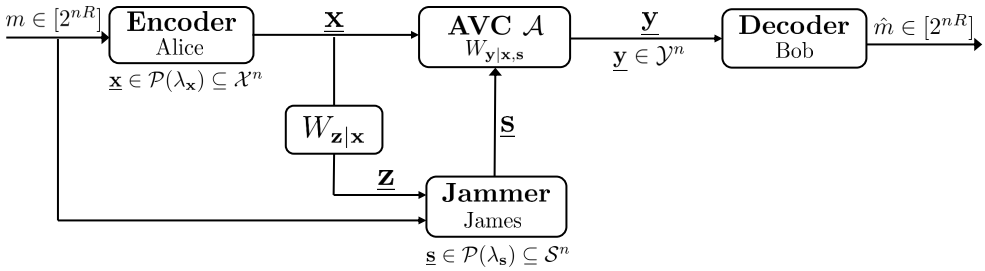


Fig. 2. Communication over a state-deterministic and myopic AVC $(\mathcal{X}, \mathcal{Z}, \mathcal{S}, \mathcal{Y}, \lambda_{\mathbf{x}}, \lambda_{\mathbf{s}}, W_{\mathbf{z}|\mathbf{x}}, W_{\mathbf{y}|\mathbf{x},\mathbf{s}})$.

Context for (in)sufficiently myopic AVCs is presented in Section IX.

We now need two distinct definitions of symmetrization/confusability – one from a converse perspective, and one from an achievability perspective. First, we present the notion relevant to our impossibility results.

Definition 19 (Strong myopic CP-Confusability set ("Converse" perspective))**.** *Given the input distribution $P_{\mathbf{x}} \in \lambda_{\mathbf{x}}$, the confusability set (for the converse perspective) $\mathcal{KCP}^{m}_{strong}(P_{\mathbf{x}})$ for an insufficiently myopic AVC $(\mathcal{A})$ is defined as*

$$\mathcal{KCP}^{m}_{strong}(P_{\mathbf{x}}) := \left\{ P_{\mathbf{x},\mathbf{x}'} \in \mathcal{CP}(P_{\mathbf{x}}) : \begin{array}{l} \forall \text{ CP-decomposition- } (P_{\mathbf{t}} P_{\mathbf{x}|\mathbf{t}}) \text{ we have,} \\ \exists\, V_{\mathbf{s}|\mathbf{z},\mathbf{x}',\mathbf{t}} \in \Delta(\mathcal{S}|\mathcal{Z} \times \mathcal{X} \times \mathcal{T}) \text{ such that} \\ \exists \text{ a constant } \delta > 0 \text{ such that } [P_{\mathbf{t}} P^{\otimes 2}_{\mathbf{x}|\mathbf{t}} W_{\mathbf{z}|\mathbf{x}} V_{\mathbf{s}|\mathbf{z},\mathbf{x}',\mathbf{t}}]_{\mathbf{s}} \in \mathrm{int}_{\delta}(\lambda_{\mathbf{s}}) \\ \bullet\ \forall (t, x, x', y) \in \mathcal{T} \times \mathcal{X}^2 \times \mathcal{Y}, \\ \sum_{z \in \mathcal{Z}} \sum_{s \in \mathcal{S}} W_{\mathbf{z}|\mathbf{x}}(z|x) V_{\mathbf{s}|\mathbf{z},\mathbf{x}',\mathbf{t}}(s|z,x',t) W_{\mathbf{y}|\mathbf{x},\mathbf{s}}(y|x,s) \\ \qquad = \sum_{z \in \mathcal{Z}} \sum_{s \in \mathcal{S}} W_{\mathbf{z}|\mathbf{x}}(z|x') V_{\mathbf{s}|\mathbf{z},\mathbf{x}',\mathbf{t}}(s|z,x,t) W_{\mathbf{y}|\mathbf{x},\mathbf{s}}(y|x',s). \end{array} \right\}$$

Similarly, the confusability set $\mathcal{KCP}^{m}_{strong}(\lambda_{\mathbf{x}})$ is defined as the union of all self-coupling spaces $\mathcal{KCP}^{m}_{strong}(P_{\mathbf{x}})$ with respect to all $P_{\mathbf{x}}$ such that $P_{\mathbf{x}} \in \lambda_{\mathbf{x}}$.

As outlined in Section IX our positive rate achievability results are based on cloud-code constructions (around a cloud-center time-sharing sequence $\underline{t}$) of pre-codewords $\underline{u}$ followed by passing the pre-codeword through an auxiliary channel $P_{\mathbf{x}|\mathbf{u},\mathbf{t}}$. Since the time-sharing sequence $\underline{t}$ is part of code design it is assumed to be known *a priori* by the jammer James; further, for insufficiently myopic AVCs the pre-codeword $\underline{u}$ can also, in general, be inferred by James (since, as indicated above, for such AVCs $I(\mathbf{x};\mathbf{y}|\mathbf{t}) \leq I(\mathbf{x};\mathbf{z}|\mathbf{t})$). Hence a different notion of symmetrizability is needed.

Definition 20 (Weak myopic symmetrizability ("Achievability" perspective))**.** *An input joint distribution $P_{\mathbf{t},\mathbf{u},\mathbf{x}} \in \Delta(\mathcal{T} \times \mathcal{U} \times \mathcal{X})$ such that $[P_{\mathbf{t},\mathbf{u},\mathbf{x}}]_{\mathbf{x}} \in \lambda_{\mathbf{x}}$ is said to be weakly myopically symmetrizable if there exists $V_{\mathbf{s}|\mathbf{t},\mathbf{u},\mathbf{u}',\mathbf{z}} \in \Delta(\mathcal{S}|\mathcal{T} \times \mathcal{U} \times \mathcal{U} \times \mathcal{Z})$ such that*

1) $\sum_{t,u,u',x,z} P_{\mathbf{t}}(t) P_{\mathbf{u}|\mathbf{t}}(u|t) P_{\mathbf{u}|\mathbf{t}}(u'|t) P_{\mathbf{x}|\mathbf{t},\mathbf{u}}(x|t,u) W_{\mathbf{z}|\mathbf{x}}(z|x) V_{\mathbf{s}|\mathbf{t},\mathbf{u},\mathbf{u}',\mathbf{z}}(s|t,u,u',z) \in \lambda_{\mathbf{s}}$;
2) *For every $(t, u, u', y) \in \mathcal{T} \times \mathcal{U}^2 \times \mathcal{Y}$,*

$$\sum_{x,z,s} P_{\mathbf{x}|\mathbf{t},\mathbf{u}}(x|t,u) W_{\mathbf{z}|\mathbf{x}}(z|x) V_{\mathbf{s}|\mathbf{t},\mathbf{u},\mathbf{u}',\mathbf{z}}(s|t,u,u',z) W_{\mathbf{y}|\mathbf{x},\mathbf{s}}(y|x,s)$$
$$= \sum_{x',z',s'} P_{\mathbf{x}|\mathbf{t},\mathbf{u}}(x'|t,u') W_{\mathbf{z}|\mathbf{x}}(z'|x') V_{\mathbf{s}|\mathbf{t},\mathbf{u},\mathbf{u}',\mathbf{z}}(s'|t,u',u,z') W_{\mathbf{y}|\mathbf{x},\mathbf{s}}(y|x',s'). \tag{11}$$

Also, for our achievability results, we restrict ourselves to a class of myopic AVCs that we call input-aware symbolwise myopic with respect to at least some input distributions $P_{\mathbf{t},\mathbf{u},\mathbf{x}}$.

Definition 21. *A channel $W_{\mathbf{z}|\mathbf{x}}$ is said to input-aware symbolwise myopic with respect to distribution $P_{\mathbf{t},\mathbf{u},\mathbf{x}}$ if $H(\mathbf{x}|\mathbf{t}=t,\mathbf{u}=u,\mathbf{z}=z)>0$ for all $(t,u,z)\in\mathcal{T}\times\mathcal{U}\times\mathcal{Z}$.*

Finally, it will help to define the following function which can be shown to be an outer bound on the capacity of general myopic AVCs.

Definition 22. *Let $(\mathcal{X},\mathcal{Z},\mathcal{S},\mathcal{Y},\lambda_{\mathbf{x}},\lambda_{\mathbf{s}},W_{\mathbf{z}|\mathbf{x}},W_{\mathbf{y}|\mathbf{x},\mathbf{s}})$ be a myopic AVC. Define the Elias–Bassalygo-type function C_{EB} as*

$$C_{\mathrm{EB}}:=\max_{P_{\mathbf{x}}\in\lambda_{\mathbf{x}}}\min_{P_{\mathbf{t}|\mathbf{z}}\in\Delta(\mathcal{T}|\mathcal{Z})} I(\mathbf{t};\mathbf{z}),$$

where

1) *the minimization is over the cardinality of $\mathcal{T}$ and $P_{\mathbf{t}|\mathbf{z}}\in\Delta(\mathcal{T}|\mathcal{Z})$ such that for every $t\in\mathcal{T}$, every $P_{\mathbf{x},\mathbf{x}'|\mathbf{t}=t}\in\mathcal{CP}(P_{\mathbf{x}|\mathbf{t}=t})$ and every decomposition $(P_{\mathbf{v}|\mathbf{t}=t},P_{\mathbf{x}|\mathbf{v},\mathbf{t}=t})\in\Delta([K])\times\Delta(\mathcal{X}|[K])$ (for some $K\in\mathbf{Z}_{\geq 1}$) satisfying $P_{\mathbf{x},\mathbf{x}'|\mathbf{t}=t}=\left[P_{\mathbf{v}|\mathbf{t}=t}P^{\otimes 2}_{\mathbf{x}|\mathbf{v},\mathbf{t}=t}\right]_{\mathbf{x},\mathbf{x}'}$, there exists $V_{\mathbf{s}|\mathbf{z},\mathbf{x}',\mathbf{t}=t}\in\Delta(\mathcal{S}|\mathcal{Z}\times\mathcal{X})$ satisfying the following two properties:*
   a) *for every $(x,x',y)\in\mathcal{X}^2\times\mathcal{Y}$,*

$$\sum_{z\in\mathcal{Z}}\sum_{s\in\mathcal{S}} W_{\mathbf{z}|\mathbf{x}}(z|x)V_{\mathbf{s}|\mathbf{z},\mathbf{x}',\mathbf{t}=t}(s|z,x')W_{\mathbf{y}|\mathbf{x},\mathbf{s}}(y|x,s)=\sum_{z\in\mathcal{Z}}\sum_{s\in\mathcal{S}} W_{\mathbf{z}|\mathbf{x}}(z|x')V_{\mathbf{s}|\mathbf{z},\mathbf{x}',\mathbf{t}=t}(s|z,x)W_{\mathbf{y}|\mathbf{x},\mathbf{s}}(y|x',s);$$

   b) *$\left[P_{\mathbf{t}}P_{\mathbf{v}|\mathbf{t}}P^{\otimes 2}_{\mathbf{x}|\mathbf{v},\mathbf{t}}W_{\mathbf{z}|\mathbf{x}}V_{\mathbf{s}|\mathbf{z},\mathbf{x}',\mathbf{t}}\right]_{\mathbf{s}}\in\lambda_{\mathbf{s}}$, where $P_{\mathbf{t}}\in\Delta(\mathcal{T})$ can be computed from the joint distribution $P_{\mathbf{x}}W_{\mathbf{z}|\mathbf{x}}P_{\mathbf{t}|\mathbf{z}}$;*
2) *the mutual information $I(\mathbf{t};\mathbf{z})$ is evaluated using the joint distribution $P_{\mathbf{x}}W_{\mathbf{z}|\mathbf{x}}P_{\mathbf{t}|\mathbf{z}}$.*

VI. Main Results

Theorem 23 (Omniscient Rate Positivity (Matching achievability and converse))**.** *The capacity of a non-state deterministic and omniscient AVC $\mathcal{A}:=(\lambda_{\mathbf{x}},\lambda_{\mathbf{s}},W_{\mathbf{y}|\mathbf{x},\mathbf{s}})$ is positive if and only if $\mathcal{CP}(\lambda_{\mathbf{x}})\not\subset\mathcal{K}(\lambda_{\mathbf{x}})$. In other words, the capacity of a non-state deterministic and omniscient AVC $\mathcal{A}:=(\lambda_{\mathbf{x}},\lambda_{\mathbf{s}},W_{\mathbf{y}|\mathbf{x},\mathbf{s}})$ is positive if and only if $\mathcal{K}_{\mathcal{CP}}(\lambda_{\mathbf{x}})\subset\mathcal{CP}(\lambda_{\mathbf{x}})$.*

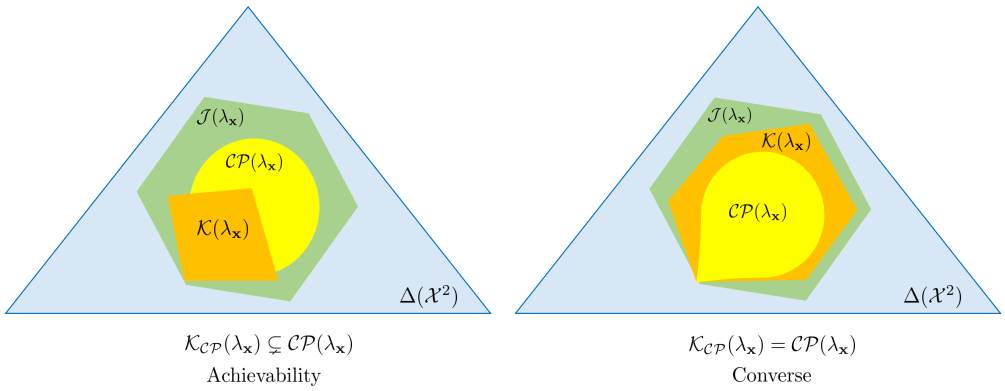


Fig. 3. Rate Positivity over non state-deterministic and omniscient AVC

Proof: The proof appears in Section VII. ∎

We generalize the well-known Elias-Bassalygo rate upper bound [Bas65] for bit-flip AVCs to general omniscient AVCs, which results in fact follows from the further generalization to general myopic AVCs (under deterministic coding) in Theorem 28, to which we refer the reader.

Theorem 24 (Capacity lower bound)**.** *The capacity of $\mathcal{A}$ is at least*

$$\max_{P^*_{\mathbf{t}}P^*_{\mathbf{x}|\mathbf{t}}}\min\left\{\min_{\substack{Q_{\mathbf{s}|\mathbf{t},\mathbf{x}}\in\Delta(\mathcal{S}|\mathcal{T}\times\mathcal{X})\\ s.t.[P_{\mathbf{t}}P_{\mathbf{x}|\mathbf{t}}Q_{\mathbf{s}|\mathbf{t},\mathbf{x}}]_{\mathbf{s}}\in\lambda_{\mathbf{s}}}} I(\mathbf{x};\mathbf{y}|\mathbf{t}),\ \min_{\substack{P_{\mathbf{t},\mathbf{x},\mathbf{x}'}\\ s.t\ [P_{\mathbf{t},\mathbf{x},\mathbf{x}'}]_{\mathbf{x},\mathbf{x}'}\in\mathcal{K}_{\mathcal{CP}}(P_{\mathbf{x}})}} I(\mathbf{x};\mathbf{x}'|\mathbf{t})\right\},\tag{12}$$

Here, the maximum is taken over the input distribution $P_{\mathbf{x}}\in\lambda_{\mathbf{x}}$ such that $\mathcal{K}_{\mathcal{CP}}(P_{\mathbf{x}})\subsetneq\mathcal{CP}(P_{\mathbf{x}})$. This implies there exists a CP-self coupling $P_{\mathbf{x},\mathbf{x}'}\notin\mathcal{K}_{\mathcal{CP}}(P_{\mathbf{x}})$ which further induces a CP decomposition $(P^*_{\mathbf{t}},P^*_{\mathbf{x}|\mathbf{t}})$ such that there does not exists a $Q_{\mathbf{s}|\mathbf{t},\mathbf{x},\mathbf{x}'}$

satisfying the symmetrizability equation in Definition 15. Therefore, for each $P_{\mathbf{x}} \in \lambda_{\mathbf{x}}$, the maximization is also implicitly taken over CP-self coupling $P_{\mathbf{x},\mathbf{x}'} \in \mathcal{CP}(P_{\mathbf{x}}) \backslash \mathcal{K}_{\mathcal{CP}}(P_{\mathbf{x}})$ and a pair $(P^*_{\mathbf{t}}, P^*_{\mathbf{x}|\mathbf{t}})$ induced by the CP-decomposition of $P_{\mathbf{x},\mathbf{x}'}$. Further, the former mutual information is evaluated w.r.t. the appropriate marginalization of the joint distribution $P_{\mathbf{t}} P_{\mathbf{x}|\mathbf{t}} Q_{\mathbf{s}|\mathbf{t},\mathbf{x}} W_{\mathbf{y}|\mathbf{x},\mathbf{s}}$, and the latter mutual information according to the appropriate marginalization of $P_{\mathbf{t}} P_{\mathbf{x}|\mathbf{t}} P_{\mathbf{x}|\mathbf{t}} V_{\mathbf{s}|\mathbf{x},\mathbf{x}',\mathbf{t}} W_{\mathbf{y}|\mathbf{x},\mathbf{s}}$ minimized over the confusability set induced by Alice's choice $P^*_{\mathbf{t}} P^*_{\mathbf{x}|\mathbf{t}}$.

Proof: The proof appears in Section VIII. ■

We will use the following shorthands in the rest of the paper to represent the mutual information terms inside the max-min expressions.

$$\mathrm{DMC}(P_{\mathbf{t}} P_{\mathbf{x}|\mathbf{t}}) \triangleq \min_{\substack{Q_{\mathbf{s}|\mathbf{t},\mathbf{x}} \in \Delta(\mathcal{S}|\mathcal{T}\times\mathcal{X}) \\ \text{s.t.}[P_{\mathbf{t}} P_{\mathbf{x}|\mathbf{t}} Q_{\mathbf{s}|\mathbf{t},\mathbf{x}}]_{\mathbf{s}} \in \lambda_{\mathbf{s}}}} I(\mathbf{x};\mathbf{y}|\mathbf{t}) \tag{13}$$

$$\mathrm{GV}(P_{\mathbf{t}} P_{\mathbf{x}|\mathbf{t}}) \triangleq \min_{\substack{P_{\mathbf{t},\mathbf{x},\mathbf{x}'} \\ \text{s.t } [P_{\mathbf{t},\mathbf{x},\mathbf{x}'}]_{\mathbf{x},\mathbf{x}'} \in \mathcal{K}_{\mathcal{CP}}(P_{\mathbf{x}})}} I(\mathbf{x};\mathbf{x}'|\mathbf{t}) \tag{14}$$

Remark 25. *Note that the former mutual information term in* (12) *(with a uniform time-sharing variable* $\mathbf{t}$*) corresponds to the list-decoding capacity of the omniscient AVC, whereas in the case of state-deterministic omniscient AVCs the latter mutual information corresponds to the generalized Gilbert-Varshamov bound (corresponding to unique decoding) implicit in the cloud code contruction in [WBBJ19], outlined in detail in [BDJ+21]. But since the list-decoding capacity of a channel is never smaller than the unique decoding capacity, hence in the case of state-deterministic AVCs the latter term is never larger than the former – this retrieves the generalized Gilbert-Varshamov bound of [WBBJ19], [BDJ+21].*

Corollary 26 (Omniscient Non-SD Bit-flip Capacity Positivity (Matching rate-positivity; rate-achievability, converse and capacity))**.** *The capacity of a non state deterministic and omniscient bit-flip AVC A-BSC*$(\frac{1}{2}, p, r)$ *is positive if and only if* $p < 1/4$*, irrespective of* $0 \leq r < 1/2$*. Furthermore, the capacity of a non state deterministic and omniscient bit-flip AVC A-BSC*$(\frac{1}{2}, p, r)$ *is given by* $1 - H(p * r)$ *for* $p \leq \frac{r}{1+2r}$*.*

Proof: The proof appears in Section XI (see Example 42). ■

For context to our results for myopic AVCs, we restate here the following impossibility result from [BDJ+20]

Theorem 27 (Impossibility)**.** *If for every feasible input distribution $P_{\mathbf{x}} \in \lambda_{\mathbf{x}}$, the CP-set $\mathcal{CP}(P_{\mathbf{x}})$ is equal to the CP-confusability set $\mathcal{K}_{\mathcal{CP}^m_{strong}}(P_{\mathbf{x}})$, the capacity $\mathbb{C}$ of a myopic AVC $\mathcal{A}$ is zero. (see fig. 4) i.e.,*

$$\mathcal{CP}(P_{\mathbf{x}}) = \mathcal{K}_{\mathcal{CP}^m_{strong}}(P_{\mathbf{x}}),\ \forall\ P_{\mathbf{x}} \in \lambda_{\mathbf{x}} \implies \mathbb{C} = 0 \tag{15}$$

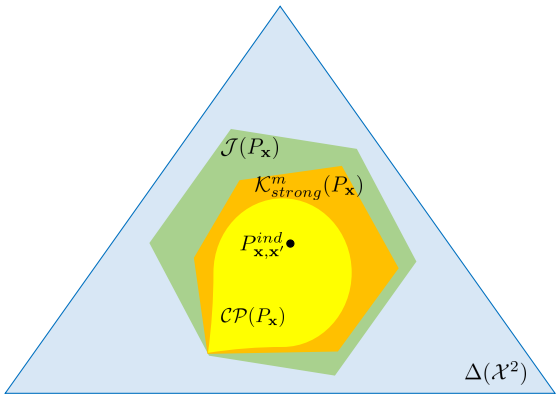


Fig. 4. Converse for myopic AVCs (27)

The above result presents the generalized Plotkin bound for myopic channels. For myopic AVCs with positive capacity, we now present a strengthened converse by generalizing the Elias-Bassalygo bound to myopic AVCs.

Theorem 28 (Elias–Bassalygo bound for myopic AVCs)**.** *For a myopic AVC $\mathcal{A}$, the capacity under deterministic encoding is at most C_{EB} as defined in Definition 22.*

Proof: The proof appears in Section X. ■

It is important to note that the above bound holds for all myopic AVCs when the encoder is not allowed to randomize.

Corollary 29. *For a state-deterministic and myopic binary alphabet AVC with parameters q and p i.e., Alice to James channel is BSC(q) and the jammer can flip upto pn bits of the transmitted codeword, the myopic EB bound implies that whenever $q \le p \le \frac{1}{4(1-q)}$, the rate $R > 1 - H(w(p,q)) = 1 - H\Big(\frac{(q+1)-\sqrt{\frac{1-q}{1-2q}}\sqrt{-4p+2q^2+q+1}}{2}\Big)$ is not achievable.*

Proof: The proof for state-deterministic and insufficiently myopic bit-flip AVC appears in Section XI (see Example 47). ∎

Finally, we present our rate positivity achievability result for AVCs under stochastic encoding.

Theorem 30 (SD Myopic Rate Positivity (Achievability))**.** *The capacity of a state-deterministic AVC $\mathcal{A}$ is positive if there exists a input-aware symbolwise myopic $P_{\mathbf{t},\mathbf{u},\mathbf{x}} \in \Delta(\mathcal{T}\times\mathcal{U}\times\mathcal{X})$ satisfying $[P_{\mathbf{t},\mathbf{u},\mathbf{x}}]_{\mathbf{x}} \in \lambda_{\mathbf{x}}$ and $P_{\mathbf{t},\mathbf{u},\mathbf{x}}$ is not weakly myopically symmetrizable.*

VII. Rate Positivity for General (Non-State Deterministic) Omniscient AVCs

A. Converse

Theorem 31 (Plotkin bound for omniscient AVCs)**.** *Let $(\mathcal{X},\mathcal{S},\mathcal{Y},\lambda_{\mathbf{x}},\lambda_{\mathbf{s}},W_{\mathbf{y}|\mathbf{x},\mathbf{s}})$ be an omniscient AVC. Suppose that for every $P_{\mathbf{x}} \in \lambda_{\mathbf{x}}$, every CP-self-coupling $P_{\mathbf{x},\mathbf{x}'} \in \mathcal{CP}(P_{\mathbf{x}})$ also satisfies $P_{\mathbf{x},\mathbf{x}'} \in \mathcal{K}_{\mathcal{CP}}(P_{\mathbf{x}})$ i.e., every decomposition $(P_{\mathbf{t}}, P_{\mathbf{x}|\mathbf{t}}) \in \Delta(\mathcal{T}) \times \Delta(\mathcal{X}|\mathcal{T})$ (for some $K \ge \mathbf{Z}_{\ge 1}$) satisfying $P_{\mathbf{x},\mathbf{x}'} = \Big[P_{\mathbf{t}}P_{\mathbf{x}|\mathbf{t}}^{\otimes 2}\Big]_{\mathbf{x},\mathbf{x}'}$, there exists $V_{\mathbf{s}|\mathbf{x},\mathbf{x}',\mathbf{t}} \in \Delta(\mathcal{S}|\mathcal{X}^2\times\mathcal{T})$ such that*

1) *there exists a constant $\delta > 0$ such that $\Big[P_{\mathbf{t}}P_{\mathbf{x}|\mathbf{t}}^{\otimes 2}V_{\mathbf{s}|\mathbf{x},\mathbf{x}',\mathbf{t}}\Big]_{\mathbf{s}} \in \mathrm{int}_\delta(\lambda_{\mathbf{s}})$;*[1]
2) *for every $(t,x,x',y) \in \mathcal{T}\times\mathcal{X}^2\times\mathcal{Y}$,*

$$\sum_{s\in\mathcal{S}} V_{\mathbf{s}|\mathbf{x},\mathbf{x}',\mathbf{t}}(s|x,x',t)W_{\mathbf{y}|\mathbf{x},\mathbf{s}}(y|x,s) = \sum_{s\in\mathcal{S}} V_{\mathbf{s}|\mathbf{x},\mathbf{x}',\mathbf{t}}(s|x',x,t)W_{\mathbf{y}|\mathbf{x},\mathbf{s}}(y|x',s).$$

Let $\mathcal{C} \subset \mathcal{X}^n$ be a code satisfying

1) *$P_{\underline{x}} \in \lambda_{\mathbf{x}}$ for every $\underline{x} \in \mathcal{C}$;*
2) *the average probability of error of $\mathcal{C}$ approaches 0 as $n \to \infty$.*

Then there exists a constant $M > 0$ that depends on δ (and on the alphabet sizes as well) but not on n such that $|\mathcal{C}| \le M$. In particular, the rate of $\mathcal{C}$ is vanishing as $n \to \infty$.

Proof: When the channel law $W_{\mathbf{y}|\mathbf{x},\mathbf{s}}$ is a zero-one law, i.e., for each $(x,s) \in \mathcal{X}\times\mathcal{S}$, there is a unique $y \in \mathcal{Y}$ such that $W_{\mathbf{y}|\mathbf{x},\mathbf{s}}(y|x,s) = 1$, the above theorem was proved in [WBBJ19]. The present proof is concerned with a general channel law $W_{\mathbf{y}|\mathbf{x},\mathbf{s}} \in \Delta(\mathcal{Y}|\mathcal{X},\mathcal{S})$.

Suppose, towards a contradiction, that $|\mathcal{C}| > M$. First, by a standard (approximate) constant composition reduction, one can find a subcode $\mathcal{C}' \subset \mathcal{C}$ such that

1) $|\mathcal{C}'| \ge c_1|\mathcal{C}|$ for some constant $c_1 > 0$ that only depends on δ and $|\mathcal{X}|$;
2) there exists $P_{\mathbf{x}} \in \lambda_{\mathbf{x}}$ such that $\big\|P_{\underline{x}} - P_{\mathbf{x}}\big\|_\infty \le \delta/10$ for every $\underline{x} \in \mathcal{C}'$.

It was then shown in [BDJ+20], [ZJB21] that if $\{\underline{\mathbf{x}}_1, \underline{\mathbf{x}}_2\}$ ($\underline{\mathbf{x}}_1 \ne \underline{\mathbf{x}}_2$) is a uniform random pair of codewords in $\mathcal{C}'$, there must exist $P_{\mathbf{x},\mathbf{x}'} \in \mathcal{CP}(P_{\mathbf{x}})$ such that with probability at least c_2, it holds that

$$\big\|P_{\underline{\mathbf{x}}_1,\underline{\mathbf{x}}_2} - P_{\mathbf{x},\mathbf{x}'}\big\|_\infty \le \delta/10. \tag{16}$$

Here the constant $c_2 > 0$ depends on δ (and the alphabet sizes) but not on n.

Consider such a distribution $P_{\mathbf{x},\mathbf{x}'} \in \mathcal{CP}(P_{\mathbf{x}})$ and any of its decomposition $(P_{\mathbf{t}}, P_{\mathbf{x}|\mathbf{t}}) \in \Delta(\mathcal{T}) \times \Delta(\mathcal{X}|\mathcal{T})$ (let $\mathcal{T} := [K]$ for some constant K that only depends on $|\mathcal{X}|$) satisfying $\Big[P_{\mathbf{t}}P_{\mathbf{x}|\mathbf{t}}^{\otimes 2}\Big]_{\mathbf{x},\mathbf{x}'} = P_{\mathbf{x},\mathbf{x}'}$. By the symmetrizability assumption in the theorem, one can find a symmetrizing distribution $V_{\mathbf{s}|\mathbf{x},\mathbf{x}',\mathbf{t}} \in \Delta(\mathcal{S}|\mathcal{X}^2\times\mathcal{T})$ satisfying Properties 1 and 2.

Now suppose a uniformly random codeword $\underline{\mathbf{x}} \in \mathcal{C}$ was transmitted by Alice and observed by James. James uniformly samples a spoofing codeword $\underline{\mathbf{x}}'$ from $\mathcal{C}'$. Then with probability at least $c_1\Big(1 - \frac{1}{c_1|\mathcal{C}|-1}\Big)c_2 > c_1\Big(1 - \frac{1}{c_1M-1}\Big)c_2$, the following three events happen simultaneously: (i) $\underline{\mathbf{x}} \in \mathcal{C}'$; (ii) $\underline{\mathbf{x}}' \ne \underline{\mathbf{x}}$; (iii) $\big\|P_{\underline{\mathbf{x}},\underline{\mathbf{x}}'} - P_{\mathbf{x},\mathbf{x}'}\big\|_2 \le \delta/10$. Given $\underline{\mathbf{x}},\underline{\mathbf{x}}'$ and the decomposition $(P_{\mathbf{t}}, P_{\mathbf{x}|\mathbf{t}})$ of $P_{\mathbf{x},\mathbf{x}'}$, James can find a sequence $\underline{t} \in \mathcal{T}^n$ such that

$$\Big\|P_{\underline{t},\underline{\mathbf{x}},\underline{\mathbf{x}}'} - P_{\mathbf{t}}P_{\mathbf{x}|\mathbf{t}}^{\otimes 2}\Big\|_2 \le \delta/5. \tag{17}$$

[1]Here $\mathrm{int}_\delta(\lambda_{\mathbf{s}})$ denotes the δ-interior of $\lambda_{\mathbf{s}}$ w.r.t. the ℓ_∞ distance.

James then samples the attack vector $\underline{\mathbf{s}} \in \mathcal{S}^n$ in the following way

$$\underline{\mathbf{s}} \sim \prod_{i=1}^{n} V_{\mathbf{s}|\mathbf{x}=\underline{\mathbf{x}}(i),\mathbf{x}'=\underline{\mathbf{x}}'(i),\mathbf{t}=\underline{t}(i)}.$$

Note that the expected type of $\underline{\mathbf{s}}$ is given by $\left[P_{\underline{t},\underline{\mathbf{x}},\underline{\mathbf{x}}'}V_{\mathbf{s}|\mathbf{x},\underline{\mathbf{x}}',\mathbf{t}}\right]_{\mathbf{s}}$. By Equation (17), the joint distribution $P_{\underline{t},\underline{\mathbf{x}},\underline{\mathbf{x}}'}V_{\mathbf{s}|\mathbf{x},\underline{\mathbf{x}}',\mathbf{t}}$ is $(\delta/5)$-close to $P_{\mathbf{t}}P^{\otimes 2}_{\mathbf{x}|\mathbf{t}}V_{\mathbf{s}|\mathbf{x},\mathbf{x}',\mathbf{t}}$. By Property 1 of the symmetrizing distribution $V_{\mathbf{s}|\mathbf{x},\mathbf{x}',\mathbf{t}}$, we have $\left[P_{\mathbf{t}}P^{\otimes 2}_{\mathbf{x}|\mathbf{t}}V_{\mathbf{s}|\mathbf{x},\mathbf{x}',\mathbf{t}}\right]_{\mathbf{s}}$ is in $\mathrm{int}_{\delta}(\lambda_{\mathbf{s}})$. Hence, by the Chernoff bound and the triangle inequality, with probability at least $1-e^{-\Omega(n)}$, the actual type $P_{\underline{\mathbf{s}}}$ is also inside $\mathrm{int}_{\delta}(\lambda_{\mathbf{s}})$.

We then argue that the channel is "symmetrized" under the attack $\underline{\mathbf{s}}$. For any $\underline{y} \in \mathcal{Y}^n$, we have

$$\begin{aligned}
\mathop{\mathbb{E}}_{\underline{\mathbf{s}}}\left[W^{\otimes n}_{\mathbf{y}|\mathbf{x},\mathbf{s}}(\underline{y}|\underline{\mathbf{x}},\underline{\mathbf{s}})\right] &= \sum_{\underline{s}\in\mathcal{S}^n}\prod_{i=1}^{n}\left(V_{\mathbf{s}|\mathbf{x},\mathbf{x}',\mathbf{t}}(\underline{s}(i)|\underline{\mathbf{x}}(i),\underline{\mathbf{x}}'(i),\underline{t}(i))W_{\mathbf{y}|\mathbf{x},\mathbf{s}}(\underline{y}(i)|\underline{\mathbf{x}}(i),\underline{s}(i))\right)\\
&= \prod_{i=1}^{n}\sum_{\underline{s}(i)\in\mathcal{S}}\left(V_{\mathbf{s}|\mathbf{x},\mathbf{x}',\mathbf{t}}(\underline{s}(i)|\underline{\mathbf{x}}(i),\underline{\mathbf{x}}'(i),\underline{t}(i))W_{\mathbf{y}|\mathbf{x},\mathbf{s}}(\underline{y}(i)|\underline{\mathbf{x}}(i),\underline{s}(i))\right)\\
&= \prod_{i=1}^{n}\sum_{\underline{s}(i)\in\mathcal{S}}\left(V_{\mathbf{s}|\mathbf{x},\mathbf{x}',\mathbf{t}}(\underline{s}(i)|\underline{\mathbf{x}}'(i),\underline{\mathbf{x}}(i),\underline{t}(i))W_{\mathbf{y}|\mathbf{x},\mathbf{s}}(\underline{y}(i)|\underline{\mathbf{x}}'(i),\underline{s}(i))\right) \qquad (18)\\
&= \mathop{\mathbb{E}}_{\underline{\mathbf{s}}'}\left[W^{\otimes n}_{\mathbf{y}|\mathbf{x},\mathbf{s}}(\underline{y}|\underline{\mathbf{x}}',\underline{\mathbf{s}}')\right],
\end{aligned}$$

where $\underline{\mathbf{s}}'$ is generated in the following way

$$\underline{\mathbf{s}}' \sim \prod_{i=1}^{n} V_{\mathbf{s}|\mathbf{x}=\underline{\mathbf{x}}'(i),\mathbf{x}'=\underline{\mathbf{x}}(i),\mathbf{t}=\underline{t}(i)}.$$

Equation (18) follows from Property 2 of the symmetrizing distribution $V_{\mathbf{s}|\mathbf{x},\mathbf{x}',\mathbf{t}}$. From here, one can easily argue (cf. Proof of Lemma 1, [CN88]) that conditioned on the events (i)-(iii), the average probability of error of $\mathcal{C}'$ is at least, say, $1/4$. By Property 1 of $\mathcal{C}'$, this further implies that the average probability of error of $\mathcal{C}$ (without conditioning on the events (i)-(iii)) is at least a constant that is independent of n. This contradicts the assumption 2 (vanishing error) in the theorem and the proof is therefore finished. ■

B. Achievability

In this section, we begin by showing the existence of a cloud code with some positive rate we first present the code construction, followed by the description of the encoder and decoder. We then conclude by presenting the probability of error analysis and showing that the error probability is negligible.

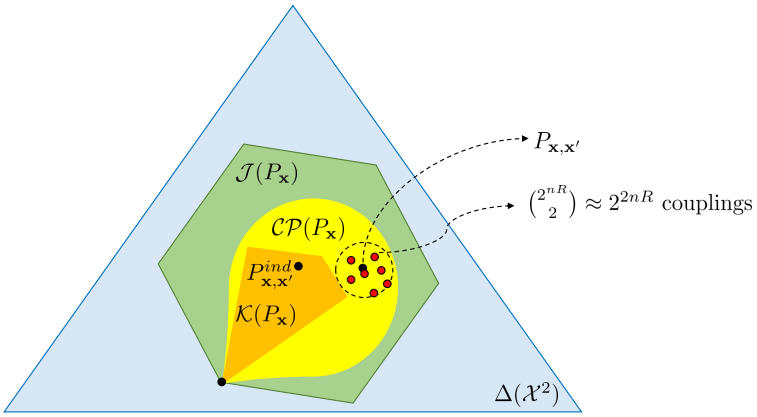
Our code construction is based on the scheme in [WBBJ19] for state-deterministic and omniscient AVCs. Note that the condition $\mathcal{K}_{\mathcal{CP}}(\lambda_{\mathbf{x}}) \subset \mathcal{CP}(\lambda_{\mathbf{x}})$ implies that there exists a input distribution $P_{\mathbf{x}} \in \lambda_{\mathbf{x}}$ such that $\mathcal{K}_{\mathcal{CP}}(P_{\mathbf{x}}) \subset \mathcal{CP}(P_{\mathbf{x}})$. This further implies that a $P_{\mathbf{x}}$ self-coupling $P_{\mathbf{x},\mathbf{x}'}$ can be chosen such that $P_{\mathbf{x},\mathbf{x}'} \in \mathcal{CP}(P_{\mathbf{x}})$ (has a CP-decomposition) and $P_{\mathbf{x},\mathbf{x}'} \notin \mathcal{K}_{\mathcal{CP}}(P_{\mathbf{x}})$.

Lemma 32. *Fix a $P_{\mathbf{x}} \in \lambda_{\mathbf{x}}$ such that $\mathcal{K}_{\mathcal{CP}}(P_{\mathbf{x}}) \subset \mathcal{CP}(P_{\mathbf{x}})$. Now, choose a $P_{\mathbf{x},\mathbf{x}'} \in \mathcal{CP}(P_{\mathbf{x}})$ such that $P_{\mathbf{x},\mathbf{x}'} \notin \mathcal{K}_{\mathcal{CP}}(P_{\mathbf{x}})$. Then, for some $c' > 0$ there exists a $(\delta, \underline{\mathbf{t}}, P_{\mathbf{x}|\mathbf{t}})$-approximate constant composition cloud code $\mathcal{C} := \{\underline{x}_i\}_{i=1}^{2^{nR}}$ with 2^{nR} codewords (with rate $R(c')$) such that for every distinct pair of codewords $\underline{\mathbf{x}}, \underline{\mathbf{x}}' \in \mathcal{C}$, we have $\|P_{\underline{\mathbf{x}},\underline{\mathbf{x}}'} - P_{\mathbf{x},\mathbf{x}'}\| \leq c'$ with very high probability.*

Proof: Since $P_{\mathbf{x},\mathbf{x}'}$ is a completely positive (CP) $P_{\mathbf{x}}$-self-coupling, there exists a time-sharing random variable $\mathbf{t}$ with distribution $P_{\mathbf{t}}$ over a finite alphabet $\mathcal{T} = \mathcal{T}$ and component random variables $\mathbf{x}_t$ for $t \in \mathcal{T}$, such that $P_{\mathbf{x},\mathbf{x}'}$ has a CP-decomposition in terms of $P_{\mathbf{t}}$ and $P_{\mathbf{x}|\mathbf{t}}$. Moreover, note that the CP-self-coupling $P_{\mathbf{x},\mathbf{x}'}$ is chosen such that $P_{\mathbf{x},\mathbf{x}'} \notin \mathcal{K}_{\mathcal{CP}}(P_{\mathbf{x}})$. Therefore, there exists some constant $c' > 0$ such that the ℓ_1 distance between $P_{\mathbf{x},\mathbf{x}'}$ and any $P_{\mathbf{x}}$-self-coupling in $\mathcal{K}_{\mathcal{CP}}(P_{\mathbf{x}})$ is at least $c' > 0$.

Fix a time-sharing sequence $\underline{\mathbf{t}}$ such that $P_{\mathbf{t}}P_{\mathbf{x}|\mathbf{t}}$ is a CP-decomposition of $P_{\mathbf{x},\mathbf{x}'}$. Each codeword $\underline{\mathbf{x}} \in \mathcal{C}$ is generated by sampling the first $nP_{\mathbf{t}}(0)$ indices i.i.d according to the distribution $P_{\mathbf{x}|\mathbf{t}=0}$, the next $nP_{\mathbf{t}}(1)$ indices are sampled i.i.d according to $P_{\mathbf{x}|\mathbf{t}=1}$ and so on until the last $nP_{\mathbf{t}}(k-1)$ are sampled according to the distribution $P_{\mathbf{x}|\mathbf{t}=k-1}$. Such a construction gives an $(\delta, P_{\mathbf{t}}, P_{\mathbf{x}|\mathbf{t}})$ approximate constant composition cloud code such that each codeword has type close to $[P_{\mathbf{t}}P_{\mathbf{x}|\mathbf{t}}]_{\mathbf{x}}$. Therefore, the expected joint type between the codewords in such a code is close to $P_{\mathbf{x},\mathbf{x}'}$.

Now, we use the Chernoff bound to show that for a given pair of codewords $(\underline{x}, \underline{x}')$ such that $\underline{x}, \underline{x}' \in \mathcal{C}$, the probability that



their joint-type $P_{\underline{x},\underline{x}'}$ differs from $P_{\mathbf{x},\mathbf{x}'}$ in ℓ_1 distance by more than c' is exponentially decaying i.e.,

$$\mathbb{P}\left(\|P_{\underline{x},\underline{x}'} - P_{\mathbf{x},\mathbf{x}'}\| > c'\right) \leq k \exp\left(-nc'^2 \min_{t\in\mathcal{T}}(P_{\mathbf{t}}(t))/3\right)$$

On taking union bound over all possible pairs of 2^{nR} codewords with in the code $\mathcal{C}$, we obtain,

$$\begin{aligned}\mathbb{P}\left(\exists\ \underline{x},\underline{x}' \in \mathcal{C} : \|P_{\underline{x},\underline{x}'} - P_{\mathbf{x},\mathbf{x}'}\| > c'\right) &\leq \binom{2^{nR}}{2} k \exp\left(-nc'^2 \min_{t\in\mathcal{T}}(P_{\mathbf{t}}(t))/3\right) \\ &< 2^{2nR} k \exp\left(-nc'^2 \min_{t\in\mathcal{T}}(P_{\mathbf{t}}(t))/3\right) \end{aligned} \tag{19}$$

Therefore, from (19), for sufficiently large blocklength n, there exists an $(\delta, \underline{\mathbf{t}}, P_{\mathbf{x}|\mathbf{t}})$-approximate constant composition cloud code with positive rate $R \leq \frac{c'^2 \min_{t_i\in\mathcal{T}}(P_{\mathbf{t}}(t_i))}{ln(2)6}$. This completes the proof of our lemma. ■

Encoder: Given a message m, the codeword $\underline{x} \in \mathcal{C}$ is transmitted over the non state-deterministic and omniscient AVC $W_{\mathbf{y}|\mathbf{x},\mathbf{s}}$.
Decoder: We use the following decoder with parameter $\eta > 0$ (we specify the exact choice of $\eta(\epsilon)$, where $\eta(\epsilon) \to 0$ as $\epsilon \to 0$, later in Lemma 34). The decoder observes $\underline{\mathbf{y}} = \underline{y}$ and outputs the estimate $\phi(\underline{y})$ according to the following rule (cf. Definition 1):

$$\phi(\underline{\mathbf{y}}) := \begin{cases} m & \text{if uniquely for } \underline{\mathbf{x}} \in \mathcal{C}, \exists\ \underline{\mathbf{s}} : P_{\underline{\mathbf{s}}} \in \lambda_{\mathbf{s}} \text{ s.t. } \max_{t\in\mathcal{T},x\in\mathcal{X},s\in\mathcal{S}} D(P_{\underline{\mathbf{y}}|\mathbf{t},\underline{\mathbf{x}},\underline{\mathbf{s}}}(\cdot|t,x,s)\|W_{\mathbf{y}|\mathbf{x},\mathbf{s}}(\cdot|x,s)) < \eta \\ 0 & \text{otherwise} \end{cases} \tag{20}$$

Before we proceed with the error analysis, we state the following useful lemmas.

Lemma 33. *Let $\epsilon > 0$ be some constant and let $P_{\mathbf{x},\mathbf{x}'}$ be a completely positive (CP) $P_{\mathbf{x}}$-self coupling (where $P_{\mathbf{x}} \in \lambda_{\mathbf{x}}$) with CP-decomposition $P_{\mathbf{t}}P_{\mathbf{x}|\mathbf{t}}$ such that $\|P_{\mathbf{x},\mathbf{x}'} - \mathcal{K}_{\mathcal{CP}}(P_{\mathbf{x}})\| = \epsilon > 0$. Then there exists an $\epsilon_1(\epsilon) > 0$, where $\epsilon_1(\epsilon) \to 0$ as $\epsilon \to 0$, such that for every refinement $P_{\mathbf{t},\mathbf{x},\mathbf{x}',\mathbf{s},\mathbf{s}'\mathbf{y}} := Q$ satisfying $[Q]_{\mathbf{x},\mathbf{x}'} = P_{\mathbf{x},\mathbf{x}'}$, $[Q]_{\mathbf{t},\mathbf{x}} = [Q]_{\mathbf{t},\mathbf{x}'} = P_{\mathbf{t}}P_{\mathbf{x}|\mathbf{t}}$, and $[Q]_{\mathbf{s}}, [Q]_{\mathbf{s}'} \in \lambda_{\mathbf{s}}$, we have*

$$\|[Q]_{\mathbf{y}|\mathbf{t},\mathbf{x},\mathbf{s}}(\cdot|t,x,s) - [Q]_{\mathbf{y}|\mathbf{t},\mathbf{x}',\mathbf{s}'}(\cdot|t,x,s)\| \geq \epsilon_1(\epsilon) \tag{21}$$

for some $(t,x,s) \in \mathcal{T}\times\mathcal{X}\times\mathcal{S}$.

Proof: Recall the definition of CP-confusability $\mathcal{K}_{\mathcal{CP}}(P_{\mathbf{x}})$ in Definition 14. As, the CP-self-coupling $P_{\mathbf{x},\mathbf{x}'} \notin \mathcal{K}_{\mathcal{CP}}(P_{\mathbf{x}})$, and, in particular, $\|P_{\mathbf{x},\mathbf{x}'} - \mathcal{K}(P_{\mathbf{x}})\| = \epsilon > 0$, it follows that no $\mathcal{A}$-refinement $P_{\mathbf{t},\mathbf{x},\mathbf{x}'\mathbf{s},\mathbf{s}',\mathbf{y}}$ of $P_{\mathbf{x},\mathbf{x}'}$ exists satisfying all the conditions in Definition 14 (otherwise, $P_{\mathbf{x},\mathbf{x}'}$ is confusable under $\mathcal{A}$ which is not the case). This implies that for every $P_{\mathbf{t},\mathbf{x},\mathbf{x}'\mathbf{s},\mathbf{s}'\mathbf{y}} := Q$ such that $[Q]_{\mathbf{x},\mathbf{x}'} = P_{\mathbf{x},\mathbf{x}'}$, $[Q]_{\mathbf{t},\mathbf{x}} = [Q]_{\mathbf{t},\mathbf{x}'} = P_{\mathbf{t}}P_{\mathbf{x}|\mathbf{t}}$, $[P_{\mathbf{t}}P_{\mathbf{x}|\mathbf{t}}]_{\mathbf{x}} \in \lambda_{\mathbf{x}}$ and $[Q]_{\mathbf{s}}, [Q]_{\mathbf{s}'} \in \lambda_{\mathbf{s}}$, we require that the 'channel consistency' condition should not hold. Thus, either $[Q]_{\mathbf{y}|\mathbf{t},\mathbf{x},\mathbf{s}}$ and $W_{\mathbf{y}|\mathbf{x},\mathbf{s}}$ are not identically same, or $[Q]_{\mathbf{y}|\mathbf{t},\mathbf{x}',\mathbf{s}'}$ and $W_{\mathbf{y}|\mathbf{x},\mathbf{s}}$ are not identically same, or both. In particular, there exists $\epsilon_a(\epsilon)$, $\epsilon_b(\epsilon)$, at least one of which is strictly positive, i.e., $\exists\ (t,x,s)$ such that

$$\begin{aligned}\|[Q]_{\mathbf{y}|\mathbf{t},\mathbf{x},\mathbf{s}}(\cdot|t,x,s) - W_{\mathbf{y}|\mathbf{x},\mathbf{s}}(\cdot|x,s)\| &\geq \epsilon_a(\epsilon) \\ \|[Q]_{\mathbf{y}|\mathbf{t},\mathbf{x}',\mathbf{s}'}(\cdot|t,x,s) - W_{\mathbf{y}|\mathbf{x},\mathbf{s}}(\cdot|x,s)\| &\geq \epsilon_b(\epsilon)\end{aligned}$$

This implies that there exists a $(t,x,s) \in \mathcal{T}\times\mathcal{X}\times\mathcal{S}$ such that the distributions $[Q]_{\mathbf{y}|\mathbf{t},\mathbf{x},\mathbf{s}}(\cdot|t,x,s)$, $[Q]_{\mathbf{y}|\mathbf{t},\mathbf{x}',\mathbf{s}'}(\cdot|t,x,s) \in \mathcal{P}(\mathcal{Y})$ are not same. Hence, for some constant $\epsilon_1(\epsilon) > 0$ there exists a tuple (t,x,s), where $\epsilon_1 \to 0$ as $\epsilon \to 0$ such that the following

holds

$$\|[Q]_{\mathbf{y}|\mathbf{t},\mathbf{x},\mathbf{s}}(\cdot|t,x,s) - [Q]_{\mathbf{y}|\mathbf{t},\mathbf{x}',\mathbf{s}'}(\cdot|t,x,s)\| \geq \epsilon_1(\epsilon)$$

This completes the proof of our lemma. ∎

Lemma 34. *Let $\epsilon > 0$ and consider a completely positive $P_{\mathbf{x}}$-self coupling (where $P_{\mathbf{x}} \in \lambda_{\mathbf{x}}$) i.e., there exists a $P_{\mathbf{x},\mathbf{x}'}$ such that $\|P_{\mathbf{x},\mathbf{x}'} - \mathcal{K}_{\mathcal{CP}}(P_{\mathbf{x}})\| = \epsilon > 0$. Then there exists an $\eta(\epsilon)$, where $\eta(\epsilon) \to 0$ as $\epsilon \to 0$, such that for an $(\delta, \underline{\mathbf{t}}, P_{\mathbf{x}|\mathbf{t}})$-approximate constant composition cloud code (as in Lemma 32) $\mathcal{C} = \{\underline{x}_i\}_{i=1}^{2^{nR}}$ with rate R, we have that for every $\underline{\mathbf{y}} \in \mathcal{Y}^n$ either*

$$\max_{t,x,s} D(P_{\underline{\mathbf{y}}|\underline{\mathbf{t}},\underline{\mathbf{x}},\underline{\mathbf{s}}}(\cdot|t,x,s)\|W_{\mathbf{y}|\mathbf{x},\mathbf{s}}(\cdot|x,s)) < \eta \tag{22}$$

or

$$\max_{t,x,s} D(P_{\underline{\mathbf{y}}|\underline{\mathbf{t}},\underline{\mathbf{x}}',\underline{\mathbf{s}}'}(\cdot|t,x,s)\|W_{\mathbf{y}|\mathbf{x},\mathbf{s}}(\cdot|x,s)) < \eta \tag{23}$$

but not both simultaneously for any distinct pair of codewords $\underline{\mathbf{x}}$, $\underline{\mathbf{x}}' \in \mathcal{C}$, i.e., exactly one codeword (i.e., either $\underline{\mathbf{x}} \in \mathcal{C}$ or $\underline{\mathbf{x}}' \in \mathcal{C}$) satisfies the decoding condition in (20)*.*

Proof: Our proof is via contradiction. Let $\underline{\mathbf{x}}, \underline{\mathbf{x}}' \in \mathcal{C}$ be such that $P_{\underline{\mathbf{x}},\underline{\mathbf{x}}'}$ is close to $P_{\mathbf{x},\mathbf{x}'}$, where $\|P_{\mathbf{x},\mathbf{x}'} - \mathcal{K}_{\mathcal{CP}}(P_{\mathbf{x}})\| = \epsilon > 0$, and let the following two equations hold simultaneously for all $\eta > 0$:

$$\max_{t,x,s} D(P_{\underline{\mathbf{y}}|\underline{\mathbf{t}},\underline{\mathbf{x}},\underline{\mathbf{s}}}(\cdot|t,x,s)\|W_{\mathbf{y}|\mathbf{x},\mathbf{s}}(\cdot|x,s)) < \eta$$
$$\max_{t,x,s} D(P_{\underline{\mathbf{y}}|\underline{\mathbf{t}},\underline{\mathbf{x}}',\underline{\mathbf{s}}'}(\cdot|t,x,s)\|W_{\mathbf{y}|\mathbf{x},\mathbf{s}}(\cdot|x,s)) < \eta$$

Then, using Pinsker's inequality (cf. Lemma 2), for a constant $c > 0$ we have

$$\max_{t,x,s} \|P_{\underline{\mathbf{y}}|\underline{\mathbf{t}},\underline{\mathbf{x}},\underline{\mathbf{s}}}(\cdot|t,x,s) - W_{\mathbf{y}|\mathbf{x},\mathbf{s}}(\cdot|x,s)\| \leq c\sqrt{\eta} \tag{24}$$

$$\max_{t,x,s} \|P_{\underline{\mathbf{y}}|\underline{\mathbf{t}},\underline{\mathbf{x}}',\underline{\mathbf{s}}'}(\cdot|t,x,s) - W_{\mathbf{y}|\mathbf{x},\mathbf{s}}(\cdot|x,s)\| \leq c\sqrt{\eta} \tag{25}$$

Using triangle inequality (note that ℓ_1 is a distance measure which satisfies the triangle inequality), it follows from (24), (25) that for every $\eta > 0$

$$\max_{t,x,s} \|P_{\underline{\mathbf{y}}|\underline{\mathbf{t}},\underline{\mathbf{x}},\underline{\mathbf{s}}}(\cdot|t,x,s) - P_{\underline{\mathbf{y}}|\underline{\mathbf{t}},\underline{\mathbf{x}}',\underline{\mathbf{s}}'}(\cdot|t,x,s))\| \leq 2c\sqrt{\eta} \tag{26}$$

However, Lemma 33 guarantees that (as a consequence of $\|P_{\mathbf{x},\mathbf{x}'} - \mathcal{K}(P_{\mathbf{x}})\| = \epsilon > 0$) there exists an $\epsilon_1(\epsilon)$ small enough such that (26) is violated, i.e., $\exists\ (t,x,s) \in \mathcal{T} \times \mathcal{X} \times \mathcal{S}$ such that

$$\|P_{\underline{\mathbf{y}}|\underline{\mathbf{t}},\underline{\mathbf{x}},\underline{\mathbf{s}}}(\cdot|t,x,s) - P_{\underline{\mathbf{y}}|\underline{\mathbf{t}},\underline{\mathbf{x}}',\underline{\mathbf{s}}'}(\cdot|t,x,s)\| \geq \epsilon_1(\epsilon)$$

for the choice of parameter $\eta(\epsilon)$ such that $0 < 2c\sqrt{\eta} \leq \epsilon_1(\epsilon)$ violates the condition (26) in our proof, and thereby establishing the contradiction. Therefore, (22) and (23) cannot hold simultaneously. This completes the proof of our lemma. ∎

We now present the error analysis. We first analyse the probability of error when there are exactly two distinct codewords, say $\underline{x}$ and $\underline{x}'$, in the codebook $\mathcal{C}$. Let $\underline{x}$ be the transmitted codeword; the decoder observes $\underline{\mathbf{y}} = \underline{y}$.

We describe two error events E_1 and E_2 based on the decoding rule in (20), which are analysed the error analysis below:

$$E_1 \triangleq \left(\exists\ (t,x,s) \in (\mathcal{T} \times \mathcal{X} \times \mathcal{S}) : D(P_{\underline{y}|\underline{t},\underline{x},\underline{s}}(\cdot|t,x,s)\|W_{\mathbf{y}|\mathbf{x},\mathbf{s}}(\cdot|x,s)) \geq \eta\right) \tag{27}$$

$$E_2 \triangleq \left(\exists\ \underline{s}' \in \mathcal{P}(\lambda_{\mathbf{s}}) : \max_{t,x,s} D(P_{\underline{y}|\underline{t},\underline{x}',\underline{s}'}(\cdot|t,x,s)\|W_{\mathbf{y}|\mathbf{x},\mathbf{s}}(\cdot|x,s)) < \eta\right) \tag{28}$$

Then, the probability of error for message m (corresponds to transmitted codeword $\underline{x}$) under the omniscient adversary's jamming state $\underline{s}$ is given by

$$\begin{aligned}
\mathbb{P}_e(m,\underline{s}) &= \mathbb{P}_{\mathbf{w}}\left((\phi(\underline{y}) = 0) \cup (\phi(\underline{y}) \neq m)\right) \\
&= \mathbb{P}_{\mathbf{w}}(E_1 \cup E_2) \\
&= \mathbb{P}_{\mathbf{w}}(E_1) + \mathbb{P}_{\mathbf{w}}(E_2 \cap E_1^c) \\
&\overset{(a)}{\leq} \mathbb{P}_{\mathbf{w}}(E_1) + \mathbb{P}_{\mathbf{w}}(E_2 \mid E_1^c) \\
&\overset{(b)}{\leq} \mathbb{P}_{\mathbf{w}}\left(\exists\ (t,x,s) \in (\mathcal{T} \times \mathcal{X} \times \mathcal{S}) : D(P_{\underline{y}|\underline{t},\underline{x},\underline{s}}(\cdot|t,x,s)\|W_{\mathbf{y}|\mathbf{x},\mathbf{s}}(\cdot|x,s)) \geq \eta\right)
\end{aligned}$$

$$+\mathbb{P}_{\mathbf{w}}\left(\exists\ \underline{s}'\in\mathcal{P}(\lambda_{\mathbf{s}}):\max_{t,x,s}D(P_{\underline{y}|\underline{t},\underline{x}',\underline{s}'}(\cdot|t,x,s)\|W_{\mathbf{y}|\mathbf{x},\mathbf{s}}(\cdot|x,s))<\eta\ \middle|\ \max_{t,x,s}D(P_{\underline{y}|\underline{t},\underline{x},\underline{s}}(\cdot|t,x,s)\|W_{\mathbf{y}|\mathbf{x},\mathbf{s}}(\cdot|x,s))<\eta\right)$$

$$\overset{(c)}{\leq}|\mathcal{T}||\mathcal{X}||\mathcal{S}|\sum_{\substack{P_{\underline{\mathbf{y}}|\underline{\mathbf{t}},\underline{\mathbf{x}},\underline{\mathbf{s}}}(\cdot|t,x,s)\in\Delta^{(n)}(\mathcal{Y})\\ \text{s.t. } D(P_{\underline{\mathbf{y}}|\underline{\mathbf{t}},\underline{\mathbf{x}},\underline{\mathbf{s}}}(\cdot|t,x,s)\|W_{\mathbf{y}|\mathbf{x},\mathbf{s}}(\cdot|x,s))\geq\eta}}\mathbb{P}_{\mathbf{w}}\left(\underline{y}\in\mathcal{P}(P_{\underline{y}|\underline{t},\underline{x},\underline{s}}(\cdot|t,x,s))\right)+0$$

$$\overset{(d)}{\leq}|\mathcal{T}||\mathcal{X}||\mathcal{S}|\sum_{\substack{P_{\underline{\mathbf{y}}|\underline{\mathbf{t}},\underline{\mathbf{x}},\underline{\mathbf{s}}}(\cdot|t,x,s)\in\Delta^{(n)}(\mathcal{Y})\\ D(P_{\underline{\mathbf{y}}|\underline{t},\underline{x},\underline{s}}(\cdot|t,x,s)\|W_{\mathbf{y}|\mathbf{x},\mathbf{s}}(\cdot|x,s))\geq\eta}}2^{-nD(P_{\underline{y}|\underline{t},\underline{x},\underline{s}}(\cdot|t,x,s)\|W_{\mathbf{y}|\mathbf{x},\mathbf{s}}(\cdot|x,s))}$$

$$\overset{(e)}{\leq}|\mathcal{T}||\mathcal{X}||\mathcal{S}|\sum_{\substack{P_{\underline{\mathbf{y}}|\underline{\mathbf{t}},\underline{\mathbf{x}},\underline{\mathbf{s}}}(\cdot|t,x,s)\in\Delta^{(n)}(\mathcal{Y})\\ D(P_{\underline{\mathbf{y}}|\underline{\mathbf{t}},\underline{\mathbf{x}},\underline{\mathbf{s}}}(\cdot|t,x,s)\|W_{\mathbf{y}|\mathbf{x},\mathbf{s}}(\cdot|x,s))\geq\eta}}2^{-n\eta}$$

$$\overset{(f)}{\leq}|\mathcal{T}||\mathcal{X}||\mathcal{S}|(n+1)^{|\mathcal{T}||\mathcal{X}||\mathcal{S}||\mathcal{Y}|}2^{-n\eta} \tag{29}$$

where $\eta\to 0$ as $n\to\infty$.

Here,

(a) follows from noting that $\mathbb{P}(A\cap B)\leq\mathbb{P}(A)$ as well as $\mathbb{P}(A\cap B)\leq\mathbb{P}(B|A)$.

(b) follows from (27) and (28).

(c) follows from Lemma 34 which guarantees that two codewords cannot simultaneously satisfy the decoding condition in (20) thereby making the second term equal to 0, and also from the union bound over all tuples $(t,x,s)\in\mathcal{T}\times\mathcal{X}\times\mathcal{S}$.

(d) follows Sanov's theorem (cf. Theorem 4) applied to $\underline{\mathbf{y}}$ generated via the channel conditional distribution $W_{\mathbf{y}|\mathbf{x},\mathbf{s}}$.

(e) follows by noting that $D(P_{\underline{\mathbf{y}}|\underline{\mathbf{t}},\underline{\mathbf{x}},\underline{\mathbf{s}}}(\cdot|t,x,s)\|W_{\mathbf{y}|\mathbf{x},\mathbf{s}}(\cdot|x,s))\geq\eta$.

(f) follows from taking union over all conditional types $P_{\underline{\mathbf{y}}|\underline{\mathbf{t}},\underline{\mathbf{x}},\underline{\mathbf{s}}}\in\Delta(\mathcal{Y}|\mathcal{T}\times\mathcal{X}\times\mathcal{S})$.

Thus, the maximum probability of error for the above two codeword code is vanishing as $n\to\infty$. Now, the maximum probability of error for the code $\mathcal{C}=\{\underline{x}_i\}_{i=1}^{2^{nR}}$ as in Lemma 32 is given by

$$\begin{aligned}\mathbb{P}_e(\mathcal{C}):=\max_m\max_{\underline{s}:\in P(\lambda_{\mathbf{s}})}\mathbb{P}_W\left(\exists\ \underline{x},\underline{x}'\in\mathcal{C}:\mathbb{P}_{\mathbf{w}}(E_1\cup E_2)\right))\\ \overset{(a)}{\leq}\binom{2^{nR}}{2}|\mathcal{T}||\mathcal{X}||\mathcal{S}|(n+1)^{|\mathcal{T}||\mathcal{X}||\mathcal{S}||\mathcal{Y}|}2^{-n\eta}\\ \leq|\mathcal{T}||\mathcal{X}||\mathcal{S}|(n+1)^{|\mathcal{T}||\mathcal{X}||\mathcal{S}||\mathcal{Y}|}2^{n(2R-\eta)}\\ \overset{(b)}{=}|\mathcal{T}||\mathcal{X}||\mathcal{S}|(n+1)^{|\mathcal{T}||\mathcal{X}||\mathcal{S}||\mathcal{Y}|}2^{-n\eta'}\end{aligned} \tag{30}$$

(a) follows from (29) and taking union over all pairs of codewords $(\underline{x},\underline{x}')\in\mathcal{C}$.

(b) follows from noting that η is chosen such that $\eta<2R$, where R is the rate of the approximate constant composition code as in Lemma 32. (here, $\eta'>0$).

Thus, the error probability $\mathbb{P}_e(\mathcal{C})$

$$\mathbb{P}_e(\mathcal{C}):=\max_{m\in\mathcal{M}}\max_{\underline{s}:\in\mathcal{P}(\lambda_{\mathbf{s}})}\mathbb{P}_e(m,\underline{s})\leq\eta''$$

is negligible in the block length n. This completes our error analysis.

VIII. Large Rate Achievability for Non State-Deterministic and Omniscient AVCs

In this section, we present a more sophisticated cloud code construction (with time-sharing) with some special properties and a two-step decoding at the receiver, inspired by the construction in [CK81]. This cloud code achieves a larger rate (R) as compared to the rate achieved by the cloud code construction in section VII.

We begin by stating a combinatorial lemma (cf. Lemma 1 [CK81]) which provides an upper bound on the number of conditionally typical sequences.

Lemma 35. *Consider three arbitrary and finite sets $\mathcal{T}$, $\mathcal{U}$, $\mathcal{X}$. Let $\underline{\mathbf{t}}\in\mathcal{T}^n$ be a sequence with type $P_{\underline{\mathbf{t}}}$. Then, for every $R>0$, $n\geq\max(|\mathcal{T}|,|\mathcal{U}|,|\mathcal{X}|)$ and $P_{\underline{\mathbf{x}}|\underline{\mathbf{t}}}\in\Delta^{(n)}(\mathcal{X}|\mathcal{T})$, there exist $\lfloor 2^{nR}\rfloor$ sequences $\underline{\mathbf{x}}_i:i\in[2^{nR}]$ and $\underline{\mathbf{x}}_i\in\mathcal{P}(P_{\underline{\mathbf{x}}|\underline{\mathbf{t}}})$ such that $\forall$ $\underline{\mathbf{u}}\in\mathcal{U}^n$ and $\forall$ $P_{\underline{\mathbf{x}}|\underline{\mathbf{u}},\underline{\mathbf{t}}}\in\Delta^{(n)}(\mathcal{X}|\mathcal{U}\times\mathcal{T})$, the following holds*

$$\left|\{i:\underline{\mathbf{x}}_i\in\mathcal{P}(P_{\underline{\mathbf{x}}|\underline{\mathbf{u}},\underline{\mathbf{t}}})\}\right|\leq 3(n+1)^{|\mathcal{T}||\mathcal{X}|}2^{\{n|R-I(\mathbf{u};\mathbf{x}|\mathbf{t})|^+\}} \tag{31}$$

with very high probability.

Proof: Let us fix a $\underline{\mathbf{u}} \in \mathcal{P}(P_{\underline{\mathbf{u}}|\underline{\mathbf{t}}})$ and the conditional type $P_{\underline{\mathbf{x}}|\underline{\mathbf{u}},\underline{\mathbf{t}}} \in \Delta^{(n)}(\mathcal{X}|\mathcal{U}\times\mathcal{T})$. We need to show that on choosing

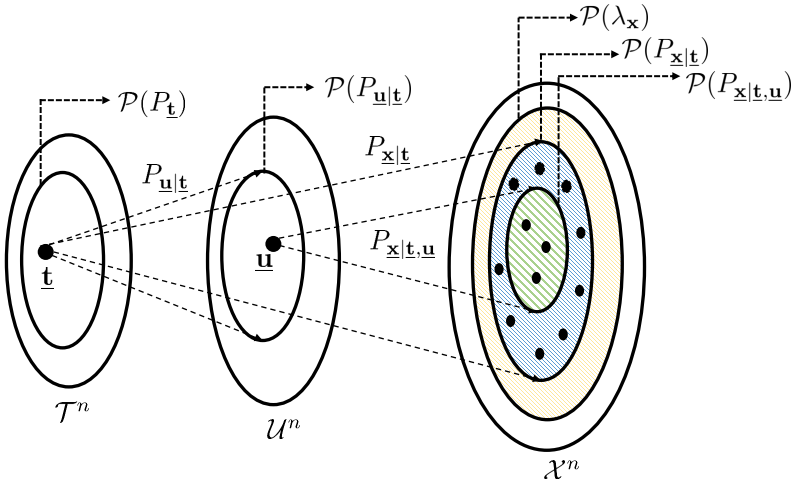


Fig. 5. Figure for Lemma 35

2^{nR} sequences from $\mathcal{P}(P_{\underline{\mathbf{x}}|\underline{\mathbf{t}}})$ uniformly at random, the number of sequences that will be from $\mathcal{P}(P_{\underline{\mathbf{x}}|\underline{\mathbf{u}},\underline{\mathbf{t}}})$ is upper bounded by $3(n+1)^{|\mathcal{T}||\mathcal{X}|}2^{\{n|R-I(\mathbf{u};\mathbf{x}|\mathbf{t})|^+\}}$.

We denote by A_i, the event that a $\underline{\mathbf{x}}_i$ such that $i \in [2^{nR}]$ chosen uniformly at random from $\mathcal{P}(P_{\underline{\mathbf{x}}|\underline{\mathbf{t}}})$ lies in $\mathcal{P}(P_{\underline{\mathbf{x}}|\underline{\mathbf{u}},\underline{\mathbf{t}}})$. Further, let us define an indicator random variable $\mathbb{I}_{A_i}$ on event A_i as follows:

$$\mathbb{I}_{A_i} = \begin{cases} 1 & \text{if } A_i \text{ occurs} \\ 0 & \text{if otherwise} \end{cases}$$

for a given $i \in [2^{nR}]$,

$$\begin{aligned} \mathbb{P}(\mathbb{I}_{A_i} = 1) &:= \frac{\text{vol}(\mathcal{P}(P_{\underline{\mathbf{x}}|\underline{\mathbf{u}},\underline{\mathbf{t}}}))}{\text{vol}(\mathcal{P}(P_{\underline{\mathbf{x}}|\underline{\mathbf{t}}}))} \\ &\leq \frac{2^{nH(\mathbf{x}|\mathbf{u},\mathbf{t})}}{(n+1)^{-|\mathcal{T}||\mathcal{X}|}2^{nH(\mathbf{x}|\mathbf{t})}} \\ &\leq (n+1)^{|\mathcal{T}||\mathcal{X}|}2^{-nI(\mathbf{u};\mathbf{x}|\mathbf{t})} \end{aligned} \tag{32}$$

Therefore, if $\xi := \sum_{i=1}^{2^{nR}} \mathbb{I}_{A_i}$

$$\begin{aligned} \mathbb{E}[\xi] &\leq 2^{nR}(n+1)^{|\mathcal{T}||\mathcal{X}|}2^{-nI(\mathbf{u};\mathbf{x}|\mathbf{t})} \\ &\leq (n+1)^{|\mathcal{T}||\mathcal{X}|}2^{n(R-I(\mathbf{u};\mathbf{x}|\mathbf{t}))} \end{aligned} \tag{33}$$

Now, we use the chernoff bound (cf. Lemma 3) to show that the probability of the event that the $\xi > 3(n+1)^{|\mathcal{T}||\mathcal{X}|}2^{\{n|R-I(\mathbf{u};\mathbf{x}|\mathbf{t})|^+\}}$ is negligible in n.

$$\begin{aligned} \mathbb{P}(\xi > (1+2)(n+1)^{|\mathcal{T}||\mathcal{X}|}2^{n|R-I(\mathbf{u};\mathbf{x}|\mathbf{t})|^+}) &< e^{-(n+1)^{|\mathcal{T}||\mathcal{X}|}2^{n|R-I(\mathbf{u};\mathbf{x}|\mathbf{t})|^+}} \\ &< e^{-(n+1)^{|\mathcal{T}||\mathcal{X}|}2^{n\delta}} \end{aligned}$$

where $\delta > 0$.
On taking union over all possible pairs of ($\underline{\mathbf{u}}$, $P_{\underline{\mathbf{x}}|\underline{\mathbf{u}},\underline{\mathbf{t}}}$), we have:

$$\begin{aligned} \mathbb{P}(\xi > 3(n+1)^{|\mathcal{T}||\mathcal{X}|}2^{n|R-I(\mathbf{u};\mathbf{x}|\mathbf{t})|^+}) &< |\mathcal{U}|^n(n+1)^{|\mathcal{T}||\mathcal{U}||\mathcal{X}|}e^{-(n+1)^{|\mathcal{X}||\mathcal{T}|}2^{n\delta}} \\ &< \epsilon_n \end{aligned}$$

where $\epsilon_n > 0$ is negligible in n i.e., $\epsilon_n \to 0$ as $n \to \infty$.
Therefore, this completes the proof of our lemma. ∎

A. Codebook Design

Lemma 36 (Code-book Construction)**.** *Let $\mathcal{T}$, $\mathcal{X}$ and $\mathcal{S}$ be finite sets. Then for every $\epsilon > 0$, $R > 0$, $n \geq n_0(|\mathcal{T}|, |\mathcal{X}|, |\mathcal{S}|, \epsilon)$, $\underline{\mathbf{t}} \in \mathcal{T}^n$ with type (close to) $P_{\mathbf{t}} \in \Delta(\mathcal{T})$ and conditional type (close to) $P_{\mathbf{x}|\mathbf{t}} \in \Delta(\mathcal{X}|\mathcal{T})$ (where $[P_{\mathbf{t}}P_{\mathbf{x}|\mathbf{t}}]_{\mathbf{x}} \in \lambda_{\mathbf{x}}$) with*

$R \leq H(\mathbf{x}|\mathbf{t})$, there exists a $(\delta, \underline{\mathbf{t}}, P_{\mathbf{x}|\mathbf{t}})$-approximate constant composition cloud code $\mathcal{C} := \{\underline{x}_i\}_{i=1}^{2^{nR}} \subseteq \mathcal{P}(P_{\mathbf{x}|\mathbf{t}}) \subset \mathcal{X}^n$ such that $\forall \underline{\mathbf{x}}_i \in \mathcal{C}, P_{\underline{\mathbf{t}},\underline{\mathbf{x}}} \approx P_{\mathbf{t}} P_{\mathbf{x}|\mathbf{t}}$, which satisfies:

(i) $\frac{1}{n}\log|\mathcal{C}| \geq R - \epsilon$

(ii) $\forall\ \underline{\mathbf{x}}, \underline{\mathbf{x}}' \in \mathcal{X}^n$, $\underline{\mathbf{s}} \in \mathcal{S}^n$ *and* $P_{\underline{\mathbf{x}}'|\underline{\mathbf{t}},\underline{\mathbf{x}},\underline{\mathbf{s}}} \in \Delta^{(n)}(\mathcal{X}|\mathcal{T}\times\mathcal{X}\times\mathcal{S})$, *we have:*

$$\left|\mathcal{P}(P_{\underline{\mathbf{x}}'|\underline{\mathbf{t}},\underline{\mathbf{x}},\underline{\mathbf{s}}}) \cap \mathcal{C}\right| \leq 3(n+1)^{|\mathcal{T}||\mathcal{X}|} 2^{\{n|R - I(\mathbf{x},\mathbf{s};\mathbf{x}'|\mathbf{t})|\}} \tag{34}$$

(iii) for every pair of codewords $\underline{\mathbf{x}}$, $\underline{\mathbf{x}}' \in \mathcal{C}$ such that $\underline{\mathbf{x}} \neq \underline{\mathbf{x}}'$

$$I(\underline{\mathbf{x}}; \underline{\mathbf{x}}'|\underline{\mathbf{t}}) < R \tag{35}$$

Proof: The second property of the code can be proved using the Lemma (35) by making the following correspondence $\mathcal{U} \leftrightarrow \mathcal{X} \times \mathcal{S}$. Therefore, we have

$$\left|\mathcal{P}(P_{\underline{\mathbf{x}}'|\underline{\mathbf{t}},\underline{\mathbf{x}},\underline{\mathbf{s}}}) \cap \mathcal{C}\right| = \left|\left\{\underline{\mathbf{x}}_i \in \mathcal{C} : \underline{\mathbf{x}}_i \in \mathcal{P}(P_{\underline{\mathbf{x}}'|\underline{\mathbf{t}},\underline{\mathbf{x}},\underline{\mathbf{s}}})\right\}\right| \leq 3(n+1)^{|\mathcal{T}||\mathcal{X}|} 2^{n|R - I(\mathbf{x},\mathbf{s};\mathbf{x}'|\mathbf{t})|^+} \tag{36}$$

Similarly, for the correspondence $\mathcal{U} \leftrightarrow \mathcal{X}$, we have the following

$$\left|\left\{\underline{\mathbf{x}}_i \in \mathcal{C} : \underline{\mathbf{x}}_i \in \mathcal{P}(P_{\underline{\mathbf{x}}'|\underline{\mathbf{t}},\underline{\mathbf{x}}})\right\}\right| \leq 3(n+1)^{|\mathcal{T}||\mathcal{X}|} 2^{n|R - I(\mathbf{x};\mathbf{x}'|\mathbf{t})|^+} \tag{37}$$

To prove the third property, we use expurgation to show that there exists a sub codebook $\mathcal{C}' \subseteq \mathcal{C}$ with almost same number of codewords as in $\mathcal{C}$ and satisfying $I(\underline{\mathbf{x}}; \underline{\mathbf{x}}'|\underline{\mathbf{t}}) < R$ for $\underline{\mathbf{x}} \neq \underline{\mathbf{x}}' \in \mathcal{C}$.

Note that the assumption $R \leq H(\mathbf{x}|\mathbf{t})$ implies that all the 2^{nR} sequences are distinct in $\mathcal{C}$. For the given conditional type $P_{\underline{\mathbf{x}}'|\underline{\mathbf{t}},\underline{\mathbf{x}}} \in \Delta^{(n)}(\mathcal{X}|\mathcal{T}\times\mathcal{X})$, and from (37) we have

$$\left|\left\{\underline{\mathbf{x}}_i \in \mathcal{C} : \underline{\mathbf{x}}_i \in \mathcal{P}(P_{\underline{\mathbf{x}}'|\underline{\mathbf{t}},\underline{\mathbf{x}}}) \text{ and } I(\underline{\mathbf{x}}_i; \underline{\mathbf{x}}|\underline{\mathbf{t}}) \geq R \text{ holds } \forall\ \underline{\mathbf{x}} \in \mathcal{C}\right\}\right| \leq 3(n+1)^{|\mathcal{X}||\mathcal{T}|}$$

On taking union over all such conditional types $P_{\underline{\mathbf{x}}'|\underline{\mathbf{t}},\underline{\mathbf{x}}} \in \Delta^{(n)}(\mathcal{X}|\mathcal{T}\times\mathcal{X})$

$$\left|\left\{\underline{\mathbf{x}}_i \in \mathcal{C} : I(\underline{\mathbf{x}}_i; \underline{\mathbf{x}}|\underline{\mathbf{t}}) \geq R\right\}\right| \leq 3(n+1)^{|\mathcal{X}||\mathcal{T}|}(n+1)^{|\mathcal{X}|^2|\mathcal{T}|}$$

$$\#\text{ bad codewords (i.e., } I(\underline{\mathbf{x}}'; \underline{\mathbf{x}}|\underline{\mathbf{t}}) > R) \leq 3(n+1)^{|\mathcal{X}||\mathcal{T}|}(n+1)^{|\mathcal{X}|^2|\mathcal{T}|}$$

Further, these bad codewords can be expurgated from the codebook $\mathcal{C}$ and we obtain a sub codebook $\mathcal{C}'$ consisting of codewords satisfying property (iii) in Lemma 36.

$$\begin{aligned} &|\mathcal{C}'|3(n+1)^{|\mathcal{X}||\mathcal{T}|}(n+1)^{|\mathcal{X}|^2|\mathcal{T}|} \geq |\mathcal{C}| \\ &\implies |\mathcal{C}'| \geq \frac{|\mathcal{C}|}{3(n+1)^{|\mathcal{X}||\mathcal{T}|}(n+1)^{|\mathcal{X}|^2|\mathcal{T}|}} \\ &\implies R_{\mathcal{C}'} \geq R_{\mathcal{C}} - \epsilon(n) \end{aligned}$$

where $\epsilon(n)$ is negligible in n. ■

B. Encoding/Decoding rules

Encoder (f): The encoder mapping is defined via deterministic function $f : \mathcal{M} \to \mathcal{C}$. Given the message $m \in \mathcal{M}$, the codeword $\underline{\mathbf{x}} = \underline{x} \in \mathcal{C}$ is transmitted over the Non state-deterministic and omniscient AVC.

Decoder (ϕ): The decoder mapping is defined via function $\phi : \mathcal{Y}^n \to \mathcal{M} \cup \{0\}$. Given the codebook $\mathcal{C}$, the time-sharing sequence $\underline{t} \in \mathcal{T}^n$ and the received vector $\underline{y} \in \mathcal{Y}^n$, the decoder outputs the codeword $\underline{\mathbf{x}} \in \mathcal{C}$ (corresponding to the message m) such that for some $\underline{\mathbf{s}} \in \mathcal{P}(\lambda_{\mathbf{s}})$ it satisfies the following conditions:

1. (Joint-typicality decoding) For some positive constant η, we have $\underline{x} \in \mathcal{L}(\underline{\mathbf{y}})$ where

$$\mathcal{L}(\underline{\mathbf{y}}) := \left\{\underline{\mathbf{x}} \in \mathcal{C} : \exists\ \underline{\mathbf{s}} \in \mathcal{P}(\lambda_{\mathbf{s}}) \text{ such that } D(P_{\underline{\mathbf{t}},\underline{\mathbf{x}},\underline{\mathbf{s}},\underline{\mathbf{y}}} \| P_{\underline{\mathbf{t}}} P_{\underline{\mathbf{x}}|\underline{\mathbf{t}}} P_{\underline{\mathbf{s}}|\underline{\mathbf{t}},\underline{\mathbf{x}}} W_{\mathbf{y}|\mathbf{x},\mathbf{s}}) < \eta\right\} \tag{38}$$

2. (Tournament decoding) $\forall\ \hat{\underline{\mathbf{x}}} \in \mathcal{L}(\underline{\mathbf{y}})$ we have

$$I(\hat{\underline{\mathbf{x}}}; \underline{\mathbf{y}}|\underline{\mathbf{t}}, \underline{\mathbf{x}}, \underline{\mathbf{s}}) < \eta \tag{39}$$

for some constant $\eta > 0$ chosen appropriately later.

The first step in the decoding rule is a "list-decoding" like step which outputs a list $\mathcal{L}(\underline{\mathbf{y}})$ of at most $\mathcal{O}(\text{poly}(n))$ codewords. The second step can be viewed as a "tournament" between all the pair of codewords within the list $\mathcal{L}(\underline{\mathbf{y}})$. If there is a unique codeword that is the champion of the tournament, the decoder outputs that codeword (message). Else, the decoder outputs 0 to declare an error.

C. Unambiguity of the decoding rules

Lemma 37. *For every $\delta > 0$ there exists an $\eta > 0$ such that for any codebook $\mathcal{C}$ in Lemma (36) with rate $R \leq GV(P_{\mathbf{t}}P_{\mathbf{x}|\mathbf{t}}) - \delta$ the two-step decoding rule defined via* (38) *and* (39) *outputs a unique $\underline{\mathbf{x}} \in \mathcal{C}$ for every received vector $\underline{y} \in \mathcal{Y}^n$.*

Proof: We use the contradiction approach to proof the following lemma. Consider two tuples $(\underline{\mathbf{t}}, \underline{\mathbf{x}}, \underline{\mathbf{s}})$ and $(\underline{\mathbf{t}}, \underline{\hat{\mathbf{x}}}, \underline{\hat{\mathbf{s}}})$ in the space $\mathcal{T}^n \times \mathcal{C} \times \mathcal{S}^n$ such that $P_{\underline{\mathbf{s}}}$, $P_{\underline{\hat{\mathbf{s}}}} \in \lambda_{\mathbf{s}}$ and both satisfy the two-step decoding rule. We use the contradiction approach to show that indeed, it can only happen when $\underline{\mathbf{x}} = \underline{\hat{\mathbf{x}}}$.
Thus, for the tuple $(\underline{\mathbf{t}}, \underline{\mathbf{x}}, \underline{\mathbf{s}})$

$$D(P_{\underline{\mathbf{t}},\underline{\mathbf{x}},\underline{\mathbf{s}},\underline{\mathbf{y}}} || P_{\underline{\mathbf{t}}} P_{\underline{\mathbf{x}}|\underline{\mathbf{t}}} P_{\underline{\mathbf{s}}|\underline{\mathbf{t}},\underline{\mathbf{x}}} W_{\mathbf{y}|\mathbf{x},\mathbf{s}}) < \eta \tag{40}$$

$$I(\underline{\hat{\mathbf{x}}}; \underline{\mathbf{y}} | \underline{\mathbf{t}}, \underline{\mathbf{x}}, \underline{\mathbf{s}}) < \eta \tag{41}$$

Similarly, for the tuple $(\underline{\mathbf{t}}, \underline{\hat{\mathbf{x}}}, \underline{\hat{\mathbf{s}}})$

$$D(P_{\underline{\mathbf{t}},\underline{\hat{\mathbf{x}}},\underline{\hat{\mathbf{s}}},\underline{\mathbf{y}}} || P_{\underline{\mathbf{t}}} P_{\underline{\hat{\mathbf{x}}}|\underline{\mathbf{t}}} P_{\underline{\hat{\mathbf{s}}}|\underline{\mathbf{t}},\underline{\hat{\mathbf{x}}}} W_{\mathbf{y}|\mathbf{x},\mathbf{s}}) < \eta \tag{42}$$

$$I(\underline{\mathbf{x}}; \underline{\mathbf{y}} | \underline{\mathbf{t}}, \underline{\hat{\mathbf{x}}}, \underline{\hat{\mathbf{s}}}) < \eta \tag{43}$$

We slightly abuse the notation to represent the vectors $(\underline{\mathbf{t}}, \underline{\mathbf{x}}, \underline{\hat{\mathbf{x}}}, \underline{\mathbf{s}}, \underline{\hat{\mathbf{s}}}, \underline{\mathbf{y}})$ using the random variables $(\mathbf{t}, \mathbf{x}, \hat{\mathbf{x}}, \mathbf{s}, \hat{\mathbf{s}}, \mathbf{y})$ such that the joint type of the vectors is close to the joint distribution of the random variables.
From (40) and (41), we have

$$\sum_{t,x,s,y} P_{\mathbf{t},\mathbf{x},\mathbf{s},\mathbf{y}}(t,x,s,y) \log \frac{P_{\mathbf{y}|\mathbf{t},\mathbf{x},\mathbf{s}}(y|t,x,s)}{W_{\mathbf{y}|\mathbf{x},\mathbf{s}}(y|x,s)} < \eta$$

$$\sum_{t,x,s,\hat{x},y} P_{\mathbf{t},\mathbf{x},\mathbf{s},\hat{\mathbf{x}},\mathbf{y}}(t,x,s,\hat{x},y) \log \frac{P_{\hat{\mathbf{x}},\mathbf{y}|\mathbf{t},\mathbf{x},\mathbf{s}}(\hat{x},y|t,x,s)}{P_{\hat{\mathbf{x}}|\mathbf{t},\mathbf{x},\mathbf{s}}(\hat{x}|t,x,s) P_{\mathbf{y}|\mathbf{t},\mathbf{x},\mathbf{s}}(y|t,x,s)} < \eta$$

On adding the above two inequalities

$$\sum_{t,x,s,\hat{x},y} P_{\mathbf{t},\mathbf{x},\mathbf{s},\hat{\mathbf{x}},\mathbf{y}}(t,x,s,\hat{x},y) \log \frac{P_{\mathbf{t},\mathbf{x},\mathbf{s},\hat{\mathbf{x}},\mathbf{y}}(t,x,s,\hat{x},y)}{P_{\mathbf{t},\mathbf{x},\mathbf{s},\hat{\mathbf{x}}}(t,x,s,\hat{x}) W_{\mathbf{y}|\mathbf{x},\mathbf{s}}(y|x,s)} < 2\eta$$

Since, the KL divergence does not increase on projecting from higher dimensional space to lower. Therefore, on projecting the simplex space from $\mathcal{T} \times \mathcal{X} \times \mathcal{S} \times \mathcal{X} \times \mathcal{Y}$ to $\mathcal{T} \times \mathcal{X} \times \mathcal{X} \times \mathcal{Y}$, we have

$$\sum_{t,x,\hat{x},y} P_{\mathbf{t},\mathbf{x},\hat{\mathbf{x}},\mathbf{y}}(t,x,\hat{x},y) \log \frac{P_{\mathbf{t},\mathbf{x},\hat{\mathbf{x}},\mathbf{y}}(t,x,\hat{x},y)}{\sum_{\mathbf{s}} P_{\mathbf{t},\mathbf{x},\mathbf{s},\hat{\mathbf{x}}}(t,x,s,\hat{x}) W_{\mathbf{y}|\mathbf{x},\mathbf{s}}(y|x,s)} < 2\eta$$

$$\sum_{t,x,\hat{x},y} P_{\mathbf{t},\mathbf{x},\hat{\mathbf{x}},\mathbf{y}}(t,x,\hat{x},y) \log \frac{P_{\mathbf{t},\mathbf{x},\hat{\mathbf{x}},\mathbf{y}}(t,x,\hat{x},y)}{P_{\mathbf{t},\mathbf{x},\hat{\mathbf{x}}}(t,x,\hat{x}) \sum_{\mathbf{s}} Q^1_{\mathbf{s}|\mathbf{t},\mathbf{x},\hat{\mathbf{x}}}(s|t,x,\hat{x}) W_{\mathbf{y}|\mathbf{x},\mathbf{s}}(y|x,s)} < 2\eta$$

$$\sum_{t,x,\hat{x},y} P_{\mathbf{t},\mathbf{x},\hat{\mathbf{x}},\mathbf{y}}(t,x,\hat{x},y) \log \frac{P_{\mathbf{t},\mathbf{x},\hat{\mathbf{x}},\mathbf{y}}(t,x,\hat{x},y)}{P_{\mathbf{t},\mathbf{x},\hat{\mathbf{x}}}(t,x,\hat{x}) V^1_{\mathbf{y}|\mathbf{t},\mathbf{x},\hat{\mathbf{x}}}(y|t,x,\hat{x})} < 2\eta \tag{44}$$

where,

$$V^1_{\mathbf{y}|\mathbf{t}|\mathbf{x},\hat{\mathbf{x}}}(y|t,x,\hat{x}) := \sum_{s \in \mathcal{S}} Q^1_{\mathbf{s}|\mathbf{t},\mathbf{x},\hat{\mathbf{x}}}(s|t,x,\hat{x}) W_{\mathbf{y}|\mathbf{x},\mathbf{s}}(y|x,s)$$

The equation (44) can be re-written in terms of KL divergence as follows

$$D(P_{\mathbf{t},\mathbf{x},\hat{\mathbf{x}},\mathbf{y}} || P_{\mathbf{t},\mathbf{x},\hat{\mathbf{x}}} V^1_{\mathbf{y}|\mathbf{t},\mathbf{x},\hat{\mathbf{x}}}) < 2\eta \tag{45}$$

using Pinsker's inequality (cf. Lemma 2),

$$\sum_{t,x,\hat{x},y} \left| P_{\mathbf{t},\mathbf{x},\hat{\mathbf{x}},\mathbf{y}}(t,x,\hat{x},y) - P_{\mathbf{t},\mathbf{x},\hat{\mathbf{x}}}(t,x,\hat{x}) V^1_{\mathbf{y}|\mathbf{t}|\mathbf{x},\hat{\mathbf{x}}}(y|t,x,\hat{x}) \right| \leq c\sqrt{2\eta} \tag{46}$$

Similarly, on solving equations (42) and (43), we can obtain the following KL divergence

$$D(P_{\mathbf{t},\mathbf{x},\hat{\mathbf{x}},\mathbf{y}}||P_{\mathbf{t},\mathbf{x},\hat{\mathbf{x}}}V^2_{\mathbf{y}|\mathbf{t},\hat{\mathbf{x}},\mathbf{x}}) < 2\eta \tag{47}$$

where,

$$V^2_{\mathbf{y}|\mathbf{t},\mathbf{x},\hat{\mathbf{x}}}(y|t,x,\hat{x}) := \sum_{\hat{s}\in\mathcal{S}} Q^2_{\hat{\mathbf{s}}|\mathbf{t},\hat{\mathbf{x}},\mathbf{x}}(\hat{s}|t,\hat{x},x)W_{\mathbf{y}|\hat{\mathbf{x}},\hat{\mathbf{s}}}(y|\hat{x},\hat{s})$$

applying Pinsker's inequality to (47), we get

$$\sum_{t,x,\hat{x},y}\left|P_{\mathbf{t},\mathbf{x},\hat{\mathbf{x}},\mathbf{y}}(t,x,\hat{x},y) - P_{\mathbf{t},\mathbf{x},\hat{\mathbf{x}}}(t,x,\hat{x})V^2_{\mathbf{y}|\mathbf{t}|\mathbf{x},\hat{\mathbf{x}}}(y|t,x,\hat{x})\right| \le c\sqrt{2\eta} \tag{48}$$

From (46) and (48), and further using the triangle inequality

$$\sum_{t,x,\hat{x}} P_{\mathbf{t},\mathbf{x},\hat{\mathbf{x}}}(t,x,\hat{x})\sum_{\mathbf{y}}\left|V^1_{\mathbf{y}|\mathbf{t}|\mathbf{x},\hat{\mathbf{x}}}(y|t,x,\hat{x}) - V^2_{\mathbf{y}|\mathbf{t}|\mathbf{x},\hat{\mathbf{x}}}(y|t,x,\hat{x})\right| \le 2c\sqrt{2\eta} \tag{49}$$

Note that from property (iii) of the code $\mathcal{C}$ in Lemma 36 and noting that $R < \text{GV}(P_{\mathbf{t}}P_{\mathbf{x}|\mathbf{t}}) - \delta$. we have

$$I(\mathbf{x};\hat{\mathbf{x}}|\mathbf{t}) \le R < \text{GV}(P_{\mathbf{t}}P_{\mathbf{x}|\mathbf{t}}) - \delta = \min_{\substack{P_{\mathbf{t},\mathbf{x},\hat{\mathbf{x}}} \\ \text{s.t } [P_{\mathbf{t},\mathbf{x},\hat{\mathbf{x}}}]_{\mathbf{x},\hat{\mathbf{x}}}\in\mathcal{K}_{\mathcal{CP}}(P_{\mathbf{x}})}} I(\mathbf{x};\hat{\mathbf{x}}|\mathbf{t}) - \delta$$

The above equation implies that for there does not exists a joint type $P_{\mathbf{t},\mathbf{x},\hat{\mathbf{x}}}$ (satisfying $[P_{\mathbf{t},\mathbf{x},\hat{\mathbf{x}}}]_{\mathbf{x}} = [P_{\mathbf{t},\mathbf{x},\hat{\mathbf{x}}}]_{\hat{\mathbf{x}}} \in \lambda_{\mathbf{x}}$), such that the following condition holds: $[P_{\mathbf{t},\mathbf{x},\hat{\mathbf{x}}}]_{\mathbf{x},\hat{\mathbf{x}}} \notin \mathcal{K}_{\mathcal{CP}}(P_{\mathbf{x}})$. Therefore

$$\min_{Q^1_{\mathbf{s}|\mathbf{t},\mathbf{x},\hat{\mathbf{x}}},Q^2_{\mathbf{s}|\mathbf{t},\mathbf{x},\hat{\mathbf{x}}}\in\Delta(\mathcal{S}|\mathcal{T}\times\mathcal{X}\times\mathcal{X})}\sum_{y\in\mathcal{Y}}\left|\sum_{s\in\mathcal{S}}Q^1_{\mathbf{s}|\mathbf{t},\mathbf{x},\hat{\mathbf{x}}}(s|t,x,\hat{x})W_{\mathbf{y}|\mathbf{x},\mathbf{s}}(y|x,s) - \sum_{\hat{s}\in\mathcal{S}}Q^2_{\hat{\mathbf{s}}|\mathbf{t},\hat{\mathbf{x}},\mathbf{x}}(\hat{s}|t,\hat{x},x)W_{\mathbf{y}|\mathbf{x},\mathbf{s}}(y|\hat{x},\hat{s})\right| > 0 \tag{50}$$

Let Q^{1*} and Q^{2*} be the minimizing distributions for $Q^1_{\mathbf{s}|\mathbf{t},\mathbf{x},\hat{\mathbf{x}}}$ and $Q^2_{\mathbf{s}|\mathbf{t},\mathbf{x},\hat{\mathbf{x}}}$ respectively and $\epsilon > 0$ be a positive constant.

$$\sum_{y\in\mathcal{Y}}\left|\sum_{s\in\mathcal{S}}Q^{1*}_{\mathbf{s}|\mathbf{t},\mathbf{x},\hat{\mathbf{x}}}(s|t,x,\hat{x})W_{\mathbf{y}|\mathbf{x},\mathbf{s}}(y|x,s) - \sum_{\hat{s}\in\mathcal{S}}Q^{2*}_{\hat{\mathbf{s}}|\mathbf{t},\hat{\mathbf{x}},\mathbf{x}}(\hat{s}|t,\hat{x},x)W_{\mathbf{y}|\hat{\mathbf{x}},\hat{\mathbf{s}}}(y|\hat{x},\hat{s})\right| \ge \epsilon$$
$$\sum_{y\in\mathcal{Y}}\left|V^1_{\mathbf{y}|\mathbf{t}|\mathbf{x},\hat{\mathbf{x}}}(y|t,x,\hat{x}) - V^2_{\mathbf{y}|\mathbf{t},\hat{\mathbf{x}},\mathbf{x}}(y|t,\hat{x},x)\right| \ge \epsilon \tag{51}$$

For two different codewords $\underline{\mathbf{x}} \neq \hat{\underline{\mathbf{x}}}$ to pass the both the conditions of the two-step decoding rule, equation (49) must hold. However, noting from (49) and (51), for η chosen sufficiently small such that $\eta < \frac{\epsilon^2}{8c^2}$, then (49) does not holds. This gives us the required contradiction for a proper choice of the parameter η.
Therefore, it proves the unique decodability of the 2-step decoder defined above.

D. Error Analysis

In this section, we show that a cloud code construction as in Lemma 36, with encoder/decoder as defined in (38) and (39) achieves a rate

$$R \le \min(\text{DMC}(P_{\mathbf{t}}P_{\mathbf{x}|\mathbf{t}}), \text{GV}(P_{\mathbf{t}}P_{\mathbf{x}|\mathbf{t}})) - \delta \tag{52}$$

for some $P_{\mathbf{t}}P_{\mathbf{x}|\mathbf{t}} \in \Delta(\mathcal{X}\times\mathcal{T})$ and a fixed $\delta > 0$ such that the maximum error probability $(\mathbb{P}_e)$ is small for sufficiently large block length n.
Note that replacing R by $R - \delta/2$ and δ by $\delta/2$ above, we have the achievable rate expression as in Theorem 24.
Now, we state a lemma which will be required later in our proof:

Lemma 38. *Given a constant $\delta > 0$ as above. Given $\mathbf{t}$ (resp. $\underline{\mathbf{t}}$) $\in \mathcal{T}$ (resp. $\mathcal{T}^n$) for every triple of random variables $\mathbf{x}', \mathbf{s}', \mathbf{y}'$ (mapped to sequences $\underline{\mathbf{x}}' \in \mathcal{C}$, $\underline{\mathbf{s}}' \in \mathcal{P}(\lambda_{\mathbf{s}})$ and $\underline{\mathbf{y}} \in \mathcal{Y}^n$) satisfying $D(P_{\mathbf{t},\mathbf{x}',\mathbf{s}',\mathbf{y}}||P_{\mathbf{t}}P_{\mathbf{x}'|\mathbf{t}}P_{\mathbf{s}'|\mathbf{x}',\mathbf{t}}W_{\mathbf{y}'|\mathbf{x}',\mathbf{s}'}) \le \eta$. Let the following hold (via assumption)*

$$I(\mathbf{x}';\mathbf{y}'|\mathbf{t}) - I(\mathbf{x}';\mathbf{y}|\mathbf{t}) \le \delta/2 \tag{53}$$

Then, an appropriate choice of the parameter $\eta(\delta)$ always exists such that (53) *holds.*

Proof: To prove the following lemma, we make the assumption that the distributions $P_{\mathbf{t}}$, $P_{\mathbf{x}}, P_{\mathbf{y}}$ and $P_{\mathbf{y}'}$ do not contain any zero element i.e., $P_{\mathbf{t}}(t) > 0\ \forall\ t \in \mathcal{T}$, $P_{\mathbf{x}}(x) > 0\ \forall\ x \in \mathcal{X}$, $P_{\mathbf{y}}(y) > 0\ \forall\ y \in \mathcal{Y}$, $P_{\mathbf{y}'}(y') > 0\ \forall\ y' \in \mathcal{Y}$. Now,

$$D(P_{\mathbf{t},\mathbf{x}',\mathbf{s}',\mathbf{y}}||P_{\mathbf{t}}P_{\mathbf{x}'|\mathbf{t}}P_{\mathbf{s}'|\mathbf{x}',\mathbf{t}}W_{\mathbf{y}'|\mathbf{x}',\mathbf{s}'}) \le \eta$$

$$\stackrel{(a)}{\Longrightarrow} D(P_{\mathbf{t},\mathbf{x}',\mathbf{y}}||P_{\mathbf{t},\mathbf{x}',\mathbf{y}'}) \leq \eta$$
$$\stackrel{(b)}{\Longrightarrow} \|P_{\mathbf{t},\mathbf{x}',\mathbf{y}'} - P_{\mathbf{t},\mathbf{x}',\mathbf{y}}\| \leq c\sqrt{\eta} \tag{54}$$

Here,
(a) follows from projecting the divergence space from $\mathcal{T}\times\mathcal{X}\times\mathcal{S}\times\mathcal{Y}$ to $\mathcal{T}\times\mathcal{X}\times\mathcal{Y}$ as it doesn't increases the KL-Divergence.
(b) using pinsker's inequality (cf. Lemma 2).

Equation (54) further implies:

$$\|P_{\mathbf{x}',\mathbf{y}'|\mathbf{t}} - P_{\mathbf{x}',\mathbf{y}|\mathbf{t}}\| \leq f(\eta) \tag{55}$$
$$\|P_{\mathbf{y}'|\mathbf{t}} - P_{\mathbf{y}|\mathbf{t}}\| \leq f^{'}(\eta) \tag{56}$$

where $f(\eta)$ and $f^{'}(\eta)$ are decaying functions in block length n. Therefore from equations (54), (55), (56) and noting that logarithm function is a continuous function, we have

$$P_{\mathbf{t},\mathbf{x}'\mathbf{y}'}\log\left(\frac{P_{\mathbf{x}',\mathbf{y}'|\mathbf{t}}}{P_{\mathbf{x}'|\mathbf{t}}P_{\mathbf{y}'|\mathbf{t}}}\right) - P_{\mathbf{t},\mathbf{x}'\mathbf{y}}\log\left(\frac{P_{\mathbf{x}',\mathbf{y}|\mathbf{t}}}{P_{\mathbf{x}'|\mathbf{t}}P_{\mathbf{y}|\mathbf{t}}}\right) \leq f^{''}(\eta)$$
$$I(\mathbf{x}';\mathbf{y}'|\mathbf{t}) - I(\mathbf{x}';\mathbf{y}|\mathbf{t}) \leq f^{''}(\eta) \tag{57}$$

Now, note from equations (53) and (57) that an appropriate choice of $\eta(\delta) > 0$ can be made such that $f^{''}(\eta) \leq \delta/2$ holds. ■

We fix a input codeword $\underline{\mathbf{x}} : P_{\underline{\mathbf{x}}} \in \lambda_{\mathbf{x}} \subseteq \Delta^{(n)}(\mathcal{X})$ and a state codeword $\underline{\mathbf{s}} : P_{\underline{\mathbf{s}}} \in \lambda_{\mathbf{s}} \subseteq \Delta^{(n)}(\mathcal{S})$. Let $\underline{\mathbf{y}} \in \mathcal{Y}^n$ be the output of the channel. Let us define two subsets $\mathbf{A}, \mathbf{B} \in \mathcal{Y}^n$:

$$\mathbf{A} := \{\underline{\mathbf{y}} : D(P_{\underline{\mathbf{t}},\underline{\mathbf{x}},\underline{\mathbf{s}},\underline{\mathbf{y}}}||P_{\underline{\mathbf{t}}}P_{\underline{\mathbf{x}}|\underline{\mathbf{t}}}P_{\underline{\mathbf{s}}|\underline{\mathbf{t}},\underline{\mathbf{x}}}W_{\underline{\mathbf{y}}|\underline{\mathbf{x}},\underline{\mathbf{s}}}) \geq \eta\}$$
$$\mathbf{B} := \{\underline{\mathbf{y}} : I(\hat{\underline{\mathbf{x}}};\underline{\mathbf{y}}|\underline{\mathbf{t}},\underline{\mathbf{x}},\underline{\mathbf{s}}) \geq \eta \text{ for some } \hat{\underline{\mathbf{x}}} \in \mathcal{L}(\underline{\mathbf{y}})\}$$

Further, let E_1 and E_2 denote the event $\underline{\mathbf{y}} \in \mathbf{A}$ and $\underline{\mathbf{y}} \in \mathbf{B}$, respectively for a given $(\underline{\mathbf{x}},\underline{\mathbf{s}})$.
The probability of error $(\mathbb{P}_e)$ is given by:

$$\mathbb{P}_{e(\underline{\mathbf{x}},\underline{\mathbf{s}})} = \mathbb{P}_{\mathbf{w}}(E_1 \cup E_2)$$
$$\leq \mathbb{P}_{\mathbf{w}}(E_1) + \mathbb{P}_{\mathbf{w}}(E_2) \tag{58}$$

From Theorem 4, we have

$$\mathbb{P}_{\mathbf{w}}(E_1) = 2^{-n[D(P_{\mathbf{t},\mathbf{x},\mathbf{s},\mathbf{y}}||P_{\mathbf{t}}P_{\mathbf{x}|\mathbf{t}}P_{\mathbf{s}|\mathbf{t},\mathbf{x}}W_{\mathbf{y}|\mathbf{x},\mathbf{s}})+H(\mathbf{y}|\mathbf{t},\mathbf{x},\mathbf{s})]}$$

On taking union over all $\underline{\mathbf{y}} \in \mathcal{P}(P_{\mathbf{y}|\mathbf{t},\mathbf{x},\mathbf{s}})$ and conditional distributions $P_{\mathbf{y}|\mathbf{t},\mathbf{x},\mathbf{s}}$, we have

$$\mathbb{P}_{\mathbf{w}}(E_1) \leq (n+1)^{|\mathcal{T}||\mathcal{X}||\mathcal{S}||\mathcal{Y}|}2^{-n[D(P_{\mathbf{t},\mathbf{x},\mathbf{s},\mathbf{y}}||P_{\mathbf{t}}P_{\mathbf{x}|\mathbf{t}}P_{\mathbf{s}|\mathbf{t},\mathbf{x}}W_{\mathbf{y}|\mathbf{x},\mathbf{s}})]}$$
$$\leq (n+1)^{|\mathcal{T}||\mathcal{X}||\mathcal{S}||\mathcal{Y}|}2^{-n\eta} \tag{59}$$

$$\mathbb{P}_{\mathbf{w}}(E_2) : = \mathbb{P}_{\mathbf{w}}(\underline{\mathbf{y}} \in \mathbf{B})$$
$$= \sum_{P_{\underline{\mathbf{y}}|\underline{\mathbf{t}},\underline{\mathbf{x}},\underline{\mathbf{s}}} \in \Delta(\mathcal{Y}|P_{\underline{\mathbf{t}},\underline{\mathbf{x}},\underline{\mathbf{s}}})} \mathbb{P}_{\mathbf{w}}(\underline{\mathbf{y}} \in \mathbf{B} \cap \mathcal{P}(P_{\underline{\mathbf{y}}|\underline{\mathbf{t}},\underline{\mathbf{x}},\underline{\mathbf{s}}})) \tag{60}$$

Here, the cardinality of the set $\mathbf{B} \cap \mathcal{P}(P_{\underline{\mathbf{y}}|\underline{\mathbf{t}},\underline{\mathbf{x}},\underline{\mathbf{s}}})$ is upper bounded by the number of $(\underline{\mathbf{x}}',\ \underline{\mathbf{y}})$ pairs in the space $\mathcal{C}\times\mathcal{Y}^n$ i.e.,

$$|\mathbf{B} \cap \mathcal{P}(P_{\underline{\mathbf{y}}|\underline{\mathbf{t}},\underline{\mathbf{x}},\underline{\mathbf{s}}})| \leq \#(\underline{\mathbf{x}}',\underline{\mathbf{y}}) \in \mathcal{C}\times\mathcal{Y}^n \tag{61}$$

such that the joint type $P_{\underline{\mathbf{t}},\underline{\mathbf{x}},\underline{\mathbf{s}},\underline{\mathbf{x}}',\underline{\mathbf{y}}}$ is close to the joint distribution $P_{\mathbf{t},\mathbf{x},\mathbf{s},\mathbf{x}',\mathbf{y}} := P$ satisfies:

- $[P]_{\mathbf{t},\mathbf{x},\mathbf{s},\mathbf{y}}(t,x,s,y) := P_{\mathbf{t},\mathbf{x},\mathbf{s}}(t,x,s)P_{\mathbf{y}|\mathbf{t},\mathbf{x},\mathbf{s}}(y|t,x,s)$ where $P_{\mathbf{t},\mathbf{x},\mathbf{s}}$ is close to $P_{\underline{\mathbf{t}},\underline{\mathbf{x}},\underline{\mathbf{s}}}$.

- for some random variable $\mathbf{x}'$ such that $[P]_{\mathbf{t},\mathbf{x}'}(t,x') := P_{\mathbf{t}}(t)P_{\mathbf{x}'|\mathbf{t}}(x'|t)$ and $[P]_{\mathbf{x}'} \in \lambda_{\mathbf{x}}$, we have the following

$$D(P_{\mathbf{t},\mathbf{x}',\mathbf{s}',\mathbf{y}}||P_{\mathbf{t}}P_{\mathbf{x}'|\mathbf{t}}P_{\mathbf{s}'|\mathbf{t},\mathbf{x}'}W_{\mathbf{y}|\mathbf{x}',\mathbf{s}'}) \leq \eta \tag{62}$$
$$I(\mathbf{x}';\mathbf{y}|\mathbf{t},\mathbf{x},\mathbf{s}) \geq \eta \tag{63}$$

holds for some $\mathbf{s}'$ such that $P_{\mathbf{s}'} \in \lambda_{\mathbf{s}}$.

Therefore, from property (ii) of the codebook in Lemma 36

$$\#\text{ sequences } \underline{\mathbf{x}}' \in \mathcal{C} \leq 3(n+1)^{|\mathcal{X}||\mathcal{T}|}2^{n|R-I(\mathbf{x},\mathbf{s};\mathbf{x}'|\mathbf{t})|^+} \tag{64}$$

where $P_{\underline{\mathbf{t}},\underline{\mathbf{x}},\underline{\mathbf{s}},\underline{\mathbf{x}}'}$ is close to the joint distribution $P_{\mathbf{t},\mathbf{x},\mathbf{s},\mathbf{x}'}$.
For every $\underline{\mathbf{x}}'$ selected according to the distribution $P_{\mathbf{t},\mathbf{x},\mathbf{s},\mathbf{x}'}$, we have

$$\#\text{ sequences } \underline{\mathbf{y}} \in \mathcal{Y}^n := |\mathcal{P}(P_{\underline{\mathbf{y}}|\underline{\mathbf{t}},\underline{\mathbf{x}},\underline{\mathbf{s}},\underline{\mathbf{x}}'})| \leq 2^{nH(\mathbf{y}|\mathbf{t},\mathbf{x},\mathbf{s},\mathbf{x}')} \tag{65}$$

where $P_{\underline{\mathbf{t}},\underline{\mathbf{x}},\underline{\mathbf{s}},\underline{\mathbf{x}}',\underline{\mathbf{y}}}$ is close to the joint distribution $P_{\mathbf{t},\mathbf{x},\mathbf{s},\mathbf{x}',\mathbf{y}}$.
Substituting the bound on number of number of sequences from (64) and (65) in (61)

$$\begin{aligned}
|\mathbf{B} \cap \mathcal{P}(P_{\underline{\mathbf{y}}|\underline{\mathbf{t}},\underline{\mathbf{x}},\underline{\mathbf{s}}})| &\leq 3(n+1)^{|\mathcal{T}||\mathcal{X}|} 2^{n(H(\mathbf{y}|\mathbf{t},\mathbf{x},\mathbf{s},\mathbf{x}')+|R-I(\mathbf{x}';\mathbf{x},\mathbf{s}|\mathbf{t})|^+)} \\
&\overset{(a)}{\leq} 3(n+1)^{|\mathcal{T}||\mathcal{X}|^2|\mathcal{S}||\mathcal{Y}|}(n+1)^{|\mathcal{T}||\mathcal{X}|} 2^{n(H(\mathbf{y}|\mathbf{t},\mathbf{x},\mathbf{s})-I(\mathbf{x}';\mathbf{y}|\mathbf{t},\mathbf{x},\mathbf{s})+|R-I(\mathbf{x}';\mathbf{x},\mathbf{s}|\mathbf{t})|^+)} \\
&\overset{(b)}{\leq} \max_{\mathbf{x}'} \left\{ 3(n+1)^{|\mathcal{T}||\mathcal{X}|^2|\mathcal{S}||\mathcal{Y}|}(n+1)^{|\mathcal{T}||\mathcal{X}|} 2^{n(H(\mathbf{y}|\mathbf{t},\mathbf{x},\mathbf{s})-I(\mathbf{x}';\mathbf{y}|\mathbf{t},\mathbf{x},\mathbf{s})+|R-I(\mathbf{x}';\mathbf{x},\mathbf{s}|\mathbf{t})|^+)} \right\} \\
&\leq 3(n+1)^{|\mathcal{T}||\mathcal{X}|^2|\mathcal{S}||\mathcal{Y}|}(n+1)^{|\mathcal{T}||\mathcal{X}|} 2^{n(H(\mathbf{y}|\mathbf{t},\mathbf{x},\mathbf{s})-\min_{\mathbf{x}'}[I(\mathbf{x}';\mathbf{y}|\mathbf{t},\mathbf{x},\mathbf{s})+|R-I(\mathbf{x}';\mathbf{x},\mathbf{s}|\mathbf{t})|^+])}
\end{aligned} \tag{66}$$

Here
(a) follows from taking union over all conditional distributions $P_{\mathbf{y}|\mathbf{t},\mathbf{x},\mathbf{s},\mathbf{x}'}$ such that $[P_{\mathbf{y}|\mathbf{t},\mathbf{x},\mathbf{s},\mathbf{x}'}]_{\mathbf{y}|\mathbf{t},\mathbf{x},\mathbf{s}} = P_{\mathbf{y}|\mathbf{t},\mathbf{x},\mathbf{s}}$.
(b) the maximum is taken over all such random variables $\mathbf{x}' \in \mathcal{C}$ satisfying the conditions in (62) and (63).

Case (i): if $R < I(\mathbf{x}';\mathbf{x},\mathbf{s}|\mathbf{t})$
equation (66) reduces to following

$$\begin{aligned}
|\mathbf{B} \cap \mathcal{P}(P_{\underline{\mathbf{y}}|\underline{\mathbf{t}},\underline{\mathbf{x}},\underline{\mathbf{s}}})| &\leq 3(n+1)^{|\mathcal{T}||\mathcal{X}|^2|\mathcal{S}||\mathcal{Y}|}(n+1)^{|\mathcal{T}||\mathcal{X}|} 2^{n(H(\mathbf{y}|\mathbf{t},\mathbf{x},\mathbf{s})-\min_{\mathbf{x}'}[I(\mathbf{x}';\mathbf{y}|\mathbf{t},\mathbf{x},\mathbf{s})])} \\
&\leq 3(n+1)^{|\mathcal{T}||\mathcal{X}|^2|\mathcal{S}||\mathcal{Y}|}(n+1)^{|\mathcal{T}||\mathcal{X}|} 2^{n(H(\mathbf{y}|\mathbf{t},\mathbf{x},\mathbf{s})-\eta)}
\end{aligned} \tag{67}$$

Case (ii): if $R > I(\mathbf{x}';\mathbf{x},\mathbf{s}|\mathbf{t})$

equation (66) reduces to following

$$\begin{aligned}
|\mathbf{B} \cap \mathcal{P}(P_{\underline{\mathbf{y}}|\underline{\mathbf{t}},\underline{\mathbf{x}},\underline{\mathbf{s}}})| &\leq 3(n+1)^{|\mathcal{T}||\mathcal{X}|^2|\mathcal{S}||\mathcal{Y}|}(n+1)^{|\mathcal{T}||\mathcal{X}|} 2^{n(H(\mathbf{y}|\mathbf{t},\mathbf{x},\mathbf{s})+R-\min_{\mathbf{x}'}[I(\mathbf{x}';\mathbf{y}|\mathbf{t},\mathbf{x},\mathbf{s})+I(\mathbf{x}';\mathbf{x},\mathbf{s}|\mathbf{t})])} \\
&\overset{(a)}{\leq} 3(n+1)^{|\mathcal{T}||\mathcal{X}|^2|\mathcal{S}||\mathcal{Y}|}(n+1)^{|\mathcal{T}||\mathcal{X}|} 2^{n(H(\mathbf{y}|\mathbf{t},\mathbf{x},\mathbf{s})+R-\min_{\mathbf{x}'} I(\mathbf{x}';\mathbf{x},\mathbf{s},\mathbf{y}|\mathbf{t}))} \\
&\overset{(b)}{\leq} 3(n+1)^{|\mathcal{T}||\mathcal{X}|^2|\mathcal{S}||\mathcal{Y}|}(n+1)^{|\mathcal{T}||\mathcal{X}|} 2^{n(H(\mathbf{y}|\mathbf{t},\mathbf{x},\mathbf{s})+R-\min_{\mathbf{x}'} I(\mathbf{x}';\mathbf{y}|\mathbf{t}))} \\
&\overset{(c)}{\leq} 3(n+1)^{|\mathcal{T}||\mathcal{X}|^2|\mathcal{S}||\mathcal{Y}|}(n+1)^{|\mathcal{T}||\mathcal{X}|} 2^{n(H(\mathbf{y}|\mathbf{t},\mathbf{x},\mathbf{s})+R-\min_{\mathbf{x}'} I(\mathbf{x}';\mathbf{y}'|\mathbf{t})+\delta/2)} \\
&\overset{(d)}{\leq} 3(n+1)^{|\mathcal{T}||\mathcal{X}|^2|\mathcal{S}||\mathcal{Y}|}(n+1)^{|\mathcal{T}||\mathcal{X}|} 2^{n(H(\mathbf{y}|\mathbf{t},\mathbf{x},\mathbf{s})-\delta/2)} \\
&\overset{(e)}{\leq} 3(n+1)^{|\mathcal{T}||\mathcal{X}|^2|\mathcal{S}||\mathcal{Y}|}(n+1)^{|\mathcal{T}||\mathcal{X}|} 2^{n(H(\mathbf{y}|\mathbf{t},\mathbf{x},\mathbf{s})-\eta)}
\end{aligned} \tag{68}$$

Here
(a) follows from chain rule for conditional mutual information.
(b) follows from the property of mutual information and noting that it is a positive quantity.
(c) follows from the assumption in Lemma 38.
(d) follows from noting that $R \leq \text{DMC}(P_{\mathbf{x}|\mathbf{t}}) - \delta$.
(e) follows from choosing η such that $\eta \leq \frac{\delta}{2}$.

from (66), (67) and (68), we have

$$|\mathbf{B} \cap \mathcal{P}(P_{\underline{\mathbf{y}}|\underline{\mathbf{t}},\underline{\mathbf{x}},\underline{\mathbf{s}}})| \leq 3(n+1)^{|\mathcal{T}||\mathcal{X}|^2|\mathcal{S}||\mathcal{Y}|}(n+1)^{|\mathcal{T}||\mathcal{X}|} 2^{n(H(\mathbf{y}|\mathbf{t},\mathbf{x},\mathbf{s})-\eta)} \tag{69}$$

from (60)

$$\begin{aligned}
\mathbb{P}_{\mathbf{w}}(E_2) &= \sum_{P_{\underline{\mathbf{y}}|\underline{\mathbf{t}},\underline{\mathbf{x}},\underline{\mathbf{s}}} \in \Delta(\mathcal{Y}|P_{\underline{\mathbf{t}},\underline{\mathbf{x}},\underline{\mathbf{s}}})} \mathbb{P}_{\mathbf{w}}(\underline{\mathbf{y}} \in \mathbf{B} \cap \mathcal{P}(P_{\underline{\mathbf{y}}|\underline{\mathbf{t}},\underline{\mathbf{x}},\underline{\mathbf{s}}})) \\
&= \sum_{P_{\underline{\mathbf{y}}|\underline{\mathbf{t}},\underline{\mathbf{x}},\underline{\mathbf{s}}}} \frac{|\mathbf{B} \cap \mathcal{P}(P_{\underline{\mathbf{y}}|\underline{\mathbf{t}},\underline{\mathbf{x}},\underline{\mathbf{s}}})|}{|\mathcal{P}(P_{\underline{\mathbf{y}}|\underline{\mathbf{t}},\underline{\mathbf{x}},\underline{\mathbf{s}}})|} \\
&\leq \sum_{P_{\underline{\mathbf{y}}|\underline{\mathbf{t}},\underline{\mathbf{x}},\underline{\mathbf{s}}}} \frac{3(n+1)^{|\mathcal{T}||\mathcal{X}|^2|\mathcal{S}||\mathcal{Y}|}(n+1)^{|\mathcal{T}||\mathcal{X}|} 2^{n(H(\mathbf{y}|\mathbf{t},\mathbf{x},\mathbf{s})-\eta)}}{(n+1)^{-|\mathcal{T}||\mathcal{X}||\mathcal{S}||\mathcal{Y}|} 2^{nH(\mathbf{y}|\mathbf{t},\mathbf{x},\mathbf{s})}}
\end{aligned}$$

$$\leq 3(n+1)^{|\mathcal{T}||\mathcal{X}|(2|\mathcal{S}||\mathcal{Y}|+|\mathcal{X}||\mathcal{S}||\mathcal{Y}|+1)}2^{-n\eta} \tag{70}$$

Further, substituting values from (59) and (70) in (58)

$$\begin{aligned}\mathbb{P}_{e(\underline{\mathbf{x}},\underline{\mathbf{s}})} &\leq \mathbb{P}_{\mathbf{w}}(E_1) + \mathbb{P}_{\mathbf{w}}(E_2) \\ &\leq (n+1)^{|\mathcal{T}||\mathcal{X}||\mathcal{S}||\mathcal{Y}|}2^{-n\eta} + 3(n+1)^{|\mathcal{T}||\mathcal{X}|(2|\mathcal{S}||\mathcal{Y}|+|\mathcal{X}||\mathcal{S}||\mathcal{Y}|+1)}2^{-n\eta} \\ &= \left[(n+1)^{|\mathcal{T}||\mathcal{X}||\mathcal{S}||\mathcal{Y}|} + 3(n+1)^{|\mathcal{T}||\mathcal{X}|(2|\mathcal{S}||\mathcal{Y}|+|\mathcal{X}||\mathcal{S}||\mathcal{Y}|+1)}\right]2^{-n\eta} \\ &= \delta_n^{'} \end{aligned} \tag{71}$$

where $\delta_n^{'} \to 0$ as $n \to \infty$. Thus, the error probability $\mathbb{P}_e$

$$\mathbb{P}_e := \max_{\underline{\mathbf{x}} \in \mathcal{C}} \max_{\underline{\mathbf{s}}: P_{\underline{\mathbf{s}}} \in \lambda_{\mathbf{s}}} \mathbb{P}_{e(\underline{\mathbf{x}},\underline{\mathbf{s}})} \leq \delta_n^{'}$$

is negligible in the blocklength n for the appropriate choice of the parameter η such that $\eta \leq \frac{\delta}{2}$ (by assumption), satisfies the uambiguity of the decoding rules in Lemma (37), and satisfies the condition in Lemma 38.
This completes our proof. ■

IX. Achievability: Rate Positivity for state-deterministic and Myopic AVCs

To give context to the definition 18:

1) The role of the time-sharing variable $\mathbf{t}$ is the same as in omniscient AVCs (where the precise characterization of rate-positivity [WBBJ19] requires such a time-sharing variable in input-constrained settings), since omniscient AVCs are a special case of myopic AVCs.

2) The role of the auxiliary channel $P_{\mathbf{x}|\mathbf{u},\mathbf{t}}$ is to allow for coding-schemes of the form where the encoder passes the "pre-codeword" $\underline{u}$ through an auxiliary channel $P_{\mathbf{x}|\mathbf{u},\mathbf{t}}$ to generate the transmitted codeword $\underline{x}$ – as shown in [Wan16], the use of such auxiliary channels can boost the achievable rate for some myopic AVCs by increasing the parameter regime where they are sufficiently myopic (and hence more benign – more like random-noise channels than the insufficiently myopic setting; achievable rates for at least some classes of sufficiently myopic AVCs can be shown [DJL19a], [BDJ$^+$21] to equal Shannon capacities of some corresponding compound DMCs, whereas the insufficiently myopic setting is in general much less well-understood, being closer to the omniscient setting where capacity characterizations are in general long-standing open problems even for "simple" bit-flip channels). This is one of at least two reasons to use an auxiliary channel in encoding (the second is outlined just below).

3) A consequence of a channel being insufficiently myopic is that for the usual ensemble of code construction techniques employed in the AVC literature, if the decoder is able to reconstruct the transmitter's intended message $\underline{u}$, then the jammer must similarly be able to do so, and it seems like that in this case the problem collapses to the omniscient setting. However, this is not quite true – there is a potential advantage for the encoder-decoder pair over the jammer in a myopic setting (as opposed to omniscient settings where $\underline{z} = \underline{x}$) lying in the fact that for some myopic AVCs, with a carefully designed communication scheme (such as in [DJL$^+$19b], [BDJ$^+$21]) the decoder may be able to reliably estimate the transmitted message without needing to reconstruct the transmitted codeword $\underline{x}$, whereas the jammer, even if it knows the encoder's intended message, may not be able to estimate $\underline{x}$ well enough to design a jamming attack that can prevent positive rate communication. Indeed, it is precisely this phenomenon that, for the binary-erasure AVC considered in [DJL$^+$19b], [BDJ$^+$21] allows for positive rate communication for some parameter values where the AVC is insufficiently myopic. This is a second reason why auxiliary encoding channels can boost throughput.

4) One can distinguish between omniscient and myopic AVCs by examining the conditional entropy $H(\mathbf{x}|\mathbf{z})$ under the given input distribution – if it is precisely zero then the AVC is omniscient, otherwise it is myopic. For our achievability arguments, however, we need the somewhat stronger condition of *input-aware symbolwise myopicity* with respect to some distribution $P_{\mathbf{t},\mathbf{u},\mathbf{x}}$ if $H(\mathbf{x}|\mathbf{t}=t,\mathbf{u}=u,\mathbf{z}=z) > 0$ for all $(t,u,z) \in \mathcal{T} \times \mathcal{U} \times \mathcal{Z}$. We need this definition to ensure that the adversary has some uncertainty about every location of $\underline{x}$. If in addition such a $P_{\mathbf{t},\mathbf{u},\mathbf{x}}$ is also myopically non-symmetrizable in the sense of Definition 20, then we can prove the exist of codes achieving positive rate.

5) To motivate our codes in this section, note that the code design in the binary-erasure AVC considered in [BDJ$^+$21] (which improved on the prior work in [DJL$^+$19b]) is an example of an AVC where, despite being insufficiently myopic, rates higher than those achievable against an omniscient adversary can be achieved. In particular, specialized to the scheme in [BDJ$^+$21], the time-sharing variable $\mathbf{t}$ is uniform and hence may be omitted, and the auxiliary input/channel $P_{\mathbf{u},\mathbf{x}}$ can be constructed in the following manner. Let $\mathbf{u} = (\mathbf{u}_1, \mathbf{u}_2)$, be uniformly distributed on the alphabet $\{0,1\} \times [m]$ (hence $\mathbf{u}_1$

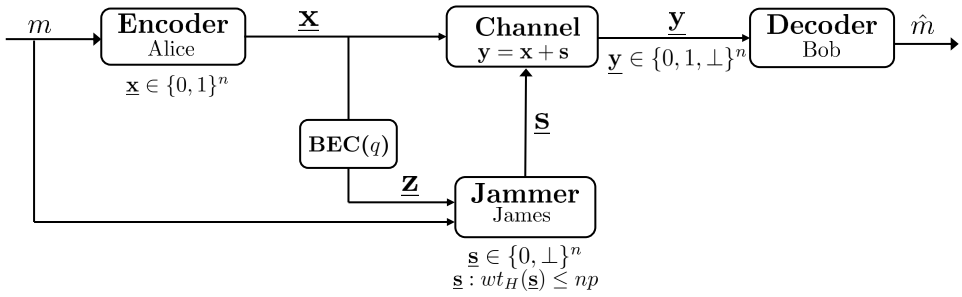


Fig. 6. Communication over a myopic and state-deterministic A$_m$-BEC(q,p).

is binary and $\mathbf{u}_2$ is m-ary), where m is a suitably large but fixed positive integer (the larger the value of q, the larger the fraction of the insufficiently myopic regime for which the resulting codes can attain a positive rate – in the limit, they attain a positive rate for all parameter regimes where a positive rate is not precluded by the symmetrizability converses in [DJL$^+$19b], [BDJ$^+$20]). The auxiliary channel $P_{\mathbf{x}|\mathbf{u}}$ in that work can be presented as

$$P_{\mathbf{x}|\mathbf{u}}(x|u_1,u_2) = \begin{cases} 1 & \text{if } u_2 = 0 \text{ and } x = u_1 \\ 1 - \frac{1}{n^{\frac{1}{2}-\frac{u_2}{m+1}}} & \text{if } u_2 > 0 \text{ and } x = u_1 \\ \frac{1}{n^{\frac{1}{2}-\frac{u_2}{m+1}}} & \text{if } u_2 > 0 \text{ and } x \neq u_1 \end{cases}$$

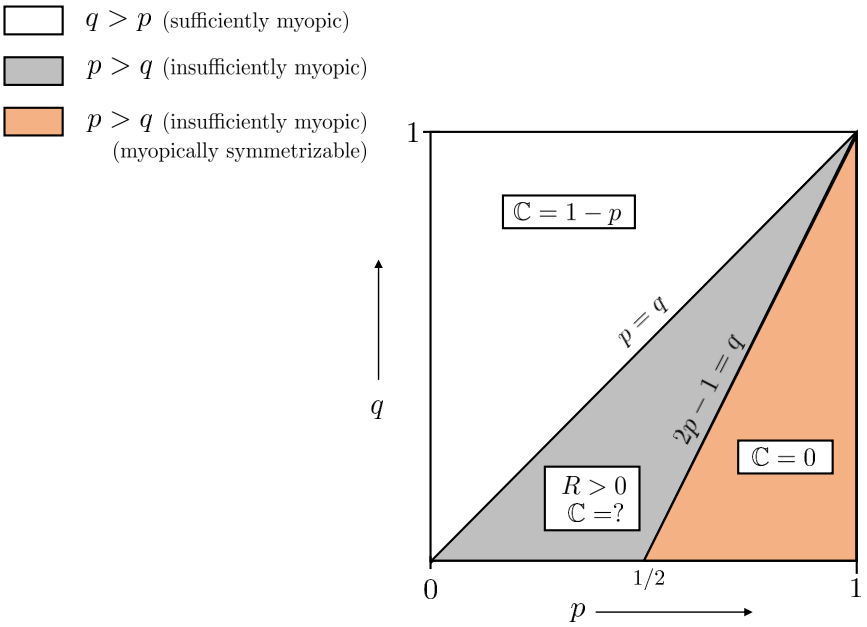


Fig. 7. Communication over a state-deterministic myopic binary-erasure AVC where $W_{\mathbf{z}|\mathbf{x}}$ is a BEC(q), and the adversary can then use $\underline{z}$ to determine which of up to pn bits of $\underline{x}$ to erase. When $q = 0$ the AVC is omniscient, and here it is known htat no positive rate is possible when $p > 1/2$. Results in [DJL19a] demonstrate that when $q > p$ (the AVC is "sufficiently myopic"), the capacity is precisely $1 - p$; results in [DJL$^+$19b] show that no positive-rate communication is possible when $2p - 1 > q$ (the AVC is "myopically symmetrizable"), and results in [BDJ$^+$21] demonstrate that even when $p > q$ (the AVC is "insufficiently myopic") as long as $2p - 1 < q$ (the AVC is "not myopically symmetrizable") a positive rate is achievable.

To summarize, time-sharing can help boost throughput in input-constrained settings, whereas auxiliary input channels may help in two distinct (and overlapping) ways – by increasing the parameter regime where the channel is sufficiently myopic (and hence more benign), and even in the parameter regime where the AVC is insufficiently myopic it can help enable/improve

communication by preventing effective jamming.

Given any input-aware symbolwise myopic distribution $P_{\mathbf{u},\mathbf{x}}$ that is not in $\mathcal{K}^m_{ach}$ (recall Definition 20) – more specifically is δ-separated from and outside the boundary of $\mathcal{K}^m_{ach}$, may be used to construct a code achieving positive rate for the given state-deterministic myopic AVC. The code design/encoding/decoding and analysis is very similar to the proof (in Section VII-B) of rate-positivity for omniscient AVCs specialized to the case of state-deterministic AVCs, with one additional complication – since the adversary's action is on $\underline{x}$ but it only views it imperfectly through $\underline{z}$ (indeed, this is where the potential advantage for Alice/Bob over James arises from) one needs to argue that the joint types of vectors behave "typically" with respect to some reference distribution. To show this, the "input-aware symbolwise myopicity" of the channel coupled with a certain Markov Concentration Lemma [GK11], [BDP17a] is useful. Below we , highlighting the key differences.

1) We start with the input-aware symbolwise myopic distribution $P_{\mathbf{u},\mathbf{x}}$ that is not in $\mathcal{K}^m_{ach}$. As in the rate-positivity code construction for omniscient AVCs, we first identify a time-sharing variable $\mathbf{t}$ (with corresponding time-sharing sequence $\underline{t}$) of the corresponding CP distribution $P_{\mathbf{u},\mathbf{u}'}$ and construct a cloud code comprising of 2^{nR} length-n $\underline{u}$ pre-codeword sequences according to the distribution $P_{\mathbf{t}}P_{\mathbf{u}|\mathbf{t}}$. By the same argument as in Lemma 32, the joint type of all pairs of pre-codewords $\underline{u}, \underline{u}'$ are "close" to the reference distribution $P_{\mathbf{u},\mathbf{u}'}$.
2) Given a message $m \in \mathcal{M}$ the encoder selects the corresponding pre-codeword $\underline{u}(m)$, and passes it through the auxiliary channel $P_{\mathbf{x}|\mathbf{u},\mathbf{t}}$ to result in the codeword $\underline{x}$ actually transmitted by Alice. This codeword then passes through the channel $W_{\mathbf{z}|\mathbf{x}}$ to result in the vector $\underline{z}$ observed by James. Since both $P_{\mathbf{x}|\mathbf{t},\mathbf{u}}$ and $W_{\mathbf{z}|\mathbf{x}}$ are independent across time-indices t, therefore via the Chernoff bound the joint type of the vectors $\underline{t}, \underline{u}, \underline{x}, \underline{z}$ is "close" to the reference distribution $P_{\mathbf{t}}(t)P_{\mathbf{u}|\mathbf{t}}(u|t)P_{\mathbf{x}|\mathbf{t},\mathbf{u}}(x|t,u)W_{\mathbf{z}|\mathbf{x}}(z|x)$. Indeed, due to 1 above, given any other ("spoofing") pre-codeword $\underline{u}'$ in the 2^{nR}-size collection generated by Alice (and known *a priori* to all parties), the joint type of all $\underline{t}, \underline{u}, \underline{u}', \underline{x}, \underline{z}$ is "close" to the reference distribution $P_{\mathbf{t}}(t)P_{\mathbf{u}|\mathbf{t}}(u|t)P_{\mathbf{u}|\mathbf{t}}(u'|t)P_{\mathbf{x}|\mathbf{t},\mathbf{u}}(x|t,u)W_{\mathbf{z}|\mathbf{x}}(z|x)$.
3) Now, using the fact that AVC is input-aware symbolwise myopic w.r.t. the distribution $P_{\mathbf{t},\mathbf{u},\mathbf{x}}$, (i.e. $H(\mathbf{x}|\mathbf{t}=t,\mathbf{u}=u,\mathbf{z}=z)>0$ for all (t,u,z)) via standard concentration arguments in the AVC literature one can demonstrate a Markov concentration argument (philosophically similar to those in [BDP17b]) arguing that for any state vector $\underline{s}$ chosen by James (as a function of $\underline{t}, \underline{u}, \underline{u}', \underline{z}$), with high probability over the $\underline{x}$ that are jointly typical w.r.t. $\underline{t}, \underline{u}, \underline{u}', \underline{z}$, the joint type of $\underline{t}, \underline{u}, \underline{u}', \underline{x}, \underline{z}, \underline{s}$ is "close" to the reference distribution $P_{\mathbf{t}}(t)P_{\mathbf{u}|\mathbf{t}}(u|t)P_{\mathbf{u}|\mathbf{t}}(u'|t)P_{\mathbf{x}|\mathbf{t},\mathbf{u}}(x|t,u)W_{\mathbf{z}|\mathbf{x}}(z|x)V_{\mathbf{s}|\mathbf{t},\mathbf{u},\mathbf{u}',\mathbf{z}}(s|t,u,u',z)$. That is, the random variable $\mathbf{s}$ corresponding to the state sequence is independent of the random variable corresponding to the transmitted codeword $\mathbf{x}$ conditioned on James' knowledge ($\mathbf{t}$)/inference of pre-codeword ($\mathbf{u}$)/choice of spoofing pre-codeword ($\mathbf{u}'$)/observation ($\mathbf{z}$).
4) With this statement relating the closeness of the joint-type to the reference distribution at hand, the remaining argument is similar to that in Section VII specialized to the state-deterministic setting, where one can use the fact that $P_{\mathbf{t},\mathbf{u},\mathbf{x}}$ satisfies weak myopic non-symmetrizability to imply that no other spoofing pre-codeword $\underline{u}'$ will have corresponding "witnesses" $\underline{x}', \underline{z}', \underline{s}'$ that would imply that the corresponding $\underline{u}'$ is typical w.r.t. the reference distribution. Hence typicality decoding suffices, and we are done.

X. Converse: Plotkin Bound and Elias-Bassalygo Bound for Myopic AVCs

A. Plotkin bound for myopic AVCs

Theorem 39 (Plotkin bound for myopic AVCs, [BDJ+20])**.** *Let $(\mathcal{X},\mathcal{Z},\mathcal{S},\mathcal{Y},\lambda_{\mathbf{x}},\lambda_{\mathbf{s}},W_{\mathbf{z}|\mathbf{x}},W_{\mathbf{y}|\mathbf{x},\mathbf{s}})$ be a myopic AVC. Suppose that for every $P_{\mathbf{x}} \in \lambda_{\mathbf{x}}$, every CP-self-coupling $P_{\mathbf{x},\mathbf{x}'} \in \mathcal{CP}(P_{\mathbf{x}})$ also satisfies $P_{\mathbf{x},\mathbf{x}'} \in \mathcal{K}^m_{\mathcal{CP}_{strong}}(P_{\mathbf{x}})$, i.e., for every decomposition $(P_{\mathbf{t}},P_{\mathbf{x}|\mathbf{t}}) \in \Delta(\mathcal{T})\times\Delta(\mathcal{X}|\mathcal{T})$ (let $\mathcal{T} := [K]$ for some $K \geq \mathbf{Z}_{\geq 1}$) satisfying $P_{\mathbf{x},\mathbf{x}'} = \left[P_{\mathbf{t}}P^{\otimes 2}_{\mathbf{x}|\mathbf{t}}\right]_{\mathbf{x},\mathbf{x}'}$, there exists $V_{\mathbf{s}|\mathbf{z},\mathbf{x}',\mathbf{t}} \in \Delta(\mathcal{S}|\mathcal{Z}\times\mathcal{X}\times\mathcal{T})$ such that*

1) *there exists a constant $\delta>0$ such that $\left[P_{\mathbf{t}}P^{\otimes 2}_{\mathbf{x}|\mathbf{t}}W_{\mathbf{z}|\mathbf{x}}V_{\mathbf{s}|\mathbf{z},\mathbf{x}',\mathbf{t}}\right]_{\mathbf{s}} \in \mathrm{int}_\delta(\lambda_{\mathbf{s}})$;*
2) *for every $(t,x,x',y) \in \mathcal{T}\times\mathcal{X}^2\times\mathcal{Y}$,*

$$\sum_{z\in\mathcal{Z}}\sum_{s\in\mathcal{S}} W_{\mathbf{z}|\mathbf{x}}(z|x)V_{\mathbf{s}|\mathbf{z},\mathbf{x}',\mathbf{t}}(s|z,x',t)W_{\mathbf{y}|\mathbf{x},\mathbf{s}}(y|x,s) = \sum_{z\in\mathcal{Z}}\sum_{s\in\mathcal{S}} W_{\mathbf{z}|\mathbf{x}}(z|x')V_{\mathbf{s}|\mathbf{z},\mathbf{x}',\mathbf{t}}(s|z,x,t)W_{\mathbf{y}|\mathbf{x},\mathbf{s}}(y|x',s).$$

Let $\mathcal{C}\subset\mathcal{X}^n$ be a code satisfying

1) *$P_{\underline{x}} \in \lambda_{\mathbf{x}}$ for every $\underline{x}\in\mathcal{C}$;*
2) *the average probability of error of $\mathcal{C}$ approaches 0 as $n\to\infty$.*

Then there exists a constant $M>0$ that depends on δ (and on the alphabet sizes as well) but not on n such that $|\mathcal{C}|\leq M$. In particular, the rate of $\mathcal{C}$ is vanishing as $n\to\infty$.

B. A covering lemma

Theorem 40. *Fix a triple of distributions $(P_{\mathbf{x}},W_{\mathbf{z}|\mathbf{x}},P_{\mathbf{t}|\mathbf{z}}) \in \Delta(\mathcal{X})\times\Delta(\mathcal{Z}|\mathcal{X})\times\Delta(\mathcal{T}|\mathcal{Z})$. Let $\lambda>0$ be a constant. Suppose that $\mathcal{C}=\{\underline{x}_i\}_{i=1}^M$ is a $(\lambda,P_{\mathbf{x}})$-constant composition code of size $M=2^{n(I(\mathbf{t};\mathbf{z})+\delta)}$ for some $\delta>0$ (sufficiently large compared*

with λ), where $I(\mathbf{t};\mathbf{z})$ is computed according to $P_{\mathbf{x}}W_{\mathbf{z}|\mathbf{x}}P_{\mathbf{t}|\mathbf{z}}$. Then for any constant $\varepsilon > 0$ such that $\lambda \ll \varepsilon \ll \delta$, we have

$$\Pr_{\underline{\mathbf{x}},\underline{\mathbf{z}},\underline{\mathbf{t}}}\Big[\Big|\Big\{\underline{x} \in \mathcal{C} : \Big\|P_{\underline{x},\underline{\mathbf{t}}} - \big[P_{\mathbf{x}}W_{\mathbf{z}|\mathbf{x}}P_{\mathbf{t}|\mathbf{z}}\big]_{\mathbf{x},\mathbf{t}}\Big\|_\infty \le \varepsilon\Big\}\Big| > 2^{n\varepsilon}\Big] \ge 1 - 2^{-n\delta/2}. \tag{72}$$

In the above probability, the randomness comes from $\underline{\mathbf{x}}$ that is uniformly distributed in $\mathcal{C}$, and $(\underline{\mathbf{z}},\underline{\mathbf{t}})$ that is ε-jointly typical with $\underline{\mathbf{x}}$, i.e., $(\underline{\mathbf{z}},\underline{\mathbf{t}}) \sim \mathrm{Unif}(\mathcal{P}_{\underline{\mathbf{z}},\underline{\mathbf{t}}|\underline{\mathbf{x}}})$, where for any $\underline{x} \in \mathcal{X}^n$ we define[2]

$$\mathcal{P}_{\underline{\mathbf{z}},\underline{\mathbf{t}}|\underline{x}} \coloneqq \big\{(\underline{z},\underline{t}) \in \mathcal{Z}^n \times \mathcal{T}^n : \big\|P_{\underline{x},\underline{z},\underline{t}} - P_{\mathbf{x}}W_{\mathbf{z}|\mathbf{x}}P_{\mathbf{t}|\mathbf{z}}\big\|_\infty \le \varepsilon\big\}. \tag{73}$$

Proof: We first study the typicality structures of the triple $(\underline{\mathbf{x}},\underline{\mathbf{z}},\underline{\mathbf{t}})$. By standard computation, for any $\underline{x} \in \mathcal{X}^n$ such that $\big\|P_{\underline{x}} - P_{\mathbf{x}}\big\|_\infty \le \lambda$ (where $\lambda \ll \varepsilon$), we have

$$2^{n(H(\mathbf{z}|\mathbf{x})+H(\mathbf{t}|\mathbf{z})+\varepsilon_*)} = 2^{n(H(\mathbf{z},\mathbf{t}|\mathbf{x})+\varepsilon_*)} \le \big|\mathcal{P}_{\underline{\mathbf{z}},\underline{\mathbf{t}}|\underline{x}}\big| \le 2^{n(H(\mathbf{z},\mathbf{t}|\mathbf{x})+\overline{\varepsilon}_*)} + 2^{n(H(\mathbf{z}|\mathbf{x})+H(\mathbf{t}|\mathbf{z})+\overline{\varepsilon}_*)}, \tag{74}$$

for some $0 < \varepsilon_* < \overline{\varepsilon}_*$ that depend on ε (and the alphabet sizes). The equality $H(\mathbf{z},\mathbf{t}|\mathbf{x}) = H(\mathbf{z}|\mathbf{x}) + H(\mathbf{t}|\mathbf{z})$ follows since since the entropy is computed according to the Markov chain $P_{\mathbf{x}}W_{\mathbf{z}|\mathbf{x}}P_{\mathbf{t}|\mathbf{z}}$. It is worth noting that the above bounds do not depend on the choice of $\underline{x}$ provided the assumption on its type.

For any $\underline{x} \in \mathcal{X}^n$ and $\underline{z} \in \mathcal{Z}^n$, we define the following conditional typical sets:

$$\mathcal{P}_{\underline{\mathbf{z}}|\underline{x}} \coloneqq \big\{\underline{z} \in \mathcal{Z}^n : \big\|P_{\underline{x},\underline{z}} - P_{\mathbf{x}}W_{\mathbf{z}|\mathbf{x}}\big\|_\infty \le \varepsilon_z\big\}, \tag{75}$$

$$\mathcal{P}_{\underline{\mathbf{t}}|\underline{x}} \coloneqq \Big\{\underline{t} \in \mathcal{T}^n : \Big\|P_{\underline{x},\underline{t}} - \big[P_{\mathbf{x}}W_{\mathbf{z}|\mathbf{x}}P_{\mathbf{t}|\mathbf{z}}\big]_{\mathbf{x},\mathbf{t}}\Big\|_\infty \le \varepsilon_t\Big\}, \tag{76}$$

where $\varepsilon_z > 0$ and $\varepsilon_t > 0$ are constants depending on ε (and the alphabet sizes) to be determined later. These sets also depend on $(P_{\mathbf{x}}, W_{\mathbf{z}|\mathbf{x}}, P_{\mathbf{t}|\mathbf{z}})$ and $\varepsilon_z, \varepsilon_t$, though we drop them in the notation. Similarly, for any $\underline{x} \in \mathcal{X}^n$ such that $\big\|P_{\underline{x}} - P_{\mathbf{x}}\big\|_\infty \le \lambda$ (where $\lambda \ll \varepsilon$), we have

$$2^{n(H(\mathbf{z}|\mathbf{x})+\varepsilon_0)} \le |\mathcal{P}_{\underline{\mathbf{z}}|\underline{x}}| \le 2^{n(H(\mathbf{z}|\mathbf{x})+\overline{\varepsilon})}, \tag{77}$$

$$2^{n(H(\mathbf{t}|\mathbf{x})+\varepsilon_0)} \le |\mathcal{P}_{\underline{\mathbf{t}}|\underline{x}}| \le 2^{n(H(\mathbf{t}|\mathbf{x})+\overline{\varepsilon})}, \tag{78}$$

for some $0 < \varepsilon_0 < \overline{\varepsilon}$ that depend on ε (and the alphabet sizes). Here $H(\mathbf{z}|\mathbf{x})$ and $H(\mathbf{t}|\mathbf{x})$ are evaluated using the joint distribution $P_{\mathbf{x}}W_{\mathbf{z}|\mathbf{x}}P_{\mathbf{t}|\mathbf{z}}$ and again the bounds do not depend on the choice of $\underline{x}$ (as long as the condition on $P_{\underline{x}}$ is satisfied).

Since joint typicality implies marginal typicality, if $(\underline{z},\underline{t}) \in \mathcal{P}_{\underline{\mathbf{z}},\underline{\mathbf{t}}|\underline{x}}$, then

$$\underline{z} \in \mathcal{P}_{\underline{\mathbf{z}}|\underline{x}} \text{ and } \underline{t} \in \mathcal{P}_{\underline{\mathbf{t}}|\underline{x}}, \tag{79}$$

provided $\varepsilon_z \ge |\mathcal{T}|\varepsilon$ and $\varepsilon_t \ge |\mathcal{Z}|\varepsilon$. This follows from the triangle inequality.

Define

$$\mathcal{F} \coloneqq \bigcup_{i=1}^{M} \mathcal{P}_{\underline{\mathbf{t}}|\underline{x}_i} \tag{80}$$

to be the set of $\mathcal{T}^n$-valued sequences that $\underline{\mathbf{t}}$ can potentially lie in. Again because joint typicality implies marginal typicality, there exists $\varepsilon' \ge |\mathcal{Z}||\mathcal{X}|\varepsilon_t$ (recall ε_t in turn is a function of ε) such that

$$\mathcal{F} \subset \big\{\underline{t} \in \mathcal{T}^n : \big\|P_{\underline{t}} - P_{\mathbf{t}}\big\|_\infty \le \varepsilon'\big\},$$

where $P_{\mathbf{t}}$ is obtained from the joint distribution $P_{\mathbf{x}}W_{\mathbf{z}|\mathbf{x}}P_{\mathbf{t}|\mathbf{z}}$. Therefore

$$|\mathcal{F}| \le 2^{n(H(\mathbf{t})+\varepsilon_1)} \tag{81}$$

for some $\varepsilon_1 > 0$ that depends on ε (and the alphabet sizes). Define also

$$\mathcal{F}' \coloneqq \Big\{\underline{t} \in \mathcal{T}^n : \Big|\Big\{\underline{x} \in \mathcal{C} : \Big\|P_{\underline{x},\underline{t}} - \big[P_{\mathbf{x}}W_{\mathbf{z}|\mathbf{x}}P_{\mathbf{t}|\mathbf{z}}\big]_{\mathbf{x},\mathbf{t}}\Big\|_\infty \le \varepsilon\Big\}\Big| \le 2^{n\varepsilon}\Big\} \tag{82}$$

to be the set of $\mathcal{T}^n$-valued sequences that only cover a small number (at most $2^{n\varepsilon}$) of codewords in $\mathcal{C}$. Under this definition, the probability in Equation (72) is equal to

$$\Pr_{\underline{\mathbf{x}},\underline{\mathbf{z}},\underline{\mathbf{t}}}[\underline{\mathbf{t}} \in \mathcal{F}'], \tag{83}$$

where the randomness comes from $\underline{\mathbf{x}}$ that is uniformly distributed over $\mathcal{C}$ and $(\underline{\mathbf{z}},\underline{\mathbf{t}})$ that is uniformly distributed over $\mathcal{P}_{\underline{\mathbf{z}},\underline{\mathbf{t}}|\underline{\mathbf{x}}}$. To upper bound the above probability, let us define the joint distribution $Q_{\underline{\mathbf{x}},\underline{\mathbf{z}},\underline{\mathbf{t}}} \in \Delta(\mathcal{X}^n \times \mathcal{Z}^n \times \mathcal{T}^n)$ as

$$Q_{\underline{\mathbf{x}},\underline{\mathbf{z}},\underline{\mathbf{t}}}(\underline{x},\underline{z},\underline{t}) \coloneqq \frac{1}{M}\mathbb{1}\{\underline{x} \in \mathcal{C}\}\frac{1}{\big|\mathcal{P}_{\underline{\mathbf{z}},\underline{\mathbf{t}}|\underline{x}}\big|}\mathbb{1}\big\{(\underline{z},\underline{t}) \in \mathcal{P}_{\underline{\mathbf{z}},\underline{\mathbf{t}}|\underline{x}}\big\}. \tag{84}$$

[2]Though $\mathcal{P}_{\underline{\mathbf{z}},\underline{\mathbf{t}}|\underline{x}}$ depends on $(P_{\mathbf{x}}, W_{\mathbf{z}|\mathbf{x}}, P_{\mathbf{t}|\mathbf{z}})$ and ε, we drop them in the notation.

Now the probability in Equation (83) can be written as

$$\begin{aligned}
\Pr_{\underline{\mathbf{x}},\underline{\mathbf{z}},\underline{\mathbf{t}}}[\underline{\mathbf{t}} \in \mathcal{F}'] &= \sum_{\underline{x}\in\mathcal{X}^n}\sum_{\underline{z}\in\mathcal{Z}^n}\sum_{\underline{t}\in\mathcal{F}'} Q_{\underline{\mathbf{x}},\underline{\mathbf{z}},\underline{\mathbf{t}}}(\underline{x},\underline{z},\underline{t}) \\
&= \frac{1}{M}\sum_{i=1}^{M}\sum_{\underline{z}\in\mathcal{Z}^n}\sum_{\underline{t}\in\mathcal{F}'}\frac{1}{\left|\mathcal{P}_{\underline{\mathbf{z}},\underline{\mathbf{t}}|\underline{x}_i}\right|}\mathbb{1}\left\{(\underline{z},\underline{t})\in\mathcal{P}_{\underline{\mathbf{z}},\underline{\mathbf{t}}|\underline{x}_i}\right\} \qquad (85)\\
&\le \frac{1}{M}\sum_{i=1}^{M}\sum_{\underline{t}\in\mathcal{F}'}\frac{1}{\left|\mathcal{P}_{\underline{\mathbf{z}},\underline{\mathbf{t}}|\underline{x}_i}\right|}\mathbb{1}\left\{\underline{t}\in\mathcal{P}_{\underline{\mathbf{t}}|\underline{x}_i}\right\}\sum_{\underline{z}\in\mathcal{Z}^n}\mathbb{1}\left\{\underline{z}\in\mathcal{P}_{\underline{\mathbf{z}}|\underline{x}_i}\right\} \qquad (86)\\
&= \frac{1}{M}\sum_{\underline{t}\in\mathcal{F}'}\sum_{i=1}^{M}\frac{\left|\mathcal{P}_{\underline{\mathbf{z}}|\underline{x}_i}\right|}{\left|\mathcal{P}_{\underline{\mathbf{z}},\underline{\mathbf{t}}|\underline{x}_i}\right|}\mathbb{1}\left\{\underline{t}\in\mathcal{P}_{\underline{\mathbf{t}}|\underline{x}_i}\right\} \\
&\le \frac{1}{M}\sum_{\underline{t}\in\mathcal{F}'}\frac{2^{n(H(\mathbf{z}|\mathbf{x})+\overline{\varepsilon})}}{2^{n(H(\mathbf{z}|\mathbf{x})+H(\mathbf{t}|\mathbf{z})+\varepsilon_*)}}\sum_{i=1}^{N}\mathbb{1}\left\{\underline{t}\in\mathcal{P}_{\underline{\mathbf{t}}|\underline{x}_i}\right\} \qquad (87)\\
&\le \frac{1}{M}\cdot|\mathcal{F}'|\cdot 2^{n(-H(\mathbf{t}|\mathbf{z})+\overline{\varepsilon}-\varepsilon_*)}2^{n\varepsilon} \qquad (88)\\
&\le 2^{-nR}\cdot|\mathcal{F}|\cdot 2^{-n(H(\mathbf{t}|\mathbf{z})-\overline{\varepsilon}+\varepsilon_*)}2^{n\varepsilon} \\
&\le 2^{n(-I(\mathbf{t};\mathbf{z})-\delta+H(\mathbf{t})+\varepsilon_1-H(\mathbf{t}|\mathbf{z})-\overline{\varepsilon}+\varepsilon_*+\varepsilon)} \qquad (89)\\
&= 2^{n(\varepsilon_1-\overline{\varepsilon}+\varepsilon_*+\varepsilon-\delta)}.
\end{aligned}$$

Equation (85) uses the definition of $Q_{\underline{\mathbf{x}},\underline{\mathbf{z}},\underline{\mathbf{t}}}$ (Equation (84)). Equation (86) follows since joint typicality implies marginal typicality (cf. Equation (79)). Equation (87) uses bounds Equations (74) and (77) on typical sets. Equation (88) follows from the definition of $\mathcal{F}'$ (Equation (82)). Equation (89) uses the bound on $|\mathcal{F}|$ (Equation (81)) and the value of R. Since $\varepsilon_1, \overline{\varepsilon}, \varepsilon_*$ all depend on ε, by taking a sufficiently large δ compared with ε and λ, we can make $\varepsilon_1 - \overline{\varepsilon} + \varepsilon_* + \varepsilon - \delta \le -\delta/2$. Then we have $\Pr[\underline{\mathbf{t}} \in \mathcal{F} \setminus \mathcal{F}'] \ge 1 - 2^{-n\delta/2}$. This finishes the proof. ∎

C. Elias–Bassalygo bound for myopic AVCs

In fact, we will prove the following theorem which implies Theorem 28 upon taking $\delta \to 0$.

Theorem 41. *Let $(\mathcal{X}, \mathcal{Z}, \mathcal{S}, \mathcal{Y}, \lambda_{\mathbf{x}}, \lambda_{\mathbf{s}}, W_{\mathbf{z}|\mathbf{x}}, W_{\mathbf{y}|\mathbf{x},\mathbf{s}})$ be a myopic AVC. Let $\mathcal{C} \subset \mathcal{X}^n$ be an arbitrary deterministic code such that*

1) *$P_{\underline{x}} \in \lambda_{\mathbf{x}}$ for every $\underline{x} \in \mathcal{C}$;*
2) *$R(\mathcal{C}) \ge C_{\mathrm{EB},\delta}$.*

Then the average probability of error of $\mathcal{C}$ is at least $\delta' > 0$ for some $\delta' \xrightarrow{\delta\to 0} 0$. The quantity $C_{\mathrm{EB},\delta}$ is defined as follows

$$C_{\mathrm{EB},\delta} \coloneqq \max_{P_{\mathbf{x}}\in\lambda_{\mathbf{x}}}\min_{P_{\mathbf{t}|\mathbf{z}}\in\Delta(\mathcal{T}|\mathcal{Z})} I(\mathbf{t};\mathbf{z}) + \delta, \qquad (90)$$

where the minimization is over the cardinality of $\mathcal{T}$ and $P_{\mathbf{t}|\mathbf{z}} \in \Delta(\mathcal{T}|\mathcal{Z})$ such that for every $t \in \mathcal{T}$, every $P_{\mathbf{x},\mathbf{x}'|\mathbf{t}=t} \in \mathcal{CP}(P_{\mathbf{x}|\mathbf{t}=t})$ and every decomposition $(P_{\mathbf{v}|\mathbf{t}=t}, P_{\mathbf{x}|\mathbf{v},\mathbf{t}=t}) \in \Delta(\mathcal{T}) \times \Delta(\mathcal{X}|\mathcal{T})$ (let $\mathcal{T} := [K]$ for some $K \in \mathbf{Z}_{\ge 1}$) satisfying $P_{\mathbf{x},\mathbf{x}'|\mathbf{t}=t} = \left[P_{\mathbf{v}|\mathbf{t}=t}P^{\otimes 2}_{\mathbf{x}|\mathbf{v},\mathbf{t}=t}\right]_{\mathbf{x},\mathbf{x}'}$, there exists $V_{\mathbf{s}|\mathbf{z},\mathbf{x}',\mathbf{t}=t} \in \Delta(\mathcal{S}|\mathcal{Z}\times\mathcal{X})$ satisfying the following two properties:

1) *for every $(t, x, x', y) \in \mathcal{T}\times\mathcal{X}^2\times\mathcal{Y}$,*

$$\sum_{z\in\mathcal{Z}}\sum_{s\in\mathcal{S}} W_{\mathbf{z}|\mathbf{x}}(z|x)V_{\mathbf{s}|\mathbf{z},\mathbf{x}',\mathbf{t}=t}(s|z,x')W_{\mathbf{y}|\mathbf{x},\mathbf{s}}(y|x,s) = \sum_{z\in\mathcal{Z}}\sum_{s\in\mathcal{S}} W_{\mathbf{z}|\mathbf{x}}(z|x')V_{\mathbf{s}|\mathbf{z},\mathbf{x}',\mathbf{t}=t}(s|z,x)W_{\mathbf{y}|\mathbf{x},\mathbf{s}}(y|x',s);$$

2) $\left[P_{\mathbf{t}}P_{\mathbf{v}|\mathbf{t}}P^{\otimes 2}_{\mathbf{x}|\mathbf{v},\mathbf{t}}W_{\mathbf{z}|\mathbf{x}}V_{\mathbf{s}|\mathbf{z},\mathbf{x}',\mathbf{t}}\right]_{\mathbf{s}} \in \mathrm{int}_\delta(\lambda_{\mathbf{s}})$.

Proof: We first pass to a large approximate constant composition subcode. Let $\mathcal{C}' \subset \mathcal{C}$ be a subcode such that for some constant $\lambda > 0$, we have

1) $|\mathcal{C}'| \ge c_1|\mathcal{C}|$ for a constant $c_1 > 0$ that depends on λ and $|\mathcal{X}|$;
2) there exists $P_{\mathbf{x}} \in \lambda_{\mathbf{x}}$ such that $\left\|P_{\underline{x}} - P_{\mathbf{x}}\right\|_\infty \le \lambda$ for every $\underline{x} \in \mathcal{C}'$.

The slack factor λ will be taken to be a sufficiently small function of δ. Suppose a uniformly random codeword $\underline{\mathbf{x}} \in \mathcal{C}$ was transmitted by Alice and James observes $\underline{\mathbf{z}}$ through the channel $W_{\mathbf{z}|\mathbf{x}}$. With probability at least c_1, the transmitted codeword $\underline{\mathbf{x}}$ comes from $\mathcal{C}'$. It therefore suffices for James to attack only $\mathcal{C}'$ instead of the whole codebook $\mathcal{C}$. We will design a jamming strategy for James such that the average probability of error of $\mathcal{C}'$ is at least a constant under this attack.

Let $\mathcal{T}$ be a certain finite alphabet and $P_{\mathbf{t}|\mathbf{z}} \in \Delta(\mathcal{T}|\mathcal{Z})$ be an auxiliary channel, both to be optimized. Pass $\underline{\mathbf{z}}$ through the auxiliary channel $P_{\mathbf{t}|\mathbf{z}}$ and obtain a sequence $\underline{\mathbf{t}} \in \mathcal{T}^n$.

By a standard typicality argument, we have

$$\Pr\left[\left\|P_{\underline{\mathbf{x}},\underline{\mathbf{z}},\underline{\mathbf{t}}} - P_{\mathbf{x}}W_{\mathbf{z}|\mathbf{x}}P_{\mathbf{t}|\mathbf{z}}\right\|_\infty \le \varepsilon\right] \ge 1 - e^{-\Omega(n\varepsilon^2)}, \tag{91}$$

where the probability is taken over $\underline{\mathbf{x}} \sim \mathrm{Unif}(\mathcal{C})$ and the Markov chain $\underline{\mathbf{x}} \xrightarrow{W_{\mathbf{z}|\mathbf{x}}} \underline{\mathbf{z}} \xrightarrow{P_{\mathbf{t}|\mathbf{z}}} \underline{\mathbf{t}}$. Define the joint distribution $P_{\underline{\mathbf{x}},\underline{\mathbf{z}},\underline{\mathbf{t}}} \in \Delta(\mathcal{X}^n \times \mathcal{Z}^n \times \mathcal{T}^n)$ as

$$P_{\underline{\mathbf{x}},\underline{\mathbf{z}},\underline{\mathbf{t}}}(\underline{x},\underline{z},\underline{t}) \coloneqq \frac{1}{M}\mathbb{1}\{\underline{x} \in \mathcal{C}\}\prod_{j=1}^n W_{\mathbf{z}|\mathbf{x}}(\underline{z}(j)|\underline{x}(j))P_{\mathbf{t}|\mathbf{z}}(\underline{t}(j)|\underline{z}(j)).$$

Conditioned on $(\underline{\mathbf{x}},\underline{\mathbf{z}})$ being jointly typical, the above distribution $P_{\underline{\mathbf{x}},\underline{\mathbf{z}},\underline{\mathbf{t}}}$ is close to the distribution $\mathrm{Unif}(\mathcal{C}) \times \mathrm{Unif}(\mathcal{P}_{\underline{\mathbf{z}},\underline{\mathbf{t}}|\underline{\mathbf{x}}})$. Here the conditional typical set $\mathcal{P}_{\underline{\mathbf{z}},\underline{\mathbf{t}}|\underline{\mathbf{x}}}$ was defined in Equation (73) and the conditional typicality is measured according to $P_{\mathbf{x}}W_{\mathbf{z}|\mathbf{x}}P_{\mathbf{t}|\mathbf{z}}$. More precisely, conditioned on the event in the probability in Equation (91), there exists a constant $\varepsilon_1 > 0$ depending on ε (and the alphabet sizes) such that for any $\underline{x} \in \mathcal{C}$ and $(\underline{z},\underline{t}) \in \mathcal{P}_{\underline{\mathbf{z}},\underline{\mathbf{t}}|\underline{x}}$,

$$1 - \varepsilon_1 \le P_{\underline{\mathbf{x}},\underline{\mathbf{t}},\underline{\mathbf{z}}}(\underline{x},\underline{z},\underline{t}) \Big/ \left(\frac{1}{M}\frac{1}{|\mathcal{P}_{\underline{\mathbf{t}},\underline{\mathbf{z}}|\underline{x}}|}\right) \le 1 + \varepsilon_1.$$

This implies

$$\sum_{\underline{t}\in\mathcal{F}'} P_{\underline{\mathbf{t}}}(\underline{t}) \le \sum_{\underline{t}\in\mathcal{F}'} Q_{\underline{\mathbf{t}}}(\underline{t})(1+\varepsilon_1) \le (1+\varepsilon_1)2^{-n\delta/2}. \tag{92}$$

Recall that $\mathcal{F}'$ was defined in Equation (82). By the notational convention, $P_{\underline{\mathbf{t}}}$ and $Q_{\underline{\mathbf{t}}}$ are the marginals of $P_{\underline{\mathbf{x}},\underline{\mathbf{z}},\underline{\mathbf{t}}}$ and $Q_{\underline{\mathbf{x}},\underline{\mathbf{z}},\underline{\mathbf{t}}}$, respectively, and the latter distribution $Q_{\underline{\mathbf{x}},\underline{\mathbf{z}},\underline{\mathbf{t}}}$ was defined in Equation (84). The second inequality in Equation (92) follows from Theorem 40. Therefore,

$$\Pr_{\underline{\mathbf{x}},\underline{\mathbf{z}},\underline{\mathbf{t}}}[\underline{\mathbf{t}} \in \mathcal{F} \setminus \mathcal{F}'] \ge 1 - (1+\varepsilon_1)2^{-n\delta/2},$$

where $\mathcal{F}$ was defined in Equation (80).

Conditioned on $\underline{\mathbf{t}} \in \mathcal{F} \setminus \mathcal{F}'$, there are at least $2^{n\varepsilon}$ codewords $\underline{x}_i$ from $\mathcal{C}$ that are covered by $\underline{\mathbf{t}}$ in the sense that $\underline{\mathbf{t}} \in \mathcal{P}_{\underline{\mathbf{t}}|\underline{x}_i}$. From now on, James restricts himself to the set of these codewords, i.e., the subcode $\mathcal{C}''$ defined as

$$\mathcal{C}'' \coloneqq \left\{\underline{x} \in \mathcal{C}' : \left\|P_{\underline{x},\underline{\mathbf{t}}} - \left[P_{\mathbf{x}}W_{\mathbf{z}|\mathbf{x}}P_{\mathbf{t}|\mathbf{z}}\right]_{\mathbf{x},\mathbf{t}}\right\|_\infty \le \varepsilon\right\}.$$

Note that $\mathcal{C}''$ depends on the realizations of $\underline{\mathbf{t}}$ and therefore is a random subcode. However, we do not make this dependency explicit in the notation. Given that $(\underline{\mathbf{x}},\underline{\mathbf{z}},\underline{\mathbf{t}})$ are jointly typical and $\underline{\mathbf{t}} \in \mathcal{F} \setminus \mathcal{F}'$, we have $|\mathcal{C}''| > 2^{n\varepsilon}$.

Now James uniformly samples a spoofing codeword $\underline{\mathbf{x}}'$ from $\mathcal{C}''$. With probability at least $1 - 2^{-n\varepsilon}$, the spoofing codeword $\underline{\mathbf{x}}'$ is different from Alice's transmitted codeword $\underline{\mathbf{x}}$. Since $|\mathcal{C}''|$ is sufficiently large (in our case at least exponentially large), it can be shown that with at least a constant $c_2 > 0$ probability (where c_2 depends on δ), a uniformly random pair $(\underline{\mathbf{x}},\underline{\mathbf{x}}')$ of (distinct) codewords from $\mathcal{C}''$ satisfies that for every $t \in \mathcal{T}$,

$$\left\|P_{\underline{\mathbf{x}},\underline{\mathbf{x}}'|\underline{\mathbf{t}}(\cdot)=t} - P_{\mathbf{x},\mathbf{x}'|\mathbf{t}=t}\right\|_\infty \le \delta/10, \tag{93}$$

for some $P_{\mathbf{x},\mathbf{x}'|\mathbf{t}} \in \Delta(\mathcal{X}^2|\mathcal{T})$ of the following form:

$$P_{\mathbf{x},\mathbf{x}'|\mathbf{t}}(x,x'|\mathbf{t}=t) = \sum_{v=1}^K P_{\mathbf{v}|\mathbf{t}=t}(v)P_{\mathbf{x}|\mathbf{v},\mathbf{t}=t}(x|v)P_{\mathbf{x}|\mathbf{v},\mathbf{t}=t}(x'|v),$$

where $K \in \mathbf{Z}_{\ge 1}$ is a constant, $P_{\mathbf{v}|\mathbf{t}} \in \Delta(\mathcal{T}|\mathcal{T})$ and $P_{\mathbf{x}|\mathbf{v},\mathbf{t}} \in \Delta(\mathcal{X}|\mathcal{T}\times\mathcal{T})$. This is a slight generalization of Equation (16) which was proved in [BDJ+20], [ZJB21]. The only difference is that in Equation (93) both distributions (or type) are conditioned on a time-sharing variable $\mathbf{t}$ (or a time-sharing sequence $\underline{\mathbf{t}}$). However, the proof follows verbatim the one for Equation (17) and we do not repeat it here.

Since $R \ge C_{\mathrm{EB},\delta}$ and by the definition of $C_{\mathrm{EB},\delta}$ (Equation (90)), for each $t \in \mathcal{T}$, there must exist a distribution $V_{\mathbf{s}|\mathbf{z},\mathbf{x}',\mathbf{t}=t} \in \Delta(\mathcal{S}|\mathcal{Z}\times\mathcal{X})$ that symmetrizes $P_{\mathbf{x},\mathbf{x}'|\mathbf{t}=t}$, i.e., satisfying Item 1 in Theorem 41. Furthermore, $V_{\mathbf{s}|\mathbf{z},\mathbf{x}',\mathbf{t}}$ satisfies the power constraint of James, i.e., Item 2 in Theorem 41. James picks such a distribution $V_{\mathbf{s}|\mathbf{z},\mathbf{x}',\mathbf{t}}$ and generates the jamming sequence $\underline{\mathbf{s}} \in \mathcal{S}^n$ as follows. For any $1 \le i \le n$,

$$\underline{\mathbf{s}}(i) \sim V_{\mathbf{s}|\mathbf{z}=\underline{\mathbf{z}}(i),\mathbf{x}'=\underline{\mathbf{x}}'(i),\mathbf{t}=\underline{\mathbf{t}}(i)}.$$

Since the type $P_{\underline{\mathbf{t}},\underline{\mathbf{x}},\underline{\mathbf{x}}'}$ is approximately $P_{\mathbf{t}}P_{\mathbf{x},\mathbf{x}'|\mathbf{t}} = P_{\mathbf{t}}\left[P_{\mathbf{v}|\mathbf{t}}P^{\otimes 2}_{\mathbf{x}|\mathbf{v},\mathbf{t}}\right]_{\mathbf{x},\mathbf{x}'|\mathbf{t}}$ (cf. Equation (93)) w.r.t. which the jamming

distribution is δ-feasible, by the same type of concentration argument as in the proof of Theorem 31, the actual jamming cost satisfies the constraint with probability $1-2^{-\Omega(n)}$.

We then argue that the channel is indeed symmetrized under the above $\underline{\mathbf{s}}$. The proof is also similar to Theorem 31.

$$\begin{aligned}
\mathop{\mathbb{E}}_{\underline{\mathbf{s}}}\Big[W^{\otimes n}_{\mathbf{y}|\mathbf{x},\mathbf{s}}(\underline{y}|\underline{\mathbf{x}},\underline{\mathbf{s}})\Big] &= \sum_{\underline{s}\in\mathcal{S}}\sum_{\underline{z}\in\mathcal{Z}^n}\prod_{i=1}^{n}\big(W_{\mathbf{z}|\mathbf{x}}(\underline{z}(i)|\underline{\mathbf{x}}(i))V_{\mathbf{s}|\mathbf{z},\mathbf{x}',\mathbf{t}}(\underline{s}(i)|\underline{\mathbf{z}}(i),\underline{\mathbf{x}}'(i),\underline{\mathbf{t}}(i))W_{\mathbf{y}|\mathbf{x},\mathbf{s}}(\underline{y}(i)|\underline{\mathbf{x}}(i),\underline{s}(i))\big)\\
&= \prod_{i=1}^{n}\sum_{\underline{z}(i)\in\mathcal{Z}}\sum_{\underline{s}(i)\in\mathcal{S}}\big(W_{\mathbf{z}|\mathbf{x}}(\underline{z}(i)|\underline{\mathbf{x}}(i))V_{\mathbf{s}|\mathbf{z},\mathbf{x}',\mathbf{t}}(\underline{s}(i)|\underline{\mathbf{z}}(i),\underline{\mathbf{x}}'(i),\underline{\mathbf{t}}(i))W_{\mathbf{y}|\mathbf{x},\mathbf{s}}(\underline{y}(i)|\underline{\mathbf{x}}(i),\underline{s}(i))\big)\\
&= \prod_{i=1}^{n}\sum_{\underline{z}(i)\in\mathcal{Z}}\sum_{\underline{s}(i)\in\mathcal{S}}\big(W_{\mathbf{z}|\mathbf{x}}(\underline{z}(i)|\underline{\mathbf{x}}'(i))V_{\mathbf{s}|\mathbf{z},\mathbf{x}',\mathbf{t}}(\underline{s}(i)|\underline{\mathbf{z}}(i),\underline{\mathbf{x}}(i),\underline{\mathbf{t}}(i))W_{\mathbf{y}|\mathbf{x},\mathbf{s}}(\underline{y}(i)|\underline{\mathbf{x}}'(i),\underline{s}(i))\big) \qquad (94)\\
&= \mathop{\mathbb{E}}_{\underline{\mathbf{s}}'}\Big[W^{\otimes n}_{\mathbf{y}|\mathbf{x},\mathbf{s}}(\underline{y}|\underline{\mathbf{x}}',\underline{\mathbf{s}}')\Big],
\end{aligned}$$

where $\underline{\mathbf{s}}'$ has the following distribution

$$\underline{\mathbf{s}}'(i)\sim V_{\mathbf{s}|\mathbf{z}=\underline{\mathbf{z}}'(i),\mathbf{x}'=\underline{\mathbf{x}}(i),\mathbf{t}=\underline{\mathbf{t}}(i)} \quad \text{and} \quad \underline{\mathbf{z}}'(i)\sim W_{\mathbf{z}|\mathbf{x}=\underline{\mathbf{x}}'(i)}.$$

In Equation (94), we used the symmetrizability property of the jamming distribution (Item 1 in Theorem 41). This will imply that the average probability of error of $\mathcal{C}'$ is at least a constant and the proof is finished. ■

XI. Examples

Example 42 (Non state-deterministic and Omniscient bit flip AVC)**.** *A non state deterministic and omniscient bit flip AVC A-BSC(w,p,r) (fig. 8) is described via an input codeword $\underline{\mathbf{x}}\in\{0,1\}^n$ such that $wt_H(\underline{\mathbf{x}})\leq nw$, state sequence $\underline{\mathbf{s}}\in\{0,1\}^n$ such that $wt_H(\underline{\mathbf{s}})\leq np$,*[3] *output sequence $\underline{\mathbf{y}}\in\{0,1\}^n$ and the index-wise decomposable and stochastic channel law $W(.|.,.)$ such that $\forall i\in[n]$:*

$$W_{\mathbf{y}|\mathbf{x},\mathbf{s}}(y_i|x_i,s_i)=\begin{cases}1-r & \text{if } y_i=x_i\oplus s_i\\ r & \text{if } y_i=x_i\oplus s_i\oplus 1\end{cases}$$

Alternatively, a non-state deterministic and omniscient bit flip AVC A-BSC(w,p,r) can also be seen as the cascade of a state

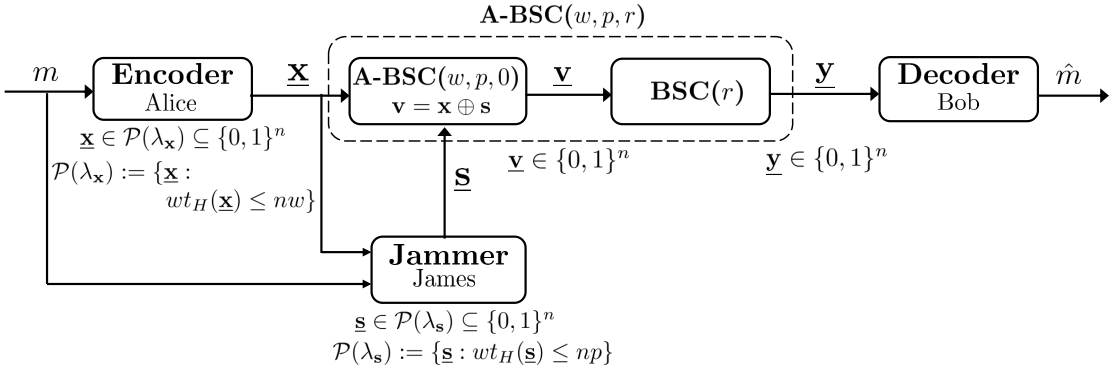


Fig. 8. Communication over a non state deterministic and bit-flip AVC A-BSC(w,p,r)

deterministic and omniscient bit-flip AVC A-BSC$(w,p,0)$ with a binary symmetric channel with crossover probability r i.e., $BSC(r)$.

Remark 43. *Note that one such example of input constraints and state constraints in binary domain is that of hamming weight constraints on the input codewords and state sequences respectively. In this case, the input constraint set $\lambda_{\mathbf{x}}$ is defined as $\lambda_{\mathbf{x}}:=\{P_{\underline{\mathbf{x}}}=(1-\alpha,\alpha):\alpha\leq w\}$. Similarly, the state constraint set $\lambda_{\mathbf{s}}$ can be defined as $\lambda_{\mathbf{s}}:=\{P_{\underline{\mathbf{s}}}=(1-\alpha,\alpha):\alpha\leq p\}$.*

[3]Note that $\underline{\mathbf{s}}$ can depend on everything that James knows in omniscient AVC model

A. Characterization of Rate Positivity

In the following analysis below, we restrict ourselves to the case of unconstrained inputs i.e., $w = \frac{1}{2}$. We proceed by presenting the converse followed by the achievability for the A-BSC$(\frac{1}{2}, p, r)$.

Converse: We already know that for the state-deterministic and omniscient bit-flip AVC A-BSC$(\frac{1}{2}, p, 0)$ (where the random noise $\underline{\mathbf{z}} = 0$ or $\mathbf{z}_i \sim Ber(0)$, $i = 1, 2, \cdots, n$), if $p > 1/4$, the capacity is zero through plotkin bound.
This immediately implies that whenever $p > 1/4$ and for any value of $0 \leq r < \frac{1}{2}$, the capacity of any non-state deterministic and omniscient bit-flip AVC A-BSC$(\frac{1}{2}, p, r)$ (where $0 \leq r < 1/2$ i.e., $\mathbf{z}_i \sim Ber(r)$ for $i = 1, 2, \cdots, n$) is also zero. This completes the proof of the converse.

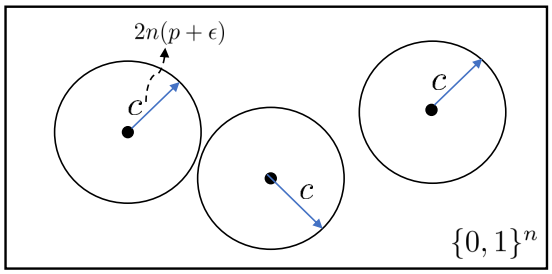
Achievability: We first present the code construction based on the GV bound; this is followed by the description of the encoder and decoder. Further, we conclude by presenting the error analysis and showing that the error probability is negligible in block length n.

Lemma 44. *For $p < \frac{1}{4}$ and any value of $0 \leq r < \frac{1}{2}$, there exists a code $\mathcal{C} := \{\underline{x}_i\}_{i=1}^{2^{nR}}$ consisting of 2^{nR} codewords, with rate $R > 0$ such that such that for every distinct pair of codewords $\underline{x}_i, \underline{x}_j \in \mathcal{C}$, we have*

$$d_H(\underline{x}_i, \underline{x}_j) = 2n(p + \epsilon) > 2pn \tag{95}$$

for some $\epsilon > 0$.

Proof: The proof of such code construction is based on the existence of the Gilbert-Varshamov (GV) type code. This can



be equivalently seen as a covering problem over the hamming space $\{0, 1\}^n$. Choose a codeword randomly $\underline{\mathbf{x}} = \underline{x}_1 \in \{0, 1\}^n$ and a hamming ball $\mathcal{B}_H$ of radius $c := 2n(p + \epsilon)$ around it i.e., $\mathcal{B}_H(\underline{x}_1, c)$. Similarly, choose other codewords $\underline{\mathbf{x}} = \underline{x}_i$ such that $i \in [2^n R]$ and cover the whole space with hamming balls $\mathcal{B}_H(\underline{x}_i, c)$ such that (95) is satisfied. Therefore, the number of hamming balls/codewords required to cover the whole space is given by:

$$\begin{aligned} \#\text{ codewords} &\geq \frac{\text{Vol }(\{0,1\}^n)}{\text{Vol }(\mathcal{B}_H(\underline{x}_i, c))} \\ 2^{nR} &\geq \frac{2^n}{2^{nH(2(p+\epsilon))+\epsilon'}} \\ R &\geq 1 - H(2p) - \epsilon'' \end{aligned}$$

where ϵ'' is a small positive constant. ■

Now, we analyse the probability of error $\mathbb{P}_e$ for the code above with rate R and show that for some optimal choice of R, the maximum error probability is negligible.

Encoder: Given the message m_i, let $\underline{x}_i$ be the transmitted codeword over the A-BSC$(\frac{1}{2}, p, r)$.

Decoder: The decoder is based on the minimum distance decoding rule. Thus, having observed the noisy sequence $\underline{\mathbf{y}} = \underline{y}$, the decoder $\phi(\underline{y})$ performs the following operation:

$$\phi(\underline{y}) = \begin{cases} m_i & \text{if } d_H(\underline{x}_i, \underline{y}) < d_H(\underline{x}_j, \underline{y}), \quad \forall j \neq i \\ 0 & \text{otherwise} \end{cases} \tag{96}$$

Probability of Error analysis: We first analyse the probability of error when there are exactly two codewords, say $\underline{x}_i$ and $\underline{x}_j$, in the codebook $\mathcal{C}$. Let $\underline{x}_i$ (corresponding to message m_i) be the transmitted codeword. Let $\underline{s}$ be James' attack noise sequence. The decoder observes the noisy sequence $\underline{y} = \underline{x}_i \oplus \underline{s} \oplus \underline{\mathbf{z}}$. Then, the probability of error for message m_i is given by

$$\begin{aligned}
\mathbb{P}_e(m_i, \underline{\mathbf{s}}) &:= \mathbb{P}_W\left(d_H(\underline{x}_i, \underline{y}) \geq d_H(\underline{x}_j, \underline{y})\right) \\
&\overset{(a)}{=} \mathbb{P}_W\left(d_H(\underline{x}_i, \underline{x}_i \oplus \underline{s} \oplus \underline{\mathbf{z}}) \geq d_H(\underline{x}_j, \underline{x}_i \oplus \underline{s} \oplus \underline{\mathbf{z}})\right) \\
&= \mathbb{P}_W\left(d_H(\underline{0}, \underline{s} \oplus \underline{\mathbf{z}}) \geq d_H(\underline{x}_j \oplus \underline{x}_i, \underline{s} \oplus \underline{\mathbf{z}})\right) \\
&\overset{(b)}{=} \mathbb{P}_W\left(wt_H(\underline{s} \oplus \underline{\mathbf{z}}) \geq wt_H(\underline{x}_j \oplus \underline{x}_i) + wt_H(\underline{s} \oplus \underline{\mathbf{z}}) - 2\left\langle \underline{x}_j \oplus \underline{x}_i, \underline{s} \oplus \underline{\mathbf{z}} \right\rangle\right) \\
&= \mathbb{P}_W\left(2\left\langle \underline{x}_j \oplus \underline{x}_i, \underline{s} \oplus \underline{\mathbf{z}} \right\rangle \geq wt_H(\underline{x}_j \oplus \underline{x}_i)\right) \\
&\overset{(c)}{=} \mathbb{P}_W\left(2\left\langle \underline{x}', \underline{s} \oplus \underline{\mathbf{z}} \right\rangle \leq wt_H(\underline{x}')\right) \\
&\overset{(d)}{=} \mathbb{P}_W\left(\left\langle \underline{x}', \underline{\mathbf{z}}' \right\rangle \geq \frac{wt_H(\underline{x}')}{2}\right) \\
&\overset{(e)}{=} \mathbb{P}_W\left(\left\langle \underline{x}', \underline{\mathbf{z}}' \right\rangle \geq n(p+\epsilon)\right)
\end{aligned} \tag{97}$$

Here

(a) follows from noting that $\underline{y} = \underline{x}_i \oplus \underline{s} \oplus \underline{\mathbf{z}}$.

(b) follows from noting that given any two sequences $\underline{u}, \underline{v} \in \{0,1\}^n$, we have $d_H(\underline{u}, \underline{v}) = wt_H(\underline{u}) + wt_H(\underline{v}) - 2\left\langle \underline{u}, \underline{v} \right\rangle$. Here $\left\langle \underline{u}, \underline{v} \right\rangle := \sum_{i=1}^n \mathbf{1}_{\{u_i = v_i = 1\}}$ determines the total number of locations where $\underline{u}$ and $\underline{v}$ are identically equal to 1.

(c) follows by defining $\underline{x}' := \underline{x}_j \oplus \underline{x}_i$.

(d) follows from defining $\underline{\mathbf{z}}' := \underline{s} \oplus \underline{\mathbf{z}}$. Note that $\underline{\mathbf{z}}'$ is a vector comprising n independent but not identically distributed Bernoulli random variables. In particular, $\underline{\mathbf{z}}'_i \sim Ber(s_i * r)$; here $a * b$ denotes convolution where $a * b := a(1-b) + (1-a)b$ for $a, b \in [0,1]$.

(e) follows from noting that the minimum hamming weight of any codeword is the minimum distance of the code.

Towards establishing an upper bound on the probability in (97), we first derive upper and lower bounds on $\mathbb{E}[\langle \underline{x}', \underline{\mathbf{z}}' \rangle]$ in the following claim.

Claim 45. *For some p, r such that $0 \leq p < 1/4$ and $0 \leq r < 1/2$, we have*

$$2(p+\epsilon)rn \leq \mathbb{E}[\langle \underline{x}', \underline{\mathbf{z}}' \rangle] \leq (p + 2\epsilon r)n \tag{98}$$

for some $\epsilon > 0$ as in Lemma 44.

Proof of Claim:

$$\begin{aligned}
\mathbb{E}[\langle \underline{x}', \underline{\mathbf{z}}' \rangle] &= \sum_{i=1}^n \mathbb{E}[x'_i \mathbf{z}'_i] \\
&= \sum_{i=1}^n x'_i \mathbb{E}[\mathbf{z}'_i] \\
&= \sum_{i \in \text{supp}(\underline{x}')} \mathbb{E}[\mathbf{z}'_i] \\
&\overset{(a)}{=} \sum_{i \in \text{supp}(\underline{x}')} s_i * r \\
&= \sum_{i \in \text{supp}(\underline{x}')} (r + s_i - 2rs_i) \\
&= \sum_{i \in \text{supp}(\underline{x}')} r + \sum_{i \in \text{supp}(\underline{x}')} s_i - \sum_{i \in \text{supp}(\underline{x}')} 2rs_i \\
&= wt_H(\underline{x}')r + (1-2r)\left(\sum_{i \in \text{supp}(\underline{x}')} s_i\right) \\
&\overset{(b)}{=} 2n(p+\epsilon)r + (1-2r)\left(\sum_{i \in \text{supp}(\underline{x}')} s_i\right)
\end{aligned} \tag{99}$$

Here

(a) follows from the fact that $\mathbf{z}'_i \sim Ber(s_i * q)$

(b) follows from noting that the minimum hamming weight of any codeword is the minimum distance of the code.

The lower bound in the Claim follows by letting $\underline{s} = \underline{0}$ (observe that the second term in the R.H.S of (99) is positive as $r < 1/2$). In particular, on substituting $\underline{s} = \underline{0}$ in (99) we get

$$\begin{aligned}\mathbb{E}[\langle \underline{x}', \underline{\mathbf{z}}' \rangle] &\geq 2(p+\epsilon)nr + (1-2r)\cdot 0 \\ &= 2(p+\epsilon)nr \end{aligned} \tag{100}$$

which corresponds to the lower bound in Claim 45. Now, for the upper bound we simplify (99) as follows

$$\begin{aligned}\mathbb{E}[\langle \underline{x}', \underline{\mathbf{z}}' \rangle] &= 2n(p+\epsilon)r + (1-2r)\left(\sum_{i \in \text{supp}(\underline{x}')} s_i\right) \\ &\leq 2n(p+\epsilon)r + (1-2r)\min\{wt_H(\underline{x}'), wt_H(\underline{s})\} \\ &\overset{(a)}{\leq} 2(p+\epsilon)nr + (1-2r)wt_H(\underline{s}) \\ &\overset{(b)}{\leq} 2(p+\epsilon)nr + (1-2r)pn \\ &= (p+2\epsilon r)n \end{aligned} \tag{101}$$

Here,

(a) follows from noting that $\min\{wt_H(\underline{x}'), wt_H(\underline{s})\} = wt_H(\underline{s})$ as $p < 1/4$.

(b) follows from noting that $wt_H(\underline{s}) \leq np$.

Therefore, from (100) and (101) we have the following

$$2(p+\epsilon)rn \leq \mathbb{E}[\langle \underline{x}', \underline{\mathbf{z}}' \rangle] \leq (p+2\epsilon r)n$$

This completes the proof of the claim. ■

Now, we state an immediate corollary below which directly follows from the above claim.

Corollary 46. *Let p, r and ϵ be given as in Claim 45. Then, we have*

$$\mathbb{E}[\langle \underline{x}', \underline{\mathbf{z}}' \rangle] < (p+\epsilon)n \tag{102}$$

The proof of this corollary follows directly from the Claim 45 by further noting that $0 \leq r < 1/2$ which implies $2r < 1$. We now use the Chernoff bound (Lemma 3) for the sum of independent but not identically distributed Bernoulli random variables to evaluate the probability in (97).

$$\begin{aligned}\mathbb{P}_e(m_i, \underline{s}) &\overset{(a)}{\leq} \mathbb{P}_W\left(\langle \underline{x}', \underline{\mathbf{z}}' \rangle \geq (1+\delta)\mathbb{E}[\langle \underline{x}', \underline{\mathbf{z}}' \rangle]\right) \\ &\overset{(b)}{\leq} e^{-\frac{\delta^2}{2+\delta}\mathbb{E}[\langle \underline{x}', \underline{\mathbf{z}}' \rangle]} \\ &\overset{(c)}{\leq} e^{-\frac{\delta^2}{2+\delta}2(p+\epsilon)rn} \end{aligned} \tag{103}$$

Here,

(a) follows from Corollary 46. In particular, we pick $\delta(\epsilon) > 0$ small enough such that $(1+\delta)\mathbb{E}[\langle \underline{x}', \underline{\mathbf{z}} \rangle] \leq n(p+\epsilon)$; it follows from Claim 45 that such a $\delta(\epsilon) > 0$ exists.

(b) follows from Lemma 3.

(c) follows from the lower bound in Claim 45.

From (103), it follows that $\mathbb{P}_e(m_i, \underline{s}) \to 0$ as $n \to \infty$ *uniformly* for every $\underline{s}$, where $wt_H(\underline{s}) \leq np$. Hence, we have

$$\mathbb{P}_e(m_i) = e^{-\frac{\delta^2}{2+\delta}2(p+\epsilon)rn} \tag{104}$$

Thus, the maximum probability of error for the above two codeword code $\mathcal{C}$ is vanishing as $n \to \infty$. Now, the maximum probability of error for the code $\mathcal{C} = \{\underline{x}_i\}_{i=1}^{2^{nR}}$ as in Lemma 44 is given by

$$\begin{aligned}\mathbb{P}_e(\mathcal{C}) &:= \max_{m_i} \max_{\underline{s}: wt_H(\underline{s}) \leq np} \mathbb{P}_W\left(\exists\ j \neq i \text{ for } i, j \in [2^{nR}] : d_H(\underline{x}_i, \underline{y}) \geq d_H(\underline{x}_j, \underline{y})\right) \\ &\leq \binom{2^{nR}}{2} \mathbb{P}_W\left(d_H(\underline{x}_i, \underline{y}) \geq d_H(\underline{x}_j, \underline{y})\right) \\ &\leq 2^{2nR} e^{-\frac{\delta^2}{2+\delta}2(p+\epsilon)rn} \\ &\leq e^{n\left(2R - \frac{\delta^2}{2+\delta}2(p+\epsilon)r\right)} \end{aligned} \tag{105}$$

Now, we choose $R(\epsilon,\delta) > 0$ such that $R < \frac{\delta^2}{2+\delta}(p+\epsilon)r$ (note that such a choice for R exists). Therefore, it follows that the maximum probability of error for the code $\mathcal{C} = \{\underline{x}_i\}_{i=1}^{2^{nR}}$ given by $\mathbb{P}_e(\mathcal{C})$ is negligible in the block length n for some positive but small rate R.

This completes the achievability analysis for a positive rate for a A-BSC$(\frac{1}{2}, p, r)$.

B. Lower Bounds, Upper Bounds, and the Capacity Characterization

Lower Bounds

An A-BSC$(\frac{1}{2}, p, r)$ is a special case of a general alphabet non state deterministic and omniscient AVC $\mathcal{A} := (\lambda_{\mathbf{x}}, \lambda_{\mathbf{s}}, W_{\mathbf{y}|\mathbf{x},\mathbf{s}})$. Therefore, on specializing the achievability result in the Theorem 24 for the A-BSC$(\frac{1}{2}, p, r)$, we obtain the following lower bound:

$$R \geq \min\{1 - H(2p), 1 - H(p * r)\} - \delta' \tag{106}$$

for some arbitrarily small constant $\delta' > 0$.

Below, we present the plot for the lower bounds for a A-BSC$(\frac{1}{2}, p, r)$:

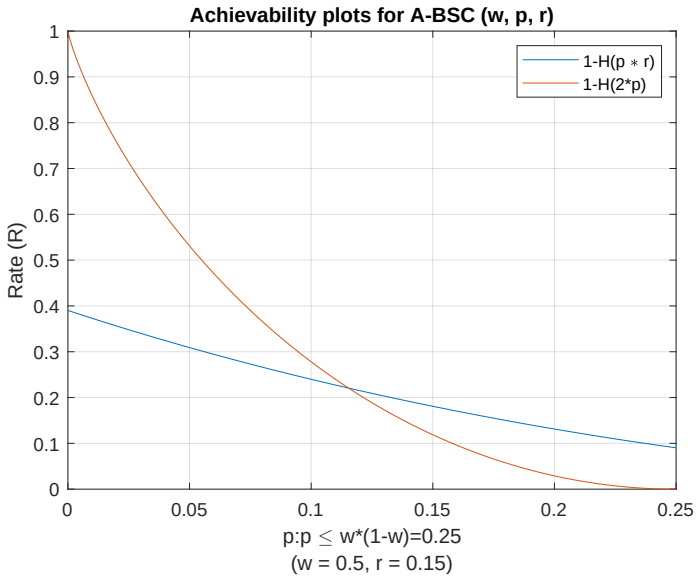


Fig. 9. Achievability plots for A-BSC$(\frac{1}{2}, p, r)$ for $r = 0.15$

Upper Bounds

The best upper bound for the non state deterministic and omniscient bit flip AVC A-BSC$(\frac{1}{2}, p, r)$ is given by analysing the MRRW2 bound and the DMC type (assuming that James attack via a stochastic noise sequence i.e., Ber(p) noise for the case of A-BSC$(\frac{1}{2}, p, r)$) bound. For completeness, we state and plot all the rate upper bounds for the A-BSC$(\frac{1}{2}, p, r)$:

Elias-Bassalygo (EB) Bound:

$$R \leq 1 - H\left(\frac{1}{2} - \frac{\sqrt{1-4p}}{2}\right) \tag{107}$$

MRRW 2 Bound:

$$R \leq H\left(\frac{1}{2}\left(1 - \sqrt{1 - (\sqrt{1-2p+4p^2} - 2p)^2}\right)\right) \tag{108}$$

DMC type Bound:

$$R \leq 1 - H\left(p * r\right) \tag{109}$$

From Equations (107), (108), (109) and the Figure 10, it can be easily seen that the best upper bound for a A-BSC$(\frac{1}{2}, p, r)$

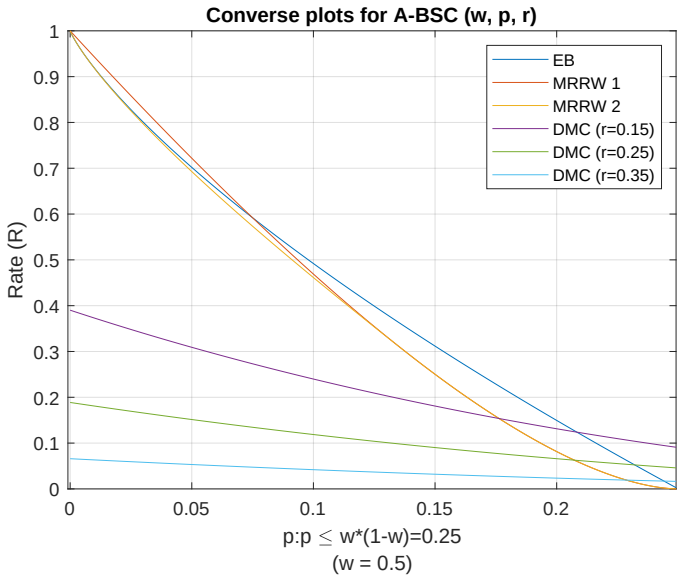


Fig. 10. Converse plots for A-BSC$(\frac{1}{2}, p, r)$ for $r = 0.15, 0.25, 0.35$

is given by the MRRW2 and DMC type bound:

$$R \leq \min\left\{H\left(\frac{1}{2}\left(1 - \sqrt{1 - (\sqrt{1-2p+4p^2} - 2p)^2}\right)\right), 1 - H\left(p * r\right)\right\} \tag{110}$$

Capacity Characterization

From the lower bound in equation (106) and the upper bound in equation (110) for a A-BSC$(\frac{1}{2}, p, r)$, it can inferred that the capacity of a A-BSC$(\frac{1}{2}, p, r)$ is well-defined for $p(r) \leq \frac{r}{1+2r}$. However, the upper bound and lower bound do not match for $p > \frac{r}{1+2r}$. Therefore, the capacity of A-BSC$(\frac{1}{2}, p, r)$ is

$$\mathbb{C}_{A-BSC} = \begin{cases} 1 - H(p * r) & \text{if } p(r) \leq p^* \\ \text{not defined} & \text{if } p(r) > p^* \end{cases}$$

where $p^* = \frac{r}{1+2r}$.

The figure below (fig. 11) shows the achievability and converse plots for a A-BSC$(\frac{1}{2}, p, r)$ for a fixed $r \in [0, \frac{1}{2})$.

Example 47 (State-deterministic and Myopic bit erasure AVC)**.** *A state-deterministic and myopic bit flip AVC A_m-BSC(w, q, p) (see fig. 12) is described via an input codeword $\underline{\mathbf{x}} \in \{0,1\}^n$ such that $wt_H(\underline{\mathbf{x}}) \leq nw$,[4] Alice to James channel BSC(q), James' input sequence $\underline{\mathbf{z}} \in \{0,1\}^n$, state sequence $\underline{\mathbf{s}} \in \{0,1\}^n$, output sequence $\underline{\mathbf{y}} \in \{0,1\}^n$ and the index-wise decomposable and deterministic channel law $W(.|.,.)$ such that $\forall i \in [n]$:*

$$W_{\mathbf{y}|\mathbf{x},\mathbf{s}}(y_i|x_i, s_i) = \begin{cases} 1 & \text{if } y_i = x_i \oplus s_i \\ 0 & \text{if } y_i = x_i \oplus s_i \oplus 1 \end{cases}$$

[4]Specifically, in this work we assume unconstrained input sequences i.e., $w = 1/2$.

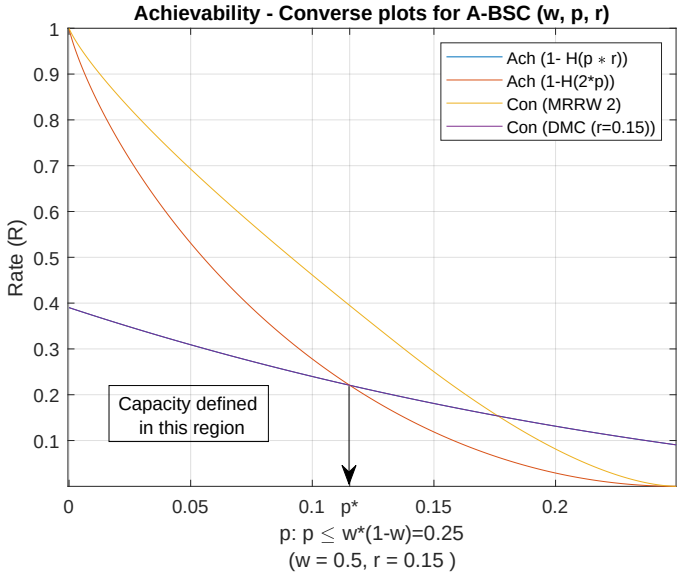


Fig. 11. Achievability and Converse plots for A-BSC$(\frac{1}{2}, p, r)$ for $r = 0.15$

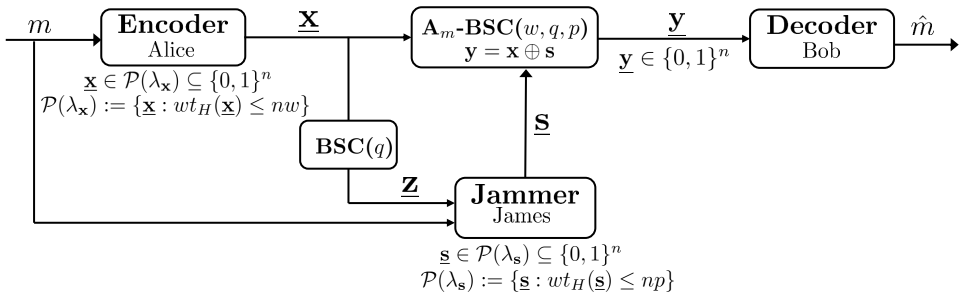


Fig. 12. Communication over a state-deterministic and myopic bit flip AVC A_m-BSC(w, q, p).

C. Elias-Bassalygo Bound

Corollary 48. *Theorem 41 implies that for a state-deterministic and myopic bit flip AVC A_m-BSC(w, q, p) whenever $q \leq p \leq \frac{1}{4(1-q)}$, the capacity of this channel under deterministic encoding is at most $C_{\text{EB}} := \max_{P_{\mathbf{x}} \in \lambda_{\mathbf{x}}} \min_{P_{\mathbf{t}|\mathbf{z}} \in \Delta(\mathcal{T}|\mathcal{Z})} I(\mathbf{t}; \mathbf{z}) = 1 - H(w(p, q)) = 1 - H(\frac{(q+1) - \sqrt{\frac{1-q}{1-2q}}\sqrt{-4p+2q^2+q+1}}{2})$.*

Proof: In the binary case, the covering variable $\mathbf{t}$ is chosen by passing $\mathbf{z}$ through a BSC(w(p,q)), where the w(p,q) is the generalized binary Plotkin bound and will be further derived.
Recall the definition of Myopic CP-confusability $\mathcal{KCP}^{m}_{strong}(P_{\mathbf{x}})$ in definition 19. If we focus on the bit-flip bit-flip case, same as Figure 12, confusability can be formulated as a linear programming problem that James the jammer, tries to confuse by using as low power as possible. We will show that the confusability condition can be formulated as a linear constraint, required power will be also a linear function and therefore we can formulate the weight constrained confusability as a linear programming problem.
To specify James's attack in the above definition, we need to specify $V_{\mathbf{s}|\mathbf{z},\mathbf{x}'}$ (t can be omitted) for each $\mathbf{s}, \mathbf{z}, \mathbf{x}' \in \{0, 1\}$. Note that $\forall \mathbf{z}, \mathbf{x}' : V_{0|\mathbf{z},\mathbf{x}'} + V_{1|\mathbf{z},\mathbf{x}'} = 1$, thus it's sufficient to find $V_{1|00}, V_{1|01}, V_{1|10}$, and $V_{1|11}$. Which are later denoted by x_{00}, x_{01}, x_{10}, and x_{11} respectively.
One can easily check that the equation in definition 19, will be vacuous whenever $\mathbf{x} = \mathbf{x}'$. Also, given that $\mathbf{y} = \mathbf{x} \oplus \mathbf{s} = \mathbf{x}' \oplus \mathbf{s}' \implies \mathbf{s} = \mathbf{y} \oplus \mathbf{x}, \mathbf{s}' = \mathbf{y} \oplus \mathbf{x}'$, this equation is equivalent to these four linear conditions:

$$\begin{cases} (\mathbf{x},\mathbf{x}',\mathbf{y})=(0,1,0): \sum_{z\in\mathcal{Z}} W_{\mathbf{z}|\mathbf{x}}(z|0)V_{\mathbf{s}|\mathbf{z},\mathbf{x}'}(0\mid z,1)=\sum_{z'\in\mathcal{Z}} W_{\mathbf{z}|\mathbf{x}}\left(z'\mid 1\right)V_{\mathbf{s}|\mathbf{z},\mathbf{x}'}\left(1\mid z',0\right) \\ (\mathbf{x},\mathbf{x}',\mathbf{y})=(0,1,1): \sum_{z\in\mathcal{Z}} W_{\mathbf{z}|\mathbf{x}}(z\mid 0)V_{\mathbf{s}|\mathbf{z},\mathbf{x}'}(1\mid z,1)=\sum_{z'\in\mathcal{Z}} W_{\mathbf{z}|\mathbf{x}}\left(z'\mid 1\right)V_{\mathbf{s}|\mathbf{z},\mathbf{x}'}\left(0\mid z',0\right) \\ (\mathbf{x},\mathbf{x}',\mathbf{y})=(1,0,0): \sum_{z\in\mathcal{Z}} W_{\mathbf{z}|\mathbf{x}}(z\mid 1)V_{\mathbf{s}|\mathbf{z},\mathbf{x}'}(1\mid z,0)=\sum_{z'\in\mathcal{Z}} W_{\mathbf{z}|\mathbf{x}}\left(z'\mid 0\right)V_{\mathbf{s}|\mathbf{z},\mathbf{x}'}\left(0\mid z',1\right) \\ (\mathbf{x},\mathbf{x}',\mathbf{y})=(1,0,1): \sum_{z\in\mathcal{Z}} W_{\mathbf{z}|\mathbf{x}}(z\mid 1)V_{\mathbf{s}|\mathbf{z},\mathbf{x}'}(0\mid z,0)=\sum_{z'\in\mathcal{Z}} W_{\mathbf{z}|\mathbf{x}}\left(z'\mid 0\right)V_{\mathbf{s}|\mathbf{z},\mathbf{x}'}\left(1\mid z',1\right) \end{cases}$$

Given that $W_{\mathbf{z}|\mathbf{x}}(.)$ is a BSC(q) these four conditions can be rewrite respectively as follows:

$$\begin{cases} (1-q)\cdot(1-x_{01})+q\cdot(1-x_{11})=q\cdot x_{00}+(1-q)\cdot x_{10} \\ (1-q)\cdot x_{01}+q\cdot x_{11}=q\cdot(1-x_{00})+(1-q)\cdot(1-x_{10}) \\ q\cdot x_{00}+(1-q)\cdot x_{10}=(1-q)\cdot(1-x_{01})+q\cdot(1-x_{11}) \\ q\cdot(1-x_{00})+(1-q)\cdot(1-x_{10})=(1-q)\cdot x_{01}+q\cdot x_{11} \end{cases}$$

It can easily seen that all these equations are equivalent to $a^\top\cdot x=b$ where $a=[q,1-q,1-q,q]$, and $b=1$.
Similarly, the power constraint ($\lambda_{\mathbf{s}}$) in definition 19, can be rewrite as follows:

$$\left[P_{\mathbf{x},\mathbf{x}'}W_{\mathbf{z}|\mathbf{x}}V_{\mathbf{s}|\mathbf{z},\mathbf{x}'}\right]_{\mathbf{s}}\in\lambda_{\mathbf{s}},\lambda_{\mathbf{s}}=B_H(p)\Leftrightarrow\sum_{x,x',z}P_{\mathbf{x}}(x)P_{\mathbf{x}}\left(x'\right)W_{\mathbf{z}|\mathbf{x}}(z\mid x)V_{\mathbf{s}|\mathbf{z},\mathbf{x}'}\left(1\mid z,x'\right)<p$$

It can easily seen that this is equivalent to $c^\top\cdot x\le p$ where:

$$\begin{cases} c_{00}=(1-w)(w)(q)+(1-w)(1-w)(1-q) \\ c_{01}=(1-w)(1-w)(q)+(1-w)(w)(1-q) \\ c_{10}=(w)(w)(q)+(w)(1-w)(1-q) \\ c_{11}=(w)(1-w)(q)+(w)(w)(1-q) \end{cases}$$

Therefore we formulate confusability as a linear programming problem as follows:

$$\begin{aligned} \text{minimize}\quad & c^Tx=\sum_{j=00}^{11}c_j\,x_j \\ \text{subject to}\quad & a^Tx=\sum_{j=00}^{11}a_jx_j=b \\ & x_j\in[0,1],\quad j=00,01,10,11 \end{aligned}$$

Where $x=[x_{00},x_{01},x_{10},x_{11}]$ specify James's attack, $a=[q,1-q,1-q,q]$ and $b=1$ are showing confusability condition and $c=[c_{00},c_{01},c_{10},c_{11}]$ is the cost vector.

Claim 49. *$x^*=[0,\frac{q}{1-q},1,0]$ is the optimizer if $q\le p\le w\le\frac{1}{2}$*

Proof: It's sufficient to show that $x^*=[0,\frac{q}{1-q},1,0]$ satisfies KKT conditions:
1. Primal constraints: $a^Tx^*=\frac{q}{1-q}\times(1-q)+1\times(1-q)=1=b$, and $x_j^*\in[0,1],j=00,01,10,11$
2. Dual constraints: $\lambda\ge 0$
3. Complementary slackness: only non-zero elements of λ are corresponds to $-x_{00}\le 0$, $x_{10}\le 1$, and $-x_{11}\le 0$ which will denoted as $\lambda_{00},\lambda_{10},\lambda_{11}$.
4. Gradient of Lagrangian with respect to x^* vanishes:

$$\begin{bmatrix} c_{00} \\ c_{01} \\ c_{10} \\ c_{11} \end{bmatrix}+\begin{bmatrix} -\lambda_{00} \\ 0 \\ \lambda_{10} \\ -\lambda_{11} \end{bmatrix}+\nu\cdot\begin{bmatrix} a_{00} \\ a_{01} \\ a_{10} \\ a_{11} \end{bmatrix}=0$$

to show that x^* is optimizer, it's sufficient to show that:

$$\begin{cases} c_{01}+\nu\cdot a_{10}=0\iff\nu=\frac{-c_{01}}{1-q} \\ c_{00}-\lambda_{00}+\nu\cdot a_{00}=0\overset{\lambda_{00}\ge 0}{\iff}c_{00}+v\cdot a_{00}\ge 0\iff 1-2q\ge 0\iff q\le\frac{1}{2} \\ c_{10}+\lambda_{10}+\nu\cdot a_{10}=0\overset{\lambda_{10}\ge 0}{\iff}c_{10}+\nu\cdot a_{10}\le 0\iff q\ge 0 \\ c_{11}-\lambda_{11}+v\cdot a_{11}=0\overset{\lambda_{11}\ge 0}{\iff}c_{11}+v\cdot a_{11}\ge 0\iff w\ge q \end{cases}$$

So, whenever $w\ge q\ge 0,q\le\frac{1}{2}$, optimizer is $x^*=[0,\frac{q}{1-q},1,0]$.

■

Claim 50. $w(p,q)=\frac{(q+1)-\sqrt{\frac{1-q}{1-2q}}\sqrt{-4p+2q^2+q+1}}{2}$

Proof:

$$p = c^T x = \frac{1-2q}{1-q} \times c_{01} + 1 \times c_{10} \iff \frac{1-2q}{1-q} \cdot w^2 + \frac{-1+q+2q^2}{1-q} \cdot w + \frac{p-pq+q^2}{1-q} = 0$$
$$\iff w = \frac{(q+1) \pm \sqrt{\frac{1-q}{1-2q}}\sqrt{-4p+2q^2+q+1}}{2} \overset{w \le \frac{1}{2}}{\iff} w = \frac{(q+1) - \sqrt{\frac{1-q}{1-2q}}\sqrt{-4p+2q^2+q+1}}{2}$$

■

Claim 51. *we are interested in the case that* $\frac{1}{2} \ge w \ge p \ge q \ge 0$*, this happens whenever* $\frac{1}{4(1-q)} \ge p \ge q \ge 0$

Proof:

$$\begin{cases} \Delta \ge 0 \iff -4p + 2q^2 + q + 1 \ge 0 \iff p \le \frac{2q^2+q+1}{4} \\ w \le \frac{1}{2} \iff q \le \sqrt{\frac{1-q}{1-2q}(-4p + 2q^2 + q + 1)} \iff p \le \frac{1}{4(1-q)} \\ w \ge p \iff -8p^2q + 8pq^2 - 4p^2 + 4q^2 \ge 0 \iff 4(p-q)(p(1-q) + q(1-p)) \ge 0 \end{cases}$$

But, one can easily check that $\frac{1}{4(1-q)} \le \frac{2q^2+q+1}{4}$, therefore it's sufficient that $\frac{1}{4(1-q)} \ge p \ge q \ge 0$ ■

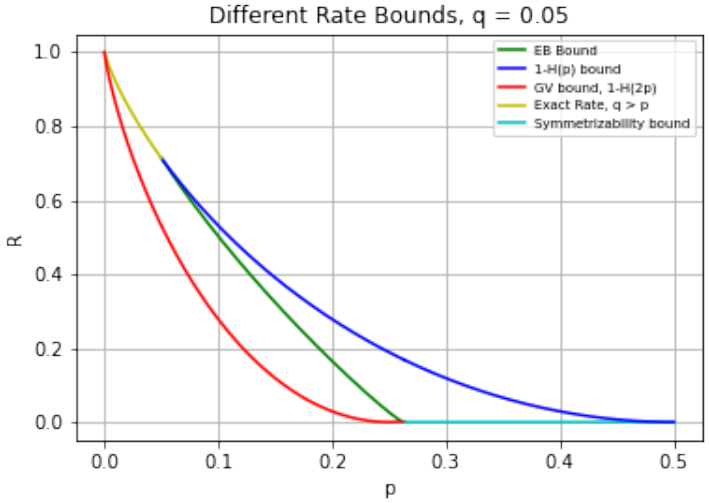


Fig. 13. Different rate bounds for 12 Channel when q = 0.05.

Remark 52. *We know that whenever* $q \le p$ *the rate is* $1 - H(p)$ *(yellow curve), and whenever* $p \ge \frac{1}{4(1-q)}$ *the rate is zero. [BDJ+20] For the* $q \le p \le \frac{1}{4(1-q)}$ *we don't know what the exact rate is. We know the GV lower bound,*$1 - H(2p)$ *, and* $1 - H(p)$ *upper bound. As Figure 13 illustrates, Myopic EB bound gives us a tighter upper bound on rate in* $q \le p \le \frac{1}{4(1-q)}$ *regime.*

■

REFERENCES

[Bas65] Leonid Alexandrovich Bassalygo. New upper bounds for error correcting codes. *Problems of Information Transmission*, 1(4):41–44, 1965.

[BBT59] David Blackwell, Leo Breiman, and A. J. Thomasian. The capacity of a class of channels. *The Annals of Mathematical Statistics*, 30(4):1229–1241, 1959.

[BDJ+20] Amitalok J. Budkuley, Bikash Kumar Dey, Sidharth Jaggi, Michael Langberg, Anand D. Sarwate, and Carol Wang. Symmetrizability for myopic avcs. In *2020 IEEE International Symposium on Information Theory (ISIT)*, pages 2103–2107, 2020.

[BDJ+21] Amitalok J. Budkuley, Bikash Kumar Dey, Sidharth Jaggi, Michael Langberg, Anand D. Sarwate, Carol Wang, and Yihan Zhang. Codes for adversaries. *Under preparation for submission to Foundations and Trends in Communications and Information Theory*, 12(1-2):1–242, 2021.

[BDP17a] Amitalok J. Budkuley, Bikash Kumar Dey, and Vinod M. Prabhakaran. Communication in the presence of a state-aware adversary. *IEEE Transactions on Information Theory*, 63(11):7396–7419, 2017.

[BDP17b] Amitalok J Budkuley, Bikash Kumar Dey, and Vinod M Prabhakaran. Communication in the presence of a state-aware adversary. *IEEE Transactions on Information Theory*, 63(11):7396–7419, 2017.

[CK81] Imre Csiszár and János Körner. On the capacity of the arbitrarily varying channel for maximum probability of error. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 57(1):87–101, 1981.

[CN88] I. Csiszar and P. Narayan. The capacity of the arbitrarily varying channel revisited: positivity, constraints. *IEEE Transactions on Information Theory*, 34(2):181–193, 1988.

[CN89] I. Csiszar and P. Narayan. Capacity and decoding rules for classes of arbitrarily varying channels. *IEEE Transactions on Information Theory*, 35(4):752–769, 1989.

[CT91] T.M. Cover and J.A. Thomas. *Elements of Information Theory*. Wiley, New York, 1991.

[DJL19a] Bikash Kumar Dey, Sidharth Jaggi, and Michael Langberg. Sufficiently myopic adversaries are blind. *IEEE Transactions on Information Theory*, 65(9):5718–5736, 2019.

[DJL+19b] Bikash Kumar Dey, Sidharth Jaggi, Michael Langberg, Anand D Sarwate, and Carol Wang. The interplay of causality and myopia in adversarial channel models. In *2019 IEEE International Symposium on Information Theory (ISIT)*, pages 1002–1006. IEEE, 2019.

[GK11] Abbas El Gamal and Young-Han Kim. *Network Information Theory*. Cambridge University Press, 2011.

[LN98] A. Lapidoth and P. Narayan. Reliable communication under channel uncertainty. *IEEE Transactions on Information Theory*, 44(6):2148–2177, 1998.

[MRRW77] Robert McEliece, Eugene Rodemich, Howard Rumsey, and Lloyd Welch. New upper bounds on the rate of a code via the delsarte-macwilliams inequalities. *IEEE transactions on Information Theory*, 23(2):157–166, 1977.

[Plo60] M. Plotkin. Binary codes with specified minimum distance. *IRE Transactions on Information Theory*, 6(4):445–450, 1960.

[Sar10] Anand Sarwate. Coding against myopic adversaries. In *Proc. IEEE Information Theory Workshop*, Dublin, Ireland, 2010.

[Sha59] C. E. Shannon. Coding theorems for a discrete source with a fidelity criterion. *IRE Nat. Conv. Rec.*, 7:142–163, 1959.

[Var57] Rom Rubenovich Varshamov. Estimate of the number of signals in error correcting codes. *Docklady Akad. Nauk, SSSR*, 117:739–741, 1957.

[Wan16] Carol Wang. On the capacity of the binary adversarial wiretap channel. In *2016 54th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 363–369. IEEE, 2016.

[WBBJ19] Xishi Wang, Amitalok J. Budkuley, Andrej Bogdanov, and Sidharth Jaggi. When are large codes possible for avcs? In *2019 IEEE International Symposium on Information Theory (ISIT)*, pages 632–636, 2019.

[ZJB20] Yihan Zhang, Sidharth Jaggi, and Amitalok J. Budkuley. Tight list-sizes for oblivious avcs under constraints, 2020.

[ZJB21] Yihan Zhang, Sidharth Jaggi, and Amitalok J. Budkuley. Tight list-sizes for oblivious avcs under constraints. In *2021 IEEE International Symposium on Information Theory (ISIT)*, pages 1439–1444, 2021.