Arbitrarily Varying Hypothesis Testing Approach To Quantum Source Compression And Classical Communication

Ayanava Dasgupta, Naqueeb Ahmad Warsi
Electronics and Communication Sciences Unit
Indian Statistical Institute, Kolkata
Kolkata 700108, India
Email: [ayanavadasgupta_r,naqueebwarsi]@isical.ac.in

Abstract

In information theory, we often encounter the problem of certain types of asymmetric hypothesis testing. In this work, we have considered the generalized problem of hypothesis testing between two different sources which changes with time arbitrarily. We mention this problem as "Arbitrarily Varying Hypothesis Testing". We consider this problem classically and provide a more intuitive proof than existing proofs with the help of a certain type of typical sets which we mention as "frequency typical sets". We also have tried to deal with this problem for classical quantum sources. As an application of our classical result, we have given proof for the achievability of deriving the capacity of a certain class of channel which we call "Arbitrarily Varying Channels" or AVC and proof for the achievability of finding the capacity region of multi-sender extension of AVC which we call "Arbitrarily Varying Multiple Access Channels" or AVMAC.

I. Introduction

Most of the results in traditional information theory are based on the assumption that the source and the channel are independent and identically distributed. However, in the real-world scenario, these assumptions may not be true. To relax these assumptions [1], [2] introduced the notion of arbitrarily varying source and channel and studied the two fundamental problems of information theory i.e. the problem of source compression for arbitrarily varying source (AVS) and the problem of reliable communication over an arbitrarily varying channel (AVC). In fact, in [3] it was shown that the capacity of an AVC may be zero under certain conditions which he termed as the "Non-symmetrizability" condition. The problem was further studied by others [4]–[7].

In [8]–[10], it has been observed that the problem of source compression and the problem of reliable communication over a noisy channel is closely related to the problem of asymmetric hypothesis testing. This motivates us to study the above-mentioned two fundamental problems from the point of view of asymmetric hypothesis testing between two arbitrarily varying distributions which is defined below:

Definition 1. *(Asymmetric hypothesis testing between two arbitrarily varying distributions) Consider two collections $\mathcal{W}_1 := \{P_s\}_{s\in\mathcal{S}}$ and $\mathcal{W}_2 := \{Q_s\}_{s\in\mathcal{S}}$ of probability distributions, defined over a non-empty finite support set $\mathcal{X}$, where $|\mathcal{S}| \leq \infty$. For any sequence $s^n := (s_1, \cdots, s_n) \in \mathcal{S}^n$, $P_{s^n} := \prod_{i=1}^n P_{s_i}$ and $Q_{s^n} := \prod_{i=1}^n Q_{s_i}$ is a distribution over $\mathcal{X}^n$. The goal here is to design a hypothesis test T that accepts $\{P_{s^n}\}_{s^n\in\mathcal{S}^n}$ (Null hypothesis) with very high probability and rejects $\{Q_{s^n}\}_{s^n\in\mathcal{S}^n}$ (Alternative hypothesis) with very high probability.*

More precisely, in [8] by Fangwei and Shiyi proved the following:

***Theorem* 1.** *[8, Theorem 4.1] If $\widehat{\mathcal{W}_1} \cap \widehat{\mathcal{W}_2} = \phi$, then for all $0 < \varepsilon < 1$, we have*

$$\lim_{n\to\infty} -\frac{1}{n}\log\ \overline{\beta}_n(\varepsilon) = \min_{P\in\widehat{\mathcal{W}_1}, Q\in\widehat{\mathcal{W}_2}} D(P||Q),$$

where,

$$\overline{\beta}_n(\varepsilon) := \min_{\substack{A\subseteq\mathcal{X}^n:\forall s^n\in\mathcal{S}^n,\\ P_{s^n}(A)\geq 1-\varepsilon}} \max_{s^n\in\mathcal{S}^n} Q_{s^n}(A),$$

and for each $i = 1, 2$, $\widehat{\mathcal{W}_i}$ is the convex hull of $\mathcal{W}_i$ (which is defined in the above problem statement)

This theorem can be interpreted as the generalization of Stein's lemma [] in the arbitrarily varying setting.

In the quantum setting a special case of the problem mentioned in Definition 1 was studied by Nötzel in [11]. In particular, the following problem was studied in [11]:

Theorem 2. *[11] Consider a collection $\mathcal{G} := \{\rho_s\}_{s\in\mathcal{S}} \subseteq \mathcal{D}(\mathcal{B})$ of quantum states (density operator) and another quantum state $\sigma \in \mathcal{D}(\mathcal{B})$, where $|\mathcal{S}| < \infty$. Then for all $0 < \varepsilon < 1$, we have*

$$\lim_{n\to\infty} -\frac{1}{n}\log\bar{\beta}_n(\varepsilon) = \min_{\rho\in\hat{\mathcal{G}}} D(\rho||\sigma),$$

where $\hat{\mathcal{G}} := conv\{\hat{\mathcal{G}}\}$ and

$$\bar{\beta}_n(\varepsilon) := \min_{\substack{0\preceq\Pi\preceq\mathbb{I}:\\ \forall s^n\in\mathcal{S}^n, \mathrm{Tr}[\Pi\rho_{s^n}]\geq 1-\varepsilon}} \mathrm{Tr}[\Pi\sigma^{\otimes n}],$$

and $\forall s^n = (s_1, \cdots, s_n) \in \mathcal{S}^n$, $\rho_{s^n} := \bigotimes_{i=1}^n \rho_{s_i}$.

It is important to observe here that, unlike Theorem 1 (wherein both the distributions were arbitrarily varying), in Theorem 2 only the null hypothesis is arbitrarily varying whereas the alternative hypothesis is fixed. Motivated by the proofs of Theorem 1 and Theorem 2, we aim to prove a quantum version of Theorem 1 for arbitrarily varying classical-quantum states, wherein the alternative hypothesis is the marginal of the null hypothesis. We define the problem formally in Definition 2:

Definition 2. *Consider a collection of density operators $\left\{\rho^{\mathcal{B}}_{x,s} : x \in \mathcal{X}, s \in \mathcal{S}\right\} \subset \mathcal{D}(\mathcal{B})$, where $\mathcal{X}, \mathcal{S}$ are two finite and non-empty sets. Now consider two collections $\Theta_1 := \left\{\rho_{s^n}^{X^n\mathcal{B}^n}\right\}_{s^n\in\mathcal{S}^n}$ and $\Theta_1 := \left\{\sigma_{s^n}^{X^n\mathcal{B}^n}\right\}_{s^n\in\mathcal{S}^n}$ of classical-quantum states, such that $\forall s^n \in \mathcal{S}^n$,*

$$\rho_{s^n}^{X^n\mathcal{B}^n} := \sum_{x^n\in\mathcal{X}^n} p(x^n)|x^n\rangle\langle x^n|^{X^n} \otimes \rho^{\mathcal{B}^n}_{x^n,s^n},$$

$$\sigma_{s^n}^{X^n\mathcal{B}^n} := \sum_{x^n\in\mathcal{X}^n} p(x^n)|x^n\rangle\langle x^n|^{X^n} \otimes \rho^{\mathcal{B}^n}_{s^n},$$

where, $\rho^{\mathcal{B}^n}_{x^n,s^n} := \bigotimes_{i=1}^n \rho^{\mathcal{B}}_{x_i,s_i}$, $\rho^{\mathcal{B}^n}_{s^n} := \bigotimes_{i=1}^n \rho^{\mathcal{B}}_{s_i}$ (where, $\forall s \in \mathcal{S}, \rho^{\mathcal{B}}_s := \sum_{x\in\mathcal{X}} p(x)\rho^{\mathcal{B}}_{x,s}$), p is a probability distribution defined over$\mathcal{X}$ and X is the input Hilbert space of dimension $|\mathcal{X}|$. The aim here is to design a hypothesis test (which is a measurement in the quantum setting and for the sake of simplicity we assume it to be projective) ${\Pi_T}^{X^n\mathcal{B}^n}$, which accepts Θ_1 (Null Hypothesis) with very large probability and reject Θ_2 (Alternative Hypothesis) with very large probability.

To give an achievability result for the problem mentioned in Definition 2, one needs to quantize the achievability proof of Theorem 1. Arguably, it appears that this quantization of the classical proof is difficult. As it involves the union of frequency typical sets and it is not very clear how to define a union of frequency typical subspaces. In this work, motivated by techniques used in the proof of achievability part of Theorem 2, we give a new proof for Theorem 1 which appears to have more potential for yielding an achievability proof for the quantum problem defined in Definition 2.

Similar to the steps used in the proof of the achievability part of Theorem 2, we use the following steps to give a new achievability proof for Theorem 1. We note here that, Step 1 and Step 2 mentioned below are required to give an achievability proof for the case when the null hypothesis is arbitrarily varying, whereas the alternative hypothesis is fixed. We need Step 3 along with Step 1 and Step 2 to give a new proof achievability proof for Theorem 1.

Step 1 (Compound vs fixed) Consider $\{P_s\}_{s\in\mathcal{S}}$, where $|\mathcal{S}|$ is finite and a probability distribution Q defined over the set $\mathcal{X}$. The aim here is to design a hypothesis test such that for large enough n, the test accepts $P_s^{\otimes n}$ with a large probability $\forall s \in \mathcal{S}$ and accept $Q^{\otimes n}$ with an exponentially small probability. See Lemma 1 for more details. Here $\otimes n$ signifies n i.i.d copies of a distribution.

Step 2 (Arbitrarily varying vs fixed) Consider $\{P_{s^n}\}_{s^n\in\mathcal{S}^n}$, where $|\mathcal{S}|$ is finite and for every $s^n \in \mathcal{S}^n, P_{s^n} := \prod_{i=1}^n P_{s_i}$ and a probability distribution $Q^{\otimes n}$ defined over the set $\mathcal{X}^n$. The aim here is to design a hypothesis test such that for large enough n, the test accepts P_{s^n} with a large probability $\forall s^n \in \mathcal{S}^n$ and accept $Q^{\otimes n}$ with an exponentially small probability. This step follows from Step 1. See Lemma 4 for more details.

Step 3 This step contains two parts and it depends on observations that frequency typical sets are permutation invariant and their intersections are either empty or permutation invariant (see Fact 1). In the first part, we show the hypothesis testing between an arbitrarily varying source and a compound source i.e. given $\{P_{s^n}\}_{s^n\in\mathcal{S}^n}$ (defined in Step 2) and another collection $\{Q_s^{\otimes n}\}_{s\in\mathcal{S}}$ of probability distributions defined over the set $\mathcal{X}^n$. The aim here is to design a hypothesis test such that for large enough n, the test accepts P_{s^n} with a large probability $\forall s^n \in \mathcal{S}^n$ and accept $Q_s^{\otimes n}$ with an exponentially small probability $\forall s \in \mathcal{S}$. This step directly follows from Step 2 by taking the intersection of the collection of hypothesis-tests coming from Step 2 considering $\{P_{s^n}\}_{s^n\in\mathcal{S}^n}$ and a fixed $Q_s^{\otimes n}$ for each $s \in \mathcal{S}$. In the quantum scenario, this idea of the intersection is highly non-trivial. See Lemma (4) and (5) for more intuitions.

In the second part we observe that for any fixed $s^n \in \mathcal{S}^n$ and any of its permutation $\sigma_n(s^n) \in \mathcal{S}^n$ the probability of any permutation invariant set T remains the same i.e. $P_{s^n}(T) = P_{\sigma_n(s^n)}(T)$ and $Q_{s^n}(T) = Q_{\sigma_n(s^n)}(T)$. Further, fix a sequence $\mathcal{S}^n$ and let q be the empirical distribution (defined over $\mathcal{S}$) obtained from s^n. Now, using the fact that T is permutation invariant, one can get a lower and an upper bound on $\widehat{Q}^{\otimes n}(T)$, where $\widehat{Q} := \sum_{s\in\mathcal{S}} q(s)Q_s$. These observations along with the first part of Step 3 yield a new achievability proof of Theorem 1. Lemma 1 below formalizes all these observations.

Lemma 1. *(**Classical Arbitrarily Varying Hypothesis Testing**) Let $\mathcal{S}$ be a finite set, $\Sigma_1 = \{P_1, P_2, .., P_{|\mathcal{S}|}\}$ and $\Sigma_2 = \{Q_1, Q_2, .., Q_{|\mathcal{S}|}\} \subset \mathcal{P}(\mathcal{X})$ be two collection of probability distribution over $\mathcal{X}$. Then, there exists a set T, which satisfies the following,*

$$\min_{s^n} P_{s^n}(T) \geq 1 - |S|2^{-n(\frac{\varepsilon^2}{2} - \frac{d+|\mathcal{S}|}{n}\log(2n))},$$
$$\max_{s^n} Q_{s^n}(T) \leq 2^{-n(\min_{P\in\widehat{\Sigma}_1, Q\in\widehat{\Sigma}_2} D(P||Q) - \frac{|\mathcal{S}|\log(2n)}{n} - \underline{\Theta}(n,\varepsilon,d,\Sigma_2))},$$

where, $\widehat{\Sigma}_1 := conv(\Sigma_1), \widehat{\Sigma}_2 := conv(\Sigma_2)$ and

$$\underline{\Theta}(n,\varepsilon,d,\Sigma_2) := \frac{d}{n}\log(2n) + \varepsilon|\log\frac{\varepsilon}{d}| + d.\varepsilon\max_{i,s}|\log(Q_s(i))|.$$

The following corollary directly follows from 1:

Corollary 1. *Consider $\mathcal{X}, \mathcal{Y}$ be two finite sets, $\mathcal{S}$ be an arbitrary set and a collection $\mathcal{W} := \{W_{Y|X,S}(\cdot|\cdot,s)\}_{s\in S}$ of conditional probability distributions (doubly stochastic $|\mathcal{X}\times\mathcal{S}|\times|\mathcal{Y}|$ matrices) and a probability distribution $P_X(\cdot)$ on $\mathcal{X}$. Given two collections of conditional probability distributions $\mathcal{W}_1 := \{W_{XY|S}(\cdot,\cdot|s)\}_{s\in S}$ and $\mathcal{W}_2 := \{W_{Y|S}(\cdot|s)\times P_X(\cdot)\}_{s\in\mathcal{S}}$, where $\forall s\in\mathcal{S}$, $W_{XY|S} = W_{Y|X,S}(\cdot|\cdot,s)\times P_X(\cdot)$ and $W_{Y|S}(\cdot|s) = \sum_{x\in\mathcal{X}} P_X(x)W_{Y|X,S}(\cdot|x,s)$. Then, for any sequence $s^n\in\mathcal{S}^n$ there exists a set $\mathcal{D}\subset\mathcal{X}^n\times\mathcal{Y}^n$ which satisfies the following*

$$W_{X^nY^n|S^n}(\mathcal{D}|s^n) \geq 1-\varepsilon, \tag{1}$$
$$W_{Y^n|S^n}\times P_{X^n}(\mathcal{D}|s^n) \leq 2^{-n(\min_Q I_{PQ}[X;Y]-\delta)}, \tag{2}$$

where,

$$I_{PQ}[X;Y] := D\Big(\sum_{s\in\mathcal{S}} Q(s)W_{XY|S}(\cdot,\cdot|s)\Big|\Big|\sum_{s\in\mathcal{S}} Q(s)W_{Y|S}(\cdot|s)\times P_X(\cdot)\Big). \tag{3}$$

Proof. Observe that, Theorem 1 has minimization over two convex hulls whereas in this special case of arbitrarily varying hypothesis testing, we have only one minimization. This is true because for $(X,Y)\sim P_{XY}$,

$$I[X:Y] = \min_{Q_Y} D(P_{XY}||P_X\times Q_Y).$$

See Appendix IX-E for more details. □

As an application of Corollary 1, we give an achievability proof for deriving the capacity of a classical AVC from the point of view of hypothesis testing. In [7], [12]–[15], a necessary and sufficient condition for the capacity of an AVC to be positive was defined. We mention this condition below in Definition 3:

Definition 3. *[7], [12]–[15] An AVC $\mathcal{W}_{\textbf{avc}} := \{W_{Y|X,S}(\cdot|\cdot,s)\}_{s\in\mathcal{S}}$ is non-symmetrizable if and only if for each probability distribution $U_{S|X}(\cdot|x)_{x\in\mathcal{X}}$ defined over $\mathcal{S}$ such that $\exists y\in\mathcal{Y}, x, x'\in\mathcal{X}$ for which*

$$\sum_{s\in\mathcal{S}} U_{S|X}(s|x')W_{Y|X,S}(y|x,s) \neq \sum_{s\in\mathcal{S}} U_{S|X}(s|x)W_{Y|X,S}(y|x',s).$$

Motivated by Corollary 1 we give a different necessary and sufficient condition for the capacity of an AVC to be positive. See Lemma 6 for further details. Further, we also give a necessary and sufficient condition for the capacity region of an arbitrarily varying multiple access channel (AVMAC) to be non-empty, which is different from the necessary and sufficient condition proposed by Jahn in [12]. See Section VI for more details in AVMAC.

In this manuscript, we aim to prove a quantum version of Corollary 1, for the problem defined in Definition 2 using the quantum version of Steps $1, 2$ and 3 mentioned above. Steps 1 and 2 have already been quantized by Nötzel [11]. We make some progress towards quantizing Step 3. In particular, in Lemma 2 below, we prove an achievability proof for the hypothesis testing between classical quantum arbitrarily varying and compound states. We note here, that the proof of Lemma 2 is arguably non-trivial and uses the tools and techniques built by Sen in [16]. Further, as mentioned in Step 3, we need an upper and a lower bound on $\widehat{Q}^{\otimes n}$. We prove a quantum version of this upper bound in Lemma 3. However, in this manuscript, we are not able to prove a quantum version of the lower bound, because it is not very clear whether the projection operator obtained in Lemma 2 is permutation invariant. If this were true then this result of the lower bound along with Lemma 2 and Lemma 3 would yield us an achievability proof of the problem defined in Definition 2, which is similar in spirit to the proof of Lemma 1. This would help us give a new achievability proof for deriving capacity region for classical-quantum arbitrarily varying channels [15] (CQ-AVC) from the point of view of asymmetric hypothesis testing. This would also help us provide achievability proof for classical-quantum arbitrarily varying multiple access channels (CQ-AVMAC) from the point of view of asymmetric hypothesis testing. Given these applications and the success of quantizing Steps $1, 2$ and 3 (except the second part of Step 3) it tempts us to propose Conjecture 1, which we plan to study in the future.

Lemma 2. *(***Hypothesis Testing of Classical-Quantum Arbitrarily Varying vs. Compound states***) Consider a collection of density operators $\{\rho^{\mathcal{B}}_{x,s} : x \in \mathcal{X}, s \in \mathcal{S}\}$ on the Hilbert Space $\mathcal{B}$, where $\mathcal{X}, \mathcal{S}$ are finite and non-empty sets. Now consider the following two collections $\left\{\rho^{X^n\mathcal{B}^n}_{s^n}\right\}_{s^n\in\mathcal{S}^n}$ and $\left\{\sigma^{X\mathcal{B}\otimes n}_{s}\right\}_{s\in\mathcal{S}}$ of classical-quantum states such that,*

$$\rho^{X^n\mathcal{B}^n}_{s^n} := \sum_{x^n\in\mathcal{X}^n} p(x^n)|x^n\rangle\langle x^n|^{X^n} \otimes \rho^{\mathcal{B}^n}_{x^n,s^n}, \quad \forall s^n \in \mathcal{S}^n,$$

$$\sigma^{X\mathcal{B}}_{s} := \sum_{x\in\mathcal{X}} p(x)|x\rangle\langle x|^{X} \otimes \rho^{\mathcal{B}}_{s}, \qquad \forall s \in \mathcal{S},$$

where, $\rho^{\mathcal{B}^n}_{x^n,s^n} := \bigotimes_{i=1}^n \rho^{\mathcal{B}}_{x_i,s_i}$, $\rho^{\mathcal{B}}_{s} := \sum_{x\in\mathcal{X}} \rho^{\mathcal{B}}_{x,s}$ and X is the input Hilbert space. Then, there exists an isometric embedding $\mathcal{T} : X^n \otimes \mathcal{B}^n \to \mathcal{X}' \otimes \mathcal{B}'$ (where $|\mathcal{X}' \otimes \mathcal{B}'| > |X^n \otimes \mathcal{B}^n|$) and a projector $(\Pi')^{\mathcal{X}'\mathcal{B}'}$, such that for any state vector $|h\rangle \in X^n \otimes \mathcal{B}$,

$$\|\mathcal{T}(|h\rangle\langle h|) - \mathcal{I}(|h\rangle\langle h|)\|_1 \leq 4.\sqrt{2^{-n\alpha\varepsilon^2 + \frac{2|\mathcal{B}|^2+|\mathcal{S}|}{n}\log(2n)}},$$

$$\mathrm{Tr}[(\Pi')^{\mathcal{X}'\mathcal{B}'}\mathcal{T}(\rho^{X^n\mathcal{B}^n}_{s^n})] \geq 1 - (24.|\mathcal{S}| + 4).\sqrt{2^{-n\alpha\varepsilon^2 + \frac{2|\mathcal{B}|^2+|\mathcal{S}|}{n}\log(2n)}}, \ \forall s^n \in \mathcal{S}^n,$$

$$\mathrm{Tr}[(\Pi')^{\mathcal{X}'\mathcal{B}'}\mathcal{T}(\sigma^{X\mathcal{B}\otimes n}_{s})] \leq 2^{-n\left\{\min\limits_{\substack{\rho\in conv\{\mathcal{G}'\}\\ s\in\mathcal{S}}} D(\rho\|\sigma^{X\mathcal{B}}_s) - \Theta\left(n,\varepsilon,|\mathcal{B}|,\sigma^{X\mathcal{B}}_s\right) - \frac{|\mathcal{S}|}{n}\log(2n)+1\right\}}, \forall s \in \mathcal{S},$$

where, $\mathcal{I} : X^n \otimes \mathcal{B}^n \to \mathcal{X}' \otimes \mathcal{B}'$ is an identity embedding and $\mathcal{G}' := conv\left\{\rho^{X\mathcal{B}}_s\right\}_{s\in\mathcal{S}}$ and $\forall s \in \mathcal{S}, \rho^{X\mathcal{B}}_s := \sum_{x\in\mathcal{X}} p(x)|x\rangle\langle x|^{X} \otimes \rho^{\mathcal{B}}_{x,s}$.

Lemma 3. *Assuming the problem-setting of Lemma 2, we can show the following:*

$$\mathrm{Tr}[(\Pi')^{\mathcal{X}'\mathcal{B}'}\mathcal{T}(\sigma^{X^n\mathcal{B}^n}_{\bar{s}^{\otimes n}})] \leq 2^{-n\left\{\min\limits_{\substack{\rho\in conv\{\mathcal{G}'\}\\ \sigma\in conv\{\mathcal{J}'\}}} D(\rho\|\sigma) - \Theta\left(n,\varepsilon,|\mathcal{B}|,\sigma^{X\mathcal{B}}_s\right) - \frac{|\mathcal{S}|}{n}\log(2n)+2\right\}},$$

where, $\sigma^{X^n\mathcal{B}^n}_{\bar{s}^{\otimes n}} := \sum_{x^n\in\mathcal{X}^n} p(x^n)|x^n\rangle\langle x^n|^{X^n} \otimes \left(\sum_{s\in\mathcal{S}} q(s)\rho^{\mathcal{B}}_s\right)^{\otimes n}$, q is a probability distribution over $\mathcal{S}$ and $\mathcal{J}' := conv\left\{\sigma^{X\mathcal{B}}_s\right\}_{s\in\mathcal{S}}$.

Proof. See Appendix IX-G for the proof. □

Conjecture 1. *Assuming the problem-setting of Lemma 2, for any $s^n \in \mathcal{S}^n$ and any of its permutation $\sigma_n(s^n) \in \mathcal{S}^n$, does $(\Pi')^{\mathcal{X}'\mathcal{B}'}$ hold permutation invariance with respect to $\mathcal{T}(\sigma^{X^n\mathcal{B}^n}_{s^n})$ and $\mathcal{T}(\sigma^{X^n\mathcal{B}^n}_{\sigma_n(s^n)})$, where, $\sigma^{X^n\mathcal{B}^n}_{s^n} := \bigotimes_{i=1}^n \sigma^{X\mathcal{B}}_{s_i}$?*

As another application of quantum arbitrary varying hypothesis testing discussed above, we also have provided an achievability proof for Schumacher's source compression [17], [18] where the quantum sources are arbitrarily varying. The derivation of this result follows from Nötzel's work [11].

II. Notations and Facts

We use $\mathcal{H}$ to denote a finite-dimensional Hilbert space, $\mathcal{D}(\mathcal{H})$ to represent the set of all state density matrix acting on $\mathcal{H}$ and $\mathcal{L}(\mathcal{H})$ represents the set of all operator over $\mathcal{H}$ ($\mathcal{D}(\mathcal{H}) \subset \mathcal{L}(\mathcal{H})$). Let $\mathcal{X}$ and $\mathcal{S}$ be two finite sets denoting the source and state alphabets. $\mathcal{P}(\mathcal{X})$ is the set of probability distribution over $\mathcal{X}$. $T_n(\mathcal{X})$ is the set of empirical distributions with denominator n. Similarly, for any distribution P defined over a set $\mathcal{X}$, we will use the notation $\mathrm{supp}(P) := \{x \in \mathcal{X} : P(x) > 0\}$. Given two distribution P and Q over the set $\mathcal{X}$ such that $\mathrm{supp}(P) \subseteq \mathrm{supp}(Q)$, the relative entropy between P and Q is defined as,

$$D(P\|Q) := \sum_{x\in\mathcal{X}} P(x)\log\left(\frac{P(x)}{Q(x)}\right).$$

Given two quantum states $\rho, \sigma \in \mathcal{D}(\mathcal{H})$ such that $\mathrm{supp}(\rho) \subseteq \mathrm{supp}(\sigma)$, the quantum relative entropy (also known to be *quantum KL-divergence*) between ρ and σ is defined as,

$$D(\rho\|\sigma) := \mathrm{Tr}[\rho(\log(\rho) - \log(\sigma)].$$

A distance measure on $\mathcal{P}(\mathcal{X})$ is defined by $\|P - Q\| = \sum_{x\in\mathcal{X}} |P(x) - Q(x)|$ where, $P, Q \in \mathcal{P}(\mathcal{X})$. For an operator $O \in \mathcal{L}(H)$, $\|O\|_p$ is the operator schatten-p norm of O. The entropy of $P \in \mathcal{P}(\mathcal{X})$ defined as $H(P) = -\sum_{x\in\mathcal{X}} P(x)\log P(x)$. For a set A, $conv(A)$ denotes the convex hull of the set A.

Definition 4 (Frequency type classes)**.** *For a non-empty finite set $\mathcal{X}$ and an integer $n > 0$, given a frequency (or, type) function $f : \mathcal{X} \to \mathbb{N}$ (such that $\sum_{i\in\mathcal{X}} f(i) = n$), we define a set $T_f \subset \mathcal{X}^n$ to be the frequency type class corresponding to f as follows,*

$$T_f := \{(x_1, \cdots, x_n) \in \mathcal{X}^n : |\{k : x_k = i\}| = f(i), \forall i \in \mathcal{X}\}.$$

The above set T_f is a permutation invariant set i.e. for every sequence $x^n : (x_1, \cdots, x_n) \in \mathcal{X}^n$, under any permutation $\sigma_n : \mathcal{X}^n \to \mathcal{X}^n$, $\sigma_n(x^n) \in T_f$ where, $\sigma_n(x^n) := (x_{\sigma_n^{-1}(1)}, \cdots, x_{\sigma_n^{-1}(n)})$.

Definition 5. *(Frequency typical set) Given a probability distribution p over a non empty set $\mathcal{X}$ and for some $0 < \delta < 1$, we define $T_p^\delta \subset \mathcal{X}^n$ as the frequency typical set with respect to p as follows:*

$$T_p^\delta := \bigcup_{f : \|\bar{f} - p\| \leq \delta} T_f,$$

where, f is a frequency function and T_f is the frequency type class corresponding tp f and $\bar{f}(i) := \frac{f(i)}{n}, \forall I \in \mathcal{X}$.

The above set T_p^δ is a permutation invariant set since it is a union of frequency type classes that are permutation invariant itself.

Fact 1. *Given two different frequency typical sets T_p^δ and $T_q^{\delta'}$ for some $0 < \delta, \delta' < 1$, then $T_p^\delta \cap T_q^{\delta'}$ is either empty or permutation invariant.*

Proof. If there exists no type f which satisfies both $\|f - p\| \leq \delta$ and $\|f - q\| \leq \delta'$ then $T_p^\delta \cap T_q^{\delta'} = \varnothing$. This follows from the fact that for to frequency functions f, g it follows directly from Definition 4 that $T_f \cap T_g = \varnothing$. Now suppose if there exists a sequence $x^n \in T_p^\delta \cap T_q^{\delta'}$ then from definition of T_p^δ and $T_q^{\delta'}$, it follows that all the permutations of x^n which follows a particular frequency function h will reside in $T_p^\delta \cap T_q^{\delta'}$. Hence $T_p^\delta \cap T_q^{\delta'}$ is a union of certain frequency type classes making it permutation invariant. □

Fact 2. *For two operators $A, B \in \mathcal{L}(H)$ we have the following:*

$$\mathrm{Tr}[AB] \leq \|AB\|_1 \leq \min\{\|A\|_\infty \|B\|_1, \|A\|_1 \|B\|_\infty\}.$$

Fact 3. *(Triangle inequality under operator norm) For two operators $A, B \in \mathcal{L}(H)$ we have the following:*

$$\|A + B\|_p \leq \|A\|_p + \|B\|_p.$$

Fact 4. *(Pinsker's inequality) Given two probability distributions $P, Q \in \mathcal{P}(\mathcal{X})$ over a non-empty finite set $\mathcal{X}$, then,*

$$D(P\|Q) \geq \frac{1}{2}\|P - Q\|^2.$$

Fact 5. *(Gao's union bound) [19], [20] Let $\Pi_1, \Pi_2, \cdots, \Pi_n$ be projectors over $\mathcal{H}$ and $\rho \in \mathcal{D}(\mathcal{H})$. Then,*

$$\mathrm{Tr}(\Pi_n \cdots \Pi_2 \Pi_1 \rho \Pi_1 \Pi_2 \cdots \Pi_n) \geq 1 - 4\sum_{i=1}^{n} \mathrm{Tr}[\Pi_i^c \rho],$$

where $\Pi_i^c = \mathbb{I} - \Pi_i$.

Fact 6. *(Gentle Measurement lemma under Ensemble of States) [21] Let $\{p(x), \rho_x\}$ be an ensemble and let $\bar{\rho} := \sum_x p(x)\rho_x$. If an operator E, where $0 \preceq E \preceq I$, has high overlap with the expected state $\bar{\rho}$ i.e. $\mathrm{Tr}[E\bar{\rho}] \geq 1 - \varepsilon$, where $\varepsilon \in (0, 1)$. Then,*

$$\mathbb{E}_X \left[\left\| \sqrt{E}\rho_X \sqrt{E} - \rho_X \right\|_1 \right] \leq 2\varepsilon^{1/2}.$$

III. Proof of Lemma 1

Before delving into the proof let's first observe the following lemma, which will help us to prove this lemma:

Lemma 4. *Let $\mathcal{S}$ be a non-empty finite set, $\{P_1, P_2, .., P_{|\mathcal{S}|}\} \subset \mathcal{P}(\mathcal{X})$ be a collection of probability distribution on a finite non-empty set $\mathcal{X}$ and another probability distribution $Q \in \mathcal{P}(\mathcal{X})$. Consider $A_n = \bigcup_{f : \overline{f} \in \Lambda_\varepsilon} T_f$, where, $\Lambda_\varepsilon = \{p \in \mathcal{P}(\mathcal{X}) : \exists s \in \mathcal{S} : \|p - P_s\| < \varepsilon\}$ and $\overline{f} \in \mathcal{P}(\mathcal{X}) : \overline{f}(d) = \frac{f(d)}{n}, \forall d \in \mathcal{X}$. Then, for any sequence $s^n \in \mathcal{S}^n$, we have,*

$$P_{s^n}(A_n) \geq 1 - 2^{-n(\frac{\varepsilon^2}{2} - \frac{d + |S|}{n}\log(2n))},$$
$$Q^n(A_n) \leq 2^{-n(\min_{P \in \widehat{\Sigma}} D(P||Q) - \Theta(n, \varepsilon, d, Q))},$$

where, $\widehat{\Sigma} := conv\big\{P_1, P_2, .., P_{|\mathcal{S}|}\big\}$.

Proof. See Appendix IX-A for the proof. □

By lemma 4, there exists a collection of permutation invariant sets $\{A_s\}_{s \in \mathcal{S}}$ such that for each $s \in \mathcal{S}$,

$$P_{s^n}(A_s) \geq 1 - 2^{-n(\frac{\varepsilon^2}{2} - \frac{d^2 + |S|}{n}\log(2n))},$$
$$Q_s^n(A_s) \leq 2^{-n(\min_{P \in \widehat{\Sigma}_1} \ D(P||Q_s) - \Theta(n, \varepsilon, d, Q_s))}.$$

Now consider the following set

$$T := \bigcap_{s\in\mathcal{S}} A_s.$$

It follows that T follows the Since for each $s \in \mathcal{S}$, A_s is a permutation invariant set, thus T is also a permutation invariant set, since for any sequence $\hat{x}^n \in T$ implies that $\forall s \in \mathcal{S}, \hat{x}^n \in A_s$. So, all permutations of $\hat{x}^n$ must belong to A_s for each $s \in \mathcal{S}$, making T a permutation invariant set. Now, for any sequence $\tilde{s}^n \in \mathcal{S}^n$, using union bound of sets, we have,

$$\begin{aligned} P_{\tilde{s}^n}(T) &\geq 1 - \sum_{s\in\mathcal{S}} P_{\tilde{s}^n}(A_s^c) \\ &= 1 - |\mathcal{S}|2^{-n(\frac{\varepsilon^2}{2} - \frac{d^2+|S|}{n}\log(2n))}. \end{aligned} \tag{4}$$

Now, from the definition of the set T, it directly follows that for any $s \in \mathcal{S}$,

$$\begin{aligned} Q_s^n(T) &\leq Q_s^n(A_s) \\ &= 2^{-n(\min_{P\in\widehat{\Sigma}_1} \ D(P||Q_s) - \Theta(n,\varepsilon,d,Q_s))}, \end{aligned} \tag{5}$$

$$\leq 2^{-n(\min\limits_{\substack{P\in\widehat{\Sigma}_1 \\ s\in\mathcal{S}}} D(P||Q_s) - \Theta(n,\varepsilon,d,Q_s))}. \tag{6}$$

For a particular sequence $\bar{s}^n$, we have a particular type $N : \mathcal{S} \to \mathbb{N}$, which $\bar{s}^n$ falls in, and the corresponding empirical (frequency) distribution $q(s) = \frac{N(s|\bar{s}^n)}{n}$. Now consider $\widehat{Q} = \sum_{s\in\mathcal{S}} q(s)Q_s$ and thus $\widehat{Q}^n = \sum_{s^n\in\mathcal{S}^n} q(s^n)Q_{s^n}$.Now we have,

$$\begin{aligned} \widehat{Q}^n(T) &\geq \sum_{s^n\in T_N} q(s^n)Q_{s^n}(T) \\ &\overset{a}{=} \sum_{s^n\in T_N} q(s^n)Q_{\bar{s}^n}(T) \\ &\geq (2n)^{-|\mathcal{S}|}Q_{\bar{s}^n}(T), \end{aligned}$$

where, a follows from (23). Therefore, we have,

$$Q_{\bar{s}^n}(T) \leq (2n)^{|\mathcal{S}|}\widehat{Q}^n(T). \tag{7}$$

Hence, we can write the following,

$$\begin{aligned} \widehat{Q}^n(T) &= 1 - \widehat{Q}^n(T^c) \\ &\overset{a}{\leq} 1 - \min_{s\in\mathcal{S}} Q_s^n(T^c) \\ &\overset{b}{\leq} 1 - 1 + \max_{s\in\mathcal{S}} 2^{-n(\min_{P\in\widehat{\Sigma}_1} \ D(P||Q_s) - \Theta(n,\varepsilon,d,Q_s))} \\ &= 2^{-n(\min_{P\in\widehat{\Sigma}_1, s\in\mathcal{S}} \ D(P||Q_s) - \underline{\Theta}(n,\varepsilon,d,\Sigma_2))} \\ &\leq 2^{-n(\min_{P\in\widehat{\Sigma}_1, Q\in\widehat{\Sigma}_2} \ D(P||Q) - \underline{\Theta}(n,\varepsilon,d,\Sigma_2))}, \end{aligned} \tag{8}$$

where a follows from Claim 1, b follows from (5). Now from (7), it follows that, for any sequence $\bar{s}^n$

$$\begin{aligned} Q_{\bar{s}^n}(T) &\leq (2n)^{|\mathcal{S}|}2^{-n(\min_{P\in\widehat{\Sigma}_1, Q\in\widehat{\Sigma}_2} \ D(P||Q) - \underline{\Theta}(n,\varepsilon,d,\Sigma_2))} \\ &= 2^{-n(\min_{P\in\widehat{\Sigma}_1, Q\in\widehat{\Sigma}_2} \ D(P||Q) - \frac{|\mathcal{S}|log(2n)}{n} - \underline{\Theta}(n,\varepsilon,d,\Sigma_2))}. \end{aligned}$$

Finally, we can say,

$$\max_{s^n\in\mathcal{S}^n} Q_{s^n}(T) \leq 2^{-n(\min_{P\in\widehat{\Sigma}_1, Q\in\widehat{\Sigma}_2} D(P||Q) - \frac{|\mathcal{S}|\log(2n)}{n} - \underline{\Theta}(n,\varepsilon,d,\Sigma_2))}.$$

Now the following corollary directly follows from the result of Lemma 4:

Corollary 2. *Consider $\Delta_1 := \left\{\sum_{s\in\mathcal{S}} p(s)P_s\right\}_{p\in T_n(\mathcal{S})}$ and $\Delta_2 := \left\{\sum_s q(s)Q_s\right\}_{q\in T_n(\mathcal{S})}$ be two collection of probability distributions, where $\{P_1, P_2, .., P_{|\mathcal{S}|}\}$ and $\{Q_1, Q_2, .., Q_{|\mathcal{S}|}\}$ are collections of probability distributions over $\mathcal{X}$. Then, there exists a set B which satisfies the following,*

$$\begin{aligned} &\min_{s^n} P_{s^n}(T) \geq 1 - |S|2^{-n(\frac{\varepsilon^2}{2} - \frac{d+|\mathcal{S}|}{n}\log(2n))}, \\ &\max_{s^n} Q_{s^n}(T) \leq 2^{-n(\min_{P\in\Delta_1, Q\in\Delta_2} D(P||Q) - \frac{|\mathcal{S}|\log(2n)}{n} - \underline{\Theta}(n,\varepsilon,d,\Delta_2))}, \end{aligned}$$

Proof. See Appendix IX-D for the proof. □

Remark 1.
Interestingly, with the help of the converse of Theorem 1 and Corollary 2 we see the following

$$\min_{P\in\Delta_1,Q\in\Delta_2} D(P||Q) = \min_{P\in\widehat{\Sigma}_1,Q\in\widehat{\Sigma}_2} D(P||Q),$$

where $\widehat{\Sigma}_1 := conv\{P_1, P_2, .., P_{|\mathcal{S}|}\}$ *and* $\widehat{\Sigma}_2 := conv\{Q_1, Q_2, .., Q_{|\mathcal{S}|}\}$*. This will drastically reduce the parameter space we are finding the optimal value of* $D(P||Q)$*.*

IV. Proof of Lemma 2

From the achievability of [11, Theorem 1 Second part] i.e. [11, Theorem 7 Achievability], it follows that for an $\varepsilon \in (0,1)$ there exists a collection of projective measurements $\left\{\Pi_s^{X^n\mathcal{B}^n}\right\}_{s\in\mathcal{S}}$ with the following properties:

$$\begin{aligned}
&\mathrm{Tr}[\Pi_s^{X^n\mathcal{B}^n}\rho_{s^n}^{X^n\mathcal{B}^n}] \geq 1 - 2^{-n\alpha\varepsilon^2 + \frac{2|\mathcal{B}|^2+|\mathcal{S}|}{n}\log(2n)}, \quad \forall s^n \in \mathcal{S}^n,\\
&\mathrm{Tr}[\Pi_s^{X^n\mathcal{B}^n}\sigma_s^{X\mathcal{B}\otimes n}] \leq 2^{n\left\{-\min\limits_{\rho\in\mathrm{conv}\{\mathcal{G}'\}} D(\rho\|\sigma_s^{X\mathcal{B}}) + \Theta\left(n,\varepsilon,|\mathcal{B}|,\sigma_s^{X\mathcal{B}}\right) + \frac{|\mathcal{S}|}{n}\log(2n)\right\}}, \forall s \in \mathcal{S},
\end{aligned}$$

where, $\mathcal{G}' := \left\{\rho_s^{X\mathcal{B}}\right\}_{s\in\mathcal{S}}$ and $\Theta\left(n,\varepsilon,|\mathcal{B}|,\sigma_s^{X\mathcal{B}}\right) := \varepsilon\left|\log\left(\frac{\varepsilon}{|\mathcal{B}|}\right)\right| + \frac{|\mathcal{B}|^2}{n}\log(2n) + |\mathcal{B}|\varepsilon.\left\|\log\sigma_s^{X\mathcal{B}}\right\|_\infty$ and $\forall s \in \mathcal{S}$, $\Pi_s^{X^n\mathcal{B}^n}$ is of the following form:

$$\Pi_s^{X^n\mathcal{B}^n} := \sum_{x^n\in\mathcal{X}^n} \overset{X^n}{|x^n\rangle\langle x^n|} \otimes \Pi_{x^n,s}^{\mathcal{B}^n}.$$

Now, before proceeding to further analysis, consider the following lemma:

Lemma 5. *Consider the two collections of classical-quantum states* $\left\{\rho_k^{X^n\mathcal{B}}\right\}_{k=1}^K$ *and* $\left\{\sigma_m^{X^n\mathcal{B}}\right\}_{m=1}^M \subset \mathcal{D}(X^n\otimes\mathcal{B})$*, given as follows:*

$$\begin{aligned}
\rho_k^{X^n\mathcal{B}} &= \sum_{x^n\in\mathcal{X}^n} p(x^n)\overset{X^n}{|x^n\rangle\langle x^n|} \otimes \rho_{k,x^n}^{\mathcal{B}}, \forall k = 1,2,\cdots,K,\\
\sigma_m^{X^n\mathcal{B}} &= \sum_{x^n\in\mathcal{X}^n} p(x^n)\overset{X^n}{|x^n\rangle\langle x^n|} \otimes \sigma_m^{\mathcal{B}}, \forall m = 1,2,\cdots,M.
\end{aligned}$$

We have two projective measurement operators $\left\{\Pi_m^{X^n\mathcal{B}}\right\}_{m=1}^M$ *which is given as follows:*

$$\Pi_m^{X^n\mathcal{B}} := \sum_{x^n\in\mathcal{X}^n} \overset{X^n}{|x^n\rangle\langle x^n|} \otimes \Pi_{m,x^n}^{\mathcal{B}}, \quad \forall m = 1,2,\cdots,M,$$

where for each $m \in \{1,2,\cdots,M\}$*,* $\Pi_m^{X^n\mathcal{B}}$ *satisfies the following properties:*

$$\begin{aligned}
&\mathrm{Tr}[\Pi_m^{X^n\mathcal{B}}\rho_k^{X^n\mathcal{B}}] \geq 1-\tilde{\varepsilon}, \quad \forall k = 1,2,\cdots,K,\\
&\mathrm{Tr}[\Pi_m^{X^n\mathcal{B}}\sigma_m^{X^n\mathcal{B}}] \leq 2^{-k_m},
\end{aligned}$$

for some $k_m > 0, \forall m \in \{1,2,\cdots,M\}$*. Then, here exists an isometric embedding* $\mathcal{T} : X^n\otimes\mathcal{B} \to \mathcal{X}'\otimes\mathcal{B}'$ *(where* $\dim(\mathcal{X}'\otimes\mathcal{B}') > \dim(X^n\otimes\mathcal{B})$*), such that for any state vector* $|h\rangle \in X^n\otimes\mathcal{B}$*,*

$$\|\mathcal{T}(|h\rangle\langle h|) - \mathcal{I}(|h\rangle\langle h|)\|_1 \leq 4\sqrt{\tilde{\varepsilon}},$$

where, $\mathcal{I} : X^n\otimes\mathcal{B} \to \mathcal{X}'\otimes\mathcal{B}'$ *is an identity embedding and for* $\mathcal{T}$ *we can show that there also exists a projector* $\Pi_{\cap M}$ *over* $\mathcal{X}'\otimes\mathcal{B}'$ *which satisfies the following properties:*

$$\mathrm{Tr}\left[\Pi_{\cap M}\mathcal{T}\left(\rho_k^{X^n\mathcal{B}}\right)\right] \geq 1 - (24.M+4).\sqrt{\tilde{\varepsilon}}, \quad \forall k = 1,2,\cdots,K, \tag{9}$$

$$\mathrm{Tr}\left[\Pi_{\cap M}\mathcal{T}\left(\sigma_m^{X^n\mathcal{B}}\right)\right] \leq 2^{-k_m+1}, \quad \forall m = 1,2,\cdots,M. \tag{10}$$

Proof. See Appendix IX-F for the proof. □

Then, similar to the approaches mentioned in the proof of Lemma 5 (see subsubsections IX-F1,IX-F2 of Appendix), we consider $\mathcal{L}_1, \cdots, \mathcal{L}_{|\mathcal{S}|}$ be a collection of $|\mathcal{S}|$ distinct mutually orthogonal Hilbert spaces isomorphic to a Hilbert space $\mathcal{L}$, where for each $s \in \mathcal{S}, |\mathcal{L}_s| = |\mathcal{L}|$ and $\mathcal{L}_s$ can be represented as

$$\mathbb{I}^{\mathcal{L}_i} := \sum_{l_i} |l_i\rangle\overset{\mathcal{L}_i}{\langle} l_i|.$$

Now consider a collection of tilting maps $\{T_{i,l_i,\delta} : \mathcal{B}^n \to \mathcal{B}_i^n\}_{i=1}^{|\mathcal{S}|}$, where for each $i \in \mathcal{S}, T_{i,l_i,\delta}$ is an isometric embedding from $\mathcal{B}^n$ to a significantly larger Hilbert space $\mathcal{B}_i^n := \mathcal{B}^n \oplus \mathcal{B}^n \otimes \mathcal{L}_i$, defined as follows:

$$T_{i,l_i,\delta} : |h\rangle \longmapsto \frac{1}{\sqrt{1+\delta^2}}(|h\rangle + \delta|h\rangle|l_i\rangle),$$

where $|l_i\rangle$ is one particular computational basis vectors of $\mathcal{L}_i$. We represent this map $T_{i,l_i,\delta}$ as it tilts a vector within $\mathcal{B}$ towards a particular orthogonal direction in $\mathcal{L}_i$. We also have another tilting map $T_{\mathcal{S},l_\mathcal{S},\delta} : \mathcal{B}^n \to \mathcal{B}'$, which is a isometric embedding from $\mathcal{B}^n$ to a significantly larger Hilbert space $\mathcal{B}'$, where $\mathcal{B}' := \mathcal{B}^n \oplus \bigoplus_{i=1}^{|\mathcal{S}|}(\mathcal{B}^n \otimes \mathcal{L}_i)$ and $l_\mathcal{S} := \{l_i\}_{i=1}^{|\mathcal{S}|}$ with $\forall i \in \mathcal{S}, |l_i\rangle$ being one particular computational basis vector of $\mathcal{L}_i$, defined as follows:

$$T_{\mathcal{S},l_\mathcal{S},\delta} : |h\rangle \longmapsto \frac{1}{\sqrt{1+|\mathcal{S}|\delta^2}}(|h\rangle + \sum_{i=1}^{|\mathcal{S}|}\delta|h\rangle|l_i\rangle).$$

We get the collection of classical-quantum states $\left\{(\rho')_{s^n}^{\mathcal{X}'\mathcal{B}'}\right\}_{s^n\in\mathcal{S}^n}$ and $\left\{(\sigma')_{s}^{\mathcal{X}'\mathcal{B}'}\right\}_{s\in\mathcal{S}}$ from $\left\{\rho_{s^n}^{X^n\mathcal{B}^n}\right\}_{s^n\in\mathcal{S}^n}$ and $\left\{\sigma_{s}^{X\mathcal{B}^{\otimes n}}\right\}_{s\in\mathcal{S}}$ given as

$$(\rho')_{s^n}^{\mathcal{X}'\mathcal{B}'} := \frac{1}{|\mathcal{L}|^{|\mathcal{S}|}} \sum_{x^n\in\mathcal{X}^n, l_\mathcal{S}} p(x^n)|x^n\rangle\langle x^n| \otimes \left(\bigotimes_{i=1}^{|\mathcal{S}|} \overset{\mathcal{L}_i}{|l_i\rangle\langle l_i|}\right) \otimes \rho_{x^n,l_\mathcal{S},s^n}^{\mathcal{B}'},$$

$$(\sigma')_{s^{\otimes n}}^{\mathcal{X}'\mathcal{B}'} := \frac{1}{|\mathcal{L}|^{|\mathcal{S}|}} \sum_{x^n\in\mathcal{X}^n, l_\mathcal{S}} p(x^n)|x^n\rangle\langle x^n| \otimes \left(\bigotimes_{i=1}^{|\mathcal{S}|} \overset{\mathcal{L}_i}{|l_i\rangle\langle l_i|}\right) \otimes \tilde{\rho}_{l_\mathcal{S},s^{\otimes n}}^{\mathcal{B}'},$$

where, $\mathcal{X}' := X^n \otimes \mathcal{L}^{\otimes|\mathcal{S}|}$, $\rho_{x^n,l_\mathcal{S},s^n}^{\mathcal{B}'} := T_{\mathcal{S},l_\mathcal{S},\delta}\left(\rho_{x^n,s^n}^{\mathcal{B}^n}\right)$ and $\tilde{\rho}_{l_\mathcal{S},s^{\otimes n}}^{\mathcal{B}'} := T_{\mathcal{S},l_\mathcal{S},\delta}\left(\rho_{s}^{\mathcal{B}^{\otimes n}}\right)$. Considering the collection $\left\{\Pi_s^{X^n\mathcal{B}^n}\right\}_{s\in\mathcal{S}}$ using exact same construction mentioned in the proof of Lemma 5 (see subsubsection IX-F2 in Appendix), we can design a projector $(\Pi')^{\mathcal{X}'\mathcal{B}'}$ on $\mathcal{X}' \times \mathcal{B}'$ and from Lemma 5, it follows that $(\Pi')^{\mathcal{X}'\mathcal{B}'}$ satisfies the following properties:

$$\begin{aligned}
\mathrm{Tr}[(\Pi')^{\mathcal{X}'\mathcal{B}'}(\rho')_{s^n}^{\mathcal{X}'\mathcal{B}'}] &\geq 1 - (24.|\mathcal{S}|+4).\sqrt{2^{-n\alpha\varepsilon^2+\frac{2|\mathcal{B}|^2+|\mathcal{S}|}{n}\log(2n)}},\ \forall s^n \in \mathcal{S}^n,\\
\mathrm{Tr}[(\Pi')^{\mathcal{X}'\mathcal{B}'}(\sigma')_{s^{\otimes n}}^{\mathcal{X}'\mathcal{B}'}] &\leq 2^{-n\left\{\min\limits_{\rho\in\mathrm{conv}\{\mathcal{G}'\}} D(\rho\|\sigma_s^{X\mathcal{B}})-\Theta\left(n,\varepsilon,|\mathcal{B}|,\sigma_s^{X\mathcal{B}}\right)-\frac{|\mathcal{S}|}{n}\log(2n)+1\right\}},\forall s \in \mathcal{S},\\
\mathrm{Tr}[(\Pi')^{\mathcal{X}'\mathcal{B}'}(\sigma')_{s^{\otimes n}}^{\mathcal{X}'\mathcal{B}'}] &\leq 2^{-n\left\{\min\limits_{\substack{\rho\in\mathrm{conv}\{\mathcal{G}'\}\\ s\in\mathcal{S}}} D(\rho\|\sigma_s^{X\mathcal{B}})-\Theta\left(n,\varepsilon,|\mathcal{B}|,\sigma_s^{X\mathcal{B}}\right)-\frac{|\mathcal{S}|}{n}\log(2n)+1\right\}},\forall s \in \mathcal{S}..
\end{aligned} \tag{11}$$

V. Application to Capacity of AVC

A discrete memory-less point-to-point arbitrarily varying channel (AVC) can be given as a collection $\mathcal{W}_{\textbf{avc}} := \left\{W_{Y|X,S}(\cdot|\cdot,s)\right\}_{s\in\mathcal{S}}$ of channels with a finite input alphabet set $\mathcal{X}$ and output alphabet set $\mathcal{Y}$ and a finite set $\mathcal{S}$, which we use to index $\mathcal{W}_{\textbf{avc}}$ can be mentioned as the set of states, wherein each i-th instance of the channel being used, one particular channel $W_{Y|X,S}(\cdot|\cdot,s)$ is arbitrarily chosen from the collection $\mathcal{W}_{\textbf{avc}}$, where s is picked from the set $\mathcal{S}$ arbitrarily. Now, for a given n-length input sequence $x^n = (x_1, \cdots, x_n) \in \mathcal{X}^n$ and a state sequence $s^n = (s_1, \cdots, s_n) \in \mathcal{S}^n$ probability of getting an output sequence $y^n = (y_1, \cdots, y_n) \in \mathcal{Y}^n$ is given as

$$W_{Y^n|X^n,S^n}^n(y^n|x^n,s^n) = \prod_{i=1}^n W_{Y|X,S}(y_i|x_i,s_i).$$

A $\left(2^{nR},n\right)$-codebook $\mathcal{C}_n$, for communicating over an AVC comprises of

- A message set $\mathcal{M}_n := \left\{1,2,\cdots,2^{nR}\right\}$.
- An enccoding operation $\mathcal{E}^{(n)} : \mathcal{M}_n \to \mathcal{X}^n$.
- A decoding operation $\mathcal{D}^{(n)} : \mathcal{Y}^n \to \mathcal{M}_n$.

Given a message m, after the encoding operation, it gets mapped to a n-length codeword $x^n(m) := \mathcal{E}^{(n)}(m)$. After sending the codeword via the n instances of the channel and decoding operation we get $\hat{M}$ as the decoded message.

For a particular message m chosen by the sender and given a codebook $\mathcal{C}_n$ and a state sequence s^n, We define the probability of error as follows

$$\begin{aligned} e(m, \mathcal{C}_n, s^n) &:= P\{\hat{M} \neq m | M = m, s^n\} \\ &= \sum_{\substack{y^n \in \mathcal{Y}^n: \\ \mathcal{D}^{(n)}(y^n) \neq m}} W^n_{Y^n|X^n,S^n}(y^n|x^n(m), s^n). \end{aligned} \tag{12}$$

We define the average error probability for a given deterministic codebook $\mathcal{C}_n$ and state sequence s^n as follows:

$$\bar{e}(\mathcal{C}_n, s^n) := \frac{1}{2^{nR}} \sum_{m=1}^{2^{nR}} e(m, \mathcal{C}_n, s^n).$$

Definition 6. *A number $R > 0$ is called an **achievable rate** for the given AVC $\mathcal{W}_{\textbf{avc}}$, if for every $\lambda > 0, \delta > 0$ and sufficiently large n, there exist a deterministic code $\mathcal{C}_n$ such that*

$$\tfrac{1}{n} \log |\mathcal{M}_n| > R - \delta', \qquad \max_{s^n \in \mathcal{S}^n} \bar{e}(\mathcal{C}_n, s^n) < \lambda.$$

The supremum over all such achievable rates is called the deterministic code capacity $C(\mathcal{W}_{\textbf{avc}})$ of the AVC. From [7, Theorem 1] we know that if $C(\mathcal{W}_{\textbf{avc}}) > 0$ then

$$C(\mathcal{W}_{\textbf{avc}}) = \max_P \min_Q I_{P,Q}[X : Y].$$

Necessary and Sufficient Condition of AVC $\mathcal{W}_{\textbf{avc}}$ for having positive capacity i.e. $C(\mathcal{W}_{\textbf{avc}}) > 0$

Corollary 1 hints the following necessary and sufficient condition for AVC $\mathcal{W}_{\textbf{avc}}$ to have a positive capacity: lemma

Lemma 6. *$C(\mathcal{W}_{\textbf{avc}}) > 0$ if and only if $\hat{\mathcal{W}}_1 \cap \hat{\mathcal{W}}_2 = \varnothing$, where $\hat{\mathcal{W}}_1 := conv\{W_{XY|S}(\cdot, \cdot|s)\}_{s \in S}$ and $\hat{\mathcal{W}}_2 := conv\{W_{Y|S}(\cdot|s) \times P_X(\cdot)\}_{s \in \mathcal{S}}$, where $\forall s \in \mathcal{S}$, $W_{XY|S} = W_{Y|X,S}(\cdot|\cdot, s) \times P_X(\cdot)$ and $W_{Y|S}(\cdot|s) = \sum_{x \in \mathcal{X}} P_X(x) W_{Y|X,S}(\cdot|x, s)$.*

Proof. From (3) and the intrinsic property of $D(\cdot\|.)$ we have the more compact form of the inverse condition for positive capacity of AVC $\mathcal{W}_{\textbf{avc}}$ ($\hat{\mathcal{W}}_1 \cap \hat{\mathcal{W}}_2 \neq \varnothing$.) i.e. there exists a probability distribution Q over $\mathcal{S}$ such that $\forall x \in \mathcal{X}, y \in \mathcal{Y}$,

$$\begin{aligned} &\sum_{s \in \mathcal{S}} Q(s) W_{XY|S}(x, y|s) = \sum_{s \in \mathcal{S}} Q(s) W_{Y|S}(y|s) \times P_X(x) \\ \Rightarrow &\sum_{s \in \mathcal{S}} Q(s) W_{Y|X,S}(y|x, s) = \sum_{s \in \mathcal{S}} Q(s) W_{Y|S}(y|s). \end{aligned} \tag{13}$$

Now, we will divide the proof into two parts i.e. we will first prove the necessity of the above condition for capacity of $\mathcal{W}_{\textbf{avc}}$ being positive and then the sufficiency of the above condition:

1) Necessity of the condition for $C(\mathcal{W}_{\textbf{avc}}) > 0$: Using the contrapositive argument, we need to show that if the AVC $\mathcal{W}_{\textbf{avc}}$ satisfies the condition given in (13), then $C(\mathcal{W}_{\textbf{avc}}) = 0$. Suppose an arbitrarily varying channel follows (13). In that case, it follows that there exists a distribution Q over $\mathcal{S}$ such that under the expectation of state sequences with respect to Q^n, the output Y^n at the receiver's end is always distributed with the same probability distribution i.e. $W^n_{Y^n}(\cdot)$ (where, $W^n_{Y^n}(\cdot) := \sum_{s^n \in c\mathcal{S}^n} W^n_{Y^n|S^n}(\cdot|s^n) Q(s^n)$), irrespective of the message transmitted. So the receiver cannot do anything but guess the output which will always give a non-zero error expected over state sequences. Hence for a particular codebook $\mathcal{C}_n$, there exists one such state sequence for which $\bar{e}(\mathcal{C}_n, s^n)$ is non-zero making the channel's capacity equal to zero according to Definition 6. The proof can be given formally as follows:

$$\begin{aligned} \mathbb{E}_{S^n \sim Q^n}[\bar{e}(\mathcal{C}_n, S^n)] &\overset{a}{=} \frac{1}{2^{nR}} \sum_{m=1}^{2^{nR}} \sum_{\substack{\hat{m} \neq m \\ s^n \in \mathcal{S}^n \\ y^n \in \mathcal{Y}^n: \\ \mathcal{D}^{(n)}(y^n) = \hat{m}}} W^n_{Y^n|X^n,S^n}(y^n|x^n(m), s^n) Q^n(s^n) \\ &\overset{b}{=} \frac{1}{2^{nR}} \sum_{m=1}^{2^{nR}} \sum_{\hat{m} \neq m} \sum_{\substack{y^n \in \mathcal{Y}^n: \\ \mathcal{D}^{(n)}(y^n) = \hat{m}}} W^n_{Y^n}(y^n). \end{aligned} \tag{14}$$

Now for a distinct pair of messages (m_1, m_2), we can show the following:

$$\begin{aligned}
&\sum_{m=m_1,m_2}\sum_{\hat{m}\neq m}\sum_{\substack{y^n\in\mathcal{Y}^n:\\ \mathcal{D}^{(n)}(y^n)=\hat{m}}} W^n_{Y^n}(y^n)\\
&=\sum_{\hat{m}\neq m_1}\sum_{\substack{y^n\in\mathcal{Y}^n:\\ \mathcal{D}^{(n)}(y^n)=\hat{m}}} W^n_{Y^n}(y^n)+\sum_{\hat{m}\neq m_2}\sum_{\substack{y^n\in\mathcal{Y}^n:\\ \mathcal{D}^{(n)}(y^n)=\hat{m}}} W^n_{Y^n}(y^n)\\
&\geq 1.
\end{aligned}$$

Since there exists exactly 2^{nR-1} distinct pairs of message, we can rewrite (14) as follows:

$$\mathbb{E}_{s^n\sim Q^n}[\bar{e}(\mathcal{C}_n, s^n)] = \frac{1}{2^{nR}}\sum_{m=1}^{2^{nR}}\sum_{\hat{m}\neq m}\sum_{\substack{y^n\in\mathcal{Y}^n:\\ \mathcal{D}^{(n)}(y^n)=\hat{m}}} W^n_{Y^n}(y^n)\geq\frac{1}{2},$$

which proves the necessity of the condition 13. Now we will show the opposite direction of the proof i.e. the sufficiency of the condition 13.

2) Sufficiency of the condition for $C(\mathcal{W}_{\textbf{avc}})>0$: Here we have to show that if $\mathcal{W}_{\textbf{avc}}$ does not satisfy the condition given in(13) then $C(\mathcal{W}_{\textbf{avc}})>0$. We prove this by devising a direct coding setup for the AVC $\mathcal{W}_{\textbf{avc}}$ and showing existence of a positive achievable rate. Now the average error probability over the choice of a random code-book $\mathcal{C}_n$ for a given state sequence is as follows

$$\bar{e}(s^n) := \mathbb{E}_{\mathcal{C}_n}\left[\frac{1}{2^{nR}}\sum_{m=1}^{2^nR} e(m,\mathcal{C}_n,s^n)\right].$$

From Corollary 1, there exists a set $\mathcal{D}$ which satisfies equation (1),(2).Then, for a given output sequence $y^n\in\mathcal{Y}^n$, we decode it to a message $\hat{m}\in\mathcal{M}_n$, if $\{x^n(\hat{m}),y^n\}\in\mathcal{D}$. Thus given y^n, the decoding error $(e(m,\mathcal{C}_n,s^n))$ for a particular message m and a codebook $\mathcal{C}_n$ can be of two types.

$$\begin{aligned}
\mathcal{E}_{1,m} &:= \{\{x^n(m),y^n\}\notin\mathcal{D}\},\\
\mathcal{E}_{2,m} &:= \{\exists\hat{m}\neq m : \{x^n(\hat{m}),y^n\}\in\mathcal{D}\}.
\end{aligned}$$

Now for any state sequence s^n we get an upper bound for $\bar{e}(s^n)$ as follows.

$$\begin{aligned}
\mathbb{E}_{\mathcal{C}_n}[\frac{1}{2^{nR}}\sum_{m=1}^{2^nR} e(m,\mathcal{C}_n,s^n)] &\leq \frac{1}{2^{nR}}\sum_{m=1}^{2^nR}\mathbb{E}_{\mathcal{C}_n}[P\{\mathcal{E}_{1,m}\}+P\{\mathcal{E}_{2,m}\}]\\
&\overset{a}{\leq}\varepsilon+2^{-n(\min_Q I_{PQ}[X;Y]+\delta-R)},
\end{aligned}$$

where a follows from Corollary 1. So from the above equation, it follows that as $n\to\infty$ and if $R<\min_Q I_{PQ}[X;Y]-\delta$, for some $\lambda\in(0,1)$ we have

$$\max_{s^n\in\mathcal{S}^n}\mathbb{E}_{\mathcal{C}_n}[\frac{1}{2^{nR}}\sum_{m=1}^{2^nR} e(m,\mathcal{C}_n,s^n)]<\frac{\lambda}{2}.$$

Now using "Elimination of correlation" technique used by Ahlswede in [5, proof of Theorem 1], we can eliminate the randomness in codewords by showing that there exists a deterministic code $\tilde{\mathcal{C}}_{(n+f(n))}$ of length $(n+f(n))$, which satisfies,

$$\max_{s^n\in\mathcal{S}^n}\bar{e}\Big(\tilde{\mathcal{C}}_{(n+f(n))},s^n\Big)<\lambda,$$

where, $f(n)$ is in polynomial of n. Since $\mathcal{W}_{\textbf{avc}}$ does not satisfy (13), it follows that $\min_Q I_{PQ}[X;Y]>0$, which means there exists a codebook $\mathcal{C}_n$ for which a positive rate for transmitting messages over the channel $\mathcal{W}_{\textbf{avc}}$ can be achieved. This proves the sufficiency of (13). □

VI. Application to Capacity of AVMAC

A discrete memory-less arbitrarily varying multiple access channel can be given as a collection $\mathcal{W}_{\textbf{m.avc}} := \left\{W_{Y|X_1,X_2,S}(\cdot|\cdot,\cdot,s)\right\}_{s\in\mathcal{S}}$ of channels with a finite input alphabet set $\mathcal{X}_1$, $\mathcal{X}_2$ and output alphabet set $\mathcal{Y}$ and a finite set $\mathcal{S}$, which we use to index $\mathcal{W}_{\textbf{m.avc}}$ can be mentioned as the set of states, wherein each i-th instance of the channel being used, one particular channel $W_{Y|X_1,X_2,S}(\cdot|\cdot,\cdot,s)$ is arbitrarily chosen from the collection $\mathcal{W}_{\textbf{m.avc}}$, where s is picked from the set $\mathcal{S}$ arbitrarily. Now, for given n-length input sequence $x_1^n=(x_{1,1},\cdots,x_{1,n})\in\mathcal{X}_1^n$, $x_2^n=(x_{2,1},\cdots,x_{2,n})\in\mathcal{X}_2^n$ and a state sequence $s^n=(s_1,\cdots,s_n)\in\mathcal{S}^n$ probability of getting an output sequence $y^n=(y_1,\cdots,y_n)\in\mathcal{Y}^n$ is given as

$$W^n_{Y^n|X_1^n,X_2^n,S^n}(y^n|x_1^n,x_2^n,s^n)=\prod_{i=1}^{n}W_{Y|X_1,X_2,S}(y_i|x_{1,i},x_{2,i},s_i).$$

A $\big(2^{nR_1},2^{nR_2},n\big)$-codebook $\mathcal{C}_n$, for communicating over an AVMAC comprises of

- A message set $\mathcal{M}_n^1:=\big\{1,2,\cdots,2^{nR_1}\big\}$ and $\mathcal{M}_n^2:=\big\{1,2,\cdots,2^{nR_2}\big\}$.
- An encoding operation $\mathcal{E}^{(n)}:\mathcal{M}_n^1\times\mathcal{M}_n^2\to\mathcal{X}_1^n\times\mathcal{X}_2^n$.
- A decoding operation $\mathcal{D}^{(n)}:\mathcal{Y}^n\to\mathcal{M}_n^1\times\mathcal{M}_n^2$.

Given a message m_1 and m_2, after the encoding operation, it gets mapped to a n-length codeword $\{x_1^n(m_1),x_2^n(m_2)\}:=\mathcal{E}^{(n)}(m_1,m_2)$. After sending the codeword via the n instances of the channel and decoding operation we get $\{\hat{M}_1,\hat{M}_2\}$ as the decoded message.

For a particular message m_1 and m_2 chosen by the first and second sender respectively and given a codebook $\mathcal{C}_n$ and a state sequence s^n, We define the probability of error as follows

$$\begin{aligned}&e(m_1,m_2,\mathcal{C}_n,s^n)\\&:=P\{(\hat{M}_1,\hat{M}_2)\neq(m_1,m_2)|(M_1,M_2)=(m_1,m_2),s^n\}\\&=\sum_{\substack{y^n\in\mathcal{Y}^n:\\ \mathcal{D}^{(n)}(y^n)\neq(m_1,m_2)}}W^n_{Y^n|X_1^n,X_2^n,S^n}(y^n|x_1^n(m_1),x_2^n(m_2),s^n).\end{aligned}$$

We define the average error probability for a given deterministic codebook $\mathcal{C}_n$ and state sequence s^n as follows:

$$\bar{e}(\mathcal{C}_n,s^n):=\frac{1}{2^{nR_1}}\frac{1}{2^{nR_2}}\sum_{m_1=1}^{2^{nR_1}}\sum_{m_2=1}^{2^{nR_2}}e(m_1,m_2,\mathcal{C}_n,s^n).$$

Now we will define achievable rate-pair as follows:

Definition 7. *a pair of numbers $(R_1>0,R_2>0)$ is called an **achievable rate-pair** for the given AVMAC $\mathcal{W}_{\textbf{m.avc}}$, if for every $\lambda>0,\delta_1>0$ and $\delta_2>0$ and sufficiently large n, there exists a $\big(2^{nR_1},2^{nR_2},n\big)$ deterministic code $\mathcal{C}_n$ such that*

$$\begin{aligned}&\frac{1}{n}\log|\mathcal{M}_n^1|>R_1-\delta_1,\\&\frac{1}{n}\log|\mathcal{M}_n^2|>R_2-\delta_2,\end{aligned}\qquad\max_{s^n\in\mathcal{S}^n}\bar{e}(\mathcal{C}_n,s^n)<\lambda.$$

Unlike AVC, we here characterize the **achievable rate region** $\mathcal{R}(\mathcal{W}_{\textbf{m.avc}})$ (which is also known as **capacity region**) as follows Given a pair of random variables (X_1,X_2) with a joint probability distribution $P_{X_1X_2}:=P_{X_1}\cdot P_{X_2}$, we define the following set

$$\mathcal{R}_{X_1,X_2}:=\left\{\forall(R_1,R_2):\begin{array}{c}0\leq R_1\leq\min_{Q_1}I_{PQ_1}[X_1:Y|X_2]\\0\leq R_2\leq\min_{Q_2}I_{PQ_2}[X_2:Y|X_1]\\R_1+R_2\leq\min_{Q_3}I_{PQ_3}[X_1X_2:Y]\end{array}\right\},$$

where,

$$\begin{aligned}I_{PQ_1}[X_i:Y|X_j]:=D\Big(&\sum_{s\in\mathcal{S}}Q_1(s)W_{Y|X_1X_2S}\times P_{X_1}\times P_{X_2}(\cdot,\cdot,\cdot|s)\|\\&\sum_{s\in\mathcal{S}}Q_1(s)W_{Y|X_jS}\times P_{X_j}(\cdot,\cdot|s)\times P_{X_i}(\cdot)\Big),\ \text{for }(i,j)=(1,2),(2,1)\end{aligned}\tag{15}$$

$$\begin{aligned}I_{PQ_3}[X_1X_2:Y]:=D\Big(&\sum_{s\in\mathcal{S}}Q_3(s)W_{Y|X_1X_2S}\times P_{X_1}\times P_{X_2}(\cdot,\cdot,\cdot|s)\|\\&\sum_{s\in\mathcal{S}}Q_3(s)W_{Y|S}(\cdot|s)\times P_{X_1}(\cdot)\times P_{X_2}(\cdot)\Big).\end{aligned}\tag{16}$$

Set $\mathcal{R}^*(\mathcal{W}_{\textbf{m.avc}})=conv\Big\{\bigcup_{P_{X_1},P_{X_2}}\mathcal{R}_{X_1,X_2}\Big\}$. Then, from [12, Theorem 1], we know,

$$\mathcal{R}(\mathcal{W}_{\textbf{m.avc}})=\mathcal{R}^*(\mathcal{W}_{\textbf{m.avc}})\quad\text{ if }int\{\mathcal{R}^*(\mathcal{W}_{\textbf{m.avc}})\}\neq\varnothing,$$

where, $int\{\mathcal{A}\}$ denotes the topological interior of a set $\mathcal{A}$.

A. Conditions of AVMAC $\mathcal{W}_{\textbf{m.avc}}$ for having non-empty capacity region i.e. $int\{\mathcal{R}^*(\mathcal{W}_{\textbf{m.avc}})\} \neq \varnothing$

Corollary 1 hints the following necessary and sufficient condition for AVMAC $\mathcal{W}_{\textbf{m.avc}}$ to have a non-empty capacity region:

Lemma 7. *$int\{\mathcal{R}^*(\mathcal{W}_{\textbf{m.avc}})\} \neq \varnothing$ if and only if*

$$\Big\{\hat{\mathcal{W}} \cap \mathcal{W}_{X_1} = \varnothing\Big\} \cap \Big\{\hat{\mathcal{W}} \cap \mathcal{W}_{X_2} = \varnothing\Big\} \cap \Big\{\hat{\mathcal{W}} \cap \mathcal{W}_{X_1,X_2} = \varnothing\Big\}, \tag{17}$$

where,

$$\begin{aligned}
\hat{\mathcal{W}} \quad &:= conv\Big\{\big\{P_{X_1} \times P_{X_2} \times W_{Y|X_1,X_2,S}(\cdot,\cdot,\cdot|s)\big\}_{s\in\mathcal{S}}\Big\},\\
\mathcal{W}_{X_1} &:= conv\Big\{\big\{P_{X_1} \times P_{X_2} \times W_{Y|X_2,S}(\cdot,\cdot,\cdot|s)\big\}_{s\in\mathcal{S}}\Big\},\\
\mathcal{W}_{X_2} &:= conv\Big\{\big\{P_{X_1} \times P_{X_2} \times W_{Y|X_1,S}(\cdot,\cdot,\cdot|s)\big\}_{s\in\mathcal{S}}\Big\},\\
\mathcal{W}_{X_1,X_2} &:= conv\Big\{\big\{P_{X_1} \times P_{X_2} \times W_{Y|X_1,X_2,S}(\cdot,\cdot,\cdot|s)\big\}_{s\in\mathcal{S}}\Big\}.
\end{aligned} \tag{18}$$

Proof. We can give a more compact form of the inverse condition for non-empty capacity region for the AVMAC $\mathcal{W}_{\textbf{m.avc}}$ ($\{\hat{\mathcal{W}}\cup\mathcal{W}_{X_1} \neq \varnothing\} \cup \{\hat{\mathcal{W}} \cap \mathcal{W}_{X_2} \neq \varnothing\} \cup \{\hat{\mathcal{W}} \cap \mathcal{W}_{X_1,X_2} \neq \varnothing\}$) given as follows:

• *Case 1 : $\hat{\mathcal{W}} \cap \mathcal{W}_{X_1} \neq \varnothing$:* From (15) and the intrinsic property of $D(\cdot\|\cdot)$, there exists a probability distribution $\{Q_1(s)\}_{s\in\mathcal{S}}$ such that,

$$\sum_s Q_1(s) W_{Y|X_1X_2S} \times P_{X_1} \times P_{X_2}(\cdot,\cdot,\cdot|s) = \sum_s Q_1(s) W_{Y|X_2S} \times P_{X_2}(\cdot,\cdot|s) \times P_{X_1}(\cdot).$$

Further we can write that $\forall x_1 \in \mathcal{X}_1, x_2 \in \mathcal{X}_2, y \in \mathcal{Y}$,

$$\sum_s Q_1(s) W_{Y|X_1X_2S}(y|x_1,x_2,s) = \sum_s Q_1(s) W_{Y|X_2S}(y|x_2,s). \tag{19}$$

This is similar to the $\mathcal{X}_1$-symmetrizability condition of AVMAC mentioned in [14].

• *Case 2 : $\hat{\mathcal{W}} \cap \mathcal{W}_{X_2} \neq \varnothing$:* From (15) and definition of $D(\cdot\|\cdot)$, there exists a probability distribution $\{Q_2(s)\}_{s\in\mathcal{S}}$ such that,

$$\sum_s Q_2(s) W_{Y|X_1X_2S} \times P_{X_1} \times P_{X_2}(\cdot,\cdot,\cdot|s) = \sum_s Q_2(s) W_{Y|X_1S} \times P_{X_1}(\cdot,\cdot|s) \times P_{X_2}(\cdot). \tag{20}$$

This is similar to the $\mathcal{X}_2$-symmetrizability condition of AVMAC mentioned in [14].

• *Case 3 : $\hat{\mathcal{W}} \cap \mathcal{W}_{X_1,X_2} \neq \varnothing$:* From (16) and definition of $D(\cdot\|\cdot)$, there exists a probability distribution $\{Q_3(s)\}_{s\in\mathcal{S}}$ such that,

$$\sum_s Q_3(s) W_{Y|X_1X_2S} \times P_{X_1} \times P_{X_2}(\cdot,\cdot,\cdot|s) = \sum_s Q_3(s) W_{Y|S}(\cdot|s) \times P_{X_1}(\cdot) \times P_{X_2}(\cdot). \tag{21}$$

This is similar to the $(\mathcal{X}_1,\mathcal{X}_2)$-symmetrizability condition of AVMAC mentioned in [14].

Now, we will divide the proof into two parts i.e. we will first prove the necessity of the above condition for capacity of $\mathcal{W}_{\textbf{m.avc}}$ being positive and then the sufficiency of the above condition:

1) Necessity of the condition for $int\{\mathcal{R}^*(\mathcal{W}_{\textbf{m.avc}})\} \neq \varnothing$: Using the contrapositive argument, we need to show that if there exists a distribution Q over $\mathcal{S}$ for which the AVMAC $\mathcal{W}_{\textbf{m.avc}}$ satisfies at least one of the three sub-conditions given in eqs. (19) to (21), then $int\{\mathcal{R}^*(\mathcal{W}_{\textbf{m.avc}})\} = \varnothing$. Suppose a multiple-access arbitrary varying channel satisfies any of the three sub-conditions given in (19),(20) and (21). If it satisfies $\mathcal{X}_1$-symmetrizability condition i.e. (19).Then, it follows that there exists a distribution Q over $\mathcal{S}$ such that under the expectation of state sequences with respect to Q^n, the output Y^n at the receiver's end is always distributed with the same probability distribution i.e. $W^n_{Y^n|X_2^n}(\cdot|x_2^n(m_2))$ (where, $W^n_{Y^n|X_2^n}(\cdot|x_2^n(m_2)) := \sum_{s^n\in cS^n} W^n_{Y^n|X_2^n,S^n}(\cdot|x_2^n(m_2),s^n)Q(s^n)$ and $x_2^n(m_2)$ is the encoding of the message m_2 transmitted by the second sender corresponding to a particular codebook $\mathcal{C}_n$), irrespective of the message transmitted by the first sender. So the reciever cannot do anything but guess the output which will always give a non-zero error expected over state sequences. Hence for $\mathcal{C}_n$, there exists one such state sequence for which $\bar{e}(\mathcal{C}_n, s^n)$ is non-zero making the interior of the channel's capacity region empty i.e. $int\{\mathcal{R}^*(\mathcal{W}_{\textbf{m.avc}})\} = \varnothing$ according to Definition 7. With similar arguments, it can be shown that if a AVMAC satisfies either (20) or (21) will imply the interior of the channel's capacity region empty. A more mathematically involved proof is given as follows:

$$\mathbb{E}_{S^n\sim Q^n}[\bar{e}(\mathcal{C}_n,S^n)] \overset{a}{=} \frac{1}{2^{n(R_1+R_2)}}\sum_{\substack{m_1=1\\m_2=1}}^{2^{nR_1}2^{nR_2}}\sum_{\substack{(\hat{m}_1,\hat{m}_2)\neq(m_1,m_2)\\ s^n\in\mathcal{S}^n\\ y^n\in\mathcal{Y}^n:\\ \mathcal{D}^{(n)}(y^n)=(\hat{m}_1,\hat{m}_2)}} W^n_{Y^n|X_1^n,X_2^n,S^n}(y^n|x_1^n(m_1),x_2^n(m_2),s^n)Q^n(s^n)$$

$$= \overbrace{\frac{1}{2^{n(R_1+R_2)}}\sum_{\substack{m_1=1\\m_2=1}}^{2^{nR_1}2^{nR_2}}\sum_{\substack{\hat{m}_1\neq m_1\\ s^n\in\mathcal{S}^n\\ y^n\in\mathcal{Y}^n:\\ \mathcal{D}^{(n)}(y^n)=(\hat{m}_1,m_2)}} W^n_{Y^n|X_1^n,X_2^n,S^n}(y^n|x_1^n(m_1),x_2^n(m_2),s^n)Q^n(s^n)}^{\textbf{(1)}}$$

$$+ \overbrace{\frac{1}{2^{n(R_1+R_2)}}\sum_{\substack{m_1=1\\m_2=1}}^{2^{nR_1}2^{nR_2}}\sum_{\substack{\hat{m}_2\neq m_2\\ s^n\in\mathcal{S}^n\\ y^n\in\mathcal{Y}^n:\\ \mathcal{D}^{(n)}(y^n)=(m_1,\hat{m}_2)}} W^n_{Y^n|X_1^n,X_2^n,S^n}(y^n|x_1^n(m_1),x_2^n(m_2),s^n)Q^n(s^n)}^{\textbf{(2)}}$$

$$+ \overbrace{\frac{1}{2^{n(R_1+R_2)}}\sum_{\substack{m_1=1\\m_2=1}}^{2^{nR_1}2^{nR_2}}\sum_{\substack{\hat{m}_1\neq m_1\\ \hat{m}_2\neq m_2\\ s^n\in\mathcal{S}^n\\ y^n\in\mathcal{Y}^n:\\ \mathcal{D}^{(n)}(y^n)=(\hat{m}_1,\hat{m}_2)}} W^n_{Y^n|X_1^n,X_2^n,S^n}(y^n|x_1^n(m_1),x_2^n(m_2),s^n)Q^n(s^n)}^{\textbf{(3)}}. \tag{22}$$

Now if for the distribution Q, using similar arguments which were used in the proof of Lemma 6, if (19) is satisfied then, it can be shown that the error term **(1)** is greater than equal to $\frac{1}{4}$ or, if (20) is satisfied then, it can be shown that the error term **(2)** is greater than equal to $\frac{1}{4}$ or if (19) is satisfied then, it can be shown that the error term **(1)** is greater than equal to $\frac{1}{4}$ or, if (21) is satisfied then, it can be shown that the error term **(3)** is greater than equal to $\frac{1}{2}$, which proves the necessity of the condition. Now we will show the opposite direction of the proof i.e. the sufficiency of the condition for $int\{\mathcal{R}^*(\mathcal{W}_{\textbf{m.avc}})\}\neq\varnothing$.

2) Sufficiency of the condition for $int\{\mathcal{R}^*(\mathcal{W}_{\textbf{m.avc}})\}\neq\varnothing$: Here we have to show that if $\mathcal{W}_{\textbf{m.avc}}$ does not satisfy any of the sub-conditions given in eqs. (19) to (21), then $int\{\mathcal{R}^*(\mathcal{W}_{\textbf{m.avc}})\}\neq\varnothing$.

We prove this by devising a direct coding setup for the AVMAC $\mathcal{W}_{\textbf{m.avc}}$ and showing the existence of a positive achievable rate. We define average error probability over the choice of a random code-book $\mathcal{C}_n$ for a given state sequence as follows

$$\bar{e}(s^n) := \mathbb{E}_{\mathcal{C}_n}\left[\frac{1}{2^{nR_1}}\frac{1}{2^{nR_2}}\sum_{m_1=1}^{2^{nR_1}}\sum_{m_2=1}^{2^{nR_2}} e(m_1,m_2,\mathcal{C}_n,s^n)\right].$$

From Corollary 1, there exists sets $\mathcal{D}_{X_1},\mathcal{D}_{X_2}$ and $\mathcal{D}_{X_1,X_2}$ which satisfies ther following

$$\mathcal{D}_{X_1} := \left\{\forall s^n\in\mathcal{S}^n: \begin{array}{c} W_{X_1^nX_2^nY^n|S^n}(\mathcal{D}|s^n)\geq 1-\varepsilon,\\ W_{X_2^nY^n|S^n}\times P_{X_1^n}(\mathcal{D}|s^n)\\ \leq 2^{-n\min_Q I_{PQ}[X_1;Y|X_2]+\delta}. \end{array}\right\},$$

$$\mathcal{D}_{X_1} := \left\{\forall s^n\in\mathcal{S}^n: \begin{array}{c} W_{X_1^nX_2^nY^n|S^n}(\mathcal{D}|s^n)\geq 1-\varepsilon,\\ W_{X_1^nY^n|S^n}\times P_{X_2^n}(\mathcal{D}|s^n)\\ \leq 2^{-n\min_Q I_{PQ}[X_2;Y|X_1]+\delta}. \end{array}\right\},$$

$$\mathcal{D}_{X_1,X_2} := \left\{\forall s^n\in\mathcal{S}^n: \begin{array}{c} W_{X_1^nX_2^nY^n|S^n}(\mathcal{D}|s^n)\geq 1-\varepsilon,\\ W_{Y^n|S^n}\times P_{X_1^n}\times P_{X_2^n}(\mathcal{D}|s^n)\\ \leq 2^{-n\min_Q I_{PQ}[X_1X_2;Y]+\delta}. \end{array}\right\}.$$

We set our decoding set to be $\hat{\mathcal{D}} := \mathcal{D}_{X_1}\cap\mathcal{D}_{X_2}\cap\mathcal{D}_{X_1,X_2}$. Then, for a given output sequence $y^n\in\mathcal{Y}^n$, we decode it to a message $\hat{m}_1\in\mathcal{M}_n^1$ and $\hat{m}_2\in\mathcal{M}_n^2$, if $\{x_1^n(\hat{m}_1),x_2^n(\hat{m}_2),y^n\}\in\hat{\mathcal{D}}$. Thus given y^n, the decoding error for a particular message m_1 and m_2

can be of two types.

$$\mathcal{E}_{1,(m_1,m_2)} := \Big\{\{x_1^n(m_1), x_2^n(m_2), y^n\} \notin \hat{\mathcal{D}}\Big\},$$
$$\mathcal{E}_{2,(m_1,m_2)} := \big\{\exists(\hat{m}_1,\hat{m}_2) \neq (m_1,m_2) : \{x_1^n(\hat{m}_1), x_2^n(\hat{m}_2), y^n\} \in \hat{\mathcal{D}}\big\}.$$

Now for any state sequence $s^n \in \mathcal{S}^n$, we can upperbound $\bar{e}(s^n)$ as follows.

$$\begin{aligned}
&\mathbb{E}_{\mathcal{C}}[\frac{1}{2^{nR_1}}\frac{1}{2^{nR_2}}\sum_{m_1=1}^{2^{nR_1}}\sum_{m_2=1}^{2^{nR_2}} e(m_1,m_2,\mathcal{C}_n,s^n)]\\
&\leq \frac{1}{2^{nR_1}}\frac{1}{2^{nR_2}}\sum_{m_1=1}^{2^{nR_1}}\sum_{m_2=1}^{2^{nR_2}} \mathbb{E}_{\mathcal{C}}[P\{\mathcal{E}_{1,(m_1,m_2)}\} + P\{\mathcal{E}_{2,(m_1,m_2)}\}]\\
&\overset{a}{\leq} 3\cdot\varepsilon + 2^{-n(\min_{Q_1} I_{PQ_1}[X_1:Y|X_2]+\delta-R_1)} + 2^{-n(\min_{Q_2} I_{PQ_2}[X_2:Y|X_1]+\delta-R_2)} + 2^{-n(\min_{Q_3} I_{PQ_3}[X_1X_2:Y]+\delta-R_1-R_2)}.
\end{aligned}$$

where, a follows from Corllary 1. So from the above equation, it follows that as $n\to\infty$ and if $R_1 < \min_{Q_1} I_{PQ_1}[X_1:Y|X_2]-\delta$, $R_2 < \min_{Q_2} I_{PQ_2}[X_2:Y|X_1]-\delta$, $R_1+R_2 < \min_{Q_3} I_{PQ_3}[X_1X_2:Y]-\delta$, for some $\mu\in(0,1)$ we have

$$\max_{s^n\in\mathcal{S}^n} \mathbb{E}_{\mathcal{C}}[\frac{1}{2^{nR_1}}\frac{1}{2^{nR_2}}\sum_{m_1=1}^{2^{nR_1}}\sum_{m_2=1}^{2^{nR_2}} e(m_1,m_2,\mathcal{C}_n,s^n)] < \frac{\mu}{3}.$$

Now using Jahn's two-user extension [12, Theorem 1 - Converse Part] of "correlation elimination" technique of Ahlswede [5], we can eliminate the randomness in codewords by showing that there exists a deterministic code $\tilde{\mathcal{C}}_{(n+g(n))}$ of length $(n+g(n))$, which satisfies,

$$\max_{s^n\in\mathcal{S}^n} \bar{e}\Big(\tilde{\mathcal{C}}_{(n+g(n))}, s^n\Big) < \mu,$$

where, $g(n)$ is in polynomial of n. Since $\mathcal{W}_{\textbf{avc}}$ satisfies none of the sub-conditions (19),(20),(21), it follows that $\min_{Q_1} I_{PQ_1}[X_1;Y|X_2] > 0, \min_{Q_2} I_{PQ_2}[X_2;Y|X_1] > 0$ and $\min_{Q_3} I_{PQ_3}[X_1,X_2;Y] > 0$, which means there exists a codebook $\mathcal{C}_n$ for which a non-zero rate-pair for transmitting messages over the channel $\mathcal{W}_{\textbf{m.avc}}$ can be achieved, which in turn proves the sufficiency of the condition for $int\{\mathcal{R}^*(\mathcal{W}_{\textbf{m.avc}})\} \neq \varnothing$. □

VII. Schumacher Compression for Quantum AVS

Consider an arbitrary varying quantum source denoted as $\mathbf{\Omega}$, which is defined as a collection of quantum states $\mathbf{\Omega} := \{\omega_s\}_{s\in\mathcal{S}}$, where $\mathcal{S} := \{1,2,\cdots,|\mathcal{S}|\}$ and $|\mathcal{S}| < \infty$, with each $\omega_s \in \mathcal{D}(\mathcal{A})$. We now define a (n,R,ε) quantum compression code as following:

Definition 8. *A (n,R,ε)-quantum compression code comprises of the following:*

- *A compression channel $\mathcal{E}_{\mathcal{A}^n\to\mathcal{W}}$ that encodes the n-fold system $\mathcal{A}^n$ to a system $\mathcal{W}$.*
- *We assume that system $\mathcal{W}$ is a Hilbert space of dimension 2^{nR}, where R is defined as the compression rate of the code, which can be given as follows:*

$$R := \frac{1}{n}\log|\mathcal{W}|.$$

  In simple words, R can be viewed as the number of qubits compressed per source system.
- *A decompression channel $\mathcal{D}_{\mathcal{W}\to\mathcal{A}^n}$ which decompresses system $\mathcal{W}$ back into original system $\mathcal{A}^n$.*

Now we define an achievable quantum compression rate as follows:

Definition 9 (Achievable Quantum Compression Rate)**.** *A number $R>0$ is called an achievable quantum compression rate if there exists an (n,R,ε)-quantum compression code such that for any sequence $s^n := (s_1,\cdots,s_n)$, $\forall\varepsilon\in(0,1)$ and sufficiently large n,*

$$\Big\|\Psi_{\omega_{s^n}}^{\mathcal{R}\mathcal{A}^n} - \mathcal{I}_{\mathcal{R}}\otimes(\mathcal{D}_{\mathcal{W}\to\mathcal{A}^n}\circ\mathcal{E}_{\mathcal{A}^n\to\mathcal{W}})(\Psi_{\omega_{s^n}}^{\mathcal{R}\mathcal{A}^n})\Big\|_1 \leq \varepsilon,$$

where $\omega_{s^n} := \bigotimes_{i=1}^n \omega_{s_i}$ and $\Psi_{\omega_{s^n}}^{\mathcal{R}\mathcal{A}^n}$ is a purification of the state ω_{s^n} with a reference system $\mathcal{R}$.

***Theorem* 3.** *(Achievability) If $R \geq \max_{\omega\in conv(\mathbf{\Omega})} S(\omega)$, then there exists a (n,R,ε) quantum data compression code wih a compression channel $\mathcal{E}_{\mathcal{A}^n\to\mathcal{W}}$ and a decompression channel $\mathcal{D}_{\mathcal{W}\to\mathcal{A}^n}$ such that for sufficiently large n, $\forall\varepsilon\in(0,1)$,*

$$\max_{s^n\in\mathcal{S}^n} ||\Psi_{\omega_{s^n}}^{\mathcal{R}\mathcal{A}^n} - \mathcal{I}_{\mathcal{R}}\otimes(\mathcal{D}_{\mathcal{W}\to\mathcal{A}^n}\circ\mathcal{E}_{\mathcal{A}^n\to\mathcal{W}})(\Psi_{\omega_{s^n}}^{\mathcal{R}\mathcal{A}^n})||_1 \leq \varepsilon$$

Proof. Say, Alice starts with the state $\omega_{s^n} := \bigotimes_{i=1}^{n} \omega_{s_i}$ and $\Psi_{\omega_{s^n}}^{\mathcal{R}\mathcal{A}^n}$ be it's purification with an inaccesible reference syatem $\mathcal{R}$. Now for each $s \in \mathcal{S}$, spectral decomposition of ω_s can be given as:

$$\omega_s := \sum_{u \in \mathcal{U}} P_{\mathcal{U},s}(u) |u\rangle\overset{\mathcal{A}}{\langle} u|,$$

where $\{|u\rangle\}_{u\in\mathcal{U}}$ is an orthonormal basis of $\mathcal{A}$ and $\{P_{\mathcal{U},s}\}_{s\in\mathcal{S}}$ is a collection of probability distribution over a support $\mathcal{U}$. Before proceeding into further analysis, we check the following lemma:

Lemma 8. *Consider a collection $\mathbf{\Omega} := \{\omega_s\}_{s\in\mathcal{S}} \subset \mathcal{D}(\mathcal{A})$, where $\mathcal{S}$ is a non-empty finite set, then $\forall \varepsilon \in (0,1)$,*

$$\lim_{n\to\infty} \frac{1}{n} \log(\gamma_n(\mathbf{\Omega}_n, \varepsilon)) = \max_{\omega \in conv(\mathbf{\Omega})} S(\omega),$$

where

$$\gamma_n(\mathbf{\Omega}_n, \varepsilon) := \min_{\substack{0\le\Pi_{\mathfrak{E}}\le\mathbb{1}^{\otimes n}:\\ \mathrm{Tr}[\Pi_{\mathfrak{E}}\rho]\ge 1-\varepsilon; \forall \rho\in\mathbf{\Omega}_n}} \dim(\Pi_{\mathfrak{E}}),$$

and $\mathbf{\Omega}_n := \{\omega_{s^n}\}_{s^n\in\mathcal{S}^n}$ with $\omega_{s^n} := \otimes_{i=1}^{n}\omega_{s_i}, \forall s_i \in \mathcal{S}, \omega_{s_i} \in \mathbf{\Omega}$.

Proof. See Appendix IX-H for the proof. □

From Lemma 8, for sufficiently large n, there exists a projector $\Pi_{\mathcal{A}^n}$ such that for some $\delta \in (0,1), \lambda > 0$

$$\begin{aligned} \mathrm{Tr}[\Pi_{\mathcal{A}^n}\omega_{s^n}] &\ge 1-\delta, \quad \forall s^n \in \mathcal{S}^n, \\ \dim(\Pi_{\mathcal{A}^n}) &= \max_{\omega\in conv(\mathbf{\Omega})} S(\omega) + \lambda. \end{aligned}$$

From the above discussion we know that $\Pi_{\mathcal{A}^n}$ projects onto a subspace of dimension at most $2^{n\left[\max_{\omega\in conv(\mathbf{\Omega})} S(\omega)+\lambda\right]}$, where for a number $a \in \mathbb{R}$, $\lfloor a \rfloor$ represents the closest integer smaller than or equal to a. For the sake of simplicity, we can write $\Pi_{\mathcal{A}^n}$ as follows,

$$\Pi_{\mathcal{A}^n} := \sum_{u^n \in T_{\mathcal{A}^n}} \overset{\mathcal{A}^n}{|u^n\rangle\langle u^n|},$$

where we denote $T_{\mathcal{A}^n}$ as the typical index set which gives us the indices of the basis vectors residing within subset of the orthonormal basis of $\mathcal{A}^n$, that spans the subspace the projector $\Pi_{\mathcal{A}^n}$ projects onto. Now consider a encoding function $\hat{\mathcal{E}} : T_{\mathcal{A}^n} \to \{0,1\}^{n\left[\max_{\omega\in conv(\mathbf{\Omega})} S(\omega)+\lambda\right]}$ which maps the set of typical indices of $T_{\mathcal{A}^n}$ to a set of binary sequences of length $n\left[\max_{\omega\in conv(\mathbf{\Omega})} S(\omega)+\lambda\right]$.

So we can design an linear map $U_{\hat{\mathcal{E}}} : \mathcal{A}^n \to \mathcal{W}$, which maps the orthonormal basis $\left\{|u^n\rangle^{\mathcal{A}^n}\right\}_{u^n\in\mathcal{U}^n}$ of $\mathcal{A}^n$ to the basis $\left\{|\hat{\mathcal{E}}(u^n)\rangle^{\mathcal{W}}\right\}$ of another Hilbert space $\mathcal{W}$, in the following way:

$$U_{\hat{\mathcal{E}}} := \sum_{u^n\in T_{\mathcal{A}^n}} |\hat{\mathcal{E}}(u^n)\rangle^{\mathcal{W}}\langle u^n|^{\mathcal{A}^n}.$$

It follows that the dimension of the Hilbert space $\mathcal{W}$ should be at most $2^{n\left[\max_{\omega\in conv(\mathbf{\Omega})} S(\omega)+\lambda\right]}$.

Now Alice designs her compression strategy by applying the map $U_{\hat{\mathcal{E}}}$ conditioned on the measurement by the projector $\Pi_{\mathcal{A}^n}$ which, for a density operator $Z_{\mathcal{A}^n} \in \mathcal{L}(\mathcal{A}^n)$ can be written as follows:

$$\mathcal{E}_{\mathcal{A}^n\to\mathcal{W}}(Z_{\mathcal{A}^n}) := U_{\hat{\mathcal{E}}}\Pi_{\mathcal{A}^n} Z_{\mathcal{A}^n}\Pi_{\mathcal{A}^n}U_{\hat{\mathcal{E}}}^{\dagger} + \mathrm{Tr}[(\mathbb{I}_{\mathcal{A}^n} - \Pi_{\mathcal{A}^n})Z_{\mathcal{A}^n}]\sigma_{\mathcal{W}},$$

where $\sigma_{\mathcal{W}}$ is some fixed density operator over $span\left\{|\hat{\mathcal{E}}(u^n)\rangle^{\mathcal{W}} : u^n \in T_{\mathcal{A}^n}\right\}$. Alice now sends the $\mathcal{W}$ part of $\mathcal{E}_{\mathcal{A}^n\to\mathcal{W}}\left(\Psi_{\omega_{s^n}}^{\mathcal{R}\mathcal{A}^n}\right)$ to Bob using $n\left[\max_{\omega\in conv(\mathbf{\Omega})} S(\omega)+\lambda\right]$ instances of a noiseless quantum channel.

Bob, on the other hand, is aware of the linear map $U_{\hat{\mathcal{E}}}$ and designs its decompression channel $\mathcal{D}_{\mathcal{W}\to\mathcal{A}^n}$ by performing the inverse of the linear map $U_{\hat{\mathcal{E}}}$, which for a density operator $V_{\mathcal{W}} \in \mathcal{L}(\mathcal{W})$ can be written as follows:

$$\mathcal{D}_{\mathcal{W}\to\mathcal{A}^n}(V_{\mathcal{W}}) := U_{\hat{\mathcal{E}}}^{\dagger}V_{\mathcal{W}}U_{\hat{\mathcal{E}}} + \mathrm{Tr}\left[\left(\mathbb{I}_{\mathcal{W}} - U_{\hat{\mathcal{E}}}U_{\hat{\mathcal{E}}}^{\dagger}\right)V_{\mathcal{W}}\right]\mu_{\mathcal{A}^n},$$

where, $\mu_{\mathcal{A}^n}$ is some fixed density operator in $\mathcal{A}^n$. Bob then applies the decompression channel over the $\mathcal{W}$ part of $\mathcal{E}_{\mathcal{A}^n\to\mathcal{W}}\left(\Psi_{\omega_{s^n}}^{\mathcal{R}\mathcal{A}^n}\right)$ which can be written as following:

$$\begin{aligned} &\mathcal{I}_{\mathcal{R}} \otimes (\mathcal{D}_{\mathcal{W}\to\mathcal{A}^n} \circ \mathcal{E}_{\mathcal{A}^n\to\mathcal{W}})(\Psi_{\omega_{s^n}}^{\mathcal{R}\mathcal{A}^n}) \\ &= \mathcal{I}_{\mathcal{R}} \otimes \mathcal{D}_{\mathcal{W}\to\mathcal{A}^n}\left(\left(\mathcal{I}_{\mathcal{R}} \otimes U_{\hat{\mathcal{E}}}\Pi_{\mathcal{A}^n}\right)\Psi_{\omega_{s^n}}^{\mathcal{R}\mathcal{A}^n}\left(\mathcal{I}_{\mathcal{R}} \otimes \Pi_{\mathcal{A}^n}U_{\hat{\mathcal{E}}}^{\dagger}\right) + \mathrm{Tr}_{\mathcal{A}^n}[(\mathcal{I}_{\mathcal{R}} \otimes (\mathbb{I}_{\mathcal{A}^n} - \Pi_{\mathcal{A}^n}))\Psi_{\omega_{s^n}}^{\mathcal{R}\mathcal{A}^n}] \otimes \sigma_{\mathcal{W}}\right) \\ &= (\mathcal{I}_{\mathcal{R}} \otimes \Pi_{\mathcal{A}^n})\Psi_{\omega_{s^n}}^{\mathcal{R}\mathcal{A}^n}(\mathcal{I}_{\mathcal{R}} \otimes \Pi_{\mathcal{A}^n}) + \mathrm{Tr}_{\mathcal{A}^n}[(\mathcal{I}_{\mathcal{R}} \otimes (\mathbb{I}_{\mathcal{A}^n} - \Pi_{\mathcal{A}^n}))\Psi_{\omega_{s^n}}^{\mathcal{R}\mathcal{A}^n}] \otimes \mathcal{D}_{\mathcal{W}\to\mathcal{A}^n}(\sigma_{\mathcal{W}}). \end{aligned}$$

Then, for any $s^n \in \mathcal{S}^n$ we have,

$$\begin{aligned}
&\left\| \Psi_{\omega_{s^n}}^{\mathcal{R}\mathcal{A}^n} - \mathcal{I}_{\mathcal{R}} \otimes (\mathcal{D}_{\mathcal{W}\to\mathcal{A}^n} \circ \mathcal{E}_{\mathcal{A}^n\to\mathcal{W}})(\Psi_{\omega_{s^n}}^{\mathcal{R}\mathcal{A}^n}) \right\|_1 \\
&\overset{a}{\leq} \left\| \Psi_{\omega_{s^n}}^{\mathcal{R}\mathcal{A}^n} - (\mathcal{I}_{\mathcal{R}} \otimes \Pi_{\mathcal{A}^n}) \Psi_{\omega_{s^n}}^{\mathcal{R}\mathcal{A}^n} (\mathcal{I}_{\mathcal{R}} \otimes \Pi_{\mathcal{A}^n}) \right\|_1 + \left\| \mathrm{Tr}_{\mathcal{A}^n}[(\mathcal{I}_{\mathcal{R}} \otimes (\mathbb{I}_{\mathcal{A}^n} - \Pi_{\mathcal{A}^n})) \Psi_{\omega_{s^n}}^{\mathcal{R}\mathcal{A}^n}] \otimes \mathcal{D}_{\mathcal{W}\to\mathcal{A}^n}(\sigma_{\mathcal{W}}) \right\|_1 \\
&\overset{b}{\leq} 2\sqrt{\delta} + \left\| \mathrm{Tr}_{\mathcal{A}^n}[(\mathcal{I}_{\mathcal{R}} \otimes (\mathbb{I}_{\mathcal{A}^n} - \Pi_{\mathcal{A}^n})) \Psi_{\omega_{s^n}}^{\mathcal{R}\mathcal{A}^n}] \right\|_1 \cdot \left\| \mathcal{D}_{\mathcal{W}\to\mathcal{A}^n}(\sigma_{\mathcal{W}}) \right\|_1 \\
&\leq 2\sqrt{\delta} + \left\| \mathrm{Tr}_{\mathcal{A}^n}[(\mathcal{I}_{\mathcal{R}} \otimes (\mathbb{I}_{\mathcal{A}^n} - \Pi_{\mathcal{A}^n})) \Psi_{\omega_{s^n}}^{\mathcal{R}\mathcal{A}^n}] \right\|_1 \\
&\leq 2\sqrt{\delta} + \mathrm{Tr}[(\mathcal{I}_{\mathcal{R}} \otimes (\mathbb{I}_{\mathcal{A}^n} - \Pi_{\mathcal{A}^n})) \Psi_{\omega_{s^n}}^{\mathcal{R}\mathcal{A}^n}] \\
&= 2\sqrt{\delta} + \mathrm{Tr}[(\mathbb{I}_{\mathcal{A}^n} - \Pi_{\mathcal{A}^n}) \omega_{s^n}] \\
&\overset{c}{\leq} 2\sqrt{\delta} + \delta,
\end{aligned}$$

where a follows from Fact 3, b follows from Fact 6 given the fact $\mathrm{Tr}\big[(\mathcal{I}_{\mathcal{R}} \otimes \Pi_{\mathcal{A}^n}) \Psi_{\omega_{s^n}}^{\mathcal{R}\mathcal{A}^n}\big] = \mathrm{Tr}[\Pi_{\mathcal{A}^n} \omega_{s^n}] \geq 1 - \delta$ and c also follows from the fact that $\mathrm{Tr}[\Pi_{\mathcal{A}^n} \omega_{s^n}] \geq 1 - \delta$. □

VIII. Conclusion and Acknowledgements

In this work, we have provided a more intuitive proof of achievability for [8, Theorem 4.1] in Lemma 1, which uses the concept of **frequency typical sets**. This lemma seems much more likely to be quantizable since Nötzel in [11] has already solved the problem of hypothesis testing between an arbitrarily varying quantum source and a discrete-memoryless quantum source with the help of **frequency typical subspaces** which is a generalization of frequency typical sets in quantum setting. We also have tried to achieve a result similar to Lemma 1 for classical-quantum states. But the best we could do is the hypothesis testing between a classical-quantum arbitrarily varying source and a classical-quantum compound source in Lemma 2. We also have shown that we are very close to having a result for hypothesis testing between two arbitrarily varying classical-quantum sources and provided the reason why we failed to do so. As an application of 1 we have shown an achievability for deriving the channel capacity of a classical point-to-point arbitrarily varying channel (AVC) in Section V and an achievability for finding the capacity region of a classical arbitrarily varying multiple access channel (AVMAC) in Section VI.

As a future work, one can try to extend the result mentioned in Lemma 2 for solving the problem of hypothesis testing between two arbitrarily varying classical-quantum sources mentioned in Conjecture 1. Solving the above will provide a solution for deriving the capacity of a classical-quantum AVC and the capacity region of a classical-quantum AVMAC in the eye of hypothesis testing. One can also try to extend the problem of hypothesis testing between two arbitrarily varying sources in a fully-quantum paradigm with or without entanglement assistance.

References

[1] J. Kiefer and J. Wolfowitz, "Channels with arbitrarily varying channel probability functions," *Information and Control*, vol. 5, no. 1, pp. 44–54, 1962. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0019995862902036

[2] T. Berger, *Rate Distortion Theory and Data Compression*. Vienna: Springer Vienna, 1975, pp. 1–39. [Online]. Available: https://doi.org/10.1007/978-3-7091-2928-9_1

[3] R. Ahlswede, "A note on the existence of the weak capacity for channels with arbitrarily varying channel probability functions and its relation to shannon's zero error capacity," *The Annals of Mathematical Statistics*, vol. 41, no. 3, pp. 1027–1033, 1970. [Online]. Available: http://www.jstor.org/stable/2239255

[4] ——, "Channels with arbitrarily varying channel probability functions in the presence of noiseless feedback," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, vol. 25, no. 3, pp. 239–252, Sep 1973. [Online]. Available: https://doi.org/10.1007/BF00535895

[5] ——, "Elimination of correlation in random codes for arbitrarily varying channels," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, vol. 44, pp. 159–175, June 1978.

[6] T. Ericson, "Exponential error bounds for random codes in the arbitrarily varying channel," *IEEE Transactions on Information Theory*, vol. 31, no. 1, pp. 42–48, 1985.

[7] I. Csiszar and P. Narayan, "The capacity of the arbitrarily varying channel revisited: positivity, constraints," *IEEE Transactions on Information Theory*, vol. 34, no. 2, pp. 181–193, 1988.

[8] F. Fu and S. Shen, "Hypothesis testing for arbitrarily varying source with exponential-type constraint," *IEEE Trans. Inf. Theory*, vol. 44, no. 2, pp. 892–895, 1998. [Online]. Available: https://doi.org/10.1109/18.661541

[9] N. A. Warsi, "Simple one-shot bounds for various source coding problems using smooth rényi quantities," *Problems of Information Transmission*, vol. 52, no. 1, pp. 39–65, Jan 2016. [Online]. Available: https://doi.org/10.1134/S0032946016010051

[10] A. Anshu, R. Jain, and N. A. Warsi, "A unified approach to source and message compression," 2019.

[11] J. Nötzel, "Hypothesis testing on invariant subspaces of the symmetric group: part i. quantum sanov's theorem and arbitrarily varying sources," *Journal of Physics A: Mathematical and Theoretical*, vol. 47, no. 23, p. 235303, may 2014. [Online]. Available: https://dx.doi.org/10.1088/1751-8113/47/23/235303

[12] J.-H. Jahn, "Coding of arbitrarily varying multiuser channels," *IEEE Transactions on Information Theory*, vol. 27, no. 2, pp. 212–226, March 1981.

[13] J. A. Gubner, "On the deterministic-code capacity of the multiple-access arbitrarily varying channel," *IEEE transactions on information theory*, vol. 36, no. 2, pp. 262–275, 1990.

[14] R. Ahlswede and N. Cai, "Arbitrarily varying multiple-access channels. i. ericson's symmetrizability is adequate, gubner's conjecture is true," *IEEE Transactions on Information Theory*, vol. 45, no. 2, pp. 742–749, 1999.

[15] R. Ahlswede and V. Blinovsky, "Classical capacity of classical-quantum arbitrarily varying channels," *IEEE Transactions on Information Theory*, vol. 53, no. 2, pp. 526–533, 2007.

[16] P. Sen, "Unions, intersections and a one-shot quantum joint typicality lemma," *Sādhanā*, vol. 46, pp. 1–44, 2021. [Online]. Available: https://api.semanticscholar.org/CorpusID:229377137

[17] R. Jozsa and B. Schumacher, "A new proof of the quantum noiseless coding theorem," *Journal of Modern Optics*, vol. 41, no. 12, pp. 2343–2349, 1994. [Online]. Available: https://doi.org/10.1080/09500349414552191
[18] B. Schumacher, "Quantum coding," *Phys. Rev. A*, vol. 51, pp. 2738–2747, Apr 1995. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevA.51.2738
[19] J. Gao, "Quantum union bounds for sequential projective measurements," *Physical Review A*, vol. 92, p. 052331, Nov. 2015. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevA.92.052331
[20] R. O'Donnell and R. Venkateswaran, "The quantum union bound made easy," in *Symposium on Simplicity in Algorithms (SOSA)*, pp. 314–320. [Online]. Available: https://epubs.siam.org/doi/abs/10.1137/1.9781611977066.25
[21] A. Winter, "The capacity of the quantum multiple-access channel," *IEEE Transactions on Information Theory*, vol. 47, no. 7, pp. 3059–3065, 2001.
[22] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Hoboken, NJ, USA: Wiley, 2006.

IX. Appendix

A. Proof of Lemma 4:

For a particular sequence $\hat{s}^n$, we have a particular type $N : \mathcal{S} \to \mathbb{N}$, which $\hat{s}^n$ falls in, and the corresponding empirical (frequency) distribution $p(s) = \frac{N(s|\hat{s}^n)}{n}$, where for any $s \in \mathcal{S}, N(s|\hat{s}^n)$ denotes the number of appearnces of s in $\hat{s}^n$. Consider $\widehat{P} = \sum_{s\in\mathcal{S}} p(s)P_s$ and thus $\widehat{P}^n = \sum_{s^n\in\mathcal{S}^n} p(s^n)P_{s^n}$.

We will first show that if two sequences s^n, t^n are in same type class N, then

$$\sum_{x^n\in T_f} P_{s^n}(x^n) = \sum_{x^n\in T_f} P_{t^n}(x^n), \tag{23}$$

where $T_f \subset \mathcal{X}^n$ is collection of sequences in $\mathcal{X}^n$ which falls in type class f. Since s^n and t^n are in same type class, there exists a permutation τ such that $s^n := \tau(t^n)$. Then,

$$\begin{aligned}
&\sum_{x^n\in T_f} P_{t^n}(x^n) = \sum_{x^n\in T_f} P_{t_1}(x_1)P_{t_2}(x_2)\dots P_{t_n}(x_n)\\
&= \sum_{x^n\in T_f} P_{t_{\tau^{-1}(1)}}(x_{\tau^{-1}(1)})P_{t_{\tau^{-1}(2)}}(x_{\tau^{-1}(2)})\dots P_{t_{\tau^{-1}(n)}}(x_{\tau^{-1}(n)})\\
&= \sum_{x^n\in T_f} P_{s_1}(x_{\tau^{-1}(1)})P_{s_2}(x_{\tau^{-1}(2)})\dots P_{s_n}(x_{\tau^{-1}(n)})\\
&\overset{a}{=} \sum_{\tau^{-1}(x^n)\in T_f} (P_{s_1}(x_{\tau^{-1}(1)})P_{s_2}(x_{\tau^{-1}(2)})\dots P_{s_n}(x_{\tau^{-1}(n)}))\\
&= \sum_{x^n\in T_f} P_{s^n}(x^n)
\end{aligned}$$

where, a follows since $(x_1,\dots,x_n) \in T_f$ if and only if $(x_{\tau^{-1}(1)},\dots,x_{\tau^{-1}(n)}) \in T_f$. Now we have,

$$\begin{aligned}
\widehat{P}^n(A_n^c) &\geq \sum_{s^n\in T_N} p(s^n)P_{s^n}(A_n^c)\\
&\overset{a}{=} \sum_{s^n\in T_N} p(s^n)P_{\hat{s}^n}(A_n^c)\\
&\geq (2n)^{-|\mathcal{S}|} P_{\hat{s}^n}(A_n^c),
\end{aligned}$$

where, a follows from the fact that the complement of a permutation invariant set is also permutation invariant and (23). Therefore, we have $P_{\hat{s}^n}(A_n^c) \leq (2n)^{|\mathcal{S}|}\widehat{P}^n(A_n^c)$.

Now, before going into the proof let's see the following two claims:

Claim 1. *Let $P_1,..,P_k$ be k probability distributions over $\mathcal{X}$ and consider $\widehat{P} := \sum_{i=1}^k \lambda_i P_i$, where $\lambda_i \in [0,1], \forall i \in \{1,\cdots,k\}$ and $\sum_{i=1}^k \lambda_i = 1$. If A is a permutation invariant subset of $\mathcal{X}^n$, then*

$$\widehat{P}^n(A) \geq \min_{s\in K} P_s^n(A),$$

where $K := \{1,\cdots,k\}$.

Proof. See Appendix IX-B for the proof. □

Lemma 9. *Let $\mathcal{S}$ be a non-empty finite set, $\{P_1, P_2, .., P_{|\mathcal{S}|}\} \subset \mathcal{P}(\mathcal{X})$ be a collection of probability distribution on a finite non-empty set $\mathcal{X}$ and another probability distribution $Q \in \mathcal{P}(\mathcal{X})$. Then, we have,*

$$\begin{aligned}
\min_{s\in\mathcal{S}} P_s^n(A_n) &\geq 1 - 2^{-n(\frac{\varepsilon^2}{2} - \frac{d}{n}\log(2n))},\\
Q^n(A_n) &\leq 2^{-n(\min_{s\in\mathcal{S}} D(P_s||Q) - \Theta(n,\varepsilon,d,Q))},
\end{aligned}$$

where,

$$\Theta(n,\varepsilon,d,Q) := \frac{d}{n}\log(2n) + \varepsilon\left|\log\frac{\varepsilon}{d}\right| + d.\varepsilon \max_i |\log(Q(i))|.$$

Proof. See Appendix IX-C for the proof. □

Now,

$$\begin{aligned}
P_{\hat{s}^n}(A_n) &= 1 - P_{\hat{s}^n}(A_n^c) \\
&\geq 1 - (2n)^{|\mathcal{S}|}\hat{P}^n(A_n^c) \\
&\overset{a}{\geq} 1 - (2n)^{|\mathcal{S}|}(1 - \min_s P_s^n(A_n)) \\
&\overset{b}{\geq} 1 - (2n)^{|\mathcal{S}|}2^{-n(\frac{\varepsilon^2}{2} - \frac{d}{n}\log(2n))} \\
&= 1 - 2^{-n(\frac{\varepsilon^2}{2} - \frac{d+|\mathcal{S}|}{n}\log(2n))}.
\end{aligned}$$

where, a follows from Claim 1 and b follows from Lemma 9. Also from Lemma 9 we have,

$$\begin{aligned}
Q^n(A_n) &\leq 2^{-n(\min_{s\in S} D(P_s||Q) - \Theta(n,\varepsilon,d,Q))} \\
&\leq 2^{-n(\min_{P\in\hat{\Sigma}} D(P||Q) - \Theta(n,\varepsilon,d,Q))}
\end{aligned}$$

B. Proof of Lemma 1:

Note that as A is permutation invariant, it suffices to prove the result for a singleton set $A := \{(x_1, ..., x_n)\}$. Without loss of generality, assume,

$$\min_{s\in\mathcal{S}} P_s^n(x_1, ..., x_n) = P_u^n(x_1, ..., x_n),$$

for some $u \in \mathcal{S}$. For the ease of calculation, we denote $P_i(x_j) := a_{i,j}, \forall i \in K, j \in \{1, \cdots, n\}$. Consider an index $b(\cdot,.) : K \times \{1, \cdots, n\} \to \{1, \cdots, n\}$, where for sequences $x^n := (x_1, \cdots, x_n)$ and $P_{s^n} := P_{s_1} \times \cdots \times P_{s_n}$, $b(v, w)$ returns the position in x^n where P_v acts on while appearing for w-th time in P_{s^n}. Thus given $a_{u,1}..a_{u,n} \leq a_{l,1}..a_{l,n}$ for any $1 \leq l \leq k$, we need to show

$$a_{u,1}..a_{u,n} \leq \widehat{P}^n(x_1, ..., x_n) = \sum_{\substack{\{\alpha_1,\cdots,\alpha_k\},\\ \{b(u,v)\}_{u,v}\\ :\sum_{i\in K}\alpha_i = n}} \lambda_1^{\alpha_1}..\lambda_k^{\alpha_k} a_{1,b(1,1)}..a_{1,b(1,\alpha_1)}...a_{k,b(k,1)}..a_{k,b(k,\alpha_k)},$$

where for each $i \in K, \alpha_i$ denotes the appearances of P_i. We will first show that for a fixed k tuple $(\alpha_1, ..., \alpha_k) : \sum_{i\in K}\alpha_i = n$,

$$\sum_{\{b(u,v)\}_{u,v}} a_{1,b(1,1)}..a_{1,b(1,\alpha_1)}...a_{k,b(k,1)}..a_{k,b(k,\alpha_k)} \geq \binom{n}{\alpha_1, .., \alpha_k} a_{1,1}..a_{1,n}. \tag{24}$$

Notice that there are $\binom{n}{\alpha_1,..,\alpha_k}$ terms in the summation above and we denote it by L. Now applying AM-GM inequality we get,

$$\begin{aligned}
\sum_{\{b(u,v)\}_{u,v}} a_{1,b(1,1)}..a_{1,b(1,\alpha_1)}...a_{k,b(k,1)}..a_{k,b(k,\alpha_k)} &\overset{a}{\geq} L\sqrt[L]{(a_{1,1}..a_{1,n})^{\frac{L\alpha_1}{n}}....(a_{k,1}..a_{k,n})^{\frac{L\alpha_k}{n}}} \\
&\geq L\sqrt[L]{(a_{1,1}..a_{1,n})^{\frac{L\alpha_1}{n}}....(a_{1,1}..a_{1,n})^{\frac{L\alpha_k}{n}}} \\
&= L\sqrt[L]{(a_{1,1}..a_{1,n})^{L}} \\
&= \binom{n}{\alpha_1, .., \alpha_k} a_{1,1}..a_{1,n},
\end{aligned}$$

where a follows from the fact that $\forall i \in K, j \in \{1, \cdots, n\}$ repetitions of $a_{i,j}$ is given as $\binom{n-1}{\alpha_1,\cdots,\alpha_i-1,\cdots,\alpha_n}$. Now by multinomial theorem, we have,

$$\widehat{P}^n(x_1, ..., x_n) \overset{a}{\geq} \sum_{\substack{\{\alpha_1,\cdots,\alpha_k\},\\ :\sum_{i\in K}\alpha_i = n}} \lambda_1^{\alpha_1}..\lambda_k^{\alpha_k}\binom{n}{\alpha_1, .., \alpha_k} a_{1,1}..a_{1,n} = a_{1,1}..a_{1,n}\left(\sum_{i=1}^{k}\lambda_i\right)^n = a_{1,1}..a_{1,n},$$

where a follows from (24).

C. Proof of Lemma 9:

For each $s \in \mathcal{S}$, we define $A_{n,s} = \bigcup_{f:\|\overline{f}-P_s\|<\varepsilon} T_f$. Note that if $\|\overline{f} - P_s\| \geq \varepsilon$, we have,

$$\begin{aligned} P_s^n(T_f) &\overset{a}{\leq} 2^{-nD(\overline{f}||P_s)} \\ &\overset{b}{\leq} 2^{-\frac{n\varepsilon^2}{2}}, \end{aligned}$$

where, a follows from Theorem 11.1.4 of [22] and b follows from Fact 4. Now,

$$\begin{aligned} 1 &= P_s^n(\cup_f T_f) \\ &= P_s^n(A_{n,s}) + P_s^n(A_{n,s}^c) \\ &\overset{a}{\leq} P_s^n(A_n) + (2n)^d \max_{f:\|\overline{f}-P_s\|\geq\varepsilon} P_s^n(T_f) \\ &\leq P_s^n(A_n) + (2n)^d.2^{-\frac{n\varepsilon^2}{2}}, \end{aligned}$$

where, a folows from the fact $A_{n,s} \subseteq A_n$. Hence we have,

$$\min_{s\in\mathcal{S}} P_s^n(A_n) \geq 1 - 2^{-n(\frac{\varepsilon^2}{2} - \frac{d}{n}\log(2n))}.$$

Now for any $\overline{f} \in \Lambda_\varepsilon$, we have,

$$\begin{aligned} &Q^n(T_f) = |T_f|.Q(1)^{f(1)}Q(2)^{f(2)}\cdots Q(|\mathcal{X}|)^{f(|\mathcal{X}|)} \\ &\leq 2^{n(H(\overline{f})+\sum_{i\in\mathcal{X}} \frac{1}{n} f(i)\log(Q(i)))} \\ &\overset{a}{\leq} 2^{n(H(P_s)+\varepsilon|\log\frac{\varepsilon}{d}|+\sum_i P_s(i)\log(Q(i))+d.\varepsilon\max_i|\log(Q(i))|)} \\ &= 2^{-n(D(P_s||Q)-\varepsilon|\log\frac{\varepsilon}{d}|-d.\varepsilon\max_i\ |\log(Q(i))|)}, \end{aligned}$$

where, a follows from the fact that $\exists s \in \mathcal{S} : |H(\overline{f}) - H(P_s)| \leq -\varepsilon\log\frac{\varepsilon}{d}$ as $\|\overline{f} - P_s\| \leq \varepsilon$. Hence we conclude,

$$\begin{aligned} Q^n(A_n) &\leq (2n)^d \max_f Q^n(T_f) \\ &\leq 2^{-n(\min_{s\in\mathcal{S}} D(P_s||Q)-\Theta(n,\varepsilon,d,Q))}, \end{aligned}$$

where,

$$\Theta(n,\varepsilon,d,Q) = \frac{d}{n}\log(2n) + \varepsilon|\log\frac{\varepsilon}{d}| + d.\varepsilon\max_i|\log(Q(i))|.$$

D. Proof of Corollary 2:

Application of Lemma 9's result on Δ_1 and a probability distribution $\widehat{Q}_q := \sum_{s\in\mathcal{S}} q(s)Q_s$ (where $q \in T_n(\mathcal{S})$ is an empirical distribution with denominator n), implies existence of a permutation invariant set B_q such that,

$$\min_{p\in T_n(\mathcal{S})} \widehat{P}_p^n(B_q) \geq 1 - 2^{-n(\frac{\varepsilon^2}{2} - \frac{d}{n}\log(2n))}, \tag{25}$$

$$\begin{aligned} \widehat{Q}_q^n(B_q) &\leq 2^{-n\left(\min_{p\in T_n(\mathcal{S})} D(\widehat{P}_p||\widehat{Q}_q)-\Theta(n,\varepsilon,d,\widehat{Q}_q)\right)} \\ &\leq 2^{-n\left(\min_{P\in\Delta_1} D(P||\widehat{Q}_q)-\Theta(n,\varepsilon,d,\widehat{Q}_q)\right)}, \end{aligned} \tag{26}$$

where $\widehat{P}_p := \sum_{s\in\mathcal{S}} p(s)P_s$. For a particular sequence $\hat{s}^n \in \mathcal{S}^n$ with its corresponding empirical (frequency) distribution $\hat{p} \in T_n(\mathcal{S})$, we have a set B_q,

$$\begin{aligned} P_{\hat{s}^n}(B_q) &= 1 - P_{\hat{s}^n}(B_q^c) \\ &\overset{a}{\geq} 1 - (2n)^{|\mathcal{S}|}\widehat{P}_{\hat{p}}^n(B_q^c) \\ &\geq 1 - \max_{p\in T_n(\mathcal{S})} (2n)^{|\mathcal{S}|}\widehat{P}_p^n(B_q^c) \\ &\overset{b}{\geq} 1 - 2^{-n(\frac{\varepsilon^2}{2} - \frac{d+|\mathcal{S}|}{n}\log(2n))}, \end{aligned}$$

where a follows from the fact that B_q^c, which is the complement of a permutation invariant set B_q, is also permutation invariant and (23) and b follows from (25). Now consider the set,

$$B := \bigcap_{q\in T_n(\mathcal{S})} B_q.$$

Observe that, B is a permutation invariant set. For a particular sequence $\bar{s}^n$, with its corresponding empirical (frequency) distribution $\hat{q}(s)$, we can write the following:

$$\begin{aligned} Q_{\bar{s}^n}(B) &\overset{a}{\leq} (2n)^{|\mathcal{S}|}\widehat{Q}^n_{\hat{q}}(B) \\ &\leq \max_{q\in T_n(\mathcal{S})}(2n)^{|\mathcal{S}|}\widehat{Q}^n_q(B) \\ &\overset{b}{\leq} (2n)^{|\mathcal{S}|}\max_{q\in T_n(\mathcal{S})}\widehat{Q}^n_q(B_q) \\ &\overset{c}{\leq} (2n)^{|\mathcal{S}|}\max_{q\in T_n(\mathcal{S})}2^{-n\left(\min_{P\in\Delta_1}D(P||\widehat{Q}_q)-\Theta(n,\varepsilon,d,\widehat{Q}_q)\right)} \\ &= 2^{-n\left(\min_{P\in\Delta_1,Q\in\Delta_2}D(P||Q)-\frac{|\mathcal{S}|\log(2n)}{n}-\underline{\Theta}(n,\varepsilon,d,\Delta_2)\right)}. \end{aligned}$$

E. Proof of Corollary 1:

Consider $\hat{\mathcal{W}}_1 := conv\{\mathcal{W}_1\}$ and $\hat{\mathcal{W}}_2 := conv\{\mathcal{W}_2\}$. Now for any set $\mathcal{D} \subset X^n \times Y^n$, the hypothesis testing errors can be written as .

$$\begin{aligned} \overline{\alpha}(\mathcal{D}) &:= \max_{s^n\in\mathcal{S}^n} W_{X^nY^n|S^n}(\mathcal{D}^c|s^n), \\ \overline{\beta}(\mathcal{D}) &:= \max_{s^n\in\mathcal{S}^n} W_{Y^n|S^n}\times P_{X^n}(\mathcal{D}|s^n). \end{aligned}$$

Thus by theorem 2.1, we get, if $\hat{\mathcal{W}}_1 \cap \hat{\mathcal{W}}_2 = \phi$, then for all $0 < \varepsilon < 1$, we have

$$\lim_{n\to\infty} -\frac{1}{n}log\ \overline{\beta}_n(\varepsilon) = \min_{P\in\hat{\mathcal{W}}_1,Q\in\hat{\mathcal{W}}_2} D(P||Q)] = \min_{\lambda,\beta\in\mathfrak{B}(\mathcal{S})} D\Big(\sum_{s\in\mathcal{S}}\lambda(s)W_{XY|S}(\cdot,\cdot|s)||\sum_{s\in\mathcal{S}}\beta(s)W_{Y|S}(\cdot|s)\times P_X(\cdot)\Big),$$

where Note that by expanding the definition of KL divergence we get that the quantity $D(P_{XY}||P_X \times Q_Y)$ is minimized when $Q_Y = P_Y$. Thus the above equality boils down to:

$$\lim_{n\to\infty} -\frac{1}{n}log\ \overline{\beta}_n(\varepsilon) = \min_Q I_{PQ}[X;Y].$$

Hence we conclude, for every $\delta > 0$, and sufficintly large n, there exists a set $\hat{\mathcal{D}} \subset \mathcal{X}^n \times \mathcal{Y}^n$ such that $\forall s^n \in S^n$,

$$\begin{aligned} W_{X^nY^n|S^n}(\mathcal{D}|s^n) &\geq 1-\varepsilon, \\ W_{Y^n|S^n}\times P_{X^n}(\mathcal{D}|s^n) &\leq 2^{-n\min_Q I_{PQ}[X;Y]+\delta}. \end{aligned}$$

F. Proof of Lemma 5:

Recently, Sen in [16], resolved the intersection issue for the case of classical-quantum states and their marginals respectively, using "tilting" and "augmenting" the Hilbert space to a significantly larger Hilbert space. Consider $\mathcal{L}_1, \mathcal{L}_2, \cdots, \mathcal{L}_M$ be M distinct mutually orthogonal Hilbert spaces isomorphic to a Hilbert space $\mathcal{L}$, where for each $m = 1, \cdots, M, |\mathcal{L}_m| = |\mathcal{L}|$ and $\mathcal{L}_m$ is orthogonal to $\mathcal{B}$ and for each $m \in \{1, \cdots, M\}$ the Hilbert space $\mathcal{L}_m$ is given as:

$$\mathbb{I}^{\mathcal{L}_m} := \sum_{l_m}^{\mathcal{L}_m} |l_m\rangle\langle l_m|,$$

where, $\{|l_m\rangle\}$ is the collection of the computational basis set of $\mathcal{L}_m$, for each $l = 1, 2, \cdots, M$. We have tilting maps $T_{m,l_m,\delta} : \mathcal{B} \to \mathcal{B}_m$, which is a isometric embedding from $\mathcal{B}$ to a significantly larger Hilbert space $\mathcal{B}_m$, where $\mathcal{B}_m := \mathcal{B} \oplus \mathcal{B} \otimes \mathcal{L}_m$ defined as follows:

$$T_{m,l_m,\delta} : |h\rangle \longmapsto \frac{1}{\sqrt{1+\delta^2}}(|h\rangle + \delta|h\rangle|l_m\rangle),$$

where $|l_m\rangle$ is one particular computational basis vectors of $\mathcal{L}_m$. We represent this map $T_{m,l_m,\delta}$ as it tilts a vector within $\mathcal{B}$ towards a particular orthogonal direction in $\mathcal{L}_m$. We also have another tilting map $T_{[M],l_{[M]},\delta} : \mathcal{B} \to \mathcal{B}'$ (here, $[M] :=$ $\{1, \cdots, M\}$), which is a isometric embedding from $\mathcal{B}$ to a significantly larger Hilbert space $\mathcal{B}'$, where $\mathcal{B}' := \mathcal{B} \oplus \bigoplus_{m=1}^M (\mathcal{B} \otimes \mathcal{L}_m)$ and $l_{[M]} := \{l_1, \cdots, l_M\}$ with each $m = 1, \cdots, M$, $|l_m\rangle$ being a particular computational basis vector of $\mathcal{L}_m$ respectively, defined as follows :

$$T_{[M],l_{[M]},\delta} : |h\rangle \longmapsto \frac{1}{\sqrt{1+M.\delta^2}}(|h\rangle + \sum_{m=1}^M \delta|h\rangle|l_m\rangle).$$

We represent the above tilting map as it tilts a vector within $\mathcal{B}$ towards $\mathcal{L}_m$ for each $m = 1, \cdots, M$ orthogonally.

1) Tilting the quantum states: Now for a particular $x^n \in X^n$, for each $k = 1, \cdots, K$, we tilt the state $\rho^{\mathcal{B}}_{k,x^n}$ towards $\{|l_1\rangle, \cdots, |l_M\rangle\}$ in the following manner:

$$\rho^{\mathcal{B}'}_{k,x^n,l_{[M]}} := T_{[M],l_{[M]},\delta}\big(\rho^{\mathcal{B}}_{k,x^n}\big).$$

So for each $k = 1, \cdots, K$ we can tilt the state $\rho_k^{X^n\mathcal{B}}$ after augmenting the input Hilbert space in the following way:

$$\rho'_k := \frac{1}{|\mathcal{L}|^M} \sum_{x^n, l_{[M]}} p(x^n) \overset{X^n}{|x^n\rangle\langle x^n|} \otimes \bigotimes_{m=1}^{M} \overset{\mathcal{L}_m}{|l_m\rangle\langle l_m|} \otimes \rho^{\mathcal{B}'}_{k,x^n,l_{[M]}}.$$

Here, for each x^n and each $k = 1, \cdots, K$ we tilt the state $\rho^{\mathcal{B}}_{k,x^n}$ towards all possible $\{|l_1\rangle, \cdots, |l_M\rangle\}$ and we need to augment or enlarge the input Hilbert space to incorporate all such tilted states in a single classical-quantum state. We perform tilting and augmentation over the states $\big\{\sigma_m^{X^n\mathcal{B}}\big\}_{m\in[M]}$ similarly and for each $m = 1, \cdots, M$ we get the following states:

$$\sigma'_m := \frac{1}{|\mathcal{L}|^M} \sum_{x^n, l_{[M]}} p(x^n) \overset{X^n}{|x^n\rangle\langle x^n|} \otimes \bigotimes_{m=1}^{M} \overset{\mathcal{L}_m}{|l_m\rangle\langle l_m|} \otimes \sigma^{\mathcal{B}'}_{m,l_{[M]}},$$

where $\sigma^{\mathcal{B}'}_{m,l_{[M]}} := T_{[M],l_{[M]},\delta}\big(\sigma^{\mathcal{B}}_m\big)$. It also follows that for each pair of $(l_{[M]})$,

$$\Big\|\rho^{\mathcal{B}'}_{k,x^n,l_{[M]}} - \rho^{\mathcal{B}}_{k,x^n}\Big\|_1 \leq 4\delta^2, \quad \forall k = 1, \cdots, K, x^n \in \mathcal{X}^n,$$
$$\Big\|\sigma^{\mathcal{B}'}_{m,l_{[M]}} - \sigma^{\mathcal{B}}_m\Big\|_1 \leq 4\delta^2, \quad \forall m = 1, \cdots, M.$$

2) Design of the "Intersection Projector": For each $m = 1, \cdots, M$ and $\forall x^n \in X^n$, we define W'_{m,x^n} be the orthogonal complement of the support of $\Pi^{\mathcal{B}}_{m,x^n}$. The objective is to prepare a projector that behaves like an intersection of $\Pi^{\mathcal{B}}_{1,x^n}, \cdots, \Pi^{\mathcal{B}}_{M,x^n}$ and the idea is to take the complement of the span of the complements of the supports of $\Pi^{\mathcal{B}}_{1,x^n}, \cdots, \Pi^{\mathcal{B}}_{M,x^n}$ respectively, which has a flavor similar to taking an intersection of two or more sets in classical scenario. But this idea has a drawback. There is a possibility that the span of the supports may cover the whole Hilbert space causing the complement of the span a zero subspace. So we need to tilt the supports very minutely in a larger Hilbert space toward the different orthogonal directions and then take the span so that the span of the tilted supports doesn't cover the larger hilbert space. So for each $m = 1, \cdots, M$ we tilt W'_{m,x^n} in the following way:

$$W'_{m,x^n,l_{[M]},\delta} := T_{m,l_m,\delta}\big(W'_{m,x^n}\big).$$

Then, we define the subspace,

$$W'_{[M],x^n,l_{[M]},\delta} := \overset{M}{\underset{m=1}{+}} W'_{m,x^n,l_{[M]},\delta},$$

to be the span of $W'_{1,x^n,l_{[M]},\delta}, \cdots, W'_{M,x^n,l_{[M]},\delta}$ (the operator '$+$' is denoted as the span of subspaces) and let $(\Pi')_{W'_{[M],x^n,l_{[M]},\delta}}$ be the orthogonal projection in $\mathcal{B}'$ onto $W'_{[M],x^n,l_{[M]},\delta}$. So now for each possible $x^n \in X^n, |l_1\rangle, \cdots, |l_m\rangle$ we have the corresponding projector as follows:

$$(\Pi')^{\mathcal{B}'}_{[M],x^n,l_{[M]},\delta} := \Big(\mathbb{I}^{\mathcal{B}'} - (\Pi')_{W'_{[M],x^n,l_{[M]},\delta}}\Big)\Pi^{\mathcal{B}'}_{\mathcal{B}}\Big(\mathbb{I}^{\mathcal{B}'} - (\Pi')_{W'_{[M],x^n,l_{[M]},\delta}}\Big),$$

where $\Pi^{\mathcal{B}'}_{\mathcal{B}}$ is an orthogonal projection in $\mathcal{B}'$ to $\mathcal{B}$(in simple words, it is the identity operator of the Hilbert space $\mathcal{B}$ in the larger Hilbert space $\mathcal{B}'$). Now we finally have to augment the input space to incorporate the collection of the projectors $\Big\{(\Pi')^{\mathcal{B}'}_{[M],x^n,l_{[M]},\delta}\Big\}_{x^n,l_{[M]}}$ in the following way:

$$\Pi_{\cap M} := \sum_{x^n, l_{[M]}} \overset{X^n}{|x^n\rangle\langle x^n|} \otimes \bigotimes_{m=1}^{M} \overset{\mathcal{L}_m}{|l_m\rangle\langle l_m|} \otimes (\Pi')^{\mathcal{B}'}_{[M],x^n,l_{[M]},\delta}.$$

Now before proceeding further into proving the desired properties of the above projector, we first observe the following result:

3) Smoothness of the tilted state: Consider a vector $|h\rangle \in \mathcal{B}$ and we apply the tilting map $T_{[M],l_{[M]},\delta}$. Then, we have,

$$\begin{aligned}
&\frac{1}{|\mathcal{L}|^{M-1}} \sum_{l_{[M]}\setminus l_1} T_{[M],l_{[M]},\delta}(|h\rangle\langle h|) \\
&= \frac{1}{|\mathcal{L}|^{M-1}(1+M.\delta^2)} \sum_{l_{[M]}\setminus l_1} \left(|h\rangle + \sum_{m=1}^{M} \delta|h\rangle|l_m\rangle\right)\left(\langle h| + \sum_{m=1}^{M} \delta\langle h|\langle l_m|\right) \\
&= \frac{1}{1+M.\delta^2}(|h\rangle + \delta|h\rangle|l_1\rangle)(\langle h| + \delta\langle h|\langle l_1|) + \frac{1}{|\mathcal{L}|^{M-1}(1+M.\delta^2)} \sum_{l_{[M]}\setminus l_1} \Big(\sum_{m=2}^{M} \delta(|h\rangle + \delta|h\rangle|l_1\rangle)\langle h|\langle l_m| \\
&\quad + \sum_{m=2}^{M} \delta|h\rangle|l_m\rangle(\langle h| + \delta\langle h|\langle l_1|) + \delta^2 \sum_{\substack{m=2\\ m'=2}}^{M,M} |h\rangle\langle h||l_m\rangle\langle l_{m'}|\Big) \\
&= \frac{1+\delta^2}{1+M\delta^2} T_{1,l_1,\delta}(|h\rangle\langle h|) + N_{1,l_1,\delta}(|h\rangle\langle h|),
\end{aligned}$$

where for any $m \in [M]$,

$$\begin{aligned}
N_{m,l_m,\delta}(|h\rangle\langle h|) := \frac{1}{|\mathcal{L}|^{M-1}(1+M\delta^2)} \sum_{l_{[M]}\setminus l_m} \Big(\delta(|h\rangle + \delta|h\rangle|l_1\rangle) \sum_{\hat{m}\neq m}^{M} \langle h|\langle l_{\hat{m}}| + \sum_{\hat{m}\neq m}^{M} \delta|h\rangle|l_{\hat{m}}\rangle(\langle h| + \delta\langle h|\langle l_1|) \\
+ \delta^2 \sum_{\substack{\hat{m}\neq m\\ m'\neq m}}^{M,M} |h\rangle\langle h||l_{\hat{m}}\rangle\langle l_{m'}|\Big).
\end{aligned}$$

We now show that for each $m \in [M]$, $N_{m,l_m,\delta}(|h\rangle\langle h|)$ is very small in terms of l_∞-norm i.e.

$$\begin{aligned}
\|N_{m,l_m,\delta}(|h\rangle\langle h|)\|_\infty &\overset{a}{\leq} \frac{1}{|\mathcal{L}|^{M-1}(1+M\delta^2)} \Bigg(\left\|\sum_{l_{[M]}\setminus l_m} \delta(|h\rangle + \delta|h\rangle|l_1\rangle) \sum_{\hat{m}\neq m}^{M} \langle h|\langle l_{\hat{m}}|\right\|_\infty + \left\|\sum_{l_{[M]}\setminus l_m} \sum_{\hat{m}\neq m}^{M} \delta|h\rangle|l_{\hat{m}}\rangle(\langle h| + \delta\langle h|\langle l_1|)\right\|_\infty \\
&\qquad + \left\|\sum_{l_{[M]}\setminus l_m} \delta^2 \sum_{\substack{\hat{m}\neq m\\ m'\neq m}}^{M,M} |h\rangle\langle h||l_{\hat{m}}\rangle\langle l_{m'}|\right\|_\infty\Bigg) \\
&\leq \frac{1}{|\mathcal{L}|^{M-1}(1+M\delta^2)} \left(2\delta.\||h\rangle + \delta|h\rangle|l_1\rangle\|_2 \cdot \sqrt{\sum_{\hat{m}\neq m}^{M} \left\|\sum_{l_{\hat{m}}} |l_{\hat{m}}\rangle\right\|_2^2} + O(\delta^2)\right) \\
&\leq \frac{O(\delta)\sqrt{(M-1).|\mathcal{L}|}}{|\mathcal{L}|^{M-1}(1+M\delta^2)} \left(\||h\rangle + \delta|h\rangle|l_1\rangle\|_2\right) \\
&\leq \frac{O(\delta)\sqrt{(M-1)}}{|\mathcal{L}|^{M-\frac{3}{2}}},
\end{aligned}$$

where a follows from Fact 2. Now we have the following properties for the tilted state i.e for each $m \in [M]$,

$$\frac{1}{|\mathcal{L}|^{M-1}} \sum_{l_{[M]}\setminus l_m} \sigma^{\mathcal{B}'}_{m,l_{[M]}} = T_{m,l_m,\delta}(\sigma^{\mathcal{B}}_m) + N_{m,l_m,\delta}(\sigma^{\mathcal{B}}_m). \tag{27}$$

Thus from (27), we can conclude that although a quantum state is tilted along two mutually orthogonal directions, its reduced (traced-out) state is primarily tilted along only one of those directions. This phenomenon is called **smoothing** a traced-out quantum state over particularly one of the directions. We now proceed to check whether we can achieve properties similar to (9) and (10) for the intersection projector $\Pi_{\cap M}$ with the tilted states $\{\rho'_j\}_{j=1,2}$ and $\{\sigma'_m\}_{m\in[M]}$.

4) Property testing of the "Intersection Projector": Let's first start with the upper-bound of error probability of not getting the actual state i.e. one of the states from the collection $\{\rho'_k\}_{k\in[K]}$(where, $[K] := \{1, \cdots, K\}$):

$$
\begin{aligned}
&\mathrm{Tr}[(\mathbb{I} - \Pi_{\cap M})\rho'_k] \\
&= |\mathcal{L}|^{-M} \sum_{x^n, l_{[M]}} p(x^n) \mathrm{Tr}\Big[\Big(\mathbb{I} - (\Pi')^{\mathcal{B}'}_{[M],x^n,l_{[M]},\delta}\Big)\rho^{\mathcal{B}'}_{k,x^n,l_{[M]}}\Big] \\
&\leq 4\delta^2 + |\mathcal{L}|^{-M} \sum_{x^n, l_{[M]}} p(x^n) \mathrm{Tr}\Big[\Big(\mathbb{I} - (\Pi')^{\mathcal{B}'}_{[M],x^n,l_{[M]},\delta}\Big)\rho^{\mathcal{B}}_{k,x^n}\Big] \\
&\overset{a}{\leq} 4\delta^2 + 4|\mathcal{L}|^{-M} \sum_{x^n, l_{[M]}} p(x^n) \Bigg(\mathrm{Tr}\Big[\Big(\mathbb{I} - \Pi^{\mathcal{B}'}_{\mathcal{B}}\Big)\rho^{\mathcal{B}}_{k,x^n}\Big] \\
&\qquad\qquad + \mathrm{Tr}\Big[(\Pi')_{W'_{[M],x^n,l_{[M]},\delta}}\rho^{\mathcal{B}}_{k,x^n}\Big]\Bigg) \\
&\overset{b}{\leq} 4\delta^2 + 4|\mathcal{L}|^{-M} 3.2. \frac{1 - \frac{\delta^2}{1+\delta^2}}{\frac{\delta^2}{1+\delta^2}} \sum_{m=1}^{M} \sum_{x^n, l_{[M]}} p(x^n) \\
&\qquad\qquad \Big(1 - \mathrm{Tr}\big[\Pi^{\mathcal{B}}_{m,x^n}\rho^{\mathcal{B}}_{k,x^n}\big]\Big) \\
&\leq 4\delta^2 + \frac{24}{\delta^2} \sum_{m=1}^{M} 1 - \mathrm{Tr}[\Pi^{X^n\mathcal{B}}_m \rho^{X^n\mathcal{B}}_k] \\
&\leq 4\delta^2 + \frac{24.M.\tilde{\varepsilon}}{\delta^2},
\end{aligned}
$$

where, a follows from Fact 5 and b follows from [16, Corollary 1]. Now if we choose $\delta := \tilde{\varepsilon}^{\frac{1}{4}}$, we have,

$$\mathrm{Tr}[\Pi_{\cap M}\rho'] \geq 1 - (24.M + 4).\sqrt{\tilde{\varepsilon}}.$$

Now we find the upper-bound of error probability of getting the wrong state i.e. one of the states from the collection $\{\sigma'_m\}_{m\in[M]}$,

$$
\begin{aligned}
&\mathrm{Tr}[\Pi_{\cap M}\sigma'_m] = |\mathcal{L}|^{-M} \sum_{x^n, l_{[M]}} p(x^n) \mathrm{Tr}\Big[(\Pi')^{\mathcal{B}'}_{[M],x^n,l_{[M]},\delta}\sigma^{\mathcal{B}'}_{m,l_{[M]}}\Big] \\
&= |\mathcal{L}|^{-M} \sum_{x^n, l_{[M]}} p(x^n) \mathrm{Tr}\Big[\Big(\mathbb{I} - (\Pi')^{\mathcal{B}'}_{W'_{m,x^n,l_{[M]},\delta}}\Big)\Pi^{\mathcal{B}'}_{\mathcal{B}}\Big(\mathbb{I} - (\Pi')^{\mathcal{B}'}_{W'_{m,x^n,l_{[M]},\delta}}\Big)\sigma^{\mathcal{B}'}_{m,l_{[M]}}\Big] \\
&= \frac{1}{|\mathcal{L}|} \sum_{x^n, l_m} p(x^n) \mathrm{Tr}\Bigg[\Big(\mathbb{I} - (\Pi')^{\mathcal{B}'}_{W'_{m,x^n,l_{[M]},\delta}}\Big)\Pi^{\mathcal{B}'}_{\mathcal{B}}\Big(\mathbb{I} - (\Pi')^{\mathcal{B}'}_{W'_{m,x^n,l_{[M]},\delta}}\Big)\Bigg(|\mathcal{L}|^{-(M-1)} \sum_{l_{[M]}\setminus l_m} \sigma^{\mathcal{B}'}_{m,l_{[M]}}\Bigg)\Bigg] \\
&\overset{a}{\leq} \frac{1}{|\mathcal{L}|} \sum_{x^n, l_m} p(x^n) \mathrm{Tr}\Big[\Big(\mathbb{I} - (\Pi')^{\mathcal{B}'}_{W'_{m,x^n,l_{[M]},\delta}}\Big)T_{m,l_m,\delta}\big(\sigma^{\mathcal{B}}_m\big)\Big] + \Big\|\Big(\mathbb{I} - (\Pi')^{\mathcal{B}'}_{W'_{m,x^n,l_{[M]},\delta}}\Big)\Pi^{\mathcal{B}'}_{\mathcal{B}}\Big(\mathbb{I} - (\Pi')^{\mathcal{B}'}_{W'_{m,x^n,l_{[M]},\delta}}\Big)N_{m,l_m,\delta}\big(\sigma^{\mathcal{B}}_m\big)\Big\|_1 \\
&\overset{b}{\leq} \frac{1}{|\mathcal{L}|} \sum_{x^n, l_m} p(x^n) \mathrm{Tr}\Big[\Big(\mathbb{I} - (\Pi')^{\mathcal{B}'}_{W'_{m,x^n,l_{[M]},\delta}}\Big)T_{m,l_m,\delta}\big(\sigma^{\mathcal{B}}_m\big)\Big] + \Big\|\Big(\mathbb{I} - (\Pi')^{\mathcal{B}'}_{W'_{m,x^n,l_{[M]},\delta}}\Big)\Big\|^2_\infty \Big\|\Pi^{\mathcal{B}'}_{\mathcal{B}}\Big\|_1 \big\|N_{m,l_m,\delta}\big(\sigma^{\mathcal{B}}_m\big)\big\|_\infty \\
&\leq \frac{1}{|\mathcal{L}|} \sum_{x^n, l_m} p(x^n) \mathrm{Tr}\Big[\Big(\mathbb{I} - (\Pi')^{\mathcal{B}'}_{T_{m,l_m,\delta}(W'_{m,x^n})}\Big)T_{m,l_m,\delta}\big(\sigma^{\mathcal{B}}_m\big)\Big] + \frac{O\Big(\tilde{\varepsilon}^{\frac{1}{4}}\Big).|\mathcal{B}|\sqrt{(M-1)}}{|\mathcal{L}|^{M-\frac{3}{2}}} \\
&\overset{c}{=} \frac{1}{|\mathcal{L}|} \sum_{x^n, l_m} p(x^n) \mathrm{Tr}\Big[\Big(\mathbb{I} - (\Pi')^{\mathcal{B}'}_{T_{m,l_m,\delta}(W'_{m,x^n})}\Big)\Pi^{\mathcal{B}'}_{T_{m,l_m,\delta}(\mathcal{B})}T_{m,l_m,\delta}\big(\sigma^{\mathcal{B}}_m\big)\Big] + \frac{O\Big(\tilde{\varepsilon}^{\frac{1}{4}}\Big).|\mathcal{B}|\sqrt{(M-1)}}{|\mathcal{L}|^{M-\frac{3}{2}}} \\
&= \frac{1}{|\mathcal{L}|} \sum_{x^n, l_m} p(x^n) \mathrm{Tr}\Bigg[\Big(\mathbb{I}^{T_{m,l_m,\delta}(\mathcal{B})} - (\Pi')^{T_{m,l_m,\delta}(\mathcal{B})}_{T_{m,l_m,\delta}(W'_{m,x^n})}\Big)T_{m,l_m,\delta}\big(\sigma^{\mathcal{B}}_m\big)\Bigg] + \frac{O\Big(\tilde{\varepsilon}^{\frac{1}{4}}\Big).|\mathcal{B}|\sqrt{(M-1)}}{|\mathcal{L}|^{M-\frac{3}{2}}} \\
&\overset{d}{=} \frac{1}{|\mathcal{L}|} \sum_{x^n, l_m} p(x^n) \mathrm{Tr}\Big[\Big(\mathbb{I}^{\mathcal{B}} - (\Pi')^{\mathcal{B}}_{W'_{m,x^n}}\Big)\sigma^{\mathcal{B}}_m\Big] + \frac{O\Big(\tilde{\varepsilon}^{\frac{1}{4}}\Big).|\mathcal{B}|\sqrt{(M-1)}}{|\mathcal{L}|^{M-\frac{3}{2}}}
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{1}{|\mathcal{L}|}\sum_{x^n,l_m} p(x^n)\mathrm{Tr}\big[\Pi^{\mathcal{B}}_{m,x^n}\sigma^{\mathcal{B}}_m\big] + \frac{O\Big(\tilde{\varepsilon}^{\frac{1}{4}}\Big).|\mathcal{B}|\sqrt{(M-1)}}{|\mathcal{L}|^{M-\frac{3}{2}}} \qquad (28)\\
&= \mathrm{Tr}[\Pi^{X^n\mathcal{B}}_m\sigma^{X^n\mathcal{B}}_m] + \frac{O\Big(\tilde{\varepsilon}^{\frac{1}{4}}\Big).|\mathcal{B}|\sqrt{(M-1)}}{|\mathcal{L}|^{M-\frac{3}{2}}}\\
&\leq 2^{-k_m} + \frac{O\Big(\tilde{\varepsilon}^{\frac{1}{4}}\Big).|\mathcal{B}|\sqrt{(M-1)}}{|\mathcal{L}|^{M-\frac{3}{2}}},
\end{aligned}
$$

where, a follows from Fact 6, b follows from Fact 2, c follows from the fact $\Pi^{\mathcal{B}'}_{T_{m,l_m,\delta}(\mathcal{B})}$ is a projector in $\mathcal{B}'$ onto $T_{m,l_m,\delta}(\mathcal{B})$ and d follows from the fact that $T_{m,l_m,\delta}$ is an isometry. Now, we can choose $|\mathcal{L}|$ large enough so that the second term in the above inequality is less than $\min_{m\in[M]}\big\{2^{-k_m}\big\}$. Finally, for each $i=1,2$, we have the following inequality:

$$\mathrm{Tr}[\Pi_{\cap M}\sigma'_m] \leq 2^{-k_m+1}.$$

G. Proof of Lemma 3

We assume the proof of Lemma 2 and Consider the following state,

$$(\sigma')^{\mathcal{X}'\mathcal{B}'}_{\bar{s}^{\otimes n}} := \frac{1}{|\mathcal{L}|^{|\mathcal{S}|}}\sum_{x^n\in\mathcal{X}^n,l_{\mathcal{S}}} p(x^n)|x^n\rangle\langle x^n| \otimes \left(\bigotimes_{i=1}^{|\mathcal{S}|}|l_i\rangle\langle l_i|^{\mathcal{L}_i}\right)\otimes\tilde{\rho}^{\mathcal{B}'}_{l_{\mathcal{S}},\bar{s}^{\otimes n}},$$

where, $\tilde{\rho}^{\mathcal{B}'}_{l_{\mathcal{S}},\bar{s}^{\otimes n}} := T_{\mathcal{S},l_{\mathcal{S}},\delta}\Big(\big(\sum_s q(s)\rho^{\mathcal{B}}_s\big)^{\otimes n}\Big)$. Now we use the similar techniques that were used in proving the upper bound of error probability discussed in IX-F4, as follows:

$$
\begin{aligned}
\mathrm{Tr}\Big[\Big(\mathbb{I}^{\mathcal{X}'\mathcal{B}'} - (\Pi')^{\mathcal{X}'\mathcal{B}'}\Big)(\sigma')^{\mathcal{X}'\mathcal{B}'}_{\bar{s}^{\otimes n}}\Big] &\overset{a}{\geq} \frac{1}{|\mathcal{L}|}\sum_{x^n,l_s} p(x^n)\mathrm{Tr}\left[\Big(\mathbb{I}^{\mathcal{B}^n} - \Pi^{\mathcal{B}^n}_{x^n,s}\Big)\left(\sum_s q(s)\rho^{\mathcal{B}}_s\right)^{\otimes n}\right] - \frac{O\Big(\bar{\varepsilon}^{\frac{1}{4}}\Big).|\mathcal{B}|\sqrt{(|\mathcal{S}|-1)}}{|\mathcal{L}|^{|\mathcal{S}|-\frac{3}{2}}}\\
&\overset{b}{\geq} \min_{s\in\mathcal{S}}\frac{1}{|\mathcal{L}|}\sum_{x^n,l_s} p(x^n)\mathrm{Tr}\Big[\Big(\mathbb{I}^{\mathcal{B}^n} - \Pi^{\mathcal{B}^n}_{x^n,s}\Big)\rho^{\mathcal{B}^{\otimes n}}_s\Big] - \frac{O\Big(\bar{\varepsilon}^{\frac{1}{4}}\Big).|\mathcal{B}|\sqrt{(|\mathcal{S}|-1)}}{|\mathcal{L}|^{|\mathcal{S}|-\frac{3}{2}}}\\
&= 1 - \max_{s\in\mathcal{S}}\mathrm{Tr}[\Pi^{X^n\mathcal{B}^n}_s\sigma^{X\mathcal{B}^{\otimes n}}_s] - \frac{O\Big(\bar{\varepsilon}^{\frac{1}{4}}\Big).|\mathcal{B}|\sqrt{(|\mathcal{S}|-1)}}{|\mathcal{L}|^{|\mathcal{S}|-\frac{3}{2}}}\\
\Rightarrow \mathrm{Tr}\Big[(\Pi')^{\mathcal{X}'\mathcal{B}'}(\sigma')^{\mathcal{X}'\mathcal{B}'}_{\bar{s}^{\otimes n}}\Big] &\leq 2^{-n\left\{\min\limits_{\substack{\rho\in\mathrm{conv}\{\mathcal{G}'\}\\ s\in\mathcal{S}}} D(\rho\|\sigma)-\Theta(n,\varepsilon,|\mathcal{B}|,\sigma)-\frac{|\mathcal{S}|}{n}\log(2n)+1\right\}} + \frac{O\Big(\bar{\varepsilon}^{\frac{1}{4}}\Big).|\mathcal{B}|\sqrt{(|\mathcal{S}|-1)}}{|\mathcal{L}|^{|\mathcal{S}|-\frac{3}{2}}}\\
&\leq 2^{-n\left\{\min\limits_{\substack{\rho\in\mathrm{conv}\{\mathcal{G}'\}\\ \sigma\in\mathrm{conv}\{\mathcal{J}'\}}} D(\rho\|\sigma)-\Theta(n,\varepsilon,|\mathcal{B}|,\sigma)+\frac{|\mathcal{S}|}{n}\log(2n)+1\right\}} - \frac{O\Big(\bar{\varepsilon}^{\frac{1}{4}}\Big).|\mathcal{B}|\sqrt{(|\mathcal{S}|-1)}}{|\mathcal{L}|^{|\mathcal{S}|-\frac{3}{2}}},
\end{aligned}
$$

where, a follows from (28), b follows from the similar arguments used in Claim 1, $\mathcal{J}' := \big\{\sigma^{X\mathcal{B}}_s\big\}_{s\in\mathcal{S}}$. We choose $|\mathcal{L}|$ large enough so that the second term in the above inequality is less than the first term. Finally, we have the following inequality:

$$\mathrm{Tr}\Big[(\Pi')^{\mathcal{X}'\mathcal{B}'}(\sigma')^{\mathcal{X}'\mathcal{B}'}_{\bar{s}^{\otimes n}}\Big] \leq 2^{-n\left\{\min\limits_{\substack{\rho\in\mathrm{conv}\{\mathcal{G}'\}\\ \sigma\in\mathrm{conv}\{\mathcal{J}'\}}} D(\rho\|\sigma)-\Theta(n,\varepsilon,|\mathcal{B}|,\sigma)-\frac{|\mathcal{S}|}{n}\log(2n)+2\right\}}.$$

This proves Lemma 3.

H. Proof of Lemma 8

Applying [11, Theorem 1] over the collection $\mathbf{\Omega}_n$ and a quantum state $\tau := \frac{1}{\dim(\mathcal{H})}\mathbb{I}_{\mathcal{H}}$ we have $\forall\varepsilon\in(0,1)$,

$$\lim_{n\to\infty}\beta_n(\mathbf{\Omega}_n,\tau,\varepsilon) = -\min_{\omega\in conv(\mathbf{\Omega})} D(\omega\|\tau),$$

where,

$$\beta_n(\mathbf{\Omega}_n, \tau, \varepsilon) := \min_{\substack{0\leq\Pi\leq\mathbb{I}_{H^{\otimes n}}: \\ \mathrm{Tr}[\Pi\omega]\geq 1-\varepsilon; \forall\omega\in\mathcal{D}(\mathcal{H})}} \mathrm{Tr}[\Pi\tau^{\otimes n}] = \frac{1}{\dim(\mathcal{H})^n} \min_{\substack{0\leq\Pi\leq\mathbb{I}_{H^{\otimes n}}: \\ \mathrm{Tr}[\Pi\omega]\geq 1-\varepsilon; \forall\omega\in\mathcal{D}(\mathcal{H})}} \mathrm{Tr}[\Pi]$$

$$= \frac{\gamma_n(\mathbf{\Omega}_n, \varepsilon)}{\dim(\mathcal{H})^n}$$

$$\Rightarrow \frac{1}{n}\log(\gamma_n(\mathbf{\Omega}_n, \varepsilon)) = \frac{1}{n}\log(\beta_n(\mathbf{\Omega}_n, \tau, \varepsilon)) + \log(\dim(\mathcal{H})).$$

Hence for sufficiently large n, we can write the following:

$$\begin{aligned}
&\lim_{n\to\infty}\frac{1}{n}\log(\gamma_n(\mathbf{\Omega}_n, \varepsilon)) \\
&= -\min_{\omega\in conv(\mathbf{\Omega})} D(\omega||\tau) + \log(\dim(\mathcal{H})) \\
&= -\min_{\omega\in conv(\mathbf{\Omega})}\left(\mathrm{Tr}\left[\omega\left(\log(\omega) - \log\left(\frac{1}{\dim(\mathcal{H})}\mathbb{I}_{\mathcal{H}}\right)\right)\right]\right) \\
&\quad + \log(\dim(\mathcal{H})) \\
&= -\min_{\omega\in conv(\mathbf{\Omega})}\left(-S(\omega) - \mathrm{Tr}\left[\omega\log\left(\frac{1}{\dim(\mathcal{H})}\mathbb{I}_{\mathcal{H}}\right)\right]\right) \\
&\quad + \log(\dim(\mathcal{H})) \\
&= -\min_{\omega\in conv(\mathbf{\Omega})}\left(-S(\omega) - \log\left(\frac{1}{\dim(\mathcal{H})}\right)\mathrm{Tr}[\omega]\right) \\
&\quad + \log(\dim(\mathcal{H})) \\
&= \max_{\omega\in conv(\mathbf{\Omega})} S(\omega).
\end{aligned}$$