Soft-covering via Constant-composition Superposition codes

Sreejith Sreekumar and Ziv Goldfeld
Dept. of Electrical and Computer Engineering, Cornell University

2021 IEEE International Symposium on Information Theory (ISIT) | 978-1-5386-8209-8/21/$31.00 ©2021 IEEE | DOI: 10.1109/ISIT45174.2021.9518258

***Abstract*—We consider the problem of soft-covering with constant composition superposition codes and characterize the optimal soft-covering exponent. A double-exponential concentration bound for deviation of the exponent from its mean is also established. We demonstrate an application of the result to achieving the secrecy-capacity region of a broadcast channel with confidential messages under a per-codeword cost constraint. This generalizes the recent characterization of the wiretap channel secrecy-capacity under an average cost constraint, highlighting the potential utility of the superposition soft-covering result to the analysis of coding problems.**

I. Introduction

Finding its roots in Wyner's seminal paper [1], soft-covering (also known as channel resolvability [2]) is by now an ubiquitous tool in information theory. It refers to the problem of simulating a target distribution by passing a uniformly chosen codeword through a noisy channel. Simulation can be attained to any desired accuracy, typically measured by the total variation (TV) distance or the Kullback-Leibler (KL) divergence, provided that the coding rate exceeds the channel input-output mutual information. The ability to simulate distributions turns out useful in various applications, including physical layer security [3]–[9], channel synthesis [10], lossy compression [11], covert communication [12], [13], and privacy [14].

Motivated by applications to multiuser scenarios with input cost constraints, we study soft-covering by superposition codes, whose inner and outer layer codewords are chosen uniformly from a constant composition ensemble [15]. We characterize the optimal soft-covering exponent, i.e., the maximum asymptotic exponential rate of the expected TV distance between the distribution induced by the codebook and a target (average) distribution. We further establish a double exponential concentration bound for the probability of deviation of this TV distance from its mean. The soft-covering results are leveraged to establish the the secrecy-capacity region of a broadcast channel (BC) with confidential messages under a per-codeword cost constraint. The capacity region recovers the secrecy-capacity of a cost constrained (CC) wiretap channel as a special case, whose characterization was recently shown in [9] to require two auxiliaries in general, even under a less stringent per-message cost constraint.

The work of S. Sreekumar is supported by the TRIPODS Center for Data Science NSF Grant CCF-1740822. The work of Z. Goldfeld is supported in part by the NSF CRII Award under Grant CCF-1947801, in part by the 2020 IBM Academic Award, and in part by the NSF CAREER Award under Grant CCF-2046018.

A. Background

The bulk of soft-covering literature focuses on single-layer random codebooks. The fundamental limit of the codebook size needed to achieve soft-covering was established in [2] for the TV distance. Lower bounds on the soft-covering exponents achievable over memoryless channels under the TV distance and the KL divergence were obtained in [3]. The TV lower bound was further improved in [10], where extensions of soft-covering to more general channels was also considered. Soft-covering in the quantum context was first explored in [16], [17], with the latter pointing out that it also holds for KL divergence (see also [18]). Double exponential concentration bounds on the deviation of KL divergence or TV distance from their means were obtained in [5], [19] and [20], respectively. More recently, [21] and [22] characterized exact soft-covering exponents with respect to (w.r.t.) KL divergence and TV distance, respectively. While the above works mostly focus on the i.i.d. ensemble, soft-covering for constant composition codebooks were studied in [21] under KL divergence and in [22] under TV distance. To the best of our knowledge, the only extensions of the soft-covering phenomena to superposition codes were given in [10] and [8], both of which focus on i.i.d. codebooks and derive achievable rates as well as concentration inequalities, but not exact exponents.

B. Notation

We use standard notation (cf. e.g., [9]). In particular, for a countable $\mathcal{X}$, the letter-typical set of n-lengthed sequences w.r.t. a probability mass function (PMF) $P\in\mathcal{P}(\mathcal{X})$ and $\delta>0$ is

$$\mathcal{T}_{\delta}^{(n)}(P):=\{\mathbf{x}\in\mathcal{X}^n:|\nu_{\mathbf{x}}(x)-P(x)|\leq\delta P(x),\ \forall x\in\mathcal{X}\},$$

where $\nu_{\mathbf{x}}(x):=\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}_{\{x_i=x\}}$ is the empirical PMF of sequence $\mathbf{x}\in\mathcal{X}^n$. The set of all n-types over an alphabet $\mathcal{X}$ is $\mathcal{P}_n(\mathcal{X}):=\cup_{\mathbf{x}\in\mathcal{X}^n}\nu_{\mathbf{x}}(x)$. An n-type variable, i.e., a random variable with PMF P for some $P\in\mathcal{P}_n(\mathcal{X})$, is denoted using an overbar notation, e.g. $\bar{X}$. For $P_{\bar{X},\bar{Y}}\in\mathcal{P}_n(\mathcal{X}\times\mathcal{Y})$, $\hat{\mathcal{T}}_n(P_{\bar{X}}):=\{\mathbf{x}\in\mathcal{X}^n:\nu_{\mathbf{x}}=P_{\bar{X}}\}$, and for $\mathbf{x}\in\hat{\mathcal{T}}_n(P_{\bar{X}})$, $\hat{\mathcal{T}}_n(P_{\bar{X},\bar{Y}}|\mathbf{x}):=\{\mathbf{y}\in\mathcal{Y}^n:\nu_{\mathbf{x},\mathbf{y}}=P_{\bar{X},\bar{Y}}\}$. The Kullback-Leibler (KL) divergence and the TV between P and Q are represented by $\mathsf{D}_{\mathsf{KL}}(P\|Q)$ and $\delta_{\mathsf{TV}}(P,Q)$, respectively. The Rényi divergence of order $\alpha\in(0,1)\cup(1,\infty)$ between $P,Q\in\mathcal{P}(\mathcal{X})$ is

$$\mathsf{D}_{\alpha}(P\|Q):=(\alpha-1)^{-1}\log\Big(\sum_{x\in\mathcal{X}}P(x)^{\alpha}Q(x)^{1-\alpha}\Big),$$

with $\lim_{\alpha\to 1}\mathsf{D}_\alpha(P\|Q)=\mathsf{D}_{\mathsf{KL}}(P\|Q)$. For $P_X\in\mathcal{P}(\mathcal{X})$ and $P_{Y|X},Q_{Y|X}\in\mathcal{P}(\mathcal{Y}|\mathcal{X})$, the conditional α-Rényi divergence is $\mathsf{D}_\alpha\big(P_{Y|X}\big\|Q_{Y|X}|P_X\big):=\mathbb{E}_{P_X}\big[\mathsf{D}_\alpha\big(P_{Y|X}(\cdot|X)\big\|Q_{Y|X}(\cdot|X)\big)\big]$.

Finally, we follow the convention that when the set over which summation/product/supremum is taken is not specified, it is assumed to be over all possible values.

II. Soft-covering via constant composition superposition codes

We first describe constant composition superposition codes. Fix $m\in\mathbb{N}$ and a joint PMF $P_{\bar U,\bar V,Z}:=P_{\bar U,\bar V}P_{Z|\bar V}$, where $P_{\bar U,\bar V}\in\mathcal{P}_m(\mathcal{U}\times\mathcal{V})$ and $P_{Z|\bar V}\in\mathcal{P}(\mathcal{Z}|\mathcal{V})$. For $n\in\{m\mathbb{N}\}$, let $\mathbb{B}_U=\{\mathbf{U}(i),\ i\in\mathcal{I}_n\}$, $|\mathcal{I}_n|=\lceil e^{nR_1}\rceil$, be a random inner layer codebook such that each codeword $\mathbf{U}(i),\ i\in\mathcal{I}_n$, is a sequence of length n chosen independently according to $\mathsf{Unif}\big(\hat{\mathcal{T}}_n(P_{\bar U})\big)$. For a fixed realization $\mathcal{B}_U$ of $\mathbb{B}_U$ and each $i\in\mathcal{I}_n$, let $\mathbb{B}_V(i):=\{\mathbf{V}(i,j),j\in\mathcal{J}_n\}$, $|\mathcal{J}_n|=\lceil e^{nR_2}\rceil$, denote a collection of n-length random sequences, each chosen independently according to $\mathsf{Unif}\big(\hat{\mathcal{T}}_n(P_{\bar U,\bar V}|\mathbf{u}(i))\big)$. Set $\mathbb{B}_V:=\{\mathbb{B}_V(i),\ i\in\mathcal{I}_n\}$, denote the random superposition codebook by $\mathbb{B}:=\{\mathbb{B}_U,\mathbb{B}_V\}$ and let $\mathcal{B}$ denote its realization. The set of all such codebooks is $\mathfrak{B}$.

Given a fixed $\mathcal{B}\in\mathfrak{B}$, an inner layer codeword $\mathbf{u}(i)$, $i\in\mathcal{I}_n$, is chosen uniformly at random; then, $\mathbf{v}(i,j)$, $j\in\mathcal{J}_n$, is uniformly chosen from the corresponding outer layer codebook and is transmitted over the channel $P_{Z|V}^{\otimes n}$. This gives rise to the following induced distribution

$$\begin{aligned}P^{(\mathcal{B})}_{I,\mathbf{U},J,\mathbf{V},\mathbf{Z}}(i,\mathbf{u},j,\mathbf{v},\mathbf{z})&=P^{(\mathcal{B}_U)}_{I,\mathbf{U}}(i,\mathbf{u})P^{(\mathcal{B}_V)}_{J,\mathbf{V},\mathbf{Z}|I,\mathbf{U}}(j,\mathbf{v},\mathbf{z}|i,\mathbf{u})\\&=\frac{\mathbb{1}_{\{\mathbf{u}=\mathbf{u}(i)\}}}{|\mathcal{I}_n|}\frac{\mathbb{1}_{\{\mathbf{v}=\mathbf{v}(i,j)\}}}{|\mathcal{J}_n|}P_{Z|V}^{\otimes n}(\mathbf{z}|\mathbf{v}).\end{aligned}\tag{1}$$

The goal of soft-covering is to approximate the induced conditional output distribution $P^{(\mathcal{B})}_{\mathbf{Z}|\mathbf{U}}(\cdot|\mathbf{u}(i))$ by the target distribution

$$P^*_{\mathbf{Z}|\mathbf{U}}\big(\mathbf{z}\big|\mathbf{u}(i)\big):=\frac{1}{\big|\hat{\mathcal{T}}_n\big(P_{\bar U,\bar V}\big|\mathbf{u}(i)\big)\big|}\sum_{\mathbf{v}\in\hat{\mathcal{T}}_n(P_{\bar U,\bar V}|\mathbf{u}(i))}P_{Z|V}^{\otimes n}(\mathbf{z}|\mathbf{v}),\tag{2}$$

for each $i\in\mathcal{I}_n$ and on average. Proximity is measured in TV:

$$\begin{aligned}\theta(\mathcal{B},i)&:=\delta_{\mathsf{TV}}\Big(P^{(\mathcal{B})}_{\mathbf{Z}|\mathbf{U}}(\cdot|\mathbf{u}(i)),P^*_{\mathbf{Z}|\mathbf{U}}(\cdot|\mathbf{u}(i))\Big),\\ \bar\theta(\mathcal{B})&:=\delta_{\mathsf{TV}}\Big(P^{(\mathcal{B}_U)}_{I,\mathbf{U}}P^{(\mathcal{B}_V)}_{\mathbf{Z}|I,\mathbf{U}},P^{(\mathcal{B}_U)}_{I,\mathbf{U}}P^*_{\mathbf{Z}|\mathbf{U}}\Big).\end{aligned}\tag{3}$$

The following theorem provides an exact characterization of the soft-covering exponent for the above setup.

Theorem 1 (Soft-covering exponent) *For any $R_1\geq 0$, $R_2>I_P(\bar V;Z|\bar U)$, and $i\in\mathcal{I}_n$, we have*

$$\begin{aligned}\lim_{n\to\infty}-\frac{1}{n}\log\mathbb{E}_\mu\left[\bar\theta(\mathbb{B})\right]&=\lim_{n\to\infty}-\frac{1}{n}\log\mathbb{E}_\mu\left[\theta(\mathbb{B},i)\right]\\&=S(P_{\bar U,\bar V,Z},R_2),\end{aligned}\tag{4}$$

$$\begin{aligned}S(P_{\bar U,\bar V,Z},R_2):=&\min_{P_{\bar Z|\bar U,\bar V}}\mathsf{D}_{\mathsf{KL}}\big(P_{\bar U,\bar V,\bar Z}\big\|P_{\bar U,\bar V}P_{Z|\bar V}\big)\\&+0.5\big[R_2-I_P(\bar V;\bar Z|\bar U)\big]^+,\end{aligned}\tag{5}$$

and μ is the PMF of $\mathbb{B}$ induced by the above codebook construction (see (13)). In particular, for $R_2>I_P(\bar V;Z|\bar U)$, there exists $\gamma>0$ such that for all $n\in\{m\mathbb{N}\}$ sufficiently large and $i\in\mathcal{I}_n$, we have

$$\mathbb{E}_\mu\left[\theta(\mathbb{B},i)\right]=\mathbb{E}_\mu\left[\bar\theta(\mathbb{B})\right]\leq e^{-n\gamma}.\tag{6}$$

The next theorem states a double exponential concentration bound for $\bar\theta(\mathbb{B})$ about its mean.

Theorem 2 (Concentration bound) *If $R_2>I_P(\bar V;Z|\bar U)$, then there exist positive constants $\gamma_1,\gamma_2>0$ such that for all sufficiently large n and $i\in\mathcal{I}_n$, we have*

$$\mathbb{P}_\mu\big(\theta(\mathbb{B},i)>e^{-n\gamma_1}\big)=\mathbb{P}_\mu\left(\bar\theta(\mathbb{B})>e^{-n\gamma_1}\right)\leq e^{-e^{n\gamma_2}}.\tag{7}$$

The following lemma which provides a variational characterization of the optimal soft-covering exponent in terms of Rényi divergence is useful in the proof of Theorem 1.

Lemma 1 (Dual characterization) *It holds that*

$$\begin{aligned}&S(P_{\bar U,\bar V,Z},R_2)\\&\quad=\max_{\lambda\in[1,2]}\frac{\lambda-1}{\lambda}\Big(R_2-\min_{Q_{Z|\bar U}}\mathsf{D}_\lambda\big(P_{Z|\bar V}\big\|Q_{Z|\bar U}|P_{\bar U,\bar V}\big)\Big).\end{aligned}\tag{8}$$

Consequently, if $R_2>I_P(\bar V;Z|\bar U)$, then $S(P_{\bar U,\bar V,Z},R_2)>0$.

The proofs of all the above results are given in Section IV.

III. Secrecy-Capacity of Cost-Constrained Broadcast Channel with Confidential Messages

Let $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{Z}$ be finite sets, $b\geq 0$ and $n\in\mathbb{N}$. Let $\mathsf{C}:\mathcal{X}\to\mathbb{R}_{\geq 0}$ be a real-valued non-negative function. The $(\mathcal{X},\mathcal{Y},\mathcal{Z},P_{Y,Z|X},\mathsf{C},b)$ CC BC with confidential messages is shown in Fig. 1, where $P_{Y,Z|X}$ is the channel transition kernel, C is the cost function and b is the cost constraint. This is the setup from [23] but with a cost constraint on the channel input. The common message to both the receivers is denoted by M_0 and the private message to Receiver 1 by M_1, each taking values in $\mathcal{M}_{0,n}=[1:2^{nR_0}]$ and $\mathcal{M}_{1,n}=[1:2^{nR_1}]$, respectively. We consider a per-codeword cost constraint:

$$\mathsf{C}_n(\mathbf{X}(m_0,m_1))\leq b\ \text{a.s.},\ \forall\ (m_0,m_1)\in\mathcal{M}_{0,n}\times\mathcal{M}_{1,n},\tag{9}$$

where, $\mathbf{X}(m_0,m_1)\sim f_n(\cdot|m_0,m_1)$ is the encoder output, and $\mathsf{C}_n(\mathbf{x}):=\frac{1}{n}\sum_{i=1}^n\mathsf{C}(x_i)$ is the n-fold extension of C. We henceforth assume $b\geq\mathsf{c}_{\min}:=\min\{\mathsf{C}(x):x\in\mathcal{X}\}$. Decoder 1 outputs the estimates $(\hat M_0,\hat M_1)$ using $g_n:\mathcal{Y}^n\to\mathcal{M}_{0,n}\times\mathcal{M}_{1,n}$, while Decoder 2 outputs $\check M_0$ from $h_n:\mathcal{Z}^n\to\mathcal{M}_{0,n}$.

A rate tuple (R_0,R_1) is said to be achievable if for every $\epsilon>0$ and sufficiently large n, there exists an (n,R_0,R_1) code $c_n=(f_n,g_n,h_n)$ that satisfies (9) and $\max\big\{e_1(c_n),e_2(c_n),\ell_{\mathsf{sem}}(c_n)\big\}\leq\epsilon$, where

$$\begin{aligned}\ell_{\mathsf{sem}}(c_n)&:=\max_{P_{M_0,M_1}}I(M_1;\mathbf{Z}),\\ e_1(c_n)&:=\max_{m_0,m_1}\sum_{\mathbf{x}}f_n(\mathbf{x}|m_0,m_1)\sum_{\mathbf{y}:\,g_n(\mathbf{y})\neq(m_0,m_1)}P_{Y|X}^{\otimes n}(\mathbf{y}|\mathbf{x}),\\ e_2(c_n)&:=\max_{m_0,m_1}\sum_{\mathbf{x}}f_n(\mathbf{x}|m_0,m_1)\sum_{\mathbf{z}:\,h_n(\mathbf{z})\neq m_0}P_{Z|X}^{\otimes n}(\mathbf{z}|\mathbf{x}).\end{aligned}$$

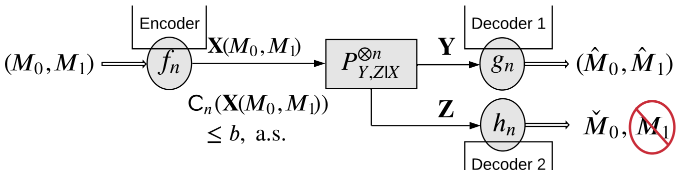

Fig. 1: The CC BC with transition kernel $P_{Y,Z|X}$.

The secrecy-capacity region $\mathcal{R}(b)$ of a per-codeword CC BC with confidential messages under semantic security (see [24]) and maximal error-probability criteria is the closure of achievable (R_0, R_1) set. We use Theorems 1-2 to characterize $\mathcal{R}(b)$.

Let $\mathcal{U}$ and $\mathcal{V}$ be finite sets. For any $P_{U,V,X} \in \mathcal{P}(\mathcal{U}\times\mathcal{V}\times\mathcal{X})$, let $\tilde{\mathcal{R}}(P_{U,V,X})$ be the set of $(R_0, R_1) \in \mathbb{R}^2_{\ge 0}$ satisfying

$$R_0 \le \min\{I_P(U;Y), I_P(U;Z)\}, \tag{10a}$$

$$R_1 \le I_P(V;Y|U) - I_P(V;Z|U), \tag{10b}$$

where $P_{U,V,X,Y,Z} = P_{U,V,X}P_{Y,Z|X}$. Set

$$\hat{\mathcal{R}}(b) := \cup_{P_{U,V,X}\in\mathcal{H}(\mathsf{C},b)}\tilde{\mathcal{R}}(P_{U,V,X}), \tag{11}$$

where, U, V, are auxiliaries with $|\mathcal{U}| \le |\mathcal{X}| + 2, |\mathcal{V}| \le |\mathcal{X}|^2 + 4|\mathcal{X}| + 2$, and

$$\mathcal{H}(\mathsf{C},b) := \big\{P_{U,V,X} : P_{U,V,X} = P_{U,V}P_{X|V}, \mathbb{E}_P\big[\mathsf{C}(X)\big] \le b\big\}. \tag{12}$$

Theorem 3 (Capacity region) *It holds that* $\mathcal{R}(b) = \hat{\mathcal{R}}(b)$.

The proof of Theorem 3 is given in Section IV-C. The achievability of $(R_0, R_1) \in \hat{\mathcal{R}}(b)$ relies on superposition coding, while the converse adapts the classic BC with confidential messages converse to accommodate the cost constraint.

IV. Proofs

A. Proof of Theorem 1

The proof is a generalization of [22, Theorem 2] to constant-composition superposition codebooks. We first prove the $\ge$ implication in (4). Denoting the set of all possible values of $\mathbb{B}_U$, $\mathbb{B}_V$, and $\mathbb{B}$ by $\mathfrak{B}_U$, $\mathfrak{B}_V$, and $\mathfrak{B}$, respectively, the codebook construction induces a PMF $\mu \in \mathcal{P}(\mathfrak{B})$, given by

$$\mu(\mathcal{B}) = \prod_{i\in\mathcal{I}_n}\big|\hat{\mathcal{T}}_n(P_{\bar{U}})\big|^{-1}\bigg(\prod_{j\in\mathcal{J}_n}\big|\hat{\mathcal{T}}_n(P_{\bar{U},\bar{V}}|\mathbf{u}(i))\big|^{-1}\bigg). \tag{13}$$

Denote the $\mathbb{B}_U$ and $\mathbb{B}_V$ marginals of μ by $\mu_{\mathbb{B}_U}$ and $\mu_{\mathbb{B}_V}$, respectively. For a fixed $\mathcal{B}_U$, we use the shorthand $\mathbb{E}_{\mu|\mathcal{B}_U}[\cdot]$ for the conditional expectation $\mathbb{E}_\mu\big[\,\cdot\,|\mathbb{B}_U = \mathcal{B}_U\big]$.

We define several quantities used throughout the proof. Fix $\mathcal{B}_U$ and $i \in \mathcal{I}_n$, henceforth, and let

$$L^{(\mathbb{B}_V)}(\mathbf{u}(i),\mathbf{z}) := \begin{cases} \frac{1}{|\mathcal{J}_n|}\sum_{j=1}^{|\mathcal{J}_n|}\frac{P^{\otimes n}_{Z|\bar{V}}(\mathbf{z}|\mathbf{V}(i,j))}{P^*_{\mathbf{Z}|\mathbf{U}}(\mathbf{z}|\mathbf{u}(i))}, & \text{if } P^*_{\mathbf{Z}|\mathbf{U}}(\mathbf{z}|\mathbf{u}(i)) > 0,\\ 1, & \text{otherwise,}\end{cases}$$

$$L^*(\mathbf{u}(i),\mathbf{z}) := \mathbb{E}_{\mu|\mathcal{B}_U}\big[L^{(\mathbb{B}_V)}(\mathbf{u}(i),\mathbf{z})\big].$$

Note that $L^*(\mathbf{u}(i),\mathbf{z}) = 1$. For $(\mathbf{u}(i),\mathbf{v},\mathbf{z}) \in \mathcal{T}_n(P_{\bar{U},\bar{V},\bar{Z}})$, set

$$\tilde{L}_{P_{\bar{U},\bar{V},\bar{Z}}} := \frac{1}{|\mathcal{J}_n|}\frac{P^{\otimes n}_{Z|\bar{V}}(\mathbf{z}|\mathbf{v})}{P^*_{\mathbf{Z}|\mathbf{U}}(\mathbf{z}|\mathbf{u}(i))},$$

$$N^{(\mathbb{B}_V)}_{P_{\bar{V}|\bar{U},\bar{Z}}}(\mathbf{u}(i),\mathbf{z}) := \Big|j \in \mathcal{J}_n : \mathbf{V}(i,j) \in \hat{\mathcal{T}}_n(P_{\bar{U},\bar{Z},\bar{V}}|\mathbf{u}(i),\mathbf{z})\Big|,$$

$$W^{(\mathbb{B}_V)}_{P_{\bar{V}|\bar{U},\bar{Z}}}(\mathbf{u}(i),\mathbf{z}) := |\mathcal{J}_n|^{-1}\,\tilde{L}_{P_{\bar{V}|\bar{U},\bar{Z}}}N^{(\mathbb{B}_V)}_{P_{\bar{V}|\bar{U},\bar{Z}}}(\mathbf{u}(i),\mathbf{z}),$$

$$W^*_{P_{\bar{V}|\bar{U},\bar{Z}}}(\mathbf{u}(i),\mathbf{z}) := \mathbb{E}_{\mu|\mathcal{B}_U}\Big[W^{(\mathbb{B}_V)}_{P_{\bar{V}|\bar{U},\bar{Z}}}(\mathbf{u}(i),\mathbf{z})\Big].$$

Lastly, for $\mathbf{u} \in \hat{\mathcal{T}}_n(P_{\bar{U}})$ and $\mathbf{V} \sim \mathsf{Unif}\big(\hat{\mathcal{T}}_n(P_{\bar{U},\bar{V}}|\mathbf{u})\big)$, define

$$q_{\bar{V}|\bar{U},\bar{Z}}(\mathbf{u},\mathbf{z}) := \mathbb{P}\big(\mathbf{V} \in \hat{\mathcal{T}}_n(P_{\bar{U},\bar{Z},\bar{V}}|\mathbf{u},\mathbf{z})\big),$$

$$F(|\mathcal{J}_n|, P_{\bar{V}|\bar{U},\bar{Z}}) := \min\big\{2q_{\bar{V}|\bar{U},\bar{Z}}(\mathbf{u},\mathbf{z}), |\mathcal{J}_n|^{-\frac{1}{2}}\,q^{\frac{1}{2}}_{\bar{V}|\bar{U},\bar{Z}}(\mathbf{u},\mathbf{z})\big\}.$$

We have the following lemma.

Lemma 2 (Bounds on intermediate quantities)

$$\mathbb{E}_{\mu|\mathcal{B}_U}\Big[\Big|W^{(\mathbb{B}_V)}_{P_{\bar{V}|\bar{U},\bar{Z}}}(\mathbf{u}(i),\mathbf{z}) - W^*_{P_{\bar{V}|\bar{U},\bar{Z}}}(\mathbf{u}(i),\mathbf{z})\Big|\Big] \le |\mathcal{J}_n|\,\tilde{L}_{P_{\bar{V}|\bar{U},\bar{Z}}}F(|\mathcal{J}_n|, P_{\bar{V}|\bar{U},\bar{Z}}), \tag{14}$$

$$q_{\bar{V}|\bar{U},\bar{Z}}(\mathbf{u}(i),\mathbf{z}) \le (n+1)^{|\mathcal{U}||\mathcal{V}|}e^{-nI(\bar{V};\bar{Z}|\bar{U})}, \tag{15}$$

$$q_{\bar{V}|\bar{U},\bar{Z}}(\mathbf{u}(i),\mathbf{z}) \ge (n+1)^{-|\mathcal{U}||\mathcal{V}||\mathcal{Z}|}e^{-nI(\bar{V};\bar{Z}|\bar{U})}. \tag{16}$$

Proof: The proof of (14) follows similar to that of [22, Lemma 3] and is omitted. To establish (15) and (16), note that $\hat{\mathcal{T}}_n(P_{\bar{U},\bar{Z},\bar{V}}|\mathbf{u}(i),\mathbf{z}) \subseteq \hat{\mathcal{T}}_n(P_{\bar{U},\bar{V}}|\mathbf{u}(i))$, which implies

$$q_{\bar{V}|\bar{U},\bar{Z}}(\mathbf{u}(i),\mathbf{z}) = \Big|\hat{\mathcal{T}}_n(P_{\bar{U},\bar{Z},\bar{V}}|\mathbf{u}(i),\mathbf{z})\Big|\,\Big|\hat{\mathcal{T}}_n(P_{\bar{U},\bar{V}}|\mathbf{u}(i))\Big|^{-1}.$$

The claims then follows from [15, Lemma 2.5]. ∎

Continuing, for $(i,\mathbf{u}) \in \mathcal{I}_n \times \hat{\mathcal{T}}_n(P_{\bar{U}})$ such that $\mathbf{u}(i) = \mathbf{u}$, we have

$$\begin{aligned}
&\rho(\mathcal{B}_U, i) := \mathbb{E}_{\mu|\mathcal{B}_U}\big[\theta\big(\{\mathcal{B}_U, \mathbb{B}_V\}, i\big)\big]\\
&= \sum_{\mathbf{z}\in\mathcal{Z}^n}P^*_{\mathbf{Z}|\mathbf{U}}(\mathbf{z}|\mathbf{u}(i))\mathbb{E}_{\mu|\mathcal{B}_U}\Big[\Big|L^{(\mathbb{B}_V)}(\mathbf{u}(i),\mathbf{z}) - 1\Big|\Big]\\
&\overset{(a)}{=} \sum_{P_{\bar{Z}|\bar{U}}\in\mathcal{P}_n(\mathcal{Z}|\mathcal{U})}\ \sum_{\mathbf{z}\in\hat{\mathcal{T}}_n(P_{\bar{U},\bar{Z}}|\mathbf{u}(i))}P^*_{\mathbf{Z}|\mathbf{U}}(\mathbf{z}|\mathbf{u}(i))\,\mathbb{E}_{\mu|\mathcal{B}_U}\Big[\Big|L^{(\mathbb{B}_V)}(\mathbf{u}(i),\mathbf{z}) - L^*(\mathbf{u}(i),\mathbf{z})\Big|\Big]\\
&= \sum_{P_{\bar{Z}|\bar{U}}\in\mathcal{P}_n(\mathcal{Z}|\mathcal{U})}\ \sum_{\mathbf{z}\in\hat{\mathcal{T}}_n(P_{\bar{U},\bar{Z}}|\mathbf{u}(i))}P^*_{\mathbf{Z}|\mathbf{U}}(\mathbf{z}|\mathbf{u}(i))\,\mathbb{E}_{\mu|\mathcal{B}_U}\Bigg[\Bigg|\sum_{P_{\bar{V}|\bar{U},\bar{Z}}}W^{(\mathbb{B}_V)}_{P_{\bar{V}|\bar{U},\bar{Z}}}(\mathbf{u}(i),\mathbf{z}) - W^*_{P_{\bar{V}|\bar{U},\bar{Z}}}(\mathbf{u}(i),\mathbf{z})\Bigg|\Bigg]\\
&= \sum_{\substack{P_{\bar{Z},\bar{V}|\bar{U}}\in\\ \mathcal{P}_n(\mathcal{Z}\times\mathcal{V}|\mathcal{U})}}\ \sum_{\mathbf{z}\in\hat{\mathcal{T}}_n(P_{\bar{U},\bar{Z}}|\mathbf{u}(i))}P^*_{\mathbf{Z}|\mathbf{U}}(\mathbf{z}|\mathbf{u}(i))\,\mathbb{E}_{\mu|\mathcal{B}_U}\Big[\Big|W^{(\mathbb{B}_V)}_{P_{\bar{V}|\bar{U},\bar{Z}}}(\mathbf{u}(i),\mathbf{z}) - W^*_{P_{\bar{V}|\bar{U},\bar{Z}}}(\mathbf{u}(i),\mathbf{z})\Big|\Big]\\
&\overset{(b)}{\le} \sum_{P_{\bar{Z},\bar{V}|\bar{U}}}\ \sum_{\mathbf{z}\in\hat{\mathcal{T}}_n(P_{\bar{U},\bar{Z}}|\mathbf{u}(i))}P^{\otimes n}_{Z|\bar{V}}(\mathbf{z}|\mathbf{v})F(|\mathcal{J}_n|, P_{\bar{V}|\bar{U},\bar{Z}})\\
&\le \sum_{P_{\bar{Z},\bar{V}|\bar{U}}}\ \sum_{\mathbf{z}\in\hat{\mathcal{T}}_n(P_{\bar{U},\bar{Z}}|\mathbf{u}(i))}e^{n\mathbb{E}_{P_{\bar{V},\bar{Z}}}\left[\log P_{Z|\bar{V}}\right]}F(|\mathcal{J}_n|, P_{\bar{V}|\bar{U},\bar{Z}})\\
&\overset{(c)}{\le} (n+1)^{\frac{3}{2}|\mathcal{U}||\mathcal{V}||\mathcal{Z}|}\max_{P_{\bar{Z}|\bar{U},\bar{V}}}e^{n\left(H(\bar{Z}|\bar{U})+\mathbb{E}_{P_{\bar{V},\bar{Z}}}\left[\log P_{Z|\bar{V}}\right]\right)}\\
&\quad(n+1)^{|\mathcal{U}||\mathcal{V}|}e^{-nI_P(\bar{V};\bar{Z}|\bar{U})}e^{-n\frac{1}{2}\left[R_2 - I_P(\bar{V};\bar{Z}|\bar{U})\right]^+}\\
&=: S_n(P_{\bar{U},\bar{V},Z}, R_2),
\end{aligned}\tag{17}$$

where (a) follows since $L^*(\mathbf{u}(i), \mathbf{z}) = 1$; (b) is due to (14) in Lemma 2; (c) follows from [15, Lemma 2.2], the definition of $F(|\mathcal{J}_n|, P_{\bar{V}|\bar{U},\bar{Z}})$ and (15)-(16) in Lemma 2. Thus, noting that $\rho(\mathcal{B}_U, i)$ is independent of i and $\mathcal{B}_U$, we have

$$\begin{aligned}\tilde{\rho}(\mathcal{B}_U) &:= \mathbb{E}_{\mu|\mathcal{B}_U}\left[\delta_{\mathsf{TV}}\Big(P^{(\mathcal{B}_U)}_{I,\mathbf{U}}P^{(\mathbb{B}_V)}_{\mathbf{Z}|I,\mathbf{U}}, P^{(\mathcal{B}_U)}_{I,\mathbf{U}}P^*_{\mathbf{Z}|\mathbf{U}}\Big)\right] \\ &= \mathbb{E}_{P^{(\mathcal{B}_U)}_{I,\mathbf{U}}}[\rho(\mathcal{B}_U, I)] \leq S_n(P_{\bar{U},\bar{V},Z}, R_2). \qquad (18)\end{aligned}$$

Taking limit and combining the resulting terms yields

$$\liminf_{n\to\infty} -\frac{1}{n}\log(\tilde{\rho}(\mathcal{B}_U)) \geq S(P_{\bar{U},\bar{V},Z}, R_2). \qquad (19)$$

Similarly, taking expectation w.r.t. $\mu_{\mathbb{B}_U}$ on both sides of (17) and (18), followed by limits leads to (4). Eqn. (6) then follows from Lemma 1.

The converse proof follows by fixing $\mathcal{B}_U$, $(I, \mathbf{U}) = (i, \mathbf{u}(i))$, and adapting the optimality argument of soft-covering exponent for single-layer codebooks from [22]. We omit further details due to space constraints.

B. Proof of Theorem 2

Fix $\mathcal{B}_U = \{\mathbf{u}(i),\ i \in \mathcal{I}_n\}$. Since $\rho(\mathcal{B}_U, i)$ is independent of $\mathcal{B}_U$ and $i \in \mathcal{I}_n$, we denote it simply by ρ. From [22, Lemma 2][1], it follows that for any $t, R_2 > 0$,

$$\mathbb{P}_\mu\Big(\theta\left(\{\mathcal{B}_U, \mathbb{B}_V\}, i\right) - \rho \geq t \mid \mathbb{B}_U = \mathcal{B}_U\Big) \leq e^{-\frac{1}{2}e^{nR_2}t^2}.$$

Taking expectation w.r.t. to $\mu_{\mathbb{B}_U}$, for any $i \in \mathcal{I}_n$, we have

$$\mathbb{P}_\mu\left(\bar{\theta}(\mathbb{B}) \geq t + \rho\right) = \mathbb{P}_\mu\left(\theta(\mathbb{B}, i) \geq t + \rho\right) \leq e^{-\frac{1}{2}e^{nR_2}t^2}.$$

Since $\rho \leq e^{-n\gamma}$ for $\gamma > 0$, by (19) and Lemma 1, if $R_2 > I_P(\bar{V}; Z|\bar{U})$, then taking $t = e^{-n\bar{\gamma}}$ for some $0 < \bar{\gamma} < \gamma$ yields

$$\mathbb{P}_\mu\big(\theta(\mathbb{B}, i) \geq 2e^{-n\bar{\gamma}}\big) = \mathbb{P}_\mu\big(\bar{\theta}(\mathbb{B}) \geq 2e^{-n\bar{\gamma}}\big) \leq e^{-\frac{1}{2}e^{n(R_2 - 2\bar{\gamma})}}.$$

Choosing $\bar{\gamma} > 0$ such that $R_2 > 2\bar{\gamma} > 0$ (possible since $R_2 > 0$ by assumption) yields the desired result.

C. Proof of Theorem 3

We first prove $\hat{\mathcal{R}}(b) \subseteq \mathcal{R}(b)$. By continuity of mutual information and the expected cost constraint in P, it suffices to show that $(R_0, R_1) \in \hat{\mathcal{R}}(b)$ is achievable, for any $b > \mathsf{c}_{\min}$.

Coding scheme: Fix $\epsilon > 0$ and a PMF $P_{U,V,X,Y,Z} := P_{U,V}P_{X|V}P_{Y,Z|X}$ such that $\mathbb{E}_P\big[\mathsf{C}(X)\big] < b$. Fix $\epsilon' \in \big(0, b - \mathbb{E}_P\big[\mathsf{C}(X)\big]\big)$. Choose $l \in \mathbb{N}$, and $Q_{\bar{U},\bar{V}} \in \mathcal{P}_l(\mathcal{U} \times \mathcal{V})$ such that $\delta_{\mathsf{TV}}\big(P_{U,V,X,Y,Z}, Q_{\bar{U},\bar{V},X,Y,Z}\big) < \epsilon'$ and $\mathbb{E}_Q\big[\mathsf{C}(X)\big] - \mathbb{E}_P\big[\mathsf{C}(X)\big] < \epsilon'$, where $Q_{\bar{U},\bar{V},X,Y,Z} = Q_{\bar{U},\bar{V}}P_{X|V}P_{Y,Z|X}$. This is possible since $\cup_{l\in\mathbb{N}}\mathcal{P}_l(\mathcal{U} \times \mathcal{V})$ is dense in $\mathcal{P}(\mathcal{U} \times \mathcal{V})$.

Let $n \in \{l\mathbb{N}\}$. Consider the random superposition codebook $\mathbb{B} := \{\mathbb{B}_U, \mathbb{B}_V\}$ constructed in Theorem 1, with $\mathcal{M}_{0,n}$ and $\mathcal{M}_{1,n} \times \mathcal{J}_n$ in place of $\mathcal{I}_n$ and $\mathcal{J}_n$, respectively. Let $\bar{\mu} \in \mathcal{P}(\mathfrak{B})$ denote the PMF induced by codebook construction as given in (13) with $Q_{\bar{U},\bar{V}}$ in place of $P_{\bar{U},\bar{V}}$. Given a codebook $\mathcal{B}$ and messages $(M_0, M_1) = (m_0, m_1)$, the encoder chooses an index pair j uniformly at random from $\mathcal{J}_n$, and transmits $\mathbf{X} = f_n(\cdot|m_0, m_1) \sim P^{\otimes n}_{X|V}\big(\cdot\,\big|\mathbf{v}(m_0, m_1, j)\big)$.

Given $\mathbf{y}$, Decoder 1 looks for a unique tuple $\big(\hat{m}_0, \hat{m}_1, \hat{j}\big) \in \mathcal{M}_{0,n} \times \mathcal{M}_{1,n} \times \mathcal{J}_n$ such that $\big(\mathbf{u}(\hat{m}_0), \mathbf{v}(\hat{m}_0, \hat{m}_1, \hat{j}), \mathbf{y}\big) \in \mathcal{T}^{(n)}_\delta(Q_{\bar{U},\bar{V},Y})$, for some $\delta > 0$. If such a unique tuple exists, it sets $g_n(\mathbf{y}) = (\hat{m}_0, \hat{m}_1)$; else, $g_n(\mathbf{y}) = (1, 1)$. Given $\mathbf{z}$, Decoder 2 looks for a unique index $\check{m}_0 \in \mathcal{M}_{0,n}$ such that $\big(\mathbf{u}(\check{m}_0), \mathbf{z}\big) \in \mathcal{T}^{(n)}_\delta(Q_{\bar{U}Z})$, and sets $h_n(\mathbf{z}) = \check{m}_0$ if its exists; else, $h_n(\mathbf{z}) = 1$. Denote the joint PMF induced by the code $c_n = (f_n, g_n, h_n)$ w.r.t. $\mathcal{B}$ by $P^{(\mathcal{B})}_{M_0,M_1,J,\mathbf{U},\mathbf{V},\mathbf{X},\mathbf{Y},\mathbf{Z},\hat{M}_0,\check{M}_0,\hat{M}_1}$.

Cost Analysis: Since for any $(m_0, m_1, j) \in \mathcal{M}_{0,n} \times \mathcal{M}_{1,n} \times \mathcal{J}_n$, $\mathbf{v}(m_0, m_1, j) \in \hat{\mathcal{T}}_n(Q_{\bar{V}})$ and $\mathbf{X} \sim P^{\otimes n}_{X|V}$,

$$\mathbb{E}_{P^{(\mathcal{B})}_{\mathbf{X}|\mathbf{V}}(\cdot|\mathbf{v}(m_0,m_1,j))}\big[\mathsf{C}_n(\mathbf{X})\big] = \mathbb{E}_{Q_X}\big[\mathsf{C}(X)\big] < b.$$

It follows that for some $\gamma' > 0$ and all $n \in \mathbb{N}$,

$$\mathbb{E}_{\bar{\mu}}\Big[\mathbb{E}_{P^{(\mathbb{B})}_{\mathbf{X}}}\big[\mathsf{C}_n(\mathbf{X})\big]\Big] \leq b - \gamma'.$$

Error probability analysis: Under conditions stated in the lemma below, the expected maximal error-probability over $\mathbb{B}_n$ decays exponentially with n. The proof is standard, and omitted due to space constraints.

Lemma 3 (Error-probability bound) *If $(R_0, R_1, R_2) \in \mathbb{R}^3_{\geq 0}$ satisfy $R_0 < I_Q(\bar{U}; Z)$, $R_1 + R_2 < I_Q(\bar{V}; Y|\bar{U})$, $R_0 + R_1 + R_2 < I_Q(\bar{U}, \bar{V}; Y)$, then there exists a $\zeta(\delta) > 0$ such that*

$$\mathbb{E}_{\bar{\mu}}\Big[\mathbb{P}_{P^{(\mathbb{B})}}\big((\hat{M}_0, \hat{M}_1) \neq (M_0, M_1)\big) + \mathbb{P}_{P^{(\mathbb{B})}}\big(\check{M}_0 \neq M_0\big)\Big] \leq e^{-n\zeta(\delta)}.$$

Security analysis: For $\mathbf{u} \in \hat{\mathcal{T}}_n(Q_{\bar{U}})$, recall the distribution $P^*_{\mathbf{Z}|\mathbf{U}}(\mathbf{z}|\mathbf{u})$ from (2). Note that $\mathbb{E}_{\bar{\mu}}\big[P^{(\mathbb{B})}_{\mathbf{Z}|\mathbf{U}}\big] = \mathbb{E}_{\bar{\mu}}\big[P^{(\mathbb{B})}_{\mathbf{Z}|M_0,M_1,\mathbf{U}}\big] = P^*_{\mathbf{Z}|\mathbf{U}}$. Following steps leading to [9, Eqns. (31)-(34)] with $P^*_{\mathbf{Z}|\mathbf{U}}$ in place of $P^{\otimes n}_{Z|U}$, it follows that for $\ell_{\mathsf{sem}}(c_n) \underset{n}{\longrightarrow} 0$ to hold, it is sufficient that there exists $\mathcal{B}$ satisfying $\max_{m_0,m_1}\tilde{\theta}(\mathcal{B}, m_0, m_1) \leq e^{-n\gamma_1}$, where $\tilde{\theta}(\mathcal{B}, m_0, m_1) := \delta_{\mathsf{TV}}\Big(P^{(\mathcal{B})}_{\mathbf{Z}|M_0,M_1,\mathbf{U}}(\cdot|m_0, m_1, \mathbf{u}(m_0)), P^*_{\mathbf{Z}|\mathbf{U}}(\cdot|\mathbf{u}(m_0))\Big)$. This existence is implied by the following lemma.

Lemma 4 (Security bound) *If $R_2 > I_Q(\bar{V}; Z|\bar{U})$, then there exists $\gamma_1, \gamma_2 > 0$ such that for all sufficiently large n,*

$$\mathbb{P}_{\bar{\mu}}\Big(\max_{(m_0,m_1)}\tilde{\theta}(\mathbb{B}, m_0, m_1) > e^{-n\gamma_1}\Big) \leq e^{-e^{n\gamma_2}}. \qquad (20)$$

The proof of (20) easily follows from (7) via the union bound by noting that $|\mathcal{M}_{0,n}| \leq e^{nR_0}$ and $|\mathcal{M}_{1,n}| \leq e^{nR_1}$.

Following the expurgation steps detailed in steps 1-3 in the proof of [9, Theorem 1] yields the existence of $\mathcal{M}'_{0,n}, \mathcal{M}'_{1,n}$, $\mathcal{B}$, f_n, g_n such that $|\mathcal{M}'_{0,n}| \geq \frac{e^{nR_0}}{4(n+2)}$, $|\mathcal{M}'_{1,n}| \geq \frac{e^{nR_1}}{4(n+2)}$ and for all $(m_0, m_1) \in \mathcal{M}'_{0,n} \times \mathcal{M}'_{1,n}$,

$$\mathbb{E}_{P^{(\mathcal{B})}}\big[\mathsf{C}_n(\mathbf{X})|(M_0, M_1) = (m_0, m_1)\big] \leq (1 + n^{-1})^2 b', \quad (21)$$

$$\begin{aligned}&\mathbb{P}_{P^{(\mathcal{B})}}\big((\hat{M}_0, \hat{M}_1) \neq (m_0, m_1)|(M_0, M_1) = (m_0, m_1)\big) \\ &+ \mathbb{P}_{P^{(\mathcal{B})}}\big(\check{M}_0 \neq m_0|M_0 = m_0\big) \leq 4(n+2)^2 e^{-n\zeta(\delta)}, \quad (22)\end{aligned}$$

$$\max_{(m_0,m_1)\in\mathcal{M}'_{0,n}\times\mathcal{M}'_{1,n}}\tilde{\theta}(\mathcal{B}, m_0, m_1) \leq e^{-n\gamma_1}. \qquad (23)$$

[1] Although [22, Lemma 2] is stated for the case of memoryless channels, the proof based on McDiarmid's inequality shows that the double exponential bound holds more generally.

The final step is to replace f_n by $\tilde{f}_n$ to satisfy the per-codeword cost constraint, where, for $(m_0, m_1, \mathbf{x}) \in \mathcal{M}'_{0,n} \times \mathcal{M}'_{1,n} \times \mathcal{T}_\delta^{(n)}(Q_X)$, the definition of $\tilde{f}_n$ is

$$\tilde{f}_n(\mathbf{x}|m_0, m_1) := \frac{\sum_j P_{X|V}^{\otimes n}\big(\mathbf{x}\big|\mathbf{v}(m_0, m_1, j)\big)}{|\mathcal{J}_n|\eta_n(m_0, m_1, \delta)},$$
$$\eta_n(m_0, m_1, \delta) := \frac{1}{|\mathcal{J}_n|}\sum_j \sum_{\mathbf{x}\in\mathcal{T}_\delta^{(n)}(Q_X)} P_{X|V}^{\otimes n}\big(\mathbf{x}\big|\mathbf{v}(m_0, m_1, j)\big),$$

and $\tilde{f}_n(\mathbf{x}|m_0, m_1) = 0$, otherwise. Since $\mathbf{v}(m_0, m_1, j) \in \hat{\mathcal{T}}_n(Q_{\bar{V}})$, [15, Lemma 2.12] implies[2] that for any $\delta > 0$, there is $\tilde{\gamma}_n \to 0$ such that $\eta_n(m_0, m_1, \delta) \geq 1 - \tilde{\gamma}_n$ for all $(m_0, m_1) \in \mathcal{M}'_{0,n} \times \mathcal{M}'_{1,n}$. The typical average lemma [25] and definition of $\tilde{f}_n$ then yield $\mathsf{C}_n(\mathbf{X}(m_0, m_1)) < b$, with probability one for all $(m_0, m_1) \in \mathcal{M}'_{0,n} \times \mathcal{M}'_{1,n}$, provided δ is sufficiently small.

Let $\tilde{P}^{(\mathcal{B})}$ denote $P^{(\mathcal{B})}$ with f_n replaced by $\tilde{f}_n$. Slightly abusing notation, we use the shorthands $p_{m_0,m_1}^{(\mathcal{B})}$ and $\tilde{p}_{m_0,m_1}^{(\mathcal{B})}$ for $P_{\mathbf{Z}|M_0,M_1,\mathbf{U}}^{(\mathcal{B})}(\cdot|m_0, m_1, \mathbf{u}(m_0))$ and $\tilde{P}_{\mathbf{Z}|M_0,M_1,\mathbf{U}}^{(\mathcal{B})}(\cdot|m_0, m_1, \mathbf{u}(m_0))$, respectively, and define

$$\theta'(\mathcal{B}, m_0, m_1) := \delta_{\mathsf{TV}}\Big(p_{m_0,m_1}^{(\mathcal{B})}, \tilde{p}_{m_0,m_1}^{(\mathcal{B})}\Big),$$
$$\kappa(\mathcal{B}, m_0, m_1) := \mathsf{D}_{\mathsf{KL}}\Big(\tilde{p}_{m_0,m_1}^{(\mathcal{B})}\Big\|p_{m_0,m_1}^{(\mathcal{B})}\Big).$$

Then, for all $(m_0, m_1) \in \mathcal{M}'_{0,n} \times \mathcal{M}'_{1,n}$, we have

$$\begin{aligned}
&\mathbb{P}_{\tilde{P}^{(\mathcal{B})}}\big((\hat{M}_0, \hat{M}_1) \neq (m_0, m_1)|(M_0, M_1) = (m_0, m_1)\big) \\
&\qquad + \mathbb{P}_{\tilde{P}^{(\mathcal{B})}}\big(\check{M}_0 \neq m_0|M_0 = m_0\big) \\
&\leq 4(n+2)^2(1-\tilde{\gamma}_n)^{-1}e^{-n\zeta(\delta)}, \\
&\max_{(m_0,m_1)\in\mathcal{M}'_{0,n}\times\mathcal{M}'_{1,n}} \delta_{\mathsf{TV}}\Big(\tilde{p}_{m_0,m_1}^{(\mathcal{B})}, P^*_{\mathbf{Z}|\mathbf{U}}(\cdot|\mathbf{u}(m_0))\Big) \\
&\overset{(a)}{\leq} \max_{(m_0,m_1)} \tilde{\theta}(\mathcal{B}, m_0, m_1) + \max_{(m_0,m_1)} \theta'(\mathcal{B}, m_0, m_1) \\
&\overset{(b)}{\leq} \max_{(m_0,m_1)} \tilde{\theta}(\mathcal{B}, m_0, m_1) + 2^{-1/2}\max_{(m_0,m_1)} \kappa(\mathcal{B}, m_0, m_1) \\
&\overset{(c)}{\leq} e^{-n\gamma_1} - \log(1-\tilde{\gamma}_n),
\end{aligned}$$

where (a) is via triangle inequality for TV metric; (b) is due to Pinsker's inequality; and (c) follows from $\eta_n(m_0, m_1, \delta) \geq 1 - \tilde{\gamma}_n$ and (23). Thus, for sufficiently large n, we have shown the existence of $\mathcal{B}$ and a $(n, R_0 - \frac{1}{n}\log(4n+8), R_1 - \frac{1}{n}\log(4n+8))$ code $c_n = (\tilde{f}_n, g_n, h_n)$ with message sets $\mathcal{M}'_{0,n}, \mathcal{M}'_{1,n}$, such that $\max\big\{e_1(c_n), e_2(c_n), \ell_{\mathsf{sem}}(c_n)\big\} \leq \epsilon$, and with probability one

$$\mathbb{E}\big[\mathsf{C}_n(\mathbf{X}(m_0, m_1))\big] \leq (1+n^{-1})^2 b' < b,$$

provided R_0, R_1, R_2 satisfy the constraints in Lemma 3 and 4.

Eliminating R_2 via the Fourier-Motzkin elimination [26] yields $R_0 < I_Q(\bar{U}; Z)$, $R_1 < I_Q(\bar{V}; Y|\bar{U}) - I_Q(\bar{V}; Z|\bar{U})$, $R_0 + R_1 < I_Q(\bar{U}, \bar{V}; Y) - I_Q(\bar{V}; Z|\bar{U})$. Since ϵ' is arbitrary, continuity of mutual information implies that $(R_0, R_1) \in \mathcal{R}(b)$ provided the above constraints hold with P in place of Q. The proof is completed by noting that the resulting rate region is equivalent to $\hat{\mathcal{R}}(b)$ and $\mathcal{R}(b)$ is a closed set by definition.

[2]This step utilizes the constant composition nature of superposition codes.

The converse proof is standard, and follows by relaxing requirements to weak secrecy and adapting the argument from [23] to accommodate the cost constraint. Details are omitted.

Remark 1 (Channel input type) *In the proof of Theorem 3, we fixed the joint type of the inner and outer layers of the superposition codebook, without restricting the type of the channel input* $\mathbf{x}$*. However, scenarios in which a fixed type of* $\mathbf{x}$ *is desired (cf. [27]–[29]) can be handled within our framework by identifying* $\bar{V} = (\bar{V}', \bar{X})$ *for some* $\bar{V}'$*, where* $P_{\bar{X}}$ *is the desired channel input type, and setting* $\mathbf{X} = f_n(\cdot|m_0, m_1) = \mathbf{x}(m_0, m_1)$.

D. Proof of Lemma 1

We extend [22, Proposition 2] to superposition codes. Set $\tilde{S}(P_{\bar{U},\bar{V},\bar{Z}}, R_2) := \frac{1}{2}\big[R_2 - I_P(\bar{V}; \bar{Z}|\bar{U})\big]$, and observe:

$$\begin{aligned}
&S(P_{\bar{U},\bar{V},Z}, R_2) \\
&:= \min_{P_{\bar{Z}|\bar{U},\bar{V}}} \mathsf{D}_{\mathsf{KL}}\big(P_{\bar{U},\bar{V},\bar{Z}}\big\|P_{\bar{U},\bar{V}}P_{Z|\bar{V}}\big) + \frac{1}{2}\big[R_2 - I_P(\bar{V}; \bar{Z}|\bar{U})\big]^+ \\
&= \min_{P_{\bar{Z}|\bar{U},\bar{V}}} \max_{\lambda\in[0,1]} \mathsf{D}_{\mathsf{KL}}\big(P_{\bar{U},\bar{V},\bar{Z}}\big\|P_{\bar{U},\bar{V}}P_{Z|\bar{V}}\big) + \lambda\tilde{S}(P_{\bar{U},\bar{V},\bar{Z}}, R_2) \\
&\overset{(a)}{=} \max_{\lambda\in[0,1]} \min_{P_{\bar{Z}|\bar{U},\bar{V}}} \mathsf{D}_{\mathsf{KL}}\big(P_{\bar{U},\bar{V},\bar{Z}}\big\|P_{\bar{U},\bar{V}}P_{Z|\bar{V}}\big) + \lambda\tilde{S}(P_{\bar{U},\bar{V},\bar{Z}}, R_2) \\
&= \max_{\lambda\in[0,1]} \min_{P_{\bar{Z}|\bar{U},\bar{V}}} \frac{\lambda R_2}{2} + (1-0.5\lambda)\,\mathsf{D}_{\mathsf{KL}}\big(P_{\bar{U},\bar{V},\bar{Z}}\big\|P_{\bar{U},\bar{V}}P_{Z|\bar{V}}\big) \\
&\qquad + 0.5\;\lambda\Big[-H_P(\bar{Z}|\bar{U}) - \mathbb{E}_{P_{\bar{V},\bar{Z}}}\big[\log P_{Z|\bar{V}}\big]\Big] \\
&\overset{(b)}{=} \max_{\lambda\in[0,1]} \min_{P_{\bar{Z}|\bar{U},\bar{V}}} \frac{\lambda R_2}{2} + (1-0.5\lambda)\,\mathsf{D}_{\mathsf{KL}}\big(P_{\bar{U},\bar{V},\bar{Z}}\big\|P_{\bar{U},\bar{V}}P_{Z|\bar{V}}\big) \\
&\qquad + 0.5\;\lambda\Big[\max_{Q_{Z|\bar{U}}} \mathbb{E}_{P_{\bar{U},\bar{Z}}}\big[\log Q_{Z|\bar{U}}\big] - \mathbb{E}_{P_{\bar{V},\bar{Z}}}\big[\log P_{Z|\bar{V}}\big]\Big] \\
&= \max_{\lambda\in[0,1]} \min_{P_{\bar{Z}|\bar{U},\bar{V}}} \max_{Q_{Z|\bar{U}}} 0.5\lambda R_2 + (1-0.5\lambda) \\
&\qquad \left[\mathsf{D}_{\mathsf{KL}}\big(P_{\bar{U},\bar{V},\bar{Z}}\big\|P_{\bar{U},\bar{V}}P_{Z|\bar{V}}\big) - \frac{\lambda}{2-\lambda}\mathbb{E}_{P_{\bar{U},\bar{V},\bar{Z}}}\left[\log\frac{P_{Z|\bar{V}}}{Q_{Z|\bar{U}}}\right]\right] \\
&= \max_{\lambda\in[0,1]} \max_{Q_{Z|\bar{U}}} \mathbb{E}_{P_{\bar{U},\bar{V}}}\Bigg[\min_{P_{\bar{Z}|\bar{U},\bar{V}}} \mathsf{D}_{\mathsf{KL}}\big(P_{\bar{Z}|\bar{U},\bar{V}}(\cdot|\bar{U},\bar{V})\big\|P_{Z|\bar{V}}(\cdot|\bar{V})\big) \\
&\qquad - \lambda(2-\lambda)^{-1}\mathbb{E}_{P_{\bar{Z}|\bar{U},\bar{V}}}\left[\log\frac{P_{Z|\bar{V}}}{Q_{Z|\bar{U}}}\right]\Bigg](1-0.5\lambda) + 0.5\lambda R_2 \\
&\overset{(c)}{=} \max_{\lambda\in[0,1]} \max_{Q_{Z|\bar{U}}} \frac{\lambda R_2}{2} - (1-0.5\lambda)\,\mathbb{E}_{P_{\bar{U},\bar{V}}}\left[\log\mathbb{E}_{P_{Z|\bar{V}}}\left[\frac{P_{Z|\bar{V}}^{\frac{\lambda}{2-\lambda}}}{Q_{Z|\bar{U}}^{\frac{\lambda}{2-\lambda}}}\right]\right] \\
&\overset{(d)}{=} \max_{\lambda\in[0,1]} \max_{Q_{Z|\bar{U}}} \frac{\lambda R_2}{2} - 0.5\lambda\;\mathsf{D}_{\frac{2}{2-\lambda}}\big(P_{Z|\bar{V}}\big\|Q_{Z|\bar{U}}|P_{\bar{U},\bar{V}}\big) \\
&= \max_{\lambda\in[1,2]} \big(1-\lambda^{-1}\big)\Big(R_2 - \min_{Q_{Z|\bar{U}}} \mathsf{D}_\lambda\big(P_{Z|\bar{V}}\big\|Q_{Z|\bar{U}}|P_{\bar{U},\bar{V}}\big)\Big), \qquad (24)
\end{aligned}$$

where (a) follows from from the minimax theorem; (b) follows since $H_P(\bar{Z}|\bar{U}) = \min_{Q_{Z|\bar{U}}} \mathbb{E}_{P_{\bar{U},\bar{Z}}}\big[-\log Q_{Z|\bar{U}}\big]$; (c) is due to [22, Lemma 20]; and (d) follows from the definition of Rényi divergence of order α. Next, note that

$$\lim_{\lambda\to 1} \min_{Q_{Z|\bar{U}}} \mathsf{D}_\lambda\big(P_{Z|\bar{V}}\big\|Q_{Z|\bar{U}}|P_{\bar{U},\bar{V}}\big) = I_P(\bar{V}; Z|\bar{U}).$$

Thus, if $R_2 > I_P(\bar{V}; Z|\bar{U})$, there exists a $\lambda \in (1, 2]$ such that RHS of (24) is strictly positive. This completes the proof.

References

[1] A. D. Wyner, "The common information of two dependent random variables," *IEEE Trans. Inf. Theory*, vol. 21, no. 2, pp. 163–179, Mar. 1975.

[2] T. S. Han and S. Verdú, "Approximation theory of output statistics," *IEEE Trans. Inf. Theory*, vol. 39, no. 3, pp. 752–772, May. 1993.

[3] M. Hayashi, "General nonasymptotic and asymptotic formulas in channel resolvability and identification capacity and their application to the wiretap channel," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1562–1575, Apr. 2006.

[4] M. R. Bloch and J. N. Laneman, "Strong secrecy from channel resolvability," *IEEE Trans. Inf. Theory*, vol. 59, no. 12, pp. 8077–8098, Dec. 2013.

[5] Z. Goldfeld, P. Cuff, and H. H. Permuter, "Semantic-security capacity for wiretap channels of type II," *IEEE Trans. Inf. Theory*, vol. 62, no. 7, pp. 3863–3879, Jul. 2016.

[6] ——, "Arbitrarily varying wiretap channels with type constrained states," *IEEE Trans. Inf. Theory*, vol. 62, no. 12, pp. 7216–7244, Dec. 2016.

[7] A. Bunin, Z. Goldfeld, H. H. Permuter, S. Shamai, P. Cuff, and P. Piantanida, "Key and message semantic-security over state-dependent channels," *IEEE Trans. Inf. Forensics and Security*, vol. 15, pp. 1541–1556, 2020.

[8] Z. Goldfeld, P. Cuff, and H. H. Permuter, "Wiretap channel with random states non-causally available at the encoder," *IEEE Trans. Inf. Theory*, vol. 66, no. 3, pp. 1497–1519, Mar. 2020.

[9] S. Sreekumar, A. Bunin, Z. Goldfeld, H. H. Permuter, and S. Shamai, "The secrecy capacity of cost-constrained wiretap channels," *IEEE Transactions on Information Theory*, 2020.

[10] P. Cuff, "Distributed channel synthesis," *IEEE Trans. Inf. Theory*, vol. 59, no. 11, pp. 7071–7096, Nov. 2013.

[11] E. C. Song, P. Cuff, and H. V. Poor, "The likelihood encoder for lossy compression," *IEEE Trans. Inf. Theory*, vol. 62, no. 4, pp. 1836–1849, Apr. 2016.

[12] J. Hou and G. Kramer, "Effective secrecy: Reliability, confusion and stealth," in *IEEE International Symposium on Information Theory (ISIT)*, 2014, pp. 601–605.

[13] M. R. Bloch, "Covert communication over noisy channels: A resolvability perspective," *IEEE Trans. Inf. Theory*, vol. 62, no. 5, pp. 2334–2354, Feb. 2016.

[14] S. Sreekumar, A. Cohen, and D. Gündüz, "Privacy-aware distributed hypothesis testing," *Entropy*, vol. 22, no. 6:665, Jun. 2020.

[15] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Cambridge University Press, 2011.

[16] R. Ahlswede and A. Winter, "Strong converse for identification via quantum channels," *IEEE Trans. Inf. Theory*, vol. 48, no. 3, pp. 569–579, Mar. 2002.

[17] A. Winter, "Secret, public and quantum correlation cost of triples of random variables," in *IEEE International Symposium on Information Theory (ISIT)*, 2005, pp. 2270–2274.

[18] J. Hou and G. Kramer, "Informational divergence approximations to product distributions," in *13th Canadian Workshop on Information Theory*, Toronto, ON, Canada, Jun. 2013, pp. 76–81.

[19] P. Cuff, "A stronger soft-covering lemma and applications," in *IEEE Conference on Communications and Network Security (CNS)*, 2015, pp. 40–43.

[20] ——, "Soft covering with high probability," in *IEEE International Symposium on Information Theory (ISIT)*, 2016, pp. 2963–2967.

[21] M. Bastani Parizi, E. Telatar, and N. Merhav, "Exact random coding secrecy exponents for the wiretap channel," *IEEE Trans. Inf. Theory*, vol. 63, no. 1, pp. 509–531, Jan. 2017.

[22] S. Yagli and P. Cuff, "Exact exponent for soft covering," *IEEE Trans. Inf. Theory*, vol. 65, no. 10, pp. 6234–6262, Oct. 2019.

[23] I. Csiszár and J. Körner, "Broadcast channels with confidential messages," *IEEE Trans. Inf. Theory*, vol. 24, no. 3, pp. 339–348, May 1978.

[24] M. Bellare, S. Tessaro, and A. Vardy, "A cryptographic treatment of the wiretap channel," in *Proc. Adv. Crypto. (CRYPTO 2012)*, Santa Barbara, CA, USA, Aug. 2012.

[25] A. E. Gamal and Y. H. Kim, *Network Information theory*. Cambridge Univ. Press, 2011.

[26] I. B. Gattegno, Z. Goldfeld, and H. H. Permuter, "Fourier-Motzkin elimination software for information theoretic inequalities," *IEEE Inf. Theory Society Newsletter*, vol. 65, no. 3, pp. 25–28, Sep. 2015, Available at http://www.ee.bgu.ac.il/∼fmeit/.

[27] J. Korner and A. Sgarro, "Universally attainable error exponents for broadcast channels with degraded message sets," *IEEE Trans. Inf. Theory*, vol. 26, no. 6, pp. 670–679, 1980.

[28] M. Hayashi and R. Matsumoto, "Universally attainable error and information exponents, and equivocation rate for the broadcast channels with confidential messages," in *49th Annual Allerton Conference on Communication, Control, and Computing*, 2011, pp. 439–444.

[29] R. Averbuch and N. Merhav, "Exact random coding exponents and universal decoders for the asymmetric broadcast channel," *IEEE Trans. Inf. Theory*, vol. 64, no. 7, pp. 5070–5086, 2018.