Wiretap Channels with Random States Non-Causally Available at the Encoder

Ziv Goldfeld, Paul Cuff and Haim H. Permuter

***Abstract*—We study the state-dependent (SD) wiretap channel (WTC) with non-causal channel state information (CSI) at the encoder. This model subsumes all instances of CSI availability as special cases, and calls for an efficient utilization of the state sequence for both reliability and security purposes. A lower bound on the secrecy-capacity, that improves upon the previously best known result published by Prabhakaran *et al.*, is derived based on a novel superposition coding scheme. Our achievability gives rise to the exact secrecy-capacity characterization of a class of SD-WTCs that decompose into a product of two WTCs, one independent of the state and the other depends only on the state. The results are derived under the strict semantic-security metric that requires negligible information leakage for all message distributions.**

***Index Terms*—Channel state information, Gelfand-Pinsker channel, semantic-security, soft-covering lemma, state-dependent channel, superposition code, wiretap channel.**

I. Introduction

Reliably transmitting a message over a noisy state-dependent (SD) channel with non-causal encoder channel state information (CSI) is a fundamental problem information-theoretic problem. Its formulation and capacity derivation date back to Gelfand and Pinsker (GP) [1]. A key virtue of the GP model is its generality. Namely, it is the most general instance of a SD point-to-point channel in which any or all of the terminals have non-causal access to CSI. Motivated by the above together with the importance of security in modern communication systems, we study the SD wiretap channel (WTC) with non-causal encoder CSI, which incorporates security versus a wiretapper into the GP channel coding paradigm.

The study of secret communication over noisy channels was pioneered by Wyner, who introduced the degraded WTC and derived its secrecy-capacity [2]. Csiszár and Körner extended Wyner's result to the non-degraded WTC [3]. These two results formed the basis for the study of physical layer security and spawned a variety of works on related topics, among which are SD-WTCs. The interest in WTCs with random states relates to the observation that knowledge of the state sequence may be exploited as an additional source of randomness to boost secrecy performance. This oftentimes involves decorrelating the transmission and the state sequence so as to avoid leaking information that might compromise security. Reliable transmission over SD channels, on the other hand, favors coherent strategies that correlate the channel input and the state. Resolving the tension between these two utilizations of the transmitter CSI is the main challenge in the considered communication scenario.

The first to consider a discrete and memoryless (DM) WTC with random states were Chen and Han Vinck [4], who studied the encoder CSI scenario. They established a lower bound on the secrecy-capacity based on a combination of wiretap coding with GP coding (see also [5] for the special case where the WTC is driven by a pair of states, one available to the encoder and the other one to the decoder). Their achievable rate, however, was shown to be suboptimal in general in a later work by Chia and El-Gamal [6]. In that work, a coding scheme that uses both wiretap coding and secret key agreement[1] was proposed for the scenario where the encoder has causal access to the state sequence, while the decoder has full CSI. Despite the restriction to use the state causally, the authors of [6] proved that their scheme can strictly outperform the adaptations of the non-causal schemes from [4], [5] to the encoder and decoder CSI setup. Subsequent related works include achievability results for the WTC with correlated sources [8], action-dependent SD-WTCs [9] and WTCs with generalized feedback [10]. The benchmark result for the SD-WTC with non-causal encoder CSI considered here is the one derived by Prabhakaran *et al.* [11], via a two layered superposition coding scheme. As a consequence of the analysis in [11], the inner layer of their superposition code is restricted to be independent of the state. However, such coding distributions are suboptimal in general.

In this paper we propose a novel superposition-based coding scheme for the SD-WTC with non-causal encoder CSI, in which both layers are correlated with the state. The scheme results in a lower bound on the secrecy-capacity, which recovers the previously best known achievability formula from [11] (as well as all preceding works) as a special case. The correlation between the inner layer of the superposition code and the state is fundamental as it allows our scheme to strictly outperform that of [11] for certain instances of the considered model. Our achievability formula also gives rise to new secrecy-capacity results. In particular, we derive the semantic-security

The work of Z. Goldfeld and H. H. Permuter was supported by the DFG via the German Israeli Project Cooperation (DIP), the Israel Science Foundation, the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement n°337752, and the Cyber Center and at Ben-Gurion University of the Negev. The work of P. Cuff was supported by the National Science Foundation (grant CCF-1350595) and the Air Force Office of Scientific Research (grant FA9550-15-1-0180).

This paper was presented in part at the 2016 IEEE International Conference on the Science of Electrical Engineers (ICSEE-2016), Eilat, Israel.

Z. Goldfeld is with the Electrical and Computer Engineering Department, Cornell University, Ithaca, NY, 14850, USA (e-mail: goldfeld@cornell.edu). H. H. Permuter is with the Department of Electrical and Computer Engineering, Ben-Gurion University of the Negev, Beer-Sheva, Israel (e-mail: haimp@bgu.ac.il). Paul Cuff is with the Department of Electrical Engineering, Princeton University, Princeton, NJ, 08544, USA (e-mail: cuff@princeton.edu).

[1] see also [7] for a related work focused solely on secret key agreement

(SS) capacity of a class of SD-WTCs that decompose into a WTC that is independent of the state and another channel that generates two noisy versions of the state, each observed either by the legitimate receiver or by the eavesdropper.

We use an over-populated superposition codebook and encode the entire confidential message at the outer layer. The transmission is correlated with the state sequence by means of the likelihood encoder [12], while security is ensured by making the eavesdropper decode the inner layer codeword that contains no confidential information. Having done so, the eavesdropper is lacking the resources to extract any information about the secret message. Superposition-based code constructions for secrecy purposes have been considered before in the context of lossy source coding in [13]–[16], where the eavesdropper was also allowed to decode a layer that contains no useful information

Our results are derived under the strict metric of SS. This criterion is a cryptographic benchmark that was adapted to suit the information-theoretic framework (of computationally unbounded adversaries) in [17]. In that work, SS was shown to be equivalent to a negligible mutual information between the message and the eavesdropper's observations, for all message distributions. In contrast to our stringent security requirement, all the aforementioned results were derived under the weak-secrecy metric, i.e., a vanishing *normalized* mutual information with respect to a *uniformly distributed* message. Nowadays, however, weak-secrecy is regarded as being insufficient, giving rise to the recent effort of upgrading information-theoretic secrecy results to strong-secrecy (by removing the normalization factor but keeping the uniformity assumption on the message). SS further strengthens both these; consequently, our achievability result outperforms the schemes from [4], [5], [11] for the SD-WTC with non-causal encoder CSI, not only in terms of the achievable secrecy rate, but also in the upgraded sense of security it guarantees.

The remainder of this paper is organized as follows. Section II provides notation, basic definitions and properties. In Section III we describe the SD-WTC with non-causal encoder CSI and state the lower bound on its SS-capacity. Section IV discusses our result, compares it to previous works, and states some tight SS-capacity results. The proof of our main theorem is provided in Section V. Section VI summarizes the main achievements and insights of this work.

II. Notation and Preliminaries

We use the following notation. $\mathbb{N}$ is the set of natural numbers (0 is not included), while $\mathbb{R}$ denotes the reals. We also define $\mathbb{R}_+ = \{x\in\mathbb{R}|x\geq 0\}$ and $\mathbb{R}_{++} = \{x\in\mathbb{R}|x>0\}$. Given $a,b\in\mathbb{R}$, we use $[a:b]$ for the set of integers $\big\{n\in\mathbb{N}\big|\lceil a\rceil \leq n\leq \lfloor b\rfloor\big\}$. Calligraphic letters denote sets, e.g., $\mathcal{X}$, the complement of $\mathcal{X}$ is $\mathcal{X}^c$, while $|\mathcal{X}|$ stands for its cardinality. $\mathcal{X}^n$ is the n-fold Cartesian product of $\mathcal{X}$. An element of $\mathcal{X}^n$ is denoted by $x^n = (x_1,x_2,\ldots,x_n)$; whenever the dimension n is clear from the context, we use boldface letters, e.g., $\mathbf{x}$, for vectors (or sequences). A substring of $\mathbf{x}\in\mathcal{X}^n$ is designated by $x_i^j = (x_i,x_{i+1},\ldots,x_j)$, for $1\leq i\leq j\leq n$; when $i=1$, the subscript is omitted. We also define $x^{n\backslash i} = (x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_n)$. Random variables are denoted by uppercase letters, e.g., X, with similar conventions for random vectors.

Let $\big(\mathcal{X},\mathcal{F},\mathbb{P}\big)$ be a probability space, where $\mathcal{X}$ is the sample space, $\mathcal{F}$ is the σ-algebra and $\mathbb{P}$ is the probability measure. Random variables over $\big(\mathcal{X},\mathcal{F},\mathbb{P}\big)$ are denoted by uppercase letters, e.g., X, with conventions for random vectors similar to those for deterministic sequences. The probability of an event $\mathcal{A}\in\mathcal{F}$ is $\mathbb{P}(\mathcal{A})$, while $\mathbb{P}(\mathcal{A}\big|\mathcal{B})$ is the conditional probability of $\mathcal{A}$ given $\mathcal{B}$. We use $\mathbb{1}_\mathcal{A}$ for the indicator function of $\mathcal{A}$, while $p_\mathcal{A}^{(U)}$ is the uniform distribution over $\mathcal{A}$. The set of all probability mass functions (PMFs) on a finite set $\mathcal{X}$ is:

$$\mathcal{P}(\mathcal{X}) \triangleq \left\{ P:\mathcal{X}\to[0,1] \,\middle|\, \sum_{x\in\mathcal{X}} P(x)=1 \right\}. \tag{1}$$

In our notation for PMFs we oftentimes use subscripts to identify the involved random variable(s) and its possible conditioning. For example, for a discrete probability space $\big(\mathcal{X},\mathcal{F},\mathbb{P}\big)$ and two random variables X and Y over that space, we use p_X, $p_{X,Y}$ and $p_{X|Y}$ to denote, respectively, the marginal PMF of X, the joint PMF of (X,Y) and the conditional PMF of X given Y. In particular, $p_{X|Y}$ is a stochastic matrix whose elements are $p_{X|Y}(x|y) = \mathbb{P}\big(X=x|Y=y\big)$. Expressions such as $p_{X,Y} = p_X p_{Y|X}$ are to be understood in the pointwise sense, i.e., $p_{X,Y}(x,y) = p_X(x)p_{Y|X}(y|x)$, for all $(x,y)\in\mathcal{X}\times\mathcal{Y}$. Accordingly, when three random variables X, Y and Z satisfy $p_{X|Y,Z} = p_{X|Y}$, they form a Markov chain, which we denote by $X-Y-Z$. Subscripts of a PMF are omitted if the arguments are lowercase versions of the random variables.

For a discrete measurable space $(\mathcal{X},\mathcal{F})$, a PMF $q\in\mathcal{P}(\mathcal{X})$ induces a probability measure on $(\mathcal{X},\mathcal{F})$, denoted by $\mathbb{P}_q$; accordingly, $\mathbb{P}_q\big(\mathcal{A}\big) = \sum_{x\in\mathcal{A}} q(x)$, for every $\mathcal{A}\in\mathcal{F}$. We use $\mathbb{E}_q$ for an expectation taken with respect to $\mathbb{P}_q$. Similarly, we use H_q and I_q for entropy or mutual information terms that are calculated with respect to q. For a sequence of random variables X^n, if the entries of X^n are drawn in an i.i.d. manner according to p_X, then for every $\mathbf{x}\in\mathcal{X}^n$ we have $p_{X^n}(\mathbf{x}) = \prod_{i=1}^n p_X(x_i)$ and we write $p_{X^n}(\mathbf{x}) = p_X^n(\mathbf{x})$. Similarly, if for every $(\mathbf{x},\mathbf{y})\in\mathcal{X}^n\times\mathcal{Y}^n$ we have $p_{Y^n|X^n}(\mathbf{y}|\mathbf{x}) = \prod_{i=1}^n p_{Y|X}(y_i|x_i)$, then we write $p_{Y^n|X^n}(\mathbf{y}|\mathbf{x}) = p_{Y|X}^n(\mathbf{y}|\mathbf{x})$. The conditional product PMF $p_{Y|X}^n$, given a specific sequence $\mathbf{x}\in\mathcal{X}^n$, is denoted by $p_{Y|X=\mathbf{x}}^n$.

The empirical PMF $\nu_\mathbf{x}$ of $\mathbf{x}\in\mathcal{X}^n$ is

$$\nu_\mathbf{x}(x) \triangleq \frac{N(x|\mathbf{x})}{n}, \tag{2}$$

where $N(x|\mathbf{x}) = \sum_{i=1}^n \mathbb{1}_{\{x_i=x\}}$. We use $\mathcal{T}_\epsilon^n(p)$ to denote the set of letter-typical sequences of length n with respect to $p\in\mathcal{P}(\mathcal{X})$ and $\epsilon>0$ [18, Chapter 3], i.e., we have

$$\mathcal{T}_\epsilon^n(p) = \left\{\mathbf{x}\in\mathcal{X}^n \,\middle|\, \big|\nu_\mathbf{x}(x)-p(x)\big| \leq \epsilon p(x),\ \forall x\in\mathcal{X}\right\}. \tag{3}$$

For a countable $\mathcal{X}$ and $p,q\in\mathcal{P}(\mathcal{X})$, the *relative entropy* (Kullback-Leibler divergence) between p and q is

$$\mathsf{D}(p||q) = \sum_{x\in\mathrm{supp}(p)} p(x)\log\left(\frac{p(x)}{q(x)}\right) \tag{4}$$

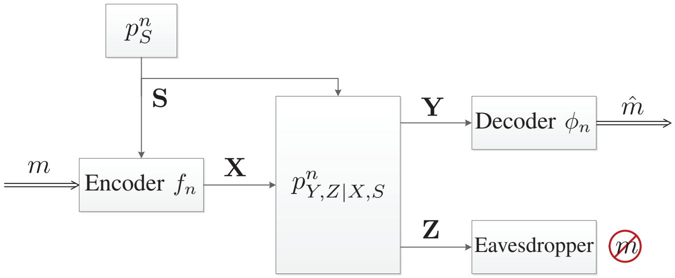

Fig. 1. The state-dependent wiretap channel with non-casual encoder channel state information.

and the *total variation* between them is

$$||p-q||_{\mathsf{TV}} = \frac{1}{2}\sum_{x\in\mathcal{X}}\big|p(x)-q(x)\big|. \tag{5}$$

Relative entropy dominates total variation through Pinsker's inequality [19, Theorem 4.1]: for any $p,q\in\mathcal{P}(\mathcal{X})$

$$||p-q||_{\mathsf{TV}} \leq \sqrt{\frac{1}{2}\mathsf{D}(p||q)}. \tag{6}$$

There is no reverse Pinsker's inequality in general, but a reverse asymptotic relation sometimes holds [20, Remark 1].

Lemma 1 (Total Variation vs. Relative Entropy) *Let $\mathcal{X}$ be a finite set and $\{p_n\}_{n\in\mathbb{N}}$ be a sequence of distributions with $p_n\in\mathcal{P}(\mathcal{X}^n)$. Let $q\in\mathcal{P}(\mathcal{X})$ and assume $p_n \ll q^n$ for every $n\in\mathbb{N}$. Then*

$$\mathsf{D}(p_n||q^n)\in O\left(\left[n+\log\frac{1}{||p_n-q^n||_{\mathsf{TV}}}\right]||p_n-q^n||_{\mathsf{TV}}\right). \tag{7}$$

III. Wiretap Channels with Random States Non-Causally Available at the Encoder

We study the SD-WTC with non-causal encoder CSI. A novel achievability bound is derived that, in some cases, strictly outperforms the previously best coding scheme.

A. Problem Setup

Let $\mathcal{S}$, $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{Z}$ be finite sets. The $(\mathcal{S},\mathcal{X},\mathcal{Y},\mathcal{Z},p_S,p_{Y,Z|X,S})$ DM SD-WTC with non-causal encoder CSI is shown in Fig. 1. A state sequence $\mathbf{s}\in\mathcal{S}^n$ that is i.i.d. according to p_S is revealed non-causally to the sender. Upon observing $\mathbf{s}$ and choosing a message $m\in\big[1:2^{nR}\big]$, the sender maps them onto a channel input sequence $\mathbf{x}\in\mathcal{X}^n$ (the mapping may be random), which is fed into a DM SD-WTC with transition probability $p_{Y,Z|X,S}$. The outputs $\mathbf{y}\in\mathcal{Y}^n$ and $\mathbf{z}\in\mathcal{Z}^n$ are observed by the receiver and the eavesdropper, respectively. Based on $\mathbf{y}$, the receiver produces an estimate $\hat{m}$ of m. The eavesdropper tries to glean whatever it can about the message from $\mathbf{z}$.

Remark 1 (Model Generality) *The considered model is the most general instance of a SD-WTC with non-causal CSI known at some or all of the terminals. The seemingly broadest setup one may consider is when the SD-WTC $p_{\tilde{Y},\tilde{Z}|X,S_1,S_2,S_3}$ is driven by a triple of correlated states $(S_1,S_2,S_3)\sim p_{S_1,S_2,S_3}$, where S_1, S_2 and S_3 are known to the transmitter, receiver and eavesdropper, respectively. However, setting $S=S_1$, $Y=(\tilde{Y},S_2)$, $Z=(\tilde{Z},S_3)$ in a SD-WTC with non-causal encoder CSI and defining the channel's transition kernel as*

$$p_{Y,Z|X,S}=p_{(\tilde{Y},S_2),(\tilde{Z},S_3)|X,S_1}=p_{S_2,S_3|S_1}p_{\tilde{Y},\tilde{Z}|X,S_1,S_2,S_3}, \tag{8}$$

one recovers this general SD-WTC from our model. The encoder CSI only model also supports a public *or a* private *bit-pipe (respectively, from the transmitter to the receiver and the eavesdropper, or to the receiver only), in addition to, or instead of, the noisy channel.*

Definition 1 (Code) *An (n,R)-code c_n for the SD-WTC with non-causal encoder CSI has a message set $\mathcal{M}_n\triangleq\big[1:2^{nR}\big]$, a stochastic encoder $f_n:\mathcal{M}_n\times\mathcal{S}^n\to\mathcal{P}(\mathcal{X}^n)$ and a decoder $\phi_n:\mathcal{Y}^n\to\hat{\mathcal{M}}_n$, where $\hat{\mathcal{M}}_n=\mathcal{M}_n\cup\{e\}$ and $e\notin\mathcal{M}_n$.*

For any message distribution $p_M\in\mathcal{P}(\mathcal{M}_n)$ and an (n,R)-code c_n, the induced joint PMF on $\mathcal{S}^n\times\mathcal{M}_n\times\mathcal{X}^n\times\mathcal{Y}^n\times\mathcal{Z}^n\times\hat{\mathcal{M}}_n$ is

$$\begin{aligned}P^{(c_n)}(\mathbf{s},m,\mathbf{x},\mathbf{y},\mathbf{z},\hat{m}) =& p_S^n(\mathbf{s})p_M(m)f_n(\mathbf{x}|m,\mathbf{s})\\ &\times p^n_{Y,Z|X,S}(\mathbf{y},\mathbf{z}|\mathbf{x},\mathbf{s})\mathbb{1}_{\big\{\hat{m}=\phi_n(\mathbf{y})\big\}}.\end{aligned} \tag{9}$$

The performance of c_n is evaluated in terms of its rate R, the maximal decoding error probability and the SS-metric.

Definition 2 (Maximal Error Probability) *The maximal error probability of an (n,R)-code c_n is*

$$e(c_n)=\max_{m\in\mathcal{M}_n}e_m(c_n), \tag{10a}$$

where

$$\begin{aligned}e_m(c_n)=&\sum_{(\mathbf{s},\mathbf{x})\in\mathcal{S}^n\times\mathcal{X}^n}p_S^n(\mathbf{s})f_n(\mathbf{x}|m,\mathbf{s})\\ &\times\sum_{\substack{(\mathbf{y},\mathbf{z})\in\mathcal{Y}^n\times\mathcal{Z}^n:\\ \phi_n(\mathbf{y})\neq m}}p^n_{Y,Z|X,S}(\mathbf{y},\mathbf{z}|\mathbf{x},\mathbf{s}).\end{aligned} \tag{10b}$$

Definition 3 (Leakage and SS Metric) *The information leakage to the eavesdropper under the (n,R)-code c_n and the message distribution $p_M\in\mathcal{P}(\mathcal{M}_n)$ is*

$$\ell(p_M,c_n)=I_{P^{(c_n)}}(M;\mathbf{Z}), \tag{11}$$

where $P^{(c_n)}$ is given in (9)*. The SS metric under c_n is*

$$\ell_{\mathsf{Sem}}(c_n)=\max_{p_M\in\mathcal{P}(\mathcal{M}_n)}\ell(p_M,c_n). \tag{12}$$

Definition 4 (Achievability) *A number $R\in\mathbb{R}_+$ is called an achievable SS-rate for the SD-WTC with non-causal encoder CSI if for every $\epsilon>0$ and sufficiently large n there exists a CR (n,R)-code c_n with*

$$e(c_n)\leq\epsilon \tag{13a}$$

$$\ell_{\mathsf{Sem}}(c_n)\leq\epsilon. \tag{13b}$$

Definition 5 (SS-Capacity) *The SS-capacity C_{Sem} of the SD-WTC with non-causal encoder CSI is the supremum of the set of achievable SS-rates.*

B. Main Result

Our main result is a novel lower bound on the SS-capacity of the SD-WTC with non-causal encoder CSI. Let $\mathcal{U}$ and $\mathcal{V}$ be finite sets and for $p_{U,V,X|S}:\mathcal{S}\to\mathcal{P}(\mathcal{U}\times\mathcal{V}\times\mathcal{X})$ define

$$R_{\mathsf{A}}\left(p_{U,V,X|S}\right)$$
$$\triangleq \min\Big\{I(V;Y|U)-I(V;Z|U), I(U,V;Y)-I(U,V;S)\Big\}, \tag{14}$$

where the mutual information terms are calculated with respect to the joint distribution $p_S p_{U,V,X|S} p_{Y,Z|X,S}$, i.e., such that $(U,V)-(X,S)-(Y,Z)$.

Theorem 1 (SD-WTC SS-Capacity Lower Bound) *The SS-capacity of the SD-WTC with non-causal encoder CSI is lower bounded by*

$$C_{\mathsf{Sem}} \geq R_{\mathsf{A}} \triangleq \max_{\substack{p_{U,V,X|S}:\\ I(U;Y)-I(U;S)\geq 0}} R_{\mathsf{A}}\left(p_{U,V,X|S}\right), \tag{15}$$

and one may restrict the cardinalities of U and V to $|\mathcal{U}|\leq|\mathcal{S}||\mathcal{X}|+5$ and $|\mathcal{V}|\leq|\mathcal{S}|^2|\mathcal{X}|^2+5|\mathcal{S}||\mathcal{X}|+3$.

The proof of Theorem 1 is given in Section V and is based on a superposition coding scheme for secrecy. The entire secret message is encoded in the *outer layer* of the superposition codebook encodes, with no confidential information carried in its inner layer. As explained next, the coding distribution is chosen so that the inner layer is better observable by the eavesdropper. This makes the eavesdropper 'waste' channel resources on decoding it, leaving insufficient resources to extract information about the secret message. The outer codebook is designed to give a physical layer advantage to the legitimate parties, thus enabling wiretap coding to conceal the confidential message. The transmission is correlated with the state sequence by means of the likelihood encoder [12]. The SS analysis relies on the soft-covering for superposition codes (Lemma 4) and an expurgation argument (see, e.g., [21, Theorem 7.7.1]).

Remark 2 (Optimal Distributions) *Note the following:*

1) *The joint distribution in* (15) *satisfies $(U,V)-(X,S)-(Y,Z)$. However, since in all the mutual information terms from* (15) *the auxiliary random variable V appears next to U or conditioned on it, we may replace V with $\tilde{V}=(U,V)$ without changing the rate. Thus, one may restrict optimization to distributions with $U-V-(X,S)-(Y,Z)$.*

2) *Feasible distributions $p_{U,V,X|S}$ in Theorem 1 must satisfy $I(U;Y)\geq I(U;S)$. This restriction can be replaced with $I(U;Z)\geq I(U;Y)\geq I(U;S)$ without changing the rate. Due to the Markov relation above, $I(U,V;Y)-I(U,V;S)$ can be rewritten as $I(V;Y)-I(V;S)$. Expanding the first expression as*

$$\begin{aligned}I(V;Y|U)&-I(V;Z|U)\\&=I(V;Y)-I(V;Z)+I(U;Z)-I(U;Y),\end{aligned} \tag{16}$$

we see that if $p_{U,V,X|S}$ satisfies $I(U;Z)<I(U;Y)$, then taking $U=0$ achieves a higher rate.

Remark 3 (Interpretation of Theorem 1) *To gain insight on the structure of R_{A}, notice that $I(V;Y|U)-I(V;Z|U)$ is the total rate of secrecy resources produced by the codebook's outer layer. That is, the outer layer can achieve a secure communication rate of $I(V;Y|U)-\max\big\{I(V;Z|U),I(V;S|U)\big\}$, and produce a secret key at a rate $\Big[I(V;S|U)-I(V;Z|U)\Big]^+$, where $[x]^+=\max\{0,x\}$. This is since some of the dummy bits needed to correlate the transmission with the state are secure for the same reason the transmission is secure.*

Also, the total amount of reliable (secured and unsecured) communication that this codebook supports in both layers combined is $I(U,V;Y)-I(U,V;S)$. Therefore, one interpretation of our encoding scheme is that the secret key produced in the outer layer (if any) is applied to the non-secure communication in the inner layer. This achieves a secure communication rate that is the minimum of the total secrecy resources $I(V;Y|U)-I(V;Z|U)$ (i.e., secure communication and secret key) and the total communication rate $I(U,V;Y)-I(U,V;S)$, corresponding to the statement of R_{A}. This effect happens naturally by the design of the superposition code, without explicitly extracting a key and applying a one-time pad.

Remark 4 (Cardinality Bounds) *The cardinality bounds on U and V in Theorem 1 follows by a standard application of the Eggleston-Fenchel-Carathéodory theorem [22, Theorem 18] twice. The details are omitted.*

Remark 5 (Recently Reported Related Result) *In [23], a lower bound on the* secret key capacity, *C_{SK}, of the SD-WTC was reported. Specifically, Theorem 1 therein states that*[2]

$$C_{\mathsf{SK}}\geq\max\Big[I(V;Y|U)-I(V;Z|U)\Big], \tag{17}$$

where the maximization is over all $p_{V,X|S}$ and $p_{U|V}$ satisfying $I(V;Y)\geq I(V;S)$. The underlying joint distribution is $p_S p_{V,X|S} p_{U|V} p_{Y,Z|S,X}$, where $U-V-(S,X)-(Y,Z)$ forms a Markov chain. The coding scheme employed in [23] is reminiscent of the superposition code proposed herein, though the analysis is different. As noted in Section I, the usage of superposition coding for SD-WTCs seems to have originated from [11] (see also [13]).

Secret key capacity is generally higher than secret message capacity, and therefore, the above lower bound does not directly apply to our model. Furthermore, [24] showed that [23,

[2] [23, Theorem 1] considers a setting with state observations at the receiver and the eavesdropper, and a public communication link. As Remark 1 explains, this is a special case of the SD-WTC. It can be verified that [23, Theorem 1] (in its original form) is recoverable from the following restatement.

Theorem 1] is missing a condition to be correct. The missing condition would have ensured that the legitimate receiver can decode the inner layer codebook without relying on the outer layer. Without it, [24] constructed a counterexample for which the right-hand side (RHS) of (17) *exceeds* C_{SK}.

C. Alternative Characterization of Achievable Rate

The formula for R_{A} can be restated in an alternative, yet equivalent, form. As before, let $\mathcal{U}$ and $\mathcal{V}$ be alphabets with cardinality bounds as in Theorem 1, and for any $p_{U,V,X|S} : \mathcal{S} \to \mathcal{P}(\mathcal{U} \times \mathcal{V} \times \mathcal{X})$ set

$$R_{\mathsf{A}}^{\mathsf{Alt}}\big(p_{U,V,X|S}\big) \triangleq \min \left\{ \begin{array}{l} I(V;Y|U) - I(V;Z|U), \\ I(U,V;Y) - I(U,V;S), \\ I(U,V;Y) - I(U;S) - I(V;Z|U) \end{array} \right\}, \tag{18}$$

where the underlying joint distribution is $p_S p_{U,V,X|S} p_{Y,Z|X,S}$.

Proposition 1 (Alternative Characterization of R_{A}) *Setting*

$$R_{\mathsf{A}}^{\mathsf{Alt}} \triangleq \max_{p_{U,V,X|S}} R_{\mathsf{A}}^{\mathsf{Alt}}\left(p_{U,V,X|S}\right), \tag{19}$$

we have $R_{\mathsf{A}}^{\mathsf{Alt}} = R_{\mathsf{A}}$.

The proof of Proposition 1 is given in Appendix D. The main challenge is showing that $R_{\mathsf{A}}^{\mathsf{Alt}} \leq R_{\mathsf{A}}$. This is since in $R_{\mathsf{A}}^{\mathsf{Alt}}$, the legitimate user may not be able to reliably decode the (inner) U layer of the superposition codebook separately. Decoding U in $R_{\mathsf{A}}^{\mathsf{Alt}}$ is possible, in general, only with the assistance of the (outer) V layer. This is evident from the second and third rate bounds in $R_{\mathsf{A}}^{\mathsf{Alt}}$. Indeed, even if an input distribution $p_{U,V,X|S}$ induces $I(U;Y) < I(U;S)$, it still might result in a positive achievable rate. In contrast, R_{A} only allows input distributions with $I(U;Y) \geq I(U;S)$, under which U is decodable on its own by the legitimate user. This discrepancy is resolved by the fact optimal input distributions for $R_{\mathsf{A}}^{\mathsf{Alt}}$ always satisfy $I(U;Y) \geq I(U;S)$ (see the proof in Appendix D).

IV. Special Cases and Examples

A. Comparison to Previous Benchmark

Theorem 1 recovers the previously best known achievable secrecy rate over SD-WTCs with non-causal encoder CSI, due to Prabhakarn *et al.* [11]. Theorem 2 of [11] established a tradeoff between achievable secret message and secret key rate pairs. Specializing their result to the secret message only scenario implies the achievability of

$$R_{\mathsf{PER}} \triangleq \max_{p_U p_{V,X|U,S}} R_{\mathsf{PER}}\left(p_U p_{V,X|U,S}\right), \tag{20a}$$

where, for any $p_U \in \mathcal{P}(\mathcal{U})$ and $p_{V,X|U,S} : \mathcal{U} \times \mathcal{S} \to \mathcal{P}(\mathcal{V} \times \mathcal{X})$,

$$R_{\mathsf{PER}}\left(p_U p_{V,X|U,S}\right) \triangleq \Big\{ I(U,V;Y) - I(U,V;S) \\ , I(V;Y|U) - I(V;Z|U) \Big\}, \tag{20b}$$

and the mutual information terms are taken with respect to $p_S p_U p_{V,X|U,S} p_{Y,Z|S,X}$, i.e., such U and S are independent and $(U,V) - (S,X) - (Y,Z)$ forms a Markov chain.

The difference between R_{PER} and R_{A} from Theorem 1 is that the former requires U to be independent of S. This essentially means that $R_{\mathbb{PER}}$ does not support GP coding in the inner layer of the superposition code. Our scheme, on the other hand, supports inner layer GP coding so long as it is decodable by the legitimate receiver. Thus, U and S may be correlated in R_{A} but must satisfy $I(U;Y) \geq I(U;S)$. As explained next, independent S and U are suboptimal in general.

Clearly, Theorem 1 recovers R_{PER} by restricting U to be independent of S in R_{A}. This restriction is valid as it satisfies $I(U;S) = 0$. Furthermore, there are instances of SD-WTC with non-causal encoder CSI for which R_{A} is strictly larger than R_{PER}. In concurrent work with collaborators [24], we construct a particular example of such a channel (see Section V-A therein). The idea is to consider a channel for which GP coding is necessary in order to attain capacity (e.g., Memory with Stuck-at-Faults) and force communication to happen in the inner layer of the codebook (by considering a strong eavesdropper). As the scheme from [11] does not allow inner layer GP coding, it achieves rates strictly below capacity. Our scheme, on the other hand, is optimal for the considered setup. This establishes the suboptimality of R_{PER} and illuminates the fundamental role of the correlation between U and S for secure transmission over SD-WTCs with non-causal encoder CSI.

Lastly, we note that the result from [11] was derived under the weak-secrecy metric.[3]. Our achievability ensures SS, and thus Theorem 1 improves upon [11, Theorem 2], not only in the rate it achieves, but also in the sense of security it guarantees.

Remark 6 (WTC with Correlated Sources) *Another related setup is that of the WTC with correlated sources [8], where the WTC* $p_{Y,Z|X}$ *is not SD and two correlated sources* $(\mathbf{S}, \mathbf{S}_1) \sim p_{S,S_1}^n$ *are observed non-causally by the encoder and the legitimate receiver, respectively. In [8],*

$$\max_{\substack{p_{W|S}p_{T,X}: \\ I(T;Y) \geq I(W;S)}} \Big\{ \Big| I(T;Y) - \max\big\{ I(T;Z), I(W;S) \big\} \Big| + I(W;S_1) \Big\}, \tag{21}$$

where the joint distribution is $p_{S,S_1} p_{S|S} p_{T,X} p_{Y,Z|X}$, *was established as a lower bound on the weak-secrecy capacity of that model. Setting* $U = 0$ *and* $V = (T,W)$ *into* $R_{\mathsf{PER}}\left(p_U p_{V,X|U,S}\right)$ *from* (20b) *and maximizing over* $p_{W|S} p_{T,X}$ *recovers* (21). *Thus, Theorem 1 also subsumes the result of [8].*

B. An Earlier Benchmark by Chen and Han Vinck

The benchmark result for the SD-WTC with non-causal encoder CSI prior to [11] is due to Chen and Han Vinck [4]. Theorem 2 of [4] lower bounds the weak-secrecy capacity by

$$R_{\mathsf{CHV}} \triangleq \max_{p_{V,X|S}} R_{\mathsf{CHV}}\left(p_{V,X|S}\right), \tag{22a}$$

where for any $p_{V,X|S} : \mathcal{S} \to \mathcal{P}(\mathcal{V} \times \mathcal{X})$,

$$R_{\mathsf{CHV}}\big(p_{V,X|S}\big) \triangleq \min\Big\{ I(V;Y) - I(V;Z), I(V;Y) - I(V;S) \Big\}, \tag{22b}$$

[3] Weak-secrecy refers to a vanishing *normalized* information leakage $\frac{1}{n} I(M;\mathbf{Z})$, where M is a uniformly distributed confidential message.

and the mutual information terms are with respect to $p_S p_{V,X|S} p_{Y,Z|X,S}$, i.e., such that $V-(X,S)-(Y,Z)$ forms a Markov Chain.

The code construction that achieves R_{CHV} combines GP coding and wiretap coding. A single-layered codebook is employed, in which the bins are large enough to simultaneously facilitate correlating the transmission with the state and confusing the eavesdropper. This construction is evident from the structure of the achievability formula by rewriting $R_{\mathsf{CHV}}\left(p_{V,X|S}\right)$ as

$$R_{\mathsf{CHV}}\left(p_{V,X|S}\right) = I(V;Y) - \max\Big\{I(V;Z), I(V;S)\Big\}. \tag{23}$$

The later work [5] studied a SD-WTC driven by a pair of pairwise i.i.d. state sequences $(\mathbf{S},\mathbf{S}_1) \sim p^n_{S,S_1}$ (the channel transition matrix is $p_{\tilde{Y},Z|X,S,S_1}$). The encoder has non-causal access to $\mathbf{S}$, while the legitimate receiver has $\mathbf{S}_1$. As explained in Remark 1, however, this instance is a special case of [4] by taking $Y = (\tilde{Y}, S_1)$ and setting $p_{Y,Z|X,S} = p_{(\tilde{Y},S_1),Z|X,S} = p_{S_1|S} p_{\tilde{Y},Z|X,S,S_1}$. The achievability of R_{CHV} is recovered from Theorem 1 (and from R_{PER} or R_{A}) by setting $U=0$.

Remark 7 (Suboptimality of [4]) *In [6], Chia and El Gamal showed that R_{CHV} is suboptimal in general. Specifically, [6] studied a SD-WTC with causal encoder CSI and full decoder CSI. Their coding scheme generated a cryptographic key from the state sequence, which is in turn used to one-time pad a part of the confidential message. The other part of the message is protected via wiretap coding. Despite the causality restriction, this strategy was shown to achieve strictly higher rates than the one from [4] for certain classes of SD-WTCs.*

C. Tight SS-Capacity Results

1) Reversely Less Noisy SD-WTC with Full Encoder and Noisy Decoder and Eavesdropper CSI: Let $\mathcal{S}_1, \mathcal{S}_2$ be finite sets and consider a SD-WTC $p_{\tilde{Y},\tilde{Z}|X,S}$ with non-causal encoder CSI, where $\tilde{Y} = (Y, S_1)$, $\tilde{Z} = (Z, S_2)$ and $p_{S_1,S_2,Y,Z|X,S} = p_{S_1,S_2|S} p_{Y,Z|X}$. Evidently, $p_{S_1,S_2,Y,Z|X,S}$ decomposes into a product of two WTCs, one independent of the state, and the other depends only on it. The legitimate receiver (respectively, the eavesdropper) observes both $\mathbf{Y}$ (respectively, $\mathbf{Z}$) from $p^n_{Y,Z|X}$ and $\mathbf{S}_1$ (respectively, $\mathbf{S}_2$). The latter is $\mathbf{S}$ passed through (the marginal of) $p^n_{S_1,S_2|S}$.

We characterize the SS-capacity of this setting when $p_{Y,Z|X}$ is reversely less noisy, i.e., when $I(U;Y) \leq I(U;Z)$, for every random variable U with $U-X-(Y,Z)$. After submitting this paper, we became aware of an independent derivation of this result under an average error probability and the weak-secrecy metric [10]. That work derived an achievable rate region based on secret key agreement for the WTC with generalized feedback. Although being different from the setup considered herein, both capture the less noisy SD-WTC as a special case.[4] Both achievability results (our Theorem 1 and [10, Theorem 1]) are tight for this instance.

To state the result, let $\mathcal{A},\mathcal{B}$ be finite sets, and for any $p_X \in \mathcal{P}(\mathcal{X})$, $p_{A|S}: \mathcal{S} \to \mathcal{P}(\mathcal{A})$ and $p_{B|A}: \mathcal{A} \to \mathcal{P}(\mathcal{B})$ define

$$\begin{aligned} &R_{\mathsf{RLN}}\left(p_X, p_{A|S}, p_{B|A}\right) \\ &\triangleq \min\Big\{I(A;S_1|B) - I(A;S_2|B), I(X;Y) - I(A;S|S_1)\Big\}, \end{aligned} \tag{24}$$

where the mutual information terms are with respect to $p_S p_{A|S} p_{B|A} p_X p_{S_1,S_2|S} p_{Y,Z|X}$, i.e., where (X,Y,Z) is independent of (S,S_1,S_2,A,B) and $A-S-(S_1,S_2)$ and $B-A-(S,S_1,S_2)$ form Markov chains (on top of the Markov relations induced by the channels).

Corollary 1 (Reversely Less Noisy SD-WTC SS-Capacity) *The SS-capacity of the reversely less noisy WTC with full encoder and noisy decoder and eavesdropper CSI is*

$$C_{\mathsf{RLN}} = \max_{p_X, p_{A|S}, p_{B|A}} R_{\mathsf{RLN}}\left(p_X, p_{A|S}, p_{B|A}\right). \tag{25}$$

A proof of Corollary 1, where the direct part is derived from Theorem 1, is given in Appendix E. Instead, one can establish achievability of (25) via an explicit coding scheme based on key agreement through multiple blocks and one-time pad operations. To gain insight, an outline of the scheme for the case where $S_2 = 0$ is given Below. It shows that in the absence of correlated observations with S at the eavesdropper's site, one may design a secure transmission strategy over a single block. Notwithstanding, a single block coding scheme is feasible even when S_2 is not a constant, via the superposition code from the proof of Theorem 1.

Explicit Achievability for Corollary 1: Observe that when $S_2 = 0$, setting $B=0$ in (25) is optimal. The resulting rate $\tilde{R}_{\mathsf{RLN}}\left(p_X, p_{A|S}\right) \triangleq \min\Big\{I(A;S_1), I(X;Y) - I(A;S|S_1)\Big\}$, for any fixed p_X and $p_{A|S}$ as before, is achieved as follows: [5]

i) Generate 2^{nR_A} a-codewords as i.i.d. samples from p^n_A.

ii) Partition the set of a-codewords into $2^{nR_{\mathsf{Bin}}}$ equal sized bins. Label each a-codeword as $\mathbf{a}(b,k)$, where $b \in \left[1 : 2^{nR_{\mathsf{Bin}}}\right]$ and $k \in \left[1 : 2^{n(R_A - R_{\mathsf{Bin}})}\right]$.

iii) Generate a point-to-point codebook that comprises $2^{n(R+R_{\mathsf{Bin}})}$ codewords $\mathbf{x}(m,b)$, where $m \in \mathcal{M}_n$ and $b \in \left[1 : 2^{nR_{\mathsf{Bin}}}\right]$, drawn from to p^n_X.

iv) Upon observing $\mathbf{s} \in \mathcal{S}^n$, the encoder searches the a-codebook for an a-codeword that is jointly-typical with $\mathbf{s}$, with respect to $p_S p_{A|S}$. Such a codeword is found with high probability if

$$R_A > I(A;S). \tag{26}$$

Let $(b,k) \in \left[1 : 2^{nR_{\mathsf{Bin}}}\right] \times \left[1 : 2^{n(R_A - R_{\mathsf{Bin}})}\right]$ be the indices of the selected a-codeword. To sent the message $m \in \mathcal{M}_n$, the encoder one-time-pads m with k to get $\tilde{m} = m \oplus k \in \mathcal{M}_n$, and transmits $\mathbf{x}(\tilde{m}, b)$ over the WTC. The one-time pad operation requires

$$R \leq R_A - R_{\mathsf{Bin}}. \tag{27}$$

[4] In fact, this is true for the slightly more general setup of the WTC with correlated sources [8].

[5] A reminiscent coding scheme was employed in [25] for the purpose of key generation (rather than secret message transmission) over the SD-WTC with non-causal encoder CSI.

v) The legitimate receiver first decodes the x-codeword using $\mathbf{y}$. Reliable decoding requires the total number of x-codewords to be less than the capacity $p_{Y|X}$, i.e.,

$$R + R_{\mathsf{Bin}} < I(X;Y). \tag{28}$$

Denoting the decoded indices by $(\hat{\tilde{m}}, \hat{b}) \in \mathcal{M}_n \times \big[1 : 2^{nR_{\mathsf{Bin}}}\big]$, the decoder then uses the noisy state observation $\mathbf{s}_1 \in \mathcal{S}_1^n$ to isolate the exact a-codeword from the $\hat{b}$-th bin. Namely, it searches for a unique index $\hat{k} \in \big[1 : 2^{n(R_A - R_{\mathsf{Bin}})}\big]$, such that $\big(\mathbf{a}(\hat{b},\hat{k}), \mathbf{s}_1\big)$ are jointly-typical with respect to p_{A,S_1} (the marginal of $p_S p_{S_1|S} p_{A|S}$). The probability of error in doing so is arbitrarily small with the blocklength, provided that

$$R_A - R_{\mathsf{Bin}} < I(A;S_1). \tag{29}$$

Having decoded $(\hat{\tilde{m}}, \hat{b})$ and $\hat{k}$, the decoder declares $\hat{m} \triangleq \hat{\tilde{m}} \oplus \hat{k}$ as the decoded message.

vi) For the eavesdropper, note that although it has the correct $(\tilde{m}, b)$ (due to the less noisy condition), it cannot decode k since it has no observation that is correlated with $\mathbf{A}$, $\mathbf{S}$ or $\mathbf{S}_1$. Security of the protocol, therefore, follows by the security of the one-time pad.

vii) Putting the above bounds together establishes the achievability of $\tilde{R}_{\mathsf{RLN}}\big(p_X, p_{A|S}\big)$.

2) Semi-Deterministic SD-WTC with Non-Causal Encoder CSI: Another observation is that R_{A} from Theorem 1 is tight when the main channel is deterministic, i.e., when $p_{Y,Z|X,S} = \mathbb{1}_{\big\{Y=y(X,S)\big\}} p_{Z|X,S}$, for some function $y : \mathcal{S} \times \mathcal{X} \to \mathcal{Y}$. In fact, the achievability results from [4], [5] are sufficient to attain optimality in this case. We state this secrecy-capacity result merely because, to the best of our knowledge, it was not explicitly stated before.

Corollary 2 (Semi-Deterministic SD-WTCM SS-Capacity) *The SS-capacity of the semi-deterministic SD-WTC with non-causal encoder CSI is*

$$C_{\mathsf{Semi-Det}} = \max_{p_{X|S}} \min\Big\{H(Y|Z), H(Y|S)\Big\}, \tag{30}$$

where the entropy terms are calculated with respect to $p_S p_{X|S} \mathbb{1}_{\big\{Y=y(X,S)\big\}} p_{Z|X,S}$.

The achievability of $C_{\mathsf{Semi-Det}}$ follows by setting $U = 0$ and $V = Y$ (a valid choice for deterministic channels) in Theorem 1. The converse follows by standard techniques – see Appendix F.

Note that the SS-capacity is unaffected by whether or not the eavesdropper's channel is deterministic. Letting $Z = z(X,S)$, for some $z : \mathcal{S} \times \mathcal{X} \to \mathcal{Z}$ does not changes the result of Corollary 2.

V. Proof of Theorem 1

Fix $\epsilon > 0$ and a conditional PMF $p_{U,V,X|S} : \mathcal{S} \to \mathcal{P}(\mathcal{U} \times \mathcal{V} \times \mathcal{X})$, for which the induced single-letter distribution

$$p \triangleq p_S p_{U,V,X|S} p_{Y,Z|X,S}, \tag{31}$$

satisfies $I(U;Y) - I(U;S) \geq 0$. Assume that $R < R_{\mathsf{A}}\big(p_{U,V,X|S}\big)$ and for any $n \in \mathbb{N}$, let $M \sim p_{\mathcal{M}_n}^{(U)}$ be a uniformly distributed message. We first prove the existence of codes with an arbitrarily small *average* error probability and a vanishing *strong secrecy* metric.[6] The expurgation method is then used to upgrade reliability to a vanishing *maximal* error probability and upgrade strong secrecy to SS.

Codebook $\boldsymbol{\mathcal{C}_n}$: We use a superposition codebook that encodes the confidential message in its outer layer. The codebook is drawn independently of the state sequence $\mathbf{S}$, but with sufficient redundancy to correlate the transmission with it.

Let I and J be independent uniform random variables over $\mathcal{I}_n \triangleq \big[1 : 2^{nR_1}\big]$ and $\mathcal{J}_n \triangleq \big[1 : 2^{nR_2}\big]$, respectively.[7] Let $\mathsf{C}_U^{(n)} \triangleq \big\{\mathbf{U}(i)\big\}_{i \in \mathcal{I}_n}$ be a random inner layer codebook – a set of random vectors of length n that are i.i.d. according to p_U^n. A realization of $\mathsf{C}_U^{(n)}$ is denoted by $\mathcal{C}_U^{(n)} \triangleq \big\{\mathbf{u}(i)\big\}_{i \in \mathcal{I}_n}$.

For the outer layer codebook, fix $\mathcal{C}_U^{(n)}$ and for every $i \in \mathcal{I}_n$ let $\mathsf{C}_V^{(n)}(i) \triangleq \big\{\mathbf{V}(i,j,m)\big\}_{(j,m) \in \mathcal{J}_n \times \mathcal{M}_n}$ be a collection of i.i.d. random vectors of length n with distribution $p^n_{V|U=\mathbf{u}(i)}$. A random outer layer codebook (with respect to an inner codebook $\mathcal{C}_U^{(n)}$) is $\mathsf{C}_V^{(n)} \triangleq \big\{\mathsf{C}_V^{(n)}(i)\big\}_{i \in \mathcal{I}_n}$. A realization of $\mathsf{C}_V^{(n)}(i)$, for $i \in \mathcal{I}_n$, is denoted by $\mathcal{C}_V^{(n)}(i) \triangleq \big\{\mathbf{v}(i,j,m)\big\}_{(j,m) \in \mathcal{J}_n \times \mathcal{M}_n}$, while $\mathcal{C}_V^{(n)}$ denotes a realization of $\mathsf{C}_V^{(n)}$. A random superposition codebook is $\mathsf{C}_n \triangleq \Big\{\mathsf{C}_U^{(n)}, \mathsf{C}_V^{(n)}\Big\}$, while $\mathcal{C}_n = \Big\{\mathcal{C}_U^{(n)}, \mathcal{C}_V^{(n)}\Big\}$ denotes a fixed codebook.

Let $\mathfrak{C}_n$ be the set of all possible outcomes of C_n. The above construction induces a PMF $\mu \in \mathcal{P}(\mathfrak{C}_n)$ over the codebook ensemble. For every $\mathcal{C}_n \in \mathfrak{C}_n$, we have

$$\mu(\mathcal{C}_n) = \prod_{i \in \mathcal{I}_b} p_U^n\big(\mathbf{u}(i)\big) \prod_{\substack{(\hat{i},j,m) \\ \in \mathcal{I}_n \times \mathcal{J}_n \times \mathcal{M}_m}} p^n_{V|U}\Big(\mathbf{v}\big(\hat{i},j,m\big)\Big|\mathbf{u}\big(\hat{i}\big)\Big). \tag{32}$$

The encoder and decoder, for a fixed superposition codebook $\mathcal{C}_n \in \mathfrak{C}_n$, are described next.

Encoder $\boldsymbol{f_{\mathcal{C}_n}}$: We utilize the likelihood-encoder [12], which enables to approximate the induced joint distribution by a simpler distribution used for the analysis.

To send $m \in \mathcal{M}_n$ upon observing $\mathbf{s} \in \mathcal{S}^n$, the encoder randomly chooses $(i,j) \in \mathcal{I}_n \times \mathcal{J}_n$ according to

$$\hat{P}^{(\mathcal{C}_n)}(i,j|m,\mathbf{s}) = \frac{p^n_{S|U,V}\big(\mathbf{s}\big|\mathbf{u}(i), \mathbf{v}(i,j,m)\big)}{\sum_{(i',j') \in \mathcal{I}_n \times \mathcal{J}_n} p^n_{S|U,V}\big(\mathbf{s}\big|\mathbf{u}(i'), \mathbf{v}(i',j',m)\big)}, \tag{33}$$

where $p_{S|U,V}$ is a conditional marginal distribution of p from (31). The channel input sequence is then generated by feeding the chosen u- and v-codewords along with the state sequence into a discrete and memoryless channel (DMC) $p_{X|U,V,S}$, i.e., it is a sample of $\mathbf{X} \sim p^n_{X|U=\mathbf{u}(i),V=\mathbf{v}(i,j,m),S=\mathbf{s}}$.

Accordingly, the (stochastic) encoding function $f_{\mathcal{C}_n} : \mathcal{M}_n \times \mathcal{S}^n \to \mathcal{P}(\mathcal{X}^n)$ is

$$f_{\mathcal{C}_n}(\mathbf{x}|m,\mathbf{s})$$

[6] Strong secrecy refers to $I(M;\mathbf{Z}) \to 0$ with the blocklength, where M is uniformly distributed.

[7] For simplicity of notation we assume that 2^{nR}, 2^{nR_1} and 2^{nR_2} are integers.

$$= \sum_{(i,j)\in\mathcal{I}_n\times\mathcal{J}_n} \hat{P}^{(\mathcal{C}_n)}(i,j|m,\mathbf{s})p^n_{X|U,V,S}\big(\mathbf{x}\big|\mathbf{u}(i),\mathbf{v}(i,j,m),\mathbf{s}\big), \tag{34}$$

for all $(m,\mathbf{s},\mathbf{x})\in\mathcal{M}_n\times\mathcal{S}^n\times\mathcal{X}^n$.

Decoder $\phi_{\mathcal{C}_n}$: Define three decoding functions:

1) $\phi_{\mathcal{C}_n}:\mathcal{Y}^n\to\hat{\mathcal{M}}_n$, which is the actual decoder of the message m.
2) $\psi^{(I)}_{\mathcal{C}_n}:\mathcal{Y}^n\to\hat{\mathcal{I}}_n$, where $\hat{\mathcal{I}}_n\triangleq\mathcal{I}_n\cup\{e\}$.
3) $\psi^{(J)}_{\mathcal{C}_n}:\mathcal{Y}^n\to\hat{\mathcal{J}}_n$, where $\hat{\mathcal{J}}_n\triangleq\mathcal{J}_n\cup\{e\}$.

Here, e is the same error symbol from the definition of $\hat{\mathcal{M}}_n$, which is assumed $e\notin\mathcal{M}_n\cup\mathcal{I}_n\cup\mathcal{J}_n$. The role of $\psi^{(I)}_{\mathcal{C}_n}$ and $\psi^{(J)}_{\mathcal{C}_n}$ is to decode the indices I and J, respectively. These functions will be used in the reliability analysis. Although, there is no reliability requirement on (I,J), the subsequently chosen codebook rates enable their decoding with high probability.

Upon observing $\mathbf{y}\in\mathcal{Y}^n$, the decoder searches for a unique triple $(\hat{i},\hat{j},\hat{m})\in\mathcal{I}_n\times\mathcal{J}_n\times\mathcal{M}_n$ such that

$$\Big(\mathbf{u}(\hat{i}),\mathbf{v}(\hat{i},\hat{j},\hat{m}),\mathbf{y}\Big)\in\mathcal{T}^n_\epsilon(p_{U,V,Y}). \tag{35}$$

If a unique triple is found, then $\phi_{\mathcal{C}_n}(\mathbf{y})=\hat{m}$, $\psi^{(I)}_{\mathcal{C}_n}(\mathbf{y})=\hat{i}$ and $\psi^{(J)}_{\mathcal{C}_n}(\mathbf{y})=\hat{j}$; otherwise, $\phi_{\mathcal{C}_n}(\mathbf{y})=\psi^{(I)}_{\mathcal{C}_n}(\mathbf{y})=\psi^{(J)}_{\mathcal{C}_n}(\mathbf{y})=e$.

The triple $(\mathcal{M}_n,f_{\mathcal{C}_n},\phi_{\mathcal{C}_n})$, defined with respect to the codebook $\mathcal{C}_n$, is an (n,R)-code c_n. The joint distribution $P^{(\mathcal{C}_n)}$ over $\mathcal{M}_n\times\mathcal{S}^n\times\mathcal{I}_n\times\mathcal{J}_n\times\mathcal{U}^n\times\mathcal{V}^n\times\mathcal{X}^n\times\mathcal{Y}^n\times\mathcal{Z}^n\times\hat{\mathcal{M}}_n$ induced by $\mathcal{C}_n$ is

$$\begin{aligned}&P^{(\mathcal{C}_n)}(\mathbf{s},m,i,j,\mathbf{u},\mathbf{v},\mathbf{x},\mathbf{y},\mathbf{z},\hat{m})\\&=p^n_S(\mathbf{s})\frac{1}{|\mathcal{M}_n|}\hat{P}^{(\mathcal{C}_n)}(i,j|m,\mathbf{s})\mathbb{1}_{\big\{\mathbf{u}=\mathbf{u}(i)\big\}\cap\big\{\mathbf{v}=\mathbf{v}(i,j,m)\big\}}\\&\quad\times p^n_{X|U,V,S}(\mathbf{x}|\mathbf{u},\mathbf{v},\mathbf{s})p^n_{Y,Z|X,S}(\mathbf{y},\mathbf{z}|\mathbf{x},\mathbf{s})\mathbb{1}_{\big\{\phi_{\mathcal{C}_n}(\mathbf{y})=\hat{m}\big\}}.\end{aligned} \tag{36}$$

Approximating Distribution: We next show that $P^{(\mathcal{C}_n)}$ is close in total variation to another distribution $Q^{(\mathcal{C}_n)}$, which we use for the reliability and security analyses. Let

$$\begin{aligned}&Q^{(\mathcal{C}_n)}(m,i,j,\mathbf{u},\mathbf{v},\mathbf{s},\mathbf{x},\mathbf{y},\mathbf{z},\hat{m})\\&=\frac{1}{|\mathcal{M}_n||\mathcal{I}_n||\mathcal{J}_n|}\mathbb{1}_{\big\{\mathbf{u}=\mathbf{u}(i)\big\}\cap\big\{\mathbf{v}=\mathbf{v}(i,j,m)\big\}}p^n_{S|U,V}(\mathbf{s}|\mathbf{u},\mathbf{v})\\&\quad\times p^n_{X|U,V,S}(\mathbf{x}|\mathbf{u},\mathbf{v},\mathbf{s})p^n_{Y,Z|X,S}(\mathbf{y},\mathbf{z}|\mathbf{x},\mathbf{s})\mathbb{1}_{\big\{\phi_{\mathcal{C}_n}(\mathbf{y})=\hat{m}\big\}}.\end{aligned} \tag{37}$$

For simplicity of notation, we sometimes abbreviate $P^{(\mathcal{C}_n)}_{\mathbf{S},M,I,J,\mathbf{U},\mathbf{V},\mathbf{X},\mathbf{Y},\mathbf{Z},\hat{M}}$ and $Q^{(\mathcal{C}_n)}_{M,I,J,\mathbf{U},\mathbf{V},\mathbf{S},\mathbf{X},\mathbf{Y},\mathbf{Z},\hat{M}}$ as $P^{(\mathcal{C}_n)}$ and $Q^{(\mathcal{C}_n)}$, respectively. The following lemma states sufficient conditions for the expected value of the total variation between $P^{(\mathsf{C}_n)}$ and $Q^{(\mathsf{C}_n)}$ to converge exponentially fast to zero.

Lemma 2 (Sufficient Conditions for Approximation) *If $(R_1,R_2)\in\mathbb{R}^2_+$ satisfy*

$$R_1>I(U;S) \tag{38a}$$

$$R_1+R_2>I(U,V;S), \tag{38b}$$

then there exist $\alpha>0$, such that for any n large enough

$$\mathbb{E}_\mu\Big|\Big|P^{(\mathsf{C}_n)}-Q^{(\mathsf{C}_n)}\Big|\Big|_{\mathsf{TV}}\leq e^{-n\alpha}. \tag{39}$$

The proof of Lemma 2 relies Lemmas 4 and 5 from Appendix A (see Appendix G for details). Lemma 2 is key in analyzing the performance of the proposed code.

Average Error Probability Analysis: For the reliability part, we first show that the average error probability can be made arbitrarily small. At the last step of this proof, the codebook is expurgated to attain a vanishing maximal error probability (in accordance with Definition 4). The main idea here is to use Lemma 2 to move away from analyzing the error probability under $P^{(\mathcal{C}_n)}$ to an analysis with respect to $Q^{(\mathcal{C}_n)}$. Analyzing the latter involves only standard typicality arguments.

The average error of a code c_n, with an underlying superposition codebook $\mathcal{C}_n$, is

$$e_a(\mathcal{C}_n)=\frac{1}{|\mathcal{M}_n|}\sum_{m\in\mathcal{M}_n}e_m(c_n)=\mathbb{P}_{P^{(\mathcal{C}_n)}}\big(\hat{M}\neq M\big), \tag{40}$$

where the subscript $P^{(\mathcal{C}_n)}$ on the RHS indicates that the probability measure is induced by $P^{(\mathcal{C}_n)}$ from (36).

We first show that a sufficient condition for the RHS of (40) to be arbitrarily small is that the average error probability induced by the $Q^{(\mathcal{C}_n)}$ PMF, i.e., $\mathbb{P}_{Q^{(\mathcal{C}_n)}}\big(\hat{M}\neq M\big)$, is small. Recall the following property of total variation (see, e.g., [12, Property (b)]). Let μ,ν be two probability measures on a $(\mathcal{X},\mathcal{F})$ and $g:\mathcal{X}\to\mathbb{R}$ be a non-negative measurable function bounded by $b\in\mathbb{R}$. It holds that

$$\big|\mathbb{E}_\mu g-\mathbb{E}_\nu g\big|\leq b\cdot\big|\big|\mu-\nu\big|\big|_{\mathsf{TV}}. \tag{41}$$

For every $n\in\mathbb{N}$, define $g_{\mathcal{C}_n}:\mathcal{M}_n\times\hat{\mathcal{M}}_n\to\mathbb{R}_+$ as $g_{\mathcal{C}_n}(m,\hat{m})=\mathbb{1}_{\{\hat{m}\neq m\}}$, and note that

$$\mathbb{E}_{P^{(\mathcal{C}_n)}}g_{\mathcal{C}_n}(M,\hat{M})=\mathbb{P}_{P^{(\mathcal{C}_n)}}\big(\hat{M}\neq M\big) \tag{42a}$$

$$\mathbb{E}_{Q^{(\mathcal{C}_n)}}g_{\mathcal{C}_n}(M,\hat{M})=\mathbb{P}_{Q^{(\mathcal{C}_n)}}\big(\hat{M}\neq M\big). \tag{42b}$$

For any $\mathcal{C}_n$ we thus have

$$\begin{aligned}&\Big|\mathbb{P}_{P^{(\mathcal{C}_n)}}\big(\hat{M}\neq M\big)-\mathbb{P}_{Q^{(\mathcal{C}_n)}}\big(\hat{M}\neq M\big)\Big|\\&\qquad\leq\Big|\Big|P^{(\mathcal{C}_n)}_{M,\hat{M}}-Q^{(\mathcal{C}_n)}_{M,\hat{M}}\Big|\Big|_{\mathsf{TV}}\\&\qquad\overset{(a)}{\leq}\Big|\Big|P^{(\mathcal{C}_n)}-Q^{(\mathcal{C}_n)}\Big|\Big|_{\mathsf{TV}},\end{aligned} \tag{43}$$

where (a) is because $\big|\big|p_X-q_X\big|\big|_{\mathsf{TV}}\leq\big|\big|p_{X,Y}-q_{X,Y}\big|\big|_{\mathsf{TV}}$, for any $p_{X,Y},q_{X,Y}\in\mathcal{P}(\mathcal{X}\times\mathcal{Y})$ with marginals p_X and q_X, respectively. Taking an expectation over the codebook ensemble, we have

$$\begin{aligned}&\mathbb{E}_\mu\mathbb{P}_{Q^{(\mathsf{C}_n)}}\big(\hat{M}\neq M\big)-\mathbb{E}_\mu\Big|\Big|P^{(\mathsf{C}_n)}-Q^{(\mathsf{C}_n)}\Big|\Big|_{\mathsf{TV}}\\&\quad\leq\mathbb{E}_\mu\mathbb{P}_{P^{(\mathsf{C}_n)}}\big(\hat{M}\neq M\big)\\&\quad\leq\mathbb{E}_\mu\mathbb{P}_{Q^{(\mathsf{C}_n)}}\big(\hat{M}\neq M\big)+\mathbb{E}_\mu\Big|\Big|\bar{P}^{(\mathsf{C}_n)}-Q^{(\mathsf{C}_n)}\Big|\Big|_{\mathsf{TV}}.\end{aligned} \tag{44}$$

Lemma 2 states that $\mathbb{E}_\mu\Big|\Big|P^{(\mathsf{C}_n)}-Q^{(\mathsf{C}_n)}\Big|\Big|$ can be made arbitrarily small with n, provided that (38) are satisfied. To show that the expected average error probability under $Q^{(\mathsf{C}_n)}$ also converges to 0 with n, consider the following arguments. For any codebook $\mathcal{C}_n\in\mathfrak{C}_n$ and $(\tilde{i},\tilde{j},\tilde{m})\in\mathcal{I}_n\times\mathcal{J}_n\times\mathcal{M}_n$,

$$
\begin{aligned}
\mathbb{E}_\mu \mathbb{P}_{Q^{(\mathsf{C}_n)}}\big(\hat{M}\neq M\big) &\overset{(a)}{\leq} \mathbb{E}_\mu \mathbb{P}_{Q^{(\mathsf{C}_n)}}\Big((\hat{M},\hat{I},\hat{J})\neq(M,I,J)\Big)\\
&\overset{(b)}{\leq} \mathbb{E}_\mu \mathbb{P}_{Q^{(\mathsf{C}_n)}}\Big((\hat{M},\hat{I},\hat{J})\neq(1,1,1)\Big|(M,I,J)=(1,1,1)\Big)\\
&\overset{(c)}{=} \mathbb{E}_\mu \mathbb{P}_{Q^{(\mathsf{C}_n)}}\Bigg(\mathcal{E}(1,1,1,\mathsf{C}_n)^c\cup\Big\{\bigcup_{\tilde{i}\neq 1}\mathcal{E}(\tilde{i},1,1,\mathsf{C}_n)\Big\}\\
&\qquad\qquad \cup\Big\{\bigcup_{(\tilde{j},\tilde{m})\neq(1,1)}\mathcal{E}(1,\tilde{j},\tilde{m},\mathsf{C}_n)\Big\}\cup\Big\{\bigcup_{(\tilde{i},\tilde{j},\tilde{m})\neq(1,1,1)}\mathcal{E}(\tilde{i},\tilde{j},\tilde{m},\mathsf{C}_n)\Big\}\Bigg)\\
&\overset{(d)}{\leq} \underbrace{\mathbb{P}_{p^n_{U,V,Y}}\Big((\mathbf{U},\mathbf{V},\mathbf{Y})\in\mathcal{T}^n_\epsilon(p_{U,V,Y})\Big)}_{P_1}+\underbrace{\sum_{\tilde{i}\neq 1}\mathbb{P}_{p^n_{U,V}\times p^n_Y}\Big((\mathbf{U},\mathbf{V},\mathbf{Y})\in\mathcal{T}^n_\epsilon(p_{U,V,Y})\Big)}_{P_2}\\
&\qquad +\underbrace{\sum_{(\tilde{j},\tilde{m})\neq(1,1)}\mathbb{P}_{p^n_{U,V}\times p^n_{Y|U}}\Big((\mathbf{U},\mathbf{V},\mathbf{Y})\in\mathcal{T}^n_\epsilon(p_{U,V,Y})\Big)}_{P_3}\\
&\qquad +\underbrace{\sum_{(\tilde{i},\tilde{j},\tilde{m})\neq(1,1,1)}\mathbb{P}_{p^n_{U,V}\times p^n_Y}\Big((\mathbf{U},\mathbf{V},\mathbf{Y})\in\mathcal{T}^n_\epsilon(p_{U,V,Y})\Big)}_{P_4},
\end{aligned}
\tag{48}
$$

define the event

$$\mathcal{E}(\tilde{i},\tilde{j},\tilde{m},\mathcal{C}_n)=\Big\{\big(\mathbf{u}(\tilde{i}),\mathbf{v}(\tilde{i},\tilde{j},\tilde{m}),\mathbf{Y}\big)\in\mathcal{T}^n_\epsilon(p_{U,V,Y})\Big\},\tag{45}$$

where $\mathbf{Y}\sim p^n_{Y|U=\mathbf{u}(\tilde{i}),V=\mathbf{v}(\tilde{i},\tilde{j},\tilde{m})}$ is the receiver's observation when $(\tilde{i},\tilde{j},\tilde{m})$ are sent over the effective DMC

$$p_{Y|U,V}(y|u,v)=\sum_{(s,x,z)\in\mathcal{S}\times\mathcal{X}\times\mathcal{Z}}p_{S|U,V}(s|u,v)p_{X|S,U,V}(x|s,u,v)\times p_{Y,Z|X,S}(y,z|x,s),\tag{46}$$

where $\forall(u,v,y)\in\mathcal{U}\times\mathcal{V}\times\mathcal{Y}$. The PMF $p_{U,V,Y}$ with respect to which we define the letter-typical set in (45) is a marginal of p from (31).

To bound the expected average error probability under $Q^{(\mathsf{C}_n)}$, for each $\mathcal{C}_n\in\mathfrak{C}_n$, we extend $Q^{(\mathcal{C}_n)}$ to the space $\mathcal{M}_n\times\mathcal{S}^n\times\mathcal{I}_n\times\mathcal{J}_n\times\mathcal{U}^n\times\mathcal{V}^n\times\mathcal{X}^n\times\mathcal{Y}^n\times\mathcal{Z}^n\times\hat{\mathcal{M}}_n\times\hat{\mathcal{I}}_n\times\hat{\mathcal{J}}_n$:

$$\begin{aligned}&Q^{(\mathcal{C}_n)}(m,i,j,\mathbf{u},\mathbf{v},\mathbf{s},\mathbf{x},\mathbf{y},\mathbf{z},\hat{m},\hat{i},\hat{j})\\&=Q^{(\mathcal{C}_n)}(m,i,j,\mathbf{u},\mathbf{v},\mathbf{s},\mathbf{x},\mathbf{y},\mathbf{z},\hat{m})\mathbb{1}_{\big\{\psi^{(I)}_{\mathcal{C}_n}(\mathbf{y})=\hat{i}\big\}\cap\big\{\psi^{(J)}_{\mathcal{C}_n}(\mathbf{y})=\hat{j}\big\}},\end{aligned}\tag{47}$$

thus allowing us to account for errors in decoding I and J. We have (48) at the top of this page, where:
(a) is because the probability of error in decoding M is upper bounded by that of decoding (I,J,M);
(b) follows by the symmetry of the code under $Q^{(\mathcal{C}_n)}$ with respect to (i,j,m);
(c) is the definition of the decoding rules $\phi_{\mathcal{C}_n}$, $\psi^{(I)}_{\mathcal{C}_n}$ and $\psi^{(J)}_{\mathcal{C}_n}$;
(d) uses the union bound and takes the expectation over the ensemble of codebooks.

By the law of large numbers $P_1\to 0$ as $n\to\infty$, while P_2, P_3 and P_4 also converge to 0 as n grows if

$$R+R_2<I(V;Y|U)\tag{49a}$$

$$R+R_1+R_2<I(U,V;Y).\tag{49b}$$

Specifically, (49a) implies that $P_3\to 0$ as $n\to\infty$, while (49b) ensures that both $P_2\to 0$ and $P_4\to 0$ as $n\to\infty$. A sufficient condition for the former is

$$R_1<I(U,V;Y).\tag{50}$$

However, (50) is redundant having (49b). Concluding, so long as (38) and (49) both hold, we have

$$\mathbb{E}_\mu e_a(\mathsf{C}_n)\xrightarrow[n\to\infty]{}0.\tag{51}$$

Security Analysis: The security analysis shows that under proper conditions the induced conditional distribution of $\mathbf{Z}$ given $(M,\mathbf{U})$ approximates the product distribution $p^n_{Z|U}$. To demonstrate this, we once again rely on the approximation of $P^{(\mathcal{C}_n)}$ through $Q^{(\mathcal{C}_n)}$. It is first shown that if strong secrecy is achieved under $Q^{(\mathcal{C}_n)}$, then it is also achieved under $P^{(\mathcal{C}_n)}$. Strong secrecy is then upgraded to SS through expurgation. Having that, it remains to show that security is attainable under $Q^{(\mathcal{C}_n)}$. The following lemma justifies that strong secrecy under $Q^{(\mathcal{C}_n)}$ implies strong secrecy under $P^{(\mathcal{C}_n)}$.

Lemma 3 (SS via Approximating Distribution) *Let $\mathcal{C}_n\in\mathfrak{C}_n$ be a superposition codebook for which there exists a*

$\beta_1 > 0$, *such that for all sufficiently large* n

$$\left|\left|P^{(\mathcal{C}_n)}_{M,\mathbf{Z}} - Q^{(\mathcal{C}_n)}_{M,\mathbf{Z}}\right|\right|_{\mathsf{TV}} \leq e^{-n\beta_1}. \tag{52}$$

Then, there exists a $\beta_2 > 0$, *such that for any* n *large enough (possibly larger than the* n *needed for* (52) *to be valid)*

$$\left|I_{P^{(\mathcal{C}_n)}}(M;\mathbf{Z}) - I_{Q^{(\mathcal{C}_n)}}(M;\mathbf{Z})\right| \leq e^{-n\beta_2}. \tag{53}$$

The proof of Lemma 3 is relegated to Appendix H. As subsequently shown, the existence of a codebook $\mathcal{C}_n$ that satisfies (52) follows by Lemma 2. For such a $\mathcal{C}_n$, we have

$$I_{P^{(\mathcal{C}_n)}}(M;\mathbf{Z}) \leq I_{Q^{(\mathcal{C}_n)}}(M;\mathbf{Z}) + e^{-n\beta_2}, \tag{54}$$

for n sufficiently large.

With that in mind, we focus on the mutual information term from the RHS of (54). For any $\mathcal{C}_n \in \mathfrak{C}_n$, we have

$$\begin{aligned}
I_{Q^{(\mathcal{C}_n)}}(M;\mathbf{Z}) &\leq I_{Q^{(\mathcal{C}_n)}}(M;I,\mathbf{U},\mathbf{Z})\\
&\overset{(a)}{=} I_{Q^{(\mathcal{C}_n)}}(M;\mathbf{Z}|I,\mathbf{U})\\
&= \mathsf{D}\Big(Q^{(\mathcal{C}_n)}_{M,\mathbf{Z}|I,\mathbf{U}}\Big|\Big|p^{(U)}_{\mathcal{M}_n}Q^{(\mathcal{C}_n)}_{\mathbf{Z}|I,\mathbf{U}}\Big|Q^{(\mathcal{C}_n)}_{I,\mathbf{U}}\Big)\\
&\overset{(b)}{=} \mathsf{D}\Big(Q^{(\mathcal{C}_n)}_{\mathbf{Z}|M,I,\mathbf{U}}\Big|\Big|Q^{(\mathcal{C}_n)}_{\mathbf{Z}|I,\mathbf{U}}\Big|p^{(U)}_{\mathcal{M}_n}Q^{(\mathcal{C}_n)}_{I,\mathbf{U}}\Big)\\
&\overset{(c)}{\leq} \mathsf{D}\Big(Q^{(\mathcal{C}_n)}_{\mathbf{Z}|M,I,\mathbf{U}}\Big|\Big|p^n_{Z|U}\Big|p^{(U)}_{\mathcal{M}_n}Q^{(\mathcal{C}_n)}_{I,\mathbf{U}}\Big),
\end{aligned} \tag{55}$$

where (a) is because M and $(I,\mathbf{U})$ are independent under $Q^{(\mathcal{C}_n)}$, (b) is by the relative entropy chain rule and because $Q^{(\mathcal{C}_n)}_{M|I,\mathbf{U}} = p^{(U)}_{\mathcal{M}_n}$, while (c) follows from

$$\begin{aligned}
&\mathsf{D}\Big(Q^{(\mathcal{C}_n)}_{\mathbf{Z}|M,I,\mathbf{U}}\Big|\Big|Q^{(\mathcal{C}_n)}_{\mathbf{Z}|I,\mathbf{U}}\Big|p^{(U)}_{\mathcal{M}_n}Q^{(\mathcal{C}_n)}_{I,\mathbf{U}}\Big)\\
&\qquad = \mathsf{D}\Big(Q^{(\mathcal{C}_n)}_{\mathbf{Z}|M,I,\mathbf{U}}\Big|\Big|p^n_{Z|U}\Big|p^{(U)}_{\mathcal{M}_n}Q^{(\mathcal{C}_n)}_{I,\mathbf{U}}\Big)\\
&\qquad\qquad - \mathsf{D}\Big(Q^{(\mathcal{C}_n)}_{\mathbf{Z}|I,\mathbf{U}}\Big|\Big|p^n_{Z|U}\Big|Q^{(\mathcal{C}_n)}_{I,\mathbf{U}}\Big)
\end{aligned} \tag{56}$$

and the non-negativity or relative entropy. The inequality from (55) is true for any $p_{Z|U}:\mathcal{U}\to\mathcal{P}(\mathcal{Z})$; we chose $p_{Z|U}$ as the conditional marginal p from (31).

Recall that $Q^{(\mathcal{C}_n)}_{I,\mathbf{U}} = p^{(U)}_{\mathcal{I}_n}\mathbb{1}_{\{\mathbf{U}=\mathbf{u}(I)\}}$ and take an expectation over the codebook ensemble on both sides of (55):

$$\begin{aligned}
&\mathbb{E}_\mu I_{Q^{(\mathsf{C}_n)}}(M;\mathbf{Z})\\
&\leq \mathbb{E}_\mu \mathsf{D}\Big(Q^{(\mathsf{C}_n)}_{\mathbf{Z}|M,I,\mathbf{U}}\Big|\Big|p^n_{Z|U}\Big|p^{(U)}_{\mathcal{M}_n\times\mathcal{I}_n}Q^{(\mathsf{C}_n)}_{\mathbf{U}|I}\Big)\\
&\overset{(a)}{=} \mathbb{E}_\mu \mathsf{D}\Big(Q^{(\mathsf{C}_n)}_{\mathbf{Z}|M=1,I=1,\mathbf{U}}\Big|\Big|p^n_{Z|U}\Big|Q^{(\mathsf{C}_n)}_{\mathbf{U}|I=1}\Big)\\
&= \mathbb{E}_\mu\left[\sum_{\mathbf{u}\in\mathcal{U}^n} Q^{(\mathsf{C}_n)}_{\mathbf{U}|I}(\mathbf{u}|1)\mathsf{D}\Big(Q^{(\mathsf{C}_n)}_{\mathbf{Z}|M=1,I=1,\mathbf{U}=\mathbf{u}}\Big|\Big|p^n_{Z|U=\mathbf{u}}\Big)\right]\\
&= \sum_{\mathbf{u}\in\mathcal{U}^n}\mathbb{E}_\mu\left[\mathbb{1}_{\{\mathbf{U}(1)=\mathbf{u}\}}\mathsf{D}\Big(Q^{(\mathsf{C}_n)}_{\mathbf{Z}|M=1,I=1,\mathbf{U}=\mathbf{u}}\Big|\Big|p^n_{Z|U=\mathbf{u}}\Big)\right]\\
&\overset{(b)}{=} \sum_{\mathbf{u}\in\mathcal{U}^n}\mathbb{E}_{\mathsf{C}^{(n)}_U}\Bigg[\mathbb{E}_{\mathsf{C}^{(n)}_V|\mathsf{C}^{(n)}_U}\Big\{\mathbb{1}_{\{\mathbf{U}(1)=\mathbf{u}\}}\\
&\qquad\qquad \times \mathsf{D}\Big(Q^{(\mathsf{C}_n)}_{\mathbf{Z}|M=1,I=1,\mathbf{U}=\mathbf{u}}\Big|\Big|p^n_{Z|U=\mathbf{u}}\Big)\Big|\mathsf{C}^{(n)}_U\Big\}\Bigg]
\end{aligned} \tag{57}$$

where (a) is by symmetry, while (b) is the law of total expectation. In step (b) above we switched from the notation $\mathbb{E}_\mu$ that emphasizes the distribution of the random codebook $\mathsf{C}_n = \left\{\mathsf{C}^{(n)}_U, \mathsf{C}^{(n)}_V\right\}$, to a notation that states the random variables themselves (and their possible conditioning).

The inner (conditional) expectation from the RHS of (57) is evaluated next. We present an argument for decorrelating the relative entropy inside the expectation and the inner layer random codebook $\mathsf{C}^{(n)}_U$. This enables removing the conditioning from the inner expectation, which simplifies the term and adjusts it to the framework of the SCL from [26, Corollary VII.5]. Applying this SCL, in turn, implies strong secrecy.

Fix $\mathbf{u}\in\mathcal{U}^n$, an inner layer codebook $\mathsf{C}^{(n)}_U = \mathcal{C}^{(n)}_U$, and consider the quantity

$$\begin{aligned}
&\mathbb{E}_{\mathsf{C}^{(n)}_V|\mathsf{C}^{(n)}_U=\mathcal{C}^{(n)}_U}\Big\{\mathbb{1}_{\{\mathbf{u}=\mathbf{u}(1)\}}\\
&\quad \times \mathsf{D}\Big(Q^{(\mathsf{C}_n)}_{\mathbf{Z}|M=1,I=1,\mathbf{U}=\mathbf{u}(1)}\Big|\Big|p^n_{Z|U=\mathbf{u}(1)}\Big)\Big|\mathsf{C}^{(n)}_U=\mathcal{C}^{(n)}_U\Big\}.
\end{aligned} \tag{58}$$

For each $\mathbf{u}\in\mathcal{U}^n$, let $\tilde{\mathsf{C}}^{(n)}_V(\mathbf{u}) \triangleq \left\{\tilde{\mathbf{V}}(\mathbf{u},j)\right\}_{j\in\mathcal{J}_n}$ be a collection of i.i.d. random vectors of length n, each distributed according to $p^n_{V|U=\mathbf{u}}$ independently of C_n. The collection $\tilde{\mathsf{C}}^{(n)}_V \triangleq \left\{\tilde{\mathsf{C}}^{(n)}_V(\mathbf{u})\right\}_{\mathbf{u}\in\mathcal{U}^n}$ is distributed according to

$$\tilde{\mu}\big(\tilde{\mathcal{C}}^{(n)}_V\big) = \prod_{\mathbf{u}\in\mathcal{U}^n}\prod_{j\in\mathcal{J}_n} p^n_{V|U}\big(\tilde{\mathbf{v}}(\mathbf{u},j)\big|\mathbf{u}\big), \tag{59}$$

where, as before, $\tilde{\mathcal{C}}^{(n)}_V(\mathbf{u}) \triangleq \left\{\tilde{\mathbf{v}}(\mathbf{u},j)\right\}_{j\in\mathcal{J}_n}$ is an outcome of $\tilde{\mathsf{C}}^{(n)}_V(\mathbf{u})$. For each $\mathbf{u}\in\mathcal{U}^n$ define a conditional PMF

$$\Pi^{\left(\tilde{\mathcal{C}}^{(n)}_V\right)}(j,\mathbf{v},\mathbf{z}|\mathbf{u}) \triangleq \frac{1}{|\mathcal{J}_n|}\mathbb{1}_{\left\{\mathbf{v}=\tilde{\mathbf{v}}(\mathbf{u},j)\right\}}p^n_{Z|U,V}(\mathbf{z}|\mathbf{u},\mathbf{v}). \tag{60}$$

Let $\mathsf{C}^{(n)}_V(1,1) \triangleq \left\{\mathbf{V}(1,j,1)\right\}_{j\in\mathcal{J}_n}$ be the collection of outer layer codewords from the codebook $\mathsf{C}^{(n)}_V(1)$ that correspond to $m = 1$. Note that the random distribution $Q^{(\mathcal{C}_n)}_{\mathbf{Z}|M=1,I=1,\mathbf{U}=\mathbf{u}(1)}$ is a function of $\mathsf{C}^{(n)}_V(1,1)$ only. Furthermore, whenever $\mathcal{C}^{(n)}_V(1,1) = \tilde{\mathcal{C}}^{(n)}_V\big(\mathbf{u}(1)\big)$, the distributions $Q^{(\mathsf{C}_n)}_{\mathbf{Z}|M=1,I=1,\mathbf{U}=\mathbf{u}(1)}$ and $\Pi^{\left(\tilde{\mathcal{C}}^{(n)}_V(\mathbf{u})\right)}_{\mathbf{Z}}$ are equal as PMFs on $\mathcal{Z}^n$. Since the set of possible outcomes of $\mathsf{C}^{(n)}_V(1,1)$ coincides with that of $\tilde{\mathsf{C}}^{(n)}_V\big(\mathbf{u}(1)\big)$, we may rewrite the conditional expectation from (58) as

$$\begin{aligned}
&\mathbb{E}_{\mathsf{C}^{(n)}_V|\mathsf{C}^{(n)}_U=\mathcal{C}^{(n)}_U}\Big\{\mathbb{1}_{\{\mathbf{u}=\mathbf{u}(1)\}}\\
&\qquad \times \mathsf{D}\Big(Q^{(\mathsf{C}_n)}_{\mathbf{Z}|M=1,I=1,\mathbf{U}=\mathbf{u}(1)}\Big|\Big|p^n_{Z|U=\mathbf{u}(1)}\Big)\Big|\mathsf{C}^{(n)}_U=\mathcal{C}^{(n)}_U\Big\}\\
&\qquad = \mathbb{1}_{\{\mathbf{u}=\mathbf{u}(1)\}}\mathbb{E}_{\tilde{\mu}}\mathsf{D}\Big(\Pi^{\left(\tilde{\mathsf{C}}^{(n)}_V\right)}_{\mathbf{Z}|\mathbf{U}=\mathbf{u}}\Big|\Big|p^n_{Z|U=\mathbf{u}}\Big).
\end{aligned} \tag{61}$$

This follows by the independence between μ and $\tilde{\mu}$. Inserting (61) into the RHS of (57), we obtain

$$\mathbb{E}_\mu I_{Q^{(\mathsf{C}_n)}}(M;\mathbf{Z})$$

$$\leq \sum_{\mathbf{u}\in\mathcal{U}^n} \mathbb{E}_{\mathsf{C}_U^{(n)}}\Bigg[\mathbb{1}_{\big\{\mathbf{U}(1)=\mathbf{u}\big\}}$$

$$\times\, \mathbb{E}_{\mathsf{C}_V^{(n)}|\mathsf{C}_U^{(n)}} \mathsf{D}\Big(Q^{(\mathsf{C}_n)}_{\mathbf{Z}|M=1,I=1,\mathbf{U}=\mathbf{u}}\Big|\Big|p^n_{Z|U=\mathbf{u}}\Big)\Big|\mathsf{C}_U^{(n)}\Bigg]$$

$$= \sum_{\mathbf{u}\in\mathcal{U}^n} \mathbb{E}_{\mathsf{C}_U^{(n)}}\Bigg[\mathbb{1}_{\big\{\mathbf{U}(1)=\mathbf{u}\big\}}\mathbb{E}_{\tilde{\mu}}\mathsf{D}\Big(\Pi^{\big(\tilde{\mathcal{C}}_V^{(n)}\big)}_{\mathbf{Z}|\mathbf{U}=\mathbf{u}}\Big|\Big|p^n_{Z|U=\mathbf{u}}\Big)\Bigg]$$

$$\overset{(a)}{=} \sum_{\mathbf{u}\in\mathcal{U}^n} q^n_U(\mathbf{u})\mathbb{E}_{\tilde{\mu}}\mathsf{D}\Big(\Pi^{\big(\tilde{\mathcal{C}}_V^{(n)}\big)}_{\mathbf{Z}|\mathbf{U}=\mathbf{u}}\Big|\Big|p^n_{Z|U=\mathbf{u}}\Big)$$

$$\overset{(b)}{=} \mathbb{E}_{\tilde{\mu}}\mathsf{D}\Big(q^n_U\Pi^{\big(\tilde{\mathcal{C}}_V^{(n)}\big)}_{\mathbf{Z}|\mathbf{U}}\Big|\Big|p^n_{U,Z}\Big) \tag{62}$$

where (a) is because $\mathbf{U}(1) \sim q^n_U$, while (b) uses the relative entropy chain rule. The RHS of (62) falls within the framework of [26, Corollary VII.5] and it converges exponentially fast to zero as $n \to \infty$, provided[8]

$$R_2 > I(V;Z|U). \tag{63}$$

Code Extraction: Summarizing the results up to this point, we have that so long as (38), (49) and (63) hold, $\mathbb{E}_\mu e_a(\mathsf{C}_n) \xrightarrow[n\to\infty]{} 0$ and, for sufficiently large n,

$$\mathbb{E}_\mu I_{Q^{(\mathsf{C}_n)}}(M;\mathbf{Z}) \leq e^{-n\tilde{\gamma}} \tag{64}$$

for some $\tilde{\gamma} > 0$ independent of n.

The Selection Lemma from [27, Lemma 5] implies the existence of a sequence of superposition codebooks $\big\{\mathcal{C}_n\big\}_{n\in\mathbb{N}}$ (giving rise to a sequence of (n,R)-codes $\big\{c_n\big\}_{n\in\mathbb{N}}$), for which

$$e_a(\mathcal{C}_n) \xrightarrow[n\to\infty]{} 0 \tag{65a}$$

$$I_{Q^{(\mathcal{C}_n)}}(M;\mathbf{Z}) \leq e^{-n\gamma}, \tag{65b}$$

where (65b) holds for n large enough and some $\gamma > 0$. Through the relation from (54), we further deduce that there exists $\delta > 0$ such that for sufficiently large n

$$I_{P^{(\mathcal{C}_n)}}(M;\mathbf{Z}) \leq e^{-n\delta}. \tag{66}$$

It is left to upgrade the vanishing average error probability and strong secrecy metric to a vanishing maximal error probability and SS. This is done by expurgating the superposition codebook [21, Theorem 7.7.1] (see also [28]). Let n be sufficiently large, so that

$$e_a(\mathcal{C}_n) = \frac{1}{\mathcal{M}_n}\sum_{m\in\mathcal{M}_n}\mathbb{P}_{P^{(\mathcal{C}_n)}}\big(\tilde{M}\neq m|M=m\big) \leq \frac{\epsilon}{3} \tag{67a}$$

$$I_{P^{(\mathcal{C}_n)}}(M;\mathbf{Z}) = \frac{1}{\mathcal{M}_n}\sum_{m\in\mathcal{M}_n}\mathsf{D}\Big(P^{(\mathcal{C}_n)}_{\mathbf{Z}|M=m}\Big|\Big|P^{(\mathcal{C}_n)}_{\mathbf{Z}}\Big) \leq \frac{\epsilon}{3}. \tag{67b}$$

The fraction of messages that induce an error probability greater than ϵ is less than $\frac{1}{3}$. Similarly, the fraction of messages with relative entropy greater than ϵ is less than $\frac{1}{3}$. Therefore, the fraction of offending messages is less than $\frac{2}{3}$. By removing them one obtains a new sequence of codes that is $\big\{\mathcal{C}^\star_n\big\}_{n\in\mathbb{N}}$, such that for every large enough n

$$\max_{m\in\mathcal{M}_n}\mathbb{P}_{P^{(\mathcal{C}^\star_n)}}\big(\tilde{M}\neq m|M=m\big) \leq \epsilon \tag{68a}$$

$$\max_{m\in\mathcal{M}_n}\mathsf{D}\Big(P^{(\mathcal{C}^\star_n)}_{\mathbf{Z}|M=m}\Big|\Big|P^{(\mathcal{C}^\star_n)}_{\mathbf{Z}}\Big) \leq \epsilon. \tag{68b}$$

The rate of the n-th code in the new sequence $\big\{\mathcal{C}^\star_n\big\}_{n\in\mathbb{N}}$ is $R - \frac{\log(3)}{n}$, and the loss is negligible for large n.

(68a) is the small maximal error probability requirement from (13a). It remains to show that (68b) implies SS. Recall that $P^{(\mathcal{C}^\star_n)}$ is induced by a uniformly distributed message, i.e., $P^{(\mathcal{C}^\star_n)} = p^{(U)}_{\mathcal{M}_n}$. For any $q \in \mathcal{P}(\mathcal{M}_n)$, let $P^{(\mathcal{C}^\star_n,q)}$ be the induced probability distribution when $M \sim q$. Namely, $P^{(\mathcal{C}^\star_n,q)}$ is given by (36), but with $q(m)$ instead of $\frac{1}{|\mathcal{M}_n|}$. For any $q \in \mathcal{P}(\mathcal{M}_n)$, consider the following:

$$I_{P^{(\mathcal{C}^\star_n,q)}}(M;\mathbf{Z})$$

$$= \sum_{m\in\mathcal{M}_n} q(m)\mathsf{D}\Big(P^{(\mathcal{C}^\star_n,q)}_{\mathbf{Z}|M=m}\Big|\Big|P^{(\mathcal{C}^\star_n,q)}_{\mathbf{Z}}\Big)$$

$$\overset{(a)}{=} \sum_{m\in\mathcal{M}_n} q(m)\Big[\mathsf{D}\Big(P^{(\mathcal{C}^\star_n,q)}_{\mathbf{Z}|M=m}\Big|\Big|P^{(\mathcal{C}^\star_n)}_{\mathbf{Z}}\Big) - \mathsf{D}\Big(P^{(\mathcal{C}^\star_n,q)}_{\mathbf{Z}}\Big|\Big|P^{(\mathcal{C}^\star_n)}_{\mathbf{Z}}\Big)\Big]$$

$$\leq \sum_{m\in\mathcal{M}_n} q(m)\max_{\tilde{m}\in\mathcal{M}_n}\mathsf{D}\Big(P^{(\mathcal{C}^\star_n,q)}_{\mathbf{Z}|M=\tilde{m}}\Big|\Big|P^{(\mathcal{C}^\star_n)}_{\mathbf{Z}}\Big)$$

$$\overset{(b)}{=} \max_{m\in\mathcal{M}_n}\mathsf{D}\Big(P^{(\mathcal{C}^\star_n)}_{\mathbf{Z}|M=m}\Big|\Big|P^{(\mathcal{C}^\star_n)}_{\mathbf{Z}}\Big)$$

$$\leq \epsilon, \tag{69}$$

where (a) follows by a similar reasoning as step (c) in the derivation of (55) (see (56)), while (b) is because $P^{(\mathcal{C}^\star_n,q)}_{\mathbf{Z}|M=m} = P^{(\mathcal{C}^\star_n)}_{\mathbf{Z}|M=m}$, for any $q \in \mathcal{P}(\mathcal{M}_n)$. Maximizing both sides of (69) over all $q \in \mathcal{P}(\mathcal{M}_n)$ establishes the SS requirement from (13b).

Finally, we apply Fourier-Motzkin Elimination on (38), (49) and (63), to eliminate R_1 and R_2. Doing so shows that any $R < R_{\mathsf{A}}\left(p_{U,V,X|S}\right)$ is achievable. Maximizing over all $p_{U,V,X|S}$ establishes Theorem 1.

Remark 8 (Alternative Security Analysis) *The security analysis shows that under the conditions* (38) *and* (63)*, the induced conditional distribution of* $\mathbf{Z}$ *given* $\mathbf{U}$ *and* M *approximates a product measure* $p^n_{Z|U}$*, on average over the messages. Since the inner layer codebook (encoded by* U*) carries no confidential information, this implies a vanishing information leakage. An alternative approach to establish this is to make the induced conditional distribution of* $\mathbf{Z}$ *given* M *(without the conditioning on* $\mathbf{U}$*) be a good proxy of* p^n_Z*. This also implies security because*

$$I_{P^{(\mathcal{C}_n)}}(M;\mathbf{Z}) \leq \frac{1}{|\mathcal{M}_n|}\sum_{m\in\mathcal{M}_n}\mathsf{D}\Big(P^{(\mathcal{C}_n)}_{\mathbf{Z}|M=m}\Big|\Big|p^n_Z\Big). \tag{70}$$

The SCL for superposition codebooks implies that the RHS of (70) *decays exponentially fast to 0, provided that*

$$R_1 > I(U;Z) \tag{71a}$$

[8] The original statement from [26, Corollary VII.5] deals with total variation rather than with relative entropy. Nonetheless, the result applies here as well due to Lemma 1. Namely, because over finite probability spaces an exponential decay of total variation implies an exponential decay of the corresponding relative entropy.

$$R_1 + R_2 > I(U,V;Z). \tag{71b}$$

Replacing (63) *with* (71) *and combining it with* (38) *and* (49), *achieves any R with*

$$\begin{aligned} R &\leq \tilde{R}_{\mathsf{A}}\left(p_{U,V,X|S}\right) \\ &\triangleq \min\Big\{ I(U,V;Y) - I(U,V;Z) \\ &\qquad\qquad ,I(V;Y|U), I(U,V;Y) - I(U,V;S)\Big\}. \end{aligned} \tag{72}$$

Seemingly, the best secrecy rates our scheme achieves is the maximum between the RHS of (72) *and $\tilde{R}_{\mathsf{A}}\left(p_{U,V,X|S}\right)$ from* (14). *However, a closer examination of $\tilde{R}_{\mathsf{A}}\left(p_{U,V,X|S}\right)$ reveals that when optimizing over all $p_{U,V,X|S}$, $\tilde{R}_{\mathsf{A}}\left(p_{U,V,X|S}\right)$ is actually redundant. To see this, notice that for any $p_{U,V,X|S}$, with $\tilde{R}_{\mathsf{A}}\left(p_{U,V,X|S}\right) \geq R_{\mathsf{A}}\left(p_{U,V,X|S}\right)$, taking $p_{\tilde{U},\tilde{V},\tilde{X}|S}$ with $\tilde{U} = 0$, $\tilde{V} = (U,V)_p$ and $p_{\tilde{X}|S,\tilde{U},\tilde{V}} = p_{X|S,U,V}$, where the subscript p in the definition of $\tilde{V}$ denotes that the random variables are distributed according to p, gives*

$$\begin{aligned} &R_{\mathsf{A}}\left(p_{\tilde{U},\tilde{V},\tilde{X}|S}\right) \\ &= \min\Big\{ I(U,V;Y) - I(U,V;Z), I(U,V;Y) - I(U,V;S)\Big\} \\ &\geq \tilde{R}_{\mathsf{A}}\left(p_{U,V,X|S}\right). \end{aligned}$$

This implies that R_{A} is at least as high as the maximal $\tilde{R}_{\mathsf{A}}\left(p_{U,V,X|S}\right)$.

Remark 9 (SS via Strong Soft-Covering) *The above proof establishes SS via expurgation. The random coding argument first produces a sequence of codes that attain strong secrecy. Then, the messages with the highest information leakage are eliminated to obtain SS. Another approach is to derive SS directly from the random coding argument using a pair of strong SCLs. Namely, using Lemma 4 one can show that the probability that the the approximation from* (39) *fails is doubly-exponentially small. Having that, the heterogeneous strong SCL from [29, Lemma 1] further implies that $P^{(\mathsf{C}_n)}_{\mathbf{Z}|M=m,\mathbf{U}}$ is close in total variation to $p^n_{Z|U}$, for each $m \in \mathcal{M}_n$ (rather than on average as above). The continuity of mutual information over discrete probability spaces with respect to total variation would then imply SS with (doubly-exponentially) high probability, with respect to the random coding ensemble. Although this approach is not necessary here, we note it because it applies in various scenarios where the expurgation argument fails. Such scenarios include compound or arbitrarily varying settings, as well as cases where instead of (or in addition to) a secret message transmission, the legitimate parties aim to agree upon a semantically secured secret key. A key is typically required to be approximately uniform; however, expurgation can alter the distribution of the key. Strong soft-covering arguments enable SS proofs in all these aforementioned instances (see [24], [27], [29]).*

VI. Summary and Concluding Remarks

We studied SD-WTCs with non-causal encoder CSI. A novel lower bound on the SS-capacity was derived. Our coding scheme is based on a superposition codebook, which encodes the confidential message in the outer layer. The codebook has sufficient redundancy to facilitate correlating both layers and the transmission with the observed state sequence. The correlation is attained using the likelihood encoder [12]. SS is ensured via distribution approximation arguments and the expurgation technique. The structure of the rate bounds for secrecy implies that the eavesdropper can decode the inner layer codeword. Since no confidential information is encoded in the inner layer, this doesn't compromise security. The gain from doing so is that decoding the inner layer exhausts the eavesdropper's channel resources. Consequently, this prevents him from inferring any information on the outer layer, which contains the confidential message.

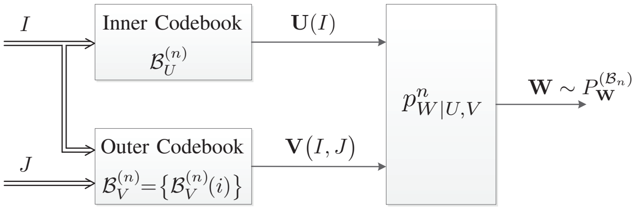

Fig. 2. Superposition soft-covering setup with the goal of making $P^{(\mathcal{B}_n)}_{\mathbf{W}} \approx p^n_W$, where $\mathcal{B}_n = \left\{\mathcal{B}^{(n)}_U, \mathcal{B}^{(n)}_V\right\}$ is a fixed superposition codebook.

Our result was compared to several previous achievability bounds from the literature. Notably, a comparison to the best past achievable scheme for the SD-WTC with non-causal encoder CSI from [11] revealed that our scheme not only captures it as a special case, but also strictly outperforms it in some cases. We further showed that our scheme achieve the SS-capacity of the reversely less noisy SD-WTC and the semi-deterministic SD-WTD, where $Y = y(X,S)$ is the legitimate receiver's observation. The latter can also be retrieved from [11], and even from the simpler achievable regions of [4], [5].

Appendix A
Soft-Covering Lemmas

A. Strong Soft-Covering Lemma for Superposition Codes

The SS analysis for the SD-WTC with non-causal encoder CSI relies on a SCL for superposition codes. Here, we give a strong version of this lemma (in the spirit of [27], [29]). The proof of Theorem 1 only uses a classic soft-covering statement (i.e., convergence in expectation). We present the stronger version for two reasons. First, the SS derivation in the proof of Theorem 1 can be preformed directly using the stronger version. Second, we believe that the sharp claim of Lemma 4 could prove useful for other research problems.

The setup is illustrated in Fig. 2, where inner and outer layer codewords are uniformly chosen and passed through a DMC to produce an output sequence. The induced distribution of the output should be asymptotically indistinguishable from a product measure. The approximation is in terms of relative entropy, which is shown to converge to 0 exponentially quickly with high probability. The negligible probability is doubly-exponentially small with the blocklength n.

Fix $p_{U,V,W} \in \mathcal{P}(\mathcal{U}\times\mathcal{V}\times\mathcal{W})$ and let I and J be independent and uniformly distributed over $\mathcal{I}_n \triangleq \left[1 : 2^{nR_1}\right]$ and $\mathcal{J}_n \triangleq$

$\left[1 : 2^{nR_2}\right]$, respectively. Let $\mathsf{B}_U^{(n)} \triangleq \left\{\mathbf{U}(i)\right\}_{i\in\mathcal{I}_n}$ be a random inner layer codebook – a set of random vectors of length n that are i.i.d. according to p_U^n. A realization of $\mathsf{B}_U^{(n)}$ is denoted by $\mathcal{B}_U^{(n)} \triangleq \left\{\mathbf{u}(i)\right\}_{i\in\mathcal{I}_n}$.

For the outer layer codebook, fix $\mathcal{B}_U^{(n)}$, and for every $i \in \mathcal{I}_n$, let $\mathsf{B}_V^{(n)}(i) \triangleq \left\{\mathbf{V}(i,j)\right\}_{j\in\mathcal{J}_n}$ be a collection of i.i.d. random vectors of length n with distribution $p^n_{V|U=\mathbf{u}(i)}$. A random outer layer codebook (with respect to an inner codebook $\mathcal{B}_U^{(n)}$) is $\mathsf{B}_V^{(n)} \triangleq \left\{\mathsf{B}_V^{(n)}(i)\right\}_{i\in\mathcal{I}_n}$. An outcome of $\mathsf{B}_V^{(n)}(i)$, for $i \in \mathcal{I}_n$ is denoted by $\mathcal{B}_V^{(n)}(i) \triangleq \left\{\mathbf{v}(i,j,m)\right\}_{j\in\mathcal{J}_n}$. We also use $\mathcal{B}_V^{(n)}$ to denote an outcome of $\mathsf{B}_V^{(n)}$. A random superposition codebook $\mathsf{B}_n \triangleq \left\{\mathsf{B}_U^{(n)}, \mathsf{B}_V^{(n)}\right\}$, while $\mathcal{B}_n \triangleq \left\{\mathcal{B}_U^{(n)}, \mathcal{B}_V^{(n)}\right\}$ denotes a fixed codebook.

Letting $\mathfrak{B}_n$ be the set of all possible outcomes of B_n, the above construction induces a distribution $\mu \in \mathcal{P}(\mathfrak{B}_n)$ over the codebook ensemble. For every $\mathcal{B}_n \in \mathfrak{B}_n$, we have

$$\nu(\mathcal{B}_n) = \prod_{i\in\mathcal{I}_b} p_U^n\big(\mathbf{u}(i)\big) \prod_{\left(\hat{i},j\right)\in\mathcal{I}_n\times\mathcal{J}_n} p^n_{V|U}\left(\mathbf{v}\big(\hat{i},j\big)\middle|\mathbf{u}\big(\hat{i}\big)\right). \tag{73}$$

For a fixed superposition code $\mathcal{B}_n$, the output sequence $\mathbf{W}$ is generated by independently drawing I and J from $\mathcal{I}_n$ and $\mathcal{J}_n$, respectively, and feeding $\mathbf{u}(i)$ and $\mathbf{v}(i,j)$ into the DMC $p^n_{W|U,V}$. The induced distribution on $\mathcal{I}_n \times \mathcal{J}_n \times \mathcal{U}^n \times \mathcal{V}^n \times \mathcal{W}^n$ by $P^{(\mathcal{B}_n)}$ is[9]

$$\begin{aligned} &P^{(\mathcal{B}_n)}(i,j,\mathbf{u},\mathbf{v},\mathbf{w}) \\ &\quad = 2^{-n(R_1+R_2)} \mathbb{1}_{\left\{\mathbf{u}=\mathbf{u}(i)\right\}\cap\left\{\mathbf{v}=\mathbf{v}(i,j)\right\}} p^n_{W|U,V}(\mathbf{w}|\mathbf{u},\mathbf{v}). \end{aligned} \tag{74}$$

Accordingly, the induced output distribution is

$$P^{(\mathcal{B}_n)}(\mathbf{w}) = 2^{-n(R_1+R_2)} \sum_{(i,j)\in\mathcal{I}_n\times\mathcal{J}_n} p^n_{W|U,V}\big(\mathbf{w}\big|\mathbf{u}(i),\mathbf{v}(i,j)\big). \tag{75}$$

We also set

$$P(\mathcal{B}_n, i, j, \mathbf{u}, \mathbf{v}, \mathbf{w}) \triangleq \mu(\mathcal{B}_n) P^{(\mathcal{B}_n)}(i,j,\mathbf{u},\mathbf{v},\mathbf{w}), \tag{76}$$

and denote by $\mathbb{P} \triangleq \mathbb{P}_P$ the probability measure induced by P. This notation is used in the remainder of this section and in the proof of the following strong SCL. When switching to other probability measures, we do so in accordance with the notation defined in Section II.

Lemma 4 (Strong Superposition SCL) *For any $p_{U,V,W}$, where $|\mathcal{W}| < \infty$, and $(R_1,R_2) \in \mathbb{R}_+^2$ with*

$$R_1 > I(U;W) \tag{77a}$$

$$R_1 + R_2 > I(U,V;W), \tag{77b}$$

there exist $\gamma_1, \gamma_2 > 0$, such that for n large enough

$$\mathbb{P}_\mu\left(\mathsf{D}\left(P_{\mathbf{W}}^{(\mathsf{B}_n)}\middle|\middle|p_W^n\right) > e^{-n\gamma_1}\right) \leq e^{-e^{n\gamma_2}}. \tag{78}$$

The proof of the lemma is relegated to Appendix B, where exact exponents is found.

B. Strong Soft-Covering Implies Classic Soft-Covering

The strong superposition SCL stated above implies the convergence to zero of the corresponding expected value [27, Lemma 2]. The expected value result is used for SS analysis in the proof of Theorem 1. For completeness, we next restate Lemma 2 from [27]; the proof is omitted.

Lemma 5 (Stronger than Classic Soft-Covering) *Under the framework of Lemma 4, let $\gamma_1, \gamma_2 > 0$ be such that (78) holds for n large enough. Then, for every such n we have*

$$\mathbb{E}_\mu \mathsf{D}\left(P_{\mathbf{W}}^{(\mathsf{B}_n)}\middle|\middle|p_W^n\right) \leq e^{-n\gamma_1} + n\log\left(\frac{1}{\mu_W}\right)e^{-e^{n\gamma_2}}, \tag{79}$$

where $\mu_W = \min_{w\in\mathrm{supp}(p_W)} p_W(w) > 0$.

Appendix B
Proof of Lemma 4

We state the proof in terms of arbitrary distributions (not necessarily discrete). When needed, we will specialize to the case where $\mathcal{W}$ is finite. For any fixed superposition codebook $\mathcal{B}_n$, denote the Radon-Nikodym derivative of the induced distribution with respect to p_W^n by

$$\Delta_{\mathcal{B}_n}(\mathbf{w}) \triangleq \frac{dP_{\mathbf{W}}^{(\mathcal{B}_n)}}{dp_W^n}(\mathbf{w}). \tag{80}$$

In the discrete case, $\Delta_{\mathcal{B}_n}$ is a ratio of PMFs. Accordingly, the relative entropy of interest, which is a function of $\mathcal{B}_n$, is

$$\mathsf{D}\left(P_{\mathbf{W}}^{(\mathcal{B}_n)}\middle|\middle|p_W^n\right) = \int dP_{\mathbf{W}}^{(\mathcal{B}_n)} \log \Delta_{\mathcal{B}_n}. \tag{81}$$

To describe the jointly-typical set over u-, v- and w-sequences, we first define information density $i_{p_{W|U}} : \mathcal{U} \times \mathcal{W} \to \mathbb{R}_+$ and $i_{p_{W|U,V}} : \mathcal{U} \times \mathcal{V} \times \mathcal{W} \to \mathbb{R}_+$ as

$$i_{p_{U,W}}(u,w) \triangleq \log\left(\frac{dp_{W|U=u}}{dp_W}(w)\right) \tag{82a}$$

$$i_{p_{U,V,W}}(u,v,w) \triangleq \log\left(\frac{dp_{W|U=u,V=v}}{dp_W}(w)\right). \tag{82b}$$

In (82), the arguments of the logarithms are the Radon-Nikodym derivatives of $p_{W|U=u}$ and $p_{W|U=u,V=v}$, respectively, with respect to p_W. Let $\epsilon_1, \epsilon_2 \geq 0$ be arbitrary (determined later) and define $\mathcal{A}_{\epsilon_1,\epsilon_2}$ as the set of all $(\mathbf{u},\mathbf{v},\mathbf{w}) \in \mathcal{U}^n \times \mathcal{V}^n \times \mathcal{W}^n$ satisfying

$$\begin{aligned} \frac{1}{n} i_{p^n_{U,W}}(\mathbf{u},\mathbf{w}) &< I(U;W) + \epsilon_1 \\ \frac{1}{n} i_{p^n_{U,V,W}}(\mathbf{u},\mathbf{v},\mathbf{w}) &< I(U,V;W) + \epsilon_2. \end{aligned} \tag{83}$$

Note that

$$i_{p^n_{U,W}}(\mathbf{u},\mathbf{w}) = \sum_{t=1}^n i_{p_{U,W}}(u_t,w_t) \tag{84a}$$

$$i_{p^n_{U,V,W}}(\mathbf{u},\mathbf{v},\mathbf{w}) = \sum_{t=1}^n i_{p_{U,V,W}}(u_t,v_t,w_t), \tag{84b}$$

[9] To simplify notation, from here on we assume that quantities of the form 2^{nR}, where $n \in \mathbb{N}$ and $R \in \mathbb{R}_+$, are integers. Otherwise, simple modifications of some of the subsequent expressions using floor operations are required.

$$\begin{aligned}\int dP_{\mathcal{B}_n,2} &= 1 - \int dP_{\mathcal{B}_n,1}\\ &= 2^{-n(R_1+R_2)} \sum_{(i,j)\in\mathcal{I}_n\times\mathcal{J}_n} \mathbb{P}_{p^n_{W|U,V}}\Big(\big(\mathbf{u}(i),\mathbf{v}(i,j,m),\mathbf{W}\big)\notin\mathcal{A}_{\epsilon_1,\epsilon_2}\Big|\mathbf{U}=\mathbf{u}(i),\mathbf{V}=\mathbf{v}(i,j)\Big). \qquad (88)\end{aligned}$$

$$\int dP_{\mathsf{B}_n,2} = 2^{-n(R_1+R_2)} \sum_{(i,j)\in\mathcal{I}_n\times\mathcal{J}_n} \mathbb{P}_{p^n_{W|U,V}}\Big(\big(\mathbf{U}(i),\mathbf{V}(i,j),\mathbf{W}\big)\notin\mathcal{A}_{\epsilon_1,\epsilon_2}\Big|\mathbf{U}=\mathbf{U}(i),\mathbf{V}=\mathbf{V}(i,j)\Big), \qquad (89)$$

$$\begin{aligned}&\mathbb{E}_\mu\mathbb{P}_{p^n_{W|U,V}}\Big(\big(\mathbf{U}(i),\mathbf{V}(i,j),\mathbf{W}\big)\notin\mathcal{A}_{\epsilon_1,\epsilon_2}\Big|\mathbf{U}=\mathbf{U}(i),\mathbf{V}=\mathbf{V}(i,j)\Big)\\ &=\mathbb{P}_{p^n_{U,V,W}}\Big(\big(\mathbf{U},\mathbf{V},\mathbf{W}\big)\notin\mathcal{A}_{\epsilon_1,\epsilon_2}\Big)\\ &=\mathbb{P}_{p^n_{U,V,W}}\left(\left\{\sum_{t=1}^n i_{p_{U,W}}(U_t,p_t)\geq n\big(I(U;W)+\epsilon_1\big)\right\}\bigcup\left\{\sum_{t=1}^n i_{p_{U,V,W}}(U_t,V_t,p_t)\geq n\big(I(U,V;W)+\epsilon_2\big)\right\}\right)\\ &\leq\mathbb{P}_{p^n_{U,V,W}}\left(2^{\lambda\sum_{t=1}^n i_{p_{U,W}}(U_t,p_t)}\geq 2^{n\lambda(I(U;W)+\epsilon_1)}\right)+\mathbb{P}_{p^n_{U,V,W}}\left(2^{\lambda\sum_{t=1}^n i_{p_{U,V,W}}(U_t,V_t,p_t)}\geq 2^{n\lambda(I(U,V;W)+\epsilon_2)}\right), \qquad (90)\end{aligned}$$

We split $P^{(\mathcal{B}_n)}_{\mathbf{W}}$ into two parts using indicator functions. For every $\mathbf{w}\in\mathcal{W}^n$, define

$$\begin{aligned}P_{\mathcal{B}_n,1}(\mathbf{v}) &\triangleq 2^{-n(R_1+R_2)}\sum_{(i,j)\in\mathcal{I}_n\times\mathcal{J}_n} p^n_{W|U,V}\big(\mathbf{w}\big|\mathbf{u}(i),\mathbf{v}(i,j)\big)\\ &\qquad\qquad\times\mathbb{1}_{\left\{\big(\mathbf{u}(i),\mathbf{v}(i,j),\mathbf{w}\big)\in\mathcal{A}_{\epsilon_1,\epsilon_2}\right\}}\\ P_{\mathcal{B}_n,2}(\mathbf{v}) &\triangleq 2^{-n(R_1+R_2)}\sum_{(i,j)\in\mathcal{I}_n\times\mathcal{J}_n} p^n_{W|U,V}\big(\mathbf{w}\big|\mathbf{u}(i),\mathbf{v}(i,j)\big)\\ &\qquad\qquad\times\mathbb{1}_{\left\{\big(\mathbf{u}(i),\mathbf{v}(i,j),\mathbf{w}\big)\notin\mathcal{A}_{\epsilon_1,\epsilon_2}\right\}}. \qquad (85)\end{aligned}$$

The measures $P_{\mathcal{B}_n,1}$ and $P_{\mathcal{B}_n,2}$ on the space $\mathcal{W}^n$ are not probability measures, but $P_{\mathcal{B}_n,1}+P_{\mathcal{B}_n,2}=P^{(\mathcal{B}_n)}_{\mathbf{W}}$ for each codebook $\mathcal{B}_n$. For every $\mathbf{w}\in\mathcal{W}^n$, we also define

$$\Delta_{\mathcal{B}_n,j}(\mathbf{w})\triangleq\frac{dP_{\mathcal{B}_n,j}}{dp^n_W}(\mathbf{w}),\quad j=1,2. \qquad (86)$$

With respect to the above definitions, Lemma 6 states an upper bound on the relative entropy of interest.

Lemma 6 *For every fixed superposition codebook $\mathcal{B}_n$, we have*

$$\begin{aligned}\mathsf{D}\Big(P^{(\mathcal{B}_n)}_{\mathbf{W}}\Big|\Big|p^n_W\Big)\leq h\left(\int dP_{\mathcal{B}_n,1}\right)&+\int dP_{\mathcal{B}_n,1}\log\Delta_{\mathcal{B}_n,1}\\ &+\int dP_{\mathcal{B}_n,2}\log\Delta_{\mathcal{B}_n,2}, \qquad (87)\end{aligned}$$

where $h(\cdot)$ is the binary entropy function.

The proof of the lemma is omitted as it follows the same steps as in the proof of [27, Lemma 3] (see Appendix B therein). Based on Lemma 6, to prove Lemma 4 it suffices to show that the probability (with respect to a random superposition codebook) of the RHS of (87) not vanishing exponentially fast to 0 as $n\to\infty$, is double-exponentially small.

Note that $P_{\mathcal{B}_n,1}$ usually contains almost all of the probability mass. That is, for fixed $\mathcal{B}_n$, we have (88) at the top of this page, which becomes (89) when the codebook is random. In (89), the RHS is an average of exponentially many i.i.d. random variables bounded between 0 and 1. The expected value of each is the exponentially small probability of correlated sequences being atypical, as seen in (90), where the last inequality uses the union bound and holds for any $\lambda\geq 0$.

We further bound the two terms from the RHS of (90) by exponentially decaying functions of n as follows. For the first term, consider:

$$\begin{aligned}&\mathbb{P}_{p^n_{U,V,W}}\left(2^{\lambda\sum_{t=1}^n i_{p_{U,W}}(U_t,p_t)}\geq 2^{n\lambda(I(U;W)+\epsilon_1)}\right)\\ &\overset{(a)}{\leq}\frac{\mathbb{E}_{p^n_{U,W}}2^{\lambda\sum_{t=1}^n i_{p_{U,W}}(U_t,p_t)}}{2^{n\lambda(I(U;W)+\epsilon_1)}}\\ &=\left(\frac{\mathbb{E}_{p_{U,W}}2^{\lambda i_{p_{U,W}}(U,W)}}{2^{\lambda(I(U;W)+\epsilon_1)}}\right)^n\\ &\overset{(b)}{=}2^{n\lambda\left(\frac{1}{\lambda}\log_2\mathbb{E}_{p_{U,W}}\left[2^{\lambda i_{p_{U,W}}(U;W)}\right]-I(U;W)-\epsilon_1\right)}\\ &\overset{(c)}{=}2^{n\lambda\left(d_{\lambda+1}(p_{U,W},p_Up_W)-I(U;W)-\epsilon_1\right)}, \qquad (91)\end{aligned}$$

where (a) is Markov's inequality, (b) follows by restricting λ to be strictly positive, while (c) is the definition of Rényi divergence of order $\lambda+1$. We use units of bits for mutual information and Rényi divergence to coincide with the base two expression of rate. For the second term from the RHS of (90), we have

$$\begin{aligned}&\mathbb{P}_{p^n_{U,V,W}}\left(2^{\lambda\sum_{t=1}^n i_{p_{U,V,W}}(U_t,V_t,p_t)}\geq 2^{n\lambda(I(U,V;W)+\epsilon_2)}\right)\\ &\qquad\leq 2^{n\lambda\left(d_{\lambda+1}(p_{U,V,W},p_{U,V}p_W)-I(U,V;W)-\epsilon_2\right)}. \qquad (92)\end{aligned}$$

Substituting $\alpha=\lambda+1$ into (91)-(92) gives

$$\begin{aligned}&\mathbb{E}_\mu\mathbb{P}_{p^n_{W|U,V}}\Big(\big(\mathbf{U}(i),\mathbf{V}(i,j),\mathbf{W}\big)\notin\mathcal{A}_{\epsilon_1,\epsilon_2}\Big|\mathbf{U}=\mathbf{U}(i),\mathbf{V}=\mathbf{V}(i,j)\Big)\\ &\leq 2^{-n\beta^{(1)}_{\alpha,\epsilon_1}}+2^{-n\beta^{(2)}_{\alpha,\epsilon_2}}, \qquad (93)\end{aligned}$$

where

$$\beta^{(1)}_{\alpha,\epsilon_1} = (\alpha-1)\big(I(U;W) + \epsilon_1 - d_\alpha(p_{U,W}, p_U p_W)\big), \quad (94a)$$

$$\beta^{(2)}_{\alpha,\epsilon_2} = (\alpha-1)\big(I(U,V;W) + \epsilon_2 - d_\alpha(p_{U,V,W}, p_{U,V} p_W)\big), \quad (94b)$$

for every $\alpha > 1$ and $\epsilon_1, \epsilon_2 \geq 0$, over which we may optimize. The optimal choices of ϵ_1 and ϵ_2 are apparent when all bounds of the proof are considered together (some yet to be derived). Proceeding, fix $\delta_1 \in \big(0, R_1 - I(U;W)\big)$ and $\delta_2 \in \big(0, R_1 + R_2 - I(U,V;W)\big)$, and for any $\alpha > 1$ set

$$\epsilon^{(1)}_{\alpha,\delta_1} = \frac{\frac{1}{2}(R_1 - \delta_1) + (\alpha-1) d_\alpha(p_{U,W}, p_U p_W)}{\frac{1}{2} + (\alpha-1)} - I(U;W), \quad (95a)$$

$$\epsilon^{(2)}_{\alpha,\delta_2} = \frac{\frac{1}{2}(R_1 + R_2 - \delta_2) + (\alpha-1) d_\alpha(p_{U,V,W}, p_{U,V} p_W)}{\frac{1}{2} + (\alpha-1)} - I(U,V;W). \quad (95b)$$

Substituting into $\beta^{(1)}_{\alpha,\epsilon_1}$ and $\beta^{(2)}_{\alpha,\epsilon_2}$ gives

$$\beta^{(1)}_{\alpha,\delta_1} \triangleq \beta^{(1)}_{\alpha,\epsilon^{(1)}_{\alpha,\delta_1}} = \frac{\alpha-1}{2\alpha-1}\big(R_1 - \delta_1 - d_\alpha(p_{U,W}, p_U p_W)\big), \quad (96a)$$

$$\beta^{(1)}_{\alpha,\delta_2} \triangleq \beta^{(2)}_{\alpha,\epsilon^{(2)}_{\alpha,\delta_2}} = \frac{\alpha-1}{2\alpha-1}\big(R_1 + R_2 - \delta_2 - d_\alpha(p_{U,V,W}, p_{U,V} p_W)\big). \quad (96b)$$

Observe that $\epsilon^{(1)}_{\alpha,\delta_1}$ and $\epsilon^{(2)}_{\alpha,\delta_2}$ in (95) are nonnegative. For example, $\epsilon^{(1)}_{\alpha,\delta_1} \geq 0$ by the $R_1 - \delta_1 > I(U;W)$ assumption, because $\alpha > 1$ and $d_\alpha(p_{U,W}, p_W p_V) \geq d_1(p_{W,V}, p_U p_W) = I(U;W)$.

The properties of Rényi divergence imply the existence of $\alpha > 1$, for which (96a) and (96b) are positive.

Lemma 7 (Strictly Positive Exponents) *There exists an $\alpha > 1$ such that $\beta^{(j)}_{\alpha,\delta_j} > 0$, for $j = 1, 2$.*

Lemma 7 is proven in Appendix C and shows that the RHS of (93) exponentially decays with n. To bound the probability (with respect to a random superposition codebook) of (89) not producing this exponential decay, we use a Chernoff bound.

Lemma 8 (Chernoff Bound) *Let $\{X_m\}_{m=1}^{M}$ be a collection of i.i.d. random variables with $X_m \in [0, B]$ and $\mathbb{E}X_m \leq \mu \neq 0$ for all $m \in [1:M]$. Then, for any c with $\frac{c}{\mu} \geq 1$*

$$\mathbb{P}\left(\frac{1}{M}\sum_{m=1}^{M} X_m \geq c\right) \leq e^{-\frac{M\mu}{B}\left(\frac{c}{\mu}\left(\ln\frac{c}{\mu} - 1\right) + 1\right)}. \quad (97a)$$

Furthermore, if $\frac{c}{\mu} \in [1, 2]$, then

$$\mathbb{P}\left(\frac{1}{M}\sum_{m=1}^{M} X_m \geq c\right) \leq e^{-\frac{M\mu}{3B}\left(\frac{c}{\mu} - 1\right)^2}. \quad (97b)$$

These bounds are proven in [27, Appendix C]. Having Lemma 8, we show that $\int dP_{\mathsf{B}_n,2}$ is exponentially small with a probability doubly-exponentially close to 1. To do so we exploit the fact that for any $j \in \mathcal{J}_n$, the structure of the superposition code implies that $\big\{\big(\mathbf{U}(i), \mathbf{V}(i,j)\big)\big\}_{i \in \mathcal{I}_n}$ comprises i.i.d. pairs of random variables. Consequently, denoting

$$f(\mathbf{u},\mathbf{v}) \triangleq \mathbb{P}_{p^n_{W|U,V}}\left((\mathbf{u},\mathbf{v},\mathbf{W}) \notin \mathcal{A}_{\epsilon^{(1)}_{\alpha,\delta_1}, \epsilon^{(2)}_{\alpha,\delta_2}} \middle| \mathbf{U} = \mathbf{u}, \mathbf{V} = \mathbf{v}\right), \quad (98)$$

we have that $\big\{f\big(\mathbf{U}(i), \mathbf{V}(i,j)\big)\big\}_{i \in \mathcal{I}_n}$ are i.i.d. for any $j \in \mathcal{J}_n$, and that

$$\mathbb{E}_\mu f\big(\mathbf{U}(i), \mathbf{V}(i,j)\big) \leq 2^{-n\beta^{(1)}_{\alpha,\delta_1}} + 2^{-n\beta^{(2)}_{\alpha,\delta_2}}, \ \forall (i,j) \in \mathcal{I}_n \times \mathcal{J}_n. \quad (99)$$

For any $c \in \mathbb{R}_+$ consider now the following:

$$\begin{aligned}
&\mathbb{P}_\mu\left(\int dP_{\mathsf{B}_n,2} \geq c\right) \\
&= \mathbb{P}_\mu\left(2^{-n(R_1+R_2)} \sum_{(i,j) \in \mathcal{I}_n \times \mathcal{J}_n} f\big(\mathbf{U}(i), \mathbf{V}(i,j)\big) \geq c\right) \\
&\leq \mathbb{P}_\mu\left(\bigcup_{j \in \mathcal{J}_n}\left\{2^{-n(R_1+R_2)} \sum_{i \in \mathcal{I}_n} f\big(\mathbf{U}(i), \mathbf{V}(i,j)\big) \geq c 2^{-nR_2}\right\}\right) \\
&\leq \sum_{j \in \mathcal{J}_n} \mathbb{P}_\mu\left(2^{-nR_1} \sum_{i \in \mathcal{I}_n} f\big(\mathbf{U}(i), \mathbf{V}(i,j)\big) \geq c\right). \quad (100)
\end{aligned}$$

Using (97b) on each summand from the RHS of (100) with $M = 2^{nR_1}$, $\mu = 2^{-n\beta^{(1)}_{\alpha,\delta_1}} + 2^{-n\beta^{(2)}_{\alpha,\delta_2}}$, $B = 1$, and $\frac{c}{\mu} = 2$:

$$\begin{aligned}
&\mathbb{P}_\mu\Big(2^{-nR_1} \sum_{i \in \mathcal{I}_n} f\big(\mathbf{U}(i), \mathbf{V}(i,j)\big) \geq 2\left(2^{-n\beta^{(1)}_{\alpha,\delta_1}} + 2^{-n\beta^{(2)}_{\alpha,\delta_2}}\right)\Big) \\
&\qquad \leq e^{-\frac{1}{3} 2^{nR_1}\left(2^{-n\beta^{(1)}_{\alpha,\delta_1}} + 2^{-n\beta^{(2)}_{\alpha,\delta_2}}\right)} \\
&\qquad \leq e^{-\frac{1}{3} 2^{n\left(R_1 - \beta^{(1)}_{\alpha,\delta_1}\right)}}. \quad (101)
\end{aligned}$$

Inserting (101) into (100), we have

$$\begin{aligned}
&\mathbb{P}_\mu\Big(\int dP_{\mathsf{B}_n,2} \geq 2\left(2^{-n\beta^{(1)}_{\alpha,\delta_1}} + 2^{-n\beta^{(2)}_{\alpha,\delta_2}}\right)\Big) \\
&\qquad \leq 2^{nR_2} \cdot e^{-\frac{1}{3} 2^{n\left(R_1 - \beta^{(1)}_{\alpha,\delta_1}\right)}}, \quad (102)
\end{aligned}$$

for which $\alpha > 1$ can be chosen to produce a double-exponential convergence to 0 of the RHS because for any $\alpha > 1$, we have

$$R_1 - \beta^{(1)}_{\alpha,\delta_1} = \frac{\alpha R_1 + (\alpha-1)\big(\delta_1 + d_\alpha(p_{U,W}, p_U p_W)\big)}{2\alpha - 1} > 0. \quad (103)$$

We next treat $\Delta_{\mathsf{B}_n,1}(\mathbf{w})$, for $\mathbf{w} \in \mathcal{W}^n$, and show that it also decays exponentially fast with a probability doubly-exponentially close to 1. To simplify notation, for each $\mathbf{w} \in \mathcal{W}^n$, let $g_{\mathbf{w}} : \mathcal{U}^n \times \mathcal{V}^n \to \mathbb{R}_+$ be

$$g_{\mathbf{w}}(\mathbf{u},\mathbf{v}) = \frac{dp_{W|U=\mathbf{u},V=\mathbf{v}}}{dp^n_W}(\mathbf{w}) \mathbb{1}_{\left\{\left(\mathbf{u},\mathbf{v},\mathbf{w}\right) \in \mathcal{A}_{\epsilon^{(1)}_{\alpha,\delta_1}, \epsilon^{(2)}_{\alpha,\delta_2}}\right\}}. \quad (104)$$

Accordingly, note that

$$\Delta_{\mathsf{B}_n,1}(\mathbf{w}) = 2^{-n(R_1+R_2)} \sum_{(i,j) \in \mathcal{I}_n \times \mathcal{J}_n} g_{\mathbf{w}}\big(\mathbf{U}(i), \mathbf{V}(i,j)\big)$$

$$
\begin{aligned}
&\mathbb{P}_\mu\Big(\Delta_{\mathsf{B}_n,1}(\mathbf{w}) \geq c\Big)\\
&= \mathbb{P}_\mu\left(2^{-n(R_1+R_2)} \sum_{(i,j)\in\mathcal{I}_n\times\mathcal{J}_n} g_{\mathbf{w}}\big(\mathbf{U}(i), \mathbf{V}(i,j)\big) \geq c\right)\\
&\leq \mathbb{P}_\mu\Big(\mathcal{D}(c')\Big) + \mathbb{P}_\mu\left(2^{-n(R_1+R_2)} \sum_{(i,j)\in\mathcal{I}_n\times\mathcal{J}_n} g_{\mathbf{w}}\big(\mathbf{U}(i), \mathbf{V}(i,j)\big) \geq c \,\middle|\, \mathcal{D}(c')^c\right)\\
&\leq \sum_{i\in\mathcal{I}_n} \mathbb{P}_\mu\left(2^{-nR_2} \sum_{j\in\mathcal{J}_n} g_{\mathbf{w}}\big(\mathbf{U}(i), \mathbf{V}(i,j)\big) \geq c' \cdot 2^{n\left(I(U;W)+\epsilon^{(1)}_{\alpha,\delta_1}\right)}\right)\\
&\qquad\qquad + \mathbb{P}_\mu\left(2^{-n(R_1+R_2)} \sum_{(i,j)\in\mathcal{I}_n\times\mathcal{J}_n} g_{\mathbf{w}}\big(\mathbf{U}(i), \mathbf{V}(i,j)\big) \geq c \,\middle|\, \mathcal{D}(c')^c\right)\\
&\leq \sum_{i\in\mathcal{I}_n} \int_{\mathbf{u}\in\mathcal{U}^n} d\,\mathbb{P}_\mu\Big(\mathbf{U}(i)=\mathbf{u}\Big) \underbrace{\mathbb{P}_\mu\left(2^{-nR_2} \sum_{j\in\mathcal{J}_n} g_{\mathbf{w}}\big(\mathbf{U}(i), \mathbf{V}(i,j)\big) \geq c' \cdot 2^{n\left(I(U;W)+\epsilon^{(1)}_{\alpha,\delta_1}\right)} \,\middle|\, \mathbf{U}(i)=\mathbf{u}\right)}_{P_1(i,\mathbf{u})}\\
&\qquad\qquad\qquad\qquad + \underbrace{\mathbb{P}_\mu\left(2^{-nR_1} \sum_{i\in\mathcal{I}_n}\left[2^{-nR_2} \sum_{j\in\mathcal{J}_n} g_{\mathbf{w}}\big(\mathbf{U}(i), \mathbf{V}(i,j)\big)\right] \geq c \,\middle|\, \forall i\in\mathcal{I},\ \mathcal{D}_i(c')^c\right)}_{P_2}. \qquad (107)
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}_\mu\Big[g_{\mathbf{w}}\big(\mathbf{U}(i,\mathsf{B}_U), \mathbf{V}(i,j)\big)\Big|\mathbf{U}(i)=\mathbf{u}\Big] &= \mathbb{E}_\mu\left[\frac{dp_{W|U=\mathbf{u},V=\mathbf{V}(i,j)}}{dp^n_W}(\mathbf{w})\mathbb{1}_{\left\{\left(\mathbf{u},\mathbf{V}(i,j),\mathbf{w}\right)\in\mathcal{A}_{\epsilon^{(1)}_{\alpha,\delta_1},\epsilon^{(2)}_{\alpha,\delta_2}}\right\}}\,\middle|\,\mathbf{U}(i)=\mathbf{u}\right]\\
&\leq \mathbb{1}_{\left\{\frac{dp^n_{W|U=\mathbf{u}}}{dp^n_W}(\mathbf{w})\leq 2^{n\left(I(U;W)+\epsilon^{(1)}_{\alpha,\delta_1}\right)}\right\}} \frac{dp_{W|U=\mathbf{u}}}{dp^n_W}(\mathbf{w})\\
&\leq 2^{n\left(I(U;W)+\epsilon^{(1)}_{\alpha,\delta_1}\right)}. \qquad (108)
\end{aligned}
$$

$$
= 2^{-nR_1} \sum_{i\in\mathcal{I}_n}\left[2^{-nR_2} \sum_{j\in\mathcal{J}_n} g_{\mathbf{w}}\big(\mathbf{U}(i), \mathbf{V}(i,j)\big)\right], \qquad (105)
$$

where the RHS is an average of 2^{nR_1} i.i.d. random variables due to the structure of the superposition codebook. Next, for any $c' \in \mathbb{R}_+$ and $i \in \mathcal{I}_n$ define

$$
\mathcal{D}_i(c') = \left\{2^{-nR_2}\sum_{j\in\mathcal{J}_n} g_{\mathbf{w}}\big(\mathbf{U}(i),\mathbf{V}(i,j)\big) \geq c' 2^{n\left(I(U;W)+\epsilon^{(1)}_{\alpha,\delta_1}\right)}\right\} \qquad (106a)
$$

and set

$$
\mathcal{D}(c') = \bigcup_{i\in\mathcal{I}_n} \mathcal{D}_i(c'). \qquad (106b)
$$

Consider the upper bound in (107) at the top of the page. on the probability that $\Delta_{\mathsf{B}_n,1}(\mathbf{w})$ is lower bounded by a constant $c \in \mathbb{R}_+$.To invoke the Chernoff bound from (97a) on $P_1(i,\mathbf{u})$, where $i \in \mathcal{I}_n$ and $\mathbf{u} \in \mathcal{U}^n$, first note that conditioned on $\mathbf{U}(i) = \mathbf{u}$, $\left\{g_{\mathbf{w}}\big(\mathbf{U}(i), \mathbf{V}(i,j)\big)\right\}_{j\in\mathcal{J}_n}$ are i.i.d. Furthermore, each $g_{\mathbf{w}}\big(\mathbf{U}(i), \mathbf{V}(i,j)\big)$ is upper bounded by $2^{n\left(I(U,V;W)+\epsilon^{(2)}_{\alpha,\delta_2}\right)}$ with probability 1, and has expected value bounded as in (108). Using (97a) with $M = 2^{nR_2}$, $\mu = 2^{n\left(I(U;W)+\epsilon^{(1)}_{\alpha,\delta_1}\right)}$, $B = 2^{n\left(I(U,V;W)+\epsilon^{(2)}_{\alpha,\delta_2}\right)}$, and $c = c' \cdot \mu$, for any $c' \geq \frac{1}{\mu}$, gives

$$
P_1(i,\mathbf{u}) \leq e^{-2^{n\left(R_2 - I(V;W|U)+\epsilon^{(1)}_{\alpha,\delta_1}-\epsilon^{(2)}_{\alpha,\delta_2}\right)}\left(c'(\ln c'-1)+1\right)}, \qquad (109)
$$

for all $(i,\mathbf{u}) \in \mathcal{I}_n \times \mathcal{U}^n$.

For P_2 we have that $\left\{2^{-nR_2}\sum_{j\in\mathcal{J}_n} g_{\mathbf{w}}\big(\mathbf{U}(i), \mathbf{V}(i,j)\big)\right\}_{i\in\mathcal{I}_n}$ are i.i.d. by the codebook construction. The conditioning on $\mathcal{D}(c')^c$ implies that each random variable $2^{-nR_2}\sum_{j\in\mathcal{J}_n} g_{\mathbf{w}}\big(\mathbf{U}(i), \mathbf{V}(i,j)\big)$, for $i \in \mathcal{I}_n$, is almost surely bounded between 0 and $c' \cdot 2^{n\left(I(U;W)+\epsilon^{(1)}_{\alpha,\delta_1}\right)}$. The expectation of each term with respect to the codebook is bounded above by one, which follows by removing the indicator function from $g_{\mathbf{w}}\big(\mathbf{U}(i), \mathbf{V}(i,j)\big)$. Setting

$M = 2^{nR_1}$, $\mu = 1$, $B = 2^{n\left(I(U;W)+\epsilon^{(1)}_{\alpha,\delta_1}\right)}$, and any $c \in [1,2]$ into (97b), gives

$$P_2 \leq e^{-\frac{1}{3}2^{n\left(R_1-I(U;W)-\epsilon^{(1)}_{\alpha,\delta_1}\right)}(c-1)^2}. \tag{110}$$

Inserting (109) and (110) into (107), we have that for any $\mathbf{w} \in \mathcal{W}^n$, $c \in [1,2]$ and $c' \geq 2^{-n\left(I(U;W)+\epsilon^{(1)}_{\alpha,\delta_1}\right)}$

$$\begin{aligned}
&\mathbb{P}_\mu\Big(\Delta_{\mathsf{B}_n,1}(\mathbf{w}) \geq c\Big)\\
&\leq 2^{nR_1}e^{-2^{n\left(R_2-I(V;W|U)+\epsilon^{(1)}_{\alpha,\delta_1}-\epsilon^{(2)}_{\alpha,\delta_2}\right)}\left(c'(\ln c'-1)+1\right)}\\
&\qquad\qquad + e^{-\frac{1}{3}2^{n\left(R_1-I(U;W)-\epsilon^{(1)}_{\alpha,\delta_1}\right)}\frac{(c-1)^2}{c'}}. \qquad (111)
\end{aligned}$$

Our next step is to choose c and c' to get the double-exponential decay on the RHS of (111). Let

$$c' = 2^{n\left(I(V;W|U)-R_2-\epsilon^{(1)}_{\alpha,\delta_1}+\epsilon^{(2)}_{\alpha,\delta_2}+2\beta^{(2)}_{\alpha,\delta_2}+\frac{\delta_2}{2}\right)} - 1, \tag{112}$$

and note that the exponent is strictly positive since

$$\begin{aligned}
&I(V;W|U) - R_2 - \epsilon^{(1)}_{\alpha,\delta_1} + \epsilon^{(2)}_{\alpha,\delta_2} + 2\beta^{(2)}_{\alpha,\delta_2} + \frac{\delta_2}{2}\\
&\overset{(a)}{=} R_1 - I(U;W) - \frac{\delta_2}{2} - \epsilon^{(1)}_{\alpha,\delta_1}\\
&= \frac{2(\alpha-1)\Big[R_1 - d_\alpha(p_{U,W}, p_U p_W) - \delta_1\Big] + \frac{2\alpha-1}{2}(2\delta_1-\delta_2)}{2\alpha-1}\\
&> 0 \qquad (113)
\end{aligned}$$

where (a) is because $\epsilon^{(2)}_{\alpha,\delta_2} + 2\beta^{(2)}_{\alpha,\delta_2} = R_1 + R_2 - I(U,V;W) - \delta_2$ and the positivity is by choosing α as in Lemma 7 and since $\delta_2 < 2\delta_1$. Consequently, $c' \to \infty$ as $n \to \infty$, and, therefore, $c' \geq 2^{-n\left(I(U;W)+\epsilon^{(1)}_{\alpha,\delta_1}\right)}$ for sufficiently large n. Since c' is unbounded (as a function of n), for n large enough we also have $\ln c' - 1 \geq 1$, which simplifies the RHS of (111) as

$$\begin{aligned}
&2^{nR_1}e^{-2^{n\left(R_2-I(V;W|U)+\epsilon^{(1)}_{\alpha,\delta_1}-\epsilon^{(2)}_{\alpha,\delta_2}\right)}\left(c'(\ln c'-1)+1\right)}\\
&\qquad \leq 2^{nR_1}e^{-2^{n\left(R_2-I(V;W|U)+\epsilon^{(1)}_{\alpha,\delta_1}-\epsilon^{(2)}_{\alpha,\delta_2}\right)}(c'+1)}\\
&\qquad = 2^{nR_1}e^{-2^{n\left(2\beta^{(2)}_{\alpha,\delta_2}+\frac{\delta_2}{2}\right)}}, \qquad (114)
\end{aligned}$$

which decays doubly-exponentially quickly to 0.

Setting $c = 1 + 2^{-n\frac{\delta_1}{4}}$, we upper bound the second term from the RHS of (111) by

$$\begin{aligned}
e^{-\frac{1}{3}2^{n\left(R_1-I(U;W)-\epsilon^{(1)}_{\alpha,\delta_1}\right)}\frac{(c-1)^2}{c'}} &\leq e^{-\frac{1}{3}2^{n\left(R_1-I(U;W)-\epsilon^{(1)}_{\alpha,\delta_1}\right)}\frac{(c-1)^2}{(c'+1)}}\\
&= e^{-\frac{1}{3}2^{n\frac{\delta_2-\delta_1}{2}}}, \qquad (115)
\end{aligned}$$

which also converges to 0 with double-exponential speed because $\delta_1 < \delta_2$.

Concluding, (111), (114) and (115) upper bound the probability of interest as

$$\begin{aligned}
&\mathbb{P}_\mu\Big(\Delta_{\mathsf{B}_n,1}(\mathbf{w}) \geq 1 + 2^{-n\frac{\delta_1}{4}}\Big)\\
&\qquad \leq 2^{nR_1}e^{-2^{n\left(2\beta^{(2)}_{\alpha,\delta_2}+\frac{\delta_2}{2}\right)}} + e^{-\frac{1}{3}2^{n\frac{\delta_2-\delta_1}{2}}}. \qquad (116)
\end{aligned}$$

At this point, we specialize to a finite $\mathcal{W}$. Consequently, $\Delta_{\mathsf{B}_n,2}$ is bounded as

$$\Delta_{\mathsf{B}_n,2}(\mathbf{w}) \leq \left(\max_{w\in\text{supp}(p_W)} \frac{1}{p_W(w)}\right)^n, \ \forall\, \mathbf{w} \in \mathcal{W}^n, \tag{117}$$

almost surely. Notice that the maximum is only over the support of p_W, which makes this bound finite. The underlying reason for this restriction is that with probability 1 a conditional distribution is absolutely continuous with respect to any of its marginals.

Having (102), (116) and (117), we can now bound the probability that the RHS of (87) is not exponentially small. Let $\mathcal{S}$ be the set of superposition codebooks $\mathcal{B}_n \in \mathfrak{B}_n$, such that all of the following are true:

$$\int dP_{\mathcal{B}_n,2} < 2 \cdot \left(2^{-n\beta^{(1)}_{\alpha,\delta_1}} + 2^{-n\beta^{(2)}_{\alpha,\delta_2}}\right), \tag{118a}$$

$$\Delta_{\mathcal{B}_n,1}(\mathbf{w}) < 1 + 2^{-n\frac{\delta_1}{4}}, \quad \forall \mathbf{w} \in \mathcal{W}^n, \tag{118b}$$

$$\Delta_{\mathcal{B}_n,2}(\mathbf{w}) \leq \left(\max_{w\in\text{supp}(p_W)} \frac{1}{p_W(w)}\right)^n, \quad \forall \mathbf{w} \in \mathcal{W}^n. \tag{118c}$$

First, we use the union bound and the fact that $\mathcal{W}^n$ is only exponentially large, to show that the probability of a random codebook not being in $\mathcal{S}$ is double-exponentially small:

$$\begin{aligned}
&\mathbb{P}_\mu(\mathsf{B}_n \notin \mathcal{S})\\
&\overset{(a)}{\leq} \mathbb{P}_\mu\left(\int dP_{\mathsf{B}_n,2} \geq 2 \cdot 2^{-n\beta_{\alpha,\delta}}\right)\\
&\quad + \sum_{\mathbf{w}\in\mathcal{W}^n} \mathbb{P}_\mu\left(\Delta_{\mathsf{B}_n,1}(\mathbf{w}) \geq 1 + 2^{-\beta_{\alpha,\delta}n}\right)\\
&\quad + \sum_{\mathbf{w}\in\mathcal{W}^n} \mathbb{P}_\mu\left(\Delta_{\mathsf{B}_n,2}(\mathbf{w}) > \left(\max_{w\in\text{supp}(p_W)} \frac{1}{p_W(w)}\right)^n\right)\\
&\overset{(b)}{\leq} 2^{nR_2} \cdot e^{-\frac{1}{3}2^{n\left(R_1-\beta^{(1)}_{\alpha,\delta_1}\right)}}\\
&\quad + |\mathcal{W}|^n\left[2^{nR_1}e^{-2^{n\left(2\beta^{(2)}_{\alpha,\delta_2}+\frac{\delta_2}{2}\right)}} + e^{-\frac{1}{3}2^{n\frac{\delta_2-\delta_1}{2}}}\right], \qquad (119)
\end{aligned}$$

where (a) is the union bound, and (b) uses (102), (116) and (117).

Next, we claim that for every codebook in $\mathcal{S}$, the RHS of (87) is exponentially small. Let $\mathcal{B}_n \in \mathcal{S}$ and consider the following. For every $x \in [0,1]$, $h(x) \leq x \log \frac{e}{x}$, which, together with (118a), implies that

$$\begin{aligned}
&h\Big(\int dP_{\mathcal{B}_n,1}\Big)\\
&\quad = h\left(\int dP_{\mathcal{B}_n,2}\right)\\
&\quad < 2\left[\log e - \log 2 \cdot \log\left(2^{-n\beta^{(1)}_{\alpha,\delta_1}} + 2^{-n\beta^{(2)}_{\alpha,\delta_2}}\right)\right]\\
&\qquad\qquad \times \left[2^{-n\beta^{(1)}_{\alpha,\delta_1}} + 2^{-n\beta^{(2)}_{\alpha,\delta_2}}\right]\\
&\quad \overset{(a)}{\leq} 4\big(\log e + 2\beta_{\alpha,\delta_1,\delta_2}\log 2\big)n2^{-n\beta_{\alpha,\delta_1,\delta_2}}, \qquad (120)
\end{aligned}$$

where (a) follows by setting $\beta_{\alpha,\delta_1,\delta_2} \triangleq \min\left\{\beta^{(1)}_{\alpha,\delta_1}, \beta^{(2)}_{\alpha,\delta_2}\right\}$. Furthermore, by (118b), we have

$$\int dP_{\mathcal{B}_n,1} \log \Delta_{\mathcal{B}_n,1} < \int dP_{\mathcal{B}_n,1} \log\left(1 + 2^{-n\frac{\delta_1}{4}}\right)$$

$$= \log\left(1+2^{-n\frac{\delta_1}{4}}\right) \overset{(a)}{\le} 2^{-n\frac{\delta_1}{4}} \log e, \tag{121}$$

where (a) is since $\log(1+x) \le x\log e$, for every $x > 0$. Finally, using (118c) and the definition of $\beta_{\alpha,\delta_1,\delta_2}$, we obtain

$$\begin{aligned}
&\int dP_{\mathcal{B}_n,2} \log \Delta_{\mathcal{B}_n,2} \\
&\qquad \le \int dP_{\mathcal{B}_n,2} \log \left(\max_{w\in\mathrm{supp}(p_W)} \frac{1}{p_W(w)} \right)^n \\
&\qquad < 2\log\left(\max_{w\in\mathrm{supp}(p_W)} \frac{1}{p_W(w)} \right) n 2^{-n\beta_{\alpha,\delta_1,\delta_2}}.
\end{aligned} \tag{122}$$

Combining (120)-(122), while setting $\gamma_{\alpha,\delta_1,\delta_2} \triangleq \min\left\{\beta_{\alpha,\delta_1,\delta_2}, \frac{\delta_1}{4}\right\}$, yields

$$\begin{aligned}
&h\left(\int dP_{\mathcal{B}_n,1}\right) + \int dP_{\mathcal{B}_n,1} \log \Delta_{\mathcal{B}_n,1} + \int dP_{\mathcal{B}_n,2} \log \Delta_{\mathcal{B}_n,2} \\
&< \Bigg(4\big(\log e + 2\beta_{\alpha,\delta_1,\delta_2} \log 2\big) + \log e \\
&\qquad + 2\log\left(\max_{w\in\mathrm{supp}(p_W)} \frac{1}{p_W(w)} \right) \Bigg) n 2^{-n\gamma_{\alpha,\delta_1,\delta_2}} \\
&\overset{(a)}{=} c_{\alpha,\delta_1,\delta_2} n 2^{-n\gamma_{\alpha,\delta_1,\delta_2}},
\end{aligned} \tag{123}$$

where (a) comes from setting

$$\begin{aligned}
c_{\alpha,\delta_1,\delta_2} &\triangleq 4\big(\log e + 2\beta_{\alpha,\delta_1,\delta_2}\log 2\big) + \log e \\
&\quad + 2\log\left(\max_{w\in\mathrm{supp}(p_W)} \frac{1}{p_W(w)} \right).
\end{aligned} \tag{124}$$

This implies that

$$\begin{aligned}
&\mathbb{P}_\mu\left(\mathsf{D}\left(P_{\mathbf{W}}^{(\mathsf{B}_n)} \middle\| p_W^n \right) \ge c_{\alpha,\delta_1,\delta_2} n 2^{-n\gamma_{\alpha,\delta_1,\delta_2}} \right) \\
&\le \mathbb{P}_\mu\Bigg(h\left(\int dP_{\mathsf{B}_n,1}\right) + \int dP_{\mathsf{B}_n,1} \log \Delta_{\mathsf{B}_n,1} \\
&\qquad + \int dP_{\mathsf{B}_n,2} \log \Delta_{\mathsf{B}_n,2} \ge c_{\alpha,\delta} n 2^{-n\beta_{\alpha,\delta}} \Bigg) \\
&\le \mathbb{P}_\mu\big(\mathsf{B}_n \notin \mathcal{S}\big) \\
&\overset{(a)}{\le} 2^{nR_2} \cdot e^{-\frac{1}{3}2^{n\left(R_1-\beta^{(1)}_{\alpha,\delta_1}\right)}} \\
&\qquad + |\mathcal{W}|^n \left[2^{nR_1} e^{-2^{n\left(2\beta^{(2)}_{\alpha,\delta_2}+\frac{\delta_2}{2}\right)}} + e^{-\frac{1}{3}2^{n\frac{\delta_2-\delta_1}{2}}} \right] \\
&\overset{(b)}{\le} 2^{nR_2} \cdot e^{-\frac{1}{3}2^{n\delta_1}} + |\mathcal{W}|^n \left[2^{nR_1} e^{-2^{n\frac{\delta_2}{2}}} + e^{-\frac{1}{3}2^{n\frac{\delta_2-\delta_1}{2}}} \right],
\end{aligned} \tag{125}$$

where (a) follows from (119), while (b) is because $\beta^{(1)}_{\alpha,\delta_1} \le \frac{1}{2}(R_1-\delta_1)$ and $\beta^{(2)}_{\alpha,\delta_2} \ge 0$. Denoting $c_{\delta_1,\delta_2} \triangleq \sup_{\alpha>1} c_{\alpha,\delta_1,\delta_2}$, (125) further gives

$$\begin{aligned}
&\mathbb{P}_\mu\left(\mathsf{D}\left(P_{\mathbf{W}}^{(\mathsf{B}_n)} \middle\| p_W^n \right) \ge c_{\delta_1,\delta_2} n 2^{-n\gamma_{\alpha,\delta_1,\delta_2}} \right) \\
&\le 2^{nR_2} \cdot e^{-\frac{1}{3}2^{n\delta_1}} + |\mathcal{W}|^n \left[2^{nR_1} e^{-2^{n\frac{\delta_2}{2}}} + e^{-\frac{1}{3}2^{n\frac{\delta_2-\delta_1}{2}}} \right].
\end{aligned} \tag{126}$$

Since (126) holds for all $\alpha > 1$ (the interesting values of α are those from Lemma 7, but the derivation is valid for all $\alpha > 1$), it must also be true, with strict inequality in the LHS, when replacing $\gamma_{\alpha,\delta_1,\delta_2}$ with $\gamma_{\delta_1,\delta_2} \triangleq \sup_{\alpha>1} \gamma_{\alpha,\delta_1,\delta_2}$, which is the exponential rate of convergence we derive for the strong SCL for superposition codes.

Concluding, if $R_1 > I(U;W)$, $R_1 + R_2 > I(U,V;W)$, then for any $\delta_1 \in \big(0, R_1 - I(U;W)\big)$ and $\delta_2 \in \big(0, R_1 + R_2 - I(U,V;W)\big)$ with $\delta_1 < \delta_2 < 2\delta_1$ we get exponential convergence of the relative entropy at rate $O(2^{-n\gamma_{\delta_1,\delta_2}})$ with doubly-exponential certainty. Discarding the precise exponents of convergence and coefficients, we state that there exist $\gamma_1, \gamma_2 > 0$, such that, for n large enough,

$$\mathbb{P}_\mu\left(\mathsf{D}\left(P_{\mathbf{W}}^{(\mathsf{B}_n)} \middle\| p_W^n \right) > e^{-n\gamma_1} \right) \le e^{-e^{n\gamma_2}}. \tag{127}$$

Appendix C
Proof of Lemma 7

The proof uses basic properties of Rényi divergence (see, e.g., [30]). First, recall that for fixed measures μ and ν, $d_\alpha(\mu,\nu)$ is monotone non-decreasing in α. Furthermore, if $\mu \ll \nu$ then $d_\alpha(\mu,\nu)$ is continuous in $\alpha \in (1,\infty]$. Since a joint distribution is always absolutely continuous with respect to the product of its marginals and by the choices of δ_1 and δ_2, there exist $\alpha_1, \alpha_2 > 1$ such that

$$\begin{aligned}
R_1 - \delta_1 &> d_{\alpha_1}(Q_{U,W}, Q_U, Q_W) \\
&\ge d_1(Q_{U,W}, Q_U, Q_W) \\
&= I(U;W), \qquad (128a) \\
R_1 + R_2 - \delta_2 &> d_{\alpha_2}(Q_{U,V,W}, Q_{U,V}, Q_W) \\
&\ge d_1(Q_{U,V,W}, Q_{U,V}, Q_W) \\
&= I(U,V;W). \qquad (128b)
\end{aligned}$$

Setting $\alpha = \min\{\alpha_1, \alpha_2\}$, we have $\beta^{(j)}_{\alpha,\delta_j} > 0$, for $j = 1, 2$.

Appendix D
Proof of Proposition 1

To prove $R_{\mathsf{A}} \le R_{\mathsf{A}}^{\mathsf{Alt}}$, note that the two first rate bounds in $R_{\mathsf{A}}^{\mathsf{Alt}}$ (see (14)) are the same as those defining R_{A}, while the third bound in $R_{\mathsf{A}}^{\mathsf{Alt}}$ is obtained by adding the first bound from R_{A} with $I(U;Y) - I(U;S)$, which we know is non-negative by (19).

For the opposite direction consider the following. Let $p^\star_{U,V,X|S} : \mathcal{S} \to \mathcal{P}(\mathcal{U} \times \mathcal{V} \times \mathcal{X})$ an optimizer of $R_{\mathsf{A}}^{\mathsf{Alt}}$ such that $R_{\mathsf{A}}^{\mathsf{Alt}} = R_{\mathsf{A}}^{\mathsf{Alt}}(p^\star_{U,V,X|S}) > 0$ (otherwise there is nothing to prove). Recall that the mutual information terms in $R_{\mathsf{A}}^{\mathsf{Alt}}(p^\star_{U,V,X|S})$ are taken with respect to $p^\star \triangleq p_S p^\star_{U,V,X|S} p_{Y,Z|X,S}$. First, note that if $p^\star_{U,V,X|S}$ is such that $I(U;Y) - I(U;S) \ge 0$, then $R_{\mathsf{A}}^{\mathsf{Alt}} \le R_{\mathsf{A}}(p^\star_{U,V,X|S}) = R_{\mathsf{A}}$ and the desired inequality holds.

Otherwise, if $p^\star_{U,V,X|S}$ has $I(U;Y) - I(U;S) < 0$, let $U' = (U, \tilde{V})$ and $V' = V$, where $\tilde{V}$ is V passed through an erasure channel, with erasures independent of all the other random variables. Denoting the probability of an erasure by $\epsilon \in [0,1]$, the joint distribution of $(S,U,V,X,Y,Z,\tilde{V},U',V')$ is

$$\begin{aligned} &p_{S,U,V,X,Y,Z,\tilde{V},U',V'} \\ &\qquad = p_S p^{\star}_{U,V,X|S} p_{Y,Z|X,S} p_{\tilde{V}|V} \mathbb{1}_{\left\{U'=(U,\tilde{V}),V'=V\right\}}, \end{aligned} \tag{129}$$

where $p_{\tilde{V}|V} : \mathcal{V} \to \mathcal{V} \cup \{?\}$, with $? \notin \mathcal{V}$, is the transition probability of a $\mathsf{BEC}(\epsilon)$. The value of ϵ will be specified later. All subsequent information measures in this proof are taken with respect to (129) or its marginals.

We first show that $\epsilon \in [0,1]$ can be chosen such that $p_{U',V',X|S}$ is a valid input distribution in R_{A}, i.e., satisfying

$$I(U';Y) - I(U',S) \geq 0. \tag{130}$$

Consider

$$\begin{aligned} &I(U';Y) - I(U';S) \\ &\qquad = I(U;Y) - I(U;S) + I(\tilde{V};Y|U) - I(\tilde{V};S|U) \\ &\qquad = I(U;Y) - I(U;S) + \bar{\epsilon}\Big[I(V;Y|U) - I(V;S|U)\Big], \end{aligned} \tag{131}$$

where $\bar{\epsilon} = 1-\epsilon$. Notice that when $\epsilon = 1$ this quantity is negative by assumption, while $\epsilon = 0$ gives

$$I(U';Y) - I(U';S) = I(U,V;Y) - I(U,V;S) > 0 \tag{132}$$

by the second rate bound in $R_{\mathsf{A}}^{\mathsf{Alt}}$. We set $\epsilon \in [0,1]$ such that $I(U';Y) - I(U';S) = 0$, thus satisfying (130).

We next evaluate $R_{\mathsf{A}}(p_{U',V',X|S})$. Starting from the second rate bound, we have

$$I(U',V';Y) - I(U',V';S) \overset{(a)}{=} I(U,V;Y) - I(U,V;S) \geq R_{\mathsf{A}}^{\mathsf{Alt}}, \tag{133}$$

where (a) uses the Markov chain $(S,U,X,Y,Z) - V - \tilde{V}$, which follows because $\tilde{V}$ is a noisy version of V.

For the first rate bound, note that

$$\begin{aligned} &I(V';Y|U') - I(V';Z|U') \\ &= I(V;Y|U,\tilde{V}) - I(V;Z|U,\tilde{V}) \\ &\overset{(a)}{=} I(V;Y|U) - I(V;Z|U) - \Big[I(\tilde{V};Y|U) - I(\tilde{V};Z|U)\Big] \\ &\overset{(b)}{=} I(V;Y|U) - I(V;Z|U) - \bar{\epsilon}\Big[I(V;Y|U) - I(V;Z|U)\Big] \\ &= \epsilon\Big[I(V;Y|U) - I(V;Z|U)\Big], \end{aligned} \tag{134}$$

where (a) and (b) follow by Markovity. A similar derivation also gives

$$I(V';Y|U') - I(V';S|U') = \epsilon\Big[I(V;Y|U) - I(V;S|U)\Big]. \tag{135}$$

We complete the proof by considering two cases. First, if $I(V;S|U) \geq I(V;Z|U)$, we obtain

$$\begin{aligned} I(V';Y|U') - I(V';Z|U') &\overset{(a)}{=} \epsilon\Big[I(V;Y|U) - I(V;Z|U)\Big] \\ &\overset{(b)}{\geq} \epsilon\Big[I(V;Y|U) - I(V;S|U)\Big] \\ &\overset{(c)}{=} I(V';Y|U') - I(V';S|U') \\ &\overset{(d)}{=} I(U',V';Y') - I(U',V';S) \\ &\overset{(e)}{\geq} R_{\mathsf{A}}^{\mathsf{Alt}}, \end{aligned} \tag{136}$$

where (a) is (134), (b) follows by the assumption that $I(V;S|U) \geq I(V;Z|U)$, (c) is (135), (d) is by choosing ϵ to satisfy $I(U';Y) - I(U';S) = 0$, while (e) uses (133).

Finally, assuming $I(V;S|U) < I(V;Z|U)$ gives:

$$\begin{aligned} &I(V';Y|U') - I(V';Z|U') \\ &\overset{(a)}{=} \epsilon\Big[I(V;Y|U) - I(V;Z|U)\Big] \\ &= I(V;Y|U) - I(V;Z|U) - \bar{\epsilon}\Big[I(V;Y|U) - I(V;Z|U)\Big] \\ &\overset{(b)}{>} I(V;Y|U) - I(V;Z|U) - \bar{\epsilon}\Big[I(V;Y|U) - I(V;S|U)\Big] \\ &\overset{(c)}{=} I(V;Y|U) - I(V;Z|U) + I(U;Y) - I(U;S) \\ &\overset{(d)}{\geq} R_{\mathsf{A}}^{\mathsf{Alt}}, \end{aligned} \tag{137}$$

where (a) is (134), (b) is by the assumption in the second case, (c) uses (131) with $I(U';Y) - I(U';S) = 0$, and, finally, (d) follows by the third rate bound in $R_{\mathsf{A}}^{\mathsf{Alt}}$.

Concluding, we obtain

$$\begin{aligned} R_{\mathsf{A}} &\geq R_{\mathsf{A}}(p_{U',V',X|S}) \\ &= \min\Big\{I(V';Y|U') - I(V';Z|U') \\ &\qquad\qquad , I(U',V';Y) - I(U',V';S)\Big\} \\ &\geq R_{\mathsf{A}}^{\mathsf{Alt}}. \end{aligned} \tag{138}$$

Appendix E
Proof of Corollary 1

A. Direct

We use Theorem 1 to derive achievability of Corollary 1. For any $q_{U,V,X|S} : \mathcal{S} \to \mathcal{U} \times \mathcal{V} \times \mathcal{X}$, replacing Y and Z in $R_{\mathsf{A}}\left(q_{U,V,X|S}\right)$ with (Y,S_1) and (Z,S_2), respectively, implies achievability of

$$\begin{aligned} &R_{\mathsf{A}}^{\mathrm{RLN}}(q_{U,V,X|S}) \\ &= \min\Big\{I(V;Y,S_1|U) - I(V;Z,S_2|U) \\ &\qquad , I(U,V;Y,S_1) - I(U,V;S) \\ &\qquad\quad , I(U,V;Y,S_1) - I(U;S) - I(V;Z,S_2|U)\Big\}. \end{aligned} \tag{139}$$

To properly define the $q_{U,V,X|S}$ that achieves (25), recall the p distribution stated after (24) that factors as $p_S p_{A|S} p_{B|A} p_X p_{S_1,S_2|S} p_{Y,Z|X}$. Let $\tilde{p}$ be a PMF over $\mathcal{S} \times \mathcal{A} \times \mathcal{B} \times \mathcal{X} \times \mathcal{Y} \times \mathcal{Z} \times \mathcal{S}_1 \times \mathcal{S}_2 \times \mathcal{B} \times \mathcal{X}$, such that

$$\tilde{p}_{S,A,B,X,S_1,S_2,Y,Z,\tilde{B},\tilde{X}} = p_{S,A,B,X,S_1,S_2,Y,Z} \mathbb{1}_{\{\tilde{B}=B\} \cap \{\tilde{X}=X\}}. \tag{140}$$

Now, fix $p_{S,A,B,X,S_1,S_2,Y,Z}$ and let $q_{U,V,X|S}$ in (14) be such that $V = (A,B)_{\tilde{p}}$, $U = (\tilde{B},\tilde{X})_{\tilde{p}}$ and $q_{X|S,U,V} = \tilde{p}_X = p_X$, where the subscript $\tilde{p}$ means that the random variables on the RHS are distributed according to their marginal from (140). Consequently, $Q_{U,V,X|S} p_{S_1,S_2|S} p_{Y,Z|X}$ is equal to the RHS of (140). We next evaluate the mutual information terms in R_{A} to show it coincides with (25). We again use the notation I_q,

$I_{\tilde{p}}$ and I_p to indicated that the underlying PMF is q, $\tilde{p}$ or p, respectively. We have

$$\begin{aligned} &I_q(V;Y,S_1|U) - I_q(V;Z,S_2|U) \\ &= I_{\tilde{p}}(A,B;Y,S_1|\tilde{B},\tilde{X}) - I_{\tilde{p}}(A,B;Z,S_2|\tilde{B},\tilde{X}) \\ &\overset{(a)}{=} I_p(A;S_1|B,X) + I_p(A;Y|B,X,S_1) - I_p(A;S_2|B,X) \\ &\qquad\qquad\qquad\qquad\qquad\qquad - I_p(A;Z|B,X,S_2) \\ &\overset{(b)}{=} I_p(A;S_1|B) - I_p(A;S_2|B), \end{aligned} \tag{141}$$

where (a) is because $\tilde{B} = B$ and $\tilde{X} = X$ almost surely and since $\tilde{p}_{S,A,B,X,S_1,S_2,Y,Z} = p_{S,A,B,X,S_1,S_2,Y,Z}$. Step (b) is because in p the chain $(Y,Z) - X - (A,B,S_1,S_2)$ is Markov.

Next, consider

$$\begin{aligned} I_q(U,V;Y,S_1) - I_q(U,V;S) &\\ = I_{\tilde{p}}(A,B,\tilde{B},\tilde{X};Y,S_1) &- I_{\tilde{p}}(A,B,\tilde{B},\tilde{X};S) \\ \overset{(a)}{=} I_p(A,B,X;Y,S_1) &- I_p(A,B,X;S) \\ \overset{(b)}{=} I_p(A,B,X;Y|S_1) &- I_p(A,B;S|S_1) \\ \overset{(c)}{=} I_p(X;Y) &- I_p(A;S|S_1), \end{aligned} \tag{142}$$

where:
(a) is for the same reason as step (a) in the derivation of (141);
(b) is because in p we have the Markov chain $(A,B,X) - S - S_1$, since X is independent of (A,B,S,S_1) and due to the chain rule;
(c) follows because under p, (X,Y) is independent of (A,B,S_1) and since $I(B;S|S_1,A) = 0$ as $B - A - (S,S_1)$ is also a Markov chain.

Finally, we shown that the third term from the RHS of (139) is since $I_q(V;S|U) \geq I_q(V;Z,S_2|U)$ for the aforementioned $q_{U,V,X|S}$. Consider

$$\begin{aligned} I_q(V;Z,S_2|U) &\overset{(a)}{=} I_p(A;S_2|B) \\ &\leq I_p(A;S,S_2|B) \\ &\overset{(b)}{=} I_p(A,B;S) - I(B;S) \\ &\overset{(c)}{=} I_p(A;S|B,X) \\ &\overset{(d)}{=} I_q(A;S|B,X), \end{aligned} \tag{143}$$

where:
(a) is due to similar arguments as those justifying (141);
(b) is because $(A,B) - S - S_2$ is a Markov chain under p;
(c) uses the independence of (A,B,S) and X;
(d) follows from the definition of the $q_{U,V,X|S}$ distribution.

Consequently, the third term in $R_{\mathsf{A}}^{\mathrm{RLN}}(q_{U,V,X|S})$ is redundant because of (142). Along with (141), this establishes the direct part of Corollary 1.

B. Converse

Let $\{c_n\}_{n\in\mathbb{N}}$ be a sequence of (n,R) semantically-secure codes for the SD-WTC with a vanishing maximal error probability. Fix $\epsilon > 0$ and let $n \in \mathbb{N}$ be sufficiently large so that (13) holds. Since both (13a) and (13b) hold for any message distribution $q \in \mathcal{P}(\mathcal{M}_n)$, in particular, they hold for a uniform $p^{(U)}_{\mathcal{M}_n}$. All the following multi-letter information measures are calculated with respect to the induced joint PMF from (9), where the channel $p_{Y,Z|X,S}$ is replaced with $p_{S_1,S_2,Y,Z|X,S}$ defined in Section IV-C1. Fano's inequality gives

$$H(M|S_1^n,Y^n) \leq 1 + n\epsilon R \triangleq n\epsilon_n, \tag{144}$$

where $\epsilon_n = \frac{1}{n} + \epsilon R$.

The security criterion from (13b) and the reversely less noisy property of the channel $p_{Y,Z|X}$ (that, respectively, justify the two following inequalities) further gives

$$\begin{aligned} \epsilon &\geq I(M;S_2^n,Z^n) \\ &= I(M;S_2^n) + \sum_{\mathbf{s}_2\in\mathcal{S}_2^n} p^n_{S_2}(\mathbf{s}_2) I(M;Z^n|S_2^n = \mathbf{s}_2) \\ &\geq I(M;S_2^n) + \sum_{\mathbf{s}_2\in\mathcal{S}_2^n} p^n_{S_2}(\mathbf{s}_2) I(M;Y^n|S_2^n = \mathbf{s}_2) \\ &= I(M;S_2^n,Y^n). \end{aligned} \tag{145}$$

Having (144) and (145), we bound R as

$$\begin{aligned} &nR \\ &= H(M) \\ &\overset{(a)}{\leq} I(M;S_1^n,Y^n) - I(M;S_2^n,Y^n) + n\delta_n \\ &= I(M;S_1^n|Y^n) - I(M;S_2^n|Y^n) + n\delta_n \\ &\overset{(b)}{=} \sum_{i=1}^n \Big[I(M;S_1^i,S_{2,i+1}^n|Y^n) - I(M;S_1^{i-1},S_{2,i}^n|Y^n) \Big] + n\delta_n \\ &\overset{(c)}{=} \sum_{i=1}^n \Big[I(M;S_{1,i}|B_i) - I(M;S_{2,i}|B_i) \Big] + n\delta_n \\ &\overset{(d)}{=} n\Big[I(M;S_{1,T}|B_T,T) - I(M;S_{2,T}|B_T,T) \Big] + n\delta_n \\ &\overset{(e)}{=} n\Big[I(A;S_1|B) - I(A;S_2|B) \Big] + n\delta_n, \end{aligned} \tag{146}$$

where:
(a) is by (144) and (145) while setting $\delta_n \triangleq \epsilon_n + \frac{\epsilon}{n}$;
(b) is a telescoping identity [31, Eqs. (9) and (11)];
(c) defines $B_i \triangleq (S_1^{i-1}, S_{2,i+1}^n, Y^n)$, for all $i \in [1:n]$;
(d) uses the standard time-sharing technique, where T is uniformly distributed over $[1:n]$ and independent of all other random variables in $P^{(c_n)}$;
(e) defines $S \triangleq S_T$, $S_1 \triangleq S_{1,T}$, $S_2 \triangleq S_{2,T}$, $X \triangleq X_T$, $Y \triangleq Y_T$, $Z \triangleq Z_T$, $B \triangleq (B_T,T)$ and $A \triangleq (M,B)$.

Another way to bound R is

$$\begin{aligned} &nR \\ &= H(M) \\ &\overset{(a)}{\leq} I(M;S_1^n,Y^n) + n\epsilon_n \\ &= I(M;S_1^n,Y^n,S^n) - I(M;S^n|S_1^n,Y^n) + n\epsilon_n \\ &\overset{(b)}{=} I(M;Y^n|S_1^n,S^n) - I(M,Y^n;S^n|S_1^n) \\ &\qquad\qquad\qquad\qquad\qquad\qquad + I(S^n;Y^n|S_1^n) + n\epsilon_n \\ &= I(M,S^n;Y^n|S_1^n) - I(M,Y^n;S^n|S_1^n) + n\epsilon_n \\ &\overset{(c)}{\leq} I(M,S^n;Y^n) - I(M,Y^n;S^n|S_1^n) + n\epsilon_n \end{aligned}$$

$$
\begin{aligned}
&\overset{(d)}{\leq} I(X^n;Y^n) - I(M,Y^n;S^n|S_1^n) + n\epsilon_n\\
&\overset{(e)}{\leq} \sum_{i=1}^{n}\Big[I(X_i;Y_i) - I(M,Y^n;S_i|S_1^n,S^{i-1})\Big] + n\epsilon_n\\
&\overset{(f)}{\leq} \sum_{i=1}^{n}\Big[I(X_i;Y_i) - I(M,Y^n,S_1^{n\backslash i},S^{i-1};S_i|S_{1,i})\Big] + n\epsilon_n\\
&\overset{(g)}{\leq} \sum_{i=1}^{n}\Big[I(X_i;Y_i) - I(M,B_i;S_i|S_{1,i})\Big] + n\epsilon_n\\
&\overset{(h)}{\leq} n\Big[I(X_T;Y_T) - I(M,B_T,T;S_T|S_{1,T})\Big] + n\epsilon_n\\
&\overset{(i)}{\leq} n\Big[I(X;Y) - I(A;S|S_1)\Big] + n\epsilon_n, \qquad (147)
\end{aligned}
$$

where:
(a) is by (144);
(b) uses the independence of M and (S_1^n,S^n) (1st term);
(c) is because conditioning cannot increase entropy and since $Y^n-(M,S^n)-S_1^n$ forms a Markov chain (1st term);
(d) uses the Markov relation $Y^n-X^n-(M,S^n)$;
(e) follows since conditioning cannot increase entropy and by the discrete and memoryless property of $p^n_{Y,Z|X}$;
(f) is because $P^{(c_n)}_{S^n,S_1^n,S_2^n}=p^n_{S,S_1,S_2}$, i.e., the marginal distribution of (S^n,S_1^n,S_2^n) are i.i.d.;
(g) is by the definition of B_i;
(h) uses time-sharing (see step (d) from (146)), the fact that conditioning cannot increase entropy and the Markov relation Y_T-X_T-T (1st term), and $\mathbb{P}\big(S_T=s,S_{1,T}=s_1,T=t\big)=p_{S,S_1}(s,s_1)p_T(t)$, for all $(s,s_1,t)\in\mathcal{S}\times\mathcal{S}_1\times[1:n]$ (2nd term);
(i) reuses the definition of the single-letter random variable from step (e) in the derivation of (146).

The joint distribution of the above variables factors as

$$
\begin{aligned}
&\mathbb{P}\big(S=s,S_1=s_1,S_2=s_2,A=a,B=b,X=x,Y=y,Z=z\big)\\
&\quad=p_S(s)p_{S_1,S_2|S}(s_1,s_2|s)\mathbb{P}\big(A=a\big|S=s,S_1=s_1,S_2=s_2\big)\\
&\qquad\times\mathbb{P}\big(B=b\big|A=a\big)\\
&\qquad\quad\times\mathbb{P}\big(X=x\big|S=s,S_1=s_1,S_2=s_2,A=a,B=b\big)\\
&\qquad\qquad\times p_{Y,Z|X}(y,z|x), \qquad (148)
\end{aligned}
$$

where the equalities $\mathbb{P}\big(S=s,S_1=s_1,S_2=s_2\big)=p_S(s)p_{S_1,S_2|S}(s_1,s_2|s)$ and $\mathbb{P}\big(Y=y,Z=z\big|S=s,S_1=s_1,S_2=s_2,A=a,B=b,X=x\big)=p_{Y,Z|X}(y,z|x)$ are straightforward from the probabilistic relations in $P^{(c_n)}$ and the definition of T, while $\mathbb{P}\big(B=b\big|S=s,S_1=s_1,S_2=s_2,A=a\big)=\mathbb{P}\big(B=b\big|A=a\big)$ follows because $A=(M,B)$. Furthermore, for every $(s,s_1,s_2,a)\in\mathcal{S}\times\mathcal{S}_1\times\mathcal{S}_2\times\mathcal{A}$, it holds that $\mathbb{P}\big(A=a\big|S=s,S_1=s_1,S_2=s_2\big)=\mathbb{P}\big(A=a\big|S=s\big)$. To see this, for any $(s^n,s_1^n,s_2^n,y^n)\in\mathcal{S}^n\times\mathcal{S}_1^n\times\mathcal{S}_2^n\times\mathcal{Y}^n$, we define the corresponding realization of A as $a=(t,m,b_t)$, where $(t,m)\in[1:n]\in\mathcal{M}_n$ and $b_t=\big(y^n,s_1^{t-1},s_{2,t+1}^n\big)$. For any $(s_t,s_{1,t},s_{2,t})\in\mathcal{S}\times\mathcal{S}_1\times\mathcal{S}_2$, we have

$$
\begin{aligned}
&\mathbb{P}\big(A=a\big|S=s_t,S_1=s_{1,t},S_2=s_{2,t}\big)\\
&\overset{(a)}{=} p_T(t)P^{(c_n)}\left(m,s_1^{t-1},s_{2,t+1}^n,y^n|s_t,s_{1,t},s_{2,t}\right)\\
&=p_T(t)\sum_{\substack{(s^{n\backslash t},x^n)\\ \in\mathcal{S}^{n-1}\times\mathcal{X}^n}}P^{(c_n)}\left(s^{n\backslash t},x^n,m,s_1^{t-1},s_{2,t+1}^n,y^n\middle|s_t,s_1^t,s_{2,t}^n,m\right)\\
&\overset{(b)}{=} p_T(t)p^{(U)}_{\mathcal{M}_n}(m)\\
&\quad\times\sum_{\substack{(s^{n\backslash t},x^n)\\ \in\mathcal{S}^{n-1}\times\mathcal{X}^n}}p_S^{n-1}\left(s^{n\backslash t}\right)p^{t-1}_{S_1|S}\left(s_1^{t-1}|s^{t-1}\right)p^{n-t}_{S_2|S}\left(s_{2,t+1}^n|s_{t+1}^n\right)\\
&\qquad\qquad\times f_n\left(x^n\middle|m,s^n\right)p^n_{Y|X}\left(y^n|x^n\right)\\
&=p_T(t)P^{(c_n)}\left(m,s_1^{t-1},s_{2,t+1}^n,y^n|s_t\right)\\
&=\mathbb{P}\big(A=a\big|S=s_t\big), \qquad (149)
\end{aligned}
$$

where (a) is because T is independent of all other random variables, while (b) uses the dependence relations in $P^{(c_n)}$ from (9) with $p_{S_1,S_2|S}p_{Y,Z|X}$ in the role of the SDWTC.

Denoting $\mathbb{P}\big(A=a\big|S=s\big)\triangleq p_{A|S}(a|s)$, $\mathbb{P}\big(B=b\big|A=a\big)\triangleq p_{B|A}(b|a)$ and $\mathbb{P}\big(X=x\big|S=s,S_1=s_1,S_2=s_2,A=a,B=b\big)\triangleq p_{X|S,S_1,S_2,A,B}(x|s,s_1,s_2,a,b)$, we have the following bound on the achievable rate:

$$
\begin{aligned}
R\leq&\frac{\min\Big\{I(A;S_1|B)-I(A;S_2|B),I(X;Y)-I(A;S|S_1)\Big\}}{1-\epsilon}\\
&\qquad+\frac{1}{(1-\epsilon)n}+\frac{\epsilon}{1-\epsilon}, \qquad (150)
\end{aligned}
$$

where the mutual information terms are with respect to $p_Sp_{S_1,S_2|S}p_{A|S}p_{B|A}p_{X|S,S_1,S_2,A,B}p_{Y,Z|X}$. Noting that in none of the mutual information terms from (150) do X and (S,S_1,S_2,A,B) appear together, we may replace $p_{X|S,S_1,S_2,A,B}$ with p_X without affecting the expressions. Taking $\epsilon\to 0$ and $n\to\infty$ completes the proof of the converse.

Appendix F
Converse Proof for Corollary 2

Let $\{c_n\}_{n\in\mathbb{N}}$ be a sequence of (n,R) codes for the SD-WTC satisfying (13). By similar arguments to those from the converse proof in Appendix E-B, we assume a uniform message distribution and note that all the following multi-letter information measures are with respect to (9). By Fano's inequality, we have

$$
H(M|Y^n)\leq 1+n\epsilon R\triangleq n\epsilon_n, \qquad (151)
$$

where $\epsilon_n=\frac{1}{n}+\epsilon R$.

First, we bound the rate R as

$$
\begin{aligned}
nR&=H(M)\\
&\overset{(a)}{\leq} I(M;Y^n)-I(M;Z^n)+n\epsilon_n'\\
&\leq I(M;Y^n|Z^n)+n\epsilon_n'\\
&\overset{(b)}{\leq}\sum_{i=1}^{n}H(Y_i|Z_i)+n\epsilon_n', \qquad (152)
\end{aligned}
$$

where (a) uses (13b) and (151) and defines $\epsilon_n'\triangleq\epsilon_n+\frac{\epsilon}{n}$, and (b) follows by the chain rule and since conditioning cannot increase entropy.

Another way to bound R is:

$$
nR=H(M)
$$

$$
\begin{aligned}
&\overset{(a)}{\leq} I(M;Y^n) - I(M;S^n) + n\epsilon_n \\
&\leq I(M;Y^n|S^n) + n\epsilon_n' \\
&\overset{(b)}{\leq} \sum_{i=1}^{n} H(Y_i|S_i) + n\epsilon_n, \qquad (153)
\end{aligned}
$$

where (a) is due to (151) and because M and S^n are independent in (9), while (b) is similar to step (b) in (152). Having (152)-(153), the converse is established by standard time-sharing arguments (as in Appendix E).

Appendix G
Proof of Lemma 2

First note that for any $\mathcal{C}_n \in \mathfrak{C}_n$ and $(i,j,m,\mathbf{s}) \in \mathcal{I}_n \times \mathcal{J}_n \times \mathcal{M}_n \times \mathcal{S}^n$, we have

$$
\begin{aligned}
Q^{(\mathcal{C}_n)}(i,j|m,\mathbf{s}) &= \frac{Q^{(\mathcal{C}_n)}(m,i,j,\mathbf{s})}{Q^{(\mathcal{C}_n)}(m,\mathbf{s})} \\
&= \frac{p^n_{S|U,V}\big(\mathbf{s}\big|\mathbf{u}(i),\mathbf{v}(i,j,m)\big)}{\sum\limits_{(i',j')\in\mathcal{I}_n\times\mathcal{J}_n} p^n_{S|U,V}\big(\mathbf{s}\big|\mathbf{u}(i'),\mathbf{v}(i',j',m)\big)} \\
&= P^{(\mathcal{C}_n)}(i,j|m,\mathbf{s}), \qquad (154)
\end{aligned}
$$

where the last step uses the definition from (33). Having (154), note that

$$
\begin{aligned}
&\Big|\Big|P^{(\mathcal{C}_n)} - Q^{(\mathcal{C}_n)}\Big|\Big|_{\mathsf{TV}} \\
&\overset{(a)}{=} \sum_{m\in\mathcal{M}_n} \frac{1}{|\mathcal{M}_n|} \\
&\quad \times \Big|\Big|P^{(\mathcal{C}_n)}_{\mathbf{S},I,J,\mathbf{U},\mathbf{V},\mathbf{X},\mathbf{Y},\mathbf{Z}|M=m} - Q^{(\mathcal{C}_n)}_{\mathbf{S},I,J,\mathbf{U},\mathbf{V},\mathbf{X},\mathbf{Y},\mathbf{Z}|M=m}\Big|\Big|_{\mathsf{TV}} \\
&\overset{(b)}{=} \frac{1}{|\mathcal{M}_n|} \sum_{m\in\mathcal{M}_n} \Big|\Big|p^n_S - Q^{(\mathcal{C}_n)}_{\mathbf{S}|M=m}\Big|\Big|_{\mathsf{TV}}, \qquad (155)
\end{aligned}
$$

where (a) uses $Q^{(\mathcal{C}_n)}_M = P^{(\mathcal{C}_n)}_M = p^{(U)}_{\mathcal{M}_n}$, while (b) is because for any $p_X, q_X \in \mathcal{P}(\mathcal{X})$ and $p_{Y|X} : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$ we have $\big|\big|p_X p_{Y|X} - q_X p_{Y|X}\big|\big|_{\mathsf{TV}} = \big|\big|p_X - q_X\big|\big|_{\mathsf{TV}}$. Combining this with (154) and the relations

$$
\begin{aligned}
Q^{(\mathcal{C}_n)}_{\mathbf{U},\mathbf{V}|I,J,\mathbf{S},M=m} &= \mathbb{1}_{\big\{\mathbf{U}=\mathbf{u}(I)\big\}\cap\big\{\mathbf{V}=\mathbf{v}(I,J,m)\big\}} \\
&= P^{(\mathcal{C}_n)}_{\mathbf{U},\mathbf{V}|I,J,\mathbf{S},M=m} \qquad (156a) \\
Q^{(\mathcal{C}_n)}_{\mathbf{X},\mathbf{Y},\mathbf{Z}|\mathbf{U},\mathbf{V},I,J,\mathbf{S},M=m} &= p^n_{X|U,V,S}\, p^n_{Y,Z|X,S} \\
&= P^{(\mathcal{C}_n)}_{\mathbf{X},\mathbf{Y},\mathbf{Z}|\mathbf{U},\mathbf{V},I,J,\mathbf{S},M=m} \qquad (156b)
\end{aligned}
$$

justifies (b).

Now, consider

$$
\begin{aligned}
&\mathbb{E}_\mu \Big|\Big|P^{(\mathsf{C}_n)} - Q^{(\mathsf{C}_n)}\Big|\Big|_{\mathsf{TV}} \\
&\quad\overset{(a)}{\leq} \mathbb{E}_\mu \frac{1}{|\mathcal{M}_n|} \sum_{m\in\mathcal{M}_n} \Big|\Big|p^n_S - Q^{(\mathcal{C}_n)}_{\mathbf{S}|M=m}\Big|\Big|_{\mathsf{TV}} \\
&\quad\overset{(b)}{=} \mathbb{E}_\mu \Big|\Big|p^n_S - Q^{(\mathcal{C}_n)}_{\mathbf{S}|M=1}\Big|\Big|_{\mathsf{TV}} \\
&\quad\overset{(c)}{\leq} \mathbb{E}_\mu \sqrt{\frac{1}{2}\mathsf{D}\Big(Q^{(\mathcal{C}_n)}_{\mathbf{S}|M=1}\Big|\Big|p^n_S\Big)} \\
&\quad\overset{(c)}{\leq} \sqrt{\frac{1}{2}\mathbb{E}_\mu \mathsf{D}\Big(Q^{(\mathcal{C}_n)}_{\mathbf{S}|M=1}\Big|\Big|p^n_S\Big)}, \qquad (157)
\end{aligned}
$$

where (a) is due to (155), (b) follws by symmetry of the codebook with respect to $m \in \mathcal{M}_n$, (c) is Pinsker's Inequality, and (d) is Jensen's inequality.

To conclude the proof, note that the expectation on the RHS of (157) falls within the framework of the SCL for superposition codes (Lemma 5), with the DMC $p^n_{S|U,V}$. Taking (R_1, R_2) as in (38) implies that there exist $\tilde{\alpha} > 0$ such that for any n large enough

$$
\mathbb{E}_\mu \mathsf{D}\Big(Q^{(\mathsf{C}_n)}_{\mathbf{S}|M=m}\Big|\Big|p^n_S\Big) \leq e^{-n\tilde{\alpha}}. \qquad (158)
$$

Combining this with (157) proves Lemma 2 with $\alpha = \frac{\tilde{\alpha}}{2}$.

Appendix H
Proof of Lemma 3

To simplify notation, throughout his proof we abbreviate $I_{P^{(\mathcal{C}_n)}}$ and $I_{Q^{(\mathcal{C}_n)}}$ as I_P and I_Q, respectively. Consider:

$$
\begin{aligned}
&\Big|I_P(M;\mathbf{Z}) - I_Q(M;\mathbf{Z})\Big| \\
&\overset{(a)}{\leq} \Big|H_P(\mathbf{Z}) - H_Q(\mathbf{Z})\Big| + \Big|H_Q(M,\mathbf{Z}) - H_P(M,\mathbf{Z})\Big| \\
&\overset{(b)}{\leq} \Big|\Big|P^{(\mathcal{C}_n)}_{\mathbf{Z}} - Q^{(\mathcal{C}_n)}_{\mathbf{Z}}\Big|\Big|_{\mathsf{TV}} \log \frac{|\mathcal{Z}^n|}{\Big|\Big|P^{(\mathcal{C}_n)}_{\mathbf{Z}} - Q^{(\mathcal{C}_n)}_{\mathbf{Z}}\Big|\Big|_{\mathsf{TV}}} \\
&\qquad + \Big|\Big|P^{(\mathcal{C}_n)}_{M,\mathbf{Z}} - Q^{(\mathcal{C}_n)}_{M,\mathbf{Z}}\Big|\Big|_{\mathsf{TV}} \log \frac{|\mathcal{M}_n| \cdot |\mathcal{Z}^n|}{\Big|\Big|P^{(\mathcal{C}_n)}_{M,\mathbf{Z}} - Q^{(\mathcal{C}_n)}_{M,\mathbf{Z}}\Big|\Big|_{\mathsf{TV}}} \\
&\overset{(c)}{\leq} e^{-n\beta_1}\Big(n\log|\mathcal{Z}| + n\log\big(2^R|\mathcal{Z}|\big)\Big) - \Big|\Big|P^{(\mathcal{C}_n)}_{\mathbf{Z}} \\
&\qquad - Q^{(\mathcal{C}_n)}_{\mathbf{Z}}\Big|\Big|_{\mathsf{TV}} \log \Big|\Big|P^{(\mathcal{C}_n)}_{\mathbf{Z}} - Q^{(\mathcal{C}_n)}_{\mathbf{Z}}\Big|\Big|_{\mathsf{TV}} \\
&\qquad - \Big|\Big|P^{(\mathcal{C}_n)}_{M,\mathbf{Z}} - Q^{(\mathcal{C}_n)}_{M,\mathbf{Z}}\Big|\Big|_{\mathsf{TV}} \log \Big|\Big|P^{(\mathcal{C}_n)}_{M,\mathbf{Z}} - Q^{(\mathcal{C}_n)}_{M,\mathbf{Z}}\Big|\Big|_{\mathsf{TV}}, \qquad (159)
\end{aligned}
$$

where (a) is because $H_P(M) = H_Q(M)$, (b) uses [32, Theorem 17.3.3], while (c) is (52).

The function $x \mapsto -x\log x$ is monotone increasing for $x \in \Big[0, 2^{-\frac{1}{\ln 2}}\Big]$ and, for large enough values of n, we have $e^{-n\beta_1} \in \Big[0, 2^{-\frac{1}{\ln 2}}\Big]$. Therefore, as $\Big|\Big|P^{(\mathcal{C}_n)}_{\mathbf{Z}} - Q^{(\mathcal{C}_n)}_{\mathbf{Z}}\Big|\Big|_{\mathsf{TV}} \leq \Big|\Big|P^{(\mathcal{C}_n)}_{M,\mathbf{Z}} - Q^{(\mathcal{C}_n)}_{M,\mathbf{Z}}\Big|\Big|_{\mathsf{TV}} \leq e^{-n\beta_1}$, we have that for sufficiently large n

$$
\begin{aligned}
&-\Big|\Big|P^{(\mathcal{C}_n)}_{\mathbf{Z}} - Q^{(\mathcal{C}_n)}_{\mathbf{Z}}\Big|\Big|_{\mathsf{TV}} \log \Big|\Big|P^{(\mathcal{C}_n)}_{\mathbf{Z}} - Q^{(\mathcal{C}_n)}_{\mathbf{Z}}\Big|\Big|_{\mathsf{TV}} \\
&\qquad - \Big|\Big|P^{(\mathcal{C}_n)}_{M,\mathbf{Z}} - Q^{(\mathcal{C}_n)}_{M,\mathbf{Z}}\Big|\Big|_{\mathsf{TV}} \log \Big|\Big|P^{(\mathcal{C}_n)}_{M,\mathbf{Z}} - Q^{(\mathcal{C}_n)}_{M,\mathbf{Z}}\Big|\Big|_{\mathsf{TV}} \\
&\leq -2e^{-n\beta_1}\log e^{-n\beta_1}. \qquad (160)
\end{aligned}
$$

Inserting (160) into (159) gives

$$
\Big|I_P(M;\mathbf{Z}) - I_Q(M;\mathbf{Z})\Big| \leq ne^{-n\beta_1}\left(2\log|\mathcal{Z}| + R + \frac{2\beta_1}{\ln 2}\right), \qquad (161)
$$

for the aforementioned values of n. This implies that (53) holds and concludes the proof of Lemma 3.

References

[1] S. I. Gelfand and M. S. Pinsker. Coding for channel with random parameters. *Problemy Pered. Inform. (Problems of Inf. Trans.)*, 9(1):19–31, 1980.

[2] A. D. Wyner. The wire-tap channel. *Bell Sys. Techn.*, 54(8):1355–1387, Oct. 1975.

[3] I. Csiszár and J. Körner. Broadcast channels with confidential messages. *IEEE Trans. Inf. Theory*, 24(3):339–348, May 1978.

[4] Y. Chen and A. J. Han Vinck. Wiretap channel with side information. *IEEE Trans. Inf. Theory*, 54(1):395–402, Jan. 2008.

[5] W. Liu and B. Chen. Wiretap channel with two-sided state information. In *Proc. 41st Asilomar Conf. Signals, Syst. Comp*, page 893897, Pacific Grove, CA, US, Nov. 2007.

[6] Y.-K. Chia and A. El Gamal. Wiretap channel with causal state information. *IEEE Trans. Inf. Theory*, 58(5):2838–2849, May 2012.

[7] A. Khisti, S. N. Diggavi, and G. W. Wornell. Secret-key agreement with channel state information at the transmitter. *IEEE Trans. Inf. Forensics Security*, 6(3):672–681, Mar. 2011.

[8] Y. Chen, N. Cai, and A. Sezgin. Wiretap channel with correlated sources. In *Proc. IEEE Int. Conf. Cloud Eng. (ICE-2014)*, pages 472–477, Istanbul, Turkey, Mar. 2014.

[9] B. Dai, A. J. Han Vinck, Y. Luo, and X. Tang. Secret-key agreement with channel state information at the transmitter. *Entropy*, 15:445473, 2013.

[10] G. Bassia, P. Piantanida, and S. Shamai. The wiretap channel with generalized feedback: Secure communication and key generation. *Submitted to IEEE Trans. Inf. Theory*, 2015. Available on ArXiv at http://arxiv.org/abs/1507.07091.

[11] V. Prabhakaran, K. Eswaran, and K. Ramchandran. Secrecy via sources and channels. *IEEE Trans. Inf. Theory*, 85(11):6747–6765, Nov. 2012.

[12] E. Song, P. Cuff, and V. Poor. The likelihood encoder for lossy compression. *IEEE Trans. Inf. Theory*, 62(4):1836–1849, Apr. 2016.

[13] J. Villard and P. Piantanida. Secure lossy source coding with side information at the decoders. In *Proc. 48th Annu. Allerton Conf. Commun., Control and Comput.*, Monticell, Illinois, United States, Sep. 2010.

[14] R. Tandon, S. Ulukus, and K. Ramchandran. Secure source coding with a helper. *IEEE Trans. Inf. Theory*, 59(4):2178–2187, Apr. 2013.

[15] M. Benammar and A. Zaidi. Lossy source and gray-wyner source coding. In *Proc. Int. Symp. Inf. Theory (ISIT-2016)*, Barcelona, Spain, Jul 2016.

[16] M. Benammar and A. Zaidi. On lossy source coding with equivocation constraints. In *Proc. Inf. Theory Workshop (ITW-2016)*, Cambridge, UK, Sep. 2016.

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

[18] J. L. Massey. *Applied Digital Information Theory*. ETH Zurich, Zurich, Switzerland, 1980-1998.

[19] I. Csiszar. Information-type measures of difference of probability distributions and indirect observations. *Studia Scientiarum Mathematicarum Hungarica*, 2:299–318, Jan. 1967.

[20] Z. Goldfeld, G. Kramer, H. H. Permuter, and P. Cuff. Strong secrecy for cooperative broadcast channels. *Submitted for publication to IEEE Trans. Inf. Theory*, 2016. Available on ArXiv at http://arxiv.org/abs/1601.01286.

[21] T. M. Cover and J. A. Thomas. *Elements of information theory*. Wiley-Interscience, New York, NY, USA, 1991.

[22] H. G. Eggleston. *Convexity*. Cambridge University Press, Cambridge, England York, 6th edition edition, 1958.

[23] A. Zibaeenejad. Key generation over wiretap models with non-causal side information. *IEEE Trans. Inf. Forensics Security*, 10(7):1456–1471, July 2015.

[24] 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 Security*, June 2018. Available on ArXiv at https://arxiv.org/abs/1708.04283.

[25] A. Khisti, S. N. Diggavi, and G. W. Wornell. Secret-key generation using correlated sources and channels. *IEEE Trans. Inf. Theory*, 58(2):652–670, Feb. 2012.

[26] P. Cuff. Distributed channel synthesis. *IEEE. Trans. Inf. Theory*, 59(11):7071–7096, Nov. 2013.

[27] Z. Goldfeld, P. Cuff, and H. H. Permuter. Semantic-security capacity for wiretap channels of type II. *IEEE Trans. Inf. Theory*, 62(7):1–17, Jul. 2016.

[28] J. Renes and R. Renner. Noisy channel coding via privacy amplification and information reconciliation. *IEEE Trans. Inf. Theory*, 57(11):7377–7385, Nov. 2011.

[29] Z. Goldfeld, P. Cuff, and H. H. Permuter. Arbitrarily varying wiretap channels with type constrained states. *Submitted to IEEE Trans. Inf. Theory*, 2016. Available on ArXiv at http://arxiv.org/abs/1601.03660.

[30] T. van Erven and P. Harremoës. Rényi divergence and Kullback-Leibler divergence. *IEEE Trans. Inf. Theory*, 60(7):3797–3820, Jul. 2014.

[31] G. Kramer. Teaching IT: An identity for the Gelfand-Pinsker converse. *IEEE Inf. Theory Society Newsletter*, 61(4):4–6, Dec. 2011.

[32] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley, New-York, 2nd edition, 2006.

Ziv Goldfeld (S'13-M'17) received his B.Sc. (summa cum laude), M.Sc. (summa cum laude) and Ph.D. degrees in Electrical and Computer Engineering from the Ben-Gurion University, Israel, in 2012, 2014 and 2017, respectively. Between 2017-2019 he was a postdoctoral fellow at the Laboratory for Information and Decision Systems (LIDS) at MIT. Since 2019 he has been an Assistant Professor of Electrical and Computer Engineering at Cornell University.

Prof. Goldfeld is a recipient of several awards, among them the are Rothschild postdoctoral fellowship, the Feder Award, a best student paper award in the IEEE 28-th Convention of Electrical and Electronics Engineers in Israel, the Lev-Zion fellowship and the Minerva Short-Term Research Grant (MRG).

Paul Cuff (S'08-M'10) received the B.S. degree in electrical engineering from Brigham Young University, Provo, UT, in 2004 and the M.S. and Ph.D. degrees in electrical engineering from Stanford University in 2006 and 2009. From 2009 to 2017 he was an assistant professor of electrical engineering at Princeton University. Since 2017 he has been a member of the general research group at Renaissance Technologies.

As a graduate student, Dr. Cuff was awarded the ISIT 2008 Student Paper Award for his work titled Communication Requirements for Generating Correlated Random Variables and was a recipient of the National Defense Science and Engineering Graduate Fellowship and the Numerical Technologies Fellowship. As faculty he received the NSF Career Award in 2014 and the AFOSR Young Investigator Program Award in 2015.

Haim H. Permuter (M'08-SM'13) received his B.Sc. (summa cum laude) and M.Sc. (summa cum laude) degrees in Electrical and Computer Engineering from the Ben-Gurion University, Israel, in 1997 and 2003, respectively, and a Ph.D. degree in Electrical Engineering from Stanford University, California in 2008.

Between 1997 and 2004, he was an officer at a research and development unit of the Israeli Defense Forces. Since 2009 he is with the department of Electrical and Computer Engineering at Ben-Gurion University where he is currently an associate professor.

Prof. Permuter is a recipient of several awards, among them are the Fullbright Fellowship, the Stanford Graduate Fellowship (SGF), Allon Fellowship, and the U.S.-Israel Binational Science Foundation Bergmann Memorial Award. He is currently serving on the editorial board of the IEEE Transactions on Information Theory.