Exact Random Coding Secrecy Exponents for the Wiretap Channel

Mani Bastani Parizi, *Student Member, IEEE,* Emre Telatar, *Fellow, IEEE,*

and Neri Merhav, *Fellow, IEEE*

Abstract

We analyze the exact exponential decay rate of the expected amount of information leaked to the wiretapper in Wyner's wiretap channel setting using wiretap channel codes constructed from both i.i.d. and constant-composition random codes. Our analysis for those sampled from i.i.d. random coding ensemble shows that the previously-known achievable secrecy exponent using this ensemble is indeed the exact exponent for an average code in the ensemble. Furthermore, our analysis on wiretap channel codes constructed from the ensemble of constant-composition random codes, leads to an exponent which, in addition to being the exact exponent for an average code, is larger than the achievable secrecy exponent that has been established so far in the literature for this ensemble (which in turn was known to be smaller than that achievable by wiretap channel codes sampled from i.i.d. random coding ensemble). We also show examples where the exact secrecy exponent for the wiretap channel codes constructed from random constant-composition codes is larger than that of those constructed from i.i.d. random codes.

Index Terms

Wiretap channel, Channel resolvability, Secrecy exponent, Resolvability exponent

I. INTRODUCTION

The problem of communication in presence of an eavesdropper wiretapping the signals sent to the legitimate receiver (see Figure 1) was first studied by Wyner [1] and later, in a broader context, by Csiszár and Körner [2], where it was shown (among others) that as long as the eavesdropper's channel is weaker than that of the legitimate receiver, reliable and *secure* communication at positive rates is feasible. More precisely, it was shown that, given any distribution on the common input alphabet of the channels, P_X, for which the mutual information developed across the legitimate receiver's channel is higher than that developed across the wiretapper's channel, that is, $I(X;Y) > I(X;Z)$, with $(X, Y, Z) \sim P_X(x)W_{\mathrm{M}}(y|x)W_{\mathrm{E}}(z|x)$ (where X, Y, and Z represent the common

input, legitimate receiver's channel output, and wiretapper's channel output, respectively), as long as the secret message rate $R_{\rm s} \triangleq \frac{1}{n} \log |\mathcal{S}_n|$ is below $I(X;Y) - I(X;Z)$ there exists a sequence of coding schemes (indexed by the block-length n) using which

$$\lim_{n \to \infty} \max_{s \in \mathcal{S}_n} \Pr\{\hat{s}_{\rm ML}(Y^n) \neq S | S = s\} = 0, \tag{1a}$$

$$\lim_{n \to \infty} \frac{1}{n} I(S; Z^n) = 0. \tag{1b}$$

In the above, S represents the secret message, $\hat{s}_{\rm ML}(Y^n)$ is the maximum-likelihood (ML) estimation of the sent message given the output sequence of the legitimate receiver's channel and Z^n represents the output sequence of the wiretapper's channel (see Figure 1).
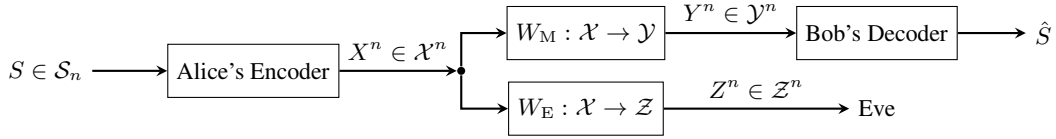

Fig. 1. Wiretap Channel

Classical codes for the wiretap channel are constructed by associating each message with a (random) code that operates at a rate R *just below* the mutual information developed across the eavesdropper's channel. To communicate a message, the stochastic encoder of Alice picks a codeword uniformly at random from the code associated to that message and transmits it via consecutive uses of the channel [1]–[3]. Such constructions, known as *capacity-based constructions* (with a slight abuse of terminology) [4], will guarantee that the normalized amount of information that Eve learns about the secret message by observing her channel output signal, $\frac{1}{n} I(S; Z^n)$, will be arbitrarily small, provided that the block-length n is sufficiently large. Recently, *resolvability-based* constructions for wiretap channel codes, namely, those associating each message with a (random) code operating at a rate *just above* the mutual information of the wiretapper's channel was shown to be more powerful than the capacity-based constructions to prove achievability results. Among other useful properties surveyed in [4], such constructions can be used to easily show that the *unnormalized* amount of information Eve learns about the secret message, $I(S; Z^n)$, vanishes as the block-length increases, namely to establish *strong secrecy* (a notion first introduced by Maurer and Wolf [5]). In particular, using the resolvability-based wiretap channel codes for stationary memoryless wiretap channels, it can be shown that the amount of information Eve learns about the secret message vanishes *exponentially fast* in the block-length, thus, it is natural to study the rate of this exponential decay.

Definition 1. Given the rate pair $(R_{\rm s}, R)$ and a pair of stationary memoryless channels $(W_{\rm M}, W_{\rm E})$, a number η is an achievable *secrecy exponent* for the wiretapper channel $W_{\rm E}$, if there exists a sequence of coding schemes of block-length n and secret message rate $R_{\rm s}$, requiring the entropy rate R at the encoder that are reliable for communication over $W_{\rm M}$ and guarantee

$$\liminf_{n \to \infty} -\frac{1}{n} \log I(S; Z^n) \geq \eta. \tag{2}$$

Hayashi [6] was first to derive a lower bound to the achievable secrecy exponents using the resolvability-based construction of wiretap channel codes from i.i.d. random codes and, later on, improved this lower bound using *privacy amplification* in [7]. More recently, it was shown (see special cases of [8, Theorem 2], [9, Theorem 3.1], or the proof given in [10]) that privacy amplification is unnecessary and the exponent derived in [7] lower-bounds the exponential decay rate of the ensemble average of the information leaked to Eve when a wiretap channel code constructed from i.i.d. random codes is used for communication.

To study the *universally achievable* (as defined in [11]) secrecy exponents in [12], constructing codes for wiretap channel from random constant-composition codes is investigated and, in conjunction with privacy amplification, a lower bound to the achievable secrecy exponent using this class of wiretap channel codes is derived. This lower bound is also shown to be smaller than the lower bound to the achievable exponent using i.i.d. random codes derived in [7].

A. Contribution and Paper Outline

In this paper we, firstly, show that the exponent derived via the method of [10] (which was first established in [7]) is indeed the *exact* secrecy exponent for an average code in the ensemble and, secondly, extend the analysis of [10] to the ensemble of constant-composition random codes (see Theorem 2 and its corollary). This, in particular, implies the previously-known lower bound to the achievable secrecy exponent using i.i.d. random codes characterizes the exact exponential decay rate of the average amount of information leaked to the eavesdropper using wiretap channel codes constructed from i.i.d. random codes. Moreover, it turns out that the exact secrecy exponent for the wiretap channel codes constructed from constant-composition random codes is larger than the lower bound derived in [12] and there are examples where this dominance is strict. Further, examples show that in general there is no ordering between the secrecy exponents of the ensembles of i.i.d. and constant-composition codes. In other words, for some channels the i.i.d. ensemble yields a better secrecy exponent, whereas in the others, the constant-composition ensemble prevails (see Section IV-B).

The analysis of [10] is based on pure random coding arguments (no privacy amplification is used) and is carried out by lower-bounding the achievable *resolvability exponents* (see Definition 5) using random codes. We will show, in this work, that this method not only proves the achievability of the exponent, but also, using very similar steps, establishes its exactness (see Definition 7). On the other side, a simple observation shows that the exact resolvability exponent equals the exact secrecy exponent for an ensemble (see Theorem 1), which in turn, allows us to conclude the exponent derived through this method is the exact secrecy exponent as well.

The remainder of this paper is organized as follows. After setting our notation conventions in Section II, we prove the equivalence of secrecy and resolvability exponents in Section III and reduce the analysis of the exact secrecy exponent for an ensemble to that of the exact resolvability exponent. We present our main result on exact secrecy exponents in Section IV, argue that the exact secrecy exponent for the ensemble of constant-composition random codes is larger than the lower bound derived in [12], and give numerical examples comparing the exponents for two ensembles of i.i.d. and constant-composition random codes. Our main result is proved in Section V. To streamline the presentation, we relegate the straightforward but tedious parts of the proof to the appendices.

B. Related Work

In addition to those cited above, [13] also presents a simple achievability proof for channel resolvability. Based on this proof the authors, in their subsequent work [14], establish strong secrecy for wiretap channel using resolvability-based constructions for wiretap channel codes. The performance of a code for the wiretap channel is measured via two figures of merit, namely, the error probability and information leakage, both of which decay exponentially in block-length when a wiretap channel code from the ensemble of random codes is employed on stationary memoryless channels (as we will also briefly discuss in Section III). The trade-off between secrecy and error exponents (as well as other generalizations of the model) is studied in [15].

Another important problem, in the realm of information-theoretic secrecy, is *secret key agreement* [16], [17]. The secrecy exponents related to this model are studied in [7], [15], [18], [19] and, in particular, in [18], [19] shown to be exact.

II. NOTATION

We use uppercase letters (like X) to denote a random variable and the corresponding lowercase version (x) for a realization of that random variable. The same convention applies to the sequences, i.e., $x^n = (x_1, \ldots, x_n)$ denotes a realization of the random sequence $X^n = (X_1, \ldots, X_n)$. We denote finite sets by script-style uppercase letters like $\mathcal{A}$. The cardinality of the set $\mathcal{A}$ is denoted by $|\mathcal{A}|$.

We write $f(n) \stackrel{.}{\leq} g(n)$ if there exists a function $p(n)$ such that $\limsup_{n \to \infty} \frac{1}{n} \log(p(n)) = 0$ and $f(n) \leq p(n)g(n)$. As noted in [20], when functions f and g depend on other variables than n it is understood that $p(n)$ can only depend on channel transition probabilities, the cardinality of its input and output alphabet, and its input distribution (and not the other parameters f and g may depend on).[1] $f(n) \stackrel{.}{=} g(n)$ means $f(n) \stackrel{.}{\leq} g(n)$ and $f(n) \stackrel{.}{\geq} g(n)$.

For $a \in \mathbb{R}$, $[a]^+ \triangleq \max\{a, 0\}$ denotes positive clipping.

We denote the set of distributions on alphabet $\mathcal{X}$ as $\mathcal{P}(\mathcal{X})$. If $P \in \mathcal{P}(\mathcal{X})$, $P^n \in \mathcal{P}(\mathcal{X}^n)$ denotes the product distribution $P^n(x^n) \triangleq \prod_{i=1}^{n} P(x_i)$ (where x^n denotes the length-n sequence $(x_1, \ldots, x_n) \in \mathcal{X}^n$). Likewise, if $V : \mathcal{X} \to \mathcal{Y}$ is a conditional distribution (that is, $\forall x \in \mathcal{X} : V(\cdot|x) \in \mathcal{P}(\mathcal{Y})$), $V^n : \mathcal{X}^n \to \mathcal{Y}^n$ denotes the conditional distribution $V^n(y^n|x^n) = \prod_{i=1}^{n} V(y_i|x_i)$. For a joint distribution $Q \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$, Q_X (respectively Q_Y) denotes its x- (respectively y-) marginal. For $P \in \mathcal{P}(\mathcal{X})$ and a stochastic matrix $V : \mathcal{X} \to \mathcal{Y}$, $P \times V \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ denotes the joint distribution $P(x)V(y|x)$ and $P \circ V \in \mathcal{P}(\mathcal{Y})$ denotes the y-marginal of the joint distribution $P \times V$, that is $(P \circ V)(y) = \sum_x P(x)V(y|x)$.

We denote the *type* of a sequence $x^n \in \mathcal{X}^n$ by $\hat{Q}_{x^n} \in \mathcal{P}(\mathcal{X})$. A distribution $P \in \mathcal{P}(\mathcal{X})$ is an *n-type* if $\forall x \in \mathcal{X} : nP(x) \in \mathbb{Z}$. We denote the set of n-types on $\mathcal{X}$ as $\mathcal{P}_n(\mathcal{X}) \subsetneq \mathcal{P}(\mathcal{X})$ and use the fact that $|\mathcal{P}_n(\mathcal{X})| \leq (n+1)^{|\mathcal{X}|}$ [21, Lemma 2.2] repeatedly. If $P \in \mathcal{P}_n(\mathcal{X})$, we denote the set of all sequences of type P as $\mathcal{T}_P^n \subset \mathcal{X}^n$.

---

[1]Let θ be a parameter that f and g depend on. If $f_\theta(n) \stackrel{.}{\leq} g_\theta(n)$ then, $\forall \theta : \limsup_{n \to \infty} \frac{1}{n} \log\left(\frac{f_\theta(n)}{g_\theta(n)}\right) \leq 0$ but the reverse is not true. In fact $f_\theta(n) \stackrel{.}{\leq} g_\theta(n)$ is equivalent to $\limsup_{n \to \infty} \sup_\theta \frac{1}{n} \log\left(\frac{f_\theta(n)}{g_\theta(n)}\right) \leq 0$ which is a stronger statement than the former.

The divergence between two distributions $P, Q \in \mathcal{P}(\mathcal{X})$ is defined as

$$D(P\|Q) \triangleq \sum_{x \in \mathcal{X}} P(x) \log \frac{P(x)}{Q(x)} \tag{3}$$

(here and in the sequel the bases of $\log$ and $\exp$ are arbitrary but the same). For two stochastic matrices $V \colon \mathcal{X} \to \mathcal{Y}$ and $W \colon \mathcal{X} \to \mathcal{Y}$, and $P \in \mathcal{P}(\mathcal{X})$, the conditional divergence is defined as

$$D(V\|W|P) \triangleq \sum_{x \in \mathcal{X}} P(x) \sum_{y \in \mathcal{Y}} V(y|x) \log \frac{V(y|x)}{W(y|x)}. \tag{4}$$

For $P \in \mathcal{P}(\mathcal{X})$,

$$H(P) \triangleq - \sum_{x \in \mathcal{X}} P(x) \log P(x). \tag{5}$$

For $Q \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$, $I(Q) \triangleq D(Q\|Q_X \times Q_Y)$. If $P \in \mathcal{P}(\mathcal{X})$ and $V \colon \mathcal{X} \to \mathcal{Y}$ is a stochastic matrix, $I(P, V) \triangleq I(P \times V)$ denotes the mutual information developed across the channel V with input distribution P.

III. Secrecy via Channel Resolvability

As we mentioned earlier, *channel resolvability* is a convenient and powerful tool for the analysis of secrecy [4]. The concept of resolvability dates back to Wyner [22] where he observed that, given a stationary memoryless channel $W \colon \mathcal{X} \to \mathcal{Z}$ and an input distribution P_X that induces the distribution $P_Z = P_X \circ W$ at its output, it is possible to well-approximate the product distribution P_Z^n at the output of W^n (the product channel corresponding to n independent uses of W) by transmitting a uniformly chosen codeword from a code of rate $R > I(X; Z)$. Indeed, if the code is sampled from the i.i.d. random coding ensemble, with very high probability the normalized divergence between the channel output distribution and P_Z^n can be made arbitrarily small by choosing n sufficiently large. Han and Verdú [23] and Hayashi [6] developed this theory further by replacing the measure of approximation by normalized variational (ℓ_1) distance and unnormalized divergence, respectively, and showed firstly, that the same limits on the code size hold in these cases and, secondly, that the distance between the output distribution and the target distribution P_Z^n vanishes exponentially fast as the block-length increases (the same result is derived in [10], [13] as well). In particular, in [6], [9], [10], [14], the exponential decay of the informational divergence is leveraged to establish an exponentially decaying upper bound on the information leaked to the eavesdropper in wiretap channel's model (using resolvability-based constructions of wiretap channel codes).

We can extend the notion of resolvability and ask for the approximation of arbitrary target distributions. Given a code $\mathcal{C} = \{x_1^n, \ldots, x_M^n\}$ and the channel $W \colon \mathcal{X} \to \mathcal{Z}$, denote by $P_\mathcal{C}$ the output distribution of W^n when a uniformly chosen codeword from $\mathcal{C}$ is transmitted, that is,

$$P_\mathcal{C}(z^n) = \frac{1}{M} \sum_{i=1}^{M} W^n(z^n | x_i^n). \tag{6}$$

Definition 2. Given a stationary memoryless channel $W \colon \mathcal{X} \to \mathcal{Z}$, a rate R, and a sequence of target distributions $\Phi = \{\Phi_n \in \mathcal{P}(\mathcal{Z}^n) : n \in \mathbb{N}\}$, a number $E^\Phi(W, R)$ is an achievable *resolvability exponent* over the channel W, at rate R, with respect to Φ if there exists a sequence of codes of block length n, $\mathcal{C}_n$, such that $\limsup_{n \to \infty} \frac{1}{n} \log |\mathcal{C}_n| \leq R$ and

$$\liminf_{n \to \infty} -\frac{1}{n} \log D(P_{\mathcal{C}_n} \| \Phi_n) \geq E^\Phi(W, R). \tag{7}$$

Definition 3. The supremum of all achievable resolvability exponents over $W \colon \mathcal{X} \to \mathcal{Z}$, at rate R, with respect to $\Phi = \{\Phi_n \in \mathcal{P}(\mathcal{Z}^n), n \in \mathbb{N}\}$ is called *the* resolvability exponent of the channel $W \colon \mathcal{X} \to \mathcal{Z}$ at rate R with respect to Φ.

It should be obvious that computing "the" resolvability exponent is a difficult task as it necessitates a search over all possible sequences of codes and find the best resolvability code. The usual way to circumvent such a difficulty is to use the probabilistic method and analyze the achievable exponents for the ensembles of random codes.

Definition 4. Given $\Pi = \{P_{X^n} \in \mathcal{P}(\mathcal{X}^n) : n \in \mathbb{N}\}$, a sequence of probability distributions on $\mathcal{X}^n$, an *ensemble of random codes* of rate R is a sequence of random codes $\mathcal{C}_n$ of block-length n and size $M = \exp(nR)$, obtained by sampling the codewords independently from the distribution P_{X^n}. In other words,

$$\Pr\{\mathcal{C}_n = \{x_1^n, \ldots, x_M^n\}\} = \prod_{i=1}^{M} P_{X^n}(x_i^n). \tag{8}$$

Definition 5. Given $\Pi = \{P_{X^n} \in \mathcal{P}(\mathcal{X}^n) : n \in \mathbb{N}\}$, a stationary memoryless channel $W \colon \mathcal{X} \to \mathcal{Z}$, a rate R, and a sequence of target distributions $\Phi \triangleq \{\Phi_n \in \mathcal{P}(\mathcal{Z}^n) : n \in \mathbb{N}\}$, a number $\underline{E}_s^\Phi(\Pi, W, R)$ is an achievable resolvability exponent for the ensemble of random codes of rate R defined by Π, over the channel $W \colon \mathcal{X} \to \mathcal{Z}$, with respect to the sequence of target distributions Φ if

$$\liminf_{n \to \infty} -\frac{1}{n} \log \mathbb{E}[D(P_{\mathcal{C}_n} \| \Phi_n)] \geq \underline{E}_s^\Phi(\Pi, W, R), \tag{9}$$

where $\mathcal{C}_n$ is a random code of size $M = \exp(nR)$ distributed according to (8).

Definition 6. The supremum of all achievable resolvability exponents for the random codes of rate R defined by $\Pi = \{P_{X^n} \in \mathcal{P}(\mathcal{X}^n) : n \in \mathbb{N}\}$, over the channel $W \colon \mathcal{X} \to \mathcal{Z}$, with respect to the sequence of target distribution $\Phi = \{\Phi_n \in \mathcal{P}(\mathcal{Z}^n) : n \in \mathbb{N}\}$ is called *the* resolvability exponent of the ensemble Π.

Remark. It is clear that the resolvability exponent of an ensemble equals $\liminf_{n \to \infty} -\frac{1}{n} \log \mathbb{E}[D(P_{\mathcal{C}_n} \| \Phi_n)]$. The reader may notice that this definition is somewhat conservative in the sense that, while it guarantees that for any E below the resolvability exponent of the ensemble there exists a sequence of codes $\mathcal{C}_n^\star$ (in the ensemble) and n_0 such that $\forall n > n_0 \colon D(P_{\mathcal{C}_n^\star} \| \Phi_n) \leq \exp(-nE)$, larger exponents may also be achievable; namely, for E' satisfying $\liminf_{n \to \infty} -\frac{1}{n} \log \mathbb{E}[D(P_{\mathcal{C}_n} \| \Phi_n)] < E' < \limsup_{n \to \infty} -\frac{1}{n} \log \mathbb{E}[D(P_{\mathcal{C}_n} \| \Phi_n)]$, there exists a sub-sequence of codes $\mathcal{C}_{n_1}^\star, \mathcal{C}_{n_2}^\star, \ldots$ (in the ensemble) such that $\forall i \colon D(P_{\mathcal{C}_{n_i}^\star} \| \Phi_{n_i}) \leq \exp(-n_i E')$. While this is a valid concern in general, we shall see that for the ensembles of interest, namely the ensembles of i.i.d. and constant-composition random codes, and specific sequences of target distributions,

$$\limsup_{n \to \infty} -\frac{1}{n} \log \mathbb{E}[D(P_{\mathcal{C}_n} \| \Phi_n)]$$
$$= \liminf_{n \to \infty} -\frac{1}{n} \log \mathbb{E}[D(P_{\mathcal{C}_n} \| \Phi_n)]. \tag{10}$$

In other words, the *exact* resolvability exponent for those ensembles exists. This excludes such circumstances.

Definition 7. The *exact* resolvability exponent of the ensemble of random codes of rate R defined with the sequence of distributions $\Pi = \{P_{X^n} \in \mathcal{P}(\mathcal{X}^n) : n \in \mathbb{N}\}$, over the channel $W \colon \mathcal{X} \to \mathcal{Z}$, with respect to the sequence of target distributions $\Phi = \{\Phi_n \in \mathcal{P}(\mathcal{Z}^n) : n \in \mathbb{N}\}$ is defined as

$$E_{\mathrm{s}}^{\Phi}(\Pi, W, R) \triangleq \lim_{n \to \infty} -\frac{1}{n} \log \mathbb{E}[D(P_{\mathcal{C}_n} \| \Phi_n)] \tag{11}$$

provided that the limit exists.

For the sake of completeness, let us also formally define the error exponent for an ensemble of random codes.

Definition 8. Given $\Pi = \{P_{X^n} \in \mathcal{P}(\mathcal{X}^n) : n \in \mathbb{N}\}$, a stationary memoryless channel $W \colon \mathcal{X} \to \mathcal{Y}$, and a rate R, a number $\underline{E}_{\mathrm{r}}(\Pi, W, R)$ is called an achievable *error exponent* of the ensemble Π at rate R on channel W, if

$$\liminf_{n \to \infty} -\frac{1}{n} \log \mathbb{E}[\Pr\{\hat{s}_{\mathrm{ML}}(Y^n) \neq S\}] \geq \underline{E}_{\mathrm{r}}(\Pi, W, R) \tag{12}$$

when $\mathcal{C}_n$, a random code of size $M = \exp(nR)$ is used to communicate a uniformly chosen message $S \in \{1, 2, \ldots, M\}$ via n independent uses of W, y^n is the output sequence of W^n, and $\hat{s}_{\mathrm{ML}}(y^n)$ is the ML estimation of S given y^n.

Remark. For the ensembles of interest in this paper, i.e., the ensembles of i.i.d. and constant-composition random codes the exact error exponents are well-known [21], [24], [25] (the exactness of the error exponent for constant-composition random codes follows from exponential tightness of the truncated union bound, cf. [26, Appendix A] for example).

Definition 9. Given a sequence of distributions on $\mathcal{X}^n$, $\Pi = \{P_{X^n} \in \mathcal{P}(\mathcal{X}^n) : n \in \mathbb{N}\}$, and a rate pair (R_{s}, R) a (random) *wiretap channel code* of secret message rate R_{s} is obtained by partitioning a random code of rate $R_{\mathrm{s}} + R$ in the ensemble into $M_{\mathrm{s}} = \exp(nR_{\mathrm{s}})$ sub-codes of rate R, denoted as $\mathcal{C}_n^s$, $s \in \{1, 2, \ldots, M_{\mathrm{s}}\}$, each associated to a message. To communicate the message s, the encoder transmits a codeword from the sub-code $\mathcal{C}_n^s$ uniformly at random (thus it requires an entropy rate of R).

Theorem 1. *Let $W_{\mathrm{M}} \colon \mathcal{X} \to \mathcal{Y}$ and $W_{\mathrm{E}} \colon \mathcal{X} \to \mathcal{Z}$ be the pair of legitimate receiver's and wiretapper's stationary memoryless channels respectively (see Figure 1). Fix a sequence of distributions $\Pi = \{P_{X^n} \in \mathcal{P}(\mathcal{X}^n) : n \in \mathbb{N}\}$ and an arbitrary sequence of target distributions $\Phi = \{\Phi_n \in \mathcal{P}(\mathcal{Z}^n) : n \in \mathbb{N}\}$. Let $\underline{E}_{\mathrm{r}}(\Pi, W_{\mathrm{M}}, R)$ be an achievable error exponent for the ensemble Π over the channel W_{M} (at rate R) and $E_{\mathrm{s}}^{\Phi}(\Pi, W_{\mathrm{E}}, R)$ be the *exact* resolvability exponent of the ensemble Π over the channel W_{E} with respect to the sequence of target distributions Φ (see Definition 7). Then for any rate pair (R_{s}, R) such that $E_{\mathrm{s}}^{\Phi}(\Pi, W_{\mathrm{E}}, R + R_{\mathrm{s}}) > E_{\mathrm{s}}^{\Phi}(\Pi, W_{\mathrm{E}}, R)$, using the ensemble of random wiretap codes constructed as in Definition 9, when the secret message S is uniformly distributed,*

$$\liminf_{n \to \infty} -\frac{1}{n} \log \mathbb{E}[\Pr\{\hat{s}_{\mathrm{ML}}(Y^n) \neq S\}] \geq \underline{E}_{\mathrm{r}}(\Pi, W_{\mathrm{M}}, R + R_{\mathrm{s}}) \tag{13}$$

$$\lim_{n \to \infty} -\frac{1}{n} \log \mathbb{E}[I(S; Z^n)] = E_{\mathrm{s}}^{\Phi}(\Pi, W_{\mathrm{E}}, R), \tag{14}$$

where $\hat{s}_{\mathrm{ML}}(y^n)$ is the ML estimation of the sent message given y^n, the output of legitimate receiver's channel. In other words, E_{s} is also the exact secrecy exponent *for the ensemble Π.*

Proof: That $\underline{E}_{\mathrm{r}}(\Pi, W_{\mathrm{M}}, R + R_{\mathrm{s}})$ is an achievable error exponent for the legitimate receiver is clear: probability of decoding the message S incorrectly is upper-bounded by probability of incorrect decoding of the sent codeword and the result follows. We shall, hence, only prove (14).

Since, to communicate a particular message $s \in \mathcal{S}_n$, the encoder transmits a codeword from the code $\mathcal{C}_n^s$ associated to the message s, conditioned on $S = s$ the output of W_{E}^n has distribution $P_{\mathcal{C}_n^s}$ and, since S is uniformly distributed, the *unconditional* output distribution of W_{E}^n will be $P_{\mathcal{C}_n}$ (cf. (6)). Therefore,

$$\mathbb{E}[I(S; Z^n)] = \mathbb{E}[D(P_{\mathcal{C}_n^s} \| \Phi_n | P_S)] - \mathbb{E}[D(P_{\mathcal{C}_n} \| \Phi_n)]. \tag{15}$$

Using the linearity of expectation and the fact that the sub-codes $\mathcal{C}_n^s$ are identically distributed,

$$\mathbb{E}[D(P_{\mathcal{C}_n^s} \| \Phi_n | P_S)] = \sum_{s=1}^{M_{\mathrm{s}}} P_S(s) \, \mathbb{E}[D(P_{\mathcal{C}_n^s} \| \Phi_n)]$$

$$= \mathbb{E}[D(P_{\mathcal{C}_n^1} \| \Phi_n)], \tag{16}$$

thus, by (11), we have

$$\lim_{n \to \infty} -\frac{1}{n} \log \mathbb{E}[D(P_{\mathcal{C}_n^s} \| \Phi_n | P_S)] = E_{\mathrm{s}}^{\Phi}(\Pi, W_{\mathrm{E}}, R), \tag{17}$$

$$\lim_{n \to \infty} -\frac{1}{n} \log \mathbb{E}[D(P_{\mathcal{C}_n} \| \Phi_n)] = E_{\mathrm{s}}^{\Phi}(\Pi, W_{\mathrm{E}}, R + R_{\mathrm{s}})$$

$$> E_{\mathrm{s}}^{\Phi}(\Pi, W_{\mathrm{E}}, R). \tag{18}$$

where the last inequality follows since $R_{\mathrm{s}} > 0$ and E_{s} is strictly increasing in R. Using (17) and (18) in (15) concludes the proof. ∎

Remark 1. That (a lower bound to) the resolvability exponent, lower-bounds the secrecy exponent is already used in [6], [9], [10]. Theorem 1 complements this result by showing that the exact resolvability exponent equals the exact secrecy exponent.

Remark 2. In the proof of Theorem 1, to show that $\underline{E}_{\mathrm{r}}$ is an achievable error exponent, we used a decoder which estimates the sent codeword and then decides to which sub-code it belongs. In [27] it has been shown that the error exponent of this decoder is the same as that of the *optimal* decoder (that computes the likelihood score for each message s by summing up the likelihoods of all codewords in $\mathcal{C}_n^s$ and then deciding on the most likely message) for an average code in the ensemble when the code sampling distribution P_{X^n} depends on x^n only through its type.

Remark 3. Using standard expurgation arguments it is easy to prove the existence of a sequence of wiretap codes (in the ensemble) using which

$$\liminf_{n \to \infty} -\frac{1}{n} \log \max_{s \in \mathcal{S}_n} \Pr\{\hat{s}_{\mathrm{ML}}(Y^n) \neq S | S = s\}$$

$$\geq E_{\mathrm{r}}(\Pi, W_{\mathrm{M}}, R + R_{\mathrm{s}}) \tag{19}$$

$$\liminf_{n \to \infty} -\frac{1}{n} \log \max_{s \in \mathcal{S}_n} D(P_{\mathcal{C}_n^s} \| \Phi_n)$$

$$\geq E_{\mathrm{s}}^{\Phi}(\Pi, W_{\mathrm{E}}, R) \tag{20}$$

where $\mathcal{S}_n \subseteq \{1, 2, \ldots, M_s\}$ is of cardinality at least $\frac{1}{2}M_s$. The second equality implies, using this sequence of codes, $\liminf\limits_{n \to \infty} -\frac{1}{n} \log I(S; Z^n) \geq E_s^\Phi(\Pi, W_E, R)$ *regardless of the distribution of secret messages* P_S (see [28, Appendix B] for more details). Moreover, as noted in [14], $\max_{s \in \mathcal{S}_n} D(P_{\mathcal{C}_n^s} \| \Phi_n)$ being small not only guarantees secrecy (that Eve learns very little about S by observing Z^n), but also implies *stealth*. Namely, Eve cannot even detect that Alice is sending useful messages over the channel (letting aside their content).

Remark 4. Equations (13) and (14) suggest a trade-off in code design in terms of the choice of $\Pi = \{P_{X^n} \in \mathcal{P}(\mathcal{X}^n) : n \in \mathbb{N}\}$. The sequence of input distributions Π that maximizes E_s may not coincide with the one that maximizes $\underline{E}_r$.

In light of Theorem 1 we shall focus on deriving the exact resolvability exponents for the ensembles of i.i.d. and constant-composition random codes.

IV. EXACT RESOLVABILITY EXPONENTS

A. Main Result

Theorem 2. *Let $\mathcal{C}_n$ be a random code of block-length n and rate R created by sampling $\exp(nR)$ codewords independently from the distribution $P_{X^n} \in \mathcal{P}(\mathcal{X}^n)$ (see (8)). Let $W \colon \mathcal{X} \to \mathcal{Z}$ be a discrete stationary memoryless channel and $P_{\mathcal{C}_n}$ (cf. (6)) denote the (random) output distribution of W^n when a uniformly chosen codeword from $\mathcal{C}_n$ is transmitted via n independent uses of W. Take the sequence of target distributions to be*

$$\Phi_n(z^n) \triangleq \mathbb{E}[P_{\mathcal{C}_n}(z^n)], \qquad n \in \mathbb{N} \tag{21}$$

(note that the above expectation is taken with respect to the randomness in codebook generation, thus the target distribution depends on P_{X^n}).

For any $P_X \in \mathcal{P}(\mathcal{X})$ such that $I(P_X, W) > 0$,

$$\lim_{n \to \infty} -\frac{1}{n} \log(\mathbb{E}[D(P_{\mathcal{C}_n} \| \Phi_n)])$$
$$= \begin{cases} E_s^{\text{i.i.d.}}(P_X, W, R), & \text{if } P_{X^n}(x^n) = P_X^n(x^n), \\ E_s^{\text{c.c.}}(P_X, W, R), & \text{if } P_{X^n}(x^n) = \frac{\mathbb{1}\{x^n \in \mathcal{T}_{P_X}^n\}}{|\mathcal{T}_{P_X}^n|}, \end{cases} \tag{22}$$

where

$$E_s^{\text{i.i.d.}}(P_X, W, R)$$
$$= \min_{Q \in \mathcal{P}(\mathcal{X} \times \mathcal{Z})} \big\{ D(Q \| P_X \times W) + [R - f(Q)]^+ \big\}, \tag{23a}$$

with

$$f(Q) \triangleq \sum_{(x,z) \in \mathcal{X} \times \mathcal{Z}} Q(x,z) \log \frac{W(z|x)}{(P_X \circ W)(z)}, \tag{23b}$$

and

$$E_s^{\text{c.c.}}(P_X, W, R)$$
$$= \min_{V \colon \mathcal{X} \to \mathcal{Z}} \big\{ D(V \| W | P_X) + [R - g(V)]^+ \big\} \tag{24a}$$

with

$$g(V) \triangleq \omega(V) + \min_{\substack{V': \mathcal{X} \to \mathcal{Z} \\ P_X \circ V' = P_X \circ V}} \{I(P_X, V') - \omega(V')\}, \tag{24b}$$

and

$$\omega(V) \triangleq \sum_{(x,z) \in \mathcal{X} \times \mathcal{Z}} P_X(x) V(z|x) \log W(z|x). \tag{24c}$$

Both exponents $E_s^{\text{i.i.d.}}$ and $E_s^{\text{c.c.}}$ are positive and strictly increasing in R for $R > I(P_X, W)$. Moreover, the value of $E_s^{\text{i.i.d.}}$ can be computed through

$$E_s^{\text{i.i.d.}}(P_X, W, R) = \max_{0 \le \lambda \le 1} \{\lambda R - F_0(P_X, W, \lambda)\} \tag{25a}$$

with

$$F_0(P_X, W, \lambda)$$
$$\triangleq \log \sum_{(x,z) \in \mathcal{X} \times \mathcal{Z}} P_X(x) W(z|x)^{1+\lambda} (P_X \circ W)(z)^{-\lambda}. \tag{25b}$$

Corollary 3. *The exponents $E_s^{\text{i.i.d.}}(P_X, W_E, R)$ and $E_s^{\text{c.c.}}(P_X, W_E, R)$ (of Theorem 2) are the exact secrecy exponents for the ensembles of random wiretap channel codes of rate pair (R, R_s) constructed from the ensembles of random i.i.d. and constant-composition codes, respectively, provided that $R_s > 0$ and $R > I(P_X, W_E)$.*

B. Comparison of Exponents

Corollary 3 states that the exponent $E_s^{\text{i.i.d.}}$ which was already derived in [7], [9], [10] is indeed the exact secrecy exponent for the ensemble of i.i.d. random codes (the exponent is expressed in the form of (25) in [7], [9], [10]). In contrast, it can be shown that $E_s^{\text{c.c.}}$, the exact secrecy exponent for the ensemble of constant-composition random codes, is larger than the previously-derived lower bound in [12]:

$$\underline{E}_s^{\text{c.c.}}(P_X, W_E, R) = \max_{0 \le \lambda \le 1} \{\lambda R - E_0(P_X, W_E, \lambda)\} \tag{26a}$$

with

$$E_0(P_X, W, \lambda)$$
$$\triangleq \log \sum_{z \in \mathcal{Z}} \left(\sum_{x \in \mathcal{X}} P_X(x) W(z|x)^{\frac{1}{1-\lambda}} \right)^{1-\lambda}, \tag{26b}$$

(note that the function E_0 in (26b) is essentially Gallager's E_0 [24]). For every discrete memoryless stationary channel $W: \mathcal{X} \to \mathcal{Z}$,

$$E_s^{\text{c.c.}}(P_X, W, R) \ge \underline{E}_s^{\text{c.c.}}(P_X, W, R). \tag{27}$$

This follows from the fact that $g(V) \le I(P_X, V)$ using similar steps as in [21, Problem 10.24] to derive Gallager-style expressions of error exponents (see Appendix A-D for a complete proof).

As for comparing the secrecy exponents $E_s^{\text{i.i.d.}}$ and $E_s^{\text{c.c.}}$, numerical examples show that in general there is no ordering between them. In particular, as shown in Figures 2 and 3, for binary symmetric channel and binary erasure channel the ensemble of constant-composition codes leads to a larger exponent than the ensemble of i.i.d. random

codes. The two exponents are equal when the input distribution is uniform. On the other side, in Figures 4 and 5 we see that for asymmetric channels (Z-channel and binary asymmetric channel) the ensemble of constant-composition random codes results in a smaller secrecy exponent compared to the ensemble of i.i.d. random codes.
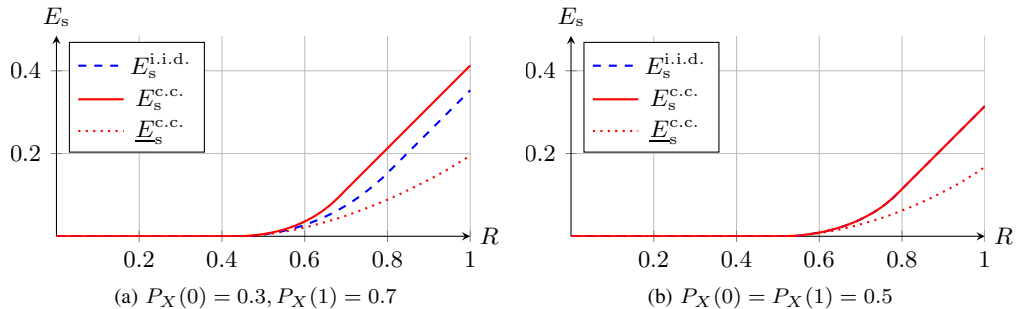

(a) $P_X(0) = 0.3, P_X(1) = 0.7$

(b) $P_X(0) = P_X(1) = 0.5$

Fig. 2. Comparison of secrecy exponents for Binary Symmetric Channel with crossover probability 0.11

(a) $P_X(0) = 0.28, P_X(1) = 0.72$

(b) $P_X(0) = P_X(1) = 0.5$
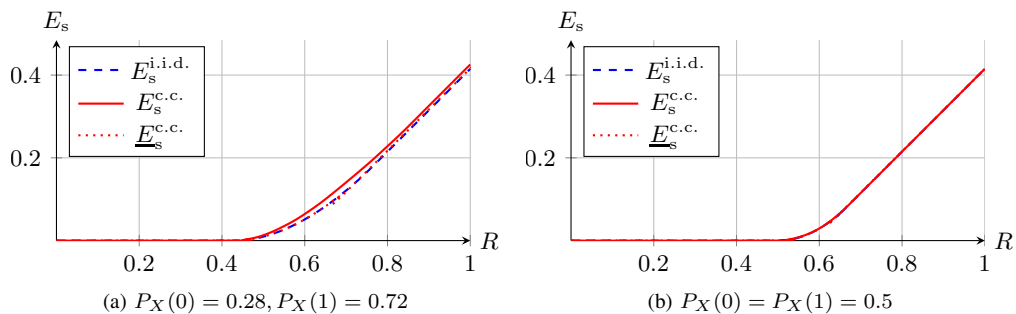
Fig. 3. Comparison of secrecy exponents for Binary Erasure Channel with erasure probability 0.5

(a) $P_X(0) = 0.36, P_X(1) = 0.64$

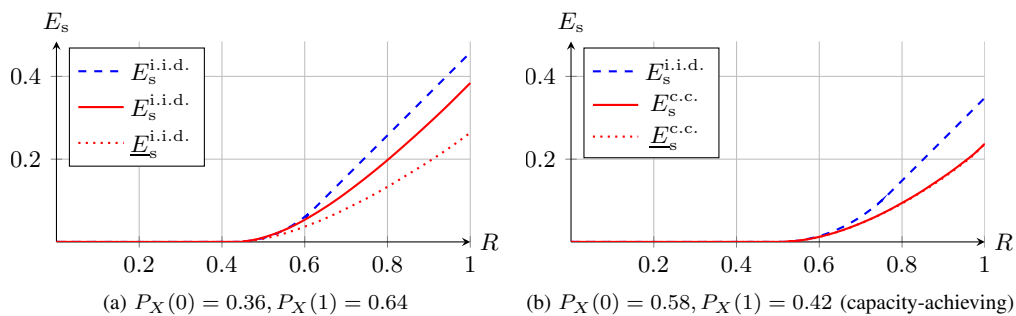(b) $P_X(0) = 0.58, P_X(1) = 0.42$ (capacity-achieving)

Fig. 4. Comparison of secrecy exponents for Z-channel with $W_{\mathrm{E}}(0|1) = 0.303$

The reader may find details on how the exponents are computed in Appendix B.

V. PROOF OF THEOREM 2

In this section we fix P_X and set $P_{XZ}(x, z) = P_X(x)W(z|x)$. Moreover we assume, without essential loss of generality, that (i) $\mathrm{supp}(P_X) = \mathcal{X}$ and (ii) for every $z \in \mathcal{Z}$, there exists at least one $x \in \mathcal{X}$ such that $W(z|x) > 0$.

(a) $P_X(0) = 0.42, P_X(1) = 0.58$ (b) $P_X(0) = 0.57, P_X(1) = 0.43$ (capacity-achieving)
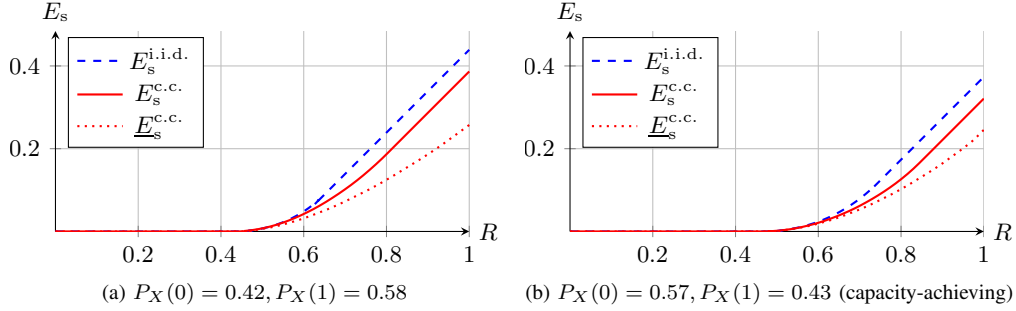
Fig. 5. Comparison of secrecy exponents for binary asymmetric channel with $W_E(1|0) = 0.01$, $W_E(0|1) = 0.303$

Recall that the setting we are considering is as follows: A random code $\mathcal{C}_n = \{X_1^n, \ldots, X_M^n\}$ of block-length n and size $M = \exp(nR)$ is generated by sampling each codeword independently from distribution P_{X^n}. A uniformly chosen codeword from this code is transmitted through the product channel W^n and the (random) distribution of its output sequence is as in (6).

Note that $P_{\mathcal{C}_n}(z^n)$ is the average of M i.i.d. random variables $W^n(z^n|X_i^n)$, $i = 1, \ldots, M$ and, hence, is naturally expected to concentrate around its mean, which is exactly the target distribution $\Phi_n(z^n)$. To prove Theorem 2 we analyze the deviations of the i.i.d. average $P_{\mathcal{C}_n}(z^n)$ from its mean for every $z^n \in \mathcal{Z}^n$.

Lemma 4. *Let Φ_n be as defined in* (21). *Then,*

(i) $P_{\mathcal{C}_n} \ll \Phi_n$ *with probability* 1.

(ii) *For both choices of P_{X^n} in* (22), $\forall z^n \in \mathrm{supp}(\Phi_n)$: $\Phi_n(z^n) > (1/\alpha)^n$ *where $\alpha > 1$ is a constant that only depends on P_X and W.*

    Proof: See Appendix A-E. ∎

Remark. While when $P_{X^n} = P_X^n$, $\Phi_n = P_Z^n$ and, hence, $\mathrm{supp}(\Phi_n) = \mathcal{Z}^n$, when P_{X^n} is the uniform distribution over the type-class $\mathcal{T}_{P_X}^n$ the support of Φ_n need not to be $\mathcal{Z}^n$. For instance, consider a binary erasure channel and P_X being uniform distribution on $\{0, 1\}$. Then Φ_n puts no mass on all-zero (and by symmetry all-one) output sequences.

Let

$$L(z^n) \triangleq \begin{cases} \frac{P_{\mathcal{C}_n}(z^n)}{\Phi_n(z^n)} & \text{if } \Phi_n(z^n) > 0, \\ 1 & \text{otherwise,} \end{cases} \tag{28}$$

denote the (random) likelihood ratio of each sequence $z^n \in \mathcal{Z}^n$. By construction,

$$\mathbb{E}[L(z^n)] = 1, \qquad \forall z^n \in \mathcal{Z}^n. \tag{29}$$

Using the linearity of expectation we have,

$$\mathbb{E}[D(P_{\mathcal{C}_n} \| \Phi_n)] = \mathbb{E}\left[\sum_{z^n \in \mathcal{Z}^n} P_{\mathcal{C}_n}(z^n) \log\left(\frac{P_{\mathcal{C}_n}(z^n)}{\Phi_n(z^n)} \right) \right] \tag{30}$$

$$= \sum_{z^n \in \mathcal{Z}^n} \mathbb{E}\left[P_{\mathcal{C}_n}(z^n) \log\left(\frac{P_{\mathcal{C}_n}(z^n)}{\Phi_n(z^n)}\right)\right] \tag{31}$$

$$= \sum_{z^n \in \mathcal{Z}^n} \Phi_n(z^n) \mathbb{E}[L(z^n) \log L(z^n)] \tag{32}$$

$$= \sum_{P \in \mathcal{P}_n(\mathcal{Z})} \sum_{z^n \in \mathcal{T}_P^n} \Phi_n(z^n) \mathbb{E}[L(z^n) \log (L(z^n))]. \tag{33}$$

For convenience, let us define

$$\mathcal{Q}_n \triangleq \left\{Q \in \mathcal{P}_n(\mathcal{X} \times \mathcal{Z}) : P_{X^n}(\mathcal{T}_{Q_X}^n) > 0\right\} \tag{34}$$

$$\mathcal{Q} \triangleq \left\{Q \in \mathcal{P}(\mathcal{X} \times \mathcal{Z}) : P_{X^n}(\mathcal{T}_{Q_X}^n) > 0\right\} \tag{35}$$

as the set of all *feasible* joint n-types and joint distributions, respectively.[2]

Theorem 2 follows as a corollary to Theorem 5.

Theorem 5. *For any $Q \in \mathcal{Q}_n$ let*

$$g_n(Q) \triangleq \omega(Q)$$

$$+ \min_{Q' \in \mathcal{Q}_n[Q_Z]} \{I(Q') + D(Q'_X \| P_X) - \omega(Q')\}, \tag{36}$$

where

$$\mathcal{Q}_n[Q_Z] \triangleq \left\{Q' \in \mathcal{Q}_n : Q'_Z = Q_Z\right\}, \quad \forall Q_Z \in \mathcal{P}_n(\mathcal{Z}), \tag{37}$$

and

$$\omega(Q) \triangleq \sum_{x,z} Q(x,z) \log W(z|x). \tag{38}$$

Then, $\forall z^n \in \mathcal{Z}^n$:

$$\Phi_n(z^n)\left(\mathbb{E}[L(z^n) \log L(z^n)] + \frac{\log(\mathrm{e})}{M}\right)$$

$$\doteq \exp\left(-n[\min\{E_1(\hat{Q}_{z^n}), E_2(\hat{Q}_{z^n})\} + H(\hat{Q}_{z^n})]\right) \tag{39}$$

where

$$E_1(Q_Z) = \min_{\substack{Q' \in \mathcal{Q}_n[Q_Z]: \\ g_n(Q') \le R + \delta_n}} \left\{D(Q' \| P_{XZ}) + R - g_n(Q')\right\}, \tag{40}$$

$$E_2(Q_Z) = \min_{\substack{Q' \in \mathcal{Q}_n[Q_Z]: \\ g_n(Q') > R + \delta_n}} D(Q' \| P_{XZ}), \tag{41}$$

and

$$\delta_n \triangleq \frac{2\log(\mathrm{e}) + 2|\mathcal{X}||\mathcal{Z}|\log(n+1)}{n}. \tag{42}$$

---

[2]More simply, $\mathcal{Q}_n = \mathcal{P}_n(\mathcal{X} \times \mathcal{Z})$ (respectively $\mathcal{Q} = \mathcal{P}(\mathcal{X} \times \mathcal{Z})$) for the i.i.d. random coding ensemble (since $\operatorname{supp}(P_X) = \mathcal{X}$), and $\mathcal{Q}_n = \{Q \in \mathcal{P}_n(\mathcal{X} \times \mathcal{Z}) : Q_X = P_X\}$ (respectively $\mathcal{Q} = \{Q \in \mathcal{P}(\mathcal{X} \times \mathcal{Z}) : Q_X = P_X\}$) for the ensemble of random constant-composition codes.

Proof of Theorem 2: Plugging (39) into (33) we get

$$\mathbb{E}[D(P_{\mathcal{C}_n}\|\Phi_n)] + \log(\mathrm{e})\frac{1}{M}$$

$$\doteq \exp\big(-n \min_{Q_Z \in \mathcal{P}_n(\mathcal{Z})} \min\{E_1(Q_Z), E_2(Q_Z)\}\big). \tag{43}$$

Moreover, since $\lim_{n\to\infty} \delta_n = 0$ and the sets of n-types are dense,

$$\lim_{n\to\infty} \min_{Q_Z \in \mathcal{P}_n(\mathcal{Z})} \min\{E_1(Q_Z), E_2(Q_Z)\}$$

$$= \min_{Q \in \mathcal{Q}}\{D(Q\|P_{XZ}) + [R - g_\star(Q)]^+\} \triangleq E_{\mathrm{s}}(P_X, W, R), \tag{44}$$

where

$$g_\star(Q) \triangleq \omega(Q)$$

$$+ \min_{\substack{Q' \in \mathcal{Q} \\ Q'_Z = Q_Z}} \{I(Q') + D(Q'_X\|P_X) - \omega(Q')\}. \tag{45}$$

It can be shown that $g_\star(P_{XZ}) = I(X; Z)$ (see (ii) of Lemma 7 in Appendix A-A). Consequently, $E_{\mathrm{s}}(P_X, W, R) \leq [R - I(X; Z)]^+ < R$ when $I(X; Z) > 0$. Using this observation in (43) shows

$$\lim_{n\to\infty} -\frac{1}{n} \log \mathbb{E}[D(P_{\mathcal{C}_n}\|\Phi_n)] = E_{\mathrm{s}}(P_X, W, R). \tag{46}$$

Also, $g_\star(P_{XZ}) = I(X; Z)$ implies E_{s}, as defined in (44), is zero for $R \leq I(X; Z)$ and strictly positive for $R > I(X; Z)$ as the objective function of (44) is the sum of two non-negative functions of Q, and is zero iff both are zero (i.e., iff $Q = P_{XZ}$ and $R \leq I(X; Z)$).

For the ensemble of i.i.d. random codes, it can be verified that $g_\star(Q) = f(Q)$ defined in (23b) (see (i) in Lemma 7 in Appendix A-A). The equivalence of (23) and (25) is shown in Appendix A-B.

Similarly, for the ensemble of constant-composition random codes, any $Q \in \mathcal{Q}$ must be of the form $P_X \times V$ for some stochastic matrix $V: \mathcal{X} \to \mathcal{Z}$ which reduces the exponent to (24).

That the exponents $E_{\mathrm{s}}^{\mathrm{i.i.d.}}$ and $E_{\mathrm{s}}^{\mathrm{c.c.}}$ are strictly increasing in R is proved in Appendix A-C. ∎

It remains to prove Theorem 5. For this we shall use the following auxiliary lemma which is proved in Appendix A-F.

Lemma 6. *Let A be an arbitrary* non-negative *random variable. Then, for any $\theta > 0$,*

$$c(\theta)\Big[\frac{\mathrm{var}(A)}{\mathbb{E}[A]} - \tau_\theta(A)\Big] \leq \mathbb{E}[A\ln(A/\mathbb{E}[A])] \leq \frac{\mathrm{var}(A)}{\mathbb{E}[A]} \tag{47}$$

where

$$\tau_\theta(A) \triangleq \mathbb{E}[A]\Big[\theta^2 \Pr\{A > (\theta + 1)\mathbb{E}[A]\}$$

$$+ 2\int_\theta^{+\infty} v \Pr\{A > (v + 1)\mathbb{E}[A]\}\,\mathrm{d}\,v\Big], \tag{48}$$

and

$$c(\theta) \triangleq \frac{(1 + \theta)\ln(1 + \theta) - \theta}{\theta^2}. \tag{49}$$

Remark. It follows from Jensen's inequality that $\mathbb{E}[A \ln(A/\mathbb{E}[A])]) \geq 0$. Lemma 6 improves this lower bound for random variables with sufficiently small tails.

Proof of Theorem 5: Assume hereafter that $z^n \in \mathcal{Z}^n$ is fixed. We, firstly, have

$$\Phi_n(z^n) = \sum_{x^n \in \mathcal{X}^n} P_{X^n}(x^n) W^n(z^n|x^n) \tag{50}$$

$$= \sum_{x^n \in \mathcal{X}^n} P_{X^n}(x^n) \sum_{Q \in \mathcal{Q}_n[\hat{Q}_{z^n}]} \mathbb{1}\{(x^n, z^n) \in \mathcal{T}_Q^n\} \exp(n\omega(Q)) \tag{51}$$

$$= \sum_{Q \in \mathcal{Q}_n[\hat{Q}_{z^n}]} \underbrace{\left\{ \sum_{x^n \in \mathcal{X}^n} P_{X^n}(x^n) \mathbb{1}\{(x^n, z^n) \in \mathcal{T}_Q^n\} \right\}}_{\triangleq p_Q} \exp(n\omega(Q)) \tag{52}$$

where $\omega(Q)$ is defined in (38). It is clear that $\sum_{Q \in \mathcal{Q}_n[\hat{Q}_{z^n}]} p_Q = 1$ (our notation is somewhat imprecise because p_Q depends on z^n through its type; but, as we have fixed z^n throughout the proof, we avoid explicitly showing this dependence for the sake of brevity). It can also be shown (see Appendix A-G) that for any distribution P_{X^n} that depends on x^n only though its type—including our cases of interest,

$$p_Q = \frac{|\mathcal{T}_Q^n|}{|\mathcal{T}_{Q_Z}^n||\mathcal{T}_{Q_X}^n|} P_{X^n}(\mathcal{T}_{Q_X}^n), \qquad \forall Q \in \mathcal{Q}_n[\hat{Q}_{z^n}]. \tag{53}$$

For both ensembles of i.i.d. and constant-composition random codes, $P_{X^n}(\mathcal{T}_{Q_X}^n) \doteq \exp(-nD(Q_X\|P_X))$, for $Q \in \mathcal{Q}_n$, thus

$$p_Q \doteq \exp\big(-n[I(Q) + D(Q_X\|P_X)]\big). \tag{54}$$

Combining the exponent in (54) and $\omega(Q)$, we have

$$\Phi_n(z^n) \doteq \exp\Big(-n \min_{Q \in \mathcal{Q}_n[\hat{Q}_{z^n}]} \{I(Q) + D(Q_X\|P_X) - \omega(Q)\}\Big) \tag{55}$$

$$= \exp\Big(-n\Big[\min_{Q \in \mathcal{Q}_n[\hat{Q}_{z^n}]} D(Q\|P_{XZ}) + H(\hat{Q}_{z^n})\Big]\Big) \tag{56}$$

Note that if $\Phi_n(z^n) = 0$, then the exponent of the above is infinity which means $\min_{Q \in \mathcal{Q}_n[\hat{Q}_{z^n}]} D(Q\|P_{XZ}) = +\infty$. This implies both exponents $E_1(\hat{Q}_{z^n})$ and $E_2(\hat{Q}_{z^n})$ (see (40) and (41)) are infinity and (39) holds. Therefore, we shall restrict our attention to the non-trivial case when $z^n \in \text{supp}(\Phi_n)$.

Using the type-enumeration method [27], [29] we have

$$P_{\mathcal{C}_n}(z^n) = \frac{1}{M} \sum_{i=1}^{M} W^n(z^n|X_i^n) \tag{57}$$

$$= \frac{1}{M} \sum_{Q \in \mathcal{Q}_n[\hat{Q}_{z^n}]} N_Q \exp\big(n\omega(Q)\big) \tag{58}$$

where

$$N_Q \triangleq \big|\{x^n \in \mathcal{C}_n : (x^n, z^n) \in \mathcal{T}_Q^n\}\big| \tag{59}$$

is the number of codewords in $\mathcal{C}_n$ that have joint type Q with z^n and $\omega(Q)$ is defined in (38). The collection $\{N_Q : Q \in \mathcal{Q}_n[\hat{Q}_{z^n}]\}$ has a multinomial distribution with cluster size M and success probabilities $\{p_Q : Q \in \mathcal{Q}_n[\hat{Q}_{z^n}]\}$ (defined in (53)). Since $z^n \in \text{supp}(\Phi_n)$,

$$L(z^n) = \frac{P_{\mathcal{C}_n}(z^n)}{\Phi_n(z^n)} = \frac{1}{M} \sum_{Q \in \mathcal{Q}_n[\hat{Q}_{z^n}]} N_Q \ell(Q), \tag{60}$$

where we have defined

$$\ell(Q) \triangleq \frac{\exp\big(n\omega(Q)\big)}{\Phi_n(z^n)}. \tag{61}$$

Using (55) we have

$$\ell(Q) \doteq \exp\big(ng_n(Q)\big), \tag{62}$$

with $g_n(Q)$ defined in (36). It also can be verified (see Appendix A-H) that explicit bounds on $\ell(Q)$ are

$$(n+1)^{-2|\mathcal{X}||\mathcal{Z}|} \exp\big(ng_n(Q)\big) \le \ell(Q)$$
$$\le (n+1)^{|\mathcal{X}||\mathcal{Z}|} \exp\big(ng_n(Q)\big). \tag{63}$$

Using the elementary properties of the multinomial distribution, it can be checked (see Appendix A-I) that if

$$A \triangleq \frac{1}{M} \sum_{Q \in \mathcal{A}} N_Q \ell(Q) \tag{64}$$

for some $\mathcal{A} \subseteq \mathcal{Q}_n[\hat{Q}_{z^n}]$, then

$$\mathbb{E}[A] = \sum_{Q \in \mathcal{A}} p_Q \ell(Q), \tag{65a}$$

$$\text{var}(A) = \frac{1}{M} \sum_{Q \in \mathcal{A}} p_Q \ell(Q)^2 - \frac{1}{M}\big(\mathbb{E}[A]\big)^2. \tag{65b}$$

Partition $\mathcal{Q}_n[\hat{Q}_{z^n}]$ as

$$\mathcal{Q}' \triangleq \{Q \in \mathcal{Q}_n[\hat{Q}_{z^n}] : g_n(Q) \le R + \delta_n\}, \tag{66}$$

$$\mathcal{Q}'' \triangleq \{Q \in \mathcal{Q}_n[\hat{Q}_{z^n}] : g_n(Q) > R + \delta_n\}, \tag{67}$$

(with δ_n defined as in (42)), and split the sum in (60) as

$$L(z^n) = \underbrace{\frac{1}{M} \sum_{Q \in \mathcal{Q}'} \ell(Q) N_Q}_{\triangleq L_1} + \underbrace{\frac{1}{M} \sum_{Q \in \mathcal{Q}''} \ell(Q) N_Q}_{\triangleq L_2}. \tag{68}$$

Using (65a) we have

$$\mathbb{E}[L_1] = \sum_{Q \in \mathcal{Q}'} \ell(Q) p_Q \triangleq \mu_1, \tag{69}$$

$$\mathbb{E}[L_2] = \sum_{Q \in \mathcal{Q}''} \ell(Q) p_Q \triangleq \mu_2. \tag{70}$$

Moreover, using (65b) we have

$$\text{var}(L_1) + \frac{1}{M}\mu_1^2 = \frac{1}{M} \sum_{Q \in \mathcal{Q}'} \ell(Q)^2 p_Q \triangleq \nu_1. \tag{71}$$

One can check (using the upper bound of (63)) that the choice of $\mathcal{Q}'$ implies

$$\nu_1 \stackrel{.}{\leq} \mu_1. \tag{72}$$

For non-negative l_1 and l_2, and $l = l_1 + l_2$,

$$l \ln(l) = l_1 \ln(l) + l_2 \ln(l) \tag{73}$$

$$= l_1 \ln(l_1) + l_1 \ln(1 + l_2/l_1) + l_2 \ln(l) \tag{74}$$

$$\leq l_1 \ln(l_1) + l_2(1 + \ln(l)) \tag{75}$$

(since $\ln(1 + l_2/l_1) \leq l_2/l_1$), thus,

$$\mathbb{E}[L(z^n) \log L(z^n)] = \log(e) \, \mathbb{E}[L(z^n) \ln L(z^n)] \tag{76}$$

$$\leq \log(e) \big(\mathbb{E}[L_1 \ln(L_1)] + \mathbb{E}[L_2(1 + \ln L(z^n))] \big) \tag{77}$$

$$\stackrel{(*)}{\leq} \log(e) \, \mathbb{E}[L_1 \ln(L_1)] + \log(e)(1 + n \ln \alpha) \, \mathbb{E}[L_2] \tag{78}$$

where $(*)$ follows from (ii) in Lemma 4 (as $L(z^n) \leq 1/\Phi_n(z^n)$). The upper bound of (47) implies

$$\mathbb{E}[L_1 \ln(L_1)] \leq \mu_1 \ln(\mu_1) + \frac{\text{var}(L_1)}{\mu_1} \stackrel{(*)}{\leq} \frac{\text{var}(L_1)}{\mu_1} \tag{79}$$

where $(*)$ follows since $\mu_1 \leq 1$. Moreover, using (71) and the fact that $\mu_1 + \mu_2 = 1$ we have

$$\frac{\text{var}(L_1)}{\mu_1} = \frac{\nu_1}{\mu_1} - \frac{\mu_1}{M} \tag{80}$$

$$= \nu_1 \frac{\mu_1 + \mu_2}{\mu_1} - \frac{1 - \mu_2}{M} \tag{81}$$

$$= \nu_1 + \mu_2 \Big(\frac{\nu_1}{\mu_1} + \frac{1}{M} \Big) - \frac{1}{M}. \tag{82}$$

Using (82) in (79) we have

$$\mathbb{E}[L_1 \ln(L_1)] + \frac{1}{M} \leq \nu_1 + \Big(\frac{\nu_1}{\mu_1} + \frac{1}{M} \Big) \mu_2 \tag{83}$$

$$\stackrel{.}{\leq} \nu_1 + \mu_2 \tag{84}$$

where the last inequality follows from (72) and the fact that $M \geq 1$.

Using (84) and (70) (and noting that $\alpha \geq 1$ only depends on P_X and W) we can further upper-bound (78) as

$$\mathbb{E}[L(z^n) \log L(z^n)] + \log(e) \frac{1}{M} \stackrel{.}{\leq} \nu_1 + \mu_2. \tag{85}$$

Using the same reasoning as that we used for deriving (55) we have

$$\Phi_n(z^n) \mu_2 = \sum_{Q \in \mathcal{Q}''} p_Q \exp\big(n\omega(Q)\big) \tag{86}$$

$$\doteq \exp\big(-n[E_2(\hat{Q}_{z^n}) + H(\hat{Q}_{z^n})]\big). \tag{87}$$

Furthermore, using (62), we have

$$\frac{1}{M} \ell(Q)^2 p_Q \doteq \exp\big(n[g_n(Q) - R]\big) \ell(Q) p_Q \tag{88}$$

which implies

$$\Phi_n(z^n)\nu_1 \doteq \exp\big(-n[E_1(\hat{Q}_{z^n}) + H(\hat{Q}_{z^n})]\big). \tag{89}$$

Plugging the (upper) bounds of (87) and (89) into (85) we get

$$\Phi_n(z^n)\Big(\mathbb{E}[L(z^n)\log L(z^n)] + \log(\mathrm{e})\frac{1}{M}\Big)$$
$$\dot{\leq} \exp\big(-n[\min\{E_1(\hat{Q}_{z^n}), E_2(\hat{Q}_{z^n})\} + H(\hat{Q}_{z^n})]\big). \tag{90}$$

Now we shall establish the lower bound counterpart of (90) to complete the proof. The choice of $\mathcal{Q}'$ implies

$$\Pr\big\{L_2 \in (0, \mathrm{e}^2)\big\} = 0. \tag{91}$$

This holds since the lower bound of (63), together with the choice of $\mathcal{Q}''$ in (67) imply $\forall Q \in \mathcal{Q}' : \ell(Q) \geq \mathrm{e}^2 \exp(nR)$. Therefore, either $\forall Q \in \mathcal{Q}' : N_Q = 0$ which implies $L_2 = 0$ or $\exists Q_0 \in \mathcal{Q}''$ such that $N_{Q_0} \geq 1$, in which case,

$$L_2 \geq \frac{1}{M}\ell(Q_0)N_{Q_0} \geq \frac{1}{M}\ell(Q_0) \geq \mathrm{e}^2. \tag{92}$$

Equation (91) implies,

$$\mathbb{E}[L_2\ln(L_2)] = \sum_{l\geq\mathrm{e}^2} l\ln(l)\Pr\{L_2 = l\} \tag{93}$$

$$\geq \ln(\mathrm{e}^2)\sum_{l\geq\mathrm{e}^2} l\Pr\{L_2 = l\} = 2\,\mathbb{E}[L_2]. \tag{94}$$

For positive l_1 and l_2, and $l = l_1 + l_2 \geq \max\{l_1, l_2\}$,

$$l\ln(l) = l_1\ln(l) + l_2\ln(l) \tag{95}$$

$$\geq l_1\ln(l_1) + l_2\ln(l_2). \tag{96}$$

Therefore,

$$\mathbb{E}[L(z^n)\ln L(z^n)] \geq \mathbb{E}[L_1\ln(L_1)] + \mathbb{E}[L_2\ln(L_2)]. \tag{97}$$

Using the lower bound of (47) (with $\tau_\theta(L_1)$ and $c(\theta)$ defined as in (48) and (49) respectively), $\forall \theta > 0$:

$$\mathbb{E}[L_1\ln(L_1)] \geq \mathbb{E}[L_1]\ln(\mathbb{E}[L_1]) + c(\theta)\Big[\frac{\mathrm{var}(L_1)}{\mathbb{E}[L_1]} - \tau_\theta(L_1)\Big] \tag{98}$$

$$\overset{(a)}{=} (1 - \mathbb{E}[L_2])\ln(1 - \mathbb{E}[L_2]) + c(\theta)\Big[\frac{\mathrm{var}(L_1)}{\mathbb{E}[L_1]} - \tau_\theta(L_1)\Big] \tag{99}$$

$$\overset{(b)}{\geq} -\mathbb{E}[L_2] + c(\theta)\Big[\frac{\mathrm{var}(L_1)}{\mathbb{E}[L_1]} - \tau_\theta(L_1)\Big]. \tag{100}$$

In the above (a) follows since $\mathbb{E}[L_1] = 1 - \mathbb{E}[L_2]$ and (b) since $(1 - \zeta)\ln(1 - \zeta) \geq -\zeta$. Using (94) and (100) in (96) shows that $\forall \theta > 0$:

$$\mathbb{E}[L(z^n)\ln L(z^n)] \geq c(\theta)\Big[\frac{\mathrm{var}(L_1)}{\mathbb{E}[L_1]} - \tau_\theta(L_1)\Big] + \mathbb{E}[L_2]. \tag{101}$$

Now we shall upper-bound $\tau_\theta(L_1)$. Starting by bounding the tail of L_1 we have

$$\Pr\{L_1 \geq (v+1)\,\mathbb{E}[L_1]\}$$

$$= \Pr\left\{\sum_{Q \in \mathcal{Q}'} \ell(Q)(N_Q - Mp_Q) \geq Mv\,\mathbb{E}[L_1]\right\} \tag{102}$$

$$\leq \Pr\left\{\bigcup_{Q \in \mathcal{Q}'} \left\{\ell(Q)(N_Q - Mp_Q) \geq \frac{Mv\,\mathbb{E}[L_1]}{|\mathcal{Q}'|}\right\}\right\} \tag{103}$$

$$\overset{(a)}{\leq} \sum_{Q \in \mathcal{Q}'} \Pr\left\{\ell(Q)(N_Q - Mp_Q) \geq \frac{Mv\,\mathbb{E}[L_1]}{|\mathcal{Q}'|}\right\} \tag{104}$$

$$\overset{(b)}{\leq} \sum_{Q \in \mathcal{Q}'} \frac{\mathbb{E}[\ell(Q)^4(N_Q - Mp_Q)^4]}{(Mv\,\mathbb{E}[L_1]/|\mathcal{Q}'|)^4} \tag{105}$$

$$= \frac{|\mathcal{Q}'|^4}{v^4(\mathbb{E}[L_1])^4} \frac{1}{M^4} \sum_{Q \in \mathcal{Q}'} \ell(Q)^4\,\mathbb{E}[(N_Q - Mp_Q)^4], \tag{106}$$

where (a) is the union bound and (b) follows by Markov inequality. For $N \sim \text{Binomial}(M, p)$,

$$\mathbb{E}[(N - Mp)^4] = Mp(1-p)[1 + 3(M-2)p(1-p)] \tag{107}$$

$$\leq \text{var}(N) + 3\,\text{var}(N)^2. \tag{108}$$

Continuing (106) we have

$$\frac{1}{M^4} \sum_{Q \in \mathcal{Q}'} \ell(Q)^4\,\mathbb{E}[(N_Q - Mp_Q)^4]$$

$$\leq \frac{1}{M^4} \sum_{Q \in \mathcal{Q}'} \ell(Q)^4\big(\text{var}(N_Q) + 3\,\text{var}(N_Q)^2\big) \tag{109}$$

$$\overset{(a)}{\dot{\leq}} \frac{1}{M^2} \sum_{Q \in \mathcal{Q}'} \ell(Q)^2\,\text{var}(N_Q) + 3\frac{1}{M^4} \sum_{Q \in \mathcal{Q}'} \ell(Q)^4\,\text{var}(N_Q)^2 \tag{110}$$

$$\overset{(b)}{\leq} \frac{1}{M^2} \sum_{Q \in \mathcal{Q}'} \ell(Q)^2\,\text{var}(N_Q)$$

$$+ 3\Big[\frac{1}{M^2} \sum_{Q \in \mathcal{Q}'} \ell(Q)^2\,\text{var}(N_Q)\Big]^2 \tag{111}$$

$$\overset{(c)}{\leq} \nu_1 + 3\nu_1{}^2 \overset{(d)}{\dot{=}} \nu_1, \tag{112}$$

where (a) follows since $\ell(Q) \leq \exp(n\delta_n)M \dot{=} M$ for $Q \in \mathcal{Q}'$, (b) since for positive summands, the sum of the squares is less than the square of the sums, (c) since $\text{var}(N_Q) \leq Mp_Q$, and (d) since $\nu_1 \overset{\cdot}{\leq} \mu_1 \leq 1$ (see (72)). Plugging (112) into (106) we get

$$\Pr\{L_1 \geq (v+1)\mathbb{E}[L_1]\} \overset{\cdot}{\leq} \frac{|\mathcal{Q}'|^4 \nu_1}{(\mathbb{E}[L_1])^4} \frac{1}{v^4} \tag{113}$$

Using the above in (48) we get

$$\tau_\theta(L_1) = \mathbb{E}[L_1]\Big[\theta^2 \Pr\{L_1 > (\theta+1)\mathbb{E}[L_1]\}$$

$$+ 2\int_\theta^{+\infty} v \Pr\{L_1 > (v+1)\mathbb{E}[L_1]\}\, \mathrm{d}\, v\Big] \tag{114}$$

$$\dot{\le} \mathbb{E}[L_1]\Big[\frac{\theta^2}{\theta^4} + 2\int_\theta^{+\infty} \frac{v}{v^4}\, \mathrm{d}\, v\Big]\frac{|\mathcal{Q}'|^4}{\mathbb{E}[L_1]^4}\nu_1 \tag{115}$$

$$\doteq \frac{\nu_1}{\mu_1{}^3}\frac{|\mathcal{Q}'|^4}{\theta^2}. \tag{116}$$

Since (116) implies $\tau_\theta(L_1) \le d(n)\frac{|\mathcal{Q}'|^4\nu_1}{\theta^2\mu_1^3}$ for some sub-exponentially increasing sequence $d(n)$ (which only depends on $|\mathcal{X}|$ and $|\mathcal{Z}|$), taking

$$\theta_n \triangleq 2\sqrt{d(n)}\frac{|\mathcal{Q}'|^2}{\mu_1}, \tag{117}$$

we will have

$$\tau_{\theta_n}(L_1) \le \frac{1}{4}\frac{\nu_1}{\mu_1}. \tag{118}$$

Using (71) and (118) in (101) we have

$$\mathbb{E}[L(z^n)\ln L(z^n)] \ge c(\theta_n)\Big[\frac{\mathrm{var}(L_1)}{\mathbb{E}[L_1]} - \tau_{\theta_n}(L_1)\Big] + \mathbb{E}[L_2] \tag{119}$$

$$\ge c(\theta_n)\Big[\frac{\nu_1}{\mu_1} - \frac{1}{M}\mu_1 - \frac{1}{4}\frac{\nu_1}{\mu_1}\Big] + \mathbb{E}[L_2] \tag{120}$$

$$\overset{(*)}{\ge} c(\theta_n)\Big[\frac{3}{4}\frac{\nu_1}{\mu_1} - \frac{1}{M}\Big] + \mathbb{E}[L_2] \tag{121}$$

(where $(*)$ follows because $\mu_1 \le 1$). Since (for $\theta > 0$), $c(\theta) \le c(0) = \frac{1}{2} < 1$, we can further lower-bound (121) as

$$\mathbb{E}[L(z^n)\ln L(z^n)] \ge \frac{3}{4}c(\theta_n)\frac{\nu_1}{\mu_1} + \mathbb{E}[L_2] - \frac{1}{M} \tag{122}$$

Moreover,

$$c(\theta_n) = \frac{1}{\theta_n}\frac{(1+\theta_n)\ln(1+\theta_n) - \theta_n}{\theta_n} \tag{123}$$

$$\overset{(a)}{\ge} \frac{1}{\theta_n}\frac{(1+\mu_1\theta_n)\ln(1+\mu_1\theta_n) - \mu_1\theta_n}{\mu_1\theta_n} \tag{124}$$

$$= \mu_1\frac{(1+\mu_1\theta_n)\ln(1+\mu_1\theta_n) - \mu_1\theta_n}{(\mu_1\theta_n)^2} \tag{125}$$

$$\overset{(b)}{\dot{\ge}} \mu_1, \tag{126}$$

where (a) follows since $\frac{(1+\theta)\ln(1+\theta)-\theta}{\theta}$ is increasing in θ and $\mu_1 \le 1$, and (b) since $\frac{(1+\theta)\ln(1+\theta)-\theta}{\theta^2}$ is decreasing in θ and $\mu_1\theta_n = 2\sqrt{d(n)}|\mathcal{Q}'|^2 \le 2\sqrt{d(n)}(n+1)^{2|\mathcal{X}||\mathcal{Z}|}$. Using this lower bound in (122) we get

$$\mathbb{E}[L(z^n)\log L(z^n)] + \log(\mathrm{e})\frac{1}{M}$$

$$= \log(\mathrm{e})\big(\mathbb{E}[L(z^n)\ln L(z^n)] + \frac{1}{M}\big) \dot{\ge} \nu_1 + \mu_2 \tag{127}$$

which, in turn, shows

$$\Phi_n(z^n)\Big(\mathbb{E}[L(z^n)\log L(z^n)] + \log(\mathrm{e})\frac{1}{M}\Big)$$
$$\dot{\geq} \exp\big(-n[\min\{E_1(\hat{Q}_{z^n}), E_2(\hat{Q}_{z^n})\} + H(\hat{Q}_{z^n})]\big), \tag{128}$$

using the (lower) bounds of (87) and (89). Combining (90) and (128) concludes the proof. ∎

VI. Conclusion and Discussion

We analyzed the *exact* exponential decay rate of the information leaked to the eavesdropper in Wyner's wiretap channel setting when an average wiretap channel code in the ensemble of i.i.d. or constant-composition random codes is used for communication. Our analysis shows that the previously-derived lower bound on the secrecy exponent of i.i.d. random codes in [7]–[10] is, indeed, tight. Moreover, our result for constant-composition random codes improves upon that of [12] (see (27) and examples in Section IV-B).

A key step in our analysis (which is applicable to any ensemble of random codes with independently sampled codewords) is to observe the equivalence of secrecy and resolvability exponents for the ensemble and, as a result, reducing the problem to the analysis of the resolvability exponent. The latter is easier as the informational divergence of interest (whose exponential decay rate is being assessed) involves a single random distribution (the output distribution) while the former involves two (the conditional and unconditional output distributions). We should emphasize that establishing secrecy via channel resolvability is a standard technique which was used in [6], [9], [10], [14] (also, in combination with privacy amplification in [7], [12]) whose advantages are discussed in [4]. Our result (Theorem 1) highlights the usefulness of this tool by showing that the resolvability exponent is not only a lower bound to the secrecy exponent but also equals the secrecy exponent.

Thanks to such a reduction, we extended the method of [10] to derive the exact resolvability exponent of random codes. It is noteworthy that, as it was already envisioned in [10], the method presented there was conveniently applicable to the ensemble of constant-composition random codes (as well as the ensemble of i.i.d. random codes already studied in [10]).

It is remarkable that, unlike the channel coding problem for which constant-composition random codes turn out to be never worse than i.i.d. random codes in terms of the exponent [21], for the resolvability problem we have examples (see Figures 4 and 5) where i.i.d. random codes are better than constant-composition codes. The examples presented in Section IV-B suggest that the superior ensemble (in terms of the secrecy exponent) depends on the channel W_{E} alone (i.e., for a given channel, either of the ensembles yields a better secrecy exponent for all input distributions). A subject for future research would be to characterize the set of channels for which the ensemble of i.i.d. random codes results in a better secrecy exponent (and vice versa).

As shown in [2], for general pairs of channels $(W_{\mathrm{M}}, W_{\mathrm{E}})$ if $I(X; Y) \leq I(X; Z)$ for all input distributions P_X, one can prefix the channel with an auxiliary channel $P_{X|U} : \mathcal{U} \to \mathcal{X}$, and by choosing P_U such that $I(U; Y) - I(U; Z) > 0$ (when $U \multimap X \multimap (Y, Z)$ have distribution $P_U(u)P_{X|U}(x|u)W_{\mathrm{M}}(y|x)W_{\mathrm{E}}(z|x)$) achieve secret message rates up to $I(U; Y) - I(U; Z)$. Channel prefixing is also proposed in [9] as a technique to treat the wiretap channels with cost constraints (the auxiliary channel $P_{X|U}$ will be chosen in such a way that its output sequence satisfies the cost

constraints for the physical channel). It is obvious that our results (as well as those of others cited) are immediately extensible to such cases.

APPENDIX A

COMPLEMENTARY PROOFS

A. Properties of $g_\star$ and g

Lemma 7. *Let $f\colon \mathcal{Q} \to \mathbb{R}$ and $g_\star\colon \mathcal{Q} \to \mathbb{R}$ be defined as in (23b) and (45) respectively, then*

(i) $\forall Q \in \mathcal{Q}$:

$$f(Q) \leq g_\star(Q) \leq I(Q) + D(Q_X \| P_X), \tag{129}$$

and the lower bound is attained if and only if $\tilde{P}_X(x) \triangleq P_X(x) \sum_z W(z|x) \frac{Q_Z(z)}{P_Z(z)} \in \mathcal{Q}$.

(ii) $g_\star(P_{X,Z}) = I(X; Z)$,

Proof: The upper bound of (129) follows since $Q' = Q$ is a feasible point in the minimization of (45). To establish the lower bound we have

$$I(Q') + D(Q'_X \| P_X) - \omega(Q')$$
$$= D(Q' \| P_{XZ}) - D(Q'_Z \| P_Z) + \sum_z Q'_Z(z) \log \frac{1}{P_Z(z)}. \tag{130}$$

Therefore,

$$\min_{Q' : Q'_Z = Q_Z} \{I(Q') + D(Q'_X \| P_X) - \omega(Q')\}$$
$$= \min_{Q' : Q'_Z = Q_Z} \{D(Q' \| P_{XZ}) - D(Q'_Z \| P_Z)\}$$
$$+ \sum_{x,z} Q(x,z) \log \frac{1}{P_Z(z)}. \tag{131}$$

By the convexity of divergence the value of the minimization on the right-hand-side of (131) is non-negative, hence,

$$g_\star(Q) \geq \sum_{x,z} Q(x,z) \log \frac{W(z|x)}{P_Z(z)}. \tag{132}$$

Moreover, the minimization (on the right-hand-side of (131)) evaluates to 0 if we can pick $Q'(x,z) = P_{X|Z}(x|z) Q_Z(z)$. Finally, (ii) follows as $f(P_{XZ}) = I(X; Z)$. ∎

To simplify the presentation let $\mathcal{V}$ denote the set of all stochastic matrices $V\colon \mathcal{X} \to \mathcal{Z}$ such that $P_X \times V \ll P_X \times W$. $\mathcal{V}$ is a compact and convex set. Moreover if $V'\colon \mathcal{X} \to \mathcal{Z}$ is not in $\mathcal{V}$, $\omega(V') = -\infty$, hence $I(P, V') - \omega(V') = +\infty$. Consequently we can rewrite (24b) as

$$g(V) = \omega(V) + \min_{\substack{V' \in \mathcal{V}: \\ P_X \circ V' = P_X \circ V}} \{I(P_X, V') - \omega(V')\}. \tag{133}$$

Note that the minima in the above is well-defined as $\mathcal{V}$ is a compact set.

Lemma 8. *The function $g\colon \mathcal{V} \to \mathbb{R}$ (as defined in (24b)) is convex and continuous in V.*

Proof: Since $\omega(V)$ (as defined in (24c)) is linear in V and for $V \in \mathcal{V}$ is continuous the claim follows if we show that the mapping

$$\psi: V \mapsto \min_{\substack{V' \in \mathcal{V}: \\ P_X \circ V' = P_X \circ V}} \{I(P_X, V') - \omega(V')\} \tag{134}$$

is convex and continuous on $\mathcal{V}$.

We first prove the convexity. Pick two stochastic matrices $V_1 \in \mathcal{V}$ and $V_2 \in \mathcal{V}$, $\lambda \in [0, 1]$, and set $V = \lambda V_1 + \overline{\lambda} V_2$ (where $\overline{\lambda} = 1 - \lambda$). Suppose the minimizer in (134) is $V_j^\star$ for $V = V_j$, $j = 1, 2$. We have,

$$\min_{\substack{V' \in \mathcal{V}: \\ P_X \circ V' = P_X \circ V}} \{I(P_X, V') - \omega(V')\}$$

$$\overset{(a)}{\leq} I(P_X, \lambda V_1^\star + \overline{\lambda} V_2^\star) - \omega(\lambda V_1^\star + \overline{\lambda} V_2^\star) \tag{135}$$

$$\overset{(b)}{=} I(P_X, \lambda V_1^\star + \overline{\lambda} V_2^\star) - \lambda \omega(V_1^\star) - \overline{\lambda} \omega(V_2^\star) \tag{136}$$

$$\overset{(c)}{\leq} \lambda \big(I(P_X, V_1^\star) - \omega(V_1^\star)\big) + \overline{\lambda} \big(I(P_X, V_2^\star) - \omega(V_2^\star)\big). \tag{137}$$

where (a) follows since $P_X \circ (\lambda V_1^\star + \overline{\lambda} V_2^\star) = \lambda P_X \circ V_1^\star + \overline{\lambda} P_X \circ V_2^\star = \lambda P_X \circ V_1 + \overline{\lambda} P_X \circ V_2 = P_X \circ V$, (b) since $\omega(V)$ is linear in V, and (c) since $I(P_X, V)$ is convex in V.

Convexity of ψ implies continuity in the interior of the set $\mathcal{V}$. The only possibility for ψ for being discontinuous is to 'jump up' at the boundaries. More precisely, to have a sequence $\{V_n \in \mathcal{V} : n \in \mathbb{N}\}$ such that $\lim_{n \to \infty} V_n = V$ (for some V on the boundaries of the set $\mathcal{V}$) but $\lim_{n \to \infty} \psi(V_n) < \psi(V)$. We shall show that this cannot happen.

Let $V_n^\star$ be the minimizer in (134) for $V = V_n$, that is,

$$V_n^\star = \arg\min_{\substack{V' \in \mathcal{V}: \\ P_X \circ V' = P_X \circ V_n}} \{I(P_X, V') - \omega(V')\}. \tag{138}$$

Consequently,

$$\lim_{n \to \infty} \psi(V_n) = \lim_{n \to \infty} \{I(P_X, V_n^\star) - \omega(V_n^\star)\}. \tag{139}$$

The sequence $\{V_n^\star \in \mathcal{V} : n \in \mathbb{N}\}$ must have a convergent subsequence and hence a limit point in $\mathcal{V}$ (as $\mathcal{V}$ is compact). Let $\lim_{n \to \infty} V_n^\star = \tilde{V}$ (by passing to the convergent subsequence if necessary). Since the mapping $V' \mapsto I(P_X, V') - \omega(V')$ is continuous on $\mathcal{V}$,

$$\lim_{n \to \infty} \{I(P_X, V_n^\star) - \omega(V_n^\star)\} = I(P_X, \tilde{V}) - \omega(\tilde{V}). \tag{140}$$

Furthermore, the projection $V' \mapsto P_X \circ V'$ is continuous, thus,

$$\lim_{n \to \infty} P_X \circ V_n^\star = P_X \circ \tilde{V}, \tag{141}$$

and

$$\lim_{n \to \infty} P_X \circ V_n = P_X \circ V. \tag{142}$$

Moreover,

$$P_X \circ V_n = P_X \circ V_n^\star, \quad \forall n \in \mathbb{N}, \tag{143}$$

by definition. Combining (141), (142), and (143) we have

$$P_X \circ \tilde{V} = P_X \circ V \tag{144}$$

Consequently,

$$\lim_{n \to \infty} \psi(V_n) = I(P_X, \tilde{V}) - \omega(\tilde{V}) \tag{145}$$

$$\geq \min_{\substack{V' \in \mathcal{V}: \\ P_X \circ V' = P_X \circ V}} I(P_X, V') - \omega(V') \tag{146}$$

$$= \psi(V). \tag{147}$$

This concludes the proof. ∎

B. Alternative form of $E_s^{\text{i.i.d.}}$

Using the fact that $\max\{a, 0\} = \max_{0 \leq \lambda \leq 1} \lambda a$,

$$\min_Q \left\{ D(Q \| P_{XZ}) + [R - f(Q)]^+ \right\}$$

$$= \min_Q \left\{ D(Q \| P_{XZ}) + \max_{0 \leq \lambda \leq 1} \lambda[R - f(Q)] \right\} \tag{148}$$

$$= \min_Q \max_{0 \leq \lambda \leq 1} \left\{ \lambda R + D(Q \| P_{XZ}) - \lambda f(Q) \right\} \tag{149}$$

$$\overset{(a)}{=} \max_{0 \leq \lambda \leq 1} \min_Q \left\{ \lambda R + D(Q \| P_{XZ}) - \lambda f(Q) \right\} \tag{150}$$

$$= \max_{0 \leq \lambda \leq 1} \left\{ \lambda R + \min_Q \left\{ D(Q \| P_{XZ}) - \lambda f(Q) \right\} \right\} \tag{151}$$

$$\overset{(b)}{=} \max_{0 \leq \lambda \leq 1} \left\{ \lambda R - F_0(P_X, W, \lambda) \right\} \tag{152}$$

where (a) follows since $D(Q \| P_{XZ}) - \lambda f(Q)$ is convex in Q (recall that $f(Q)$ is linear in Q) and (b) since

$$D(Q \| P_{XZ}) - \lambda f(Q)$$

$$= \sum_{x,z} Q(x, z) \log \frac{Q(x, z)}{P_{XZ}(x, z)^{1+\lambda} P_X(x)^{-\lambda} P_Z(z)^{-\lambda}} \tag{153}$$

$$\overset{(*)}{\geq} -\log \sum_{x,z} P_{XZ}(x, z)^{1+\lambda} P_X(x)^{-\lambda} P_Z(z)^{-\lambda} \tag{154}$$

$$= F_0(P_X, W, \lambda), \tag{155}$$

with equality in $(*)$ iff $Q(x, z) \propto P_{XZ}(x, z)^{1+\lambda} P_X(x)^{-\lambda} P_Z(z)^{-\lambda}$. ∎

C. Strict Monotonicity of $E_s^{\text{i.i.d.}}$ and $E_s^{\text{c.c.}}$ in R

That $E_s^{\text{i.i.d.}}$ is strictly increasing in R for $R > I(X; Z)$ can be easily seen through the form of (25): $E_s^{\text{i.i.d.}}$ is the supremum of affine functions of R thus is convex in R. On the other side, since $F_0(P_X, W, \lambda)$ is a convex function of λ passing through the origin with slope $I(X; Z)$, $E_s^{\text{i.i.d.}}(P_X, W, R)$ starts to increase above 0 once R exceeds $I(X; Z)$ which means it will be strictly increasing for $R > I(X; Z)$.

We only need to prove the claim for $E_{\mathrm{s}}^{\mathrm{c.c.}}$. (This proof may also be used to show $E_{\mathrm{s}}^{\mathrm{i.i.d.}}$ is strictly increasing in R, replacing $g(V)$ with $f(Q)$.) Note that

$$E_{\mathrm{s}}^{\mathrm{c.c.}}(P_X, W, R) = \min\Big\{ \min_{V:g(V)\geq R} D(V\|W|P_X),$$

$$\min_{V:g(V)\leq R}\{D(V\|W|P_X) + R - g(V)\}\Big\}. \tag{156}$$

We first show that for $R > I(X;Z)$,

$$E_{\mathrm{s}}^{\mathrm{c.c.}}(P_X, W, R) = \min_{V:g(V)\leq R}\{D(V\|W|P_X) + R - g(V)\} \tag{157}$$

$$= R + \min_{V:g(V)\leq R}\{D(V\|W|P_X) - g(V)\} \tag{158}$$

This follows since for $R > I(X;Z)$,

$$\min_{V:g(V)\geq R} D(V\|W|P_X) = \min_{V:g(V)=R} D(V\|W|P_X) \tag{159}$$

Let us first prove (159): Suppose this is not the case, i.e., there exists $V^\star$ with $g(V^\star) > R$ such that $D(V^\star\|W|P_X) \leq D(V\|W|P_X)$ for every V with $g(V) \geq R$. We can safely assume that $P_X \times V^\star \ll P_X \times W$ (otherwise $D(V\|W|P_X) = +\infty$ for all V such that $g(V) \geq R$ and (158) automatically follows). Let $V_\lambda \triangleq \lambda V^\star + (1-\lambda)W$, for $\lambda \in [0,1]$. It is easy to check that $\forall \lambda \in [0,1]: P_X \times V_\lambda \ll P_X \times W$, thus the mapping $\lambda \mapsto g(V_\lambda)$ is convex and continuous by the convexity and continuity of g (see Lemma 8) on the interval $[0,1]$. We know that $g(V_1) = g(V_\star) > R$ and $g(V_0) = g(W) = I(X;Z) < R$. Therefore, there exists $\beta \in (0,1)$ for which $g(V_\beta) = R$. On the other side, the convexity of divergence implies

$$D(V_\beta\|W|P_X) \leq \beta D(V^\star\|W|P_X) + (1-\beta)D(W\|W|P_X) \tag{160}$$

$$< D(V^\star\|W|P_X) \tag{161}$$

since $\beta < 1$. This contradicts the optimality of $V^\star$.

Now, we show that $E_{\mathrm{s}}^{\mathrm{c.c.}}(P_X, W, R') > E_{\mathrm{s}}^{\mathrm{c.c.}}(P_X, W, R)$ for $R' > R > I(X;Z)$. Let

$$V^* = \underset{V:g(V)\leq R'}{\arg\min}\{D(V\|W|P_X) - g(V)\}. \tag{162}$$

If $g(V^*) \leq R$, then

$$E_{\mathrm{s}}^{\mathrm{c.c.}}(P_X, W, R') = R' + D(V^*\|W|P_X) - g(V^*) \tag{163}$$

$$= R' + \min_{V:g(V)\leq R}\{D(V\|W|P_X) - g(V)\} \tag{164}$$

$$> R + \min_{V:g(V)\leq R}\{D(V\|W|P_X) - g(V)\} \tag{165}$$

$$= E_{\mathrm{s}}^{\mathrm{c.c.}}(P_X, W, R) \tag{166}$$

which proves the claim.

Otherwise, we have $R < g(V^*) \leq R'$. Consider once again the family of stochastic matrices defined as $V_\lambda \triangleq \lambda V^* + (1-\lambda)W$. We know $P_X \times V^* \ll P_X \times W$ (for if it is not, $D(V^*\|W|P_X) = +\infty$ and $g(V^*) = -\infty$ which means the exponent is infinity which is contradiction since $E_{\mathrm{s}}^{\mathrm{c.c.}}(P_X, W, R') \leq R' - I(X;Z)$ by taking $V = W$

in (158)). Using the same reasoning as above, since $g(V_1) > R$ and $g(V_0) = I(X; Z) < R$ one can find $\beta \in (0, 1)$ such that $g(V_\beta) = R$ and

$$D(V_\beta \| W | P_X) \leq \beta D(V^\star \| W | P_X).$$ (167)

Moreover, we know that

$$D(V_\beta \| W | P_X) = R + [D(V_\beta \| W | P_X) - g(V_\beta)]$$ (168)

$$\geq R + \min_{V: g(V) \leq R} \{D(V \| W | P_X) - g(V)\}$$ (169)

$$= E_s^{\text{c.c.}}(P_X, W, R).$$ (170)

One the other side,

$$E_s^{\text{c.c.}}(P_X, W, R') = R' + D(V^* | W | P_X) - g(V^*)$$ (171)

$$\overset{(a)}{\geq} D(V^* \| W | P_X)$$ (172)

$$\overset{(b)}{\geq} \frac{1}{\beta} D(V_\beta \| W | P_X)$$ (173)

$$\overset{(c)}{\geq} \frac{1}{\beta} E_s^{\text{c.c.}}(P_X, W, R)$$ (174)

$$\overset{(*)}{>} E_s^{\text{c.c.}}(P_X, W, R),$$ (175)

where (a) follows since $g(V^\star) \leq R'$, (b) follows from (167) and (c) from (170) and finally $(*)$ holds since $\beta < 1$ and $E_s^{\text{c.c.}}(P_X, W, R) > 0$. ∎

D. Proof of (27)

By Lemma 7 we have $g(V) \leq I(P_X, V)$, thus,

$$R - g(V) \geq R - I(P_X, V).$$ (176)

Therefore,

$$E_s^{\text{c.c.}}(P_X, W, R) = \min_V \{D(V \| W | P_X) + [R - g(V)]^+\}$$ (177)

$$\geq \min_V \{D(V \| W | P_X) + [R - I(P_X, V)]^+\}$$ (178)

$$\overset{(a)}{=} \min_V \{D(V \| W | P_X) + \max_{0 \leq \lambda \leq 1} \{\lambda R - \lambda I(P_X, V)\}\}$$ (179)

$$\overset{(b)}{=} \max_{0 \leq \lambda \leq 1} \{\lambda R + \min_V \{D(V \| W | P_X) - \lambda I(P_X, V)\}\}$$ (180)

where (a) follows since $[a]^+ = \max_{0 \le \lambda \le 1} \lambda a$ and (b) by observing that $D(V\|W|P_X) - \lambda I(P_X, V)$ is convex in V for $\lambda \le 1$ (and linear in λ). The latter holds since $I(P_X, V) = \min_{Q_Z \in \mathcal{P}(\mathcal{Z})} D(V\|Q_Z|P_X)$, therefore,

$$D(V\|W|P_X) - \lambda I(P_X, V)$$

$$= \max_{Q_Z \in \mathcal{P}(\mathcal{Z})} \{D(V\|W|P_X) - \lambda D(V\|Q_Z|P_X)\} \tag{181}$$

$$= \max_{Q_Z} \sum_{x,z} P_X(x) V(z|x) \log \frac{V(z|x)^{1-\lambda}}{W(z|x) Q_Z(z)^{-\lambda}} \tag{182}$$

$$= \frac{1}{t} \max_{Q_Z} \sum_{x,z} P_X(x) V(z|x) \log \frac{V(z|x)}{W(z|x)^t Q_Z(z)^{1-t}}. \tag{183}$$

where we have defined $t \triangleq \frac{1}{1-\lambda}$ in the last step. The objective function inside the $\max$ in (183) is convex in V and since the supremum of convex functions is still convex, the convexity of $D(V\|W|P_X) - \lambda I(P_X, V)$ in V follows. It can also be seen that the objective function is concave in Q_Z for $\lambda > 0$ (i.e. $t > 1$). Using this observation we have

$$\min_V \{D(V\|W|P_X) - \lambda I(P_X, V)\}$$

$$= \frac{1}{t} \min_V \max_{Q_Z} \sum_{x,z} P_X(x) V(z|x) \log \frac{V(z|x)}{W(z|x)^t Q_Z(z)^{1-t}} \tag{184}$$

$$= \frac{1}{t} \max_{Q_Z} \min_V \sum_{x,z} P_X(x) V(z|x) \log \frac{V(z|x)}{W(z|x)^t Q_Z(z)^{1-t}} \tag{185}$$

$$\overset{(a)}{=} \max_{Q_Z} \left\{ -\frac{1}{t} \sum_x P_X(x) \log \sum_z W(z|x)^t Q_Z(z)^{1-t} \right\} \tag{186}$$

$$\overset{(b)}{\ge} \max_{Q_Z} \left\{ -\frac{1}{t} \log \sum_x P_X(x) \sum_z W(z|x)^t Q_Z(z)^{1-t} \right\} \tag{187}$$

$$= -\min_{Q_Z} \left\{ \frac{1}{t} \log \sum_z Q_Z(z)^{1-t} \sum_x P_X(x) W(z|x)^t \right\} \tag{188}$$

where (a) and (b) follow by the concavity of logarithm. KKT conditions imply the solution to the minimization of (188) is

$$Q_Z(z) = c \left(\sum_x P_X(x) W(z|x)^t \right)^{1/t} \tag{189}$$

with $c^{-1} = \sum_z \left(\sum_x P_X(x) W(z|x)^t \right)^{1/t}$. Plugging this into the objective function of (188) and replacing $t = \frac{1}{1-\lambda}$ we have

$$\min_V \{D(V\|W|P_X) - \lambda I(P_X, V)\}$$

$$= -\log \sum_z \left(\sum_x P_X(x) W(z|x)^{\frac{1}{1-\lambda}} \right)^{1-\lambda} \tag{190}$$

$$= -E_0(P_X, W, \lambda). \tag{191}$$

Plugging (191) into (180) proves the claim. ∎

E. Proof of Lemma 4

$\Phi_n(z^n)$ is the expectation of a non-negative random variable $P_{\mathcal{C}_n}(z^n)$. Therefore, $\Phi_n(z^n) = 0$ implies $P_{\mathcal{C}_n}(z^n) = 0$ almost surely. This proves (i).

We have

$$\Phi_n(z^n) = \sum_{x^n \in \mathcal{X}^n} P_{X^n}(x^n) W^n(z^n|x^n) \tag{192}$$

Let

$$\xi \triangleq \min_{x \in \mathcal{X}} P_X(x) \tag{193}$$

and

$$\zeta \triangleq \min_{z \in \mathcal{Z}} \min_{x:W(z|x)>0} W(z|x) \tag{194}$$

be two strictly positive and finite constants that depend only on P_X and W. $\Phi_n(z^n) > 0$ implies there exists at least one sequence $x_0^n \in \mathrm{supp}(P_{X^n})$ for which $W^n(z^n|x_0^n) > 0$. Therefore, $W^n(z^n|x_0^n) > \zeta^n$. Thus (192) yields

$$\Phi_n(z^n) \geq P_{X^n}(x_0^n)\zeta^n. \tag{195}$$

Moreover, for both choices of P_{X^n} in (22) if $x^n \in \mathrm{supp}(P_{X^n})$, $P_{X^n}(x^n) \geq P_X^n(x^n) \geq \xi^n$. Using this observation in (195) proves (ii) (with $\alpha = \frac{1}{\xi\zeta}$). ∎

F. Proof of Lemma 6

Take $U \triangleq \frac{A}{\mathbb{E}[A]}$ so that $\mathbb{E}[U] = 1$. We shall prove that

$$c(\theta)\left(\mathrm{var}(U) - \tau_\theta(U)\right) \leq \mathbb{E}[U\ln(U)] \leq \mathrm{var}(U). \tag{196}$$

The claim then follows by noting that $\mathbb{E}[A\ln(A/\mathbb{E}[A])] = \mathbb{E}[A]\,\mathbb{E}[U\ln(U)]$ and $\mathrm{var}(A) = \mathrm{var}(U)/(\mathbb{E}[A])^2$.

We firstly have

$$\mathbb{E}[U\ln(U)] = \mathbb{E}[U\ln(U) - (U-1)] \tag{197}$$

$$\leq \mathbb{E}[(U-1)^2] = \mathrm{var}(U), \tag{198}$$

since $u\ln(u) - (u-1) \leq (u-1)^2$. Moreover,

$$u\ln(u) - (u-1) \geq c(\theta)(u-1)^2 \mathbb{1}\{u \leq \theta+1\}. \tag{199}$$

This follows by observing that $\frac{u\ln(u)-(u-1)}{(u-1)^2}$ is a decreasing function of u. Thus,

$$\mathbb{E}[U\ln(U)] \geq c(\theta) \int_0^{\theta+1} (u-1)^2 \, \mathrm{d}\,F_U(u). \tag{200}$$

where $F_U(u)$ is the cumulative distribution function of u.

Furthermore,

$$\int_0^{\theta+1} (u-1)^2 \, \mathrm{d}\,F_U(u) = \mathrm{var}(U) - \int_{\theta+1}^{+\infty} (u-1)^2 \, \mathrm{d}\,F_U(u) \tag{201}$$

Let $v \triangleq u - 1$ for the sake of brevity and denote by $\bar{F}_V(v) \triangleq \Pr\{V > v\} = \Pr\{U > v + 1\}$ the complementary distribution function of V. Then,

$$\int_{\theta+1}^{+\infty} (u-1)^2 \, \mathrm{d}\, F_U(u) = \int_{\theta}^{+\infty} v^2 \, \mathrm{d}\, F_V(v) \tag{202}$$

$$= \left[-v^2 \bar{F}_V(v)\right]_{\theta}^{+\infty} + 2 \int_{\theta}^{+\infty} v \bar{F}_V(v) \, \mathrm{d}\, v \tag{203}$$

$$\stackrel{(*)}{=} \theta^2 \bar{F}_V(\theta) + 2 \int_{\theta}^{+\infty} v \bar{F}_V(v) \, \mathrm{d}\, v. \tag{204}$$

The equality in $(*)$ follows since we assumed the variance of U exists. This proves (196). ∎

G. Proof of (53)

We have

$$p_Q = \sum_{x^n \in \mathcal{X}^n} \mathbb{1}\{(x^n, z^n) \in \mathcal{T}_Q^n\} P_{X^n}(x^n) \tag{205}$$

$$= \frac{P_{X^n}(\mathcal{T}_{Q_X}^n)}{|\mathcal{T}_{Q_X}^n|} \sum_{x^n \in \mathcal{X}^n} \mathbb{1}\{(x^n, z^n) \in \mathcal{T}_Q^n\} \tag{206}$$

since $P_{X^n}(x^n)$ only depends on the type of x^n. On the other side, we have

$$|\mathcal{T}_Q^n| = \sum_{z^n \in \mathcal{Z}^n} \sum_{x^n \in \mathcal{X}^n} \mathbb{1}\{(x^n, z^n) \in \mathcal{T}_Q^n\} \tag{207}$$

The value of the inner sum in (207) only depends on the type of $z^{n\,3}$ and, clearly, is zero if $Q_Z \neq \hat{Q}_{z^n}$. Thus

$$|\mathcal{T}_Q^n| = |\mathcal{T}_{\hat{Q}_{z^n}}^n| \mathbb{1}\{Q_Z = \hat{Q}_{z^n}\} \sum_{x^n \in \mathcal{X}^n} \mathbb{1}\{(x^n, z^n) \in \mathcal{T}_Q^n\}. \tag{208}$$

Plugging (208) into (206) yields (53). ∎

H. Derivation of (63)

For both ensembles of interest we have

$$\exp\left(-n[H(Q_X) + D(Q_X \| P_X)]\right) \leq \frac{P_{X^n}(\mathcal{T}_{Q_X}^n)}{|\mathcal{T}_{Q_X}^n|}$$

$$\leq (n+1)^{|\mathcal{X}|} \exp\left(-n[H(Q_X) + D(Q_X \| P_X)]\right). \tag{209}$$

Moreover,

$$(n+1)^{-|X||Z|} \exp\left(nH(Q)\right) \leq \left|\mathcal{T}_Q^n\right| \leq \exp\left(nH(Q)\right), \tag{210}$$

and

$$\exp\left(-nH(Q_Z)\right) \leq \frac{1}{\left|\mathcal{T}_{Q_Z}^n\right|} \leq (n+1)^{|\mathcal{Z}|} \exp\left(-nH(Q_Z)\right) \tag{211}$$

---

[3]for if $z^n \neq \tilde{z}^n$ have the same type, by permuting the letters of $\tilde{z}^n$ we can obtain z^n. Now, if we apply the same permutation to every $x^n \in \mathcal{X}^n$ to obtain $\tilde{x}^n$,

$$\sum_{x^n \in \mathcal{X}^n} \mathbb{1}\{(x^n, z^n) \in \mathcal{T}_Q^n\} = \sum_{\tilde{x}^n \in \mathcal{X}^n} \mathbb{1}\{(\tilde{x}^n, \tilde{z}^n) \in \mathcal{T}_Q^n\}.$$

Multiplying the above, we have

$$(n+1)^{-|\mathcal{X}||\mathcal{Z}|} \exp\big(-n[I(Q) + D(Q_X\|P_X)]\big) \leq p_Q$$

$$\leq (n+1)^{|\mathcal{X}||\mathcal{Z}|} \exp\big(-n[I(Q) + D(Q_X\|P_X)]\big) \tag{212}$$

Plugging the above in (52) and using the fact that $|\mathcal{Q}_n[\hat{Q}_{z^n}]| \leq (n+1)^{|\mathcal{X}||\mathcal{Z}|}$ we obtain Equation 213 at the bottom of the next page which, in turn, yields (63). ∎

I. Proof of (65)

We only prove (65b).

$$\mathrm{var}(A) = \sum_{Q\in\mathcal{A}} \frac{1}{M^2} \ell(Q)^2 \, \mathrm{var}(N_Q)$$

$$+ \sum_{\substack{(Q_1,Q_2)\in\mathcal{A}^2 \\ Q_1 \neq Q_2}} \frac{1}{M^2} \ell(Q_1)\ell(Q_2) \, \mathrm{cov}(N_{Q_1}, N_{Q_2}) \tag{214}$$

$$\overset{(\star)}{=} \frac{1}{M} \sum_{Q\in\mathcal{A}} \ell(Q)^2 p_Q(1 - p_Q)$$

$$- \frac{1}{M} \sum_{\substack{(Q_1,Q_2)\in\mathcal{A}^2 \\ Q_1 \neq Q_2}} \ell(Q_1)\ell(Q_2) p_{Q_1} p_{Q_2}, \tag{215}$$

where $(\star)$ follows since $\mathrm{var}(N_Q) = M p_Q(1 - p_Q)$ and $\mathrm{cov}(N_{Q_1}, N_{Q_2}) = -M p_{Q_1} p_{Q_2}$. Moreover,

$$\sum_{\substack{(Q_1,Q_2)\in\mathcal{A}^2 \\ Q_1 \neq Q_2}} \ell(Q_1)\ell(Q_2) p_{Q_1} p_{Q_2}$$

$$= \sum_{Q_1\in\mathcal{A}} \ell(Q_1) p_{Q_1} \sum_{Q_2\in\mathcal{A}\setminus\{Q_1\}} \ell(Q_2) p_{Q_2} \tag{216}$$

$$= \sum_{Q_1\in\mathcal{A}} \ell(Q_1) p_{Q_1} \Big(\mathbb{E}[A] - p_{Q_1}\ell(Q_1) \Big) \tag{217}$$

where the last equality follows from (65a). Using the above in (215) we get,

$$\mathrm{var}(A)$$

$$= \frac{1}{M} \sum_{Q\in\mathcal{A}} \ell(Q) p_Q \Big[(1 - p_Q)\ell(Q) - \big(\mathbb{E}[A] - p_Q\ell(Q) \big) \Big] \tag{218}$$

$$= \frac{1}{M} \sum_{Q\in\mathcal{A}} \ell(Q) p_Q \big[\ell(Q) - \mathbb{E}[A] \big] \tag{219}$$

$$(n+1)^{-|\mathcal{X}||\mathcal{Z}|} \exp\Big(-n \min_{Q\in\mathcal{Q}_n[\hat{Q}_{z^n}]} \{I(Q) + D(Q_X\|P_X) - \omega(Q)\}\Big) \leq \Phi_n(z^n)$$

$$\leq (n+1)^{2|\mathcal{X}||\mathcal{Z}|} \exp\Big(-n \min_{Q\in\mathcal{Q}_n[\hat{Q}_{z^n}]} \{I(Q) + D(Q_X\|P_X) - \omega(Q)\}\Big) \tag{213}$$

$$= \frac{1}{M} \sum_{Q \in \mathcal{A}} \ell(Q)^2 p_Q - \frac{1}{M} \mathbb{E}[A]^2. \qquad \blacksquare$$

APPENDIX B
NUMERICAL EVALUATION OF THE SECRECY EXPONENTS

A. Computing $E_{\mathrm{s}}^{\mathrm{i.i.d.}}$ and $\underline{E}_{\mathrm{s}}^{\mathrm{c.c.}}$

Both $E_{\mathrm{s}}^{\mathrm{i.i.d.}}$ and $\underline{E}_{\mathrm{s}}^{\mathrm{c.c.}}$ can be easily evaluated via the expressions (25) and (26) using the fact that both F_0 and E_0 (defined in (25b) and (26b) respectively) are convex in λ, and pass through the origin with slope $I(X; Z)$.

For instance to evaluate $E_{\mathrm{s}}^{\mathrm{i.i.d.}}$ we know that

1) for $R \leq I(X; Z) = \frac{\partial}{\partial \lambda} F_0(P_X, W, \lambda)\big|_{\lambda=0}$, $E_{\mathrm{s}}(P_X, W, R) = 0$;

2) for $I(X; Z) < R < \frac{\partial}{\partial \lambda} F_0(P_X, W, \lambda)\big|_{\lambda=1}$, the pairs $R, E_{\mathrm{s}}^{\mathrm{i.i.d.}}$ are related parametrically as

$$R(\lambda) = \frac{\partial}{\partial \lambda} F_0(P_X, W, \lambda) \qquad (220a)$$

$$E_{\mathrm{s}}(\lambda) = \lambda R(\lambda) - F_0(P_X, W, \lambda) \qquad (220b)$$

for the range of $\lambda \in [0, 1]$;

3) finally, if $R \geq F_0'(1)$,

$$E_{\mathrm{s}}(P_X, W, R) = R - F_0(P_X, W, 1). \qquad (221)$$

It is clear that to evaluate $\underline{E}_{\mathrm{s}}^{\mathrm{c.c.}}$ one has to precisely follow the same steps replacing F_0 with E_0.

B. Computing $E_{\mathrm{s}}^{\mathrm{c.c.}}$

To compute $E_{\mathrm{s}}^{\mathrm{c.c.}}$ (defined in (24)) one has to solve two minimizations. Namely, that of (24a) and that of (24b) (to compute $g(V)$). The latter turns out to be efficiently solvable using standard convex optimization tools.

Fix $Q_Z \in \mathcal{P}(\mathcal{Z})$ (to be set to $P_X \circ V$ to compute $g(V)$). Also note that $I(P_X, V') - \omega(V') = D(V'\|W|P_X) + H(P_X \circ V')$, thus, the minimization problem of (24b) is equivalent to minimizing $D(V'\|W|P_X)$ under the constraint $P_X \circ V' = Q_Z$. We have:

$$\min_{V': P_X \circ V' = Q} D(V'\|W|P_X) = \min_{V'} \Big\{ D(V'\|W|P_X)$$
$$+ \max_{\rho \in \mathbb{R}^{|\mathcal{Z}|}} \sum_z \rho_z \left[Q_Z(z) - (P_X \circ V')(z) \right] \Big\} \qquad (222)$$

$$= \max_{\rho \in \mathbb{R}^{|\mathcal{Z}|}} \Big\{ \min_{V'} \Big\{ D(V'\|W|P_X) - \sum_{x,z} P_X(x) V'(z|x) \rho_z \Big\}$$
$$+ \sum_z \rho_z Q_Z(z) \Big\}, \qquad (223)$$

where $\rho \triangleq (\rho_1, \ldots, \rho_{|\mathcal{Z}|})$ and the last equality follows since $D(V\|W|P_X)$ is convex in V and the second term is linear in V. Moreover, the inner unconstrained minimization has the value

$$\min_{V'} \left\{ D(V'\|W|P_X) - \sum_{x,z} P_X(x)V'(z|x)\rho_z \right\}$$

$$= \min_{V'} \sum_{x,z} P_X(x)V'(z|x) \log \frac{V'(z|x)}{W(z|x)\exp(\rho_z)} \tag{224}$$

$$= -\sum_x P_X(x) \log \sum_z W(z|x)\exp(\rho_z), \tag{225}$$

by choosing $V'(z|x) \propto W(z|x)\exp(\rho_z)$. Plugging this into (223) we get

$$\min_{V':P_X \circ V'=Q} D(V'\|W|P_X) = \max_{\rho \in \mathbb{R}^{|\mathcal{Z}|}} \left\{ \sum_z \rho_z Q_Z(z) \right.$$

$$\left. - \sum_x P_X(x) \log \sum_z W(z|x)\exp(\rho_z) \right\}. \tag{226}$$

Remark. Using Hölder's inequality, it can be checked that the objective function of (226) is concave in ρ thus can be efficiently maximized using standard numerical methods.

Proof: Since the first sum in the objective function of (226) is linear in ρ it is sufficient to prove that the function

$$\rho \mapsto \sum_x P_X(x) \log \left(W(z|x)\exp(\rho_z) \right) \tag{227}$$

is convex in ρ. Fix $t \in [0,1]$ and $\rho, \rho' \in \mathbb{R}^{|\mathcal{Z}|}$. For every $x \in \mathcal{X}$, Hölder's inequality implies

$$\sum_z W(z|x)\exp(t\rho_z + (1-t)\rho_z')$$

$$= \sum_z W(z|x)^t \exp(t\rho_z) \times W(z|x)^{1-t} \exp((1-t)\rho_z') \tag{228}$$

$$\leq \left(\sum_z W(z|x)\exp(\rho_z) \right)^t \times \left(\sum_x W(z|x)\exp(\rho_z') \right)^{1-t} \tag{229}$$

Taking the logarithm of both sides, multiplying by $P_X(x)$, and finally summing over x proves the claim. ∎

Finally, for small alphabet sizes that we have considered in Section IV-B we can solve the minimization of (24a) via exhaustive search.

References

[1] A. D. Wyner, "The wire-tap channel," *Bell System Technical Journal*, vol. 54, no. 8, pp. 1355–1387, 1975.

[2] I. Csiszár and J. Körner, "Broadcast channels with confidential messages," *IEEE Transactions on Information Theory*, vol. 24, no. 3, pp. 339–348, 1978.

[3] J. L. Massey, "A simplified treatment of wyner's wire-tap channel." in *Proceedings of Annual Allerton Conference on Communication, Control, and Computing*, Monticello, IL, Oct. 1983, pp. 268–276.

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

[5] U. Maurer and S. Wolf, "Information-theoretic key agreement: From weak to strong secrecy for free," in *Advances in Cyptology — EUROCRYPT 2000*, ser. Lecture Notes in Computer Science, B. Preneel, Ed., vol. 1807. Springer-Verlag, May 2000, pp. 351–368.

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

[7] ——, "Exponential decreasing rate of leaked information in universal random privacy amplification," *IEEE Transactions on Information Theory*, vol. 57, no. 6, pp. 3989–4001, Jun. 2011.

[8] M. Hayashi and R. Matsumoto, "Secure multiplex coding with dependent and non-uniform multiple messages," in *Proceedings of Annual Allerton Conference on Communication, Control, and Computing*, Oct. 2012, pp. 954–959.

[9] T. S. Han, H. Endo, and M. Sasaki, "Reliability and secrecy functions of the wiretap channel under cost constraint," *IEEE Transactions on Information Theory*, vol. 60, no. 11, pp. 6819–6843, Nov. 2014.

[10] M. Bastani Parizi and E. Telatar, "On the secrecy exponent of the wire-tap channel," in *Proceedings of IEEE Information Theory Workshop (ITW)*, Oct. 2015, pp. 287–291.

[11] J. Körner and A. Sgarro, "Universally attainable error exponents for broadcast channels with degraded message sets," *IEEE Transactions on Information Theory*, vol. 26, no. 6, pp. 670–679, Nov. 1980.

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

[13] J. Hou and G. Kramer, "Informational divergence approximations to product distributions," in *Proceedings of Canadian Workshop on Information Theory (CWIT)*, Jun. 2013, pp. 76–81.

[14] ——, "Effective secrecy: Reliability, confusion and stealth," in *Proceedings of IEEE International Symposium on Information Theory (ISIT)*, Jun. 2014, pp. 601–605.

[15] T.-H. Chou, V. Y. F. , Tan, and S. C. Draper, "The sender-excited secret key agreement model: Capacity, reliability, and secrecy exponents," *IEEE Transactions on Information Theory*, vol. 61, no. 1, pp. 609–627, Jan. 2015.

[16] U. M. Maurer, "Secret key agreement by public discussion from common information," *IEEE Transactions on Information Theory*, vol. 39, no. 3, pp. 733–742, May 1993.

[17] R. Ahlswede and I. Csiszár, "Common randomness in information theory and cryptography—part I: Secret sharing," *IEEE Transactions on Information Theory*, vol. 39, no. 4, pp. 1121–1132, Jul. 1993.

[18] M. Hayashi, "Tight exponential analysis of universally composable privacy amplification and its applications," *IEEE Transactions on Information Theory*, vol. 59, no. 11, pp. 7728–7746, Nov. 2013.

[19] M. Hayashi and V. Y. F. Tan, "Equivocations and exponents under various rényi information measures," in *Proceedings of IEEE International Symposium on Information Theory (ISIT)*, Jun. 2015, pp. 281–285.

[20] I. Csiszár, "The method of types," *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2505–2523, Oct. 1998.

[21] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, 2nd ed.   Cambridge University Press, 2011.

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

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

[24] R. G. Gallager, *Information Theory and Reliable Communication*.   New York, NY, USA: John Wiley & Sons, Inc., 1968.

[25] ——, "The random coding bound is tight for the average code," *IEEE Transactions on Information Theory*, vol. 19, no. 2, pp. 244–246, Mar. 1973.

[26] N. Shulman, "Communication over an unknown channel via common broadcasting," Ph.D. dissertation, Department of Electrical Engineering Systems, Tel Aviv University, 2003.

[27] N. Merhav, "Exact random coding error exponents of optimal bin index decoding," *IEEE Transactions on Information Theory*, vol. 60, no. 10, pp. 6024–6031, Oct. 2014.

[28] M. Bastani Parizi and E. Telatar, "On the secrecy exponent of the wire-tap channel," *arXiv e-prints*, vol. abs/1501.06287v3, 2015. [Online]. Available: http://arxiv.org/abs/1501.06287v3

[29] N. Merhav, "Statistical physics and information theory," *Foundations and Trends in Communications and Information Theory*, vol. 6, no. 1–2, pp. 1–212, 2009. [Online]. Available: http://dx.doi.org/10.1561/0100000052