The Secrecy Capacity of Cost-Constrained Wiretap Channels

Sreejith Sreekumar, Alexander Bunin, Ziv Goldfeld, *Member, IEEE*, Haim H. Permuter, *Senior Member, IEEE*, and Shlomo Shamai, *Life Fellow, IEEE*

***Abstract*— In many information-theoretic channel coding problems, adding an input cost constraint to the operational setup amounts to restricting the optimization domain in the capacity formula. This paper shows that, in contrast to common belief, such a simple modification does not hold for the cost-constrained (CC) wiretap channel (WTC). The secrecy-capacity of the discrete memoryless (DM) WTC without cost constraints is described by a single auxiliary random variable. For the CC DM-WTC, however, we show that two auxiliaries are necessary to achieve capacity. Specifically, we first derive the secrecy-capacity formula, proving the direct part via superposition coding. Then, we provide an example of a CC DM-WTC whose secrecy-capacity cannot be achieved using a single auxiliary. This establishes the fundamental role of superposition coding over CC WTCs.**

***Index Terms*— Cost constraint, physical-layer security, secrecy-capacity, superposition coding, wiretap channel.**

I. Introduction

PHYSICAL-LAYER security (PLS), rooted in information-theoretic principles, dates back to Wyner's landmark 1975 paper [1], where the wiretap channel (WTC) was introduced. This model formulates reliable and secure communication over noisy channels in the presence of an eavesdropper (see Fig. 1). By harnessing randomness from the noisy channel and combining it with proper physical layer coding, Wyner characterized the fundamental limit of reliable and secure communication, termed the *secrecy-capacity*. For a memoryless WTC $P_{Y,Z|X}$, the secrecy-capacity is [2]

$$C_{\mathsf{WTC}}(P_{Y,Z|X}) = \max_{P_{V,X}} I(V;Y) - I(V;Z), \tag{1}$$

where the joint distribution is $P_{V,X}P_{Y,Z|X}$ (i.e., $V-X-(Y,Z)$ forms a Markov chain) and V is an auxiliary random variable. PLS guarantees protection against computationally-unbounded adversaries without using shared keys. As such, it has attracted continuous attention in the information-theoretic literature, as surveyed in, e.g., [3]–[6].

In this work, we revisit the classic WTC with an input cost constraint, and show that *two-layered coding is necessary* for achieving its secrecy-capacity. In many information-theoretic communication problems, adding an input cost constraint amounts to restricting the optimization domain in the capacity expression (e.g., the set of feasible $P_{V,X}$ in (1)). We show that this reasoning is *not* valid for cost-constrained (CC) WTCs. To do so, we characterize the CC secrecy-capacity using *two* auxiliary variables and prove that a single-auxiliary formula is strictly suboptimal. For the latter, an example of a CC WTC is provided for which a two-layered scheme strictly outperforms any single-layered code. This establishes the fundamental role of two-layered (superposition) coding for the CC WTC.

The necessity of two auxiliaries to achieve the secrecy-capacity of CC WTCs is perhaps surprising. This is evident from the fact that non-exact expressions for it have been used in recent works [7, Corollary 2], [8, Theorem 3.7]. The requirement of two auxiliaries is even more remarkable when one considers the recently established analogy between WTC and Gelfand-Pinsker (GP) channels [9] without cost constraints. Indeed, for the GP channel, the CC and the unconstrained capacities are given by the same expression up to adding the proper restriction to feasible input distributions [10]–[12]. To the best of our knowledge, the WTC is the only point-to-point communication scenario for which the capacity formula itself changes due to the addition of an input cost constraint.

A. Background

The secrecy-capacity of a degraded WTC was established in [1], under the so-called weak-secrecy criterion[1].

[1]Weak-secrecy refers to a vanishing information leakage rate $\frac{1}{n}I(M;Z^n)$ as the blocklength $n \to \infty$, where M is secret message and Z^n is the eavesdropper's observation.

Manuscript received April 8, 2020; revised November 5, 2020; accepted November 13, 2020. Date of publication November 23, 2020; date of current version February 17, 2021. The work of Sreejith Sreekumar was supported by the Transdisciplinary Research in Principles of Data Science (TRIPODS) Center for Data Science National Science Foundation under Grant CCF-1740822. The work of Ziv Goldfeld was supported in part by the National Science Foundation under Grant CCF-1947801 and in part by the 2020 IBM Academic Award. The work of Haim H. Permuter was supported in part by the WIN Consortium via the Israel Ministry of Economy and Science, in part by the Deutsche Forschungsgemeinschaft (DFG) via the Deutsch-Israelische Projektkooperation (DIP), in part by the Israel Science Foundation, and in part by the Cyber Center at Ben-Gurion University of the Negev. The work of Shlomo Shamai was supported in part by the WIN Consortium via the Israel Ministry of Economy and Science and in part by the European Union's Horizon 2020 Research and Innovation Programme under Grant 694630. *(Corresponding author: Sreejith Sreekumar.)*

Sreejith Sreekumar and Ziv Goldfeld are with the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY 14850 USA (e-mail: sreejithsreekumar@cornell.edu; goldfeld@cornell.edu).

Alexander Bunin is with the Advanced Flash Solutions Laboratory, Samsung, Ramat Gan 5251003, Israel (e-mail: albun@tx.technion.ac.il).

Shlomo Shamai is with the Department of Electrical Engineering, Technion–Israel Institute of Technology, Haifa 3200003, Israel (e-mail: sshlomo@ee.technion.ac.il).

Haim H. Permuter is with the Department of Electrical and Computer Engineering, Ben-Gurion University of the Negev, Be'er-Sheva 8410501, Israel (e-mail: haimp@bgu.ac.il).

Communicated by L. Wang, Associate Editor for Shannon Theory.

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TIT.2020.3039916.

Digital Object Identifier 10.1109/TIT.2020.3039916

The secrecy-capacity of the general WTC was proved in [2]. The main result of [2] in fact accounted for a generalization of the WTC to the broadcast channel (BC) with confidential messages. The secrecy-capacity of the latter is characterized using a pair of auxiliary random variables, and the direct proof uses superposition wiretap coding. It was also shown therein that reducing the BC result to the WTC (by nullifying the common message) and requiring perfect secrecy on the private message, one of the auxiliaries can be taken to be a constant[2]. This gave rise to the secrecy-capacity formula given in (1).

Interestingly, this formula is quite robust to the security metric being used. In recent years, upgrading weak-secrecy to more stringent metrics gained much popularity. Strong-secrecy removes the normalization by the blocklength from the weak-secrecy metric, requiring that the information leakage itself vanishes [13], [14]. A further strengthening to the semantic-security metric was introduced in [15]. While weak- and strong-secrecy both assume a uniform distribution on the messages (i.e., security on average), semantic-security demands a vanishing information leakage for all probability distributions over the message set (i.e., worst-case). Despite these increasingly stringent requirement, the secrecy-capacity with strong and semantic-security metrics remains unchanged [16, Theorem 17.11] [15] compared to (1).

In practice, transmitted signals are often bound to cost (e.g., power) constraints. Therefore, various communication scenarios originally explored without such constraints were later adapted to the CC case. This includes point-to-point channels [17, Chapter 7], the GP channel [11], and the multiple-access channel [18, Problem 4.8], to name a few. Several different types of cost constraints exist in the literature such as average cost, maximal (per-codeword) cost, and peak power constraint (for continuous alphabet input channels). For all these settings, the capacity under a CC is given by the same expression as in the unconstrained case, but with an added restriction on the optimization domain (see [16], [19] and the references above).[3] As will be shown herein, such a simple adaptation of (1) to the CC case is not valid for the WTC.

B. Contributions

We consider a discrete and memoryless (DM) WTC with an input cost constraint and establish a single-letter characterization of its secrecy-capacity. In contrast to (1), our characterization uses two auxiliary random variables, which we show are necessary in general. As discussed above, this differentiates the WTC from other channel coding setups, where an introduction of an additional auxiliary is not needed when imposing an input cost constraint on the operational problem. We consider all three aforementioned security metrics, that is, weak-secrecy, strong-secrecy and semantic-security, and show that the secrecy-capacity is the same for them all. This is done by proving achievability under semantic-security (strongest among the three), while deriving the converse with respect to weak-secrecy.

The achievability proof uses a superposition wiretap code that carries the entire confidential message in its outer layer. The inner layer encodes only random bits purposed to confuse the eavesdropper. The cost, reliability and security analyses rely on standard random coding arguments. However, due to the presence of a cost constraint, the expected value analysis (over the codebook ensemble) does not automatically imply the existence of a deterministic codes sequence with the desired performance. We resolve this issue via a novel two-step expurgation technique that first prunes 'bad' codebooks, and only then disposes of 'bad' messages. A careful analysis shows that the inflicted rate loss is negligible, giving rise to a deterministic codebook that satisfies the desired cost, reliability and security requirements.

We then turn to show that two auxiliaries are necessary to achieve the CC WTC secrecy-capacity. This is done by constructing an example for which superposition coding attains a strictly higher secrecy rate than standard wiretap coding. The necessity of two auxiliaries can be understood by viewing the inner layer auxiliary as a "time-sharing" variable that leaks no information about the message to the eavesdropper. In a time-shared scheme, the cost constraint needs to be satisfied only on average (over the participating schemes). Thus, individual schemes could possibly violate the cost constraint, and indeed, it may be beneficial to consider such schemes for achieving higher secrecy rates. In particular, such a situation could occur if the mutual information term $I(V;Y)-I(V;Z)$ from the secrecy-capacity expression in (1) is a convex function over the CC optimization domain.

C. Organization

The remainder of this paper is organized as follows. Section II provides preliminary definitions and sets up the operational problem. The main results are stated and discussed in Section III, while their proofs are furnished in Section IV. Finally, concluding remarks are given in Section V.

II. Preliminaries and Problem Setup

A. Notation

We use the following notation. $\mathbb{N}$, $\mathbb{R}$ and $\mathbb{R}_{\geq 0}$ denotes the set of natural numbers, real numbers and non-negative real numbers, respectively. For $a, b \in \mathbb{R}_{\geq 0}$, $[a : b] := \{n \in \mathbb{N} : \ a \leq n \leq b\}$. Calligraphic letters, e.g., $\mathcal{X}$, denote sets while $|\mathcal{X}|$ stands for its cardinality. For $n \in \mathbb{N}$, $\mathcal{X}^n$ denotes the n-fold Cartesian product of $\mathcal{X}$, and $x^n = (x_1, \cdots, x_n)$ denotes an element of $\mathcal{X}^n$. Whenever the dimension n is clear from the context, bold-face letters denotes vectors or sequences, e.g., $\mathbf{x}$ for x^n. For $i, j \in \mathbb{N}$ such that $i \leq j$, $x_i^j := (x_i, x_{i+1}, \cdots, x_j)$; the subscript is omitted when $i = 1$.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, where Ω, $\mathcal{F}$ and $\mathbb{P}$ are the sample space, σ-algebra and probability measure, respectively. Random variables over $(\Omega, \mathcal{F}, \mathbb{P})$ are denoted by uppercase letters, e.g., X, with similar conventions as above for random vectors. We use $\mathbb{1}_{\mathcal{A}}$ for the indicator function of $\mathcal{A} \in \mathcal{F}$. The set of all probability mass functions (PMFs) on

[2]Namely, U in [2, Corollary 2] can be taken to be a constant u^*, where $u^* = \arg\max_{u \in \mathcal{U}} I(V;Y|U=u) - I(V;Z|U=u)$. As we later show, this argument cannot be applied in the CC WTC setting, as taking a constant U may violate the cost constraint.

[3]We also mention that, apart from secrecy-capacity, other notions of the secrecy-cost trade-off exists, such as capacity per unit cost [20]–[23].

a finite set $\mathcal{X}$ (always endowed with the power set σ-algebra) is denoted by $\mathcal{P}(\mathcal{X})$.

The joint PMF of two discrete random variables X and Y on $(\Omega, \mathcal{F}, \mathbb{P})$ is denoted by $P_{X,Y}$; the corresponding marginals are P_X and P_Y. The conditional PMF of X given Y is represented by $P_{X|Y}$. Expressions such as $P_{X,Y} = P_X P_{Y|X}$ are to be understood as pointwise equality, i.e., $P_{X,Y}(x,y) = P_X(x)P_{Y|X}(y|x)$, for all $(x,y) \in \mathcal{X} \times \mathcal{Y}$. When the joint distribution of a triple (X,Y,Z) factors as $P_{X,Y,Z} = P_{X,Y}P_{Z|Y}$, these variables form a Markov chain $X - Y - Z$. When X and Y are statistically independent, we write $X \perp\!\!\!\perp Y$. If the entries of X^n are drawn in an independent and identically distributed (i.i.d.) manner, i.e., if $P_{X^n}(x^n) = \prod_{i=1}^n P_X(x_i)$, $\forall\, x^n \in \mathcal{X}^n$, then the PMF P_{X^n} is denoted by $P_X^{\otimes n}$. Similarly, if $P_{Y^n|X^n}(y^n|x^n) = \prod_{i=1}^n P_{Y|X}(y_i|x_i)$, then we write $P_{Y|X}^{\otimes n}$ for $P_{Y^n|X^n}$. The conditional product PMF given a fixed $x^n \in \mathcal{X}^n$ is designated by $P_{Y|X}^{\otimes n}(\cdot|x^n)$.

For a discrete measurable space $(\mathcal{X}, \mathcal{F})$, the probability measure induced by a PMF $P \in \mathcal{P}(\mathcal{X})$ is denoted by $\mathbb{P}_P$; namely $\mathbb{P}_P(\mathcal{A}) = \sum_{x \in \mathcal{A}} P(x)$, for all $A \in \mathcal{F}$. The corresponding expectation is designated by $\mathbb{E}_P$. Similarly, mutual information and entropy with an underlying PMF P are denoted as I_P and H_P, respectively. When the PMF is clear from the context, the subscript is omitted. We use $\mathcal{T}_\delta^{(n)}(P)$ to denote the set of letter-typical sequences of length n with respect to a PMF $P \in \mathcal{P}(\mathcal{X})$ and a non-negative δ:

$$\mathcal{T}_\delta^{(n)}(P) := \{\mathbf{x} \in \mathcal{X}^n : |\nu_{\mathbf{x}}(x) - P(x)| \le \delta P(x),\ \forall x \in \mathcal{X}\}, \quad (2)$$

where $\nu_{\mathbf{x}}(x) := \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{\{x_i = x\}}$ is the empirical PMF of sequence $\mathbf{x} \in \mathcal{X}^n$. Finally, for a countable sample space $\mathcal{X}$ and PMFs $P, Q \in \mathcal{P}(\mathcal{X})$, the Kullback-Leibler (KL) divergence between P and Q is

$$\mathsf{D}_{\mathsf{KL}}(P\|Q) := \sum_{x \in \mathcal{X}} P(x) \log\left(\frac{P(x)}{Q(x)}\right), \quad (3)$$

and the total variation is

$$\delta_{\mathsf{TV}}(P,Q) := \frac{1}{2}\sum_{x \in \mathcal{X}} \big|P(x) - Q(x)\big|. \quad (4)$$

B. Problem Setup

Let $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{Z}$ be finite sets, $b \ge 0$ and $n \in \mathbb{N}$. Let $\mathsf{C} : \mathcal{X} \to \mathbb{R}_{\ge 0}$ be a real-valued non-negative function. The $(\mathcal{X}, \mathcal{Y}, \mathcal{Z}, P_{Y,Z|X}, \mathsf{C}, b)$ CC DM-WTC (henceforth referred to as CC WTC) is shown in Fig. 1, where $P_{Y,Z|X} : \mathcal{X} \to \mathcal{P}(\mathcal{Y} \times \mathcal{Z})$ is the channel transition kernel, C is the cost function and b is the cost constraint. The encoder chooses a message $m \in \mathcal{M}_n := [1 : 2^{nR}]$, $R \ge 0$, and maps it onto a channel input sequence $\mathbf{x} \in \mathcal{X}^n$. The codeword $\mathbf{x}$ is transmitted over the n-fold WTC $P_{Y,Z|X}^{\otimes n}$, which outputs sequences $\mathbf{y} \in \mathcal{Y}^n$ and $\mathbf{z} \in \mathcal{Z}^n$. The decoder observes $\mathbf{y}$, based on which it produces an estimate $\hat{m} \in \mathcal{M}_n$ of m. The eavesdropper observes $\mathbf{z}$, from which it tries to extract information about the transmitted message m.

Definition 1 (Code): An (n,R)-code c_n for a CC WTC $(\mathcal{X}, \mathcal{Y}, \mathcal{Z}, P_{Y,Z|X}, \mathsf{C}, b)$ with a message set $\mathcal{M}_n$ is a pair of functions (f_n, g_n) such that

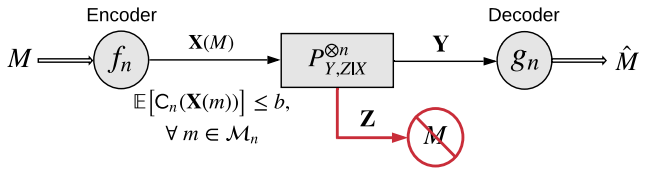

Fig. 1. The CC WTC with transition kernel $P_{Y,Z|X}$. The encoder outputs $\mathbf{X}(m)$ for message $M = m$, and the decoder and the eavesdropper observes $\mathbf{Y}$ and $\mathbf{Z}$, respectively. The cost constraint $\mathbb{E}[\mathsf{C}_n(\mathbf{X}(m))] \le b$ is imposed for each message $m \in \mathcal{M}_n$.

(i) $f_n : \mathcal{M}_n \to \mathcal{P}(\mathcal{X}^n)$ is a stochastic encoder that satisfies the per-message cost constraint given by

$$\mathbb{E}[\mathsf{C}_n(\mathbf{X}(m))] := \sum_{\mathbf{x} \in \mathcal{X}^n} f_n(\mathbf{x}|m)\mathsf{C}_n(\mathbf{x}) \le b,\ \forall\, m \in \mathcal{M}_n, \quad (5)$$

where, $\mathbf{X}(m) \sim f_n(\cdot|m)$, and $\mathsf{C}_n(\mathbf{x}) := \frac{1}{n}\sum_{i=1}^n \mathsf{C}(x_i)$, $\forall\, \mathbf{x} \in \mathcal{X}^n$, is the n-fold extension of C;

(ii) $g_n : \mathcal{Y}^n \to \mathcal{M}_n$ is the decoding function.

Remark 1 (Minimal Cost): The per-message constraint in (5) can be satisfied only if $b \ge \mathsf{c}_{\min} := \min\{\mathsf{C}(x) : x \in \mathcal{X}\}$. Henceforth, we will assume this condition holds without further mention.

A message PMF $P_M \in \mathcal{P}(\mathcal{M}_n)$ and a code $c_n = (f_n, g_n)$ induces a PMF on $\mathcal{M}_n \times \mathcal{X}^n \times \mathcal{Y}^n \times \mathcal{Z}^n \times \mathcal{M}_n$ given by

$$P^{(c_n)}(m, \mathbf{x}, \mathbf{y}, \mathbf{z}, \hat{m}) = P_M(m)\ f_n(\mathbf{x}|m)\ P_{Y,Z|X}^{\otimes n}(\mathbf{y}, \mathbf{z}|\mathbf{x})\ \mathbb{1}_{\{\hat{m} = g_n(\mathbf{y})\}}. \quad (6)$$

The performance of c_n is evaluated in terms of the maximal decoding error probability and a chosen security metric.

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

$$e(c_n) := \max_{m \in \mathcal{M}_n} e_m(c_n), \quad (7a)$$

where for any $m \in \mathcal{M}_n$,

$$\begin{aligned} e_m(c_n) &:= \mathbb{P}_{P^{(c_n)}}\big(\hat{M} \ne m \big| M = m\big) \\ &= \sum_{\mathbf{x} \in \mathcal{X}^n} f_n(\mathbf{x}|m) \sum_{\substack{\mathbf{y} \in \mathcal{Y}^n : \\ g_n(\mathbf{y}) \ne m}} P_{Y|X}^{\otimes n}(\mathbf{y}|\mathbf{x}). \end{aligned} \quad (7b)$$

Definition 3 (Security Metrics): For a given (n,R)-code c_n and message PMF P_M, the information leakage to the eavesdropper is $\ell(P_M, c_n) := I_{P^{(c_n)}}(M; \mathbf{Z})$. The weak-secrecy, strong-secrecy and semantic-security metrics with respect to (w.r.t.) the code c_n are, respectively, defined as

$$\ell_{\mathsf{weak}}(c_n) := \frac{1}{n}\ell(\bar{P}_M, c_n) \quad (8)$$

$$\ell_{\mathsf{str}}(c_n) := \ell(\bar{P}_M, c_n) \quad (9)$$

$$\ell_{\mathsf{sem}}(c_n) := \max_{P_M \in \mathcal{P}(\mathcal{M}_n)} \ell(P_M, c_n), \quad (10)$$

where $\bar{P}_M$ denotes the uniform distribution on $\mathcal{M}_n$.

Definition 4 (Achievability): A rate $R \ge 0$ is said to be achievable with semantic-security for a CC WTC

$(\mathcal{X}, \mathcal{Y}, \mathcal{Z}, P_{Y,Z|X}, \mathsf{C}, b)$ if for every $\epsilon > 0$ and sufficiently large n, there exists an (n, R)-code c_n such that (5) is satisfied, and

$$\max\left\{e(c_n), \ell_{\mathsf{sem}}(c_n)\right\} \leq \epsilon. \tag{11}$$

Achievability w.r.t. the weak- or strong-secrecy metrics is defined by replacing ℓ_{sem} with ℓ_{weak} or ℓ_{str}, respectively.

Definition 5 (Secrecy Capacity): The semantic-security capacity of a CC WTC $(\mathcal{X}, \mathcal{Y}, \mathcal{Z}, P_{Y,Z|X}, \mathsf{C}, b)$, denoted by $C_{\mathsf{sem}}(P_{Y,Z|X}, \mathsf{C}, b)$, is the supremum of the set of all rates achievable with semantic-security. The strong-secrecy-capacity $C_{\mathsf{str}}(P_{Y,Z|X}, \mathsf{C}, b)$ and weak-secrecy-capacity $C_{\mathsf{weak}}(P_{Y,Z|X}, \mathsf{C}, b)$ are defined similarly w.r.t. the corresponding notion of achievability.

III. Main Results

We give a single-letter characterization of the weak-secrecy, strong-secrecy and semantic-security capacities of the CC WTC, all of which are shown to be equal. The characterization involves two auxiliary random variables. Both auxiliaries are necessary to achieve capacity in general. We first state the capacity result, and then provide an example for which any single-auxiliary scheme is suboptimal.

A. Secrecy-Capacity Results

Let $\mathcal{U}$ and $\mathcal{V}$ be finite sets, and for any $P_{U,V,X} \in \mathcal{P}(\mathcal{U} \times \mathcal{V} \times \mathcal{X})$, set

$$\tilde{C}(P_{U,V,X}, P_{Y,Z|X}) := I_P(V;Y|U) - I_P(V;Z|U), \tag{12}$$

where the mutual information terms are taken w.r.t. $P = P_{U,V,X} P_{Y,Z|X}$. Also, let

$$\mathcal{H}(\mathsf{C}, b) := \left\{ P_{U,V,X} \in \mathcal{P}(\mathcal{U} \times \mathcal{V} \times \mathcal{X}) \colon \begin{matrix} P_{U,V,X} = P_{U,V} P_{X|V}, \\ \mathbb{E}_P\big[\mathsf{C}(X)\big] \leq b \end{matrix} \right\}, \tag{13}$$

and define

$$\bar{C}(P_{Y,Z|X}, \mathsf{C}, b) := \sup_{P_{U,V,X} \in \mathcal{H}(\mathsf{C},b)} \tilde{C}(P_{Y,Z|X}, P_{U,V,X}). \tag{14}$$

Our main result is given next. For simplicity of presentation, we will suppress $P_{Y,Z|X}$ and C from the notation $C_{\mathsf{sem}}(P_{Y,Z|X}, \mathsf{C}, b)$, $C_{\mathsf{str}}(P_{Y,Z|X}, \mathsf{C}, b)$, $C_{\mathsf{weak}}(P_{Y,Z|X}, \mathsf{C}, b)$, $\tilde{C}(P_{Y,Z|X}, P_{U,V,X})$, $\mathcal{H}(\mathsf{C}, b)$ and $\bar{C}(P_{Y,Z|X}, \mathsf{C}, b)$, henceforth denoting them by $C_{\mathsf{sem}}(b)$, $C_{\mathsf{str}}(b)$, $C_{\mathsf{weak}}(b)$, $\tilde{C}(P_{U,V,X})$, $\mathcal{H}(b)$ and $\bar{C}(b)$, respectively.

Theorem 1 (Secrecy-Capacity): The secrecy-capacity of a CC WTC $(\mathcal{X}, \mathcal{Y}, \mathcal{Z}, P_{Y,Z|X}, \mathsf{C}, b)$ under weak-secrecy, strong-secrecy and semantic-security is the same, and is given by

$$C_{\mathsf{sem}}(b) = C_{\mathsf{str}}(b) = C_{\mathsf{weak}}(b) = \bar{C}(b). \tag{15}$$

The proof of Theorem 1 is given in Section IV-A. The achievability of (15) relies on a superposition wiretap coding, while the converse adapts the classic WTC converse to accommodate the cost constraint. We note that the CC WTC's secrecy-capacity expression involves two auxiliary random variables U and V. In contrast, the secrecy-capacity formula of a WTC (without a cost constraint), given in (1), uses only a single auxiliary. In Section III-B we show that a reduction of $\bar{C}(b)$ to a single auxiliary is impossible, in general.

Remark 2 (Per-Codeword Cost Constraint): In addition to the per-message cost constraint in (5), Theorem 1 can be extended to the more stringent scenario of a per-codeword cost constraint, i.e.,

$$\mathsf{C}_n(\mathbf{x}) \leq b, \forall\, \mathbf{x} \in \mathcal{X}^n \text{ s.t. } \exists\, m \in \mathcal{M}_n \text{ with } f_n(\mathbf{x}|m) > 0. \tag{16}$$

In other words, re-defining achievability by replacing (5) with (16), the same $\bar{C}(b)$ is obtained as the per-codeword cost secrecy-capacity under weak-secrecy, strong-secrecy and semantic security. The argument relies on a natural modification of the achievability proof of Theorem 1, by employing constant-composition superposition wiretap codes instead of the currently used i.i.d. ensemble (the converse follows from the per-message cost case). A central ingredient for the modified security analysis is a version of Lemma 3 for the constant-composition ensemble, which can be obtained by extending the results in [24] to superposition codes.

The following lemma provides additional properties of $\bar{C}(b)$. These properties are used in the proof of Theorem 1.

Lemma 1 (Structural Properties): In the definition of $\bar{C}(b)$ in (14), it suffices to consider auxiliary alphabets $\mathcal{U}$ and $\mathcal{V}$ with $|\mathcal{U}| \leq |\mathcal{X}|$ and $|\mathcal{V}| \leq |\mathcal{X}|^2$. Moreover, $\bar{C}(b)$ is a non-decreasing and concave (for $b \geq \mathsf{c}_{\min}$) function of b , and the supremum in (14) is achieved, i.e.,

$$\begin{aligned} \bar{C}(b) &= \max_{P_{U,V,X} \in \mathcal{H}(b)} \tilde{C}(P_{U,V,X}) \\ &:= \max_{P_{U,V,X} \in \mathcal{H}(b)} I_P(V;Y|U) - I_P(V;Z|U). \end{aligned} \tag{17}$$

The proof of Lemma 1 is provided in Appendix A for completeness.

B. Two Auxiliaries Are Necessary

Comparing (1) and (17), one might ask whether a reduction to a single auxiliary random variable in Theorem 1 is possible. We show that the answer is negative in general. To this end, we provide an example of a CC WTC $P_{Y,Z|X}$, for which

$$\begin{aligned} &\max_{P_{U,V,X} \in \mathcal{H}(b)} I_P(V;Y|U) - I_P(V;Z|U) \\ &\qquad > \max_{\substack{P_{V,X}: \\ \mathbb{E}_P[\mathsf{C}(X)] \leq b}} I_P(V;Y) - I_P(V;Z), \end{aligned} \tag{18}$$

where the mutual information terms on the right hand side (RHS) are w.r.t. $P_{V,X} P_{Y,Z|X}$. To explain briefly, the example incorporates a WTC setup in which the transmitter (encoder) is connected to the receiver (decoder) via a noiseless private data link. The transmitter can choose the content as well as timing of the transmission, however, it is constrained to use the link at most half of the time. The receiver observes the data (error-free) when transmission occurs, and random noise otherwise. On the other hand, the eavesdropper has no access to the data link, but perfectly knows the timing of the transmission. We next describe the details of the WTC setup.

Consider the $(\mathcal{X}, \mathcal{Y}, \mathcal{Z}, P_{Y,Z|X}, \mathsf{C}, b)$ CC WTC shown in Fig. 2 that is defined as follows:

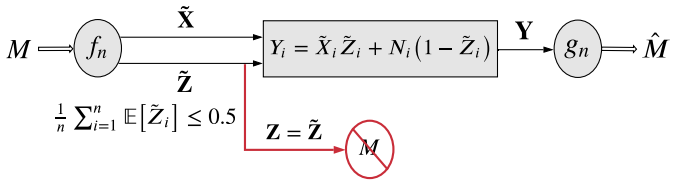

Fig. 2. The CC WTC with transition kernel $P_{Y,Z|X}$ used in Proposition 1. The encoder outputs $\mathbf{X} = (\tilde{\mathbf{X}}, \tilde{\mathbf{Z}})$ such that the cost constraint $\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(\tilde{Z}_i) \le 0.5$ is satisfied. The eavesdropper's observation $\mathbf{Z} = \tilde{\mathbf{Z}}$ is controlled by the encoder, and the realization of $\tilde{Z}_i$ decides whether the decoder observes $\tilde{X}_i$ (when $\tilde{Z}_i = 1$) or noise N_i (when $\tilde{Z}_i = 0$).

- Let $\tilde{\mathcal{X}} = \mathcal{Y} = \mathcal{Z} = \{0,1\}$, and $\mathcal{X} = \tilde{\mathcal{X}} \times \mathcal{Z}$.
- The channel input is $X = (\tilde{X}, \tilde{Z}) \sim P_{\tilde{X},\tilde{Z}}$, where $\tilde{X}$ and $\tilde{Z}$ take values in $\tilde{\mathcal{X}}$ and $\mathcal{Z}$, respectively, and both are controlled by the encoder.
- Consider the cost function $\mathsf{C}(x) = \mathsf{C}(\tilde{x}, \tilde{z}) = \tilde{z}$, for all $x = (\tilde{x}, \tilde{z}) \in \tilde{\mathcal{X}} \times \mathcal{Z}$, and set the cost constraint to $b = 0.5$. Thus, the input must satisfy
$$\mathbb{E}\big[\mathsf{C}(X)\big] = \mathbb{E}\big[\tilde{Z}\big] \le \frac{1}{2}.$$
- Let $N \sim \mathsf{Ber}(0.5)$ be independent of $X = (\tilde{X}, \tilde{Z})$ and set
$$Y = \tilde{X}\tilde{Z} + N(1 - \tilde{Z}).$$
Let $P_{Y|X} = P_{Y|\tilde{X},\tilde{Z}}$ denote the transition kernel from X to Y induced by the above relation.
- The WTC $P_{Y,Z|X}$ is given by $P_{Y,Z|\tilde{X},\tilde{Z}} = P_{Y|\tilde{X},\tilde{Z}} \mathbb{1}_{\{Z=\tilde{Z}\}}$.

We have the following proposition whose proof is given in Section IV-B.

Proposition 1 (Necessity of Two Auxiliaries): For the $(\mathcal{X}, \mathcal{Y}, \mathcal{Z}, P_{Y,Z|X}, \mathsf{C}, 0.5)$ CC WTC described above,
$$\begin{aligned}\bar{C}(0.5) &:= \max_{P_{U,V,X} \in \mathcal{H}(0.5)} I_P(V;Y|U) - I_P(V;Z|U) \\ &\ge 0.5 > \max_{\substack{P_{V,X}: \\ \mathbb{E}_P[\mathsf{C}(X)] \le 0.5}} I_P(V;Y) - I_P(V;Z). \end{aligned} \tag{19}$$

For proving Proposition 1, we choose a $P_{U,V,X} \in \mathcal{H}(0.5)$ such that $I_P(V;Y|U) - I_P(V;Z|U) = 0.5$, thus establishing $\bar{C}(0.5) \ge 0.5$. This is done by selecting $U = \tilde{Z} \sim \mathsf{Ber}(0.5)$, $\tilde{X} \sim \mathsf{Ber}(0.5) \perp\!\!\!\perp \tilde{Z}$, and $V = X := (\tilde{X}, \tilde{Z})$. Intuitively, such a choice of auxiliaries correspond to a communication scheme of transmitting data half of the time (say, every alternate channel use). Subsequently, we show that the RHS of (19) is strictly below 0.5. This is shown by starting with the assumption that there exists a $P_{V,X}$ such that $\mathbb{E}_P[\mathsf{C}(X)] \le 0.5$ and $I_P(V;Y) - I_P(V;Z) \ge 0.5$, and then arguing that it leads to a contradiction.

Achieving the CC WTC secrecy-capacity in the above example requires two auxiliary random variables. While a reduction in the number of auxiliaries is impossible in general, we next show that no auxiliaries are needed when the WTC is less noisy. Namely, the condition is that Y is *less noisy* than Z, i.e., $I_P(U;Y) \ge I_P(U;Z)$, for all $P_{U,X,Y,Z} = P_{U,X} P_{Y,Z|X}$ [25]. This is similar to the state of affairs for WTC without a cost constraint [2].

Corollary 1 (Less Noisy WTCs): If Y is less noisy than Z, then
$$\bar{C}(b) = \max_{P_X:\ \mathbb{E}_P[\mathsf{C}(X)] \le b} I_P(X;Y) - I_P(X;Z). \tag{20}$$
If Z is less noisy than Y, then $\bar{C}(b) = 0$.

The proof of Corollary 1 is provided in Section IV-C.

IV. Proofs

A. Proof of Theorem 1

We will show that $C_{\mathsf{weak}}(b) \le \bar{C}(b)$ and $C_{\mathsf{sem}}(b) \ge \bar{C}(b)$ for any $b \ge 0$. This combined with the fact that $C_{\mathsf{sem}}(b) \le C_{\mathsf{str}}(b) \le C_{\mathsf{weak}}(b)$ will imply the desired result.

1) Converse: Recall that $\bar{P}_M$ denotes the uniform distribution on $\mathcal{M}_n$. It suffices to show that for any $\delta > 0$, a rate R achievable under the weak-secrecy metric satisfies $R \le \bar{C}(b) + \delta$, for large enough n. Fix $\epsilon > 0$ and let c_n be an (n, R)-code with $\max\{e(c_n), \ell_{\mathsf{weak}}(c_n)\} \le \epsilon$. Any such code must also satisfy the weaker constraint $\max\{\mathbb{E}_{\bar{P}_M}[e_M(c_n)], \frac{1}{n}\ell(\bar{P}_M, c_n)\} \le \epsilon$. Accordingly, we may assume without loss of generality that $P_M = \bar{P}_M$. From Fano's inequality [26], it follows that
$$H(M|\mathbf{Y}) \le 1 + \epsilon n R. \tag{21}$$

Now, we can write
$$\begin{aligned}
nR &= H(M) \\
&\overset{(a)}{\le} I(M;\mathbf{Y}) + 1 + \epsilon n R \\
&\overset{(b)}{\le} I(M;\mathbf{Y}) - I(M;\mathbf{Z}) + 1 + \epsilon n R + n\epsilon \\
&= \sum_{i=1}^{n} I(M;Y_i|Y^{i-1}) - I(M;Z_i|Z_{i+1}^n) + 1 + \epsilon n(R+1) \\
&\overset{(c)}{=} \sum_{i=1}^{n} I(M;Y_i|Y^{i-1}) - I(M;Z_i|Z_{i+1}^n) + \epsilon n(R+1) \\
&\qquad + I(Z_{i+1}^n;Y_i|M,Y^{i-1}) - I(Y^{i-1};Z_i|M,Z_{i+1}^n) + 1 \\
&= \sum_{i=1}^{n} I(M,Z_{i+1}^n;Y_i|Y^{i-1}) - I(M,Y^{i-1};Z_i|Z_{i+1}^n) \\
&\qquad + 1 + \epsilon n(R+1) \\
&\overset{(d)}{=} \sum_{i=1}^{n} I(M;Y_i|Y^{i-1},Z_{i+1}^n) - I(M;Z_i|Y^{i-1},Z_{i+1}^n) \\
&\qquad + 1 + \epsilon n(R+1) \\
&= \sum_{i=1}^{n} I(M,Y^{i-1},Z_{i+1}^n;Y_i|Y^{i-1},Z_{i+1}^n) + 1 + \epsilon n(R+1) \\
&\qquad - I(M,Y^{i-1},Z_{i+1}^n;Z_i|Y^{i-1},Z_{i+1}^n) \\
&\overset{(e)}{=} \sum_{i=1}^{n} I(V_i;Y_i|U_i) - I(V_i;Z_i|U_i) + 1 + \epsilon n(R+1)
\end{aligned}$$

$$\overset{(f)}{\leq} \sum_{i=1}^{n} \bar{C}\Big(\mathbb{E}\big[\mathsf{C}(X_i)\big]\Big) + 1 + \epsilon n(R+1)$$

$$\overset{(g)}{\leq} n\bar{C}\left(\sum_{i=1}^{n} \frac{1}{n}\mathbb{E}\big[\mathsf{C}(X_i)\big]\right) + 1 + \epsilon n(R+1)$$

$$\overset{(h)}{=} n\left(\bar{C}(b) + \epsilon(R+1) + \frac{1}{n}\right),$$

where

(a) follows from (21);
(b) is because $l(\bar{P}_M, c_n) \leq n\epsilon$;
(c) and (d) use the Csiszár-sum identity [18];
(e) is due to the auxiliary random variable identification $U_i = (Y^{i-1}, Z_{i+1}^n)$ and $V_i = (M, Y^{i-1}, Z_{i+1}^n)$;
(f) follows from the definition of $\bar{C}(\cdot)$ in (14) since $U_i - V_i - X_i - (Y_i, Z_i)$ form a Markov chain with the above auxiliary variable identification;
(g) is due to the concavity of $\bar{C}(\cdot)$ proved in Lemma 1;
(h) is because $\bar{C}(\cdot)$ is non-decreasing and X^n satisfies $\mathbb{E}\Big[\frac{1}{n}\sum_{i=1}^{n}\mathsf{C}(X_i)\Big] \leq b$ due to (5).

Thus,

$$R \leq \frac{\bar{C}(b) + \epsilon + \frac{1}{n}}{1-\epsilon}.$$

The claim follows by taking n sufficiently large and $\epsilon > 0$ small enough.

2) Achievability: By the continuity[4] of $\bar{C}(b)$, it suffices to show that for any $b > \mathsf{c}_{\min}$ and $\epsilon > 0$, there exists an (n, R) code c_n that satisfies (11), provided that $R < \bar{C}(b)$ and n is sufficiently large. To this end, we construct an ensemble of superposition wiretap codes and show that the expected (over the ensemble) cost, error probability and semantic-security metric satisfy average versions of the constraints. Then, through a sequence of codebook and message expurgation steps, we show the existence of a code c_n that satisfies (5) and (11), as required.

Fix $\epsilon > 0$ and a joint PMF $P_{U,V,X,Y,Z} := P_{U,V}P_{X|V}P_{Y,Z|X} \in \mathcal{P}(\mathcal{U}\times\mathcal{V}\times\mathcal{X}\times\mathcal{Y}\times\mathcal{Z})$ such that $\mathbb{E}_P\big[\mathsf{C}(X)\big] < b$.

Codebook $\mathcal{B}_n$: We use a superposition codebook such that the inner layer adds redundancy to confuse the eavesdropper, while the outer layer carries the information about the message.

Define the index sets $\mathcal{I}_n := [1 : 2^{nR_1}]$ and $\mathcal{J}_n := [1 : 2^{nR_2}]$. Let $\mathbb{B}_U^{(n)} := \{\mathbf{U}(i),\ i \in \mathcal{I}_n\}$ be a random inner layer codebook such that each codeword $\mathbf{U}(i),\ i \in \mathcal{I}_n$, is a sequence of length n generated independently according to $P_U^{\otimes n}$. Denote a realization of $\mathbb{B}_U^{(n)}$ by $\mathcal{B}_U^{(n)} := \{\mathbf{u}(i),\ i \in \mathcal{I}_n\}$.

For a fixed $\mathcal{B}_U^{(n)}$ and each $i \in \mathcal{I}_n$, let $\mathbb{B}_V^{(n)}(i) := \{\mathbf{V}(i,j,m), (j,m) \in \mathcal{J}_n \times \mathcal{M}_n\}$ denote a collection of n-length random sequences, each generated independently according to $P_{V|U}^{\otimes n}\big(\cdot\big|\mathbf{u}(i)\big)$. Denote a realization of $\mathbb{B}_V^{(n)}(i)$ by $\mathcal{B}_V^{(n)}(i) := \{\mathbf{v}(i,j,m), (j,m) \in \mathcal{J}_n \times \mathcal{M}_n\}$. Also, set $\mathbb{B}_V^{(n)} := \{\mathbb{B}_V^{(n)}(i),\ i \in \mathcal{I}_n\}$, and denote its possible outcome by $\mathcal{B}_V^{(n)}$. With the above, the random superposition codebook is given by $\mathbb{B}_n := \{\mathbb{B}_U^{(n)}, \mathbb{B}_V^{(n)}\}$, and its realization is denoted by $\mathcal{B}_n$.

[4]This follows from the concavity of $\bar{C}(b)$; see Lemma 1.

Denoting the set of all possible values of $\mathbb{B}_n$ by $\mathfrak{B}_n$, the codebook construction described above induces a PMF $\mu \in \mathcal{P}(\mathfrak{B}_n)$, given by

$$\mu(\mathcal{B}_n) := \mu\left(\mathcal{B}_U^{(n)}, \mathcal{B}_V^{(n)}\right)$$

$$:= \prod_{i\in\mathcal{I}_n}\left[P_U^{\otimes n}\big(\mathbf{u}(i)\big) \prod_{(j,m)\in\mathcal{J}_n\times\mathcal{M}_n} P_{V|U}^{\otimes n}\big(\mathbf{v}(i,j,m)\big|\mathbf{u}(i)\big)\right].$$

Encoder f_n: Given a codebook $\mathcal{B}_n$ and message $M = m$, the encoder chooses an index pair (i,j) uniformly at random from the set $\mathcal{I}_n \times \mathcal{J}_n$, and transmits $\mathbf{X} \sim P_{\mathbf{X}|\mathbf{V}}^{\otimes n}\big(\cdot\big|\mathbf{v}(i,j,m)\big)$. The induced encoding function $f_n : \mathcal{M}_n \to \mathcal{P}(\mathcal{X}^n)$ is given

$$f_n(\mathbf{x}|m) = \frac{1}{|\mathcal{I}_n||\mathcal{J}_n|}\sum_{(i,j)\in\mathcal{I}_n\times\mathcal{J}_n} P_{\mathbf{X}|\mathbf{V}}^{\otimes n}\big(\mathbf{x}\big|\mathbf{v}(i,j,m)\big),$$
$$\forall\ (m,\mathbf{x}) \in \mathcal{M}_n \times \mathcal{X}^n. \tag{22}$$

Decoder g_n: Upon observing $\mathbf{y} \in \mathcal{Y}^n$, the decoder looks for a unique tuple $(\hat{i},\hat{j},\hat{m}) \in \mathcal{I}_n \times \mathcal{J}_n \times \mathcal{M}_n$ such that $\Big(\mathbf{u}(\hat{i}), \mathbf{v}(\hat{i},\hat{j},\hat{m}), \mathbf{y}\Big) \in \mathcal{T}_\delta^{(n)}(P_{U,V,Y})$, for some $\delta > 0$. If such a unique tuple exists, the decoder sets $g_n(\mathbf{y}) = \hat{m}$; else, $g_n(\mathbf{y}) = 1$.

Induced distribution: Denote the pair (f_n, g_n) w.r.t. the codebook $\mathcal{B}_n$ by $c_n(\mathcal{B}_n)$. For a given codebook $\mathcal{B}_n$, $c_n(\mathcal{B}_n)$ induces a joint PMF $P_{M,I,J,\mathbf{U},\mathbf{V},\mathbf{X},\mathbf{Y},\mathbf{Z},\hat{M}}^{(\mathcal{B}_n)} \in \mathcal{P}(\mathcal{M}_n \times \mathcal{I}_n \times \mathcal{J}_n \times \mathcal{U}^n \times \mathcal{V}^n \times \mathcal{X}^n \times \mathcal{Y}^n \times \mathcal{Z}^n \times \mathcal{M}_n)$, given by

$$P_{M,I,J,\mathbf{U},\mathbf{V},\mathbf{X},\mathbf{Y},\mathbf{Z},\hat{M}}^{(\mathcal{B}_n)}(m,i,j,\mathbf{u},\mathbf{v},\mathbf{x},\mathbf{y},\mathbf{z},\hat{m})$$
$$= P_M(m)\frac{1}{|\mathcal{I}_n||\mathcal{J}_n|}\mathbb{1}_{\{\mathbf{u}=\mathbf{u}(i),\mathbf{v}=\mathbf{v}(i,j,m)\}}P_{X|V}^{\otimes n}(\mathbf{x}|\mathbf{v})$$
$$P_{YZ|X}^{\otimes n}(\mathbf{y},\mathbf{z}|\mathbf{x})\mathbb{1}_{\{\hat{m}=g_n(\mathbf{y})\}}. \tag{23}$$

Henceforth, we use $P^{(\mathcal{B}_n)}$ and $P^{(\mathbb{B}_n)}$ as shorthands for $P_{M,I,J,\mathbf{U},\mathbf{V},\mathbf{X},\mathbf{Y},\mathbf{Z},\hat{M}}^{(\mathcal{B}_n)}$ and $P_{M,I,J,\mathbf{U},\mathbf{V},\mathbf{X},\mathbf{Y},\mathbf{Z},\hat{M}}^{(\mathbb{B}_n)}$, respectively. We will also denote the probability measure induced by $P^{(\mathcal{B}_n)}$ and $P^{(\mathbb{B}_n)}$ by $\mathbb{P}_{P^{(\mathcal{B}_n)}}$ and $\mathbb{P}_{P^{(\mathbb{B}_n)}}$, respectively. Note that $P^{(\mathbb{B}_n)}$ is a random PMF and $\mathbb{P}_{P^{(\mathbb{B}_n)}}$ is a random probability measure.

Cost Analysis: We analyze the expectation (w.r.t. the random codebook) of the cost averaged over messages. For any $m \in \mathcal{M}_n$, note that $\mathbb{E}_\mu\Big[P_{\mathbf{X}|M}^{(\mathbb{B}_n)}(\mathbf{x}|m)\Big] = P_X^{\otimes n}(\mathbf{x})$, $\forall\ \mathbf{x} \in \mathcal{X}^n$, which readily implies that

$$\mathbb{E}_\mu\Big[\mathbb{E}_{P_{\mathbf{X}|M}^{(\mathbb{B}_n)}(\cdot|m)}\big[\mathsf{C}_n(\mathbf{X})\big]\Big] = \mathbb{E}_{P_X^{\otimes n}}\big[\mathsf{C}_n(\mathbf{X})\big]$$
$$= \mathbb{E}_{P_X}\big[\mathsf{C}(X)\big] < b.$$

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

$$\mathbb{E}_\mu\Big[\mathbb{E}_{P_{\mathbf{X}}^{(\mathbb{B}_n)}}\big[\mathsf{C}_n(\mathbf{X})\big]\Big] \leq b - \gamma'. \tag{24}$$

Average error probability analysis: We analyze the expected error probability averaged over messages. For any $\mathcal{B}_n \in \mathfrak{B}_n$ and $(i,j,m) \in \mathcal{I}_n \times \mathcal{J}_n \times \mathcal{M}_n$, let $\mathbf{Y} \sim P_{\mathbf{Y}|I,J,M}^{(\mathcal{B}_n)}(\cdot|i,j,m)$, and define the following error

events:

$$\mathcal{E}_1(i,j,m) := \Big\{\big(\mathbf{u}(i),\mathbf{v}(i,j,m),\mathbf{Y}\big) \notin \mathcal{T}_\delta^{(n)}(P_{U,V,Y})\Big\}$$

$$\mathcal{E}_2(i,j,m) := \Big\{\exists (j',m') \in \mathcal{J}_n \times \mathcal{M}_n,\ (j',m') \neq (j,m) : \big(\mathbf{u}(i),\mathbf{v}(i,j',m'),\mathbf{Y}\big) \in \mathcal{T}_\delta^{(n)}(P_{U,V,Y})\Big\}$$

$$\mathcal{E}_3(i,j,m) := \Big\{\exists \big(i',j',m'\big) \in \mathcal{I}_n \times \mathcal{J}_n \times \mathcal{M}_n,\ i' \neq i : \big(\mathbf{u}(i'),\mathbf{v}\big(i',j',m'\big),\mathbf{Y}\big) \in \mathcal{T}_\delta^{(n)}(P_{U,V,Y})\Big\}.$$

Due to the symmetry of the random codebook $\mathbb{B}_n$, encoder f_n and decoder g_n, the expected error probability over $\mathbb{B}_n$, i.e., $\mathbb{E}_\mu\Big[\mathbb{P}_{P^{(\mathbb{B}_n)}}(\hat{M} \neq M)\Big]$, is the same for any realization of (I,J,M). Thus, we may assume without loss of generality that $(I,J,M) = (1,1,1)$. We have

$$\begin{aligned}
&\mathbb{E}_\mu\Big[\mathbb{P}_{P^{(\mathbb{B}_n)}}(\hat{M} \neq M)\Big] \\
&= \mathbb{E}_\mu\Big[\mathbb{P}_{P^{(\mathbb{B}_n)}}\big(\hat{M} \neq M \big| (I,J,M) = (1,1,1)\big)\Big] \\
&\leq \mathbb{E}_\mu\Big[\mathbb{P}_{P^{(\mathbb{B}_n)}}\big(\mathcal{E}_1(1,1,1) \cup \mathcal{E}_2(1,1,1) \cup \mathcal{E}_3(1,1,1)\big| \\
&\qquad (I,J,M) = (1,1,1)\big)\Big]. \qquad (25)
\end{aligned}$$

To upper bound the the RHS of (25), we use the following lemma whose proof is given in Appendix B.

Lemma 2: If $(R,R_1,R_2) \in \mathbb{R}^3_{\geq 0}$ satisfy

$$R_2 + R < I_P(V;Y|U) \qquad (26)$$

$$R_1 + R_2 + R < I_P(U,V;Y), \qquad (27)$$

then there exists a $\zeta(\delta) > 0$ such that

$$\mathbb{E}_\mu\Big[\mathbb{P}_{P^{(\mathbb{B}_n)}}\big(\mathcal{E}_1(1,1,1) \cup \mathcal{E}_2(1,1,1) \cup \mathcal{E}_3(1,1,1)\big| (I,J,M) = (1,1,1)\big)\Big] \leq e^{-n\zeta(\delta)}. \qquad (28)$$

Thus, from (25) and (28), it follows that

$$\mathbb{E}_\mu\Big[\mathbb{P}_{P^{(\mathbb{B}_n)}}\big(\hat{M} \neq M\big)\Big] \leq e^{-n\zeta(\delta)} \underset{n}{\longrightarrow} 0, \qquad (29)$$

provided (26) and (27) holds.

Security analysis: First consider

$$\begin{aligned}
&I_{P^{(\mathcal{B}_n)}}(M;\mathbf{Z}) \\
&\leq I_{P^{(\mathcal{B}_n)}}(M;I,\mathbf{U},\mathbf{Z}) \\
&= \mathsf{D}_{\mathsf{KL}}\left(P^{(\mathcal{B}_n)}_{M,I,\mathbf{U},\mathbf{Z}}\middle\| P^{(\mathcal{B}_n)}_M P^{(\mathcal{B}_n)}_{I,\mathbf{U},\mathbf{Z}}\right) \\
&= \mathsf{D}_{\mathsf{KL}}\left(P^{(\mathcal{B}_n)}_{M,I,\mathbf{U},\mathbf{Z}}\middle\| P^{(\mathcal{B}_n)}_{M,I,\mathbf{U}} P^{(\mathcal{B}_n)}_{\mathbf{Z}|I,\mathbf{U}}\right) \\
&\overset{(a)}{=} \mathsf{D}_{\mathsf{KL}}\left(P^{(\mathcal{B}_n)}_{M,I,\mathbf{U}} P^{(\mathcal{B}_n)}_{\mathbf{Z}|M,I,\mathbf{U}}\middle\| P^{(\mathcal{B}_n)}_{M,I,\mathbf{U}} P^{\otimes n}_{Z|U}\right) \\
&\qquad - \mathsf{D}_{\mathsf{KL}}\left(P^{(\mathcal{B}_n)}_{I,\mathbf{U},\mathbf{Z}}\middle\| P^{(\mathcal{B}_n)}_{I,\mathbf{U}} P^{\otimes n}_{Z|U}\right) \\
&\overset{(b)}{\leq} \mathsf{D}_{\mathsf{KL}}\left(P^{(\mathcal{B}_n)}_{M,I,\mathbf{U}} P^{(\mathcal{B}_n)}_{\mathbf{Z}|M,I,\mathbf{U}}\middle\| P^{(\mathcal{B}_n)}_{M,I,\mathbf{U}} P^{\otimes n}_{Z|U}\right) \qquad (30) \\
&\overset{(c)}{\leq} \max_{m\in\mathcal{M}_n} \mathsf{D}_{\mathsf{KL}}\left(P^{(\mathcal{B}_n)}_{I,\mathbf{U}} P^{(\mathcal{B}_n)}_{\mathbf{Z}|M,I,\mathbf{U}}(\cdot|m,\cdot,\cdot)\middle\| P^{(\mathcal{B}_n)}_{I,\mathbf{U}} P^{\otimes n}_{Z|U}\right), \qquad (31)
\end{aligned}$$

where,

(a) and (c) is since $M \perp\!\!\!\perp (I,\mathbf{U})$;

(b) uses the non-negativity of KL divergence.

Maximizing w.r.t. $P_M \in \mathcal{P}(\mathcal{M}_n)$ on both sides of (31), we obtain that

$$\begin{aligned}
&\max_{P_M\in\mathcal{P}(\mathcal{M}_n)} I_{P^{(\mathcal{B}_n)}}(M;\mathbf{Z}) \\
&\leq \max_{m\in\mathcal{M}_n} \mathsf{D}_{\mathsf{KL}}\left(P^{(\mathcal{B}_n)}_{I,\mathbf{U}} P^{(\mathcal{B}_n)}_{\mathbf{Z}|M,I,\mathbf{U}}(\cdot|m,\cdot,\cdot)\middle\| P^{(\mathcal{B}_n)}_{I,\mathbf{U}} P^{\otimes n}_{Z|U}\right). \qquad (32)
\end{aligned}$$

Note that $P^{(\mathcal{B}_n)}_{I,\mathbf{U}} P^{(\mathcal{B}_n)}_{\mathbf{Z}|M,I,\mathbf{U}}(\cdot|m,\cdot,\cdot) \ll P^{(\mathcal{B}_n)}_{I,\mathbf{U}} P^{\otimes n}_{Z|U}$, where $\ll$ denotes absolute continuity of measures. For PMFs P,Q with a finite support such that $P \ll Q$, $\mathsf{D}_{\mathsf{KL}}(P\|Q)$ can be upper bounded by the total variation distance $\delta_{\mathsf{TV}}(P,Q)$ [27, Equation 30]. Defining

$$\theta(m,\mathcal{B}_n) := \delta_{\mathsf{TV}}\left(P^{(\mathcal{B}_n)}_{I,\mathbf{U}} P^{(\mathcal{B}_n)}_{\mathbf{Z}|M,I,\mathbf{U}}(\cdot|m,\cdot,\cdot), P^{(\mathcal{B}_n)}_{I,\mathbf{U}} P^{\otimes n}_{Z|U}\right),$$

and applying [28, Lemma 9], we obtain

$$\begin{aligned}
&\max_{m\in\mathcal{M}_n} \mathsf{D}_{\mathsf{KL}}\left(P^{(\mathcal{B}_n)}_{I,\mathbf{U}} P^{(\mathcal{B}_n)}_{\mathbf{Z}|M,I,\mathbf{U}}(\cdot|m,\cdot,\cdot)\middle\| P^{(\mathcal{B}_n)}_{I,\mathbf{U}} P^{\otimes n}_{Z|U}\right) \\
&\leq \max_{m\in\mathcal{M}_n} \theta(m,\mathcal{B}_n)\Big(n\log|\mathcal{Z}| - \log\theta(m,\mathcal{B}_n) \\
&\qquad + n\log P^{(\mathsf{min})}_{Z|U}\Big), \qquad (33)
\end{aligned}$$

where

$$P^{(\mathsf{min})}_{Z|U} := \min\{P_{Z|U}(z|u),\ (z,u)\in\mathcal{Z}\times\mathcal{U} : P_{Z|U}(z|u) > 0\}.$$

Thus, showing that there exist $\mathcal{B}_n \in \mathfrak{B}_n$ and $\gamma_1 > 0$ such that $\max_{m\in\mathcal{M}_n}\theta(m,\mathcal{B}_n) \leq e^{-n\gamma_1}$ for large enough n, is sufficient (by (32) and (33)) to get

$$\max_{P_M\in\mathcal{P}(\mathcal{M}_n)} I_{P^{(\mathcal{B}_n)}}(M;\mathbf{Z}) \underset{n}{\longrightarrow} 0. \qquad (34)$$

The existence of such a $\mathcal{B}_n$ is implied by the following lemma. The lemma restates the outcome of the secrecy analysis from [29], providing a double-exponential bound on the probability of an exponentially small deviation of $\max_{m\in\mathcal{M}_n}\theta(m,\mathbb{B}_n)$ from zero.

Lemma 3 (Lemma 4 From [29]): If

$$R_2 > I_P(V;Z|U), \qquad (35)$$

then there exists $\gamma_1,\gamma_2 > 0$ such that for all sufficiently large n,

$$\mathbb{P}_\mu\left(\max_{m\in\mathcal{M}_n}\theta(m,\mathbb{B}_n) > e^{-n\gamma_1}\right) \leq e^{-e^{n\gamma_2}}. \qquad (36)$$

Lemma 3 follows from the proof of Lemma 4 in [29], which is a stronger version of the superposition soft-covering lemma [27]. The double-exponential bound in (36) is an implication of Chernoff bound applied to the collection of an exponential number of i.i.d. codewords in the random superposition codebook.

Summary of random coding argument: Combining (24), (29) and (36), we have shown that

$$\mathbb{E}_\mu\Big[\mathbb{E}_{P^{(\mathbb{B}_n)}_{\mathbf{X}}}\big[\mathsf{C}_n(\mathbf{X})\big]\Big] \leq b' := b - \gamma', \qquad (37a)$$

$$\mathbb{E}_\mu\Big[\mathbb{P}_{P^{(\mathbb{B}_n)}}(\hat{M} \neq M)\Big] \leq e^{-n\zeta(\delta)}, \qquad (37b)$$

$$\mathbb{E}_\mu\Big[\mathbb{1}_{\{\max_{m\in\mathcal{M}_n}\theta(m,\mathbb{B}_n) > e^{-n\gamma_1}\}}\Big] \leq e^{-e^{n\gamma_2}}, \qquad (37c)$$

provided (26), (27) and (35) hold.

We next perform a sequence of expurgation steps: first, w.r.t. codebooks and then w.r.t. messages. At the end of this process, we deduce the existence of a single codebook $\mathcal{B}_n$ that satisfies (5) and (11). Note that the selection lemma of [30] or [29] is not applicable here as the RHS of (37a) is a constant which does not vanish to zero as required by the lemma.

Expurgation: For any $\mathfrak{B}'_n \subseteq \mathfrak{B}_n$, let

$$\bar{\mu}(\mathfrak{B}'_n) := \sum_{\mathcal{B}_n \in \mathfrak{B}'_n} \mu(\mathcal{B}_n) \tag{38}$$

be the probability measure on $\mathfrak{B}_n$ induced by μ. Our expurgation technique on the codebooks $\mathcal{B}_n \in \mathfrak{B}_n$ is performed for each fixed n (sufficiently large) as described in the following steps:

1) **Codebook expurgation to satisfy average (over messages) cost:** Expurgate codebooks $\mathcal{B}_n \in \mathfrak{B}_n$ with the highest cost $\mathbb{E}_{P_{\mathbf{X}}^{(\mathcal{B}_n)}}\big[\mathsf{C}_n(\mathbf{X})\big]$ to obtain a set $\mathfrak{B}'_n \subset \mathfrak{B}_n$ such that $\frac{1}{n+2} \leq \bar{\mu}(\mathfrak{B}'_n) < \frac{1}{n+1}$. This is possible for large n since each codebook $\mathcal{B}_n$ has exponentially small probability. We now show that all the codebooks $\mathcal{B}_n \in \mathfrak{B}'_n$ satisfy $\mathbb{E}_{P_{\mathbf{X}}^{(\mathcal{B}_n)}}\big[\mathsf{C}_n(\mathbf{X})\big] \leq \frac{n+1}{n} b'$. Assume otherwise that there exists $\mathcal{B}_n^* \in \mathfrak{B}'_n$ such that $\mathbb{E}_{P_{\mathbf{X}}^{(\mathcal{B}_n^*)}}\big[\mathsf{C}_n(\mathbf{X})\big] > \frac{n+1}{n} b'$. Then, we can write

$$\begin{aligned}
\mathbb{E}_{\mu}\Big[\mathbb{E}_{P_{\mathbf{X}}^{(\mathbb{B}_n)}}\big[\mathsf{C}_n(\mathbf{X})\big]\Big] &= \sum_{\mathcal{B}_n \in \mathfrak{B}_n} \mu(\mathcal{B}_n) \mathbb{E}_{P_{\mathbf{X}}^{(\mathcal{B}_n)}}\big[\mathsf{C}_n(\mathbf{X})\big] \\
&\geq \sum_{\mathcal{B}_n \in \mathfrak{B}_n \setminus \mathfrak{B}'_n} \mu(\mathcal{B}_n) \mathbb{E}_{P_{\mathbf{X}}^{(\mathcal{B}_n)}}\big[\mathsf{C}_n(\mathbf{X})\big] \\
&\overset{(a)}{\geq} \sum_{\mathcal{B}_n \in \mathfrak{B}_n \setminus \mathfrak{B}'_n} \mu(\mathcal{B}_n) \mathbb{E}_{P_{\mathbf{X}}^{(\mathcal{B}_n^*)}}\big[\mathsf{C}_n(\mathbf{X})\big] \\
&> \sum_{\mathcal{B}_n \in \mathfrak{B}_n \setminus \mathfrak{B}'_n} \mu(\mathcal{B}_n) \frac{n+1}{n} b' \\
&\overset{(b)}{\geq} \frac{n}{n+1} \frac{n+1}{n} b' = b',
\end{aligned} \tag{39}$$

where

(a) is because by the expurgation procedure, $\mathbb{E}_{P_{\mathbf{X}}^{(\mathcal{B}_n)}}\big[\mathsf{C}_n(\mathbf{X})\big] \geq \mathbb{E}_{P_{\mathbf{X}}^{(\mathcal{B}_n^*)}}\big[\mathsf{C}_n(\mathbf{X})\big]$ for every $\mathcal{B}_n \in \mathfrak{B}_n \setminus \mathfrak{B}'_n$;

(b) is since $\bar{\mu}(\mathfrak{B}_n \setminus \mathfrak{B}'_n) \geq \frac{n}{n+1}$.

Eqn. (39) contradicts (37a), and hence $\mathbb{E}_{P_{\mathbf{X}}^{(\mathcal{B}_n)}}\big[\mathsf{C}_n(\mathbf{X})\big] \leq \frac{n+1}{n} b'$ should hold for all $\mathcal{B}_n \in \mathfrak{B}'_n$.

Define a PMF $\mu_1 \in \mathcal{P}(\mathfrak{B}_n)$ and its induced probability measure $\bar{\mu}_1$ on $\mathfrak{B}_n$ by

$$\mu_1(\mathcal{B}_n) := \begin{cases} \frac{\mu(\mathcal{B}_n)}{\bar{\mu}(\mathfrak{B}'_n)}, & \text{if } \mathcal{B}_n \in \mathfrak{B}'_n, \\ 0, & \text{otherwise,} \end{cases}$$

and

$$\bar{\mu}_1(\tilde{\mathfrak{B}}_n) := \sum_{\mathcal{B}_n \in \tilde{\mathfrak{B}}_n} \mu_1(\mathcal{B}_n), \ \tilde{\mathfrak{B}}_n \subseteq \mathfrak{B}_n,$$

respectively. Then, we have

$$\begin{aligned}
&\mathbb{E}_{P_{\mathbf{X}}^{(\mathcal{B}_n)}}\big[\mathsf{C}_n(\mathbf{X})\big] \leq \frac{n+1}{n} b', \ \forall\ \mathcal{B}_n \in \mathfrak{B}'_n, \\
&\mathbb{E}_{\mu_1}\big[\mathbb{P}_{P^{(\mathbb{B}_n)}}(\hat{M} \neq M)\big] \leq \frac{1}{\bar{\mu}(\mathfrak{B}'_n)} \mathbb{E}_{\mu}\big[\mathbb{P}_{P^{(\mathbb{B}_n)}}(\hat{M} \neq M)\big] \\
&\qquad\qquad \leq \frac{e^{-n\zeta(\delta)}}{\bar{\mu}(\mathfrak{B}'_n)} \leq (n+2) e^{-n\zeta(\delta)}, \\
&\mathbb{E}_{\mu_1}\Big[\mathbb{1}_{\{\max_{m \in \mathcal{M}_n} \theta(m, \mathbb{B}_n) > e^{-n\gamma_1}\}}\Big] \\
&\leq \frac{1}{\bar{\mu}(\mathfrak{B}'_n)} \mathbb{E}_{\mu}\Big[\mathbb{1}_{\{\max_{m \in \mathcal{M}_n} \theta(m, \mathbb{B}_n) > e^{-n\gamma_1}\}}\Big] \\
&\leq \frac{e^{-e^{n\gamma_2}}}{\bar{\mu}(\mathfrak{B}'_n)} \leq (n+2) e^{-e^{n\gamma_2}},
\end{aligned}$$

2) **Codebook expurgation to satisfy average cost, average error probability and semantic-security:** Expurgate codebooks $\mathcal{B}_n \in \mathfrak{B}'_n$ with the highest average error probability to obtain a set $\mathfrak{B}''_n \subset \mathfrak{B}'_n$ such that $\frac{1}{3} \leq \bar{\mu}_1(\mathfrak{B}''_n) < \frac{1}{2}$. Then, it follows similarly to step 1 that

$$\mathbb{P}_{P^{(\mathcal{B}_n)}}(\hat{M} \neq M) \leq 2(n+2) e^{-n\zeta(\delta)}, \ \forall\ \mathcal{B}_n \in \mathfrak{B}''_n.$$

Define another PMF $\mu_2 \in \mathcal{P}(\mathfrak{B}_n)$ by

$$\mu_2(\mathcal{B}_n) := \begin{cases} \frac{\mu_1(\mathcal{B}_n)}{\bar{\mu}_1(\mathfrak{B}''_n)}, & \text{if } \mathcal{B}_n \in \mathfrak{B}''_n, \\ 0, & \text{otherwise.} \end{cases}$$

Then, we have

$$\begin{aligned}
&\mathbb{E}_{P_{\mathbf{X}}^{(\mathcal{B}_n)}}\big[\mathsf{C}_n(\mathbf{X})\big] \leq \frac{n+1}{n} b', \ \forall\ \mathcal{B}_n \in \mathfrak{B}''_n, \\
&\mathbb{P}_{P^{(\mathcal{B}_n)}}(\hat{M} \neq M) \leq 2(n+2) e^{-n\zeta(\delta)}, \ \forall\ \mathcal{B}_n \in \mathfrak{B}''_n, \\
&\mathbb{E}_{\mu_2}\Big[\mathbb{1}_{\{\max_{m \in \mathcal{M}_n} \theta(m, \mathbb{B}_n) > e^{-n\gamma_1}\}}\Big] \leq 3(n+2) e^{-e^{n\gamma_2}}.
\end{aligned}$$

Perform one more expurgation step similar to the previous step to obtain a non-empty set of codebooks $\mathfrak{B}'''_n$ such that for each codebook $\mathcal{B}_n \in \mathfrak{B}'''_n$ and sufficiently large n,

$$\mathbb{E}_{P_{\mathbf{X}}^{(\mathcal{B}_n)}}\big[\mathsf{C}_n(\mathbf{X})\big] \leq \frac{n+1}{n} b', \ \forall\ \mathcal{B}_n \in \mathfrak{B}'''_n, \tag{40}$$

$$\mathbb{P}_{P^{(\mathcal{B}_n)}}(\hat{M} \neq M) \leq 2(n+2) e^{-n\zeta(\delta)}, \ \forall\ \mathcal{B}_n \in \mathfrak{B}'''_n, \tag{41}$$

$$\max_{m \in \mathcal{M}_n} \theta(m, \mathcal{B}_n) \leq e^{-n\gamma_1}, \ \forall\ \mathcal{B}_n \in \mathfrak{B}'''_n. \tag{42}$$

3) **Message expurgation to satisfy per-message cost, maximal error probability and semantic-security:** Now, fixing a codebook $\mathcal{B}_n \in \mathfrak{B}'''_n$, perform expurgation on the set of messages $\mathcal{M}_n$ to obtain upper bounds on the per-message cost and maximal error probability, in place of the average (over messages) cost and average (over messages) error probability given in (40) and (41), respectively. Note that (40) and (41) hold for any message distribution P_M. Consider $P_M = \bar{P}_M$. Let $\alpha \in [\frac{1}{n+2}, \frac{1}{n+1})$. Similar to step 1, by expurgating a $(1-\alpha)$ fraction of the messages $m \in \mathcal{M}_n$ with the highest cost $\mathbb{E}_{P_{\mathbf{X}|M}^{(\mathcal{B}_n)}(\cdot|m)}\big[\mathsf{C}_n(\mathbf{X})\big]$ to obtain a set $\mathcal{M}'_n \subset \mathcal{M}_n$, and defining for all $(m, i, j, \mathbf{u}, \mathbf{v}, \mathbf{x}, \mathbf{y}, \mathbf{z}, \hat{m}) \in \mathcal{M}_n \times \mathcal{I}_n \times \mathcal{J}_n \times \mathcal{U}^n \times \mathcal{V}^n \times \mathcal{X}^n \times \mathcal{Y}^n \times \mathcal{Z}^n \times \hat{\mathcal{M}}_n$, a PMF $\tilde{P}^{(\mathcal{B}_n)}$

given by

$$\tilde{P}^{(\mathcal{B}_n)}_{M,I,J,\mathbf{U},\mathbf{V},\mathbf{X},\mathbf{Y},\mathbf{Z},\hat{M}}(m,i,j,\mathbf{u},\mathbf{v},\mathbf{x},\mathbf{y},\mathbf{z},\hat{m}) := \begin{cases} \frac{|\mathcal{M}_n|}{|\mathcal{M}'_n|} P^{(\mathcal{B}_n)}_{M,I,J,\mathbf{U},\mathbf{V},\mathbf{X},\mathbf{Y},\mathbf{Z},\hat{M}}(m,i,j,\mathbf{u},\mathbf{v},\mathbf{x},\mathbf{y},\mathbf{z},\hat{m}), & \text{if } m \in \mathcal{M}'_n, \\ 0, & \text{otherwise,} \end{cases} \tag{43}$$

it follows that

$$\mathbb{E}_{\tilde{P}^{(\mathcal{B}_n)}_{\mathbf{X}|M}(\cdot|m)}\big[\mathsf{C}_n(\mathbf{X})\big] \le \frac{(n+1)^2}{n^2} b',\ \forall\ m \in \mathcal{M}'_n,$$
$$\mathbb{P}_{\tilde{P}^{(\mathcal{B}_n)}}(\hat{M} \ne M) \le 2(n+2)^2 e^{-n\zeta(\delta)},$$
$$\max_{m \in \mathcal{M}'_n} \theta(m, \mathcal{B}_n) \le e^{-n\gamma_1}.$$

Finally, for $\beta \in [\frac{1}{3}, \frac{1}{2})$, expurgating a $1-\beta$ fraction of the messages $m \in \mathcal{M}'_n$ with the highest error probability $\mathbb{P}_{\tilde{P}^{(\mathcal{B}_n)}}(\hat{M} \ne m | M = m)$ and denoting the resulting set by $\mathcal{M}''_n$ ($\mathcal{M}''_n \subset \mathcal{M}'_n$), it follows similar to step 3 that

$$\mathbb{E}_{\tilde{P}^{(\mathcal{B}_n)}_{\mathbf{X}|M}(\cdot|m)}\big[\mathsf{C}_n(\mathbf{X})\big] \le \frac{(n+1)^2}{n^2} b',\ \forall\ m \in \mathcal{M}''_n, \tag{44}$$

$$\mathbb{P}_{\tilde{P}^{(\mathcal{B}_n)}_{\hat{M}|M}(\cdot|m)}(\hat{M} \ne m) \le 4(n+2)^2 e^{-n\zeta(\delta)},\ \forall m \in \mathcal{M}''_n, \tag{45}$$

$$\max_{m \in \mathcal{M}''_n} \theta(m, \mathcal{B}_n) \le e^{-n\gamma_1}. \tag{46}$$

Summary of expurgation steps:

Note that $|\mathcal{M}''_n| = \beta\alpha|\mathcal{M}_n| \ge \frac{e^{nR}}{3(n+2)}$. Thus, for sufficiently large n, we have shown the existence of a codebook $\mathcal{B}_n$ and a $(n, R - \frac{1}{n}\log(3n+6))$ code $c_n(\mathcal{B}_n) = (f_n, g_n)$ with message set $\mathcal{M}''_n$, such that

$$\mathbb{E}\big[\mathsf{C}_n(\mathbf{X}(m)\big] \le \frac{(n+1)^2}{n^2} b' = b - \gamma' + \frac{2n+1}{n^2}(b-\gamma') < b,\ \forall\ m \in \mathcal{M}''_n, \tag{47}$$

and

$$\max\left\{ \max_{m \in \mathcal{M}''_n} e_m\left(c_n(\mathcal{B}_n)\right), \max_{P_M \in \mathcal{P}(\mathcal{M}''_n)} \ell\left(P_M, c_n(\mathcal{B}_n)\right) \right\} \le \epsilon, \tag{48}$$

provided (26), (27) and (35) holds.

Eliminating R_1 and R_2 from (26), (27) and (35) via the Fourier-Motzkin elimination yields $R < I_P(V;Y|U) - I_P(V;Z|U)$. The proof of (15) is completed by noting that $C_{\mathsf{sem}}(b)$ is a closed set by definition, and $I_P(V;Y|U) - I_P(V;Z|U)$ is a continuous function of P.

Remark 3 (Invariance of Secrecy-Capacity): Theorem 1 states the invariance of CC WTC secrecy-capacity to the employed secrecy metric. The converse proof further shows that capacity remains unchanged if the maximal error probability and/or per message cost constraints are relaxed to average (over messages) constraints of the form $\mathbb{E}_{\bar{P}_M}\big[e_M(c_n)\big] \le \epsilon$ and $\mathbb{E}_{\bar{P}_M}\big[\mathsf{C}_n(\mathbf{X}(M))\big] \le b$, respectively.

Remark 4 (Selection Lemma): We note that the conclusion of Step 2 of the expurgation procedure can be deduced via the Selection Lemma of [30]. However, Steps 1 and 3 seem to require the expurgation argument.

B. Proof of Proposition 1

Since the WTC transition kernel is $P_{Y,Z|\tilde{X}\tilde{Z}} = P_{Y|\tilde{X},\tilde{Z}}\mathbb{1}_{\{Z=\tilde{Z}\}}$, $Z = \tilde{Z}$ with probability one. We henceforth identify Z with $\tilde{Z}$. We start by showing that

$$\max_{P_{U,V,X} \in \mathcal{H}(0.5)} I_P(V;Y|U) - I_P(V;Z|U) \ge 0.5. \tag{49}$$

Set $U = \tilde{Z} \sim \mathsf{Ber}(0.5)$, $\tilde{X} \sim \mathsf{Ber}(0.5)$, $\tilde{X} \perp\!\!\!\perp \tilde{Z}$, and $V = X = (\tilde{X}, \tilde{Z})$. This choice satisfies $P_{U,V,X} \in \mathcal{H}(0.5)$, since $U - V - X = (\tilde{X}, \tilde{Z}) - (Y, \tilde{Z})$ and $\mathbb{E}_P\big[\mathsf{C}(X)\big] := \mathbb{E}_P\big[\tilde{Z}\big] = 0.5$. Moreover,

$$\begin{aligned} I_P(V;Y|U) - I_P(V;\tilde{Z}|U) &= I_P(X;Y|\tilde{Z}) - I_P(X;\tilde{Z}|\tilde{Z}) \\ &= I_P(X;Y|\tilde{Z}) \\ &= 0.5, \end{aligned} \tag{50}$$

where (50) is because $I_P(X;Y|\tilde{Z}) = P_{\tilde{Z}}(1)H(\tilde{X}) = 0.5$. Hence, (49) holds.

Next, we establish that

$$\max_{\substack{P_{V,X}: \\ \mathbb{E}_P[\mathsf{C}(X)] \le 0.5}} I_P(V;Y) - I_P(V;\tilde{Z}) < 0.5. \tag{51}$$

For any $P_{V,X} = P_{V,\tilde{X},\tilde{Z}}$ such that $\mathbb{E}_P\big[\tilde{Z}\big] \le 0.5$, we have the following chain of inequalities[5]:

$$\begin{aligned} &I(V;Y) - I(V;\tilde{Z}) \\ &= I(\tilde{X},\tilde{Z},V;Y) - I(\tilde{X},\tilde{Z};Y|V) - I(V;\tilde{Z}) \\ &\overset{(a)}{=} I(\tilde{X},\tilde{Z};Y) - I(\tilde{X},\tilde{Z};Y|V) - I(V;\tilde{Z}) \\ &= I(\tilde{X};Y|\tilde{Z}) - I(\tilde{X};Y|\tilde{Z},V) - I(V;\tilde{Z}|Y) \\ &\overset{(b)}{=} P_{\tilde{Z}}(0)I(\tilde{X};N|\tilde{Z}=0) + P_{\tilde{Z}}(1)I(Y;Y|\tilde{Z}=1) \\ &\quad - I(\tilde{X};Y|\tilde{Z},V) - I(V;\tilde{Z}|Y) \\ &\overset{(c)}{=} P_{\tilde{Z}}(1)H(Y|\tilde{Z}=1) - I(\tilde{X};Y|\tilde{Z},V) - I(V;\tilde{Z}|Y) \\ &\overset{(d)}{\le} 0.5H(Y|\tilde{Z}=1) - I(\tilde{X};Y|\tilde{Z},V) - I(V;\tilde{Z}|Y) \qquad (52) \\ &\overset{(e)}{\le} 0.5, \qquad (53) \end{aligned}$$

where

(a) is due to the Markov chain $V - (\tilde{X},\tilde{Z}) - (Y,\tilde{Z})$;
(b) follows from the definition of Y;
(c) is because $N \perp\!\!\!\perp (\tilde{X},\tilde{Z})$;
(d) uses the cost constraint $\mathbb{E}\big[\tilde{Z}\big] \le 0.5$;
(e) is by the non-negativity of mutual information and since Y is binary.

Thus,

$$\max_{\substack{P_{V,X}: \\ \mathbb{E}[\mathsf{C}(X)] \le 0.5}} I(V;Y) - I(V;\tilde{Z}) \le 0.5. \tag{54}$$

Consequently, (51) is violated only if there exist some $X = (\tilde{X},\tilde{Z})$ and a joint PMF $P_{V,X,Y,\tilde{Z}} = P_{V,X}P_{Y,\tilde{Z}|X}$ such

[5]We omit the subscript P in the subsequent mutual information and entropy terms as the PMF is $P = P_{V,X}P_{Y,\tilde{Z}|X}$ throughout.

that $\mathbb{E}\big[\tilde{Z}\big] \leq 0.5$, and the inequalities in (52) and (53) hold with equality, i.e.,

$$I(V;Y) - I(V;\tilde{Z}) = 0.5. \tag{55}$$

For this to be possible, the following conditions must hold:

1) $P_{\tilde{Z}}(1) = 0.5$;
2) $H(Y|\tilde{Z}=1) = 1$ which means that given $\tilde{Z}=1$, $Y \sim \mathsf{Ber}(0.5)$;
3) $I(\tilde{X};Y|\tilde{Z},V) = 0$ which implies that $\tilde{X} - (\tilde{Z},V) - Y$ forms a Markov chain;
4) $I(V;\tilde{Z}|Y) = 0$ which implies that $V - Y - \tilde{Z}$ forms a Markov chain.

Now, notice that $P_{Y|\tilde{Z}=0} = P_{Y|\tilde{Z}=1} = \mathsf{Ber}(0.5)$. Hence, $Y \perp\!\!\!\perp \tilde{Z}$, which further implies via Condition (4) that $(V,Y) \perp\!\!\!\perp \tilde{Z}$. Finally,

$$I(V;Y|\tilde{Z}=1) \geq I(\tilde{X};Y|\tilde{Z}=1) = H(Y|\tilde{Z}=1) = 1. \tag{56}$$

The inequality in (56) is due to Condition (3) above, while the last equality is due to Condition (2). Since Y is binary, $I(V;Y|\tilde{Z}=1) \leq 1$, and therefore the inequality in (56) is an equality.

To conclude, observe that $(V,Y) \perp\!\!\!\perp \tilde{Z}$ (shown above) implies that

$$I(V;Y) - I(V;\tilde{Z}) = I(V;Y|\tilde{Z}=1) - 0 = 1. \tag{57}$$

However, $I(V;Y) - I(V;\tilde{Z}) \leq 0.5$ from (53). This leads to a contradiction, and so (55) is invalid. Via (54), this implies that (51) holds. Combining (51) with (49) proves Proposition 1.

C. Proof of Corollary 1

Fix $P_{U,V,X,Y,Z} = P_{U,V}P_{X|V}P_{Y,Z|X}$, where Y is less noisy than Z. We have

$$\begin{aligned}
&I_P(V;Y|U) - I_P(V;Z|U)\\
&\overset{(a)}{=} I_P(V;Y) - I_P(V;Z) - (I_P(U;Y) - I_P(U;Z))\\
&\overset{(b)}{=} I_P(X;Y) - I_P(X;Z) - (I_P(X;Y|V) - I_P(X;Z|V))\\
&\quad - (I_P(U;Y) - I_P(U;Z))\\
&\overset{(c)}{\leq} I_P(X;Y) - I_P(X;Z),
\end{aligned}$$

where, (a) and (b) are due to the Markov chain $U - V - X - (Y,Z)$ that holds under PMF $P_{U,V,X,Y,Z}$; and (c) is due to the less noisy assumption which implies that $I_P(U;Y) - I_P(U;Z) \geq 0$ and $I_P(X;Y|V) - I_P(X;Z|V) \geq 0$ (due to $V - X - (Y,Z)$ under $P_{U,V,X,Y,Z}$).

Thus, it follows that

$$\bar{C}(b) \leq \max_{P_X:\mathbb{E}_P[\mathsf{C}(X)]\leq b} I_P(X;Y) - I_P(X;Z). \tag{58}$$

The reverse inequality follows trivially by selecting $V = X$ and $U = \emptyset$ in (17), thus proving (20).

If Z is less noisy than Y, then $I_P(V;Y|U) - I_P(V;Z|U) \leq 0$ for any $P_{U,V,X,Y,Z} = P_{U,V}P_{X|V}P_{Y,Z|X}$ which gives $\bar{C}(b) = 0$, due to its non-negativity.

V. Concluding Remarks

This paper revisited the classical wiretap channel (WTC) setting with a cost constraint, and showed that achieving its secrecy-capacity generally requires two-layer coding. To do so, we proved optimality of superposition wiretap coding under a cost constraint, and provided a WTC example for which single-layer coding is strictly suboptimal. This stands in contrast to the classic WTC secrecy-capacity result without a cost constraint, that is characterized using a single auxiliary random variable. In many other communication scenarios, imposing a cost constraint on the input amounts to a simple adaptation of the unconstrained case capacity; namely, restricting the optimization to those input distributions that satisfy the constraint in expectation. Our result provides an example where this commonly observed behavior does not hold, and a second auxiliary must be introduced as a result of the added cost constraint. Our main goal was to highlight this important fact and put forth the correct cost constrained (CC) WTC secrecy-capacity characterization, for which non-exact expressions exist in the literature.

An analogy can be drawn between the secrecy-capacity of a CC WTC and the capacity of channels with action-dependent states [31]. The capacity of the latter with transition kernels $P_{S|A}$ for the action-state pair (A,S) and $P_{Y|X,S}$ for the state-dependent channel is (see [31, Theorem 1])

$$\begin{aligned}
C_{\mathsf{ADGP}} &= \max I_P(V;Y) - I_P(V;S|A)\\
&= \max I_P(A;Y) + I_P(V;Y|A) - I_P(V;S|A),
\end{aligned} \tag{59}$$

where the maximization is taken w.r.t. the joint distribution $P_{A,S,V,X,Y} = P_A P_{S|A} P_{V|A,S} P_{X|V,S} P_{Y|X,S}$. Here, $I_P(A;Y)$ can be interpreted as the information gain at the receiver due to the choice of action A (with PMF P_A), while the remaining term in (59) can be thought of as the communication rate achievable over a GP channel $P_{Y|X,S}$ with the action-induced state S distributed as $P_S(s) = \sum_{a\in\mathcal{A}} P_A(a)P_{S|A}(s|a)$. Here, A takes the role of U in our setting, while the role of V is identical in both settings as the main information carrying auxiliary. However, while A can also be used to convey extra information to the receiver in addition to conditioning the state-dependent channel, U in the CC WTC essentially corresponds to noise added to confuse the eavesdropper, and thus does not carry any information. Accordingly, subtracting $I_P(A;Y)$ to accommodate this fact leads to the desired analogy between the capacity expressions in the two scenarios.

While this work focused on the discrete setting, an interesting future avenue is to examine the necessity of two auxiliaries for achieving the secrecy-capacity of continuous-alphabet CC WTCs. Since the canonical example of a Gaussian WTC under an average power constraint satisfies the less noisy property (either Y is less noisy than Z or vice-versa), we expect that the capacity expression without any auxiliaries presented in Corollary 1 extends to this case via existing techniques in the literature [32]–[34]. Thus, non-Gaussian channel models, and in particular, ones that cannot be classified as less noisy in general, are the interesting ones to explore. Another possible direction in the continuous channel setup is imposing a peak power constraint on the input. For point-to-point channels,

it is known that the peak-constrained capacity achieving distribution is discrete with a finite support [35]–[38]. Exploring the properties of optimal distributions when an eavesdropper is present seems like a natural extension. It would also be interesting to investigate the above questions in adversarial WTC setting [39]–[45]. A good starting point is to compare the secrecy-capacity with and without cost constraints in special cases where a single-letter solution (without cost) is known [40], [41], [45], [46]. Finally, examining the effect of a cost constraint on more complicated networks with secrecy requirements, such as broadcast channels or cloud C-RANs with rate-limited backhaul links (see, e.g., [47] and [48]) is another potential objective.

Appendix A
Proof of Lemma 1

The proof of the cardinality bounds $|\mathcal{U}| \leq |\mathcal{X}|$ and $|\mathcal{V}| \leq |\mathcal{X}|^2$ follows via a standard application of the Eggleston-Fenchel-Carathéodory Theorem [49, Theorem 18], and is omitted.

Given that $|\mathcal{U}|$ and $|\mathcal{V}|$ are finite, the set $\mathcal{H}(b)$ is non-empty (whenever $b \geq \mathsf{c}_{\min}$) and compact. Since $\bar{C}(b)$ is the supremum of a continuous function $I(V;Y|U) - I(V;Z|U)$ of $P_{U,V,X}$ over $\mathcal{H}(b)$, the supremum is achieved and thus a maximum exists. The fact that $\bar{C}(b)$ is monotonic and non-decreasing in b follows by its definition.

Finally, to show the concavity of $\bar{C}(b)$ for $b \geq \mathsf{c}_{\min}$, consider the following. For $i = 0, 1$, let $b_i \geq \mathsf{c}_{\min}$ and $P_{U_i,V_i,X_i,Y_i,Z_i} \in \mathcal{P}(\mathcal{U} \times \mathcal{V} \times \mathcal{X} \times \mathcal{Y} \times \mathcal{Z})$ be a PMF for which

$$P_{U_i,V_i,X_i} \in \mathcal{H}(b_i), \tag{60}$$

$$P_{U_i,V_i,X_i,Y_i,Z_i} := P_{U_i,V_i,X_i} P_{Y_i,Z_i|X_i} := P_{U_i,V_i,X_i} P_{Y,Z|X},$$

$$\bar{C}(b_i) := I_P(V_i;Y_i|U_i) - I_P(V_i;Z_i|U_i). \tag{61}$$

Also, let $\tau \in [0,1]$, $Q \sim \mathsf{Ber}(\tau)$ with $\mathcal{Q} = \{0,1\}$, $U_\tau := (U_Q, Q)$, $V_\tau := (V_Q, Q)$ and $X_\tau := X_Q$, and $P_{U_\tau,V_\tau,X_\tau,Y_\tau,Z_\tau} \in \mathcal{P}(\mathcal{U}_\tau \times \mathcal{V}_\tau \times \mathcal{X} \times \mathcal{Y} \times \mathcal{Z})$ be a PMF defined by

$$\begin{aligned} P_{U_\tau,V_\tau,X_\tau,Y_\tau,Z_\tau} &:= P_{U_\tau,V_\tau} P_{X_\tau|V_\tau} P_{Y_\tau,Z_\tau|X_\tau} \\ &:= P_{U_\tau,V_\tau} P_{X_\tau|V_\tau} P_{Y,Z|X}. \end{aligned} \tag{62}$$

Note that $U_\tau - V_\tau - X_\tau - (Y_\tau, Z_\tau)$ holds under $P_{U_\tau,V_\tau,X_\tau,Y_\tau,Z_\tau}$, and $P_{U_\tau,V_\tau,X_\tau} \in \mathcal{H}((1-\tau)b_0 + \tau b_1)$ (by redefining $\mathcal{H}(\cdot)$ in (13) using $\mathcal{U} \times \mathcal{Q}$ and $\mathcal{V} \times \mathcal{Q}$ in place of $\mathcal{U}$ and $\mathcal{V}$, respectively) since

$$\mathbb{E}_P\left[\mathsf{C}(X_\tau)\right] \leq (1-\tau)b_0 + \tau b_1. \tag{63}$$

Then, we have

$$\begin{aligned} &(1-\tau)\bar{C}(b_0) + \tau\bar{C}(b_1) \\ &= (1-\tau)\left(I_P(V_0;Y_0|U_0) - I_P(V_0;Z_0|U_0)\right) \\ &\quad + \tau\left(I_P(V_1;Y_1|U_1) - I_P(V_1;Z_1|U_1)\right) \\ &= I_P(V_Q,Q;Y_Q|U_Q,Q) - I_P(V_Q,Q;Z_Q|U_Q,Q) \\ &= I_P(V_\tau;Y_\tau|U_\tau) - I_P(V_\tau;Z_\tau|U_\tau) \\ &\leq \bar{C}((1-\tau)b_0 + \tau b_1), \end{aligned}$$

which establishes the concavity of $\bar{C}(\cdot)$.

Appendix B
Proof of Lemma 2

The proof is standard, however, we provide it for completeness. Let $P^{(\mu)}(\cdot) := \mathbb{E}_\mu\left[P^{(\mathbb{B}_n)}(\cdot)\right]$ and $\mathbb{P}_{P^{(\mu)}}$ denote the random coding PMF and its induced probability measure, respectively. Also, define

$$\begin{aligned} \zeta_1^{(n)}(\delta) := \inf_{\substack{\nu_{\mathbf{u},\mathbf{v},\mathbf{y}}: \\ (\mathbf{u},\mathbf{v},\mathbf{y}) \notin \mathcal{T}_\delta^{(n)}(P_{U,V,Y})}} & \mathsf{D}_{\mathsf{KL}}\left(\nu_{\mathbf{u},\mathbf{v},\mathbf{y}} \| P_{U,V,Y}\right) \\ & - |\mathcal{U}||\mathcal{V}||\mathcal{Y}| \frac{\log(n+1)}{n}, \end{aligned}$$

where $\nu_{\mathbf{u},\mathbf{v},\mathbf{y}}$ denotes the empirical PMF of $(\mathbf{u},\mathbf{v},\mathbf{y}) \in \mathcal{U}^n \times \mathcal{V}^n \times \mathcal{Y}^n$. Note that for $\delta > 0$ and n sufficiently large, $\zeta_1^{(n)}(\delta) > 0$.

First, consider the probability of the error event $\mathcal{E}_1(1,1,1)$ averaged over the random codebook $\mathbb{B}_n$. We have

$$\begin{aligned} &\mathbb{E}_\mu\left[\mathbb{P}_{P^{(\mathbb{B}_n)}}\left(\mathcal{E}_1(1,1,1)\middle|(I,J,M) = (1,1,1)\right)\right] \\ &= \mathbb{P}_{P^{(\mu)}}\left(\left(\mathbf{U}(1), \mathbf{V}(1,1,1), \mathbf{Y}\right) \notin \mathcal{T}_\delta^{(n)}(P_{U,V,Y})\right) \\ &\leq e^{-n\zeta_1^{(n)}(\delta)} \underset{n}{\longrightarrow} 0, \end{aligned} \tag{64}$$

where the inequality in (64) follows from Lemmas 2.2 and 2.6 in [16].

Next, we analyze the probability of the error event $\mathcal{E}_2(1,1,1)$ averaged over $\mathbb{B}_n$. Note that for $(j,m) \neq (1,1)$ and sufficiently large n,

$$\begin{aligned} &\mathbb{P}_{P^{(\mu)}}\Big(\left(\mathbf{U}(1), \mathbf{V}(1,j,m), \mathbf{Y}\right) \in \mathcal{T}_\delta^{(n)}(P_{U,V,Y})\Big| \\ &\qquad (I,J,M) = (1,1,1)\Big) \\ &= \sum_{(\mathbf{u},\mathbf{v},\mathbf{y}) \in \mathcal{T}_\delta^{(n)}(P_{U,V,Y})} P_{U,V}^{\otimes n}(\mathbf{u},\mathbf{v}) P_{Y|U}^{\otimes n}(\mathbf{y}|\mathbf{u}) \\ &\leq \sum_{(\mathbf{u},\mathbf{v},\mathbf{y}) \in \mathcal{T}_\delta^{(n)}(P_{U,V,Y})} e^{-n(H_P(U,V) + H_P(Y|U) - O(\delta))} \\ &\leq e^{-n(H_P(U,V) + H_P(Y|U) - H_P(U,V,Y) - O(\delta))} \\ &= e^{-n(I_P(V;Y|U) - O(\delta))}. \end{aligned}$$

Hence,

$$\begin{aligned} &\mathbb{E}_\mu\left[\mathbb{P}_{P^{(\mathbb{B}_n)}}\left(\mathcal{E}_2(1,1,1)\middle|(I,J,M) = (1,1,1)\right)\right] \\ &\leq \sum_{\substack{(j,m) \in \mathcal{J}_n \times \mathcal{M}_n: \\ (j,m) \neq (1,1)}} \mathbb{P}_{P^{(\mu)}}\Big(\left(\mathbf{U}(1), \mathbf{V}(1,j,m), \mathbf{Y}\right) \in \\ &\qquad \mathcal{T}_\delta^{(n)}(P_{U,V,Y})\Big|(I,J,M) = (1,1,1)\Big) \\ &\leq e^{n(R_2 + R - I_P(V;Y|U) + O(\delta))}. \end{aligned}$$

This implies that for δ sufficiently small and n large enough, there exists $\zeta_2(\delta) > 0$ such that

$$\begin{aligned} &\mathbb{E}_\mu\left[\mathbb{P}_{P^{(\mathbb{B}_n)}}\left(\mathcal{E}_2(1,1,1)\middle|(I,J,M) = (1,1,1)\right)\right] \\ &\leq e^{-n\zeta_2(\delta)} \underset{n}{\longrightarrow} 0, \end{aligned} \tag{65}$$

provided (26) holds.

Finally, consider the third error event $\mathcal{E}_3(1,1,1)$. We have for $i \neq 1$ and sufficiently large n that

$$\begin{aligned}
&\mathbb{P}_{P^{(\mu)}}\Big(\big(\mathbf{U}(i),\mathbf{V}(i,j,m),\mathbf{Y}\big) \in \mathcal{T}_\delta^{(n)}(P_{U,V,Y})\big| \\
&\qquad (I,J,M)=(1,1,1)\Big) \\
&= \sum_{(\mathbf{u},\mathbf{v},\mathbf{y}) \,\in\, \mathcal{T}_\delta^{(n)}(P_{U,V,Y})} P_{U,V}^{\otimes n}(\mathbf{u},\mathbf{v}) P_Y^{\otimes n}(\mathbf{y}) \\
&\leq \sum_{(\mathbf{u},\mathbf{v},\mathbf{y}) \,\in\, \mathcal{T}_\delta^{(n)}(P_{U,V,Y})} e^{-n(H_P(U,V)+H_P(Y)-O(\delta))} \\
&\leq e^{-n(H_P(U,V)+H_P(Y)-H_P(U,V,Y)-O(\delta))} \\
&= e^{-n(I_P(U,V;Y)-O(\delta))}.
\end{aligned}$$

Hence,

$$\begin{aligned}
&\mathbb{E}_\mu\Big[\mathbb{P}_{P^{(\mathbb{B}_n)}}\big(\mathcal{E}_3(1,1,1)\big|(I,J,M)=(1,1,1)\big)\Big] \\
&\leq \sum_{\substack{(i,j,m)\in \mathcal{I}_n\times\mathcal{J}_n\times\mathcal{M}_n: \\ i\neq 1}} \mathbb{P}_{P^{(\mu)}}\Big(\big(\mathbf{U}(i),\mathbf{V}(i,j,m),\mathbf{Y}\big) \in \\
&\qquad\qquad \mathcal{T}_\delta^{(n)}(P_{U,V,Y})\big|(I,J,M)=(1,1,1)\Big) \\
&\leq e^{n(R_1+R_2+R-I_P(U,V;Y)+O(\delta))}.
\end{aligned}$$

Thus, it follows that for δ sufficiently small and n large enough, there exists $\zeta_3(\delta) > 0$ such that

$$\begin{aligned}
&\mathbb{E}_\mu\Big[\mathbb{P}_{P^{(\mathbb{B}_n)}}\big(\mathcal{E}_3(1,1,1)\big|(I,J,M)=(1,1,1)\big)\Big] \\
&\leq e^{-n\zeta_3(\delta)} \underset{n}{\longrightarrow} 0, \qquad (66)
\end{aligned}$$

provided (27) holds. The claim in the lemma follows from (64), (65) and (66) via the union bound on probability applied to the left hand side of (28).

REFERENCES

[1] A. D. Wyner, "The wire-tap channel," *Bell Syst. Tech. J.*, vol. 54, no. 8, pp. 1355–1387, Oct. 1975.

[2] I. Csiszar and J. Korner, "Broadcast channels with confidential messages," *IEEE Trans. Inf. Theory*, vol. IT-24, no. 3, pp. 339–348, May 1978.

[3] A. Mukherjee, S. A. A. Fakoorian, J. Huang, and A. L. Swindlehurst, "Principles of physical layer security in multiuser wireless networks: A survey," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 3, pp. 1550–1573, 3rd Quart., 2014.

[4] Y. Wu, A. Khisti, C. Xiao, G. Caire, K.-K. Wong, and X. Gao, "A survey of physical layer security techniques for 5G wireless networks and challenges ahead," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 4, pp. 679–695, Apr. 2018.

[5] J. M. Hamamreh, H. M. Furqan, and H. Arslan, "Classifications and applications of physical layer security techniques for confidentiality: A comprehensive survey," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 2, pp. 1773–1828, 2nd Quart., 2019.

[6] Y. Liang, H. V. Poor, and S. Shamai, "Information theoretic security," *Found. Trends Commun. Inf. Theory*, vol. 5, nos. 4–5, pp. 355–580, Apr. 2009.

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

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

[9] Z. Goldfeld and H. H. Permuter, "Wiretap and Gelfand-Pinsker channels analogy and its applications," *IEEE Trans. Inf. Theory*, vol. 65, no. 8, pp. 4979–4996, Aug. 2019.

[10] S. I. Gel'fand and M. S. Pinsker, "Coding for channels with random parameters," *Probl. Control Inf. Theory*, vol. 9, no. 1, pp. 19–31, 1980.

[11] R. J. Barron, B. Chen, and G. W. Wornell, "The duality between information embedding and source coding with side information and some applications," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1159–1180, May 2003.

[12] S. S. Pradhan, J. Chou, and K. Ramchandran, "Duality between source coding and channel coding and its extension to the side information case," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1181–1203, May 2003.

[13] U. Maurer, "The strong secret key rate of discrete random triples," in *Communications and Cryptography: Two Sides One Tapestry*. Norwell, MA, USA: Springer US, 1994, pp. 271–285.

[14] I. Csiszár, "Almost independence and secrecy capacity," *Problems Inf. Transmiss.*, vol. 32, no. 1, pp. 40–47, 1996.

[15] M. Bellare, S. Tessaro, and A. Vardy, "A cryptographic treatment of the wiretap channel," in *Proc. Adv. Crypto. (CRYPTO)*, Santa Barbara, CA, USA, Aug. 2012, pp. 1–31.

[16] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Cambridge, U.K.: Cambridge Univ. Press, 2011.

[17] R. G. Gallager, *Information Theory and Reliable Communication*. New York, NY, USA: Wiley, 1968.

[18] A. E. Gamal and Y. H. Kim, *Network Information Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2011.

[19] T. S. Han, *Information-Spectrum Methods in Information Theory*. Berlin, Germany: Springer-Verlag, 2003.

[20] S. Verdu, "On channel capacity per unit cost," *IEEE Trans. Inf. Theory*, vol. 36, no. 5, pp. 1019–1030, Sep. 1990.

[21] V. V. Prelov and E. C. van der Meulen, "An asymptotic expression for the information and capacity of a multidimensional channel with weak input signals," *IEEE Trans. Inf. Theory*, vol. 39, no. 5, pp. 1728–1735, Sep. 1993.

[22] B. Hajek and V. G. Subramanian, "Capacity and reliability function for small peak signal constraints," *IEEE Trans. Inf. Theory*, vol. 48, no. 4, pp. 828–839, Apr. 2002.

[23] A. El Gamal, M. Mohseni, and S. Zahedi, "Bounds on capacity and minimum energy-per-bit for AWGN relay channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1545–1561, Apr. 2006.

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

[25] J. Körner and K. Marton, "Comparison of two noisy channels," in *Colloquia Mathematica Societatis János Bolyai*. Keszthely, Hungary: North-Holland, 1977, pp. 411–423.

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

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

[28] Z. Goldfeld, P. Cuff, and H. H. Permuter, "Wiretap channels with random states non-causally available at the encoder," 2018, *arXiv:1608.00743v2*. [Online]. Available: http://arxiv.org/abs/1608.00743v2

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

[30] M. Bloch and J. Barros, *Physical-Layer Security: From Information Theory to Security Engineering*. Cambridge, U.K.: Cambridge Univ. Press, 2011.

[31] T. Weissman, "Capacity of channels with action-dependent states," *IEEE Trans. Inf. Theory*, vol. 56, no. 11, pp. 5396–5411, Nov. 2010.

[32] S. Leung-Yan-Cheong and M. Hellman, "The Gaussian wire-tap channel," *IEEE Trans. Inf. Theory*, vol. IT-24, no. 4, pp. 451–456, Jul. 1978.

[33] H. Tyagi and A. Vardy, "Explicit capacity-achieving coding scheme for the Gaussian wiretap channel," in *Proc. IEEE Int. Symp. Inf. Theory*, Honolulu, HI, USA, Jun. 2014, pp. 956–960.

[34] C. Ling, L. Luzzi, J.-C. Belfiore, and D. Stehle, "Semantically secure lattice codes for the Gaussian wiretap channel," *IEEE Trans. Inf. Theory*, vol. 60, no. 10, pp. 6399–6416, Oct. 2014.

[35] J. G. Smith, "The information capacity of amplitude- and variance-constrained sclar Gaussian channels," *Inf. Control*, vol. 18, no. 3, pp. 203–219, Apr. 1971.

[36] S. Shamai and I. Bar-David, "The capacity of average and peak-power-limited quadrature Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 41, no. 4, pp. 1060–1071, Jul. 1995.

[37] W. Huleihel, Z. Goldfeld, T. Koch, M. Madiman, and M. Medard, "Design of discrete constellations for peak-power-limited complex Gaussian channels," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Vail, CO, USA, Jun. 2018, pp. 556–560.

[38] A. Dytso, S. Yagli, H. V. Poor, and S. S. Shitz, "The capacity achieving distribution for the amplitude constrained additive Gaussian channel: An upper bound on the number of mass points," *IEEE Trans. Inf. Theory*, vol. 66, no. 4, pp. 2006–2022, Apr. 2020.
[39] H. Boche and R. F. Schaefer, "Capacity results and super-activation for wiretap channels with active wiretappers," *IEEE Trans. Inf. Forensics Security*, vol. 8, no. 9, pp. 1482–1496, Sep. 2013.
[40] P. Wang and R. Safavi-Naini, "A model for adversarial wiretap channels," *IEEE Trans. Inf. Theory*, vol. 62, no. 2, pp. 970–983, Feb. 2016.
[41] Z. Goldfeld, P. Cuff, and H. H. Permuter, "Arbitrarily varying wiretap channels with type constrained states," *IEEE Trans. Inf. Theory*, vol. 62, no. 12, pp. 7216–7244, Dec. 2016.
[42] J. Nötzel, M. Wiese, and H. Boche, "The arbitrarily varying wiretap channel—Secret randomness, stability, and super-activation," *IEEE Trans. Inf. Theory*, vol. 62, no. 6, pp. 3504–3531, Apr. 2016.
[43] M. Wiese, J. Nötzel, and H. Boche, "A channel under simultaneous jamming and eavesdropping attack—Correlated random coding capacities under strong secrecy criteria," *IEEE Trans. Inf. Theory*, vol. 62, no. 7, pp. 3844–3862, May 2016.
[44] B. K. Dey, S. Jaggi, and M. Langberg, "Sufficiently myopic adversaries are blind," *IEEE Trans. Inf. Theory*, vol. 65, no. 9, pp. 5718–5736, Sep. 2019.
[45] M. Tahmasbi, M. R. Bloch, and A. Yener, "Learning an adversary's actions for secret communication," *IEEE Trans. Inf. Theory*, vol. 66, no. 3, pp. 1607–1624, Mar. 2020.
[46] Z. Goldfeld, P. Cuff, and H. H. Permuter, "Semantic-security capacity for wiretap channels of type II," *IEEE Trans. Inf. Theory*, vol. 62, no. 7, pp. 3863–3879, Jul. 2016.
[47] S. Zou, Y. Liang, L. Lai, H. V. Poor, and S. Shamai, "Broadcast networks with layered decoding and layered secrecy: Theory and applications," *Proc. IEEE*, vol. 103, no. 10, pp. 1841–1856, Oct. 2015.
[48] M. Zeyde, O. Sineone, and S. Shamai, "Confidential communication in C-RAN systems with infrastructure sharing," in *Proc. IEEE Int. Conf. Sci. Electr. Eng. Isr. (ICSEE)*, Elat, Israel, Dec. 2018, pp. 1–5.
[49] H. G. Eggleston, *Convexity*, 6th ed. Cambridge, U.K.: Cambridge Univ. Press, 1958.

Sreejith Sreekumar received the B.Tech. degree in electrical engineering from the National Institute of Technology, Calicut, in 2011, the M.Tech. degree in communication engineering from the Indian Institute of Technology, Bombay, in 2013, and the Ph.D. degree in electrical engineering from Imperial College London in 2019. From 2013 to 2015, he worked as a Systems Design Engineer with Broadcom Communications Pvt. Ltd., Bengaluru, India. He is currently a Post-Doctoral Research Associate with the Department of Electrical and Computer Engineering, Cornell University, Ithaca, NY, USA.

Alexander Bunin received the B.Sc. degree *(summa cum laude)* in electrical engineering, the B.Sc. degree *(summa cum laude)* in physics, and the M.Sc. degree *(summa cum laude)* in electrical engineering from the Technion–Israel Institute of Technology in 2010 and 2018, respectively. He is currently with the Advanced Flash Solutions Laboratory, Samsung, Israel, where his research interest includes advanced applications of machine learning algorithms to flash memory.

Ziv Goldfeld (Member, IEEE) received the B.Sc., M.Sc., and Ph.D. degrees in electrical and computer engineering from Ben-Gurion University of the Negev, Israel, in 2012, 2014, and 2017, respectively. From 2017 to 2019, he was a Post-Doctoral Fellow with the Laboratory for Information and Decision Systems (LIDS), MIT. Since 2019, he has been an Assistant Professor of electrical and computer engineering, Cornell University. He was a recipient of several awards, among them are the 2020 NSF CRII Award, the 2020 IBM Academic Award, the Rothschild Postdoctoral Fellowship, the Feder Award, and the Best Student Paper Award in the IEEE 28th Convention of Electrical and Electronics Engineers in Israel.

Haim H. Permuter (Senior Member, IEEE) received the B.Sc. and M.Sc. degrees *(summa cum laude)* in electrical and computer engineering from Ben-Gurion University of the Negev, Israel, in 1997 and 2003, respectively, and the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA, USA, in 2008. From 1997 to 2004, he was an Officer with the Research and Development Unit of the Israeli Defense Forces. Since 2009, he has been with the Department of Electrical and Computer Engineering, Ben-Gurion University of the Negev, where he is currently a Professor and the Luck-Hille Chair of electrical engineering. He also serves as the Head of the communication track in his department. He was a recipient of several awards, among them the Fullbright Fellowship, the Stanford Graduate Fellowship (SGF), the Allon Fellowship, and the U.S.–Israel Binational Science Foundation Bergmann Memorial Award. He has served on the editorial boards for the IEEE TRANSACTIONS ON INFORMATION THEORY from 2013 to 2016.

Shlomo Shamai (Shitz) (Life Fellow, IEEE) is currently with the Department of Electrical Engineering, Technion–Israel Institute of Technology, where he is also a Technion Distinguished Professor, and holds the William Fondiller Chair of Telecommunications. He is also an URSI Fellow, a member of the Israeli Academy of Sciences and Humanities, and a Foreign Member of the U.S. National Academy of Engineering. He was a recipient of the 2011 Claude E. Shannon Award, the 2014 Rothschild Prize in Mathematics/Computer Sciences and Engineering, and the 2017 IEEE Richard W. Hamming Medal. He was a co-recipient of the 2018 Third Bell Labs Prize for Shaping the Future of Information and Communications Technology. He was also a recipient of numerous technical and paper awards and recognitions of the IEEE (Donald G. Fink Prize Paper Award), Information Theory, Communications and Signal Processing Societies, and EURASIP. He is listed as a Highly Cited Researcher (Computer Science) for the years 2004, 2005, 2006, 2007, 2008, and 2013. He has served as an Associate Editor for the Shannon Theory of the IEEE TRANSACTIONS ON INFORMATION THEORY. He has also served twice on the Board of Governors for the Information Theory Society. He has also served on the Executive Editorial Board for the IEEE TRANSACTIONS ON INFORMATION THEORY, the IEEE Information Theory Society Nominations and Appointments Committee, and the IEEE Information Theory Society, Shannon Award Committee.