Arbitrarily Varying Wiretap Channels With Type Constrained States

Ziv Goldfeld, *Student Member, IEEE*, Paul Cuff, *Member, IEEE*, and Haim H. Permuter, *Senior Member, IEEE*

Abstract—Determining a single-letter secrecy-capacity formula for the arbitrarily varying wiretap channel (AVWTC) is an open problem largely because of two main challenges. Not only does it capture the difficulty of the compound wiretap channel (another open problem), it also requires that secrecy is ensured with respect to exponentially many possible channel state sequences. By extending the strong soft-covering lemma, recently derived by the authors, to the heterogeneous scenario, this paper accounts for the exponential number of secrecy constraints while only imposing single-letter constraints on the communication rate. Through this approach, we derive a single-letter characterization of the correlated-random (CR)-assisted semantic-security (SS) capacity of an AVWTC with a type constraint on the allowed state sequences. The allowed state sequences are the ones in a typical set around a single constraining type. The stringent SS requirement is established by showing that the mutual information between the message and the eavesdropper's observations is negligible even when maximized over all message distributions, choices of state sequences, and realizations of the CR-code. Both the achievability and the converse proofs of the type-constrained coding theorem rely on stronger claims than actually required. The direct part establishes a novel single-letter lower bound on the CR-assisted SS-capacity of an AVWTC with state sequences constrained by any convex and closed set of state probability mass functions. This bound achieves the best known single-letter secrecy rates for a corresponding compound wiretap channel over the same constraint set. In contrast to other single-letter results in the AVWTC literature, this paper does not assume the existence of a best channel to the eavesdropper. Instead, SS follows by leveraging the heterogeneous version of the strong soft-covering lemma and a CR-code reduction argument. Optimality is a consequence of a max-inf upper bound on the CR-assisted SS-capacity of an AVWTC with state sequences constrained to any collection of type-classes. When adjusted to the aforementioned compound WTC, the upper bound simplifies to a max–min structure, thus strengthening the previously best known single-letter upper bound by Liang *et al.* that has a min–max form. The proof of the upper bound uses a novel distribution coupling argument. The capacity formula shows that the legitimate users effectively see an averaged main channel, while security must be ensured versus an eavesdropper with perfect channel state information. An example visualizes our single-letter results, and their relation to the past multi-letter secrecy-capacity characterization of the AVWTC is highlighted.

Index Terms—Arbitrarily varying wiretap channel, distribution coupling, information theoretic security, physical layer security, soft-covering lemma.

I. INTRODUCTION

MODERN communication systems usually present an architectural separation between error correction and data encryption. The former is typically realized at the physical layer by transforming the noisy communication channel into a reliable "bit pipe". The data encryption is implemented on top of that by applying cryptographic principles. The cryptographic approach assumes no knowledge on the quality of the eavesdropper's channel and relies solely on restricting the computational power of the eavesdropper. The looming prospect of quantum computers (QCs) (some companies have recently reported a working prototype of a QC with over than 1000 qbits [1]–[5]), however, would boost computational abilities, rendering some critical cryptosystems insecure and weakening others.[1] Post-QC cryptography offers partial solutions that rely on larger keys, but even now considerable efforts are made to save this expensive resource. Nonetheless, cryptography remains the main practical tool for protecting data, at least for the time being.

Physical Layer Security, rooted in information-theoretic principles, is an alternative approach to provably secure communication that dates back to Wyner's celebrated paper on the wiretap channel (WTC) [9]. Essentially, Wyner's main idea was to exploit the noise of the communication channel along with proper physical layer coding to guarantee secrecy against a computationally-unlimited eavesdropper. Protection against such an eavesdropper, however, comes at a price of assuming that the eavesdropper's channel is perfectly known to the legitimate parties and stays fixed during the transmission. Many of the information-theoretic secrecy results that followed

[1] More specifically, asymmetric ciphers that rely on the hardness of integer factorization or discrete logarithms can be completely broken using QCs via Shor's algorithm (or a variant thereof) [6], [7]. Symmetric encryption, on the other hand, would be weakened by QC attacks but could regain its strength by increasing the size of the key [8]. This essentially follows since a QC can search through a space of size 2^n in time $2^{\frac{n}{2}}$, so by doubling the size of the key a symmetric cryptosystem would offer the same protection versus a QC attack, as the original system did versus a classic attack.

relied on extending Wyner's ideas, and therefore, are derived under the same hypothesis. Much of the critique by the cryptographic community towards information-theoretic security is aimed exactly at that assumption.

Practical systems suffer from limited channel state information (CSI) due to inaccuracies in the channel's estimation process and imperfect feedback. Furthermore, adversarial eavesdroppers will refrain from providing the legitimate parties with any information about their channels to make securing the data even harder. Accordingly, limited CSI (especially about the eavesdropper's channel) must be assumed to successfully model a practical communication system. The model of an arbitrarily varying WTC (AVWTC), that is the focus of this work, does just that. The AVWTC combines the WTC [9], [10] and the arbitrarily varying channel (AVC) [11]–[15]. It consists of a collection of discrete-memoryless WTCs indexed by elements in a finite state space. The state at each time instance is chosen in an arbitrary manner and is unknown to the legitimate parties. Being aware of the state space, however, the legitimate users can place the actual channel realization within a certain uncertainty set, which models their limited eavesdropper's CSI. A relaxed scenario where the main channel is fixed in time but the eavesdropper's channel is varying and unknown was studied in [16]. The authors of [16] considered the MIMO Gaussian case and proved the existence of a universal secure coding scheme.

Inspired by wiretap channel instances that involve active eavesdroppers [17]–[20], the AVWTC was first introduced in [21]. The author of [21] derived single-letter lower and upper bounds on the correlated-random (CR) assisted weak secrecy-capacity of the AVWTC. The relation between the CR-assisted weak secrecy-capacity and the uncorrelated weak secrecy-capacity was also established in Theorem 5 of that work (see also [22] for similar results for strong secrecy). It turns out that it makes a difference whether CR codes or their uncorrelated counterparts are used. In particular, the uncorrelated secrecy-capacity may be zero (if the main channel is symmetrizable [14, Definition 2]), while the CR-assisted secrecy-capacity is positive. Thus, viewing CR as an additional resource for communication, this resource can make communication possible where it is impossible without, as long as the choice of the state sequence is independent of the realization of the CR. On the other hand, CR should not be viewed as a cryptographic key to be exploited for secrecy, and therefore, it is assumed to be known to the eavesdropper. A single-letter characterization of the CR-assisted secrecy-capacity remains an open problem and only a multi-letter description has been established [23], [24]. Despite the computational infeasibility of that formula, it was used in [24] to prove that the CR-assisted secrecy-capacity of the AVWTC is a continuous function of the uncertainty set. In contrast, the work of [25] showed that the same in not true for uncorrelated secrecy-capacity, by exemplifying a discontinuity point.

The challenge presented by the AVWTC is twofold. First, it subsumes the difficulty of the compound WTC (where the channel's state is constant in time), for which a single-letter secrecy-capacity characterization is also an open problem [26]–[33]. While a multi-letter description of the compound WTC's secrecy-capacity was found in [32], it is currently unknown how to single-letterize this expression. The underlying gap is that while reliability must be ensured with respect to the worst main channel, security is measured under the best eavesdropper channel; a single channel state under which these extremes simultaneously materialize, however, does not necessarily exist. The second difficulty concerning AVWTCs is that security must be ensured under all possible state sequences, whose number grows exponentially with the blocklength. To get single-letter results, the latter is usually dealt with by assuming the existence of *a best channel to the eavesdropper* and establishing secrecy with respect to that channel only (see, e.g., [21], [34]). Yet, the only single-letter secrecy-capacity characterization for an AVWTC that the authors are aware of assumes even more [21, Th. 4]. On top of the existence of such a best channel, the derivation of [21, Th. 4] also relies on the AVWTC being strongly-degraded and having independent (main channel and eavesdropper channel) states. A related setting for which a single-letter formula is known is an AVWTC where the CR is used as a secret key and there is a sufficient amount thereof [22]. Although such a model slightly deviates from the operational meaning of CR as considered in this work, formally, it can be viewed as another scenario for which the multi-letter formula from [24] is single-letterizable.

We consider a *general* AVWTC with a type constraint on the allowed state sequences, and establish in Theorem 1 a single-letter characterization of its CR-assisted semantic-security (SS) capacity. Our approach relies neither on the existence of a best channel to the eavesdropper nor on the benefit of secret CR. Instead, we show that the exponential number of security requirements can be satisfied, even while using a random codebook construction under single-letter constraints on the communication rate. This is done via a finer analysis that uses a heterogeneous strong soft-covering lemma, on which we expand the discussion subsequently. A full characterization of both the CR-assisted and the uncorrelated capacities of the classic AVC with a pair of linear constraints on the state and input sequences is due to Csiszár and Narayan [14], [15]. The extension of this setting to the AVWTC scenario is the focus of [35], where a multi-letter description of the CR-assisted secrecy-capacity is given. In our case, the type constraint essentially means that the viable state sequences are only the ones of the prescribed type. However, since a fixed distribution (even if rational) is not a valid type for all blocklengths, we define achievability by allowing the empirical distribution of the state sequences to be within a small gap from the type. By doing so, the type constrained AVWTC is well defined for all blocklengths. As a consequence, our uncertainty set is a typical set around the allowed type, which still contains exponentially many state sequences. The structure of the CR-assisted SS-capacity formula suggests that the legitimate users effectively see the averaged channel (i.e., the expectation of the main channels with respect to the type) while security must be ensured versus an eavesdropper with perfect CSI. A specific instance of a type constrained AVWTC that is related to binary symmetric - binary erasure (BS-BE) WTC that was studied in [36] is used to visualize the result.

The results are derived while adopting the prescription of [37] to replace the commonly used strong secrecy metric with the stricter SS metric [38], [39]. The authors of [37] advocate SS as the new standard for information-theoretic security, because from a cryptographic point of view, strong secrecy is insufficient to provide security of applications. Its main drawback lies in the assumption that the message is random and uniformly distributed, as real-life messages are neither (messages may be files, votes or any type of structured data, often with low entropy). In turn, the uniformly distributed message makes the strong secrecy metric an average quantity, that might converge even when many[2] of the messages are actually not secured. Furthermore, to eliminate the benefit of CR for secrecy purposes, we demand that SS holds for each realization of the CR (a similar approach was taken in [24] and [25] with respect to the strong secrecy metric). This essentially means that the transmission is semantically-secure even if the choice of the state sequence depends on the realization of the CR.

In Lemma 1 we develop a heterogeneous soft-covering analysis tool that is key in ensuring SS under the exponentially many state sequences of the AVWTC. By means of the Chernoff bound, the lemma guarantees a double-exponential decay of the probability that soft-covering fails to occur under the relative entropy metric. The probability is taken with respect to a random codebook and the convergence occurs as long as the rate of the codebook is greater than the mutual information between the channel's input and output random variables. In turn, this allows us to furnish a single-letter achievability result without assuming that a best channel to the eavesdropper exists. Doubly-exponentially decaying probabilities coming from the Chernoff bound were previously used in the context of secrecy in, e.g., [22]–[24], [32], [34], [40]. In particular, claims similar to these presented in this work (but under the total variation metric) were used in [22], [24], and [25] for the security analysis under the AVWTC scenario. Similar concentration results also previously appeared in quantum information theory sources [41], [42]. Nevertheless, we emphasize the significance of the strong soft-covering lemma as a stand-alone claim because of the effectiveness of soft-covering in proofs of secrecy, resolvability [43], [44], and channel synthesis [45]. Furthermore, the convergence to 0 of the relative entropy implied by Lemma 1 naturally relates to the definition of SS that uses mutual information.

To prove our coding theorem for the type constrained AVWTC (i.e., the main result in Theorem 1), we provide both a stronger achievability and a stronger converse than is actually required. The broader achievability claim, found in Theorem 2, is a lower bound on the CR-assisted SS-capacity of an AVWTC with state sequences constrained by any convex and closed set of state PMFs. This bound shows that the best known achievable single-letter secrecy rates for a similarly constrained compound WTC [28], [32] can be achieved also in

the AVWTC.[3] The lower bound is derived by first generating a CR-code over a large family of uncorrelated codes, whose size grows doubly-exponentially with the blocklength, and establish reliability by arguments similar to those used for the classic AVC with constrained states [15]. Then, we invoke a Chernoff bound to show that a uniform CR-code over a family that is no more than polynomial in size is sufficient. Having this, SS follows via the union bound and the strong soft-covering lemma, because the combined number of codes, state sequences and messages grows only exponentially with the blocklenght. The lemma is still sharp enough to imply that the probability of a random codebook violating security is doubly-exponentially small. The obtained single-letter achievability formula is shown to be recoverable from the multi-letter CR-assisted secrecy-capacity description from [24] and [25] when specialized to the constrained states scenario.

The polynomial size of the reduced CR-code is of consequence for the uncorrelated scenario as well. Provided that the uncorrelated SS-capacity is strictly positive, the relatively small CR-code allows to replace the shared randomness between the legitimate parties (used for selecting a code from the family) with a local randomness at the transmitter. The transmitter may select the code and inform the receiver which code is in use by sending its index as a short prefix. The positivity of the uncorrelated capacity is essential to allow the reliable transmission of the short prefix with a vanishing rate. Thus, the missing piece for establishing the uncorrelated SS-capacity of the type constrained AVWTC is a dichotomy result (in the spirit of [12] and [15]) based on a condition that distinguishes weather its uncorrelated and CR-assisted secrecy-capacities are equal or not. Such results are available for AVWTCs without constraints on the state space [21], [22], [34]. CR SS-capacity being the focus of this work, we pose the dichotomy result and the corresponding threshold property for the constrained states scenario as questions for future research.

To prove the converse part of the main result, we claim in Theorem 3 an upper bound on CR-assisted SS-capacity of an AVWTC with state sequences from any collection of type-classes. The upper bound is of a max-inf form, i.e., first an infimum over the constraint set is taken, and then the result is maximized over the input distributions. When specializing the result to the aforementioned compound WTC, it produces an upper bound that improves upon the previously best known single-letter upper bound for this setting [28, Th. 2]. The latter result has a min-max structure, while our upper bound has a max-min form. This strengthening is due to a derivation that is uniform over the constraint set. The analysis is preformed per each type in the set and shows that reliability and SS under states from even a single type-class imply similar performance limits as the same channel but where the state sequence is independently and identically distributed (i.i.d.) according to the type. The main challenge is in upper bounding the normalized equivocation of the message given an output sequence that is generated by the average channel. This step relies on the equivocation being continuous in the set of viable

---

[2]The number of unsecured messages may even grow exponentially with the blocklength, while still having a converging strong secrecy metric.

[3]This connection is expounded in Remark 15.

state sequences. A novel distribution coupling argument is used to establish this desired property.

This paper is organized as follows. Section II provides definitions and basic properties. In Section III we state and prove the heterogeneous strong soft-covering lemma. The AVWTC with type constrained states is studied in Section IV, where the setup is defined and the CR-assisted SS-capacity is characterized and proven. Section IV also states the lower and upper bounds for the more general setup, relates the achievability result to past multi-letter descriptions of the CR-assisted secrecy-capacity and provides an example. The lower and upper bounds are proven in Sections V and VI, respectively. Finally, Section VII summarizes the main achievements and insights of this work.

II. NOTATIONS AND PRELIMINARIES

We use the following notations. As customary, $\mathbb{N}$ is the set of natural numbers (which does not include 0), $\mathbb{Q}$ denotes the rational numbers, while $\mathbb{R}$ are the reals. We further define $\mathbb{R}_+ = \{x \in \mathbb{R} | x \geq 0\}$ and $\mathbb{R}_{++} = \{x \in \mathbb{R} | x > 0\}$. Given two real numbers a, b, we denote by $[a:b]$ the set of integers $\{n \in \mathbb{N} | \lceil a \rceil \leq n \leq \lfloor b \rfloor\}$. Calligraphic letters denote sets, e.g., $\mathcal{X}$, the complement of $\mathcal{X}$ is denoted by $\mathcal{X}^c$, while $|\mathcal{X}|$ stands for its cardinality. $\mathcal{X}^n$ denotes the n-fold Cartesian product of $\mathcal{X}$. An element of $\mathcal{X}^n$ is denoted by $x^n = (x_1, x_2, \ldots, x_n)$; whenever the dimension n is clear from the context, vectors (or sequences) are denoted by boldface letters, e.g., $\mathbf{x}$. For any $\mathcal{S} \subseteq [1:n]$, we use $\mathbf{x}^{\mathcal{S}} \triangleq (x_i)_{i \in \mathcal{S}}$ to denote the substring of x^n defined by $\mathcal{S}$, with respect to the natural ordering of $\mathcal{S}$. For instance, if $\mathcal{S} = [i:j]$, where $1 \leq i < j \leq n$, then $\mathbf{x}^{\mathcal{S}} = (x_i, x_{i+1}, \ldots, x_j)$. For $\mathcal{S} = [i:j]$ as before we sometimes write x_i^j instead of $\mathbf{x}^{\mathcal{S}}$; when $i = 1$, the subscript is omitted. We also use $x^{n \setminus i}$ instead $\mathbf{x}^{\mathcal{S}}$, when $\mathcal{S} = [1:i-1] \cup [i+1:n]$, for $1 \leq i \leq n$.

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

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

PMFs are denoted by the uppercase letters such as P or Q, with a subscript that identifies the random variable and its possible conditioning. For example, for a discrete probability space $(\mathcal{X}, \mathcal{F}, \mathbb{P})$ and two correlated random variables X and Y over that space, we use P_X, $P_{X,Y}$ and $P_{X|Y}$ to denote, respectively, the marginal PMF of X, the joint PMF of (X, Y) and the conditional PMF of X given Y. In particular, $P_{X|Y}$ represents the stochastic matrix whose elements are given by $P_{X|Y}(x|y) = \mathbb{P}(X = x|Y = y)$. Expressions such as $P_{X,Y} = P_X P_{Y|X}$ are to be understood as $P_{X,Y}(x, y) = P_X(x) P_{Y|X}(y|x)$, for all $(x, y) \in \mathcal{X} \times \mathcal{Y}$. Accordingly, when

three random variables X, Y and Z satisfy $P_{X|Y,Z} = P_{X|Y}$, they form a Markov chain, which we denote by $X - Y - Z$. We omit subscripts if the arguments of a PMF are lowercase versions of the random variables. The support of a PMF P and the expectation of a real-valued random variable X are denoted by $\mathrm{supp}(P)$ and $\mathbb{E}[X]$, respectively. If $X \sim P$, we emphasize that an expectation is taken with respect to the distribution on X by writing $\mathbb{E}_X$ or $\mathbb{E}_P$ (choosing the simpler of the two). Similarly, we use H_P and I_P to indicate that an entropy or a mutual information term are calculated with respect to a PMF P.

For a discrete measurable space $(\mathcal{X}, \mathcal{F})$, a PMF $Q \in \mathcal{P}(\mathcal{X})$ gives rise to a probability measure on $(\mathcal{X}, \mathcal{F})$, which we denote by $\mathbb{P}_Q$; accordingly, $\mathbb{P}_Q(\mathcal{A}) = \sum_{x \in \mathcal{A}} Q(x)$, for every $\mathcal{A} \in \mathcal{F}$. For a sequence of random variable X^n, if the entries of X^n are drawn in an independent and identically distributed (i.i.d.) manner according to P_X, then for every $\mathbf{x} \in \mathcal{X}^n$ we have $P_{X^n}(\mathbf{x}) = \prod_{i=1}^n P_X(x_i)$ and we write $P_{X^n}(\mathbf{x}) = P_X^n(\mathbf{x})$. Similarly, if for every $(\mathbf{x}, \mathbf{y}) \in \mathcal{X}^n \times \mathcal{Y}^n$ we have $P_{Y^n|X^n}(\mathbf{y}|\mathbf{x}) = \prod_{i=1}^n P_{Y|X}(y_i|x_i)$, then we write $P_{Y^n|X^n}(\mathbf{y}|\mathbf{x}) = P_{Y|X}^n(\mathbf{y}|\mathbf{x})$. The conditional product PMF $P_{Y|X}^n$ given a specific sequence $\mathbf{x} \in \mathcal{X}^n$ is denoted by $P_{Y|X=\mathbf{x}}^n$.

The type $\nu_{\mathbf{x}}$ of a sequence $\mathbf{x} \in \mathcal{X}^n$ is

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

where $N(x|\mathbf{x}) = \sum_{i=1}^n \mathbb{1}_{\{x_i = x\}}$. The subset of $\mathcal{P}(\mathcal{X})$ that contains all possible types of sequences $\mathbf{x} \in \mathcal{X}^n$ is denoted by $\mathcal{P}_n(\mathcal{X})$. By [46, Lemma II.1],

$$|\mathcal{P}_n(\mathcal{X})| = \binom{n + |\mathcal{X}| - 1}{|\mathcal{X}| - 1} \leq (n+1)^{|\mathcal{X}|}. \tag{3}$$

For $P \in \mathcal{P}_n(\mathcal{X})$, the type-class $\{\mathbf{x} \in \mathcal{X}^n | \nu_{\mathbf{x}} = P\}$ is denoted by $\mathcal{T}_P^n$. We use $\mathcal{T}_\epsilon^n(P)$ to denote the set of letter-typical sequences with respect to the PMF $P \in \mathcal{P}(\mathcal{X})$ and the positive number ϵ defined by

$$\mathcal{T}_\epsilon^n(P) = \left\{ \mathbf{x} \in \mathcal{X}^n \middle| |\nu_{\mathbf{x}}(x) - P(x)| \leq \frac{\epsilon}{|\mathcal{X}|} \mathbb{1}_{\{P(x) > 0\}} \right\}. \tag{4}$$

This definition of the letter-typical set resembles this from [47, Ch. 2], with the only difference being the normalization of ϵ by $|\mathcal{X}|$. This gives rise to an upper bound on the probability of an i.i.d. sequence being atypical that is uniform in the underlying i.i.d. distribution. Namely, by a simple adaptation of [47, Lemma 2.12], if X^n is i.i.d. according to $P \in \mathcal{P}(\mathcal{X})$, then

$$\mathbb{P}_{P^n}\left(X^n \notin \mathcal{T}_\epsilon^n(P) \right) \leq 2|\mathcal{X}| e^{-2n \frac{\epsilon^2}{|\mathcal{X}|^2}}. \tag{5}$$

This uniform bound plays an important role in the proof of Theorem 3, where an upper bound on the CR-assisted SS-capacity of the AVWTC is established.

Definition 1 (Relative Entropy): Let $(\mathcal{X}, \mathcal{F})$ be a measurable space and let P and Q be two probability measures on $\mathcal{F}$, with $P \ll Q$ (i.e., P is absolutely continuous with respect to Q). The relative entropy between P and Q is

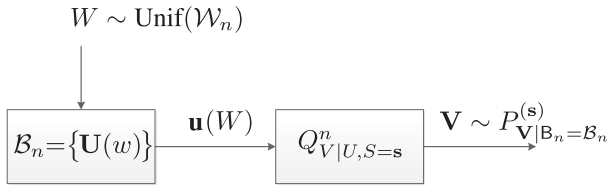$$D(P \| Q) = \int_{\mathcal{X}} dP \log \left(\frac{dP}{dQ} \right), \tag{6}$$

Fig. 1. Coding problem with the goal of making $P_{\mathbf{V}|\mathsf{B}_n=\mathcal{B}_n}^{(\mathbf{s})}$ resemble $Q_{V|S=\mathbf{s}}^n$.

where $\frac{dP}{dQ}$ denotes the Radon-Nikodym derivative of P with respect to Q. If the sample space $\mathcal{X}$ is countable, (6) reduces to

$$D(P\|Q) = \sum_{x\in\mathrm{supp}(P)} P(x) \log\left(\frac{P(x)}{Q(x)}\right). \qquad (7)$$

Definition 2 (Total Variation): Let $(\mathcal{X}, \mathcal{F})$ be a measurable space and P and Q be two probability measures on $\mathcal{F}$. The total variation between P and Q is

$$\|P - Q\|_{TV} = \sup_{\mathcal{A}\in\mathcal{F}} |P(\mathcal{A}) - Q(\mathcal{A})|. \qquad (8)$$

If the sample space $\mathcal{X}$ is countable, (8) reduces to

$$\|P - Q\|_{TV} = \frac{1}{2} \sum_{x\in\mathcal{X}} |P(x) - Q(x)|. \qquad (9)$$

III. HETEROGENEOUS STRONG SOFT-COVERING LEMMA

We present here a generalization of the original *strong soft-covering lemma* first established by the authors in [48, Lemma 1]. The lemma in this work is a heterogeneous version on the original homogeneous claim. [50, Lemma 1] considers a discrete-memoryless channel (DMC) that does not change throughout the block transmission, while here the memoryless channel may vary from symbol to symbol. This variation is modeled as an n-fold state-dependent channel $Q_{V|U,S}^n$ over which a codeword $\mathbf{u} \in \mathcal{U}^n$ is transmitted under a state sequence $\mathbf{s} \in \mathcal{S}^n$. Thus, in each time instance $i \in [1:n]$, the i-th symbol of $\mathbf{u}$ is transmitted over the channel $Q_{V|U,S=s_i}$.

Let $\mathcal{B}_n$ be a randomly generated codebook of u-sequence, one of which is selected uniformly at random and passed through the channel $Q_{V|U,S=\mathbf{s}}^n$. Lemma 1 gives a sufficient condition for the induced conditional distribution of the channel output given the state to result in a good approximation of $Q_{V|S=\mathbf{s}}^n$ in the limit of large n, for any $\mathbf{s} \in \mathcal{S}^n$ (Fig. 1). The proximity between the induced and the desired distributions is measured in terms of relative entropy. Specifically, we show that as long as the codebook is of size $|\mathcal{B}_n| = 2^{nR}$ with $R > I(U;V|S)$, where the mutual information is calculated with respect to the empirical PMF $\nu_{\mathbf{s}}$ of the state sequence, the relative entropy vanishes exponentially quickly with the blocklength n, with high probability with respect to the random codebook. Via the Chernoff bound, the negligible probability of the random set not producing this desired result is doubly-exponentially small.

The heterogeneous strong soft-covering lemma is subsequently invoked for the SS analysis of the AVWTC, where the double-exponential decay it provides plays a key role. Similar claims that use total variation were previously made in

the context of AVWTCs in [22], [24], and [25] (in particular, see [22, Lemma 1]), though the codebook design was slightly different in those works. The stronger notion of soft-covering was also previously observed in works on quantum information theory [41], [42]. We emphasize Lemma 1 as a stand-alone tool due to its simplicity, and consequently, the prospect of it coming in handy for other proofs of secrecy, channel resolvability, channel synthesis, etc.

A. Soft-Covering Setup and Result

Let $\mathcal{S}$ be a finite set and let $\mathbf{s} \in \mathcal{S}^n$ be a sequence with an empirical PMF $\nu_{\mathbf{s}}$. Let $\mathsf{B}_n = \{\mathbf{U}(w)\}_{w\in\mathcal{W}_n}$, where[4] $\mathcal{W}_n = [1 : 2^{nR}]$ with $R \in \mathbb{R}_+$, be a set of random vectors that are i.i.d. according to $Q_{U|S=\mathbf{s}}^n$, where $Q_{U|S} : \mathcal{S} \to \mathcal{P}(\mathcal{U})$. We refer to B_n as the random codebook, denote by $\mathfrak{B}_n$ the set of all its possible realizations, while a specific realization is denoted by $\mathcal{B}_n = \{\mathbf{u}(w)\}_{w\in\mathcal{W}_n}$. For every $\mathcal{B}_n \in \mathfrak{B}_n$, a sequence $\mathbf{u}(w)$ is randomly and uniformly selected and passed through a memoryless non-stationary channel $Q_{V|U,S=\mathbf{s}}^n$. For each $\mathbf{s} \in \mathcal{S}^n$, the induced joint distribution of B_n, W and $\mathbf{V}$ is

$$P_{\mathsf{B}_n,W,\mathbf{V}}^{(\mathbf{s})}(\mathcal{B}_n, w, \mathbf{v})$$
$$= \left[\prod_{w'\in\mathcal{W}_n} Q_{U|S}^n(\mathbf{u}(w')|\mathbf{s})\right] 2^{-nR} \cdot Q_{V|U,S}^n(\mathbf{v}|\mathbf{u}(w),\mathbf{s}), \quad (10)$$

which gives rise to a probability measure denoted by $\mathbb{P}$.[5] When switching to other probability measures, we do so in accordance to the notations defined in Section II. On account of (10), the induced output distribution conditioned on a codebook $\mathcal{B}_n \in \mathfrak{B}_n$ for each state sequence $\mathbf{s} \in \mathcal{S}^n$ is:

$$P_{\mathbf{V}|\mathsf{B}_n}^{(\mathbf{s})}(\mathbf{v}|\mathcal{B}_n) = 2^{-nR} \sum_{w\in\mathcal{W}_n} Q_{V|U,S}^n(\mathbf{v}|\mathbf{u}(w),\mathbf{s}). \qquad (11)$$

The following lemma states that as long as the codebook is of size 2^{nR}, with[6] $R > I_{\nu_{\mathbf{s}}Q}(U;V|S)$, the induced output PMF constitutes a good approximation of $Q_{V|S=\mathbf{s}}^n$ in the limit of large n, with high probability. Namely, the probability that the relative entropy between the induced PMF and product PMF vanishes exponentially quickly with the blocklength n, is double exponentially close to 1.

Lemma 1 (Heterogeneous Strong Soft-Covering Lemma): For $\mathbf{s} \in \mathcal{S}^n$ with empirical PMF $\nu_{\mathbf{s}}$, and any $\zeta > 0$, $Q_{U,V|S} : \mathcal{S} \to \mathcal{P}(\mathcal{U} \times \mathcal{V})$, and $R > I_{\nu_{\mathbf{s}}Q}(U;V|S) + \zeta$, where $|\mathcal{S}|, |\mathcal{V}| < \infty$, there exist $\gamma_1, \gamma_2 > 0$, such that for n large enough

$$\mathbb{P}\left(D\left(P_{\mathbf{V}|\mathsf{B}_n}^{(\mathbf{s})} \middle\| Q_{V|S=\mathbf{s}}^n\right) > e^{-n\gamma_1}\right) \leq e^{-e^{n\gamma_2}}. \qquad (12)$$

---

[4]To simplify notation, we assume that 2^{nR} is an integer, for all $n \in \mathbb{N}$. Otherwise, simple modifications of some of the subsequent expressions using floor operations are needed.

[5]Since $\mathbf{s} \in \mathcal{S}^n$ stays fixed throughout this section, the notation $\mathbb{P}$ omits the dependence of the state sequence.

[6]The subscript $\nu_{\mathbf{s}}Q$ indicates that the mutual information is calculated with respect to $\nu_{\mathbf{s}}Q_{U|S}Q_{V|U,S}$

More precisely, for any $n \in \mathbb{N}$ and $\delta \in \left(0, R - I_{v_s Q}(U; V|S)\right)$

$$\mathbb{P}\left(D\left(P^{(s)}_{\mathbf{V}|\mathcal{B}_n} \middle\| Q^n_{V|S=s}\right) > c_\delta n 2^{-n\gamma_\delta}\right) \leq \left(1 + |\mathcal{V}|^n\right) e^{-\frac{1}{3} 2^{n\delta}}, \tag{13}$$

where

$$\gamma_\delta = \sup_{\eta > 1} \frac{\eta - 1}{2\eta - 1}\left(R - \delta - \max_{s \in \mathcal{S}} d_\eta(Q_{U,V|S=s}, Q_{U|S=s} Q_{V|S=s})\right) \tag{14a}$$

$$c_\delta = 3 \log e + \gamma_\delta 2 \log 2 + 2 \log \left(\max_{\substack{(s,v) \in \mathcal{S} \times \mathcal{V}: \\ Q_{V|S}(v|s) > 0}} \frac{1}{Q_{V|S}(v|s)}\right), \tag{14b}$$

and $d_\eta(\mu, v) = \frac{1}{\eta - 1} \log_2 \int d\mu \left(\frac{d\mu}{dv}\right)^{1-\eta}$ is the Rényi divergence of order η.

The proof of Lemma 1 bears close resemblance to the proof of the homogeneous version of the strong soft-covering lemma from [48, Lemma 1]. The main difference, is in the bound on the expected value of the probability of atypical sequences. To avoid verbatim repetition of the arguments from [48], we summarize (most of) the technical parts from the proof of the homogeneous case in our Lemma 2 and invoke it to establish Lemma 1.

The important quantity in the lemma above is γ_δ, which is the exponent that soft-covering achieves. We see in (13) that the double-exponential convergence of probability occurs with exponent $\delta > 0$. Thus, the best soft-covering exponent that the lemma achieves with confidence, over all $\delta > 0$, is

$$\begin{aligned}\gamma^* &= \sup_{\delta > 0} \gamma_\delta \\ &= \gamma_0 \\ &= \sup_{\eta > 1} \frac{\eta - 1}{2\eta - 1}\left(R - \max_{s \in \mathcal{S}} d_\eta(Q_{U,V|S=s}, Q_{U|S=s} Q_{V|S=s})\right). \end{aligned} \tag{15}$$

The double-exponential confidence rate δ acts as a reduction in codebook rate R in the definition of γ_δ. Consequently, $\gamma_\delta = 0$ for $\delta \geq R - I_{v_s Q}(U; V|S)$.

Remark 1: The role of ζ in the statement of Lemma 1 is merely to ensure that a fixed $\delta \in \left(0, R - I_{v_s Q}(U; V|S)\right)$ can be found for all n. Since the mutual information is calculated with respect to v_s, its value may vary with n. Taking $R > I_{v_s Q}(U; V|S) + \zeta$ implies that $(0, \zeta) \subseteq \left(0, R - I_{v_s Q}(U; V|S)\right)$, and therefore, a fixed value of δ as needed exists.

Remark 2 (Total Variation Exponent of Decay): The strong soft-covering lemma can be reproduced while replacing the relative entropy with total variation. Although, relative entropy can be used to bound total variation via Pinsker's inequality, this approach causes a loss of a factor of 2 in the exponent of decay. Alternatively, the proof of Lemma 1 can be modified to produce the bound on the total variation instead of the relative entropy. This direct method keeps the error exponents the same for the total variation case as it is for relative entropy.

Proof of Lemma 1: We state the proof in terms of arbitrary distributions $Q_{U|S}$ and $Q_{V|U,S}$ (not necessarily discrete).

We assume $|\mathcal{S}| < \infty$, and will specialize to a finite output alphabet $\mathcal{V}$ only when needed.

First, define conditional information density $i_{Q_{U,V|S=s}}$, which is a function on the space $\mathcal{U} \times \mathcal{V}$ specified by

$$i_{Q_{U,V|S=s}}(u, v) \triangleq \log \left(\frac{dQ_{V|U=u,S=s}}{dQ_{V|S=s}}(v)\right). \tag{16}$$

In (16), the argument of the logarithm is the Radon-Nikodym derivative between $Q_{V|U=u,S=s}$ and $Q_{V|S=s}$. Let $\epsilon \geq 0$ be arbitrary, and define the conditional jointly typical set of u- and v-sequences given $\mathbf{s}$ as

$$\mathcal{A}_\epsilon(\mathbf{s}) \triangleq \left\{(\mathbf{u}, \mathbf{v}) \in \mathcal{U}^n \times \mathcal{V}^n \middle| \frac{1}{n} i_{Q^n_{U,V|S=\mathbf{s}}}(\mathbf{u}, \mathbf{v}) < I(U; V|S) + \epsilon\right\} \tag{17}$$

and note that

$$i_{Q^n_{U,V|S=\mathbf{s}}}(\mathbf{u}, \mathbf{v}) = \sum_{t=1}^{n} i_{Q_{U,V|S=s_t}}(u_t, v_t). \tag{18}$$

For brevity, in (17) and henceforth, we use $I(U; V|S)$ instead of $I_{v_s Q}(U; V|S)$.

Next, for every $\mathcal{B}_n \in \mathfrak{B}_n$, we split $P^{(s)}_{\mathbf{V}|\mathcal{B}_n = \mathcal{B}_n}$ into two parts, making use of the indicator function. For every $\mathbf{v} \in \mathcal{V}^n$, define

$$P^{(1)}_{\mathcal{B}_n, \mathbf{s}}(\mathbf{v}) \triangleq 2^{-nR} \sum_{w \in \mathcal{W}_n} Q^n_{V|U,S}\left(\mathbf{v}|\mathbf{u}(w), \mathbf{s}\right) \mathbf{1}_{\left\{\left(\mathbf{u}(w), \mathbf{v}\right) \in \mathcal{A}_\epsilon(\mathbf{s})\right\}} \tag{19a}$$

$$P^{(2)}_{\mathcal{B}_n, \mathbf{s}}(\mathbf{v}) \triangleq 2^{-nR} \sum_{w \in \mathcal{W}_n} Q^n_{V|U,S}\left(\mathbf{v}|\mathbf{u}(w), \mathbf{s}\right) \mathbf{1}_{\left\{\left(\mathbf{u}(w), \mathbf{v}\right) \notin \mathcal{A}_\epsilon(\mathbf{s})\right\}}. \tag{19b}$$

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

$$\Delta_{\mathcal{B}_n, \mathbf{s}}(\mathbf{v}) = \frac{dP^{(s)}_{\mathbf{V}|\mathcal{B}_n = \mathcal{B}_n}}{dQ^n_{V|S=\mathbf{s}}}(\mathbf{v}), \tag{20}$$

where the right-hand side (RHS) is the Radon-Nikodym derivative between $P^{(s)}_{\mathbf{V}|\mathcal{B}_n = \mathcal{B}_n}$ and $Q^n_{V|S=\mathbf{s}}$, and also split it into $\Delta_{\mathcal{B}_n, \mathbf{s}}(\mathbf{v}) = \Delta^{(1)}_{\mathcal{B}_n, \mathbf{s}}(\mathbf{v}) + \Delta^{(2)}_{\mathcal{B}_n, \mathbf{s}}(\mathbf{v})$, where

$$\Delta^{(1)}_{\mathcal{B}_n, \mathbf{s}}(\mathbf{v}) \triangleq \frac{dP^{(1)}_{\mathcal{B}_n, \mathbf{s}}}{dQ^n_{V|S=\mathbf{s}}}(\mathbf{v}) \tag{21a}$$

$$\Delta^{(2)}_{\mathcal{B}_n, \mathbf{s}}(\mathbf{v}) \triangleq \frac{dP^{(2)}_{\mathcal{B}_n, \mathbf{s}}}{dQ^n_{V|S=\mathbf{s}}}(\mathbf{v}). \tag{21b}$$

To see that the RHS of (20) indeed exists, on could easily verify that $P^{(s)}_{\mathbf{V}|\mathcal{B}_n = \mathcal{B}_n}$ is absolutely continuous with respect to $Q^n_{V|S=\mathbf{s}}$. This essentially follows since if $Q^n_{V|S=\mathbf{s}}(\mathcal{A}) = 0$, for some $\mathcal{A} \subseteq \mathcal{V}^n$, then $Q^n_{V|U=\mathbf{u}, S=\mathbf{s}}(\mathcal{A}) = 0$, for every $\mathbf{u} \in \text{supp}\left(Q^n_{U|S=\mathbf{s}}\right)$. The structure of $P^{(s)}_{\mathbf{V}|\mathcal{B}_n = \mathcal{B}_n}$ given in (11) then implies that $P^{(s)}_{\mathbf{V}|\mathcal{B}_n = \mathcal{B}_n}(\mathcal{A}) = 0$.

Note that $\int dP_{\mathsf{B}_n,\mathbf{s}}^{(2)}$ is an average of exponentially many i.i.d. random variables bounded between 0 and 1, given by

$$\int dP_{\mathsf{B}_n,\mathbf{s}}^{(2)}$$
$$= \sum_{w \in \mathcal{W}_n} 2^{-nR} \cdot \mathbb{P}_{Q_{V|U,S=\mathbf{s}}^n}\Big(\big(\mathbf{U}(w),\mathbf{V}\big) \notin \mathcal{A}_\epsilon(\mathbf{s})\Big|\mathbf{U}(w)\Big). \quad (22)$$

With respect to (22) and the above definitions, the heterogeneous strong soft-covering lemma is established by the following technical lemma.

Lemma 2: Let $|\mathcal{S}|, |\mathcal{V}| < \infty$ and $\epsilon \geq 0$. *If there exist* $\alpha, \beta_\epsilon > 0$ *such that*

$$\mathbb{E}_{\mathsf{B}_n}\mathbb{P}_{Q_{V|U,S=\mathbf{s}}^n}\Big(\big(\mathbf{U}(w),\mathbf{V}\big) \notin \mathcal{A}_\epsilon(\mathbf{s})\Big|\mathbf{U}(w)\Big) \leq 2^{-n\beta_\epsilon} \quad (23a)$$

$$\mathbb{P}\Big(\Delta_{\mathsf{B}_n,\mathbf{s}}^{(2)}(\mathbf{v}) \leq \alpha^n\Big) = 1, \quad \forall \mathbf{v} \in \mathcal{V}^n, \quad (23b)$$

then

$$\mathbb{P}\Big(D\Big(P_{\mathbf{V}|\mathsf{B}_n}^{(s)}\Big|\Big|Q_{V|S=\mathbf{s}}^n\Big) \geq c_{\beta,\epsilon,\alpha}n2^{-n\beta_\epsilon}\Big)$$
$$\leq e^{-\frac{1}{3}2^{n(R-\beta_\epsilon)}} + |\mathcal{V}|^n e^{-\frac{1}{3}2^{n(R-I(U;V|S)-\epsilon-2\beta_\epsilon)}}, \quad (24)$$

where $c_{\beta,\epsilon,\alpha} = 3\log e + 2\beta_\epsilon \log 2 + 2\log \alpha$.

Lemma 2 essentially follows from the proof of [50, Lemma 1]. More specifically, the derivation repeats the steps between Equations (18) and (40) in the proof of [48, Lemma 1] and is, therefore, omitted. Having this, it remains to be shown that (23) holds for certain positive β_ϵ and α. For (23a), observe that

$$\mathbb{E}_{\mathsf{B}_n}\mathbb{P}_{Q_{V|U,S=\mathbf{s}}^n}\Big(\big(\mathbf{U}(w),\mathbf{V}\big) \notin \mathcal{A}_\epsilon(\mathbf{s})\Big|\mathbf{U}(w)\Big)$$
$$= \mathbb{P}_{Q_{U,V|S=\mathbf{s}}^n}\Big(\big(\mathbf{U},\mathbf{V}\big) \notin \mathcal{A}_\epsilon(\mathbf{s})\Big)$$
$$= \mathbb{P}_{Q_{U,V|S=\mathbf{s}}^n}\Big(\sum_{t=1}^n i_{Q_{U,V|S=s_t}}(U_t,V_t) \geq n\big(I(U;V|S)+\epsilon\big)\Big)$$
$$\overset{(a)}{=} \mathbb{P}_{Q_{U,V|S=\mathbf{s}}^n}\Big(2^{\lambda\sum_{t=1}^n i_{Q_{U,V|S=s_t}}(U_t,V_t)} \geq 2^{n\lambda(I(U;V|S)+\epsilon)}\Big)$$
$$\overset{(b)}{\leq} \frac{\mathbb{E}_{Q_{U,V|S=\mathbf{s}}^n}2^{\lambda\sum_{t=1}^n i_{Q_{U,V|S=s_t}}(U_t,V_t)}}{2^{n\lambda(I(U;V|S)+\epsilon)}}, \quad (25)$$

where (a) is true for any $\lambda \geq 0$ and (b) is Markov's inequality. For the numerator on the RHS of (25), we have

$$\mathbb{E}_{Q_{U,V|S=\mathbf{s}}^n}2^{\lambda\sum_{t=1}^n i_{Q_{U,V|S=s_t}}(U_t,V_t)}$$
$$\overset{(a)}{=} \prod_{t=1}^n \mathbb{E}_{Q_{U,V|S=s_t}}2^{\lambda i_{Q_{U,V|S=s_t}}(U_t,V_t)}$$
$$\leq \Big(\max_{t\in[1:n]} \mathbb{E}_{Q_{U,V|S=s_t}}2^{\lambda i_{Q_{U,V|S=s_t}}(U_t,V_t)}\Big)^n$$
$$\overset{(b)}{\leq} \Big(\max_{s\in\mathcal{S}} \mathbb{E}_{Q_{U,V|S=s}}2^{\lambda i_{Q_{U,V|S=s}}(U_s,V_s)}\Big)^n, \quad (26)$$

where (a) uses the independence across time, while (b) follows because $|\mathcal{S}| < \infty$ and by defining $(U_s,V_s) \sim Q_{U,V|S=s}$.

Plugging (26) back into (25), gives

$$\mathbb{E}_{\mathsf{B}_n}\mathbb{P}_{Q_{V|U,S=\mathbf{s}}^n}\Big(\big(\mathbf{U}(w),\mathbf{V}\big) \notin \mathcal{A}_\epsilon(\mathbf{s})\Big|\mathbf{U}(w)\Big)$$
$$\leq \left(\frac{\max_{s\in\mathcal{S}}\mathbb{E}_{Q_{U,V|S=s}}2^{\lambda i_{Q_{U,V|S=s}}(U_s,V_s)}}{2^{\lambda(I(U;V|S)+\epsilon)}}\right)^n$$
$$= \left(\frac{2^{\log_2\Big(\max_{s\in\mathcal{S}}\mathbb{E}_{Q_{U,V|S=s}}2^{\lambda i_{Q_{U,V|S=s}}(U_s,V_s)}\Big)}}{2^{\lambda(I(U;V|S)+\epsilon)}}\right)^n$$
$$\overset{(a)}{=} \left(\frac{2^{\lambda\max_{s\in\mathcal{S}}\frac{1}{\lambda}\log_2\Big(\mathbb{E}_{Q_{U,V|S=s}}2^{\lambda i_{Q_{U,V|S=s}}(U_s,V_s)}\Big)}}{2^{\lambda(I(U;V|S)+\epsilon)}}\right)^n$$
$$= 2^{n\lambda\Big(\max_{s\in\mathcal{S}}\frac{1}{\lambda}\log_2\mathbb{E}_{Q_{U,V|S=s}}\big[2^{\lambda i_{Q_{U,V|S=s}}(U_s,V_s)}\big]-I(U;V|S)-\epsilon\Big)}$$
$$\overset{(b)}{=} 2^{n\lambda\Big(\max_{s\in\mathcal{S}}d_{\lambda+1}(Q_{U,V|S=s},Q_{U|S=s}Q_{V|S=s})-I(U;V|S)-\epsilon\Big)}, \quad (27)$$

where (a) is because the logarithm is non-decreasing and by restricting λ to be strictly positive, while (b) is from the definition of the Rényi divergence of order $\lambda+1$. Substituting $\eta = \lambda + 1$ into (27) yields

$$\mathbb{E}_{\mathsf{B}_n}\mathbb{P}_{Q_{V|U,S=\mathbf{s}}^n}\Big(\big(\mathbf{U}(w),\mathbf{V}\big) \notin \mathcal{A}_\epsilon(\mathbf{s})\Big|\mathbf{U}(w)\Big) \leq 2^{-n\beta_{\eta,\epsilon}}, \quad (28)$$

where

$$\beta_{\eta,\epsilon} = (\eta-1)\Big(I(U;V|S)+\epsilon$$
$$- \max_{s\in\mathcal{S}}d_\eta(Q_{U,V|S=s},Q_{U|S=s}Q_{V|S=s})\Big), \quad (29)$$

for every $\eta > 1$ and $\epsilon \geq 0$. Thus (23a) holds with $\beta_{\eta,\epsilon}$ in the role of β_ϵ.

The relation in (23b) follows by noting that $|\mathcal{S}|, |\mathcal{V}| < \infty$ implies that for any $\mathbf{v} \in \mathcal{V}^n$ and $\mathcal{B}_n \in \mathfrak{B}_n$, we have

$$\Delta_{\mathcal{B}_n,\mathbf{s}}^{(2)}(\mathbf{v}) \leq \left(\max_{\substack{(s,v)\in\mathcal{S}\times\mathcal{V}:\\Q_{V|S}(v|s)>0}}\frac{1}{Q_{V|S}(v|s)}\right)^n. \quad (30)$$

Notice that the maximum is only over the pair (s,v) for which $Q_{V|S}(v|s) > 0$, which makes this bound finite. The underlying reason for this restriction is that with probability one a conditional distribution is absolutely continuous with respect to its associated marginal distribution. By (30), we obtain

$$\mathbb{P}\Big(\Delta_{\mathcal{B}_n,\mathbf{s}}^{(2)}(\mathbf{v}) \leq \alpha^n\Big) = 1, \quad \forall \mathbf{v} \in \mathcal{V}^n, \quad (31)$$

with

$$\alpha = \max_{\substack{(s,v)\in\mathcal{S}\times\mathcal{V}:\\Q_{V|S}(v|s)>0}}\frac{1}{Q_{V|S}(v|s)}. \quad (32)$$

Thus, by Lemma 2 we have (24) with $\beta_{\eta,\epsilon}$ and α from (29) and (32), respectively. Recalling that we may optimize over $\eta > 1$ and $\epsilon \geq 0$, we fix $\delta \in \big(0, R-I(U;V|S)\big)$ and set

$$\epsilon_{\eta,\delta} = \frac{\frac{1}{2}(R-\delta)+(\eta-1)\max_{s\in\mathcal{S}}d_\eta(Q_{U,V|S=s},Q_{U|S=s}Q_{V|S=s})}{\frac{1}{2}+(\eta-1)}$$
$$- I(U;V|S). \quad (33)$$

Substituting into $\beta_{\eta,\epsilon}$ gives

$$\beta_{\eta,\delta} \triangleq \beta_{\eta,\epsilon_{\eta,\delta}}$$
$$= \frac{\eta-1}{2\eta-1}\left(R - \delta - \max_{s\in\mathcal{S}} d_\eta(Q_{U,V|S=s}, Q_{U|S=s}Q_{V|S=s})\right). \tag{34}$$

Plugging α and $\beta_{\eta,\delta}$ into $c_{\beta,\epsilon,\alpha}$, which we relabel as $c_{\eta,\delta}$, we have

$$c_{\eta,\delta} = 3\log e + 2\beta_{\eta,\delta}\log 2 + 2\log\left(\max_{\substack{(s,v)\in\mathcal{S}\times\mathcal{V}:\\Q_{V|S}(v|s)>0}}\frac{1}{Q_{V|S}(v|s)}\right). \tag{35}$$

Observe that $\epsilon_{\eta,\delta}$ in (33) is non-negative under the assumption that $R - \delta > I(U; V|S)$, because $\eta > 1$ and

$$\max_{s\in\mathcal{S}} d_\eta(Q_{U,V|S=s}, Q_{U|S=s}Q_{V|S=s})$$
$$\geq \max_{s\in\mathcal{S}} d_1(Q_{U,V|S=s}, Q_{U|S=s}Q_{V|S=s})$$
$$\geq I(U; V|S). \tag{36}$$

Reevaluating (24) based on (33)-(35) gives

$$\mathbb{P}\left(D\left(P^{(\mathbf{s})}_{\mathsf{V}|\mathsf{B}_n}\Big\|Q^n_{V|S=\mathbf{s}}\right) \geq c_{\eta,\delta}n2^{-n\beta_{\eta,\delta}}\right)$$
$$\leq e^{-\frac{1}{3}2^{n(R-\beta_{\eta,\epsilon})}} + |\mathcal{V}|^n e^{-\frac{1}{3}2^{n\left(R-I(U;V|S)-\epsilon_{\eta,\delta}-2\beta_{\eta,\delta}\right)}}$$
$$= e^{-\frac{1}{3}2^{n\left(R-\beta_{\eta,\delta}\right)}} + |\mathcal{V}|^n \cdot e^{-\frac{1}{3}2^{n\delta}}$$
$$\overset{(a)}{\leq} \left(1+|\mathcal{V}|^n\right)e^{-\frac{1}{3}2^{n\delta}}, \tag{37}$$

where (a) is because $\beta_{\eta,\delta} \leq \frac{1}{2}(R-\delta)$. Denoting $c_\delta \triangleq \sup_{\eta>1} c_{\eta,\delta}$, (37) further gives

$$\mathbb{P}\left(D\left(P^{(\mathbf{s})}_{\mathsf{V}|\mathsf{B}_n}\Big\|Q^n_{V|S=\mathbf{s}}\right) \geq c_\delta n2^{-n\beta_{\eta,\delta}}\right) \leq \left(1+|\mathcal{V}|^n\right)e^{-\frac{1}{3}2^{n\delta}}. \tag{38}$$

Since (38) is true for all $\eta > 1$, it must also be true, with strict inequality in the LHS, when replacing $\beta_{\eta,\delta}$ with

$$\gamma_\delta \triangleq \sup_{\eta>1}\beta_{\eta,\delta}$$
$$= \sup_{\eta>1}\frac{\eta-1}{2\eta-1}\left(R - \delta - \max_{s\in\mathcal{S}} d_\eta(Q_{U,V|S=s}, Q_{U|S=s}Q_{V|S=s})\right)$$

which is the exponential rate of convergence stated in (14a) that we derive for the heterogeneous strong soft-covering lemma. This establishes the statement from (13) and proves Lemma 1.

Concluding, if $R > I(U; V|S)+\zeta$, for any $\zeta > 0$ arbitrarily small, then for any $\delta \in \left(0, R-I(U; V|S)\right)$, we get exponential convergence of the relative entropy at rate $O(2^{-n\gamma_\delta})$ with double-exponential certainty. Discarding the precise exponents of convergence and coefficients, we state that there exist $\gamma_1, \gamma_2 > 0$, such that for n large enough

$$\mathbb{P}\left(D\left(P^{(\mathbf{s})}_{\mathsf{V}|\mathsf{B}_n}\Big\|Q^n_{V|S=\mathbf{s}}\right) > e^{-n\gamma_1}\right) \leq e^{-e^{n\gamma_2}}. \tag{39}$$

∎

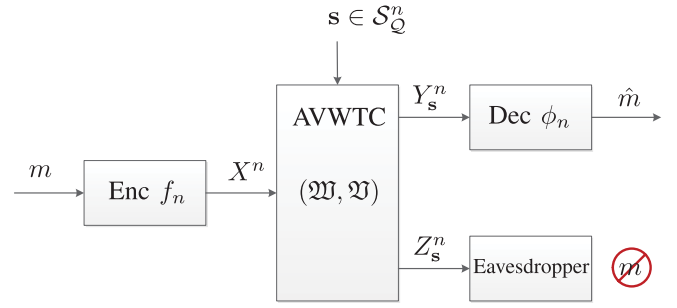

Fig. 2. The AVWTC with $\mathcal{Q}$-constrained states, i.e., when the allowed state sequences have empirical PMFs that belong to $\mathcal{Q}$.

IV. ARBITRARILY VARYING WIRETAP CHANNELS WITH TYPE CONSTRAINED STATES

A. Problem Setup and Definitions

Let $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{Z}$ and $\mathcal{S}$ be finite sets. A discrete-memoryless (DM) arbitrarily varying wiretap channel (AVWTC), as illustrated in Fig. 2, is defined by a pair $(\mathfrak{W}, \mathfrak{V})$ of families of channels $\mathfrak{W} = \left\{W_s : \mathcal{X} \to \mathcal{P}(\mathcal{Y})\big|s \in \mathcal{S}\right\}$ and $\mathfrak{V} = \left\{V_s : \mathcal{X} \to \mathcal{P}(\mathcal{Z})\big|s \in \mathcal{S}\right\}$, from $\mathcal{X}$ to $\mathcal{Y}$ and $\mathcal{Z}$, respectively. Thus, $s \in \mathcal{S}$ denotes the state of the channels and can be interpreted as an index identifying a particular pair $(W, V) \in \mathfrak{W} \times \mathfrak{V}$.

The n-th extension of the channel laws for input $\mathbf{x} \in \mathcal{X}^n$ and outputs $\mathbf{y} \in \mathcal{Y}^n$ and $\mathbf{z} \in \mathcal{Z}^n$, under the state sequence $\mathbf{s} \in \mathcal{S}^n$ are

$$W^n_{\mathbf{s}}(\mathbf{y}|\mathbf{x}) \triangleq \prod_{i=1}^n W_{s_i}(y_i|x_i) \tag{40a}$$

$$V^n_{\mathbf{s}}(\mathbf{z}|\mathbf{x}) \triangleq \prod_{i=1}^n V_{s_i}(z_i|x_i). \tag{40b}$$

The families of channels $W^n_{\mathbf{s}} : \mathcal{X}^n \to \mathcal{P}(\mathcal{Y}^n)$ and $V^n_{\mathbf{s}} : \mathcal{X}^n \to \mathcal{P}(\mathcal{Z}^n)$, for $\mathbf{s} \in \mathcal{S}^n$, are denoted by $\mathfrak{W}^n$ and $\mathfrak{V}^n$, respectively, and $(\mathfrak{W}^n, \mathfrak{V}^n)$ is referred to as the (n-fold) AVWTC. The random variables representing the outputs of the AVWTC $(\mathfrak{W}^n, \mathfrak{V}^n)$ observed by the legitimate user and by the eavesdropper under the state sequence $\mathbf{s} \in \mathcal{S}^n$ are denoted by $Y^n_{\mathbf{s}}$ and $Z^n_{\mathbf{s}}$, respectively.

For any $\mathcal{Q} \subseteq \mathcal{P}(\mathcal{S})$ define

$$\mathcal{S}^n_{\mathcal{Q}} \triangleq \left\{\mathbf{s} \in \mathcal{S}^n\big|\nu_{\mathbf{s}} \in \mathcal{Q}\right\}. \tag{41}$$

We impose a constraint $\mathcal{Q}$ on the allowed state sequences, i.e., only $\mathbf{s} \in \mathcal{S}^n_{\mathcal{Q}}$ are permitted. The triple $(\mathfrak{W}^n, \mathfrak{V}^n, \mathcal{Q})$ is referred to as the (n-fold) $\mathcal{Q}$-constrained AVWTC. We subsequently focus on the *type constrained AVWTC* formally defined in Definitions 9 and 10. Intuitively, one may think of the type constrained scenario as corresponding to $\mathcal{Q}$ being a singleton, i.e., $\mathcal{Q} = \{Q_S\}$, for some $Q_S \in \mathcal{P}(\mathcal{S})$. However, such a setting would not be well-defined if Q_S is not a rational distribution. Even if Q_S is rational, it is not a valid type for all $n \in \mathbb{N}$, thus restricting the feasible blocklengths for the AVWTC. To circumvent these pathologies, in Definitions 9 and 10 we restrict the state sequences to have types that are close to Q_S, but not necessarily Q_S itself.

Remark 3: One easily verifies that defining an AVWTC in terms of the pair $(\mathfrak{W}^n, \mathfrak{V}^n)$ is without loss of generality. In general, any state-input pair $(s, x) \in \mathcal{S} \times \mathcal{X}$ induces a joint conditional output PMF $U_s(\cdot, \cdot|x) \in \mathcal{P}(\mathcal{Y} \times \mathcal{Z})$. However, the performance of any of the codes defined below is measured with respect to the marginal output PMFs $W_s(\cdot|x) \in \mathcal{P}(\mathcal{Y})$ and $V_s(\cdot|x) \in \mathcal{P}(\mathcal{Z})$. Thus, under the framework presented here, all AVWTCs with the same marginals W and V are equivalent.

Definition 3 (Uncorrelated Code): An uncorrelated (n, M_n)-code c_n for the AVWTC $(\mathfrak{W}^n, \mathfrak{V}^n)$ has a message set $\mathcal{M}_n = [1 : M_n]$, a stochastic encoder $f_n : \mathcal{M}_n \to \mathcal{P}(\mathcal{X}^n)$ and decoder $\phi_n : \mathcal{Y}^n \to \hat{\mathcal{M}}_n$, where $\hat{\mathcal{M}}_n \triangleq \mathcal{M}_n \cup \{e\}$ and $e \notin \mathcal{M}_n$ is an error symbol.

For any uncorrelated (n, M_n)-code c_n and state sequence $\mathbf{s} \in \mathcal{S}^n$, the induced joint PMF on $\mathcal{M}_n \times \mathcal{X}^n \times \mathcal{Y}^n \times \mathcal{Z}^n \times \hat{\mathcal{M}}$ is

$$P^{(c_n, \mathbf{s})}_{M, \mathbf{X}, \mathbf{Y}_\mathbf{s}, \mathbf{Z}_\mathbf{s}, \hat{M}}(m, \mathbf{x}, \mathbf{y}, \mathbf{z}, \hat{m})$$
$$\triangleq P_M(m) f_n(\mathbf{x}|m) W^n_\mathbf{s}(\mathbf{y}|\mathbf{x}) V^n_\mathbf{s}(\mathbf{z}|\mathbf{x}) \mathbb{1}_{\{\hat{m} = \phi_n(\mathbf{y})\}}, \quad (42)$$

where $P_M \in \mathcal{P}(\mathcal{M}_n)$. The performance of c_n on the type constrained AVWTC $(\mathfrak{W}^n, \mathfrak{V}^n, Q_S)$ is evaluated in terms of its rate $\frac{1}{n} \log M_n$, the maximal decoding error probability and the SS-metric. Reliability and security must be ensured with respect to every allowed constrained state sequence.

Definition 4 (Message Error Probability): Let c_n be an uncorrelated (n, M_n)-code for the AVWTC $(\mathfrak{W}^n, \mathfrak{V}^n)$. For any $m \in \mathcal{M}_n$ and $\mathbf{s} \in \mathcal{S}^n$, let $e_m(W^n_\mathbf{s}, c_n)$ be the error probability in decoding m under the state sequence $\mathbf{s}$, given by

$$e_m(W^n_\mathbf{s}, c_n) = \sum_{\mathbf{x} \in \mathcal{X}^n} f_n(\mathbf{x}|m) \sum_{\substack{\mathbf{y} \in \mathcal{Y}^n : \\ \phi_n(\mathbf{y}) \neq m}} W^n_\mathbf{s}(\mathbf{y}|\mathbf{x}). \quad (43)$$

Definition 5 (SS Metric): Let c_n be an uncorrelated (n, M_n)-code for the AVWTC $(\mathfrak{W}^n, \mathfrak{V}^n)$. The information leakage to the eavesdropper under the state sequence $\mathbf{s} \in \mathcal{S}^n$ and the message PMF $P_M \in \mathcal{P}(\mathcal{M}_n)$ is

$$\ell(V^n_\mathbf{s}, P_M, c_n) = I_{c_n}(M; \mathbf{Z}_\mathbf{s}), \quad (44)$$

where the subscript c_n denotes that the mutual information term is calculated with respect to the induced joint distribution $P^{(c_n, \mathbf{s})}_{M, \mathbf{Z}_\mathbf{s}}$ from (42). For any $Q \subseteq \mathcal{P}(\mathcal{S})$, the SS metric with respect to c_n and the Q-constrained AVWTC $(\mathfrak{W}^n, \mathfrak{V}^n, Q)$ is[7]

$$\ell_{\text{Sem}}(\mathfrak{V}^n, Q, c_n) = \max_{\substack{\mathbf{s} \in \mathcal{S}^n_Q, \\ P_M \in \mathcal{P}(\mathcal{M}_n)}} \ell(V^n_\mathbf{s}, P_M, c_n). \quad (45)$$

Remark 4: We use the convention that the maximum over an empty set is $-\infty$. Accordingly if Q contains no rational distributions then $\ell_{\text{Sem}}(\mathfrak{V}^n, Q, c_n) = -\infty$, for all $n \in \mathbb{N}$. Even when there exists $Q_S \in \mathcal{P}_n(\mathcal{S})$ such that $Q_S \in Q$, there are blocklengths n for which $v_\mathbf{s} \neq Q_S$ for every $\mathbf{s} \in \mathcal{S}^n$, and consequently, $\ell_{\text{Sem}}(\mathfrak{V}^n, Q, c_n) = -\infty$ for these values of n as well.

Remark 5: SS requires that the uncorrelated code c_n works well for all message PMFs. This means that the mutual

[7] $\ell_{\text{Sem}}(\mathfrak{V}^n, Q, c_n)$ is actually the mutual-information-security (MIS) metric, which is equivalent to SS by [37]. We use the representation in (45) rather than the formal definition of SS (see, e.g., [37, eq. (4)]) out of analytical convenience.

information term in (45) is maximized over P_M when c_n is known. In other words, although not stated explicitly, the optimal P_M is a function of c_n.

We proceed with defining correlated random (CR) codes, their associated maximal error probability and SS-metric, CR-assisted achievability and CR-assisted secrecy-capacity.

Definition 6 (CR Code, Error Probability and SS Metric): A CR (n, M_n, K_n)-code $\mathbf{C}_n$ for the AVWTC $(\mathfrak{W}^n, \mathfrak{V}^n)$ is given by a family of uncorrelated (n, M_n)-codes $\mathcal{C}_n = \{c_n(\gamma)\}_{\gamma \in \Gamma_n}$, where $\Gamma_n = [1 : K_n]$, and a PMF $\mu_n \in \mathcal{P}(\Gamma_n)$. For any $m \in \mathcal{M}_n$ and $\mathbf{s} \in \mathcal{S}^n$, the associated error probability with respect to $\mathbf{C}_n$ is

$$\mathcal{E}_m(W^n_\mathbf{s}, \mathbf{C}_n) = \sum_{\gamma \in \Gamma_n} \mu_n(\gamma) e_m(W^n_\mathbf{s}, c_n(\gamma)) \quad (46)$$

The maximal error probability and SS-metric of $\mathbf{C}_n$ for the Q-constrained AVWTC $(\mathfrak{W}^n, \mathfrak{V}^n, Q)$ are defined as

$$\mathcal{E}(\mathfrak{W}^n, Q, \mathbf{C}_n) = \max_{\substack{\mathbf{s} \in \mathcal{S}^n_Q, \\ m \in \mathcal{M}_n}} \mathcal{E}_m(W^n_\mathbf{s}, \mathbf{C}_n) \quad (47a)$$

$$\mathcal{L}_{\text{Sem}}(\mathfrak{V}^n, Q, \mathbf{C}_n) = \max_{\gamma \in \Gamma_n} \ell_{\text{Sem}}(\mathfrak{V}^n, Q, c_n(\gamma))$$
$$= \max_{\substack{\gamma \in \Gamma_n, \\ \mathbf{s} \in \mathcal{S}^n_Q, \\ P_M \in \mathcal{P}(\mathcal{M}_n)}} \ell(V^n_\mathbf{s}, P_M, c_n(\gamma)). \quad (47b)$$

Remark 6: The choice of encoder-decoder in a CR code is based on a realization of a random experiment that is available to the transmitted and the legitimate receiver. However, this CR the legitimate users share should not be viewed as a cryptographic key to be exploited for secrecy. This is accounted for in (47b) by requiring that every uncorrelated code in the family $\mathcal{C}_n$ is semantically-secure. The choice of the state sequence, on the other hand, may depend on the family $\mathcal{C}_n$ and the PMF μ_n, but not on the realization itself.

Definition 7 (CR-Assisted Achievability): A number $R \in \mathbb{R}_+$ is called an achievable CR-assisted SS-rate for the Q-constrained AVWTC $(\mathfrak{W}^n, \mathfrak{V}^n, Q)$, if for every $\epsilon > 0$ and sufficiently large n, there exists a CR (n, M_n, K_n)-code $\mathbf{C}_n$ with

$$\frac{1}{n} \log M_n > R - \epsilon \quad (48a)$$
$$\mathcal{E}(\mathfrak{W}^n, Q, \mathbf{C}_n) \leq \epsilon \quad (48b)$$
$$\mathcal{L}_{\text{Sem}}(\mathfrak{V}^n, Q, \mathbf{C}_n) \leq \epsilon. \quad (48c)$$

Remark 7: Note that if there are no types in Q then any rate is achievable. Consequently, if $Q_1 \subseteq Q_2 \subseteq \mathcal{P}(\mathcal{S})$, then any R that is achievable for the Q_2-constrained AVWTC is also achievable for the Q_1-constrained AVWTC. The achievable rates are therefore an increasing set as the constraint set decreases. When specializing to the type constrained AVWTC, we allow state sequences with types that are δ-close to the constraining distribution (see Definition 10). Consequently, the set of feasible state sequences is never empty, for large enough values of n. Nonetheless, the aforementioned monotonicity of the type constrained CR-assisted capacity still holds.

Remark 8: Our achievability proof shows that $\mathcal{L}_{\text{Sem}}(\mathfrak{V}^n, \mathcal{Q}, \mathsf{C}_n)$ *vanishes exponentially fast. This is a standard requirement in cryptography, commonly referred to as strong-SS (see, e.g., [37, Sec. 3.2]).*

Definition 8 (CR-Assisted Capacity): The CR-assisted SS-capacity $C_{\text{R}}(\mathfrak{W}, \mathfrak{V}, \mathcal{Q})$ *of the* $\mathcal{Q}$*-constrained AVWTC is the supremum of the set of achievable CR-assisted SS-rates.*

Our main goal is solving the type constrained AVWTC $(\mathfrak{W}^n, \mathfrak{V}^n, Q_S)$, for $Q_S \in \mathcal{P}_n(\mathcal{S})$. However, since a fixed rational distribution Q_S is not a valid type for all blocklengths, the definitions of the type constrained performance metrics and its achievability use a relaxation parameter. For any $Q_S \in \mathcal{P}(\mathcal{S})$ and $\delta > 0$, let $\mathcal{Q}_\delta(Q_S) \triangleq \left\{ \nu_{\mathbf{s}} \in \mathcal{P}_n(\mathcal{S}) \Big| \mathbf{s} \in \mathcal{T}_\delta^n(Q_S) \right\}$. The definitions of the error probability and the SS-metric for the type constrained AVWTC repeat those from Definition 6 with $\mathcal{Q}_\delta(Q_S)$ instead of $\mathcal{Q}$.

Definition 9 (Type Constrained Error Probability & SS): For any $Q_S \in \mathcal{P}(\mathcal{S})$ *and* $\delta > 0$*, the maximal error probability and SS-metric of a CR* (n, M_n, K_n)*-code* C_n *for the type constrained AVWTC* $(\mathfrak{W}^n, \mathfrak{V}^n, Q_S)$ *with relaxation* δ *are* $\mathcal{E}(\mathfrak{W}^n, \mathcal{Q}_\delta(Q_S), \mathsf{C}_n)$ *and* $\mathcal{L}_{\text{Sem}}(\mathfrak{V}^n, \mathcal{Q}_\delta(Q_S), \mathsf{C}_n)$*, respectively (see (47)).*

Definition 10 (Type Constrained Achievability & Capacity): A number $R \in \mathbb{R}_+$ *is called an achievable CR-assisted SS-rate for the type constrained AVWTC* $(\mathfrak{W}^n, \mathfrak{V}^n, Q_S)$*, if there exists a* $\delta > 0$ *such that for every* $\epsilon > 0$ *and sufficiently large n, there exists a CR* (n, M_n, K_n)*-code* C_n *with*

$$\frac{1}{n} \log M_n > R - \epsilon \qquad (49a)$$

$$\mathcal{E}(\mathfrak{W}^n, \mathcal{Q}_\delta(Q_S), \mathsf{C}_n) \le \epsilon \qquad (49b)$$

$$\mathcal{L}_{\text{Sem}}(\mathfrak{V}^n, \mathcal{Q}_\delta(Q_S), \mathsf{C}_n) \le \epsilon. \qquad (49c)$$

The CR-assisted SS-capacity $C_{\text{R}}(\mathfrak{W}, \mathfrak{V}, Q_S)$ *of the type constrained AVWTC* $(\mathfrak{W}^n, \mathfrak{V}^n, Q_S)$ *is the supremum of the set of achievable CR-assisted SS-rates.*

Remark 9: The definition of type constrained achievability allows the empirical PMFs of the state sequences to be within a $\delta > 0$ *gap from* Q_S*. By doing so, the type constrained AVWTC is well-defined for all sufficiently large blocklengths* $n \in \mathbb{N}$*, even when* Q_S *is not actually a type but a PMF on* $\mathcal{S}$*.*

Remark 10: If R is an achievable CR-assisted SS-rate for the type constrained AVWTC $(\mathfrak{W}^n, \mathfrak{V}^n, Q_S)$ *and* $\delta > 0$ *is its corresponding parameter, then (49b)-(49c) also hold for any* $\delta' \in (0, \delta)$*. This implies that* $C_{\text{R}}(\mathfrak{W}, \mathfrak{V}, Q_S) = \sup_{\delta > 0} C_{\text{R}}(\mathfrak{W}, \mathfrak{V}, \mathcal{Q}_\delta(Q_S))$*.*

B. Single-Letter CR-Capacity Results

Our main result is a single-letter characterization of the CR-assisted SS-capacity $C_{\text{R}}(\mathfrak{W}, \mathfrak{V}, Q_S)$ of the type constrained AVWTC $(\mathfrak{W}^n, \mathfrak{V}^n, Q_S)$, for any $Q_S \in \mathcal{P}(\mathcal{S})$. A multi-letter characterization of the CR-assisted strong secrecy-capacity of the AVWTC without constraints on the state sequences was found in [24]. The uncorrelated secrecy-capacity was then derived in [22] by relating it to the CR-assisted secrecy-capacity via the corresponding coding result for the classic AVC [15]. To the best of our knowledge,

the only single-letter characterization of a secrecy-capacity of an AVWTC outside the current work [21, Th. 4] is under the following assumptions: (i) security under the weak secrecy metric (as shown in [24, Corollary 1] an upgrade to strong secrecy under the same conditions (ii)-(iv) is possible); (ii) the state space decomposes as $\mathcal{S} = \mathcal{S}_y \times \mathcal{S}_z$, where $s_y \in \mathcal{S}_y$ and $s_z \in \mathcal{S}_z$ are the states of the main AVC and of the AVC to the eavesdropper, respectively; (iii) the eavesdroppers output is a degraded version of the output of the main AVC under any pair of state, i.e., $X - Y_{s_y} - Z_{s_z}$ forms a Markov chain, for all $(s_y, s_z) \in \mathcal{S}_y \times \mathcal{S}_z$; (iv) there exists a best channel to the eavesdropper, i.e., the exists $s_z^\star \in \mathcal{S}_z$ such that $X - Z_{s_z^\star} - Z_{s_z}$ forms a Markov chain, for all $s_z \in \mathcal{S}_z$.[8]

Our single-letter CR-capacity characterization is derived without assuming any of the above, while upgrading the secrecy metric to SS. Constrained state sequences were considered in the context of the classic point-to-point (PTP) AVC and the corresponding CR-assisted and uncorrelated capacities were derived in [15] and [14], respectively. An AVWTC with linear peak constraints on the input and the state sequences was studied in [35], where a multi-letter description of its CR-assisted strong secrecy-capacity was found.

Theorem 1 (AVWTC CR-Assisted SS-Capacity): For any $Q_S \in \mathcal{P}(\mathcal{S})$*, the CR-assisted SS-capacity of the type constrained AVWTC* $(\mathfrak{W}^n, \mathfrak{V}^n, Q_S)$ *is*

$$C_{\text{R}}(\mathfrak{W}, \mathfrak{V}, Q_S) = \max_{Q_{U,X}} \left[I(U; Y) - I(U; Z|S) \right], \qquad (50)$$

where the mutual information terms are calculated with respect to a joint PMF $Q_{U,X} Q_S Q_{Y|X,S} Q_{Z|X,S}$ *with* $Q_{Y|X,S}(y|x,s) = W_s(y|x)$ *and* $Q_{Z|X,S}(z|x,s) = V_s(z|x)$*, for all* $(s, x, y, z) \in \mathcal{S} \times \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$*, and* $|\mathcal{U}| \le |\mathcal{X}|$*.*

Theorem 1 is a consequence of two other stronger results that state a lower and upper bound on the CR-assisted SS-capacity of a general $\mathcal{Q}$-constrained AVWTC. These bounds match when specialized to the type constrained scenario. The lower and upper bounds are given in Theorem 2 and 3, respectively. Theorem 1 is proven in Section IV-E.

Remark 11 (SS-Capacity Interpretation): The characterization of the CR-assisted SS-capacity $C_{\text{R}}(\mathfrak{W}, \mathfrak{V}, Q_S)$ *in (50) has the common structure of two subtracted mutual information terms. The first term, which corresponds to the capacity of the main channel, suggests that the legitimate users effectively see the averaged DMC* $W_Q : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$ *defined by* $W_Q(y|x) \triangleq \sum_{s \in \mathcal{S}} Q_S(s) W_s(y|x)$*. In general, the capacity of the averaged channel is no larger than the capacities of the main channels* W_s *associated with each* $s \in \mathcal{S}$*. Namely, denoting the capacity of a PTP channel* $W : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$ *by* $C(W)$*, it holds that* $C(W_Q) \le \min_{s \in \mathcal{S}} C(W_s)$*. This is due to the convexity of the mutual information in the conditional PMF (for a fixed marginal) and Jensen's inequality.*

The second (subtracted) mutual information term is the loss in capacity induced by the secrecy requirement. The

---

[8]An even stronger version of assumptions (iii) and (iv) was used in [21]. Specifically, the degraded property and the existence of a best channel to the eavesdropper were assumed to hold not only for every pair of original states, but also for any pair of averaged states (defined as convex combinations of the original ones).

independence of U and S allows one to rewrite the conditional mutual information as $I(U; S, Z)$, which implies that security must be ensured versus an eavesdropper with perfect CSI.

Remark 12 (Relation to the IID State Scenario): The for-mula in (50) can be viewed as the secrecy-capacity of the WTC with state variables that are i.i.d. according to Q_S, when no CSI is available to the legitimate users while the eavesdropper has full CSI. For simplicity, we outline a proof of this claim under the strong secrecy metric; an upgrade to SS is possible by means of the Lemma 1 herein. The direct part follows by constructing a classic WTC code using i.i.d. samples of a random variable $U \sim Q_U$ that is independent of $S \sim Q_S$. Setting the total number of codewords just below $2^{nI(U;Y)}$ ensures reliable decoding, while strong secrecy follows by standard soft-covering arguments as long as each subcode-book has a rate that is at least $I(U; Z, S)$ (see, [45, Corollay VII.5]). The converse essentially follows by repeating the steps between Equations (134)-(138) from Section VI, while omitting the conditioning on the random variable C_n in (134).

Remark 13 (Cardinality Bound): The cardinality bound on $\mathcal{U}$ in Theorem 1 is established using a standard application of the Eggleston-Fenchel-Carathéodory Theorem [49, Th. 18]. The details are omitted.

We have the following lower bound on the CR-assisted SS-capacity of a $\mathcal{Q}$-constrained AVWTC.

Theorem 2 (Achievability with $\mathcal{Q}$-Constrained States): For any convex and closed $\mathcal{Q} \subseteq \mathcal{P}(\mathcal{S})$, the CR-assisted SS-capacity of the $\mathcal{Q}$-constrained AVWTC $(\mathfrak{W}^n, \mathfrak{V}^n, \mathcal{Q})$ is lower bounded as

$$C_{\mathrm{R}}(\mathfrak{W}, \mathfrak{V}, \mathcal{Q}) \geq \max_{Q_{U,X}} \left[\min_{Q_S^{(1)} \in \mathcal{Q}} I(U; Y) - \max_{Q_S^{(2)} \in \mathcal{Q}} I(U; Z|S) \right],$$
(51)

where the mutual information terms are calculated with respect to joint PMFs $Q_{U,X} Q_S^{(j)} Q_{Y|X,S} Q_{Z|X,S}$, for $j = 1, 2$, with $Q_{Y|X,S}(y|x,s) = W_s(y|x)$ and $Q_{Z|X,S}(z|x,s) = V_s(z|x)$, for all $(s, x, y, z) \in \mathcal{S} \times \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$, and $|\mathcal{U}| \leq |\mathcal{X}|$.

The proof of Theorem 2 is given in Section V. It relies on the approach from [15] for the error probability analysis of a CR-code over a family of codes that grows doubly-exponentially with the blocklength. Since this family of codes is too large to establish SS in the sense of (47b), we use the Chernoff bound to show that a sub-family with no more than polynomially many codes is sufficient for reliability. Having that, the double-exponential decay that Lemma 1 provides is leveraged to establish SS over the reduced CR-code. The fact that reliability and security must hold with respect to the worst case choice in $\mathcal{Q}$ is expressed in the minimization of $I(U; Y)$ over all $Q_S^{(1)}$ PMFs and the maximization of $I(U; Z|S)$ over $Q_S^{(2)}$.

Although no converse proof accompanies Theorem 2, the lower bound it states is stronger than existing single-letter achievability results in the literature as it assumes no 'best channel to the eavesdropper', doesn't impose any specific structure on the state space, and ensures SS.

Remark 14 (AVWTC Main Challenges): The difficulty in obtaining single-letter results for the AVWTC is twofold. First,

the AVWTC must (in particular) satisfy all the performance requirements of the compound WTC (where the channel remains constant throughout the block transmission). The sec-ond difficulty is in ensuring security under exponentially many possible state sequences. Setting sights on single-letter results, the common workaround for the latter problem is to assume the existence of 'a best channel to the eavesdropper'. Formally, it means that there exists a PMF $Q^\star \in \mathcal{P}(\mathcal{S})$, such that the averaged channel $\sum_{s \in \mathcal{S}} Q^\star(s) V_s(y|x)$ is better than all other averaged channels in the sense that the corresponding outputs form a Markov chain with the channel input X. Secrecy is then guaranteed with respect to this 'best channel' only. As Theorem 2 is derived without any assumptions on the AVWTC, it highlights the strength of Lemma 1 in proving that exponentially many secrecy constraints (strongly related to the soft-covering phenomenon) are simultaneously satisfied, while only single-letter rate bounds are needed.

Remark 15 (Relation to Compound WTCs): Theorem 2 establishes that the AVWTC is no worse than the best known single-letter secrecy rates for the compound WTC. Take the $\mathcal{Q}$-constrained AVWTC from Theorem 2 with some convex and closed $\mathcal{Q} \subseteq \mathcal{P}(\mathcal{S})$. Consider a compound WTC derived from this AVWTC. The state of the compound WTC is any point $Q_S \in \mathcal{Q}$. The compound WTC itself follows the probability law of the AVWTC, with the arbitrarily varying state S^n replaced by an i.i.d. state according to Q_S and the S^n sequence included in the channel output to the eavesdropper. For this compound WTC, the RHS of (51) coincides with the sharpest single-letter lower bound on the secrecy-capacity of the compound WTC in the literature (see [28, Th. 1] and [32, Th. 3.6]).

A general upper bound on the CR-assisted SS-capacity of a $\mathcal{Q}$-constrained AVWTC is given next. To state the result, for any countable alphabet $\mathcal{X}$ we define $\mathcal{P}_{\mathbb{Q}}(\mathcal{X})$ as the set of rational PMFs on $\mathcal{X}$. Namely,

$$\mathcal{P}_{\mathbb{Q}}(\mathcal{X}) \triangleq \left\{ P \in \mathcal{P}(\mathcal{X}) \Big| P(x) \in \mathbb{Q}, \quad \forall x \in \mathcal{X} \right\}. \quad (52)$$

Theorem 3 (Upper Bound with $\mathcal{Q}$-constrained States): For any $\mathcal{Q} \subseteq \mathcal{P}(\mathcal{S})$, the CR-assisted SS-capacity of the $\mathcal{Q}$-constrained AVWTC $(\mathfrak{W}^n, \mathfrak{V}^n, \mathcal{Q})$ is upper bounded as

$$C_{\mathrm{R}}(\mathfrak{W}, \mathfrak{V}, \mathcal{Q})$$
$$\leq \max_{Q_{V,U,X}} \inf_{Q_S \in \mathcal{Q} \cap \mathcal{P}_{\mathbb{Q}}(\mathcal{S})} \left[I(U; Y|V) - I(U; S, Z|V) \right], \quad (53)$$

where the mutual information terms are calculated with respect to a joint PMF $Q_{V,U,X} Q_S Q_{Y|X,S} Q_{Z|X,S}$ with $Q_{Y|X,S}(y|x,s) = W_s(y|x)$ and $Q_{Z|X,S}(z|x,s) = V_s(z|x)$, for all $(s, x, y, z) \in \mathcal{S} \times \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$. Furthermore, one may restrict $|\mathcal{U}| \leq |\mathcal{X}|$ and $|\mathcal{V}| \leq |\mathcal{X}|^2 - 1$.

The proof of Theorem 3 is given in Section VI. The max-inf structure of the RHS of (53) calls for a derivation that is uniform in $Q_S \in \mathcal{Q}$. The infimum is taken over $\mathcal{Q} \cap \mathcal{P}_{\mathbb{Q}}(\mathcal{S})$ (rather than over $\mathcal{Q}$) because the proof effectively considers only the rational distributions in $\mathcal{Q}$ while leveraging the monotonicity of the CR-assisted SS-capacity with respect to $\mathcal{Q}$ (see Remark 7). The proof relies on a novel argument based on distribution coupling. We show that for each $Q_S \in \mathcal{Q} \cap \mathcal{P}_{\mathbb{Q}}(\mathcal{S})$, reliability and SS under a type constraint Q_S imply similar

performance for the same channel but where the state sequence is i.i.d. according to Q_S. The main difficulty is in showing that even when transmitting over a DMC obtained by averaging the $W_s \in \mathfrak{W}$ with respect to Q_S, the normalized equivocation of the message given the output sequence at the *legitimate user* is still small.

Usually, Fano's inequality is sufficient for $\frac{1}{n} H(M|Y^n)$ to become arbitrarily small with n. This, however, is not the case here. The reliability criterion from (48b) and Fano's inequality imply that $\max_{\mathbf{s} \in \mathcal{S}_Q^n} \frac{1}{n} H(M|Y_{\mathbf{s}}^n)$ is small, but Theorem 3 needs this to hold for $\frac{1}{n} H(M|Y^n)$. In general, the former must not imply the latter because for any $\mathbf{s} \in \mathcal{S}_Q^n$, the channel $W_{\mathbf{s}} \in \mathfrak{W}$ is at least as good as W_Q, meaning that the averaged channel induces a possibly larger equivocation. We establish the desired convergence of the equivocation at the legitimate receiver by a continuity property which we derive via the coupling idea.

Remark 16 (Relation to the Converse of Theorem 1): The converse for Theorem 1 is derived based on Theorem 3. Since the latter is valid for any $\mathcal{Q} \subset \mathcal{P}(\mathcal{S})$, combining it with basic continuity arguments implies the optimality of the RHS of (50). However, Theorem 3 encapsulates an even stronger claim: the RHS of (50) is the best achievable CR-assisted SS-rate even if the state sequence is constrained to a single type (that potentially might allow higher rates), rather than a vanishing typical set.

Although clearly sufficient, the strong claim of Theorem 3 is not necessary for the converse of Theorem 1. In fact, one can circumvent the main difficulty in proving Theorem 3 (as described above), by establishing the optimality of the RHS of (50) based on arguments similar to those used for the converse of the classic AVC with constrained states [14, Lemma 3.2]. Specifically, the probability of a decoding error under a random state sequence that is i.i.d. according to Q_S can be shown to be small simply by splitting the analysis to typical and atypical state sequences. The definition of the type constrained CR-assisted achievability (Definition 10), that accounts for a typical set around Q_S, takes care of the typical part, while the atypical part is bounded above by the exponentially decaying probability of the atypical set. Having that, a standard application of Fano's inequality implies that $\frac{1}{n} H(M|Y^n)$ converges to 0 with the blocklength, as required.

This simple argument, however, does not apply when there is an actual type constraint (rather than a typical set constraint) on the state sequences. This is because for any $Q_S \in \mathcal{P}_n(\mathcal{S})$, the probability of an i.i.d. Q_S sequence of states not being in $\mathcal{T}_{Q_S}^n$ approaches 1 when n grows. As a consequence, the corresponding decoding error probability might not be small. The proof of Theorem 3 does not directly deals with the error probability. Instead, the aforementioned distribution coupling arguments and continuity of entropy are used to show that the normalized equivocation converges for state sequences in the entire typical set, as long as it converges for at least one specific type in that typical set. Although the proof Theorem 3 is cumbersome and requires several non-trivial steps, we take this route (rather than a converse tailored for Theorem 1) due to the stronger and insightful claim it produces.

Remark 17 (Time-Sharing Random Variable V): The conditioning on V in the RHS of (53) effectively allows the legitimate user to choose a random mixture of $Q_{U,X}$ distributions. The advantage in doing so is that there might not exist a single state distribution that is bad for the whole mixture. This is reminiscent of a two-player zero-sum game, where the player who fixes the strategy first often benefits from a mixed strategy. When specializing to the type constrained scenario, however, the time-sharing random variable is removed. This is since when only one state distribution is allowed, the aforementioned distribution mixing outcomes with no gain.

Remark 18 (Relation to Compound WTCs): The best previously known single-letter upper bound on the secrecy-capacity of the compound WTC is due to Liang et al. [28, Th. 2]. That upper bound has a min-max structure, and it is derived by claiming that the secrecy-capacity of the compound WTC is bounded above by that of the worst WTC in the set. This type of bounds are commonly related to knowledge of the channel's state at the transmitter (cf. e.g., [50]). Indeed, as shown in [32], the upper bound from [28] is tight for the compound WTC with encoder CSI.

Specializing the max-inf upper bound from Theorem 3 to the compound WTC described in Remark 15 (i.e., over an appropriate constraint set), results in a strengthening of the claim from [28, Th. 2]. The obtained bound first minimizes the difference of mutual information terms from the RHS of (53) over the constraint set, and then maximizes the outcome over the input distribution. It is easily observed the difference between the two bounds can be strict. In fact, for the special case of a PTP compound channel (i.e., without an eavesdropper) our bound is the actual capacity, while the bound from [28] is loose. A simple example is a channel that consists of two orthogonal binary channels: one is noise free while the other one is purely noise (i.e., binary symmetric channel with crossover probability $\frac{1}{2}$). The state determines which channel is noisy, and the transmitter selects a binary input, which is unknown to the receiver, specifying which channel to use (both channels give an output each time, with one being pure noise). For this instance, the compound capacity is $\frac{1}{2}$ [bit/use], but the looser min-max bound gives 1 [bit/use].

Theorem 3 essentially says that the $\mathcal{Q}$-constrained AVWTC is no better than the compound WTC under the corresponding constraints. Although this point seems intuitively obvious, it actually requires some careful attention. At first glance, the compound channel (on a constraint set) seems like an AVWTC with a stricter restriction on the eavesdropper, that now must choose an i.i.d. state sequence from the constraint set (rather than an arbitrary one). However, this perspective is misleading since the i.i.d. state sequence might not fall within the type constraint, especially when dealing with a single type.

Remark 19 (Cardinality Bound): The cardinality bound on $\mathcal{U}$ and $\mathcal{V}$ in Theorem 3 follow by applying the Eggleston-Fenchel-Carathéodory Theorem [49, Th. 18] twice. The details are omitted.

A simple consequence of Theorem 3 is the following.

Corollary 1 (Upper Bound When $\mathcal{Q}$ is Open): If $\mathcal{Q} \subseteq \mathcal{P}(\mathcal{S})$ is an open set, then the CR-assisted SS-capacity of

the $\mathcal{Q}$-constrained AVWTC $(\mathfrak{W}^n, \mathfrak{V}^n, \mathcal{Q})$ is upper bounded as

$$C_R(\mathfrak{W}, \mathfrak{V}, \mathcal{Q}) \leq \max_{Q_{V,U,X}} \inf_{Q_S \in \mathcal{Q}} \left[I(U; Y|V) - I(U; Z|S, V) \right],$$
(54)

where joint PMF is as described in Theorem 3.

When $\mathcal{Q}$ is an open set, the domain of the infimum requires no intersection with $\mathcal{P}_{\mathbb{Q}}(\mathcal{S})$. This essentially follows because the rational numbers are dense in the reals and the mutual information is continuous in the underlying distribution. The full details are omitted.

C. An Example

We give a simple numerical example that visualizes the SS-capacity result of Theorem 1. Let $\mathcal{X} = \mathcal{Y} = \{0, 1\}$ and $\mathcal{Z} = \{0, 1, ?\}$, where ? is an erased symbol. Further assume that the state space $\mathcal{S}$ decomposes as $\mathcal{S} = \mathcal{S}_1 \times \mathcal{S}_2$, where $\mathcal{S}_j = \{0, 1\}$, for $j = 1, 2$. Let $(\mathfrak{W}, \mathfrak{V})$ be an AVWTC, where the elements of $\mathfrak{W}$ and $\mathfrak{V}$ are indexed by $s_1 \in \mathcal{S}_1$ and $s_2 \in \mathcal{S}_2$, respectively. Define the main channel $W_{s_1} : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$, for $s_1 \in \mathcal{S}_1$, as $W_{s_1}(y|x) = \mathbb{1}_{\{y = x \oplus s_1\}}$, where $\oplus$ denotes the modulo 2 addition. For the eavesdropper, let $V_0 : \mathcal{X} \to \mathcal{P}(\mathcal{Z})$ be a noiseless channel, while $V_1 : \mathcal{X} \to \mathcal{P}(\mathcal{Z})$ outputs the symbol ? with probability 1. Namely,

$$V_{s_2}(z|x) = \begin{cases} \mathbb{1}_{\{z = x\}}, & s_2 = 0 \\ \mathbb{1}_{\{z = ?\}}, & s_2 = 1. \end{cases}$$
(55)

Finally, we introduce a type constraint $Q_{S_1, S_2} = Q_{S_1} Q_{S_2}$ on the state sequences, where $Q_{S_1}(1) = \epsilon$ and $Q_{S_2}(1) = \alpha$, for some $\epsilon \in [0, \frac{1}{2}]$ and $\alpha \in [0, 1]$. Denote the CR-assisted SS-capacity of this AVWTC by $C_R(\epsilon, \alpha)$.

By Theorem 1, The CR-assisted SS-capacity is

$$C_R(\epsilon, \alpha) = \max_{Q_{U,X}} \left[I(U; Y) - I(U; Z|S_2) \right],$$
(56)

where the mutual information terms are calculated with respect to the joint distribution $Q_{S_1}(s_1) Q_{S_2}(s_2) Q_{U,X}(u, x) W_{s_1}(y|x) V_{s_2}(z|x)$.

Note that for any $Q_{U,X} \in \mathcal{P}(\mathcal{U} \times \mathcal{X})$, we have

$$
\begin{aligned}
& I(U; Z|S_2) \\
&= Q_{S_2}(0) I(U; Z|S_2 = 0) + Q_{S_2}(1) I(U; Z|S_2 = 1) \\
&\overset{(a)}{=} (1 - \alpha) I(U; X),
\end{aligned}
$$
(57)

where (a) is because $Z = ?$ whenever $S_2 = 1$ (thus nullifying the second mutual information term), while given on $S_2 = 0$, we have $Z = X$ and the conditioning is removed due to the independence of S_2 and (U, X). Consequently, (56) reduces to

$$C_R(\epsilon, \alpha) = \max_{Q_{U,X}} \left[I(U; Y) - (1 - \alpha) I(U; X) \right],$$
(58)

which is calculated with respect to $Q_{S_1}(s_1) Q_{U,X}(u, x) W_{s_1}(y|x)$. Now, since S_1 does not appear in any of the mutual information terms, their value remains unchanged if the above joint distribution is replaced with $Q_{U,X}(u, x) W_{Q_1}(y|x)$, where $W_{Q_1} : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$ is the average DMC $W_{Q_1}(y|x) = \sum_{s_1 \in \mathcal{S}_1} Q_{S_1}(s_1) W_{s_1}(y|x)$ (see Remark 11). Noting that the DMC W_{Q_1} is a binary

symmetric channel with crossover probability ϵ (BSC(ϵ)), we have that $C_R(\epsilon, \alpha)$ is the secrecy-capacity of BS-BE WTC with a BSC(ϵ) between the legitimate users and a binary erasure channel with erasure probability α (BEC(α)) to the eavesdropper [36].

Remark 20: Interestingly, (58) *is also the SS-capacity of the WTC of type II (WTCII) with a BSC(ϵ) to the legitimate user and an eavesdropper who can actively choose any $\lfloor n(1 - \alpha) \rfloor$ of the transmitted symbols to observe [48]. This is not surprising since the WTCII with a noisy main channel is a particular instance of a type constrained AVWTC. Consequently, its SS-capacity (stated in [50, Th. 3]) is recovered from Theorem 1 by steps similar to those presented between Equations* (56)*-*(58)*. Namely, this is done by letting the AVC between the legitimate users be a DMC and treating the eavesdropper's AVC as in* (57)*. In fact, the type constrained AVWTC captures as a special case also the generalization of the WTCII from [51], where the subset of symbols chosen by the eavesdropper is further corrupted by noise (i.e., by passing it through another DMC). The only actual difference between these WTCII models and the AVWTC is that the main channel in the formers is a DMC (and not an AVC), which makes CR-assisted codes unnecessary. However, this is the case for any AVWTC with a main DMC.*

Fig. 3 depicts the CR-assisted SS-capacity of the considered AVWTC as a function of type constraints on the main and on the eavesdropper's channels. The variation of $C_R(\epsilon, \alpha)$ as a function of $Q_{S_1}(1) = \epsilon$ for a fixed $\alpha = 0.4$ is shown in Fig. 3(a), while Fig. 3(b) presents the SS-capacity as a function of $Q_{S_2}(1) = \alpha$ when $\epsilon = 0.1$ is fixed. The curves are plotted by parametrizing the joint PMF of the binary random variables U and X and spanning over the possible probability values. As mentioned before, $C_R(\epsilon, \alpha)$ is also the secrecy-capacity of a BS-BE WTC.

This WTC was studied in [36], where it was shown that the secrecy-capacity is zero if $\alpha < 4\epsilon(1 - \epsilon)$. When $\epsilon = 0.1$ the threshold value of α is 0.36. Indeed, Fig. 3(b) reveals that $C_R(0.1, \alpha) = 0$ for any $\alpha < 0.36$. Beyond 0.36, the SS-capacity monotonically increases with α, since the larger the probability of an erasure, the worse the channel to the eavesdropper is. From the opposite perspective, a fixed $\alpha = 0.4$ induces two real solutions to the equation $0.4 = 4\epsilon(1 - \epsilon)$, which are $\epsilon_1 \approx 0.1127$ and $\epsilon_2 \approx 0.8872$. The condition $0.4 < 4\epsilon(1 - \epsilon)$ is then satisfied for any $\epsilon \in (\epsilon_1, \epsilon_2)$, which gives a zero SS-capacity in that region in Fig. 3(b). Also observe that as a function of ϵ, $C_R(\epsilon, 0.4)$ grows as the crossover probability approaches the extreme values of 0 or 1.

D. Relation Between Theorem 1 and the Multi-Letter CR-Assisted Secrecy-Capacity Characterization

We compare the single-letter result of Theorem 1 to the multi-letter description of the secrecy-capacity as derived in [24]. Reciting the result of [24], the (strong) secrecy-capacity of an AVWTC $(\mathfrak{W}, \mathfrak{V})$ with unconstrained states is given in (59) at the bottom of next page,[9] where the subscript

---

[9]We use macrons in the following notation of random variables not because they are multi-dimensions, but to distinguish them from other random variable that have yet to be introduced.
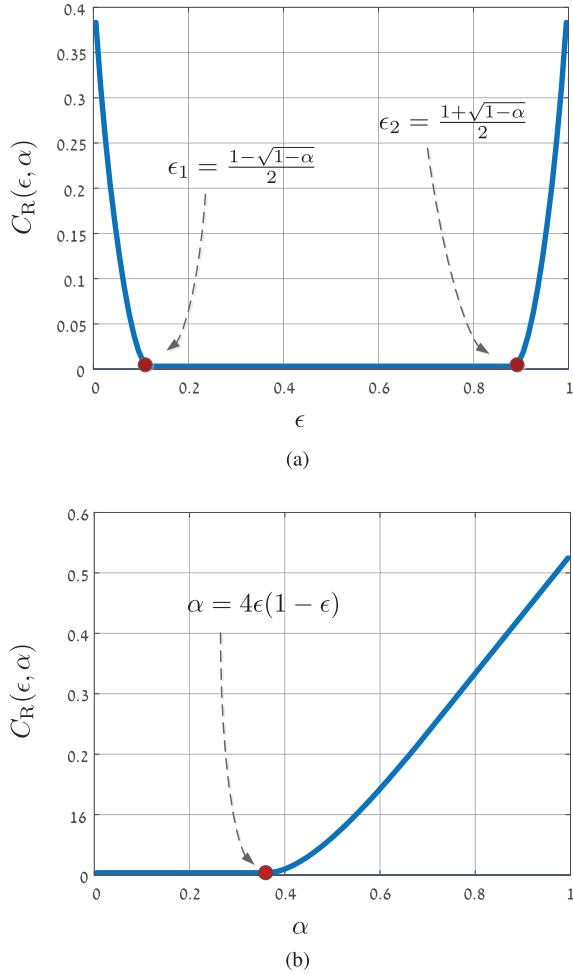
Fig. 3. CR-assisted SS-capacity $C_R(\epsilon, \alpha)$ versus: (a) the proscribed type for the main channel $Q_{S_1}(1) = \epsilon$, which corresponds to the portion of flipped symbols in the BS-BE WTC; (b) the proscribed type for the eavesdropper's channel $Q_{S_2}(1) = \alpha$, which corresponds to the portion of erasures in the BS-BE WTC.

P indicates that the mutual information terms are calculated with respect to a joint distribution that for each $k \in \mathbb{N}$, $Q \in \mathcal{P}(\mathcal{S})$ and $s^k \in \mathcal{S}^k$, is given by (60) at the bottom of this page and the cardinality of $\mathcal{V}_k$ is bounded above as $|\mathcal{V}_k| \leq |\mathcal{X}|^k$.

To adapt $R_S^*(\mathfrak{W}, \mathfrak{V})$ to the type constrained AVWTC $(\mathfrak{W}, \mathfrak{V}, Q_S)$, for $Q_S \in \mathcal{P}_\mathbb{Q}(\mathcal{S})$ (see (52)), we first account for the fact that $\mathcal{T}_{Q_S}^k$ is empty for several values of $k \in \mathbb{N}$. Denote the non-zero entries of Q_S by $Q_S(s) = \frac{a_s}{b_s}$, where

$s \in \text{supp}(Q_S)$, and define $\ell = \text{lcm}\{b_s\}_{s \in \text{supp}(Q_S)}$, where lcm $\mathcal{A}$, for $\mathcal{A} \subset \mathbb{N}$ with $|\mathcal{A}| < \infty$, is the *least common multiple* of the elements in $\mathcal{A}$. Denote $\mathbb{N}_\ell \triangleq \{n \cdot \ell \,|\, n \in \mathbb{N}\}$ and henceforth consider only blocklengths that belong to $\mathbb{N}_\ell$. Having this, the adaptation of $R_S^*(\mathfrak{W}, \mathfrak{V})$ to the AVWTC with a type constraint $Q_S \in \mathcal{P}_\mathbb{Q}(\mathcal{S})$, is given in (61) at the bottom of this page, where each $P_{\bar{V}_{n\ell}, \bar{X}^{n\ell}} \in \mathcal{P}(\mathcal{V}_{n\ell} \times \mathcal{X}^{n\ell})$ induces the joint distribution $P \triangleq P_{\bar{V}_{n\ell}, \bar{X}^{n\ell}, \bar{Y}_{Q_S}^{n\ell}, \bar{Z}_{s^{n\ell}}^{n\ell}}^{(Q_S)}$. As a disclaimer, we remark that the we did not directly prove[10] that $R_S^*(\mathfrak{W}, \mathfrak{V}, Q_S)$ is indeed a multi-letter description of the type constrained AVWTC secrecy-capacity, and we first state it merely as an educated guess. Nonetheless, as the following proposition shows, this is actually the case since $R_S^*(\mathfrak{W}, \mathfrak{V}, Q_S)$ is equal to the CR-assisted SS-capacity formula from Theorem 1.

Proposition 1 (Multi-Letter CR-Assisted SS-Capacity): For any $Q_S \in \mathcal{P}_\mathbb{Q}(\mathcal{S})$ it holds that

$$R_S^*(\mathfrak{W}, \mathfrak{V}, Q_S) = C_R(\mathfrak{W}, \mathfrak{V}, Q_S). \tag{62}$$

Proof: We first show that $R_S^*(\mathfrak{W}, \mathfrak{V}, Q_S) \geq C_R(\mathfrak{W}, \mathfrak{V}, Q_S)$. Let $\bar{Q}_{U,X} \in \mathcal{P}(\mathcal{U} \times \mathcal{X})$ be the extremum achieving distributions in $C_R(\mathfrak{W}, \mathfrak{V}, Q_S)$, and denote $\bar{Q} \triangleq \bar{Q}_{U,X} Q_S Q_{Y|X,S} Q_{Z|X,S}$, where $Q_{Y|X,S}$ and $Q_{Z|X,S}$ are defined in Theorem 1. For each $n \in \mathbb{N}$ and $s^{n\ell} \in \mathcal{T}_{Q_S}^{n\ell}$, denote $k = n\ell$ and define

$$\bar{P}_{\bar{U}^k, \bar{X}^k, \bar{Y}_{Q_S}^k, \bar{Z}_{s^k}^k}(u^k, x^k, y^k, z^k)$$
$$\triangleq \prod_{i=1}^k \bar{Q}_{U,X}(u_i, x_i) W_{Q_S}(y_i|x_i) V_{s_i}(z_i|x_i), \tag{63}$$

where $W_{Q_S} : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$ is given by $W_{Q_S}(y|x) = \sum_{s \in \mathcal{S}} Q_S(s) W_s(y|x)$. We abbreviate $\bar{P}_{\bar{U}^k, \bar{X}^k, \bar{Y}_{Q_S}^k, \bar{Z}_{s^k}^k}$ as $\bar{P}$ and note that it corresponds to taking $\mathcal{V}_k = \mathcal{U}^k$ and setting $P_{\bar{V}_k, \bar{X}^k} = \bar{Q}_{U,X}^k$ in $P_{\bar{V}_k, \bar{X}^k, \bar{Y}_{Q_S}^k, \bar{Z}_{s^k}^k}^{(Q_S)}$. As a last preliminary technical step we define $(U^k, X^k, S^k, Y^k, Z^k)$ as independent copies of $(U, X, S, Y, Z) \sim \bar{Q}$, i.e., $(U^k, X^k, S^k, Y^k, Z^k) \sim \bar{Q}^k$, and note that $\bar{P}_{\bar{U}^k, \bar{X}^k, \bar{Y}_{Q_S}^k} = \bar{Q}_{U,X,Y}^k$ and that $\bar{P}_{\bar{U}^k, \bar{X}^k, \bar{Z}_{s^k}^k} = \bar{Q}_{U,X,Z|S=s^k}^k$, for each $s^k \in \mathcal{S}^k$.

Now, evaluating the first mutual information term from (61) under $\bar{P}$ gives

$$I_{\bar{P}}(\bar{U}^k; \bar{Y}_{Q_S}^k) = I_{\bar{Q}^k}(U^k; Y^k) \overset{(a)}{=} k \cdot I_{\bar{Q}}(U; Y), \tag{64}$$

[10]By adapting the proof steps from [24].

$$R_S^*(\mathfrak{W}, \mathfrak{V}) \triangleq \lim_{k \to \infty} \frac{1}{k} \sup_{P_{\bar{V}_k, \bar{X}^k} \in \mathcal{P}(\mathcal{V}_k \times \mathcal{X}^k)} \left[\min_{Q \in \mathcal{P}(\mathcal{S})} I_{P^{(Q)}}(\bar{V}_k; \bar{Y}_Q^k) - \max_{s^k \in \mathcal{S}^k} I_{P^{(Q)}}(\bar{V}_k; \bar{Z}_{s^k}^k) \right] \tag{59}$$

$$P_{\bar{V}_k, \bar{X}^k, \bar{Y}_Q^k, \bar{Z}_{s^k}^k}^{(Q)}(v_k, x^k, y^k, z^k) = P_{\bar{V}_k, \bar{X}^k}(v_k, x^k) \prod_{i=1}^k \left(\sum_{s \in \mathcal{S}} Q(s) W_s(y_i|x_i) \right) V_{s_i}(z_i|x_i) \tag{60}$$

$$R_S^*(\mathfrak{W}, \mathfrak{V}, Q_S) \triangleq \lim_{n \to \infty} \frac{1}{n\ell} \sup_{P_{\bar{V}_{n\ell}, \bar{X}^{n\ell}} \in \mathcal{P}(\mathcal{V}_{n\ell} \times \mathcal{X}^{n\ell})} \left[I_P(\bar{V}_{n\ell}; \bar{Y}_{Q_S}^{n\ell}) - \max_{s^{n\ell} \in \mathcal{T}_{Q_S}^{n\ell}} I_P(\bar{V}_{n\ell}; \bar{Z}_{s^{n\ell}}^{n\ell}) \right], \tag{61}$$

where (a) is because $\left\{(U_i, Y_i)\right\}_{i=1}^k$ are a sequence of i.i.d. pairs according to the marginal distribution of U and Y with respect to $\bar{Q}$. For the subtracted term from (61) when calculated with respect to $\bar{P}$, we have

$$
\begin{aligned}
\max_{s^k \in \mathcal{T}_{Q_S}^k} I_{\bar{P}}\left(\bar{U}_k; \bar{Z}_{s^k}^k\right) &\overset{(a)}{=} \max_{s^k \in \mathcal{T}_{Q_S}^k} I_{\bar{Q}^k}(U^k; Z^k | S^k = s^k) \\
&\overset{(b)}{=} \max_{s^k \in \mathcal{T}_{Q_S}^k} \sum_{i=1}^k I_{\bar{Q}^k}(U_i; Z_i | S_i = s_i) \\
&\overset{(c)}{=} \max_{s^k \in \mathcal{T}_{Q_S}^k} k \cdot \sum_{s \in \mathcal{S}} Q_S(s) I_{\bar{Q}}(U; Z | S = s) \\
&= k \cdot I_{\bar{Q}}(U; Z | S) \qquad (65)
\end{aligned}
$$

where (a) is because $\bar{P}_{\bar{U}^k, \bar{X}^k, \bar{Z}_{s^k}^k} = \bar{Q}_{U,X,Z|S=s^k}^k$, for every $s^k \in \mathcal{S}^k$, (b) uses the product structure of $\bar{Q}^k$, while (c) follows since $v_{s^k} = Q_S$ for every $s^k \in \mathcal{T}_{Q_S}^k$.

Based on (64) and (65) we get the desired inequality since

$$
\begin{aligned}
&R_S^*(\mathfrak{W}, \mathfrak{V}, Q_S) \\
&\geq \lim_{\substack{k \to \infty: \\ k \in \mathbb{N}_\ell}} \frac{1}{k} \left[I_{\bar{P}}\left(\bar{U}^k; \bar{Y}_{Q_S}^k\right) - \max_{s^k \in \mathcal{T}_{Q_S}^k} I_{\bar{P}}\left(\bar{U}^k; \bar{Z}_{s^k}^k\right) \right] \\
&= I_{\bar{Q}}(U; Y) - I_{\bar{Q}}(U; Z | S) \\
&= C_R(\mathfrak{W}, \mathfrak{V}, Q_S). \qquad (66)
\end{aligned}
$$

The opposite inequality, that is $R_S^*(\mathfrak{W}, \mathfrak{V}, Q_S) \leq C_R(\mathfrak{W}, \mathfrak{V}, Q_S)$, follows by repeating the steps from the proof of Theorem 3 (with the proper update of notation). To avoid verbatim repetition of the same arguments, we give only an outline of the proof while describing the required adjustments. To this end, for each $k \in \mathbb{N}_\ell$ and $P_{\bar{V}_k, \bar{X}^k} \in \mathcal{P}(\mathcal{V}_k \times \mathcal{X}^k)$, it is convenient to define a new distribution $\hat{P}_{V_k, X^k, S^k, Y^k, Z^k}$ given by

$$
\begin{aligned}
&\hat{P}_{V_k, X^k, S^k, Y^k, Z^k}\left(v_k, x^k, s^k, y^k, z^k\right) \\
&\triangleq P_{\bar{V}_k, \bar{X}^k}\left(v_k, x^k\right) Q_S^k\left(s^k\right) W_{s^k}^k\left(y^k | x^k\right) V_{s^k}^k\left(z^k | x^k\right). \quad (67)
\end{aligned}
$$

Noting that $\hat{P}_{V_k, X^k, Y^k, Z^k} = P_{\bar{V}_k, \bar{X}^k, \bar{Y}_{Q_S}^k}$ and $\hat{P}_{V_k, X^k, Z^k | S^k = s^k} = P_{\bar{V}_k, \bar{X}^k, \bar{Z}_{s^k}^k}$, for every $s^k \in \mathcal{S}^k$, one may rewrite (61) as (68) from the bottom of this page.

Having this, to establish $C_R(\mathfrak{W}, \mathfrak{V}, Q_S)$ as an upper bound, fix $k \in \mathbb{N}_\ell$ and $P_{\bar{V}_k, \bar{X}^k} \in \mathcal{P}(\mathcal{V}_k \times \mathcal{X}^k)$ and consider the following. Due to the structure of $\hat{P}$ one may treat V_k as the message M in the proof of Theorem 3 and invoke simple adaptations of Lemmas 7 and 8 to get

$$
\max_{s^k \in \mathcal{T}_{Q_S}^k} I_{\hat{P}}\left(V_k; Z^k | S^k = s^k\right) \geq I_{\hat{P}}(V_k; Z^k | S^k) - k\xi_{k,\alpha}, \quad (69)
$$

where $\xi_{k,\alpha} = \log |\mathcal{Z}| \left(\alpha + 2|\mathcal{S}| e^{-2k \frac{\alpha^2}{|\mathcal{S}|^2}} \right)$ and α is any number in $(0, 1]$. Using (69), for any $k \in \mathbb{N}_\ell$ and $P_{\bar{V}_k, \bar{X}^k} \in$

$\mathcal{P}(\mathcal{V}_k \times \mathcal{X}^k)$, we have

$$
\begin{aligned}
&I_{\hat{P}}\left(V_k; Y^k\right) - \max_{s^k \in \mathcal{T}_{Q_S}^k} I_{\hat{P}}\left(V_k; Z^k | S^k = s^k\right) \\
&\leq I_{\hat{P}}\left(V_k; Y^k\right) - I_{\hat{P}}\left(V_k; Z^k | S^k\right) + k\xi_{k,\alpha}. \qquad (70)
\end{aligned}
$$

Applying standard manipulations on the RHS of (70) (similar to those in the derivation of (134)), further shows that

$$
\begin{aligned}
&I_{\hat{P}}\left(V_k; Y^k\right) - \max_{s^k \in \mathcal{T}_{Q_S}^k} I_{\hat{P}}\left(V_k; Z^k | S^k = s^k\right) \\
&\leq k \cdot \max_{Q_{U,X} \in \mathcal{P}(\mathcal{U} \times \mathcal{X})} \left[I_Q(U; Y) - I_Q(U; Z | S) \right] + k\xi_{k,\alpha}, \quad (71)
\end{aligned}
$$

where the mutual information terms are taken with respect to the joint distribution from Theorem 1. The derivation of (71) also relies on claims similar to those from Lemmas 10 and 11 from Section VI. Dividing both sides by k, taking the supremum of the LHS over all $P_{\bar{V}_k, \bar{X}^k} \in \mathcal{P}(\mathcal{V}_k \times \mathcal{X}^k)$ and replacing $\alpha \in (0, 1]$ with a sequence $\{\alpha_k\}_{k \in \mathbb{N}_\ell}$ that decays sufficiently slowly to zero, the proof is completed by letting $k \to \infty$. ∎

E. Proof of Theorem 1 From Theorems 2 and 3

Achievability: For the direct part, denote the RHS of (50) by $C_R^\star(\mathfrak{W}, \mathfrak{V}, Q_S)$ and assume that $R < C_R^\star(\mathfrak{W}, \mathfrak{V}, Q_S)$. We show that there exists $\delta_0 > 0$, such that for any $\epsilon > 0$ there is a CR (n, M_n, K_n)-code C_n that satisfies (49).

For any $\delta > 0$, define

$$
\mathcal{P}_\delta(Q_S) \triangleq \left\{ P \in \mathcal{P}(\mathcal{S}) \,\middle|\, |P(s) - Q_S(s)| \leq \delta Q_S(s), \ \forall s \in \mathcal{S} \right\}, \quad (72)
$$

and note that $\mathcal{P}_\delta(Q_S)$ is convex and closed. Applying Theorem 2 with $\mathcal{Q} = \mathcal{P}_\delta(Q_S)$ and recalling Definition 7 of $\mathcal{Q}$-constrained achievability, yields that if[11]

$$
R < \max_{Q_{U,X}} \left[\min_{Q_1 \in \mathcal{P}_\delta(Q_S)} I_{Q_1}(U; Y) - \max_{Q_2 \in \mathcal{P}_\delta(Q_S)} I_{Q_2}(U; Z | S) \right], \quad (73)
$$

then there exists a CR (n, M_n, K_n)-code C_n that satisfies (49). Thus, to establish the achievability of R it suffices to show that there exists $\delta_0 > 0$ for which (73) holds. The following lemma, that is proven in Appendix A using continuity, fills that gap.

Lemma 3: The following limiting relation holds:

$$
\begin{aligned}
&\max_{Q_{U,X}} \left[\min_{Q_1 \in \mathcal{P}_\delta(Q_S)} I_{Q_1}(U; Y) - \max_{Q_2 \in \mathcal{P}_\delta(Q_S)} I_{Q_2}(U; Z | S) \right] \\
&\qquad \nearrow C_R^\star(\mathfrak{W}, \mathfrak{V}, Q_S), \qquad (74)
\end{aligned}
$$

as $\delta \searrow 0$.

---

[11] For simplicity of notation, throughout the proof of Theorem 1 we write Q_j instead of $Q_S^{(j)}$, for $j = 1, 2$.

$$
R_S^*(\mathfrak{W}, \mathfrak{V}, Q_S) = \lim_{\substack{k \to \infty: \\ k \in \mathbb{N}_\ell}} \frac{1}{k} \sup_{P_{\bar{V}_k, \bar{X}^k} \in \mathcal{P}(\mathcal{V}_k \times \mathcal{X}^k)} \left[I_{\hat{P}}\left(V_k; Y^k\right) - \max_{s^k \in \mathcal{T}_{Q_S}^k} I_{\hat{P}}\left(V_k; Z^k | S^k = s^k\right) \right]. \qquad (68)
$$

Converse: Assume that R is an achievable CR-assisted SS-rate for the type constrained AVWTC $(\mathfrak{W}^n, \mathfrak{V}^n, Q_S)$. Then, there exists $\delta > 0$, such that for all $\epsilon > 0$ and sufficiently large n, there exists a CR (n, M_n, K_n)-code C_n that satisfies (49). Define

$$\mathcal{R}_\delta(Q_S) \triangleq \mathcal{P}_\delta(Q_S) \cap \mathcal{P}_\mathbb{Q}(\mathcal{S}), \tag{75}$$

(see (52) and (72)) which is the set of all rational PMFs on $\mathcal{S}$ that are element-wise δ-close to Q_S. Recall the definition of $\mathcal{S}_\mathcal{Q}^n$ from (41), where $\mathcal{Q} \subseteq \mathcal{P}(\mathcal{S})$ is any subset of PMFs. Taking $\mathcal{Q} = \mathcal{R}_\delta(Q_S)$ gives $\mathcal{S}_\mathcal{Q}^n = \mathcal{T}_\delta^n(Q_S)$, for any $n \in \mathbb{N}$.

Using Theorem 3 with $\mathcal{Q} = \mathcal{R}_\delta(Q_S)$ gives the following upper bound:

$$R \leq \max_{Q_{V,U,X}} \inf_{\hat{Q}_S \in \mathcal{R}_\delta(Q_S)} \left[I_{\hat{Q}}(U; Y|V) - I_{\hat{Q}}(U; S, Z|V) \right], \tag{76}$$

where $I_{\hat{Q}}$ denotes that the mutual information terms are calculated with respect to the marginals of $Q_{V,U,X}\hat{Q}_S Q_{Y|X,S} Q_{Z|X,S}$ for some $\hat{Q}_S \in \mathcal{R}_\delta(Q_S)$, where $Q_{Y|X,S}(y|x,s) = W_s(y|x)$ and $Q_{Z|X,S}(z|x,s) = V_s(z|x)$, for all $(s, x, y, z) \in \mathcal{S} \times \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$.

We first show that any $\hat{Q}_S \in \mathcal{R}_\delta(Q_S)$ in the joint distribution of (V, U, X, S, Y, Z) can be replaced with the Q_S while causing only a small change in the value of the mutual information terms. Let $Q_{V,U,X}^\star \in \mathcal{P}(\mathcal{V} \times \mathcal{U} \times \mathcal{X})$ be the maximizer of the RHS of (76). With some abuse of notation we denote by $I_{\hat{Q}}$ and I_Q a mutual information term calculated with respect to $Q_{V,U,X}^\star \hat{Q}_S Q_{Y|X,S} Q_{Z|X,S}$ or $Q_{V,U,X}^\star Q_S Q_{Y|X,S} Q_{Z|X,S}$, respectively. By the definition of $\mathcal{R}_\delta(Q_S)$, for any $\hat{Q}_S \in \mathcal{R}_\delta(Q_S)$ we have

$$\left| \hat{Q}_S(s) - Q_S(s) \right| \leq \frac{\delta}{|\mathcal{S}|}, \quad \forall s \in \mathcal{S}. \tag{77}$$

The continuity of the mutual information implies that there exists a function $f(\delta)$, such that $\lim_{\delta \to 0} f(\delta) = 0$ is independent of $Q_{V,U,X}^\star$ and

$$I_{\hat{Q}}(U; Y|V) - I_{\hat{Q}}(U; S, Z|V)$$
$$\leq I_Q(U; Y|V) - I_Q(U; S, Z|V) + f(\delta). \tag{78}$$

Further notice that Definition 10 of the type constrained achievability gives

$$\mathcal{E}\left(\mathfrak{W}^n, \mathcal{Q}_{\delta'}(Q_S), \mathsf{C}_n\right) \leq \epsilon \tag{79a}$$
$$\mathcal{L}_{\text{Sem}}\left(\mathfrak{V}^n, \mathcal{Q}_{\delta'}(Q_S), \mathsf{C}_n\right) \leq \epsilon. \tag{79b}$$

for any $\delta' \in (0, \delta)$ (see Remark 10). Taking $\delta' \to 0$ in (78), while noting that the rational distributions are dense in $\mathcal{P}(\mathcal{S})$ and that $I_{\hat{Q}}(U; Y|V) - I_{\hat{Q}}(U; S, Z|V)$ is continuous in $\hat{Q}_S$ gives

$$R \leq I_Q(U; Y|V) - I_Q(U; S, Z|V). \tag{80}$$

Our last step is to remove the conditioning on V. The structure of the joint distribution $Q_{V,U,X,S,Y,Z} = Q_{V,U,X} Q_S Q_{Y|X,S} Q_{Z|X,S}$ implies that for any $v \in \mathcal{V}$, the conditional distribution of (U, X, S, Y, Z) given $V = v$ factors as $Q_{U,X,S,Y,Z|V=v} = Q_{U,X|V=v} Q_S Q_{Y|X,S} Q_{Z|X,S}$. Denoting by I_{Q_v} a mutual information term taken with respect

to $Q_{U,X|V=v} Q_S Q_{Y|X,S} Q_{Z|X,S}$, we further upper bound R from (80) as

$$R \leq I_Q(U; Y|V) - I_Q(U; S, Z|V)$$
$$= \sum_{v \in \mathcal{V}} Q_V(v) \Big[I_Q(U; Y|V=v) - I_Q(U; S, Z|V=v) \Big]$$
$$\leq \max_{v \in \mathcal{V}} \Big[I_Q(U; Y|V=v) - I_Q(U; S, Z|V=v) \Big]$$
$$= \max_{v \in \mathcal{V}} \Big[I_{Q_v}(U; Y) - I_{Q_v}(U; S, Z) \Big]$$
$$\overset{(a)}{=} \max_{v \in \mathcal{V}} \Big[I_{Q_v}(U; Y) - I_{Q_v}(U; Z|S) \Big]$$
$$\leq \max_{Q_{U,X}} \Big[I_Q(U; Y) - I_Q(U; Z|S) \Big], \tag{81}$$

where (a) is because U and S are independent under Q_v, for every $v \in \mathcal{V}$. This completes the proof.

V. PROOF OF THEOREM 2

The proof first constructs a reliable CR-code over a family of doubly-exponentially many uncorrelated codes. Specifically, the family consists of all realizations of a random wiretap code with i.i.d. codewords. Reliability is then established via a simple adaptation of the standard AVC error probability analysis [15]. Being double-exponential in size, however, the original family of codes is too large to derive SS in the sense of (47b). Therefore, a Chernoff bound is used to show that only a polynomial sub-family of codes is sufficient for reliability, and having that, the double-exponential decay that Lemma 1 provides is leveraged to prove SS over the reduced CR-code.

Without loss of generality, we assume that $\mathcal{Q}$ contains at least one rational distribution; otherwise, there is nothing to prove (see Remark 7). Consequently, we henceforth refer only to blocklengths $n \in \mathbb{N}$ for which $\mathcal{S}_\mathcal{Q}^n \neq \emptyset$. We show that (51) is achievable when $U = X$. Then, using standard channel prefixing arguments, the RHS of (51) is achieved.

Fix $\epsilon > 0$, a PMF $Q_{U,X} \in \mathcal{P}(\mathcal{U} \times \mathcal{X})$ and let Q_X be its marginal on $\mathcal{X}$. For any $Q_S \in \mathcal{Q}$, the joint distribution of (X, S, Y, Z) is $Q_{X,S,Y,Z} = Q_X Q_S Q_{Y|X,S} Q_{Z|X,S}$, where $Q_{Y|X,S}$ and $Q_{Z|X,S}$ are given in Theorem 2. Since Q_X, $Q_{Y|X,S}$ and $Q_{Z|X,S}$ stay fixed throughout the proof, we use I_{Q_S} to denote a mutual information term taken with respect to $Q_X Q_S Q_{Y|X,S} Q_{Z|X,S}$. Let W be a random variable uniformly distributed over $\mathcal{W}_n \triangleq [1 : 2^{n\tilde{R}}]$, where $\tilde{R} \in \mathbb{R}_+$, that is chosen independently of the message M; W stands for the stochastic part of the encoder.

Random Codebook B_n: A random codebook is a collection of independent random vectors $\mathsf{B}_n = \{\mathbf{X}(m, w)\}_{(m,w) \in \mathcal{M}_n \times \mathcal{W}_n}$, each distributed according to Q_X^n. Set $\tilde{K}_n = |\mathcal{X}|^{n|\mathcal{M}_n\|\mathcal{W}_n|}$ and index (with respect to some arbitrary order) all possible realizations of B_n by $\tilde{\Gamma}_n = [1 : \tilde{K}_n]$ to obtain the set of codebooks $\mathfrak{B}_n = \{\mathcal{B}_n^{(\gamma)}\}_{\gamma \in \tilde{\Gamma}_n}$, where $\mathcal{B}_n^{(\gamma)} = \{\mathbf{x}(m, w, \gamma)\}_{(m,w) \in \mathcal{M}_n \times \mathcal{W}_n}$ is the γ-th element of $\mathfrak{B}_n$. Further, we define the measure $\tilde{\mu}_n$ on $\tilde{\Gamma}_n$ as

$$\tilde{\mu}_n(\gamma) = \prod_{(m,w) \in \mathcal{M}_n \times \mathcal{W}_n} Q_X^n(\mathbf{x}(m, w, \gamma)), \quad \forall \gamma \in \tilde{\Gamma}_n. \tag{82}$$

Stochastic Encoder $f_\gamma^{(n)}$: Fix $\gamma \in \tilde{\Gamma}_n$. To send the message $m \in \mathcal{M}_n$ the encoder randomly and uniformly chooses w from $\mathcal{W}_n$ and feeds $\mathbf{x}(m, w, \gamma) \in \mathcal{B}_n^{(\gamma)}$ into the AVWTC. The stochastic encoder $f_\gamma^{(n)} : \mathcal{M}_n \to \mathcal{P}(\mathcal{X}^n)$ is thus defined by

$$f_\gamma^{(n)}(\mathbf{x}|m) = \sum_{w \in \mathcal{W}_n} 2^{-n\tilde{R}} \mathbb{1}_{\left\{\mathbf{x}(m,w,\gamma)=\mathbf{x}\right\}}. \tag{83}$$

Decoder $\phi_\gamma^{(n)}$: Both m and w are decoded by the legitimate user. With some abuse of notation, for every $\gamma \in \tilde{\Gamma}_n$ we consider a decoding rule $\phi_\gamma^{(n)} : \mathcal{Y}^n \to (\mathcal{M}_n \times \mathcal{W}_n) \cup \{e\}$ that is defined in terms of a non-negative-valued function $d : \mathcal{X}^n \times \mathcal{Y}^n \to \mathbb{R}_+$ as

$$\phi_\gamma^{(n)}(\mathbf{y}) = \begin{cases} (m, w), & \substack{\max \\ (m',w') \in \mathcal{M}_n \times \mathcal{W}_n: \\ (m',w') \neq (m,w)} \, d\big(\mathbf{x}(m', w', \gamma), \mathbf{y}\big) \\ & \qquad < d\big(\mathbf{x}(m, w, \gamma), \mathbf{y}\big) \\ e, & \text{no } (m, w) \text{ as above exists} \end{cases} \tag{84}$$

Although the decoding rule is given in terms of an arbitrary function d, we soon limit ourselves to a specific choice with respect to which we establish reliability. We use an arbitrary d for now to state Lemma 4 in its most general form, thus emphasizing the generality of this decoding rule.

For every $\gamma \in \tilde{\Gamma}_n$ denote $c_n^{(\gamma)} \triangleq \left(f_\gamma^{(n)}, \phi_\gamma^{(n)}\right)$ as the associated uncorrelated (n, M_n)-code, and let $\tilde{\mathcal{C}}_n \triangleq \left\{c_n^{(\gamma)}\right\}_{\gamma \in \tilde{\Gamma}_n}$. The CR $(n, M_n, \tilde{K}_n)$-code $\tilde{\mathsf{C}}_n$ is thus defined by the family $\tilde{\mathcal{C}}_n$, the index set $\tilde{\Gamma}_n$ of size $\tilde{K}_n$, and the measure $\tilde{\mu}_n \in \mathcal{P}(\tilde{\Gamma}_n)$ from (82). Note that $\tilde{\mathsf{C}}_n$ is a CR-code over a family of doubly-exponentially many (n, M_n)-codes.

Error Probability Analysis: We first show the $\tilde{\mathsf{C}}_n$ is reliable. Having that, we use a Chernoff bound to reduce our CR-code to be only polynomial (with the blocklength) in size and then establish SS. The reliability of $\tilde{\mathsf{C}}_n$ relies on the following lemma, which is effectively an adaptation of [47, Lemma 12.9]. For completeness, we prove the lemma in Appendix B.

Lemma 4: If $\mathbb{E}_{Q_X^n} d(X, \mathbf{y}) \leq 1$, for all $\mathbf{y} \in \mathcal{Y}^n$, then for every channel $W_n : \mathcal{X}^n \to \mathcal{P}(\mathcal{Y}^n)$, $\eta > 0$ and $(m, w) \in \mathcal{M}_n \times \mathcal{W}_n$, we have

$$\mathcal{E}_{m,w}(W_n, \tilde{\mathsf{C}}_n) \leq \mathbb{P}_{Q_X^n W_n}\left(d(X, Y) < \frac{|\mathcal{M}_n||\mathcal{W}_n|}{\eta}\right) + \eta, \tag{85}$$

where

$$\mathcal{E}_{m,w}(W_n, \tilde{\mathsf{C}}_n) = \sum_{\gamma \in \tilde{\Gamma}_n} \tilde{\mu}_n(\gamma) \sum_{\substack{\mathbf{y} \in \mathcal{Y}^n: \\ \phi_\gamma^{(n)}(\mathbf{y}) \neq (m,w)}} W_n\big(y|x(m, w, \gamma)\big) \tag{86}$$

is the expected error probability in decoding (m, w) over the ensemble $\tilde{\mathcal{C}}_n$.

Assume without loss of generality that all entries of the matrices $W \in \mathfrak{W}$ are bounded below by a positive constant $\upsilon > 0$. Indeed, consider a modified family of channels $\mathfrak{W}_\upsilon$, formally defined by $W_\upsilon(y|x) = (1 - \upsilon|\mathcal{Y}|)W(y|x) + \upsilon$, where $W \in \mathfrak{W}$. Replacing $\mathfrak{W}$ with $\mathfrak{W}_\upsilon$ causes a negligible change in $I(U; Y) - I(U; Z|S)$ if υ is small, for any $Q_S Q_{U,X}$. Further,

any sequence of CR codes that is reliable with respect to $\mathfrak{W}_\upsilon$ is trivially modified at the decoder to give the same maximal error probability under $\mathfrak{W}$ as the original one did for $\mathfrak{W}_\upsilon$. The modified decoder simulates the output of the AVC $\mathfrak{W}$ to look as if it was generated by $\mathfrak{W}_\upsilon$. Specifically, upon observing an output sequence $\mathbf{y} \in \mathcal{Y}$ generated by $\mathfrak{W}$, for each time instance $i \in [1 : n]$, the modified decoder draws a Bernoulli$(\upsilon|\mathcal{Y}|)$ distributed random variable: If the outcome is 0, then the observed y_i is preserved. If, on the other hand, the outcome is 1, the new decoder draws a symbol y uniformly from $\mathcal{Y}$ and replaces the observed y_i with the uniformly chosen symbol. This makes the new y-sequence look like it was generated by the AVC $\mathfrak{W}_\upsilon$, and the original decoder is then used.

Our next steps up until Lemma 5 (included) follow close resemblance to [47, Lemma 12.10]. For any $Q \in \mathcal{P}(\mathcal{S})$ define $W_Q : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$, the averaged DMC under Q, as

$$W_Q(y|x) \triangleq \sum_{s \in \mathcal{S}} Q(s) W_s(y|x), \quad \forall (x, y) \in \mathcal{X} \times \mathcal{Y}. \tag{87}$$

Let $\tilde{Q} \in \mathcal{Q}$ be a minimizer of $\min_{Q \in \mathcal{Q}} I_Q(X; Y)$ and set $d : \mathcal{X}^n \times \mathcal{Y}^n \to \mathbb{R}_+$ as

$$d(\mathbf{x}, \mathbf{y}) = \frac{W_{\tilde{Q}}^n(\mathbf{y}|\mathbf{x})}{\tilde{Q}_Y^n(\mathbf{y})} = \prod_{i=1}^n \frac{W_{\tilde{Q}}(y_i|x_i)}{\tilde{Q}_Y(y_i)}, \tag{88}$$

for all $(\mathbf{x}, \mathbf{y}) \in \mathcal{X}^n \times \mathcal{Y}^n$ with $\tilde{Q}_Y^n(\mathbf{y}) \neq 0$, where

$$\tilde{Q}_Y(y) = \sum_{x \in \mathcal{X}} Q_X(x) W_{\tilde{Q}}(y|x). \tag{89}$$

If $\tilde{Q}_Y^n(\mathbf{y}) = 0$, set $d(\mathbf{x}, \mathbf{y}) = 1$, for all $\mathbf{x} \in \mathcal{X}^n$. Clearly

$$\mathbb{E}_{Q_X^n} d(\mathbf{X}, \mathbf{y}) = 1, \quad \forall \mathbf{y} \in \mathcal{Y}^n. \tag{90}$$

Lemma 4 implies that for every $\eta > 0$, $\mathbf{s} \in \mathcal{S}^n$ and $(m, w) \in \mathcal{M}_n \times \mathcal{W}_n$,

$$\mathcal{E}_{m,w}(W_{\mathbf{s}}^n, \tilde{\mathsf{C}}_n) \leq \mathbb{P}_{Q_X^n W_{\mathbf{s}}^n}\left(\frac{W_{\tilde{Q}}^n(\mathbf{Y}_{\mathbf{s}}|\mathbf{X})}{\tilde{Q}_Y^n(\mathbf{Y}_{\mathbf{s}})} < \frac{2|\mathcal{M}_n||\mathcal{W}_n|}{\eta}\right) + \frac{\eta}{2}, \tag{91}$$

where $(\mathbf{X}, \mathbf{Y}_{\mathbf{s}}) \sim Q_X^n W_{\mathbf{s}}^n$.

Fix $\mathbf{s} \in \mathcal{S}^n$ and define the random variable $L_n(\mathbf{X}, \mathbf{Y}_{\mathbf{s}}) \triangleq \log \frac{W_{\tilde{Q}}^n(\mathbf{Y}_{\mathbf{s}}|\mathbf{X})}{\tilde{Q}_Y^n(\mathbf{Y}_{\mathbf{s}})}$. Observe that

$$\mathbb{E}_{Q_X^n W_{\mathbf{s}}^n} L_n(\mathbf{X}, \mathbf{Y}_{\mathbf{s}})$$

$$= \sum_{(\mathbf{x},\mathbf{y}) \in \mathcal{X}^n \times \mathcal{Y}^n} Q_X^n(\mathbf{x}) W_{\mathbf{s}}^n(\mathbf{y}|\mathbf{x}) \log\left(\prod_{i=1}^n \frac{W_{\tilde{Q}}(y_i|x_i)}{\tilde{Q}_Y(y_i)}\right)$$

$$= \sum_{i=1}^n \sum_{(x,y) \in \mathcal{X} \times \mathcal{Y}} Q_X(x) W_{s_i}(y|x) \log\left(\frac{W_{\tilde{Q}}(y|x)}{\tilde{Q}_Y(y)}\right). \tag{92}$$

For any $\mathbf{s} \in \mathcal{S}^n$ and $(x, y) \in \mathcal{X} \times \mathcal{Y}$, denote

$$\sum_{i=1}^n W_{s_i}(y|x) = n \sum_{s \in \mathcal{S}} \nu_{\mathbf{s}}(s) W_s(y|x) \triangleq n W_{\nu_{\mathbf{s}}}(y|x), \tag{93}$$

where $\nu_{\mathbf{s}}$ is the empirical PMF of $\mathbf{s}$, as defined in (2). Consequently, we have

$$\mathbb{E}_{Q_X^n W_{\mathbf{s}}^n} L_n(\mathbf{X}, \mathbf{Y}_{\mathbf{s}})$$
$$= n \cdot \sum_{(x,y) \in \mathcal{X} \times \mathcal{Y}} Q_X(x) W_{\nu_{\mathbf{s}}}(y|x) \log\left(\frac{W_{\tilde{Q}}(y|x)}{\tilde{Q}_Y(y)}\right). \quad (94)$$

We next show that if $\mathbf{s} \in \mathcal{S}_{\mathcal{Q}}^n$, then the RHS of (94) is lower bounded by $I_{\tilde{Q}}(X;Y)$.

Lemma 5: For any $Q \in \mathcal{Q}$, the following relation holds

$$\sum_{(x,y) \in \mathcal{X} \times \mathcal{Y}} Q_X(x) W_Q(y|x) \log\left(\frac{W_{\tilde{Q}}(y|x)}{\tilde{Q}_Y(y)}\right) \geq I_{\tilde{Q}}(X;Y). \quad (95)$$

Lemma 5 is proven in Appendix C. Now, if $\mathbf{s} \in \mathcal{S}_{\mathcal{Q}}^n$, then $\nu_{\mathbf{s}} \in \mathcal{Q}$. By Lemma 5 and (94) this gives

$$\mathbb{E}_{Q_X^n W_{\mathbf{s}}^n} L_n(\mathbf{X}, \mathbf{Y}_{\mathbf{s}}) \geq n I_{\tilde{Q}}(X;Y), \quad \forall \mathbf{s} \in \mathcal{S}_{\mathcal{Q}}^n. \quad (96)$$

Furthermore, since $W_s(y|x) > \upsilon > 0$ for every $(s, x, y) \in \mathcal{S} \times \mathcal{X} \times \mathcal{Y}$, it holds that

$$\left| \log \frac{W_{\tilde{Q}}(y|x)}{\tilde{Q}_Y(y)} \right| \leq \left| \log \frac{1}{\upsilon} \right| = -\log \upsilon, \quad (97)$$

which implies that

$$\left| \log \frac{W_{\tilde{Q}}(Y|X)}{\tilde{Q}_Y(Y)} \right| \leq -\log \upsilon \quad (98)$$

is true with probability 1, for any $(X, Y) \sim P_{X,Y} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$. This yields

$$\text{var}\left(\log \frac{W_{\tilde{Q}}(Y|X)}{\tilde{Q}_Y(Y)} \right) \leq \log^2 \upsilon. \quad (99)$$

Having (91), (96) and (99), the proof of reliability is concluded as follows. Recall that $\mathcal{M}_n = [1 : M_n]$ and set

$$M_n = \left\lfloor 2^{\left(\min_{Q \in \mathcal{Q}} I_Q(X;Y) - \tilde{R} - \frac{\delta}{2}\right)} \right\rfloor, \quad (100)$$

for some $\delta > 0$ to be specified later. Using (91) with $\eta_n = \frac{1}{n}$ in the role of η, we have that for n sufficiently large and all $(m, w) \in \mathcal{M}_n \times \mathcal{W}_n$ and $\mathbf{s} \in \mathcal{S}_{\mathcal{Q}}^n$, we have

$$\mathcal{E}_{m,w}(W_{\mathbf{s}}^n, \tilde{\mathsf{C}}_n)$$
$$\leq \mathbb{P}_{Q_X^n W_{\mathbf{s}}^n}\left(\frac{W_{\tilde{Q}}^n(\mathbf{Y}_{\mathbf{s}}|\mathbf{X})}{\tilde{Q}_Y^n(\mathbf{Y}_{\mathbf{s}})} < \frac{2|\mathcal{M}_n\|\mathcal{W}_n|}{\eta_n}\right) + \frac{\eta_n}{2}$$
$$= \mathbb{P}_{Q_X^n W_{\mathbf{s}}^n}\left(L_n(\mathbf{X}, \mathbf{Y}_{\mathbf{s}}) < \log|\mathcal{M}_n\|\mathcal{W}_n| - \log\frac{\eta_n}{2}\right) + \frac{\eta_n}{2}$$
$$\overset{(a)}{\leq} \mathbb{P}_{Q_X^n W_{\mathbf{s}}^n}\left(L_n(\mathbf{X}, \mathbf{Y}_{\mathbf{s}}) < n I_{\tilde{Q}}(X;Y) - \frac{n\delta}{2} - \log\frac{\eta_n}{2}\right) + \frac{\eta_n}{2}$$
$$\overset{(b)}{\leq} \mathbb{P}_{Q_X^n W_{\mathbf{s}}^n}\left(\left|\mathbb{E}L_n(\mathbf{X}, \mathbf{Y}_{\mathbf{s}}) - L_n(\mathbf{X}, \mathbf{Y}_{\mathbf{s}})\right| > \frac{n\delta}{2}\right) + \frac{\eta_n}{2}$$
$$\overset{(c)}{\leq} \frac{4\text{var}(L_n(\mathbf{X}, \mathbf{Y}_{\mathbf{s}}))}{n^2 \delta^2} + \frac{\eta_n}{2}$$
$$\overset{(d)}{\leq} \frac{4\log^2 \upsilon}{n\delta^2} + \frac{\eta_n}{2}$$
$$\overset{(e)}{=} \frac{c_{\delta,\upsilon}}{n} \quad (101)$$

where (a) and (b) use (100) and (96), respectively, (c) is Chebyshev's inequality, (d) follows by the pairwise independence of $(\mathbf{X}, \mathbf{Y}_{\mathbf{s}})$ across time and (99), while (e) is by setting $c_{\delta,\upsilon} = \frac{1}{2} + \frac{4\log^2 \upsilon}{\delta^2}$. Concluding, (101) yields

$$\max_{\substack{\mathbf{s} \in \mathcal{S}_{\mathcal{Q}}^n, \\ (m,w) \in \mathcal{M}_n \times \mathcal{W}_n}} \mathcal{E}_{m,w}(W_{\mathbf{s}}^n, \tilde{\mathsf{C}}_n) \leq \frac{c_{\delta,\upsilon}}{n}, \quad (102)$$

which implies the reliability of CR-code $\tilde{\mathsf{C}}_n$.

CR Reduction for Reliability: Our next step is to reduce the CR-code $\tilde{\mathsf{C}}_n$ over the family $\tilde{\mathcal{C}}_n$ of size $\tilde{K}_n = |\mathcal{X}|^{n|\mathcal{M}_n\|\mathcal{W}_n|}$, to be over a family of codes that is no more than polynomial in size (see [12] for Ahlswede's original CR elimination argument for the classic AVC). This reduction is crucial for the subsequent security analysis. To do so, let $\{G_k\}_{k=1}^{K_n}$ be a collection of $K_n \in \mathbb{N}$ i.i.d. random variables with values in $\tilde{\Gamma}_n$ and a common distribution $\tilde{\mu}_n$. Each realization $\gamma_k \in \tilde{\Gamma}_n$ of G_k, $k \in [1 : K_n]$, corresponds to a codebook $\mathcal{B}_n^{(\gamma)} \in \mathfrak{B}_n$ which, in turn, induces a code $c_n^{(\gamma)}$.

We show that averaging the error probabilities associated with each random code $C_n(k) \triangleq c_n^{(G_k)}$ results in a vanishing term, with arbitrarily high probability. In a later stage, we extract a realization of $\{C_n(k)\}_{k=1}^{K_n}$ that is both reliable and semantically-secure, and define our CR-code to be uniformly distributed over the codes in the realization.

Thus, for each $\mathbf{s} \in \mathcal{S}^n$ and $(m, w) \in \mathcal{M}_n \times \mathcal{W}_n$, consider the random variable

$$\frac{1}{K_n} \sum_{k=1}^{K_n} e_{m,w}(W_{\mathbf{s}}^n, C_n(k)), \quad (103)$$

where a possible value of each $e_{m,w}(W_{\mathbf{s}}^n, C_n(k))$, $k \in [1 : K_n]$ and $\gamma_k \in \tilde{\Gamma}_n$, is

$$e_{m,w}(W_{\mathbf{s}}^n, c_n(k)) \triangleq e_{m,w}(W_{\mathbf{s}}^n, c_n^{(\gamma_k)})$$
$$= \sum_{\substack{\mathbf{y} \in \mathcal{Y}^n: \\ \phi_{\gamma_k}^{(n)}(\mathbf{y}) \neq (m,w)}} W_{\mathbf{s}}^n(\mathbf{y}|\mathbf{x}(m, w, \gamma_k)). \quad (104)$$

This is an average of K_n independent random variables, each bounded between 0 and 1. Furthermore, the expected value of each equals to $\mathcal{E}_{m,w}(W_{\mathbf{s}}^n, \tilde{\mathsf{C}}_n)$, and therefore, (102) implies that

$$\mathbb{E}_{\tilde{\mu}_n} e_{m,w}(W_{\mathbf{s}}^n, C_n(k)) \leq \frac{c_{\delta,\upsilon}}{n}, \quad (105)$$

for each $\mathbf{s} \in \mathcal{S}_{\mathcal{Q}}^n$, $(m, w) \in \mathcal{M}_n \times \mathcal{W}_n$ and $k \in [1 : K_n]$. The probability of (103) not decaying to zero under any $\mathbf{s} \in \mathcal{S}_{\mathcal{Q}}^n$ and $(m, w) \in \mathcal{M}_n \times \mathcal{W}_n$ is upper bounded using the version of the Chernoff bound from [48, Lemma 4].

Lemma 6 (Chernoff Bound [48]): Let $\{X_\ell\}_{\ell=1}^L$ be a collection of i.i.d. random variables with common distribution P, such that $\text{supp}(P) \subseteq [0, B]$ and $\mathbb{E}X_\ell \leq \mu \neq 0$, for all $\ell \in [1 : L]$. Then for any c with $\frac{c}{\mu} \in [1, 2]$,

$$\mathbb{P}_{P^L}\left(\frac{1}{L} \sum_{\ell=1}^L X_\ell \geq c\right) \leq e^{-\frac{L\mu}{3B}\left(\frac{c}{\mu} - 1\right)^2}. \quad (106)$$

Setting $K_n = n^3$ and using (106) with $L = K_n$, $\mu = \frac{c_{\delta,v}}{n}$, $B = 1$, and $\frac{c}{\mu} = 2$, assures that $\frac{1}{K_n}\sum_{k=1}^{K_n} e_{m,w}(W_{\mathbf{s}}^n, C_n(k))$ is arbitrarily small with probability super-exponentially close to 1. That is, for each $\mathbf{s} \in \mathcal{S}_Q^n$ and $(m, w) \in \mathcal{M}_n \times \mathcal{W}_n$, we have

$$\mathbb{P}_{\tilde{\mu}_n^n}\left(\frac{1}{K_n}\sum_{k=1}^{K_n} e_{m,w}(W_{\mathbf{s}}^n, C_n(k)) \geq \frac{2c_{\delta,v}}{n}\right) \leq e^{-\frac{1}{3}\frac{c_{\delta,v}K_n}{n}}$$
$$= e^{-\frac{c_{\delta,v}n^2}{3}}. \tag{107}$$

By (107) and the union bound, we have

$$\mathbb{P}_{\tilde{\mu}_n^n}\left(\left\{\max_{\substack{\mathbf{s}\in\mathcal{S}_Q^n,\\(m,w)\\\in\mathcal{M}_n\times\mathcal{W}_n}} \frac{1}{K_n}\sum_{k=1}^{K_n} e_{m,w}(W_{\mathbf{s}}^n, C_n(k)) < \frac{2c_{\delta,v}}{n}\right\}^c\right)$$

$$= \mathbb{P}_{\tilde{\mu}_n^n}\left(\max_{\substack{\mathbf{s}\in\mathcal{S}_Q^n,\\(m,w)\\\in\mathcal{M}_n\times\mathcal{W}_n}} \frac{1}{K_n}\sum_{k=1}^{K_n} e_{m,w}(W_{\mathbf{s}}^n, C_n(k)) \geq \frac{2c_{\delta,v}}{n}\right)$$

$$\leq \sum_{\mathbf{s}\in\mathcal{S}_Q^n}\sum_{\substack{(m,w)\\\in\mathcal{M}_n\times\mathcal{W}_n}} \mathbb{P}_{\tilde{\mu}_n^n}\left(\frac{1}{K_n}\sum_{k=1}^{K_n} e_{m,w}(W_{\mathbf{s}}^n, C_n(k)) \geq \frac{2c_{\delta,v}}{n}\right)$$

$$\leq \left|\mathcal{S}_Q^n\right| \cdot |\mathcal{M}_n| \cdot |\mathcal{W}_n| \cdot e^{-\frac{c_{\delta,v}n^2}{3}}$$

$$\overset{(a)}{\leq} |S|^n \cdot 2^{n\left(\min_{Q\in\mathcal{Q}} I_Q(X;Y) - \frac{\delta}{2}\right)} \cdot e^{-\frac{c_{\delta,v}n^2}{3}}$$

$$\triangleq \kappa_n^{(1)}, \tag{108}$$

where (a) uses (100) and $\left|\mathcal{S}_Q^n\right| \leq |S|^n$. Note that $\kappa_n^{(1)} \to 0$ as $n \to \infty$.

Security Analysis: We show that the probability of $\{C_n(k)\}_{k=1}^{K_n}$ violating the SS requirement is arbitrarily small. First, for any $\gamma \in \tilde{\Gamma}_n$, $P_M \in \mathcal{P}(\mathcal{M}_n)$ and $\mathbf{s} \in \mathcal{S}^n$, let $P_{M,W,\mathbf{X},\mathbf{Z}_{\mathbf{s}}}^{(\gamma,\mathbf{s})}$ be the induced joint distribution over $\mathcal{M}_n \times \mathcal{Z}^n$, which is given by (see (83))

$$P_{M,\mathbf{Z}_{\mathbf{s}}}^{(\gamma,\mathbf{s})}(m, \mathbf{z}) = P_M(m)\frac{1}{|\mathcal{W}_n|}\sum_{w\in\mathcal{W}_n} V_{\mathbf{s}}^n(\mathbf{z}|\mathbf{x}(m, w, \gamma)). \tag{109a}$$

Accounting also for the random codebook construction, we define $G_n \sim \tilde{\mu}_n$ as a random variable taking values in $\tilde{\Gamma}_n$ and set

$$P_{G_n,M,\mathbf{Z}_{\mathbf{s}}}^{(\mathbf{s})}(\gamma, m, \mathbf{z}) = \tilde{\mu}_n(\gamma)P_{M,\mathbf{Z}_{\mathbf{s}}}^{(\gamma,\mathbf{s})}(m, \mathbf{z}). \tag{109b}$$

For any $\mathbf{s} \in \mathcal{S}^n$ and $\gamma \in \Gamma_n$, have

$$\max_{\substack{\mathbf{s}\in\mathcal{S}_Q^n,\\P_M\in\mathcal{P}(\mathcal{M}_n)}} D\left(P_{\mathbf{Z}_{\mathbf{s}}|M,G_n=\gamma}^{(\mathbf{s})}\,\middle\|\,P_{\mathbf{Z}_{\mathbf{s}}|G_n=\gamma}^{(\mathbf{s})}\,\middle|\,P_M\right)$$

$$\overset{(a)}{\leq} \max_{\substack{\mathbf{s}\in\mathcal{S}_Q^n,\\P_M\in\mathcal{P}(\mathcal{M}_n)}} D\left(P_{\mathbf{Z}_{\mathbf{s}}|M,G_n=\gamma}^{(\mathbf{s})}\,\middle\|\,Q_{Z|S=\mathbf{s}}^n\,\middle|\,P_M\right)$$

$$= \max_{\substack{\mathbf{s}\in\mathcal{S}_Q^n,\\P_M\in\mathcal{P}(\mathcal{M}_n)}} \sum_{m\in\mathcal{M}_n} P_M(m)D\left(P_{\mathbf{Z}_{\mathbf{s}}|M=m,G_n=\gamma}^{(\mathbf{s})}\,\middle\|\,Q_{Z|S=\mathbf{s}}^n\right)$$

$$\leq \max_{\substack{\mathbf{s}\in\mathcal{S}_Q^n,\\m\in\mathcal{M}_n}} D\left(P_{\mathbf{Z}_{\mathbf{s}}|M=m,G_n=\gamma}^{(\mathbf{s})}\,\middle\|\,Q_{Z|S=\mathbf{s}}^n\right), \tag{110}$$

where (a) is because for any $\mathbf{s} \in \mathcal{S}_Q^n$ and $P_M \in \mathcal{P}(\mathcal{M}_n)$

$$D\left(P_{\mathbf{Z}_{\mathbf{s}}|M,G_n=\gamma}^{(\mathbf{s})}\,\middle\|\,P_{\mathbf{Z}_{\mathbf{s}}|G_n=\gamma}^{(\mathbf{s})}\,\middle|\,P_M\right)$$
$$= D\left(P_{\mathbf{Z}_{\mathbf{s}}|M,G_n=\gamma}^{(\mathbf{s})}\,\middle\|\,Q_{Z|S=\mathbf{s}}^n\,\middle|\,P_M\right) - D\left(P_{\mathbf{Z}_{\mathbf{s}}|G_n=\gamma}^{(\mathbf{s})}\,\middle\|\,Q_{Z|S=\mathbf{s}}^n\right)$$
$$\leq D\left(P_{\mathbf{Z}_{\mathbf{s}}|M,G_n=\gamma}^{(\mathbf{s})}\,\middle\|\,Q_{Z|S=\mathbf{s}}^n\,\middle|\,P_M\right). \tag{111}$$

Furthermore, by Lemma 1, for any $m \in \mathcal{M}_n$ and $\mathbf{s} \in \mathcal{S}^n$ with empirical PMF $\nu_{\mathbf{s}}$, taking $\tilde{R} > I_{\nu_{\mathbf{s}}Q}(X;Z|S) + \zeta$ for any $\zeta > 0$, implies that there exist $\gamma_1, \gamma_2 > 0$ (uniform in $\mathbf{s}$), such that for n large enough

$$\mathbb{P}_{\tilde{\mu}_n}\left(D\left(P_{\mathbf{Z}_{\mathbf{s}}|M=m,G_n}^{(\mathbf{s})}\,\middle\|\,Q_{Z|S=\mathbf{s}}^n\right) > e^{-n\gamma_1}\right) \leq e^{-e^{n\gamma_2}}. \tag{112}$$

As all $\mathbf{s} \in \mathcal{S}_Q^n$ have $\nu_{\mathbf{s}} \in \mathcal{Q}$, setting

$$\tilde{R} = \max_{Q\in\mathcal{Q}} I_Q(X;Z|S) + \frac{\delta}{2} \tag{113}$$

gives (112) for every $\mathbf{s} \in \mathcal{S}_Q^n$.

Now, with respect to the collection of the i.i.d. random variables $\{G_n(k)\}_{k=1}^{K_n}$ from before (which are, in fact, i.i.d. copies of G_n), with $K_n = n^3$, we have that for n sufficiently large[12]

$$\mathbb{P}_{\tilde{\mu}_n^n}\left(\max_{\substack{k\in[1:K_n],\\\mathbf{s}\in\mathcal{S}_Q^n,\\P_M\in\mathcal{P}(\mathcal{M}_n)}} D\left(P_{\mathbf{Z}_{\mathbf{s}}|M,G_n(k)}^{(\mathbf{s})}\,\middle\|\,P_{\mathbf{Z}_{\mathbf{s}}|G_n(k)}^{(\mathbf{s})}\,\middle|\,P_M\right) > e^{-n\gamma_1}\right)$$

$$\overset{(a)}{\leq} \mathbb{P}_{\tilde{\mu}_n^n}\left(\max_{\substack{k\in[1:K_n],\\\mathbf{s}\in\mathcal{S}_Q^n,\\m\in\mathcal{M}_n}} D\left(P_{\mathbf{Z}_{\mathbf{s}}|M=m,G_n(k)}^{(\mathbf{s})}\,\middle\|\,Q_{Z|S=\mathbf{s}}^n\right) > e^{-n\gamma_1}\right)$$

$$\overset{(b)}{\leq} \sum_{k=1}^{K_n}\sum_{\mathbf{s}\in\mathcal{S}_Q^n}\sum_{m\in\mathcal{M}_n} \mathbb{P}_{\tilde{\mu}_n}\left(D\left(P_{\mathbf{Z}_{\mathbf{s}}|M=m,G_n}^{(\mathbf{s})}\,\middle\|\,Q_{Z|S=\mathbf{s}}^n\right) > e^{-n\gamma_1}\right)$$

$$\overset{(c)}{\leq} n^3 \cdot |S|^n \cdot 2^{nR} \cdot e^{-e^{n\gamma_2}}$$

$$\triangleq \kappa_n^{(2)}, \tag{114}$$

---

[12]In the following chain of inequalities, we consider conditional marginal distributions of a joint distribution $P_{G_n(k),M,W,\mathbf{X},\mathbf{Z}_{\mathbf{s}}}^{(\mathbf{s})}$, where $k \in [1:K_n]$, each defined exactly like in (109), but with $G_n(k)$ in the role of G_n.

where (a) uses (110), (b) follows by the union bound and because $\{G_n(k)\}_{k=1}^{K_n}$ being i.i.d. copies of $G_n \sim \tilde{\mu}_n$, which implies

$$\mathbb{P}_{\tilde{\mu}_n^n}\left(D\left(P_{\mathbf{Z}_\mathbf{s}|M=m,G_n(k)}^{(\mathbf{s})}\middle\|Q_{Z|S=\mathbf{s}}^n\right) > e^{-n\gamma_1}\right)$$
$$= \mathbb{P}_{\tilde{\mu}_n}\left(D\left(P_{\mathbf{Z}_\mathbf{s}|M=m,G_n}^{(\mathbf{s})}\middle\|Q_{Z|S=\mathbf{s}}^n\right) > e^{-n\gamma_1}\right), \quad \forall k \in [1:K_n],$$

while (c) is by (112)-(113). The double-exponential decay of probability that Lemma 1 provides yields $\kappa_n^{(2)} \to 0$ as $n \to \infty$.

Realization Extraction and CR-code Construction: As long as the rate constraints in (100) and (113) hold, Equations (108) and (114) along with the Selection Lemma from [48, Lemma 5], imply the existence of a realization of $\{G_n(k)\}_{k=1}^{K_n}$, denoted by $\{\gamma_k\}_{k=1}^{K_n}$, that for any n sufficiently large satisfies

$$\max_{\substack{\mathbf{s}\in\mathcal{S}_\mathcal{Q}^n, \\ (m,w)\in\mathcal{M}_n\times\mathcal{W}_n}} \frac{1}{K_n}\sum_{k=1}^{K_n} e_{m,w}\left(W_\mathbf{s}^n, c_n(k)\right) \leq \frac{2c_{\delta,\nu}}{n} \quad \text{(115a)}$$

$$\max_{\substack{k\in[1:K_n], \\ \mathbf{s}\in\mathcal{S}_\mathcal{Q}^n, \\ P_M\in\mathcal{P}(\mathcal{M}_n)}} D\left(P_{\mathbf{Z}_\mathbf{s}|M}^{(\gamma_k,\mathbf{s})}\middle\|P_{\mathbf{Z}_\mathbf{s}}^{(\gamma_k,\mathbf{s})}\middle|P_M\right) \leq e^{-n\gamma_1}. \quad \text{(115b)}$$

where, as defined in the CR-reduction part of the proof, $c_n(k) \triangleq c_n^{(\gamma_k)}$.

Set $\Gamma_n = [1:K_n]$, $\mathcal{C}_n \triangleq \{c_n(k)\}_{k\in\Gamma_n}$ and $\mu_n(k) = K_n^{-1}$. Associating a CR (n, M_n, K_n)-code C_n with Γ_n, $\mathcal{C}_n$ and μ_n, (115b) is clearly equivalent to

$$\mathcal{L}_{\text{Sem}}(\mathfrak{V}^n, \mathcal{Q}, \mathsf{C}_n) \leq e^{-n\gamma_1}. \quad \text{(116)}$$

Next, since for every $\mathbf{s} \in \mathcal{S}_\mathcal{Q}^n$ and $(m, w) \in \mathcal{M}_n \times \mathcal{W}_n$, we have

$$\mathcal{E}_{m,w}(W_\mathbf{s}^n, \mathsf{C}_n) \geq \sum_{k\in\Gamma_n}\mu_n(k)\sum_{\mathbf{x}\in\mathcal{X}^n} f_{\gamma_k}(\mathbf{x}|m)\sum_{\substack{\mathbf{y}\in\mathcal{Y}^n: \\ \phi_{\gamma_k}^{(n)}(\mathbf{y})\neq m}} W_\mathbf{s}^n(\mathbf{y}|\mathbf{x})$$
$$= \mathcal{E}_m(W_\mathbf{s}^n, \mathsf{C}_n), \quad \text{(117)}$$

(115a) implies

$$\mathcal{E}(\mathfrak{W}^n, \mathcal{Q}, \mathsf{C}_n) = \max_{\substack{\mathbf{s}\in\mathcal{S}_\mathcal{Q}^n, \\ m\in\mathcal{M}_n}} \mathcal{E}_m(W_\mathbf{s}^n, \mathsf{C}_n) \leq \frac{2c_{\delta,\nu}}{n}. \quad \text{(118)}$$

The proof is concluded by combining (100) with (113) to eliminate $\tilde{R}$, which leaves us with

$$\frac{1}{n}\log M_n \leq \min_{Q_1\in\mathcal{Q}} I_{Q_1}(X;Y) - \max_{Q_2\in\mathcal{Q}} I_{Q_2}(X;Z|S) - \delta,$$
$$\text{(119)}$$

(116) and (118). As $\delta > 0$ was arbitrary, this implies the existence of a sufficiently large n for which (48) is satisfied.

Remark 21 (Relation to Uncorrelated SS-Capacity): Recall that $K_n = n^3$, i.e., our reduced CR-code C_n is only polynomial in size. This has implication to the uncorrelated scenario because if the uncorrelated SS-capacity is strictly positive, one may replace the shared randomness between the legitimate parties with local randomness at the transmitter (which is always available in WTC scenarios). In a CR-code, the shared

randomness is used for selecting which code $c_n(\gamma)$, where $\gamma \in \Gamma_n$, from $\mathcal{C}_n$ will be employed thorough the transmission. Instead, the transmitter may select $\gamma \in \Gamma_n$ and communicate it to the receiver as a prefix. Since $K_n = n^3$, the positivity of the uncorrelated capacity ensures the reliable transmission of γ with a vanishing rate. A condition that differentiates between $\mathcal{Q}$-constrained AVWTCs with zero and non-zero uncorrelated capacities and a dichotomy result (stating that the uncorrelated capacity is either zero or equal to the CR-assisted capacity) are thus the missing pieces in telling whether the RHS of (51) lower bounds the uncorrelated SS-capacity of a given AVWTC. Such a dichotomy result [21] based on a certain threshold property (namely, the symmetrizability of the main AVC) [22] is known for the scenario with unconstrained states. Therefore, Theorem 2 holds for uncorrelated codes when $\mathcal{Q} = \mathcal{P}(\mathcal{S})$ and the considered AVWTC satisfies the condition from [22] for having a positive uncorrelated SS-capacity.

VI. Proof for Theorem 3

Fix $\emptyset \neq \mathcal{Q} \subset \mathcal{P}(\mathcal{S})$ (if $\mathcal{Q} = \emptyset$ there is nothing to prove) and assume without loss of generality that $\mathcal{Q} \subseteq \mathcal{P}_\mathbb{Q}(\mathcal{S})$, i.e., that it contains only rational PMFs. Otherwise, the CR-assisted SS-capacity being a monotone non-increasing function of the constraint set (see Remark 7), implies

$$C_\text{R}(\mathfrak{W}, \mathfrak{V}, \mathcal{Q}) \leq C_\text{R}\left(\mathfrak{W}, \mathfrak{V}, \mathcal{Q}\cap\mathcal{P}_\mathbb{Q}(\mathcal{S})\right). \quad \text{(120)}$$

Let $R \in \mathbb{R}_+$ be an achievable CR-assisted SS-rate for the $\mathcal{Q}$-constrained AVWTC $(\mathfrak{W}^n, \mathfrak{V}^n, \mathcal{Q})$. Then, for all $\epsilon > 0$ and sufficiently large n, there exists a CR (n, M_n, K_n)-code C_n that satisfies (48). To get the max-inf upper bound of Theorem 3, we derive an upper bound on $C_\text{R}(\mathfrak{W}, \mathfrak{V}, \mathcal{Q})$ that is uniform in $Q_S \in \mathcal{Q}$.

Fix $\epsilon > 0$ and let C_n be the corresponding CR (n, M_n, K_n)-code that satisfies (48) for some sufficiently large n. Further let $\mathcal{C}_n = \{c_n(\gamma)\}_{\gamma\in\Gamma_n}$, where $|\Gamma_n| = K_n$, and $\mu_n \in \mathcal{P}(\Gamma_n)$ be the associated family of (n, M_n)-codes and the PMF over this family, respectively. For any $Q_S \in \mathcal{Q}$ the reliability and the security constrains for achievability stated in (48b)-(48c) continue to hold when restricting the state sequences to $\mathcal{T}_{Q_S}^n$ (instead of allowing any $\mathbf{s} \in \mathcal{S}^n$ with empirical PMF $\nu_\mathbf{s} \in \mathcal{Q}$). Thus, for any $Q_S \in \mathcal{Q}$ and sufficiently large n we have

$$\max_{\substack{\mathbf{s}\in\mathcal{T}_{Q_S}^n, \\ m\in\mathcal{M}_n}} \mathcal{E}_m(W_\mathbf{s}^n, \mathsf{C}_n) \leq \mathcal{E}(\mathfrak{W}^n, \mathcal{Q}, \mathsf{C}_n) \leq \epsilon \quad \text{(121a)}$$

$$\max_{\substack{\gamma\in\Gamma_n, \\ \mathbf{s}\in\mathcal{T}_{Q_S}^n, \\ P_M\in\mathcal{P}(\mathcal{M}_n)}} \ell\left(V_\mathbf{s}^n, P_M, c_n(\gamma)\right) \leq \mathcal{L}_\text{Sem}(\mathfrak{V}^n, \mathcal{Q}, \mathsf{C}_n) \leq \epsilon. \quad \text{(121b)}$$

Although the value of n beyond which (121) becomes valid may depend on $\mathcal{Q}$, it is independent of any certain $Q_S \in \mathcal{Q}$.

Fix $Q_S \in \mathcal{Q}$ and recall that if $n \in \mathbb{N}$ is a blocklength for which $\mathcal{T}_{Q_S}^n = \emptyset$, then (48b)-(48c) are trivially satisfied. We avoid these trivial blocklengths by henceforth only considering values of n that belong to $\mathbb{N}_\ell$, as defined in Section IV-D. To remind the reader, $\mathbb{N}_\ell \triangleq \{n \cdot \ell | n \in \mathbb{N}\}$, where ℓ is the least common denominator of all the non-zero entries of Q_S.

Since for any $Q_S \in \mathcal{Q}$, (121a) ensures that $\mathcal{E}_m\big(W_{\mathbf{s}}^n, \mathsf{C}_n\big) \leq \epsilon$, for all $\mathbf{s} \in \mathcal{T}_{Q_S}^n$ and $m \in \mathcal{M}_n$, we have

$$\bar{\mathcal{E}}(\mathfrak{W}^n, Q_S, \mathsf{C}_n) \triangleq \max_{\mathbf{s} \in \mathcal{T}_{Q_S}^n} \sum_{\gamma \in \Gamma_n} \mu_n(\gamma) \frac{1}{M_n} \sum_{m \in \mathcal{M}_n} e_m\big(W_{\mathbf{s}}^n, c_n(\gamma)\big)$$
$$\leq \epsilon, \tag{122}$$

for n large enough. Similarly, by (121b) it also holds that

$$\mathcal{L}(\mathfrak{W}^n, Q_S, \mathsf{C}_n) \triangleq \max_{\mathbf{s} \in \mathcal{T}_{Q_S}^n} \sum_{\gamma \in \Gamma_n} \mu_n(\gamma) \ell\big(V_{\mathbf{s}}^n, P_M^{(U)}, c_n(\gamma)\big) \leq \epsilon, \tag{123}$$

where $P_M^{(U)}$ is the uniform PMF on $\mathcal{M}_n$. In other words, a small maximal error probability and SS (for all the codes in $\mathcal{C}_n$) imply small average error probability and strong secrecy (when taking the expectation over the ensemble $\mathcal{C}_n$).

For any $\gamma \in \Gamma_n$ let $\Upsilon^{(\gamma)}$ be a PMF on $\mathcal{S}^n \times \mathcal{M}_n \times \mathcal{X}^n \times \mathcal{Y}^n \times \mathcal{Z}^n \times \hat{\mathcal{M}}_n$ defined by

$$\Upsilon_{\mathbf{S},M,\mathbf{X},\mathbf{Y},\mathbf{Z},\hat{M}}^{(\gamma)}(\mathbf{s}, m, \mathbf{x}, \mathbf{y}, \mathbf{z}, \hat{m})$$
$$\triangleq Q_S^n(\mathbf{s}) P_{M,\mathbf{X},\mathbf{Y}_{\mathbf{s}},\mathbf{Z}_{\mathbf{s}},\hat{M}}^{(\gamma,\mathbf{s})}(m, \mathbf{x}, \mathbf{y}, \mathbf{z}, \hat{m}), \tag{124}$$

where $P^{(\gamma,\mathbf{s})}$ is an abbreviation of $P^{(c_n(\gamma),\mathbf{s})}$ from (42) and we set $P_M^{(\gamma,\mathbf{s})} = P_M^{(U)}$, for all $\gamma \in \Gamma_n$ and $\mathbf{s} \in \mathcal{S}^n$. Thus, for every $\mathbf{s} \in \mathcal{S}^n$ with $Q_S^n(\mathbf{s}) > 0$ and any $\gamma \in \Gamma_n$, the conditional PMF of $\Upsilon^{(\gamma)}$ given $\mathbf{S} = \mathbf{s}$ equals the corresponding induced PMF $P^{(\gamma,\mathbf{s})}$. Furthermore, let C_n be a random variable that describes the choice of an (n, M_n)-code $c_n(\gamma)$, $\gamma \in \Gamma_n$, from the family $\mathcal{C}_n$ according to the distribution μ_n. We now set

$$\Upsilon_{C_n,\mathbf{S},M,\mathbf{X},\mathbf{Y},\mathbf{Z},\hat{M}}\big(c_n(\gamma), \mathbf{s}, m, \mathbf{x}, \mathbf{y}, \mathbf{z}, \hat{m}\big)$$
$$\triangleq \mu_n(\gamma) \Upsilon_{\mathbf{S},M,\mathbf{X},\mathbf{Y},\mathbf{Z},\hat{M}}^{(\gamma)}(\mathbf{s}, m, \mathbf{x}, \mathbf{y}, \mathbf{z}, \hat{m}) \tag{125a}$$

$$P_{C_n,M,\mathbf{X},\mathbf{Y}_{\mathbf{s}},\mathbf{Z}_{\mathbf{s}},\hat{M}}^{(\mathbf{s})}\big(c_n(\gamma), m, \mathbf{x}, \mathbf{y}, \mathbf{z}, \hat{m}\big)$$
$$\triangleq \mu_n(\gamma) P_{M,\mathbf{X},\mathbf{Y}_{\mathbf{s}},\mathbf{Z}_{\mathbf{s}},\hat{M}}^{(\gamma,\mathbf{s})}(m, \mathbf{x}, \mathbf{y}, \mathbf{z}, \hat{m}). \tag{125b}$$

Henceforth, we use $I_\Upsilon(\cdot)$ and $I_P(\cdot)$ to indicate that a mutual information term is calculated with respect to Υ or $P^{(\mathbf{s})}$ from (125). We now present three technical lemmas that are essential in establishing the result of Theorem 3. For the proofs of Lemmas 7, 8 and 9 see Appendices D, E and F, respectively.

Lemma 7 (Leakage under Typical State Sequence): For any $Q_S \in \mathcal{P}(\mathcal{S})$, $\alpha \in (0, 1]$, $n \in \mathbb{N}_\ell$ and $s_1 \in \mathcal{T}_\alpha^n(Q_S)$, there exists $s_2 \in \mathcal{T}_{Q_S}^n$, such that

$$\Big| I_\Upsilon(M; Z^n | S^n = s_1, C_n) - I_\Upsilon(M; Z^n | S^n = s_2, C_n) \Big|$$
$$\leq n\alpha \log |\mathcal{Z}|. \tag{126}$$

Lemma 8 (Average Leakage under Υ): For any $Q_S \in \mathcal{Q}$, $\alpha \in (0, 1]$ and $n \in \mathbb{N}_\ell$ sufficiently large that is independent of Q_S and α, the following relation holds

$$I_\Upsilon(M; Z^n | S^n, C_n) \leq n\eta_{n,\alpha}^{(1)}, \tag{127}$$

where $\eta_{n,\alpha}^{(1)} \triangleq \frac{\epsilon}{n} + \log |\mathcal{Z}| \left(\alpha + 2|\mathcal{S}|e^{-2n\frac{\alpha^2}{|\mathcal{S}|^2}} \right)$.

Recall the definition of the averaged DMC $W_Q : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$ from (87), given by

$$W_Q(y|x) = \sum_{s \in \mathcal{S}} Q_S(s) W_s(y|x), \quad \forall (x, y) \in \mathcal{X} \times \mathcal{Y}. \tag{128}$$

The n-fold extension of W_Q satisfies

$$W_Q^n(\mathbf{y}|\mathbf{x}) = \prod_{i=1}^n \sum_{s \in \mathcal{S}} Q_S(s) W_s(y_i|x_i)$$
$$= \sum_{\mathbf{s} \in \mathcal{S}^n} Q_S^n(\mathbf{s}) W_{\mathbf{s}}^n(\mathbf{y}|\mathbf{x})$$
$$= \Upsilon(\mathbf{y}|\mathbf{x}), \quad \forall (\mathbf{x}, \mathbf{y}) \in \mathcal{X}^n \times \mathcal{Y}^n. \tag{129}$$

Thus, the conditional marginal PMF $\Upsilon_{\mathbf{Y}|\mathbf{X}}$ of Υ from (125) describes an n-length block transmission over the average channel W_Q^n. As subsequently shown, the derivation of our single-letter upper bound relies on the normalized equivocation of the message M given an output sequence Y^n of the average DMC W_Q^n being small. Commonly, Fano's inequality implies that quantities such as $\frac{1}{n}H(M|Y^n)$ can be made arbitrarily small with n. Here, however, this equivocation term is not directly related to the performance criteria defining CR-assisted achievability. A brute force application of Fano's inequality based on (121a) gives

$$\max_{\mathbf{s} \in \mathcal{T}_{Q_S}^n} H_P(M|Y_{\mathbf{s}}^n, C_n) \leq 1 + \epsilon \log M_n. \tag{130}$$

However, it remains to be shown that (130) implies that $\frac{1}{n}H_\Upsilon(M|Y^n, C_n)$ is small. In general, for any $\mathbf{s} \in \mathcal{T}_{Q_S}^n$, the channel $W_{\mathbf{s}} \in \mathfrak{W}$ is at least as good as W_Q, meaning that the averaged channel induces a possibly larger equivocation. Nonetheless, the equivocation of the message given the output sequence of W_Q^n is upper bounded in Lemma 9.

Lemma 9 (Equivocation under Averaged Channel): For any $Q_S \in \mathcal{Q}$, $\alpha \in \left(0, \frac{1}{2}\right]$ and $n \in \mathbb{N}_\ell$ sufficiently large that is independent of Q_S and α, the following relation holds

$$H_\Upsilon(M|Y^n, C_n) \leq n\eta_{n,\alpha}^{(2)}, \tag{131}$$

where $\eta_{n,\alpha}^{(2)} = \frac{1}{n} + \frac{1}{n} \log M_n \left(\epsilon + 2|\mathcal{S}|e^{-2n\frac{\alpha^2}{|\mathcal{S}|^2}} \right) + \alpha \log |\mathcal{Y}| + 2h(\alpha) + |\mathcal{S}|\frac{\log(n+1)}{n}$ and h is the binary entropy function.

Fix $d \in \left(0, \frac{1}{2}\right)$ and let $\alpha_n = n^{-\left(\frac{1}{2}-d\right)}$, for $n \in \mathbb{N}_\ell$. Accordingly, $\{\alpha_n\}_{n \in \mathbb{N}_\ell}$ vanishes to 0 slower than $\frac{1}{\sqrt{n}}$, which means that for sufficiently large n, random noise is typical with respect to α_n with arbitrarily high probability. Furthermore, $\alpha_n \in \left(0, \frac{1}{2}\right]$, for every $n \geq 2^{\frac{2}{1-2d}}$, so replacing α from Lemmas 8 and 9 with α_n, for sufficiently large n we have

$$I_\Upsilon(M; Z^n | S^n, C_n) \leq n\eta_n^{(1)} \tag{132a}$$
$$H_\Upsilon(M|Y^n, C_n) \leq n\eta_n^{(2)}, \tag{132b}$$

where

$$\eta_n^{(1)} \triangleq \eta_{n,\alpha_n}^{(1)} = \frac{\epsilon}{n} + \log |\mathcal{Z}| \left(\alpha_n + 2|\mathcal{S}|e^{-\frac{2n^{2d}}{|\mathcal{S}|^2}} \right) \quad (133a)$$

$$\eta_n^{(2)} \triangleq \eta_{n,\alpha_n}^{(2)} = \frac{1}{n} + \frac{1}{n} \log M_n \left(\epsilon + 2|\mathcal{S}|e^{-\frac{2n^{2d}}{|\mathcal{S}|^2}} \right) + \alpha_n \log |\mathcal{Y}|$$
$$+ 2h(\alpha_n) + |\mathcal{S}| \frac{\log(n+1)}{n}, \quad (133b)$$

and consequently $\lim_{n \to \infty} \eta_n^{(j)} = 0$, for $j = 1, 2$. Note that the independence of n and α is essential for applying the Lemmas with the vanishing sequence $\{\alpha_n\}_{n \in \mathbb{N}_\ell}$. Furthermore, (132) uniformly hold for all $Q_S \in \mathcal{Q}$.

Having (132)-(133), we proceed with upper bounding the achievable rate R. Unless explicitly stated otherwise, all subsequent information measures are taken with respect to Υ, which is therefore omitted from the notation of mutual information. For any Q_S and $n \in \mathbb{N}_\ell$ sufficiently large (in particular, larger than $2^{\frac{2}{1-2d}}$), we have

$$\log M_n$$
$$\overset{(a)}{\leq} I(M; Y^n | C_n) - I(M; S^n, Z^n | C_n) + n\eta_n$$
$$\overset{(b)}{=} \sum_{i=1}^n \Big[I(M; Y^i, S_{i+1}^n, Z_{i+1}^n | C_n)$$
$$- I(M; Y^{i-1}, S_i^n, Z_i^n | C_n) \Big] + n\eta_n$$
$$= \sum_{i=1}^n \Big[I(M; Y_i | Y^{i-1}, S_{i+1}^n, Z_{i+1}^n, C_n)$$
$$- I(M; S_i, Z_i | Y^{i-1}, S_{i+1}^n, Z_{i+1}^n, C_n) \Big] + n\eta_n$$
$$\overset{(c)}{=} \sum_{i=1}^n \Big[I(M; Y_i | V_i) - I(M; S_i, Z_i | V_i) \Big] + n\eta_n, \quad (134)$$

where:

(a) uses (132) and the independence of (M, S^n, C_n), while defining $\eta_n \triangleq \eta_n^{(1)} + \eta_n^{(2)}$;
(b) follows by a telescoping identity [52, eqs. (9) and (11)] and the independence of C_n, M and S^n;
(c) is by defining $V_i \triangleq (Y^{i-1}, S_{i+1}^n, Z_{i+1}^n, C_n)$, for all $i \in [1 : n]$. The identification of V_i is uniform in $Q_S \in \mathcal{Q}$.

The bound in (134) is rewritten by introducing a time-sharing random variable T that is uniformly distributed over the set $[1 : n]$ and is independent of (S^n, M, X^n, Y^n, Z^n):

$$\frac{1}{n} \log M_n$$
$$\leq \frac{1}{n} \sum_{t=1}^n \Big[I(M; Y_t | V_t) - I(M; S_t, Z_t | V_t) \Big] + \eta_n$$
$$= \sum_{t=1}^n \mathbb{P}\big(T = t\big) \Big[I(M; Y_t | V_t) - I(M; S_t, Z_t | V_t) \Big] + \eta_n$$
$$= I(M; Y_T | V_T, T) - I(M; S_T, Z_T | V_T, T) + \eta_n. \quad (135)$$

Denoting $S_T \triangleq S$, $V \triangleq (V_T, T)$, $U \triangleq (M, V)$, $X \triangleq X_T$, $Y \triangleq Y_T$ and $Z \triangleq Z_T$. Lemma 10 establishes an independence property under Υ, which is key in deriving the factorization

property of the distribution of (S, V, U, X, Y, Z) (induced by Υ), stated in Lemma 11. Both lemmas are proven in Appendix G.

Lemma 10: For any $Q_S \in \mathcal{Q}$, and $i \in [1 : n]$, S_i and $(C_n, S^{n \backslash i}, M, X^n, Y^{n \backslash i}, Z^{n \backslash i})$ are independent under Υ from (125a).

Lemma 11: For any $Q_S \in \mathcal{Q}$ and $(s, v, u, x, y, z) \in \mathcal{S} \times \mathcal{V} \times \mathcal{U} \times \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$, where $\mathcal{V}$ and $\mathcal{U}$ are the alphabets that correspond to the definitions of V and U stated above, the following factorization holds

$$\mathbb{P}_\Upsilon(V = v, U = u, X = x, S = s, Y = y, Z = z)$$
$$= \mathbb{P}_\Upsilon(V = v, U = u, X = x) Q_S(s) W_s(y|x) V_s(z|x). \quad (136)$$

Denoting $\mathbb{P}_\Upsilon(V = v, U = u, X = x) \triangleq Q_{V,U,X}(v, u, x)$, for all $(v, u, x) \in \mathcal{V} \times \mathcal{U} \times \mathcal{X}$, Lemma 11 shows that the joint distribution of (S, V, U, X, Y, Z) factors as stated in Theorem 3.

Finally, we substitute $\eta_n = \eta_n^{(1)} + \eta_n^{(2)}$, while using the definition of $\eta_n^{(2)}$ and (48a), to get that for any $Q_S \in \mathcal{Q}$ and n sufficiently large

$$R < \frac{I_Q(U; Y|V) - I_Q(U; S, Z|V)}{1 - \epsilon - 2|\mathcal{S}|e^{-2n^{2d}}}$$
$$+ \frac{\eta_n^{(1)} + \epsilon + \frac{1}{n} + \alpha_n \log |\mathcal{Y}| + 2h(\alpha_n) + |\mathcal{S}| \frac{\log(n+1)}{n}}{1 - \epsilon - 2|\mathcal{S}|e^{-\frac{2n^{2d}}{|\mathcal{S}|^2}}} + \epsilon, \quad (137)$$

where I_Q denotes that the underlying distribution of the mutual information terms is $Q_{V,U,X} Q_S Q_{Y|X,S} Q_{Z|X,S}$, where $Q_{Y|X,S}(y|x,s) = W_s(y|x)$ and $Q_{Z|X,S}(z|x,s) = V_s(z|x)$, for all $(s, x, y, z) \in \mathcal{S} \times \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$. Letting $n \to \infty$ (which takes α_n and $\eta_n^{(1)}$ to 0) and $\epsilon \to 0$ gives

$$R \leq I_Q(U; Y|V) - I_Q(U; S, Z|V), \quad \forall Q_S \in \mathcal{Q}. \quad (138)$$

Taking an infimum of the RHS (138) over all $Q_S \in \mathcal{Q}$ further gives

$$R \leq \inf_{Q_S \in \mathcal{Q}} \Big[I_Q(U; Y|V) - I_Q(U; S, Z|V) \Big]. \quad (139)$$

Finally, upper bounding the RHS of (139) by maximizing it over all $Q_{V,U,X} \in \mathcal{P}(\mathcal{V} \times \mathcal{U} \times \mathcal{X})$ concludes the proof.

VII. SUMMARY AND CONCLUDING REMARKS

We derived the CR-assisted SS-capacity of the AVWTC with type constrained states. The constraint allows only the state sequences whose empirical distribution is within a small gap from the prescribed type. Achievability relies on a general single-letter lower bound on the capacity of the $\mathcal{Q}$-constrained AVWTC that does not assume the existence of a best channel to the eavesdropper. To establish SS under each of the exponentially many possible states, the mutual information between the message and the eavesdropper's observations was shown to be negligible even when maximized over all message distributions, choices of state sequences and realizations of the CR-code. The SS analysis was based on a heterogeneous version of the strong soft-covering lemma that was recently

presented in [48]. The lemma showed that the probability (with respect to a randomly generated codebook) of the soft-covering phenomenon happening is doubly-exponentially close to one, when transmitting over a state-dependent channel with a certain state sequence realization. The condition for the above is that the rate of the codebook is above the conditional mutual information between the input and output given the state. An application of the union bound combined with a CR-code reduction argument (based on a Chernoff bound) then establishes SS. The resulting reliable and semantically-secure reduced CR-code is over a family of (uncorrelated) codes that is only polynomial in size.

The converse for the type constrained scenario used a general upper bound on $C_R(\mathfrak{W}, \mathfrak{V}, \mathcal{Q})$. Derived uniformly over the constraint set, the upper bound has a max-inf form, and when specialized to a compound WTC over a corresponding constraint set, it improves upon the previously best known single-letter upper bound for that problem [28, Th. 2]. The proof of the upper bound showed that reliability and SS under all state sequences in any type-class imply similar performance when the state sequence is i.i.d. according to the type. The main challenge was in proving that the normalized equivocation of the message given the output sequence is negligible for outputs generated by the averaged main channel. This step required a continuity property that was derived via a novel distribution coupling argument. Combining our upper and lower bounds with some continuity arguments established the SS-capacity of the type constrained AVWTC. The formula has the structure of two subtracted mutual information terms. The first term suggests that the legitimate users effectively transmit over the averaged DMC, which is in general no better than any of the main channels associated with each state. The second (subtracted) mutual information term corresponds to ensuring secrecy versus an eavesdropper with perfect CSI.

Our main goal was to find a single-letter description of the admissible secrecy-rate in an AVWTC scenario while accounting for each of its exponential number of security constraints (instead of relying on assumptions that degenerate the scenario to a single dominating constraint). The heterogeneous strong soft-covering lemma allowed us to do just that, while upgrading to SS. Our achievability proof showed the existence of CR-assisted SS-capacity achieving CR-code of polynomial size. Consequently, combining our code construction with a condition that identifies whether a given type constrained AVWTC has zero or non-zero uncorrelated capacity, will suffice for characterizing the uncorrelated SS-capacity. Such a differentiating condition being currently unknown, we pose it as a question for future research.

APPENDIX A
PROOF OF LEMMA 3

Let $Q_{U,X} \in \mathcal{P}(\mathcal{U} \times \mathcal{X})$ and denote

$$\mathcal{I}(Q_S, Q_{U,X}) \triangleq I_{Q_S}(U; Y) - I_{Q_S}(U; Z|S), \quad (140a)$$

$$\mathcal{I}_\delta(Q_S, Q_{U,X}) \triangleq \min_{Q_1 \in \mathcal{P}_\delta(Q_S)} I_{Q_1}(U; Y)$$
$$- \max_{Q_2 \in \mathcal{P}_\delta(Q_S)} I_{Q_2}(U; Z|S), \quad (140b)$$

where I_Q stands for the mutual information term being calculated with respect to Q as the state distribution. Note that for any $Q_{U,X}$ and $\delta > 0$ we have

$$\mathcal{I}_\delta(Q_S, Q_{U,X}) \leq \mathcal{I}(Q_S, Q_{U,X}), \quad (141)$$

and therefore

$$\mathcal{I}_\delta^*(Q_S) \triangleq \max_{Q_{U,X}} \mathcal{I}_\delta(Q_S, Q_{U,X})$$
$$\leq \max_{Q_{U,X}} \mathcal{I}(Q_S, Q_{U,X})$$
$$= C_R^\star(\mathfrak{W}, \mathfrak{V}, Q_S). \quad (142)$$

Fix $Q_{U,X} \in \mathcal{P}(\mathcal{U} \times \mathcal{X})$. The continuity of mutual information implies that for every $Q_1, Q_2 \in \mathcal{P}_\delta(Q_S)$, we have

$$\left| I_{Q_S}(U; Y) - I_{Q_1}(U; Y) \right| \leq f_1(\delta) \quad (143a)$$

$$\left| I_{Q_S}(U; Z|S) - I_{Q_2}(U; Z|S) \right| \leq f_2(\delta), \quad (143b)$$

where $\lim_{\delta \to 0} f_j(\delta) = 0$, for $j = 1, 2$, uniformly in $Q_{U,X}$ (i.e., f_1 and f_2 are independent of $Q_{U,X}$). For any $\delta > 0$, if $Q_1^\star \in \mathcal{P}_\delta(Q_S)$ achieves $\min_{Q_1 \in \mathcal{P}_\delta(Q_S)} I_{Q_1}(U; Y)$, then (143a) implies

$$I_{Q_S}(U; Y) \leq I_{Q_1^\star}(U; Y) + f_1(\delta)$$
$$= \min_{Q_1 \in \mathcal{P}_\delta(Q_S)} I_{Q_1}(U; Y) + f_1(\delta). \quad (144)$$

Similarly, we also have

$$I_{Q_S}(U; Z|S) \geq \max_{Q_2 \in \mathcal{P}_\delta(Q_S)} I_{Q_2}(U; Z|S) - f_2(\delta). \quad (145)$$

Combining (144) and (145) shows that

$$\mathcal{I}(Q_S, Q_{U,X}) \leq \mathcal{I}_\delta(Q_S, Q_{U,X}) + f_1(\delta) + f_2(\delta), \quad (146)$$

for every $Q_{U,X} \in \mathcal{P}(\mathcal{U} \times \mathcal{X})$ and $\delta > 0$, which, in turn, implies

$$C_R^\star(\mathfrak{W}, \mathfrak{V}, Q_S) \leq \mathcal{I}_\delta^*(Q_S) + f_1(\delta) + f_2(\delta), \quad \forall \delta > 0. \quad (147)$$

Since $\lim_{\delta \to 0} f_j(\delta) = 0$, for $j = 1, 2$, (142) and (147) produce (74).

APPENDIX B
PROOF OF LEMMA 4

Let $\eta > 0$ be arbitrary and with respect to the random codebook B_n, for every $\mathbf{y} \in \mathcal{Y}^n$, define the random variable

$$\Phi_n(\mathbf{y}) = \begin{cases} (m, w), & \max_{\substack{(m', w') \in \mathcal{M}_n \times \mathcal{W}_n: \\ (m', w') \neq (m, w)}} d\big(\mathbf{X}(m', w', \gamma), \mathbf{y}\big) \\ & \qquad\qquad < d\big(\mathbf{X}(m, w, \gamma), \mathbf{y}\big) \\ e, & \text{no } (m, w) \text{ as above exists} \end{cases}$$

$$(148)$$

Furthermore, for every $(m, w) \in \mathcal{M}_n \times \mathcal{W}_n$ and $\mathbf{y} \in \mathcal{Y}^n$, also set $Z(\mathbf{y}, m, w) = \mathbb{1}_{\left\{ \Phi_n(\mathbf{y}) \neq (m, w) \right\}}$. With respect to the measure

$\tilde{\mu}_n$ from (82), we have

$$
\begin{aligned}
&\mathcal{E}_{m,w}(W_n, \tilde{C}_n) \\
&= \mathbb{E}_{\tilde{\mu}_n} \sum_{\mathbf{y} \in \mathcal{Y}^n} W_n(\mathbf{y}|\mathbf{X}(m,w)) Z(\mathbf{y}, m, w) \\
&\overset{(a)}{=} \sum_{\mathbf{x} \in \mathcal{X}^n} Q_X^n(\mathbf{x}) \sum_{\mathbf{y} \in \mathcal{Y}^n} W_n(\mathbf{y}|\mathbf{x}) \\
&\quad \times \mathbb{E}_{\tilde{\mu}_n}\Big[Z(\mathbf{y}, m, w) \Big| \mathbf{X}(m,w) = \mathbf{x} \Big] \\
&= \sum_{(\mathbf{x},\mathbf{y}) \in \mathcal{X}^n \times \mathcal{Y}^n} Q_X^n(\mathbf{x}) W_n(\mathbf{y}|\mathbf{x}) \\
&\quad \times \mathbb{P}_{\tilde{\mu}_n}\Big(\Phi_n(\mathbf{y}) \neq (m,w) \Big| \mathbf{X}(m,w) = \mathbf{x} \Big) \\
&\overset{(b)}{=} \sum_{(\mathbf{x},\mathbf{y}) \in \mathcal{X}^n \times \mathcal{Y}^n} Q_X^n(\mathbf{x}) W_n(\mathbf{y}|\mathbf{x}) \\
&\quad \times \mathbb{P}_{\tilde{\mu}_n}\Bigg(\max_{\substack{(m',w') \in \mathcal{M}_n \times \mathcal{W}_n: \\ (m',w') \neq (m,w)}} d\big(\mathbf{X}(m', w'), \mathbf{y}\big) \geq d(\mathbf{x}, \mathbf{y}) \Bigg),
\end{aligned}
$$
$$ \tag{149} $$

where (a) is the law of total expectation (by first taking a conditional expectation on $\mathbf{X}(m,w)$), while (b) uses the definition of $\Phi_n(\mathbf{y})$ and the independence of the random vectors in the collection $\mathbf{B}_n$.

For all $(\mathbf{x}, \mathbf{y}) \in \mathcal{X}^n \times \mathcal{Y}^n$ with $d(\mathbf{x}, \mathbf{y}) \geq \frac{|\mathcal{M}_n\|\mathcal{W}_n|}{\eta}$, we upper bound the probability on the RHS of (149) as

$$
\begin{aligned}
&\mathbb{P}_{\tilde{\mu}_n}\Bigg(\max_{\substack{(m',w') \in \mathcal{M}_n \times \mathcal{W}_n: \\ (m',w') \neq (m,w)}} d\big(\mathbf{X}(m', w'), \mathbf{y}\big) \geq d(\mathbf{x}, \mathbf{y}) \Bigg) \\
&\leq \mathbb{P}_{\tilde{\mu}_n}\Bigg(\max_{\substack{(m',w') \in \mathcal{M}_n \times \mathcal{W}_n: \\ (m',w') \neq (m,w)}} d\big(\mathbf{X}(m', w'), \mathbf{y}\big) \geq \frac{|\mathcal{M}_n\|\mathcal{W}_n|}{\eta} \Bigg) \\
&= \mathbb{P}_{\tilde{\mu}_n}\Bigg(\bigcup_{\substack{(m',w') \in \mathcal{M}_n \times \mathcal{W}_n: \\ (m',w') \neq (m,w)}} \Big\{ d\big(\mathbf{X}(m', w'), \mathbf{y}\big) \geq \frac{|\mathcal{M}_n\|\mathcal{W}_n|}{\eta} \Big\} \Bigg) \\
&\overset{(a)}{\leq} \sum_{\substack{(m',w') \in \mathcal{M}_n \times \mathcal{W}_n: \\ (m',w') \neq (m,w)}} \mathbb{P}_{Q_X^n}\Big(d(\mathbf{X}, \mathbf{y}) \geq \frac{|\mathcal{M}_n\|\mathcal{W}_n|}{\eta} \Big) \\
&\overset{(b)}{\leq} \eta \sum_{\substack{(m',w') \in \mathcal{M}_n \times \mathcal{W}_n: \\ (m',w') \neq (m,w)}} \frac{\mathbb{E}_{Q_X^n} d(\mathbf{X}, \mathbf{y})}{|\mathcal{M}_n\|\mathcal{W}_n|} \\
&\overset{(c)}{\leq} \eta
\end{aligned}
$$
$$ \tag{150} $$

where (a) uses the union bound and the fact that $\mathbf{X}(m', w') \sim Q_X^n$, for all $(m', w') \in \mathcal{M}_n \times \mathcal{W}_n$, (b) is Markov's inequality and (c) follows by the assumption that $\mathbb{E}_{Q_X^n} d(\mathbf{X}, \mathbf{y}) \leq 1$, for all $\mathbf{y} \in \mathcal{Y}^n$.

Plugging (150) back into (149) completes the proof:

$$
\begin{aligned}
&\mathcal{E}_{m,w}(W_n, \tilde{C}_n) \\
&\leq \sum_{\substack{(\mathbf{x},\mathbf{y}) \in \mathcal{X}^n \times \mathcal{Y}^n: \\ d(\mathbf{x},\mathbf{y}) < \frac{|\mathcal{M}_n\|\mathcal{W}_n|}{\eta}}} Q_X^n(\mathbf{x}) W_n(\mathbf{y}|\mathbf{x}) \cdot 1 \\
&\quad + \sum_{\substack{(\mathbf{x},\mathbf{y}) \in \mathcal{X}^n \times \mathcal{Y}^n: \\ d(\mathbf{x},\mathbf{y}) \geq \frac{|\mathcal{M}_n\|\mathcal{W}_n|}{\eta}}} Q_X^n(\mathbf{x}) W_n(\mathbf{y}|\mathbf{x}) \cdot \eta \\
&\leq \mathbb{P}_{Q_X^n W_n}\Big(d(\mathbf{X}, \mathbf{Y}) < \frac{|\mathcal{M}_n\|\mathcal{W}_n|}{\eta} \Big) + \eta.
\end{aligned}
$$
$$ \tag{151} $$

APPENDIX C
PROOF OF LEMMA 5

First define

$$
\mathfrak{W}_Q \triangleq \Big\{ W_Q : \mathcal{X} \to \mathcal{P}(\mathcal{Y}) \Big| Q \in \mathcal{Q} \Big\}, \tag{152}
$$

and note that the convexity of $\mathcal{Q}$ implies that $\mathfrak{W}_Q$ is also a convex set. Throughout this proof we make use of a slightly modified notation of mutual information. Specifically, we represent the mutual information between a pair of random variables in terms of their underlying joint distribution, i.e., for any $P \in \mathcal{P}(\mathcal{X})$ and $W : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$, let $I(P, W) \triangleq I(X; Y)$, where $(X, Y) \sim P \times W$. Accordingly, we may write

$$
\min_{Q \in \mathcal{Q}} I_Q(X; Y) = \min_{W \in \mathfrak{W}_Q} I(Q_X, W). \tag{153}
$$

Let $\tilde{W} \in \mathfrak{W}_Q$ be a channel that achieves the RHS of (153), i.e., with respect to the notation in the error probability analysis from Section V, we have

$$
I(Q_X, \tilde{W}) = I_{\tilde{Q}}(X; Y). \tag{154}
$$

The convexity of $\mathfrak{W}_Q$ implies that for every $W \in \mathfrak{W}_Q$ and $\alpha \in [0, 1]$

$$
I(Q_X, \alpha W + (1-\alpha)\tilde{W}) \geq I(Q_X, \tilde{W}), \tag{155}
$$

and therefore,

$$
\lim_{\alpha \searrow 0} \frac{\partial}{\partial \alpha} I(Q_X, \alpha W + (1-\alpha)\tilde{W}) \geq 0. \tag{156}
$$

Similarly to [47, eq. (12.19)], since

$$
\begin{aligned}
&\frac{\partial}{\partial \alpha} I(Q_X, \alpha W + (1-\alpha)\tilde{W}) \\
&= \sum_{(x,y) \in \mathcal{X} \times \mathcal{Y}} Q_X(x)\big(W(y|x) - \tilde{W}(y|x)\big) \\
&\quad \times \log\left(\frac{\alpha W(y|x) + (1-\alpha)\tilde{W}(y|x)}{\alpha Q_Y(y) + (1-\alpha)\tilde{Q}_Y(y)} \right),
\end{aligned}
$$
$$ \tag{157} $$

where $Q_Y(y) = \sum_{x \in \mathcal{X}} Q_X(x) W(y|x)$ and $\tilde{Q}_Y(y) = \sum_{x \in \mathcal{X}} Q_X(x) \tilde{W}(y|x)$, it follows that

$$
\sum_{(x,y) \in \mathcal{X} \times \mathcal{Y}} Q_X(x) W(y|x) \log\left(\frac{\tilde{W}(y|x)}{\tilde{Q}_Y(y)}\right)
$$
$$
\geq I(Q_X, \tilde{W})
$$
$$
= I_{\tilde{Q}}(X; Y). \tag{158}
$$

APPENDIX D
PROOF OF LEMMA 7

Fix $Q_S \in \mathcal{Q}$, $\alpha \in (0, 1]$, $n \in \mathbb{N}_\ell$ and $\mathbf{s}_1 \in \mathcal{T}_\alpha^n(Q_S)$. Clearly, there exists an $\mathbf{s}_2 \in \mathcal{T}_{Q_S}^n$, such that

$$
d_H(\mathbf{s}_1, \mathbf{s}_2) \leq n\alpha, \tag{159}
$$

where $d_H : \mathcal{S}^n \times \mathcal{S}^n \to [0 : n]$ is the Hamming distance function. Let $\mathcal{A}$ be the set of indices for which the components of $\mathbf{s}_1$ and $\mathbf{s}_2$ coincide, i.e.,

$$
\mathcal{A} = \{i \in [1 : n] | s_{1,i} = s_{2,i}\}. \tag{160}
$$

Note that (159) implies that $|\mathcal{A}^c| \leq n\alpha$.

Recall that for any subset $\emptyset \neq \mathcal{A} \subseteq [1 : n]$ and any n-dimensional vector $\mathbf{x} \in \mathcal{X}^n$, we denote the vector of elements from $\mathbf{x}$ with indices in $\mathcal{A}$ by $\mathbf{x}^{\mathcal{A}}$, that is, $\mathbf{x}^{\mathcal{A}} = (x_i)_{i \in \mathcal{A}}$. Similar convention is used for random vectors, while using uppercase letters. By the mutual information chain rule, the absolute value of the difference of mutual information terms from (126) is upper bounded as follows:

$$
\left| I_\Upsilon(M; Z^n | S^n = \mathbf{s}_1, C_n) - I_\Upsilon(M; Z^n | S^n = \mathbf{s}_2, C_n) \right|
$$
$$
\leq \left| I_\Upsilon(M; \mathbf{Z}^{\mathcal{A}} | S^n = \mathbf{s}_1, C_n) - I_\Upsilon(M; \mathbf{Z}^{\mathcal{A}} | S^n = \mathbf{s}_2, C_n) \right|
$$
$$
+ \Big| I_\Upsilon(M; \mathbf{Z}^{\mathcal{A}^c} | \mathbf{Z}^{\mathcal{A}}, S^n = \mathbf{s}_1, C_n)
$$
$$
- I_\Upsilon(M; \mathbf{Z}^{\mathcal{A}^c} | \mathbf{Z}^{\mathcal{A}}, S^n = \mathbf{s}_2, C_n) \Big|
$$
$$
\overset{(a)}{\leq} \left| I_\Upsilon(M; \mathbf{Z}^{\mathcal{A}} | S^{\mathcal{A}} = \mathbf{s}_1^{\mathcal{A}}, C_n) - I_\Upsilon(M; \mathbf{Z}^{\mathcal{A}} | S^{\mathcal{A}} = \mathbf{s}_2^{\mathcal{A}}, C_n) \right|
$$
$$
+ \max \left\{ \begin{array}{l} I_\Upsilon(M; \mathbf{Z}^{\mathcal{A}^c} | \mathbf{Z}^{\mathcal{A}}, S = \mathbf{s}_1, C_n), \\ I_\Upsilon(M; \mathbf{Z}^{\mathcal{A}^c} | \mathbf{Z}^{\mathcal{A}}, S = \mathbf{s}_2, C_n) \end{array} \right\}.
$$
$$
\overset{(b)}{\leq} \max \left\{ H_\Upsilon(\mathbf{Z}^{\mathcal{A}^c} | S^n = \mathbf{s}_1), H_\Upsilon(\mathbf{Z}^{\mathcal{A}^c} | S^n = \mathbf{s}_2) \right\}
$$
$$
\overset{(c)}{\leq} n\alpha \log |\mathcal{Z}| \tag{161}
$$

where:

- (a) is because for any $m \in \mathcal{M}_n$, $\mathbf{z}^{\mathcal{A}} \in \mathcal{Z}^{|\mathcal{A}|}$ and $\mathbf{s}_j$, where $j = 1, 2$, (124) implies

$$
\Upsilon(m, \mathbf{z}^{\mathcal{A}} | \mathbf{s}_j)
$$
$$
= \sum_{\gamma \in \Gamma_n} \mu_n(\gamma) \frac{1}{M_n} \sum_{\mathbf{x} \in \mathcal{X}^n} f_\gamma(\mathbf{x}|m) V_{\mathbf{s}_j^{\mathcal{A}}}^{|\mathcal{A}|}(\mathbf{z}^{\mathcal{A}} | \mathbf{x}^{\mathcal{A}})
$$
$$
= \Upsilon(m, \mathbf{z}^{\mathcal{A}} | \mathbf{s}_j^{\mathcal{A}}); \tag{162}
$$

- (b) is since $\mathbf{s}_1^{\mathcal{A}} = \mathbf{s}_2^{\mathcal{A}}$, because conditioning cannot increase entropy and the memoryless property from (162);
- (c) holds since entropy is maximized by the uniform distribution and because $|\mathcal{A}^c| \leq n\alpha$.

APPENDIX E
PROOF OF LEMMA 8

Fix $Q_S \in \mathcal{Q}$ and $\alpha \in (0, 1]$. First observe that for any $n \in \mathbb{N}_\ell$, we have

$$
\max_{\mathbf{s} \in \mathcal{T}_\alpha^n(Q_S)} I_\Upsilon(M; Z^n | S^n = \mathbf{s}, C_n)
$$
$$
\overset{(a)}{\leq} \max_{\mathbf{s} \in \mathcal{T}_{Q_S}^n} I_\Upsilon(M; Z^n | S^n = \mathbf{s}, C_n) + n\alpha \log |\mathcal{Z}|
$$
$$
\overset{(b)}{=} \mathcal{L}(\mathfrak{V}^n, Q_S, \mathsf{C}_n) + n\alpha \log |\mathcal{Z}|, \tag{163}
$$

where (a) follows from Lemma 7, while (b) is because $\Upsilon_{M,\mathbf{Z}|S=\mathbf{s}} = P_{M,\mathbf{Z}_\mathbf{s}}^{(\mathbf{s})}$, for every $\mathbf{s} \in \mathcal{S}^n$, and the notation in (123). On account of (123), for sufficiently large values of n that are independent of $Q_S \in \mathcal{Q}$, it holds that

$$
\mathcal{L}(\mathfrak{V}^n, Q_S, \mathsf{C}_n) \leq \epsilon. \tag{164}
$$

Consequently, for those values of n, we have[13]

$$
I_\Upsilon(M; Z^n | S^n, C_n)
$$
$$
= \sum_{\mathbf{s} \in \mathcal{S}^n} Q_S^n(\mathbf{s}) I_\Upsilon(M; Z^n | S^n = \mathbf{s}, C_n)
$$
$$
\leq \sum_{\mathbf{s} \in \mathcal{T}_\alpha^n(Q_S)} Q_S^n(\mathbf{s}) I_\Upsilon(M; Z^n | S^n = \mathbf{s}, C_n)
$$
$$
+ \sum_{\mathbf{s} \notin \mathcal{T}_\alpha^n(Q_S)} Q_S^n(\mathbf{s}) n \log |\mathcal{Z}|
$$
$$
\leq \max_{\mathbf{s} \in \mathcal{T}_\alpha^n(Q_S)} I_\Upsilon(M; Z^n | S^n = \mathbf{s}, C_n)
$$
$$
+ n \log |\mathcal{Z}| \cdot \mathbb{P}_{Q_S^n}\big(S^n \notin \mathcal{T}_\alpha^n(Q_S)\big)
$$
$$
\overset{(a)}{\leq} \mathcal{L}(\mathfrak{V}^n, Q_S, \mathsf{C}_n) + n \log |\mathcal{Z}| \left(\alpha + 2|\mathcal{S}| e^{-2n \frac{\alpha^2}{|\mathcal{S}|^2}}\right)
$$
$$
\overset{(b)}{\leq} n \eta_{n,\alpha}^{(1)}, \tag{165}
$$

where (a) uses (163) and the upper bound on the probability of drawing an atypical sequence from (5), while (b) follows from (164).

APPENDIX F
PROOF OF LEMMA 9

Fix $Q_S \in \mathcal{Q}$. For any $n \in \mathbb{N}_\ell$, let $\mathcal{E}_U(\mathfrak{W}^n, Q_S, \mathsf{C}_n)$ be the average error probability under the CR-code C_n when the state sequence is chosen randomly and uniformly over $\mathcal{T}_{Q_S}^n$. Clearly, for any n sufficiently large (that is independent of Q_S), we have

$$
\mathcal{E}_U(\mathfrak{W}^n, Q_S, \mathsf{C}_n) \leq \bar{\mathcal{E}}(\mathfrak{W}^n, Q_S, \mathsf{C}_n) \overset{(a)}{\leq} \epsilon, \tag{166}
$$

where (a) is on account of (122).

For each $\gamma \in \Gamma_n$, the PMF on $\mathcal{S}^n \times \mathcal{M}_n \times \mathcal{X}^n \times \mathcal{Y}^n \times \mathcal{Z}^n \times \hat{\mathcal{M}}_n$ describing the random experiment where the state sequence is uniformly drawn from $\mathcal{T}_{Q_S}^n$ is

$$
\Lambda_1^{(\gamma)}(\mathbf{s}, m, \mathbf{x}, \mathbf{y}, \mathbf{z}, \hat{m}) \triangleq \frac{\mathbb{1}_{\{\mathbf{s} \in \mathcal{T}_{Q_S}^n\}}}{|\mathcal{T}_{Q_S}^n|} P^{(\gamma, \mathbf{s})}(m, \mathbf{x}, \mathbf{y}, \mathbf{z}, \hat{m}), \tag{167}
$$

---

[13] Only the last step relies on n being sufficiently large; all other steps are valid for every $n \in \mathbb{N}_\ell$

for all $(\mathbf{s}, m, \mathbf{x}, \mathbf{y}, \mathbf{z}, \hat{m}) \in \mathcal{S}^n \times \mathcal{M}_n \times \mathcal{X}^n \times \mathcal{Y}^n \times \mathcal{Z}^n \times \hat{\mathcal{M}}_n$. As before, we set

$$\Lambda_1\big(c_n(\gamma), \mathbf{s}, m, \mathbf{x}, \mathbf{y}, \mathbf{z}, \hat{m}\big) \triangleq \mu_n(\gamma)\Lambda_1^{(\gamma)}(\mathbf{s}, m, \mathbf{x}, \mathbf{y}, \mathbf{z}, \hat{m}). \tag{168}$$

By (166) and Fano's inequality, we have

$$H_{\Lambda_1}(M|Y^n, C_n) \leq 1 + \mathcal{E}_U(\mathfrak{W}^n, Q_S, \mathsf{C}_n) \cdot \log M_n$$
$$\leq 1 + \epsilon \log M_n, \tag{169}$$

where the last inequality holds for the aforementioned sufficiently large n values.

To upper bound $H_\Upsilon(M|Y^n, C_n)$ in terms of $H_{\Lambda_1}(M|Y^n, C_n)$, we first index all the types in $\mathcal{P}_n(\mathcal{S})$ by $i \in \mathcal{B} \triangleq [1 : |\mathcal{P}_n(\mathcal{S})|]$. Set $Q_1 = Q_S$, and associate with every $Q_i \in \mathcal{P}_n(\mathcal{S})$, $i \neq 1$, a PMF Λ_i on $\mathcal{C}_n \times \mathcal{S}^n \times \mathcal{M}_n \times \mathcal{X}^n \times \mathcal{Y}^n \times \mathcal{Z}^n \times \hat{\mathcal{M}}_n$, that is defined analogously to Λ_1 from (168). Thus, with respect to Λ_i, the state sequence is uniformly chosen from $\mathcal{T}_{Q_i}^n$.

Let B be a random variable over $\mathcal{B}$ that takes the value $B = i$, for $i \in \mathcal{B}$, if an i.i.d. state sequence $S^n \sim Q_S^n$ satisfies $S^n \in \mathcal{T}_{Q_i}^n$. First note that

$$I_\Upsilon(M; Y^n|C_n) = I_\Upsilon(M; B, Y^n|C_n) - I_\Upsilon(M; B|Y^n, C_n)$$
$$\geq I_\Upsilon(M; B, Y^n|C_n) - \log|\mathcal{B}|$$
$$\geq I_\Upsilon(M; B, Y^n|C_n) - |\mathcal{S}|\log(n+1), \tag{170}$$

which implies

$$H_\Upsilon(M|Y^n, C_n) \leq H_\Upsilon(M|Y^n, B, C_n) + |\mathcal{S}|\log(n+1). \tag{171}$$

Next, we expand the conditional entropy from the RHS of (171) with respect to B, while splitting it into typical and atypical realizations of B. Fix $\alpha \in \left(0, \frac{1}{2}\right]$ and define

$$\mathcal{I}(Q_S, \alpha) = \left\{i \in \mathcal{B} \,\middle|\, \mathcal{T}_{Q_i}^n \subset \mathcal{T}_\alpha^n(Q_S)\right\}. \tag{172}$$

For any $n \in \mathbb{N}_\ell$, we have

$$H_\Upsilon(M|Y^n, B, C_n)$$
$$\leq \sum_{i \in \mathcal{I}(Q_S, \alpha)} \mathbb{P}_{Q_S^n}\big(S^n \in \mathcal{T}_{Q_i}^n\big) H_\Upsilon(M|Y^n, B = i, C_n)$$
$$\quad + \mathbb{P}_{Q_S^n}\big(S^n \notin \mathcal{T}_\alpha^n(Q_S)\big) \cdot \log M_n$$
$$\overset{(a)}{\leq} \sum_{i \in \mathcal{I}(Q_S, \alpha)} \mathbb{P}_{Q_S^n}\big(S^n \in \mathcal{T}_{Q_i}^n\big) H_\Upsilon\big(M\big|Y^n, S^n \in \mathcal{T}_{Q_i}^n, C_n\big)$$
$$\quad + 2|\mathcal{S}|e^{-2n\frac{\alpha^2}{|\mathcal{S}|^2}} \log M_n, \tag{173}$$

where (a) uses (5) and the definition of B. Recall that

$$\mathbb{P}_{Q_S^n}\big(S^n = \mathbf{s}\big|S^n \in \mathcal{T}_{Q_i}^n\big) = \frac{\mathbb{1}_{\left\{\mathbf{s} \in \mathcal{T}_{Q_i}^n\right\}}}{|\mathcal{T}_{Q_i}^n|}, \quad \forall \mathbf{s} \in \mathcal{S}^n, \ \forall i \in \mathcal{B}, \tag{174}$$

and therefore,

$$H_\Upsilon(M|Y^n, B = i, C_n) = H_\Upsilon\big(M\big|Y^n, S^n \in \mathcal{T}_{Q_i}^n, C_n\big)$$
$$= H_{\Lambda_i}(M|Y^n, C_n). \tag{175}$$

Having this, we remark that the main idea in upper bounding $H_\Upsilon(M|Y^n, C_n)$ is to use (171) and (173), while arguing that the conditional entropy $H_{\Lambda_i}(M|Y^n, C_n)$ is small for any $1 \neq i \in \mathcal{I}(Q_S, \alpha)$ (i.e., for any random state sequence in the typical set) as long as we know it is small for one single type in the typical set (i.e., as long as (169) holds). To do so, we show that $H_{\Lambda_i}(M|Y^n, C_n)$, for all $i \in \mathcal{I}(Q_S, \alpha)$, can be replaced with $H_{\Lambda_1}(M|Y^n, C_n)$ with a correction term that can be made arbitrarily small when normalized by n. With some abuse of notation, we henceforth denote by Λ_1 the marginal PMF of (168) on $\mathcal{S}^n \times \mathcal{M}_n \times \mathcal{X}^n \times \mathcal{Y}^n$. Similarly, for any $i \in \mathcal{I}(Q_S, \alpha)$, Λ_i denotes the marginal on $\mathcal{S}^n \times \mathcal{M}_n \times \mathcal{X}^n \times \mathcal{Y}^n$ of the original Λ_i.

Fix $i \in \mathcal{I}(Q_S, \alpha)$ and let $\Lambda_{1,i}$ be a coupling of Λ_1 and Λ_i, which is a PMF on $\mathcal{S}^n \times \mathcal{M}_n \times \mathcal{X}^n \times \mathcal{Y}^n \times \mathcal{S}^n \times \mathcal{Y}^n$ that is defined by the following steps:

1) Similarly to the definition of the set $\mathcal{A}$ from (160), for any $(\mathbf{s}, \tilde{\mathbf{s}}) \in \mathcal{S}^n \times \mathcal{S}^n$ set

$$\mathcal{A}(\mathbf{s}, \tilde{\mathbf{s}}) \triangleq \big\{j \in [1 : n]\big|s_j = \tilde{s}_j\big\}. \tag{176}$$

2) For any $\gamma \in \Gamma_n$, define

$$\Lambda_{1,i}^{(\gamma)}(\mathbf{s}, m, \mathbf{x}, \mathbf{y}, \tilde{\mathbf{s}}, \tilde{\mathbf{y}})$$
$$\triangleq \Lambda_1^{(\gamma)}(\mathbf{s}, m, \mathbf{x}, \mathbf{y})\Lambda_{1,i}(\tilde{\mathbf{s}}|\mathbf{s})\Lambda_{1,i}(\tilde{\mathbf{y}}|\mathbf{s}, \mathbf{x}, \mathbf{y}, \tilde{\mathbf{s}}), \tag{177}$$

where

$$\Lambda_{1,i}(\tilde{\mathbf{y}}|\mathbf{s}, \mathbf{x}, \mathbf{y}, \tilde{\mathbf{s}}) = \mathbb{1}_{\bigcap_{j \in \mathcal{A}(\mathbf{s}, \tilde{\mathbf{s}})}\left\{\tilde{y}_j = y_j\right\}} \prod_{j \in \mathcal{A}(\mathbf{s}, \tilde{\mathbf{s}})^c} W_{\tilde{s}_j}(\tilde{y}_j|x_j), \tag{178}$$

and $\Lambda_{1,i}(\tilde{\mathbf{s}}|\mathbf{s})$ is defined by the pseudo-algorithm given at the next page.

Namely, in each step the algorithm first uniformly chooses an index $j \in [1 : n]$, such that the number of appearances of s_j in $\mathbf{s}$ is above the quota allowed by Q_i. Then, s_j is replaced with a symbol $s \in \mathcal{S}$ that is uniformly chosen from the set of symbols whose number of appearances in $\mathbf{s}$ is below the quote subscribed by Q_i. The procedure repeats itself until the modified sequence belongs to $\mathcal{T}_{Q_i}^n$. Clearly, the algorithm stops after a finite number of cycles, since in each cycle the sequence $\mathbf{s}'$ is adjusted so that its empirical PMF of $\nu_{\mathbf{s}'}$ is closer to Q_i. We give a formal justification for the finite running time argument subsequently.

3) As a last step, we set

$$\Lambda_{1,i}\big(c_n(\gamma), \mathbf{s}, m, \mathbf{x}, \mathbf{y}, \tilde{\mathbf{s}}, \tilde{\mathbf{y}}\big)$$
$$= \mu_n(\gamma)\Lambda_{1,i}^{(\gamma)}(\mathbf{s}, m, \mathbf{x}, \mathbf{y}, \tilde{\mathbf{s}}, \tilde{\mathbf{y}}). \tag{179}$$

For any $1 \neq i \in \mathcal{I}(Q_S, \alpha)$, the symmetry in constructing $\Lambda_{1,i}(\tilde{\mathbf{s}}|\mathbf{s})$, the uniformity of $\Lambda_{1,i}(\mathbf{s})$ over $\mathcal{T}_{Q_S}^n$, and the fact that sequences of the same type are merely permutations of one another, imply that the marginal PMF of $\tilde{S}^n$ with respect to $\Lambda_{1,i}$ is uniform over $\mathcal{T}_{Q_i}^n$, i.e.,

$$\Lambda_{1,i}(\tilde{\mathbf{s}}) = \Lambda_i(\tilde{\mathbf{s}}) = \frac{\mathbb{1}_{\left\{\tilde{\mathbf{s}} \in \mathcal{T}_{Q_i}^n\right\}}}{|\mathcal{T}_{Q_i}^n|}. \tag{180}$$

Algorithm 1 Construction of $\Lambda_{1,i}(\tilde{\mathbf{s}}|\mathbf{s})$

1: $\mathbf{s}' \leftarrow \mathbf{s}$
2: **while** $\mathbf{s}' \notin \mathcal{T}_{Q_i}^n$ **do**
3:     $\mathcal{J}(\mathbf{s}') := \left\{ j \in [1:n] \,\big|\, N(s'_j|\mathbf{s}') > nQ_i(s'_j) \right\}$
4:     $\mathcal{L}(\mathbf{s}') := \left\{ s \in \mathcal{S} \,\big|\, N(s|\mathbf{s}') < nQ_i(s) \right\}$
5:     Draw $j \sim \text{Unif}\left(\mathcal{J}(\mathbf{s}')\right)$
6:     Draw $s \sim \text{Unif}\left(\mathcal{L}(\mathbf{s}')\right)$
7:     $s'_j \leftarrow s$
8: **end while**

Recalling the definition of $\Lambda_{1,i}(\tilde{\mathbf{y}}|\mathbf{s}, \mathbf{x}, \mathbf{y}, \tilde{\mathbf{s}})$ from (178), we thus obtain

$$\Lambda_{1,i}(\tilde{\mathbf{s}}, m, \mathbf{x}, \tilde{\mathbf{y}}) = \Lambda_i(\tilde{\mathbf{s}}, m, \mathbf{x}, \tilde{\mathbf{y}}), \tag{181}$$

which shows that $\Lambda_{1,i}$ is a valid coupling of Λ_1 and Λ_i.

Some additional properties of the algorithms output are needed. Denote by $K \in \mathbb{N} \cup \{\infty\}$ the number of cycles it takes until Algorithm 1 terminates (i.e., for now, K may be infinite). For each $k \in [1:K]$ (if $K = \infty$ then $k \in [1:K]$ is to be understood as $k \in \mathbb{N}$), denote by $\mathbf{s}'_k$ the $\mathbf{s}'$ sequence obtained after the k-th cycle. Accordingly, $\mathbf{s}'_0 = \mathbf{s}$, and if indeed $K < \infty$, then $\mathbf{s}'_K = \tilde{\mathbf{s}}$. To analyse the algorithm's operation, for each $k \in [1:K]$, define

$$\mathcal{L}_k^{(h)} = \left\{ s \in \mathcal{S} \,\big|\, N(s|\mathbf{s}'_k) > nQ_i(s) \right\} \tag{182a}$$

$$\mathcal{L}_k^{(l)} = \left\{ s \in \mathcal{S} \,\big|\, N(s|\mathbf{s}'_k) < nQ_i(s) \right\}, \tag{182b}$$

and further set

$$N_k^{(h)} = \sum_{s \in \mathcal{L}_k^{(h)}} \left| N(s|\mathbf{s}'_k) - nQ_i(s) \right| \tag{183a}$$

$$N_k^{(l)} = \sum_{s \in \mathcal{L}_k^{(l)}} \left| N(s|\mathbf{s}'_k) - nQ_i(s) \right|. \tag{183b}$$

Note that $N_k^{(h)} = N_k^{(l)}$, for every $k \in [1:K]$, and that in each iteration both $N_k^{(h)}$ and $N_k^{(l)}$ reduce by 1, i.e.,

$$N_k^{(h)} = N_{k-1}^{(h)} - 1, \quad k \in [1:K]. \tag{184}$$

Clearly, Algorithm 1 terminates once $\mathcal{L}_k^{(h)} = \mathcal{L}_k^{(l)} = \emptyset$, or equivalently, once $N_k^{(h)} = N_k^{(l)} = 0$. When combined with (184), this characterizes $K = N_0^{(h)}$, thus justifying the finite running time of the algorithm. Consequently, if $\left(S^n, \tilde{S}^n\right) \sim \Lambda_{1,i}$, then

$$\mathbb{P}_{\Lambda_{1,i}}\left(\left(S^n, \tilde{S}^n\right) \in \mathcal{T}_{Q_S}^n \times \mathcal{T}_{Q_i}^n \right) = 1. \tag{185}$$

Another important outcome of the algorithm's operation is that the sequences S^n and $\tilde{S}^n$ jointly distributed according to $\Lambda_{1,i}$ are almost surely within a Hamming distance of at most $n\alpha$. Namely, we claim that

$$\mathbb{P}_{\Lambda_{1,i}}\left(d_H\left(S^n, \tilde{S}^n\right) \leq n\alpha \right) = 1. \tag{186}$$

To see the validity of (186), note that in each iteration one symbol of the current $\mathbf{s}'_k$ is altered. Furthermore, an altered symbol is never modified again in any of the succeeding iterations. As a consequence, this observation implies that

$$d_H(\mathbf{s}'_{k-1}, \mathbf{s}'_k) = 1, \quad \forall k \in [1:K], \tag{187}$$

and when combined with the fact the $d_H(\mathbf{s}, \mathbf{s}'_0) = 0$, we obtain

$$d_H(\mathbf{s}, \mathbf{s}'_k) = d_H(\mathbf{s}, \mathbf{s}'_{k-1}) + 1, \quad \forall k \in [1:K]. \tag{188}$$

Thus, to show that for $d_H(\mathbf{s}, \tilde{\mathbf{s}}) \leq n\alpha$, it suffices to show that the number of cycles $K \leq n\alpha$ (keeping in mind that $\mathbf{s}'_K = \tilde{\mathbf{s}}$). Indeed, we have

$$K = N_0^{(h)} \overset{(a)}{=} \sum_{s \in \mathcal{L}_0^{(h)}} \left| N(s|\mathbf{s}) - nQ_i(s) \right| \overset{(b)}{\leq} n \sum_{s \in \mathcal{L}_0^{(h)}} \frac{\alpha}{|\mathcal{S}|} \overset{(c)}{\leq} n\alpha, \tag{189}$$

where (a) uses (183a), (b) is because $\mathcal{T}_{Q_i}^n \subset \mathcal{T}_\alpha^n(Q_S)$, while (c) follows since $\left|\mathcal{L}_0^{(h)}\right| \leq |\mathcal{S}|$.

Having (185)-(186), let $\mathcal{A}_i\left(S^n, \tilde{S}^n\right)$ be a random variable defined by (176), where $\left(S^n, \tilde{S}^n\right) \sim \Lambda_{1,i}$ and $1 \neq i \in \mathcal{I}(Q_S, \alpha)$. We abbreviate $\mathcal{A}_i\left(S^n, \tilde{S}^n\right)$ as $\mathcal{A}_i$ (and further omit the index i, when it is clear from the context). Define the random variable $\mathcal{A}_i^c \triangleq [1:n] \setminus \mathcal{A}_i$, and denote its alphabet by $\mathfrak{A}_i^c$. As a consequence of (185), for any $n \in \mathbb{N}_\ell$, the cardinality of $\mathcal{A}_i^c$ is upper bounded by [53, Sec. 3.1]

$$|\mathfrak{A}_i^c| = \sum_{j=0}^{\lfloor n\alpha \rfloor} \binom{n}{j} \leq 2^{nh(\alpha)}, \quad \forall 1 \neq i \in \mathcal{I}(Q_S, \alpha), \tag{190}$$

where h is the binary entropy function. Since $\mathcal{A}_i$ and $\mathcal{A}_i^c$ uniquely define one another, (190) yields

$$H_{\Lambda_{1,i}}(\mathcal{A}) = H_{\Lambda_{1,i}}(\mathcal{A}^c) \leq \log |\mathfrak{A}_i^c| \leq nh(\alpha), \quad \forall i \in \mathcal{I}(Q_S, \alpha), \tag{191}$$

We are now ready to link the mutual information between the message and the legitimate user's output sequence under Λ_i, to the corresponding term under Λ_1. This, in turn, yields an upper bound on $H_{\Lambda_i}(M|Y^n, C_n)$ in terms of $H_{\Lambda_1}(M|Y^n, C_n)$, that when combined with (173), suffices to establish Lemma 9. In the following chain of inequalities $(\tilde{S}^n, \tilde{Y}^n)$ and (S^n, Y^n) denote the state sequence and legitimate channel output when distributed according to Λ_i and Λ_1, respectively. Fix $1 \neq i \in \mathcal{I}(Q_S, \alpha)$ and for any $n \in \mathbb{N}_\ell$ consider:

$$
\begin{aligned}
&I_{\Lambda_i}(M; \tilde{Y}^n|C_n) \\
&\overset{(a)}{\geq} I_{\Lambda_{1,i}}(M; \tilde{Y}^n|\mathcal{A}, C_n) - H_{\Lambda_{1,i}}(\mathcal{A}) \\
&\geq I_{\Lambda_{1,i}}(M; \tilde{\mathbf{Y}}^{\mathcal{A}}|\mathcal{A}, C_n) - H_{\Lambda_{1,i}}(\mathcal{A}) \\
&\overset{(b)}{=} I_{\Lambda_{1,i}}(M; \mathbf{Y}^{\mathcal{A}}|\mathcal{A}, C_n) - H_{\Lambda_{1,i}}(\mathcal{A}) \\
&= I_{\Lambda_{1,i}}(M; \mathbf{Y}^{\mathcal{A}}|\mathcal{A}, C_n) + I_{\Lambda_{1,i}}(M; \mathbf{Y}^{\mathcal{A}^c}|\mathbf{Y}^{\mathcal{A}}, \mathcal{A}, C_n) \\
&\quad - I_{\Lambda_{1,i}}(M; \mathbf{Y}^{\mathcal{A}^c}|\mathbf{Y}^{\mathcal{A}}, \mathcal{A}, C_n) - H_{\Lambda_{1,i}}(\mathcal{A}) \\
&\overset{(c)}{\geq} I_{\Lambda_{1,i}}(M; Y^n|\mathcal{A}, C_n) - n\alpha \log |\mathcal{Y}| - H_{\Lambda_{1,i}}(\mathcal{A}) \\
&\overset{(d)}{\geq} I_{\Lambda_1}(M; Y^n|C_n) - n\alpha \log |\mathcal{Y}| - 2H_{\Lambda_{1,i}}(\mathcal{A}) \\
&\overset{(e)}{\geq} I_{\Lambda_1}(M; Y^n|C_n) - n\alpha \log |\mathcal{Y}| - 2nh(\alpha), \tag{192}
\end{aligned}
$$

where:

(a) and (d) follow by similar arguments as in the lower bound from (170);

(b) is since since $\tilde{Y}_j = Y_j$ almost surely, for all $j \in \mathcal{A}$ (cf. (178));

(c) is because for every $a \in \mathfrak{A}_i$, we have

$$I_{\Lambda_{1,i}}(M; \mathbf{Y}^{a^c} | \mathbf{Y}^a, \mathcal{A} = a, C_n) \leq |a^c| \log |\mathcal{Y}| \leq n\alpha \log |\mathcal{Y}|;$$

(e) uses (191).

Clearly, (192) yields

$$H_{\Lambda_i}(M | \tilde{Y}^n, C_n) \leq H_{\Lambda_1}(M | Y^n, C_n) + n\alpha \log |\mathcal{Y}| + 2nh(\alpha), \tag{193}$$

for each $1 \neq i \in \mathcal{I}(Q_S, \alpha)$. Inserting this back into (173), while keeping (171) and (175) in mind, for any n sufficiently large (that is independent of Q_S and α), we have

$$H_{\Upsilon}(M | Y^n, C_n)$$
$$\leq \sum_{i \in \mathcal{I}(Q_S, \alpha)} \mathbb{P}_{Q_S^n}\left(S^n \in \mathcal{T}_{Q_i}^n\right) H_{\Lambda_i}(M | Y^n, C_n)$$
$$\quad + 2 \log M_n |\mathcal{S}| e^{-2n\frac{\alpha^2}{|\mathcal{S}|^2}} + |\mathcal{S}| \log(n+1)$$
$$\overset{(a)}{\leq} H_{\Lambda_1}(M | Y^n, C_n) + n\alpha \log |\mathcal{Y}| + 2nh(\alpha)$$
$$\quad + 2 \log M_n |\mathcal{S}| e^{-2n\frac{\alpha^2}{|\mathcal{S}|^2}} + |\mathcal{S}| \log(n+1)$$
$$\overset{(b)}{\leq} n\eta_{n,\alpha}^{(2)}, \tag{194}$$

where (a) uses (193), while (b) follows by (169) and by setting
$\eta_{n,\alpha}^{(2)} = \frac{1}{n} + \frac{1}{n} \log M_n \left(\epsilon + 2|\mathcal{S}| e^{-2n\frac{\alpha^2}{|\mathcal{S}|^2}}\right) + \alpha \log |\mathcal{Y}| + 2h(\alpha) + |\mathcal{S}| \frac{\log(n+1)}{n}$.

APPENDIX G
PROOF OF LEMMAS 10 AND 11

A. Lemma 10

Fix $Q_S \in \mathcal{Q}$, $\gamma \in \Gamma_n$ and $i \in [1:n]$. For any $s_i \in \mathcal{S}$ and $\left(c_n(\gamma), s^{n\backslash i}, m, x^n, y^{n\backslash i}, z^{n\backslash i}\right) \in \mathcal{C}_n \times \mathcal{S}^{n-1} \times \mathcal{M}_n \times \mathcal{X}^n \times \mathcal{Y}^{n-1} \times \mathcal{Z}^{n-1}$, for some $\gamma \in \Gamma_n$, we have

$$\Upsilon\left(c_n(\gamma), s^{n\backslash i}, m, x^n, y^{n\backslash i}, z^{n\backslash i} \big| s_i\right)$$
$$= \frac{\Upsilon\left(c_n(\gamma), s_i, s^{n\backslash i}, m, x^n, y^{n\backslash i}, z^{n\backslash i} \big| s_i\right)}{\Upsilon(s_i)}. \tag{195}$$

The marginal distribution of S_i with respect to Υ from (125a), is

$$\Upsilon(s_i) = Q_S(s_i), \tag{196}$$

while for the numerator, we have

$$\Upsilon\left(c_n(\gamma), s_i, s^{n\backslash i}, m, x^n, y^{n\backslash i}, z^{n\backslash i}\right)$$
$$= \mu_n(\gamma) \Upsilon^{(\gamma)}(s_i) \Upsilon^{(\gamma)}\left(s^{n\backslash i}, m, x^n, y^{n\backslash i}, z^{n\backslash i} \big| s_i\right)$$
$$= \mu_n(\gamma) Q_S(s_i) Q_S^{n-1}\left(s^{n\backslash i}\right) \frac{1}{M_n} f_\gamma(x^n | m)$$
$$\quad \times W_{s^{n\backslash i}}^{n-1}\left(y^{n\backslash i} \big| x^{n\backslash i}\right) V_{s^{n\backslash i}}^{n-1}\left(z^{n\backslash i} \big| x^{n\backslash i}\right)$$
$$= \mu_n(\gamma) Q_S(s_i) \Upsilon^{(\gamma)}\left(s^{n\backslash i}\right) \Upsilon^{(\gamma)}\left(m \big| s^{n\backslash i}\right)$$
$$\quad \times \Upsilon^{(\gamma)}\left(x^n \big| s^{n\backslash i}, m\right) \Upsilon^{(\gamma)}\left(y^{n\backslash i}, z^{n\backslash i} \big| s^{n\backslash i}, m, x^n\right)$$
$$= \Upsilon(s_i) \Upsilon\left(c_n(\gamma), s^{n\backslash i}, m, x^n, y^{n\backslash i}, z^{n\backslash i}\right). \tag{197}$$

Inserting (196) and (197) back into (195) completes the proof.

B. Lemma 11

Again, fix $Q_S \in \mathcal{Q}$ and recall that $V = (V_T, T)$, where $V_T = (Y^{T-1}, S_{T+1}^n, Z_{T+1}^n, C_n)$. Therefore, we represent a realization v of V as $v = (\tilde{v}_t, t)$, where $\tilde{v}_t \triangleq \left(y^{t-1}, s_{t+1}^n, z_{t+1}^n, c_n(\gamma)\right) \in \mathcal{Y}^{t-1} \times \mathcal{S}^{n-t} \times \mathcal{Z}^{n-t} \times \mathcal{C}_n$, for some $\gamma \in \Gamma_n$, and $t \in [1:n]$. For any $(v, u, x, s, y, z) \in \mathcal{V} \times \mathcal{U} \times \mathcal{X} \times \mathcal{S} \times \mathcal{Y} \times \mathcal{Z}$, we have

$$\mathbb{P}_{\Upsilon}(S = s | V = v, U = u, X = x)$$
$$\overset{(a)}{=} \mathbb{P}_{\Upsilon}\left(S_T = s \big| (V_T, T) = (\tilde{v}_t, t), M = m, X_T = x\right)$$
$$\overset{(c)}{=} \mathbb{P}_{\Upsilon^{(\gamma)}}(S_t = s)$$
$$= Q_S(s), \tag{198}$$

where (a) is because $U = (M, V_T, T)$, while (b) uses the independence of T and (M, X^n, S^n, Y^n, Z^n) and the independence relation from Lemma 10.

By similar steps to those in the derivation of (198), we also obtain

$$\mathbb{P}_{\Upsilon}(Y = y, Z = z | V = v, U = u, X = x, S = s)$$
$$\overset{(a)}{=} \mathbb{P}_{\Upsilon^{(\gamma)}}(Y_t = y, Z_t = z | X_t = x, S_t = s)$$
$$= W_s(y | x) V_s(z | x), \tag{199}$$

where (a) also relies on the Markov relation induced by the channel.

REFERENCES

[1] R. Harris *et al.*, "Experimental investigation of an eight-qubit unit cell in a superconducting optimization processor," *Phys. Rev. B*, vol. 82, no. 2, p. 024511, Jul. 2010.

[2] M. W. Johnson *et al.*, "A scalable control system for a superconducting adiabatic quantum optimization processor," *Supercond. Sci. Technol.*, vol. 23, no. 6, p. 065004, Apr. 2010.

[3] A. J. Berkley *et al.*, "A scalable readout system for a superconducting adiabatic quantum optimization system," *Supercond. Sci. Technol.*, vol. 23, no. 10, p. 105014, Sep. 2010.

[4] M. W. Johnson *et al.*, "Quantum annealing with manufactured spins," *Nature*, vol. 473, no. 7346, pp. 194–198, May 2011.

[5] N. Jones. (2013). *Google and NASA Snap Up Quantum Computer D-Wave Two*. [Online]. Available: http://www.scientificamerican.com/article.cfm?id=google-nasa-snap-up-quantum-computer-dwave-two

[6] P. W. Shor, "Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer," *SIAM Rev.*, vol. 41, no. 2, pp. 303–332, Apr. 1999.

[7] D. J. Bernstein, "Introduction to post-quantum cryptography," in *Post-Quantum Cryptography*. Heidelberg, Germany: Springer, 2009, pp. 1–14.

[8] R. A. Perlner and D. A. Cooper, "Quantum resistant public key cryptography: A survey," in *Proc. Symp. Identity Trust Internet (IDtrust)*, Gaithersburg, Maryland, Apr. 2009, pp. 85–93.

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

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

[11] D. Blackwell, L. Breiman, and A. J. Thomasian, "The capacities of certain channel classes under random coding," *Ann. Math. Statist.*, vol. 31, no. 3, pp. 558–567, 1960.

[12] R. Ahlswede, "Elimination of correlation in random codes for arbitrarily varying channels," *Zeitschrift Wahrscheinlichkeitstheorie Verwandte Gebiete*, vol. 44, no. 2, pp. 159–175, 1978.

[13] J. H. Jahn, "Coding of arbitrarily varying multiuser channels," *IEEE Trans. Inf. Theory*, vol. 27, no. 2, pp. 212–226, Mar. 1981.

[14] I. Csiszár and P. Narayan, "Arbitrarily varying channels with constrained inputs and states," *IEEE Trans. Inf. Theory*, vol. 34, no. 1, pp. 27–34, Jan. 1988.

[15] I. Csiszár and P. Narayan, "The capacity of the arbitrarily varying channel revisited: Positivity, constraints," *IEEE Trans. Inf. Theory*, vol. 34, no. 2, pp. 181–193, Mar. 1988.

[16] X. He and A. Yener, "MIMO wiretap channels with unknown and varying eavesdropper channel states," *IEEE Trans. Inf. Theory*, vol. 60, no. 11, pp. 6844–6869, Nov. 2014.

[17] M. J. Mihaljević, "On message protection in cryptosystems modelled as the generalized wire-tap channel II," in *Error Control, Cryptology, and Speech Compression* (Lecture Notes in Computer Science). Berlin, Germany: Springer-Verlag, 1994, pp. 13–24.

[18] Y. Luo, C. Mitrpant, A. J. H. Vinck, and K. Chen, "Some new characters on the wire-tap channel of type II," *IEEE Trans. Inf. Theory*, vol. 51, no. 3, pp. 1222–1229, Mar. 2005.

[19] R. Liu, Y. Liang, H. V. Poor, and P. Spasojevic, "Secure nested codes for type II wiretap channels," in *Proc. IEEE Inf. Theory Workshop (ITW)*, Lake Tahoe, CA, USA, Sep. 2007, pp. 337–342.

[20] V. Aggarwal, L. Lai, A. R. Calderbank, and H. V. Poor, "Wiretap channel type II with an active eavesdropper," in *Proc. Int. Symp. Inf. Theory (ISIT)*, Seoul, South Korea, Jun./Jul. 2009.

[21] E. MolavianJazi, "Secure communication over arbitrarily varying wiretap channels," M.S. thesis, Graduate School Univ. Notre Dame, Notre Dame, IN, USA, Dec. 2009.

[22] J. Nötzel, M. Wiese, and H. Boche, "The arbitrarily varying wiretap channel—Randomness, stability, and super-activation," *IEEE Trans. Inf. Theory*, vol. 62, no. 6, pp. 3504–3531, Jun. 2016.

[23] M. Wiese, J. Nötzel, and H. Boche, "The arbitrarily varying wiretap channel—Communication under uncoordinated attacks," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Hong Kong, Jun. 2015, pp. 2146–2150.

[24] 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, Jul. 2016.

[25] H. Boche, R. F. Schaefer, and H. V. Poor, "On the continuity of the secrecy capacity of compound and arbitrarily varying wiretap channels," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 12, pp. 2531–2546, Dec. 2015.

[26] D. Blackwell, L. Breiman, and A. J. Thomasian, "The capacity of a class of channels," *Ann. Math. Statist.*, vol. 30, no. 4, pp. 1229–1241, Dec. 1959.

[27] J. Wolfowitz, "Simultaneous channels," *Arch. Rational Mech. Anal.*, vol. 4, no. 4, pp. 371–386, 1960.

[28] Y. Liang, G. Kramer, V. H. Poor, and S. Shamai (Shitz), "Compound wiretap channels," *EURASIP J. Wireless Commun. Netw.*, vol. 2009, no. 1, Oct. 2009, Art. no. 142374.

[29] E. Ekrem and S. Ulukus, "On Gaussian MIMO compound wiretap channels," in *Proc. Conf. Inf. Sci. Syst.*, Baltimore, MD, USA, Mar. 2010, pp. 1–6.

[30] A. Khisti, "On the MISO compound wiretap channel," in *Proc. Inf. Theory Appl. Workshop (ITA)*, San Diego, CA, USA, Jan. 2010, pp. 1–7.

[31] A. Khisti, "Interference alignment for the multiantenna compound wiretap channel," *IEEE Trans. Inf. Theory*, vol. 57, no. 5, pp. 2976–2993, May 2014.

[32] I. Bjelaković, H. Boche, and J. Sommerfeld, "Secrecy results for compound wiretap channels," *Problems Inf. Transmiss.*, vol. 49, no. 1, pp. 73–98, Jan. 2013.

[33] R. F. Schaefer and S. Loyka, "The secrecy capacity of a compound MIMO Gaussian channel," in *Proc. IEEE Inf. Theory Workshop (ITW)*, Seville, Spain, Sep. 2013, pp. 104–108.

[34] I. Bjelaković, H. Boche, and J. Sommerfeld, "Capacity results for arbitrarily varying wiretap channels," in *Information Theory, Combinatorics, and Search Theory*. New York, NY, USA: Springer, 2013, pp. 123–144.

[35] C. R. Janda, M. Wiese, J. Nötzel, H. Boche, and E. A. Jorswieck, "Wiretap-channels under constrained active and passive attacks," in *Proc. IEEE Conf. Commun. Netw. Secur. (CNS)*, Jun. 2015, pp. 16–21.

[36] O. Ozel and S. Ulukus, "Wiretap channels: Roles of rate splitting and channel prefixing," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Saint Petersburg, Russia, Jul./Aug. 2011, pp. 628–632.

[37] M. Bellare, S. Tessaro, and A. Vardy, "A cryptographic treatment of the wiretap channel," in *Proc. Adv. Crypto. (CRYPTO)*, Santa Barbara, CA, USA, Jan. 2012. [Online]. Available: http://arxiv.org/abs/1201.2205

[38] S. Goldwasser and S. Micali, "Probabilistic encryption," *J. Comput. Syst. Sci.*, vol. 28, no. 2, pp. 270–299, Apr. 1984.

[39] M. Bellare, A. Desai, E. Jokipii, and P. Rogaway, "A concrete security treatment of symmetric encryption," in *Proc. IEEE 38th Symp. Found. Comput. Sci.*, Miami, FL, USA, Oct. 1997, pp. 394–403.

[40] I. Csiszár, "Almost independence and secrecy capacity," *Problems Inf. Transmiss.*, vol. 32, no. 1, pp. 48–57, Mar. 1996.

[41] I. Devetak and A. Winter, "Classical data compression with quantum side information," *Phys. Rev. A*, vol. 68, no. 4, p. 042301, Oct. 2003.

[42] M. M. Wilde, *Quantum Information Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2013.

[43] 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.

[44] J. Hou and G. Kramer, "Effective secrecy: Reliability, confusion and stealth," in *Proc. IEEE Int. Symp. Inf. Theory*, Honolulu, HI, USA, Jun./Jul. 2014. [Online]. Available: http://arxiv.org/abs/1311.1411

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

[46] I. Csiszár, "The method of types," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2505–2523, Oct. 1998.

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

[48] 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. 1–17, Jul. 2016.

[49] H. G. Eggleston, *Convexity*, 6th ed. Cambridge, U.K.: Cambridge Univ. Press, 1958.

[50] B. Shrader and H. Permuter, "Feedback capacity of the compound channel," *IEEE Trans. Inf. Theory*, vol. 55, no. 8, pp. 3629–3644, Aug. 2009.

[51] M. Nafea and A. Yener, "A new wiretap channel model and its strong secrecy capacity," in *Proc. Int. Symp. Inf. Theory (ISIT)*, Barcelona, Spain, Jul. 2016, pp. 2804–2808.

[52] G. Kramer, "Teaching IT: An identity for the Gelfand–Pinsker converse," *IEEE Inf. Theory Soc. Newslett.*, vol. 61, no. 4, pp. 4–6, Dec. 2011.

[53] D. Galvin. (2014). "Three tutorial lectures on entropy and counting." [Online]. Available: https://arxiv.org/abs/1406.7872

Ziv Goldfeld (S'13) received his B.Sc. (summa cum laude) and M.Sc. (summa cum laude) degrees in Electrical and Computer Engineering from the Ben-Gurion University, Israel, in 2012 and 2014, respectively. He is currently a student in the direct Ph.D. program for honor students in Electrical and Computer Engineering at that same institution.

Between 2003 and 2006, he served in the intelligence corps of the Israeli Defense Forces.

Ziv is a recipient of several awards, among them are the Dean's List Award, the Basor Fellowship, the Lev-Zion fellowship, IEEEI-2014 best student paper award, a Minerva Short-Term Research Grant (MRG), and a Feder Family Award in the national student contest for outstanding research work in the field of communications technology.

Paul Cuff (S'08–M'10) received the B.S. degree in electrical engineering from Brigham Young University, Provo, UT, in 2004 and the M.S. and Ph.D. degrees in electrical engineering from Stanford University in 2006 and 2009. Since 2009 he has been an Assistant Professor of Electrical Engineering at Princeton University.

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

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

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

Prof. Permuter is a recipient of several awards, among them the Fullbright Fellowship, the Stanford Graduate Fellowship (SGF), Allon Fellowship, and the U.S.-Israel Binational Science Foundation Bergmann Memorial Award. Haim is currently serving on the editorial board of the IEEE TRANSACTIONS ON INFORMATION THEORY.