Arbitrarily Varying Wiretap Channels With Non-Causal Side Information at the Jammer

Carsten Rudolf Janda, *Member, IEEE*, Moritz Wiese, *Member, IEEE*, Eduard Axel Jorswieck, *Fellow, IEEE*, and Holger Boche, *Fellow, IEEE*

***Abstract*— Secure communication in a potentially hostile environment is becoming more and more critical. The Arbitrarily Varying Wiretap Channel (AVWC) provides information-theoretical bounds on how much information can be exchanged even in the presence of an active attacker. If the active attacker has non-causal side information, situations in which a legitimate communication system has been hacked can be modeled. We investigate the AVWC with non-causal side information at the jammer for the case that there exists a best channel to the eavesdropper. Non-causal side information means that the transmitted codeword is known to an active adversary before it is transmitted. By considering the maximum error criterion, we also allow messages to be known at the jammer before the corresponding codeword is transmitted. A single-letter formula for the Common Randomness (CR)-assisted secrecy capacity is derived. Additionally, we provide a formula for the CR-assisted secrecy capacity for the cases where the channel to the eavesdropper is strongly degraded, strongly noisier, or strongly less capable with respect to the main channel. Furthermore, we compare our results to the CR-assisted secrecy capacity for the cases of maximum error criterion but without non-causal side information at the jammer (blind adversary), maximum error criterion with non-causal side information of the messages at the jammer (semi-blind adversary), and the case of average error criterion without non-causal side information at the jammer (blind adversary).**

***Index Terms*— Active eavesdroppers, arbitrarily varying wiretap channel, non-causal side information at the jammer, maximum error probability, physical layer secrecy.**

Manuscript received 14 April 2022; revised 27 October 2022; accepted 1 February 2023. Date of publication 15 February 2023; date of current version 17 March 2023. The work of Carsten Rudolf Janda was supported in part by the German Research Foundation [Deutsche Forschungsgemeinschaft (DFG)] through the Project Play Scate under Grant DFG JO 801/21-1. The work of Moritz Wiese was supported in part by the German Federal Ministry of Education and Research [Bundesministerium für Bildung und Forschung (BMBF)] in the program "Souverän. Digital. Vernetzt." within the research hub 6G-life under Grant 16KISK002, in part by the BMBF through the Project "Post Shannon Communication–NewCom" under Grant 16KIS1003K, and in part by the German Research Foundation (DFG) through the Project Play Scate under Grant DFG BO 1734/36-1. The work of Eduard Axel Jorswieck was supported in part by the Federal Ministry of Education and Research (BMBF, Germany) in the program of "Souverän. Digital. Vernetzt." joint project 6G-RIC under Grant 16KISK020K and Grant 16KISK031. The work of Holger Boche was supported in part by the German Federal Ministry of Education and Research (BMBF) in the program "Souverän. Digital. Vernetzt." within the Research Hub 6G-Life under Grant 16KISK002 and in part by the BMBF within the project "Post Shannon Communication–NewCom" under Grant 16KIS1003K. An earlier version of this paper was presented in part at the 2020 IEEE Conference on Communications and Network Security (CNS) [DOI: 10.1109/CNS48642.2020.9162323] and the 2020 IEEE International Symposium on Information Theory (ISIT) [DOI: 10.1109/ISIT44484.2020.9173973]. *(Corresponding author: Carsten Rudolf Janda.)*

Carsten Rudolf Janda and Eduard Axel Jorswieck are with the Department of Information Theory and Communication Systems, Institute of Communications Technology, TU Braunschweig, Braunschweig, 38106 Lower Saxony, Germany (e-mail: c.janda@tu-braunschweig.de; e.jorswieck@tu-braunschweig.de).

Moritz Wiese is with the Chair of Theoretical Information Technology, Munich University of Technology, München, 80333 Bavaria, Germany (e-mail: wiese@tum.de).

Holger Boche is with the Chair of Theoretical Information Technology and the Federal Ministry of Education and Research (BMBF, Germany) Research Hub 6G-Life, Munich University of Technology, München, 80333 Bavaria, Germany, also with the Excellence Cluster Cyber Security in the Age of Large-Scale Adversaries, Ruhr University Bochum, 44801 Bochum, Germany, also with the Munich Center for Quantum Science and Technology (MCQST), 80799 München, Germany, and also with the Munich Quantum Valley (MQV), 80799 München, Germany (e-mail: boche@tum.de).

Communicated by L. Wang, Associate Editor for Shannon Theory and Information Measures.

Digital Object Identifier 10.1109/TIT.2023.3245722

I. Introduction

SECRECY in an adversarial environment is an essential requirement in modern communication systems. It was Wyner [3] who considered secure communications over noisy channels and introduced the **W**ire**t**ap **C**hannel (WTC). Later, his work was extended by [4] to the broadcast channel with confidential messages and in [5] to the Gaussian WTC. In [6], Ozarow and Wyner introduced the wiretap channel of type II [1]. The secrecy metrics in the works above are considered "weak". There exist other secrecy metrics, such as strong secrecy or semantic secrecy. In [7], the authors investigated the ordinary WTC[2] and the so-called WTC of type II. They provided achievable semantic secrecy rates for the WTC and gave a single-letter formula for the semantic secrecy capacity for the WTC of type II. In [8], the authors presented a generalized WTC model. This model consists of a mixture of a WTC and the WTC of type II. During one fraction of the transmission of one codeword, the eavesdropping channel behaves like a WTC. In the remaining time instances, it behaves like a WTC of type II. For this model, [8] contributed a single-letter secrecy capacity formula under the strong secrecy criterion. The previous works combat a passive eavesdropper by cleverly taking the physical properties of the transmission medium into account and coming up with a coding strategy that can guarantee information-theoretic security, confidentiality, and reliable communication simultaneously. By introducing channel states, active adversaries who can arbitrarily modify

[1] Essentially, the eavesdropper is able to perfectly receive a fraction of the transmitted codeword. In contrast to a "random" erasure channel, here the eavesdropper can choose the exact symbols he wants to obtain.

[2] Throughout the paper, we use the term WTC when we mean the ordinary WTC.

the channel state can be modeled by the **A**rbitrarily **V**arying **C**hannel (AVC). For the AVC, different code concepts are introduced in [9]. In [10], the discussion of [9] is extended for different error criteria. It can be shown that the CR-assisted[3] capacity of an AVC under the average error criterion equals its CR-assisted capacity under the maximum error criterion. The author also stated a dichotomous result (Ahlswede's Dichotomy) for the deterministic code capacity under the average error criterion. The latter equals its CR-assisted capacity or equals zero. Even though [10] provides a necessary and sufficient condition for the deterministic code capacity under the maximum error criterion to be positive, the question of the exact formula remains an open problem.

In [11], CR-assisted codes for the AVC with a limited amount of CR are studied. The author limited the amount of CR to increase only exponentially with respect to the block length. Furthermore, an exponential error bound is considered. Additionally, the author provided a sufficient condition for when the deterministic code capacity is zero. This condition is called symmetrizability. The author proved that if the symmetrizability condition is fulfilled, the (average) error probability is bounded away from zero and is lower bounded by $\frac{1}{4}$.

In the literature, several cases of side information at the transmitter and/or the jammer have been considered. In [12], an AVC is considered, where the jammer has non-causal access to the channel input and the message. Since the message is known non-causally at the jammer, the considered error probability has to be the maximum error probability. The authors used a list code under the maximum error criterion approach to prove the CR-assisted capacity for this model. In [13], the authors investigated the AVC with non-causal side information at the jammer, i.e., the jammer has non-causal access to the transmitted codeword and the message. In contrast to [12] and [13] used random coding arguments instead of list codes. Furthermore, the authors imposed peak input and state constraints and derived the CR-assisted code capacity under the average and the maximum error criteria, and compared these results. They limited the amount of CR that is needed to achieve the capacity and stated that non-causal knowledge of the channel input at the jammer leads to lower capacity than non-causal knowledge of the messages. In [14], the situation of "nosy noise" where the channel input is perfectly known at the jammer [12], is generalized to a "myopic adversary", where a jammer has a noisy version of the channel input as side information. Furthermore, a CR-assisted capacity formula under the maximum error criterion is derived. In [15], a version of the AVC is considered, where the jammer and the transmitter have non-causal knowledge about the messages and the channel state (here, not controlled by the jammer). Based on this knowledge, the jammer can adopt its jamming signal. Simultaneously, the transmitter uses Gel'fand Pinsker or dirty paper coding to optimize the CR-assisted capacity under the maximum error criterion. For the dirty paper AVC, it was shown that a memoryless Gaussian jamming strategy is the jammer's optimal choice. In [16], an **A**rbitrarily **V**arying **C**lassical-**Q**uantum **C**hannel (AVCQC) is investigated, where the jammer has side information about the channel input or both the channel input and the message. The authors determined the CR-assisted capacity for both average and maximum error criteria and established a strong converse. Furthermore, all derived capacities are equal. The additional knowledge of the message does not decrease the capacity further.

Various works have considered input and state constraints. Since we do not consider constraints in our work, we only mention one fundamental result. In [17], the AVC with peak constraints is considered. The authors introduced a "cost"-function and have shown that if the jammer cannot symmetrize the channel because of his state peak constraint, the deterministic code capacity might be positive but less than the CR-assisted capacity. Furthermore, the authors proved that the symmetrizability condition from [11] is not only sufficient but also necessary for the deterministic code capacity of an AVC to be zero.

If confidentiality requirements are combined with active attacks on communication systems, the AVWC is the correct channel model. In the case where the channel state is determined by nature and there are secrecy requirements, the **C**ompound **W**iretap **C**hannel (CWC) is an appropriate model. In the following, we give a very brief literature overview of the CWC and the AVWC. In [18], CR-assisted codes for the AVWC are considered. The authors presented a single-letter formula for achievable CR-assisted secrecy rates. Furthermore, the authors provided a single-letter formula for the CR-assisted secrecy capacity for the strongly degraded case with independent states. In [19], the AVWC under the average error criterion is investigated. The authors combined strong secrecy requirements with Ahlswede's **E**limination **T**echnique (ET) and were able to derive a single-letter formula for the CR-assisted achievable secrecy rates. Additionally, the authors presented a multi-letter formula for the deterministic code secrecy capacity. In [20], continuity properties of the secrecy capacities of CWCs and AVWCs are studied. The authors showed that for the CWC, the secrecy capacity is continuous with respect to the channel states. In contrast to the compound case, the authors proved that the deterministic code secrecy capacity of an AVWC possesses discontinuity properties with respect to the channel state. The authors presented an example in which the deterministic code secrecy capacity is continuous for almost all convex combinations of channel states (for all $\lambda \in (0, 1]$, where λ is the linear coefficient of an AVC with two elements, parameterizing the convex combination of channel matrices). Furthermore, the authors derived a zero deterministic code secrecy capacity for $\lambda = 0$, while calculating $\lim_{\lambda \searrow 0}$, the deterministic code secrecy capacity remains strictly lager than zero. In [21], the AVWC is investigated, and multi-letter formulas for the CR-assisted and deterministic code secrecy capacities for the case that the eavesdropper is ignorant about the CR are derived. The authors proved that even though the deterministic code secrecy capacity possesses discontinuities, it is still stable around its positivity points. Furthermore, the authors provided a complete

[3]In the literature, the terms CR-assisted and random are used, e.g., in [10] the terms random code and random code capacity are used, while we use the terms CR-assisted code and CR-assisted capacity

characterization of AVWCs, which possess the super-activation property. In [22], a multi-letter formula for the CR-assisted secrecy capacity in the general case and a single-letter formula for the CR-assisted secrecy capacity in the strongly degraded case are proved. The authors considered both average and maximum error criteria and showed that the capacities are equivalent under both criteria.

In the literature, several cases of side information at the transmitter and/or the jammer have been considered with secrecy constraints. In [23], the binary WTC of type II with an active eavesdropper, who observes a fraction of the transmitted codeword causally, is considered. The authors specifically investigated the cases, where the eavesdropper erases his observed symbols, and where the eavesdropper flips his observed symbols. For these models, achievable secrecy rates are proved. In [24], an AVWC, where the active adversary has access to the CR, is studied. This work relates the dichotomy behavior of the deterministic code capacity of AVC to the case with secrecy requirements. The authors showed that the if the AVWC is symmetrizable, then the CR secrecy capacity of the AVWC with knowledge of the common randomness at the active adversary equals zero. Otherwise, it equals the CR secrecy capacity of the AVWC. In [25], the deterministic list code secrecy capacity of an AVWC is investigated. The authors provided a multi-letter formula and presented a symmetrizability condition on the list size for the secrecy capacity to be zero. In [26], a WTC with non-causal **C**hannel **S**tate **I**nformation (CSI)T is investigated. Under the maximum error and semantic security criteria, a single-letter formula for the achievable secrecy rate is derived. In [27], the AVWC with input and state peak constraints is investigated. The authors derived a multi-letter formula for the achievable secrecy rate. In [28], the author scrutinized a variation of the AVWC, in which an adversary receives a fraction of the codeword perfectly (in terms of WTC of type II) and modifies another fraction of the codeword, where the adversary can use his observed side information. The author determined upper and lower bounds on the semantic secrecy capacity. In [29], the authors used a strong soft covering lemma to derive a single-letter formula of the CR-assisted semantic secrecy capacity of an AVWC with type-constrained states. In [30], deterministic wiretap-codes for the AVWC with input and state peak constraints are considered. The authors provided a single-letter formula for achievable secrecy rates.

A. Contribution

This work considers the AVWC with non-causal side information at the jammer. Non-causal side information means that codewords are known at an active adversary before they are transmitted. We provide the single-letter CR-assisted secrecy capacity under the maximum error criterion for the case that there exists a best channel to the eavesdropper. By considering the maximum error criterion, we also allow the active attacker to know the messages. We use methods of [16], hence random coding arguments instead of list codes [12], which might be an alternative approach. When considering WTCs, the secrecy capacity formula depends in general on auxiliary **R**andom **V**ariables (RVs). One of these RVs can be interpreted as adding a "prefix" channel as part of the encoding process. We derive a CR-assisted secrecy capacity formula for the cases in which the eavesdropping channel is strongly degraded, strongly noisier, or strongly less capable with respect to the main channel, that does not depend on such a prefix auxiliary RV. Furthermore, compare our results to the CR-assisted secrecy capacity for the cases of maximum error criterion but without any side information at the jammer (blind adversary), maximum error criterion with non-causal side information of the messages at the jammer (semi-blind adversary), and the case of average error criterion without any side information at the jammer (blind adversary). By considering this model, we can describe situations in which a communication system is subject to two simultaneous attacks, eavesdropping and jamming attacks. For both, we individually assume worst case scenarios. By requiring a best channel to the eavesdropper, we also consider the case of colluding jammer and eavesdropper. Even though the jammer obtains a perfect version of the channel input (and also has knowledge about the messages), we can show by our secrecy analysis that the eavesdropper does not obtain any information about the messages and can also not be informed about the channel input by the jammer, using corresponding state sequences. The complete link between the messages and the eavesdropper is secured. The eavesdropper obtains a perfect observation of the CR shared between the legitimate communication partners. Hence, the CR cannot be used as a key to encrypt the data. In Table I, we set our work into context. For this overview, we only considered state dependent channels with secrecy requirements, whose states are influenced by an external entity. But keep in mind that there are publications without secrecy requirements, which are still highly related to this work, i.e, [12], [13], [14]. Since our work does not include constraints on the input or states, we excluded those works from the table, as well.

The paper is organized as follows. We present the system model in Section II and state our main result in Section III. Finally, in Section IV, we compare our results to the standard AVWC, provide an example, and close with a discussion. The proofs of the main results can be found in the appendices.

Notation: We follow the notation of [22], and a list of the used symbols and their meanings can be found in Appendix H. In particular, all logarithms are taken to base 2. Equivalently, the $\exp\{.\}$ function means $2^{\{.\}}$. Sets are denoted by calligraphic letters. The cardinality of a set $\mathcal{U}$ is denoted by $|\mathcal{U}|$. The set of all probability measures on a set $\mathcal{U}$ is denoted by $\mathcal{P}(\mathcal{U})$. For $p \in \mathcal{P}(\mathcal{U})$ we define $p^n \in \mathcal{P}(\mathcal{U}^n)$ as $p^n(x^n) = \prod_i p(x_i)$. The entropies and mutual information terms will be written in terms of the involved probability functions or in terms of the involved random variables. For example

$$H(W|p) := -\sum_{x,y} p(x)W(y|x)\log W(y|x)$$
$$I(p;W) := H(pW) - H(W|p).$$

Furthermore, let the type of a sequence $s^n = (s_1, s_2, \ldots, s_n)$ be the probability measure $q \in \mathcal{P}(\mathcal{S})$ defined by

TABLE I

LITERATURE OVERVIEW RELATED TO THE PRESENTED MANUSCRIPT (WITHOUT CONSTRAINTS AND WITH SECRECY REQUIREMENTS). NOTATION: SIDE INFORMATION - D-CSI (DIFFERENT CSI CASES AT THE TRANSMITTER AND RECEIVER), MII/CII/MCII (MESSAGE / CHANNEL INPUT / MESSAGE AND CHANNEL INPUT NON-CAUSALLY KNOWN AT THE JAMMER), PCI (A FRACTION OF THE CHANNEL INPUT CAUSALLY KNOWN AT THE JAMMER). ERROR - A (AVERAGE ERROR CRITERION), M (MAXIMUM ERROR CRITERION). RESULT - SL (SINGLE-LETTER), ML (MULTI-LETTER), AR (ACHIEVABLE RATE), C (CAPACITY), R (RANDOMNESS ASSISTED), D (DETERMINISTIC), SD (STRONGLY DEGRADED)

Reference	CWC/AVWC	Side Information	Error	Result
[18]	AVWC	-	a	sl - r AR, r sd C
[23]	BAC, type II	CI	a	sl - d AR
[19]	AVWC	-	a	sl - r AR, ml - d C
[31]	CWC	d-CSI	a	ml - C, sl - C for special cases
[22]	AVWC	-	a/m	ml - r C, sl - r sd C
[28]	BAC, type I/II	CI	a	sl - d AR
[32]	AVWC	CSIT	a	r C
[16]	CQAVC	MII/CII/MCII	a/m	sl C
[26]	AVWC	non-causal CSIT	m	sl - r AR, C (sp. cases)
[33]	AVWC, MAC	-	a	sl - r AR, r C (sp. cases)
This work	AVWC	MII/CII/MCII	a/m	sl - r C

$q(a) = \frac{1}{n}N(a|s^n)$, where $N(a|s^n)$ denotes the number of occurrences of a in the sequence s^n. The set of all possible types of sequences of length n is denoted by $\mathcal{P}_0^n(\mathcal{S})$.[4] Additionally, for a $p \in \mathcal{P}(\mathcal{X})$ and $\delta > 0$, we define the typical set $\mathcal{T}_{p,\delta}^n \subset \mathcal{X}^n$ as the set of sequences $x^n \in \mathcal{X}^n$ satisfying for all $a \in \mathcal{X}$ the conditions

$$\left|\frac{1}{n}N(a|x^n) - p(a)\right| \leq \delta, \quad \text{if } p(a) > 0,$$
$$\text{and } N(a|x^n) = 0 \quad \text{if } p(a) = 0.$$

Similarly, for a $W \in \mathcal{P}(\mathcal{Y}|\mathcal{X})$ and a $\delta > 0$ we define the conditionally typical set $\mathcal{T}_{W,\delta}^n(x^n) \subset \mathcal{Y}^n$ as the set of sequences $y^n \in \mathcal{Y}^n$ satisfying for all $a \in \mathcal{X}$, $b \in \mathcal{Y}$ the conditions

$$\left|\frac{1}{n}N(a,b|x^n,y^n) - W(b|a)\frac{1}{n}N(a|x^n)\right| \leq \delta, \quad \text{if } W(b|a) > 0,$$
$$N(a,b|x^n,y^n) = 0 \quad \text{if } W(b|a) = 0.$$

See also [34, Chapter 2] for the method of types and the definitions of typical sequences.

II. SYSTEM MODEL

We consider a CR-assisted AVWC as depicted in Fig. 1. A transmitter Alice tries to communicate reliably and securely with a legitimate receiver Bob in the presence of an eavesdropper Eve. The communication is done via state dependent **D**iscrete **M**emoryless **C**hannels (DMCs) $W^n(y^n|x^n,s^n)$ and $V^n(z^n|x^n,s^n)$, where s^n is the channel state, x^n is the channel input, and y^n and z^n are the received sequences at Bob and Eve, respectively. Alice, Bob, and Eve have access to a common source of randomness $\mathcal{U}_n$, whose realization can not be used as a key for encryption, since Eve also has access to it. The channel state s^n is controlled by an external jammer Jim, who has non-causal access to the channel input X_u^n. The channel input of length n is dependent on the the CR realization, and hence indexed by it. Note that this system model is considered without secrecy constraints by Sarwate [12], using a connection between deterministic list codes and CR-assisted codes. Furthermore, this system model also is considered without secrecy constraints for the classical-quantum case by Boche et al. [16]. In the latter case, the authors use random coding arguments.

Remark 1: By requiring a best channel to the eavesdropper, we can show that the jammer is not able to encode information about the channel input into the choice of the state sequence. Hence, if there is no other channel between the jammer and the eavesdropper, we also cover the situation of colluding attackers.

*Definition 1 (**A**rbitrarily **V**arying **W**iretap **C**hannel):* We describe an **Arbitrarily Varying Wiretap Channel** by $(\mathcal{X},\mathcal{S},\mathcal{W},\mathcal{V},\mathcal{Y},\mathcal{Z})$. Let $\mathcal{X},\mathcal{S},\mathcal{Y},\mathcal{Z}$ be finite sets. The family of channels to the legitimate receiver is described by $\mathcal{W} = \{(W_s : \mathcal{X} \to \mathcal{P}(\mathcal{Y})) : \ s \in \mathcal{S}\}$. The family of channels to the illegitimate receiver is described by $\mathcal{V} = \{(V_s : \mathcal{X} \to \mathcal{P}(\mathcal{Z})) : \ s \in \mathcal{S}\}$. The channel is memoryless in the sense that the probability of receiving the sequences $y^n = (y_1, y_2, \ldots, y_n)$ and $z^n = (z_1, z_2, \ldots, z_n)$, when sending $x^n = (x_1, x_2, \ldots, x_n)$ is

$$\begin{aligned} W^n(y^n|x^n,s^n) &= \prod_{i=1}^n W(y_i|x_i,s_i) = \prod_{i=1}^n W_{s_i}(y_i|x_i) \\ &= W_{s^n}^n(y^n|x^n), \\ V^n(z^n|x^n,s^n) &= \prod_{i=1}^n V(z_i|x_i,s_i) = \prod_{i=1}^n V_{s_i}(z_i|x_i) \\ &= V_{s^n}^n(z^n|x^n). \end{aligned}$$

By $(\mathcal{W},\mathcal{V})$, we mean the AVWC defined above.

[4] When emphasizing that p is a *single-letter* distribution defined by the empirical distribution of sequences of length n, we equivalently use the notation $p \in \mathcal{P}_0(\mathcal{S}^n)$. However, note that $\tilde{p} \in \mathcal{P}(\mathcal{S}^n)$ is a multi-letter distribution on $\mathcal{S}^n$.

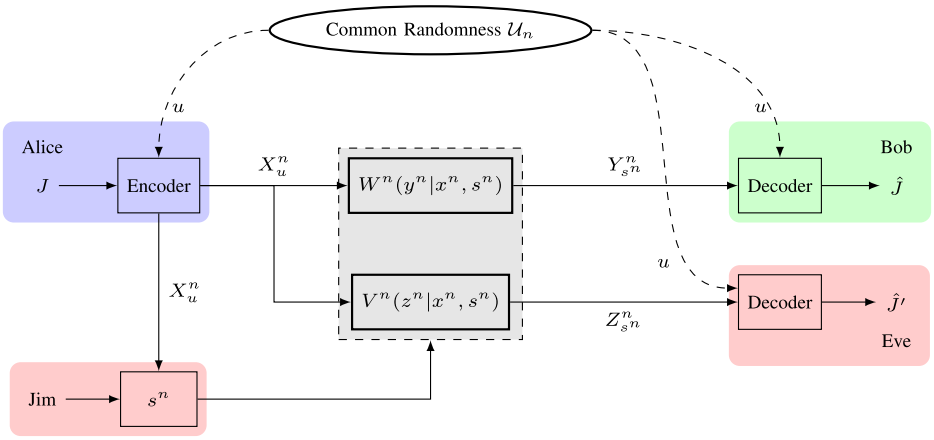

Fig. 1. System model. Jammer has non-causal knowledge about the channel input.

Definition 2 (Deterministic Wiretap-Code): An (n, J_n) **deterministic wiretap-code** $\mathcal{K}_n$ consists of a stochastic encoder $E : \mathcal{J}_n \to \mathcal{P}(\mathcal{X}^n)$ and mutually disjoint decoding sets $\mathcal{D}_j \subset \mathcal{Y}^n$, $\mathcal{D}_j \cap \mathcal{D}_{j'} = \emptyset$, $j, j' \in \mathcal{J}_n$. We define $EW_{s^n}^n : \mathcal{J}_n \to \mathcal{P}(\mathcal{Y}^n)$ by

$$EW_{s^n}^n(y^n|j) = \sum_{x^n \in \mathcal{X}^n} E(x^n|j) W^n(y^n|x^n, s^n).$$

The maximum error $e(\mathcal{K}_n)$ for the AVWC can be expressed as

$$e(\mathcal{K}_n) := \max_{s^n \in \mathcal{S}^n} \max_{j \in \mathcal{J}_n} \sum_{x^n \in \mathcal{X}^n} E(x^n|j) W^n(\mathcal{D}_j^c|x^n, s^n)$$

If the jammer has non-causal knowledge about the channel input x^n, then the maximum error probability has to be expressed as

$$\hat{e}(\mathcal{K}_n) := \max_{f \in \mathcal{F}} \max_{j \in \mathcal{J}_n} \sum_{x^n \in \mathcal{X}^n} E(x^n|j) W^n(\mathcal{D}_j^c|x^n, f(x^n)),$$

for all deterministic jamming functions $\mathcal{F} : \mathcal{X}^n \to \mathcal{S}^n$. We define $EV_{s^n}^n : \mathcal{J}_n \to \mathcal{P}(\mathcal{Z}^n)$ and $EV_f^n : \mathcal{J}_n \to \mathcal{P}(\mathcal{Z}^n)$ by

$$EV_{s^n}^n(z^n|j) = \sum_{x^n \in \mathcal{X}^n} E(x^n|j) V^n(z^n|x^n, s^n),$$
$$EV_f^n(z^n|j) = \sum_{x^n \in \mathcal{X}^n} E(x^n|j) V^n(z^n|x^n, f(x^n)).$$

Definition 3 (Common Randomness Assisted Wiretap-Code): An $(n, J_n, \mathcal{U}_n, p_U)$ **CR-assisted wiretap-code** $\mathcal{K}_n^{\text{ran}}$ consists of a family of stochastic encoders $\mathcal{E} = \{(E_u : \mathcal{J}_n \to \mathcal{P}(\mathcal{X}^n)) : u \in \mathcal{U}_n\}$ and mutually disjoint (for fixed u) decoding sets $\mathcal{D}_{j,u} \subset \mathcal{Y}^n$, $\mathcal{D}_{j,u} \cap \mathcal{D}_{j',u} \neq \emptyset$, $j, j' \in \mathcal{J}_n$, $u \in \mathcal{U}_n$ with message set $\mathcal{J}_n := \{1, \ldots, J_n\}$, and $p_U \in \mathcal{P}(\mathcal{U}_n)$. Note that for different realizations of the CR $\mathcal{U}_n$, $u \neq u'$, the decoding sets do not have to be disjoint, $\mathcal{D}_{j,u} \cap \mathcal{D}_{j',u'} \neq \emptyset$. The maximum error probability averaged over all possible randomly chosen deterministic wiretap-codebooks $e(\mathcal{K}_n^{\text{ran}})$ can be written as

$$e(\mathcal{K}_n^{\text{ran}}) := \max_{s^n \in \mathcal{S}^n} \max_{j \in \mathcal{J}_n} \sum_{u \in \mathcal{U}_n} p_U(u) \sum_{x^n \in \mathcal{X}^n} E_u(x^n|j) W^n(\mathcal{D}_{j,u}^c|x^n, s^n).$$

Here, the jammer does not know the channel input non-causally.

We define the channel $p_{X^n U|J} : \mathcal{J}_n \to \mathcal{P}(\mathcal{X}^n \times \mathcal{U})$ as

$$p_{X^n U|J}(x^n, u|j) = p_{X^n|JU}(x^n|j, u) p_U(u) = E_u(x^n|j) p_U(u).$$

Let $\mathcal{F} : \mathcal{X}^n \to \mathcal{S}^n$ describe the family of all deterministic mappings from $\mathcal{X}^n$ to $\mathcal{S}^n$. If the jammer has non-causal knowledge of the channel input x^n, then the maximum error probability has to be adapted to

$$\hat{e}(\mathcal{K}_n^{\text{ran}}) := \max_{f \in \mathcal{F}} \max_{j \in \mathcal{J}_n} \sum_{x^n \in \mathcal{X}^n} p_{X^n|J}(x^n|j) \sum_{u \in \mathcal{U}_n} p_{U|X^n,J}(u|x^n, j) W^n(\mathcal{D}_{j,u}^c|x^n, f(x^n)).$$

Remark 2: In contrast to the standard AVWC, here in the case of non-causal knowledge at the jammer the worst case choice of s^n is done within each term of the sum. Since the jammer knows the channel input, he can adopt to that specific codeword choice. Furthermore, let $\mathcal{F}'$ be the family of all deterministic mappings $\mathcal{J}_n \times \mathcal{X}^n \to \mathcal{S}^n$, and $\mathcal{F}''$ be the family of all deterministic mappings $\mathcal{J}_n \to \mathcal{S}^n$. From Lemma 2 below, we have

$$\begin{aligned} e(\mathcal{K}_n) &= \max_{s^n \in \mathcal{S}^n} \max_{j \in \mathcal{J}_n} \sum_{x^n \in \mathcal{X}^n} E(x^n|j) W^n(\mathcal{D}_j^c|x^n, s^n) \\ &= \max_{j \in \mathcal{J}_n} \max_{f'' \in \mathcal{F}''} \sum_{x^n \in \mathcal{X}^n} E(x^n|j) W^n(\mathcal{D}_j^c|x^n, f''(j)), \text{ and} \\ \hat{e}(\mathcal{K}_n) &= \max_{f \in \mathcal{F}} \max_{j \in \mathcal{J}_n} \sum_{x^n \in \mathcal{X}^n} E(x^n|j) W^n(\mathcal{D}_j^c|x^n, f(x^n)) \\ &= \max_{j \in \mathcal{J}_n} \max_{f' \in \mathcal{F}'} \sum_{x^n \in \mathcal{X}^n} E(x^n|j) W^n(\mathcal{D}_j^c|x^n, f'(x^n, j)). \end{aligned}$$

That implies the following statement. Considering the maximum error probability (with respect to the messages) corresponds to the case, where the jammer additionally knows the messages, because the order of maximization can be exchanged according to Lemma 2 (see also [13]). Furthermore, the inner optimization is done for fixed parameter of the outer optimization. That means for each given message $j \in \mathcal{J}_n$, the worst case state sequence will be considered. This implies the above equalities. Equivalent statements hold for the CR-assisted codes.

Definition 4 (Achievable Common Randomness Assisted Secrecy Rates and Common Randomness Assisted Secrecy Capacities): A nonnegative number R_S is called an **achievable CR-assisted secrecy rate** for the AVWC if there exists a sequence $(\mathcal{K}_n^{\text{ran}})_{n=1}^{\infty}$ of $(n, J_n, \mathcal{U}_n, p_U)$ CR-assisted codes for uniformly distributed messages, such that the following requirements are fulfilled

$$\liminf_{n\to\infty} \frac{1}{n} \log J_n \geq R_S, \tag{1}$$

$$\lim_{n\to\infty} e(\mathcal{K}_n^{\text{ran}}) = 0, \tag{2}$$

$$\lim_{n\to\infty} \max_{s^n\in\mathcal{S}^n} \max_{u\in\mathcal{U}_n} I(p_J; E_u V_{s^n}^n) = 0. \tag{3}$$

A nonnegative number $\widehat{\widehat{R}}_S$ **is called an achievable CR-assisted secrecy rate for the AVWC with non-causal knowledge of the channel input at the jammer** if there exists a sequence $(\mathcal{K}_n^{\text{ran}})_{n=1}^{\infty}$ of $(n, J_n, \mathcal{U}_n, p_U)$ CR-assisted codes for uniformly distributed messages, such that the following requirements are fulfilled

$$\liminf_{n\to\infty} \frac{1}{n} \log J_n \geq \widehat{\widehat{R}}_S, \tag{4}$$

$$\lim_{n\to\infty} \hat{e}(\mathcal{K}_n^{\text{ran}}) = 0, \tag{5}$$

$$\lim_{n\to\infty} \max_{f\in\mathcal{F}} \max_{u\in\mathcal{U}_n} I(p_J; E_u V_f^n) = 0. \tag{6}$$

The supremum of all achievable CR-assisted secrecy rates for the AVWC is called the **CR-assisted secrecy capacity** of the AVWC $(\mathcal{W}, \mathcal{V})$ and is denoted by $\widehat{C}_S^{\text{ran}}(\mathcal{W}, \mathcal{V})$, when the jammer has no knowledge about the channel input, and $\widehat{\widehat{C}}_S^{\text{ran}}(\mathcal{W}, \mathcal{V})$ if the jammer has non-causal knowledge of the channel input.

The secrecy capacity $\widehat{C}_S^{\text{ran}}(\mathcal{W}, \mathcal{V})$ is lower bounded by $\widehat{\widehat{C}}_S^{\text{ran}}(\mathcal{W}, \mathcal{V})$. Note that the eavesdropper has access to the CR, too. Hence, the randomness cannot be used as a key ensuring secure communication between Alice and Bob. We explicitly do not bound the cardinality of the CR. In [16], the authors provide capacity formulas for quantum channels with an informed jammer but without secrecy constraints. The authors additionally relate and compare the capacity formulas for the cases that the jammer knows the messages additionally and that the jammer does not know the messages.

Lemma 1: Let $\mathcal{P}(\mathcal{S}^n|\mathcal{X}^n)$ be the set of all conditional probability distributions of the state sequences $s^n \in \mathcal{S}^n$ given the channel input $x^n \in \mathcal{X}^n$.

We can, in fact, equivalently consider the maximization over $\theta \in \mathcal{P}(\mathcal{S}^n|\mathcal{X}^n)$ or the maximization over all deterministic mappings $\mathcal{F} : \mathcal{X}^n \to \mathcal{S}^n$.

Proof of Lemma 1: See Appendix E. ∎

Remark 3: Since the mutual information is a convex (row convex) function with respect to the conditional probability function of the output given the input for fixed input distribution, the optimal jamming strategy with respect to the secrecy constraint is deterministic.

$$\theta^{*,n}(s^n|x^n) = \mathbb{1}_{s^{*,n}}(x^n)$$

In other words, the optimal state sequence (in terms of the secrecy constraint) results in a boundary point of $\widetilde{\mathcal{V}}^n$. Moreover, taking convex combinations of channel states does not increase the mutual information. The reason for the equivalence of the consideration of stochastic and deterministic jamming strategies with respect to the reliability criterion is as follows. Although the considered space is larger when allowing stochastic mappings instead of only deterministic mappings to the state space, the error probability is upper bounded by the worst case state sequence.

Definition 5 (Convex Closure and Row Convex Closure [10]): Let $p \in \mathcal{P}(\mathcal{S})$ and $\hat{p} \in \mathcal{P}(\mathcal{S}|\mathcal{X})$ be probability measures. The **convex closure** and the **row convex closure** of the AVC are defined as

$$\widehat{\mathcal{W}} := \left\{ W_p(\cdot|\cdot) : \sum_{s\in\mathcal{S}} p(s) W(\cdot|\cdot, s), \quad p \in \mathcal{P}(\mathcal{S}) \right\} \tag{7}$$

$$\widehat{\widehat{\mathcal{W}}} := \left\{ W_{\hat{p}}(\cdot|x) : \sum_{s\in\mathcal{S}} \hat{p}(s|x) W(\cdot|x, s), \right. \\ \left. \hat{p}(s|x) \in \mathcal{P}(\mathcal{S}|\mathcal{X}), x \in \mathcal{X}, \right\} \tag{8}$$

Example 1: Let $\mathcal{X} = \mathcal{Y} = \mathcal{S} = \{0, 1\}$, and

$$W(\cdot|\cdot, S = 0) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad W(\cdot|\cdot, S = 1) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

The convex closure and the row convex closure are given respectively as

$$\widehat{\mathcal{W}} = \left\{ W(\cdot|\cdot) : \begin{pmatrix} \alpha & 1-\alpha \\ 1-\alpha & \alpha \end{pmatrix}, \quad \alpha \in [0, 1] \right\},$$

$$\widehat{\widehat{\mathcal{W}}} = \left\{ W(\cdot|\cdot) : \begin{pmatrix} \alpha & 1-\alpha \\ 1-\beta & \beta \end{pmatrix}, \quad \alpha, \beta \in [0, 1] \right\}.$$

Remark 4 (Notation): With slight abuse of notation, we use the subscripts of V and W to show the dependence on the state sequence s^n, the deterministic mapping $f \in \mathcal{F}$, $\mathcal{F} : \mathcal{X}^n \to \mathcal{S}^n$ and stochastic mappings $\theta \in \mathcal{P}(\mathcal{S}^n|\mathcal{X}^n)$. Since we use certain notations interchangeably, we clarify them in the following (shown for V). For the AVC $\mathcal{V}$ to the eavesdropper with channel input x^n, and channel state s^n, the probability of obtaining the channel output z^n is denoted by $V^n(z^n|x^n, s^n)$ or interchangeably by $V_{s^n}^n(z^n|x^n)$. If the jammer has non-causal access to the channel input, he can apply the deterministic mapping (or strategy) $f \in \mathcal{F}$, $\mathcal{F} : \mathcal{X}^n \to \mathcal{S}^n$. In that case, we write $V^n(z^n|x^n, f(x^n))$, or equivalently $V_f^n(z^n|x^n)$. If the jammer uses a stochastic mapping (strategy) $\theta \in \mathcal{P}(\mathcal{S}^n|\mathcal{X}^n)$ instead of a deterministic mapping (strategy), we write $V_\theta^n(z^n|x^n)$, with $V_\theta^n(z^n|x^n) = \sum_{s^n\in\mathcal{S}^n} \theta(s^n|x^n) V^n(z^n|x^n, s^n)$. Hence, we consider the averaged channel with respect to the channel state s^n in dependence on the channel input x^n. We denote the conditional probability of obtaining the output sequence z^n under the conditions that we transmitted the secure message $j \in \mathcal{J}_n$ and that the jammer applies the deterministic jamming strategy $f \in \mathcal{F}$ as $V_{f,u}^n(z^n|j)$, with $V_{f,u}^n(z^n|j) = E_u V_f^n(z^n|j) = \sum_{x^n\in\mathcal{X}^n} E_u(x^n|j) V^n(z^n|x^n, f(x^n))$. Since we use the stochastic encoder E_u, we average with respect to the channel input $x^n \in \mathcal{X}^n$. Correspondingly, if the jammer applies a stochastic jamming strategy $\theta \in \mathcal{P}(\mathcal{S}^n|\mathcal{X}^n)$ instead of a deterministic mapping, we write $V_{\theta,u}^n(z^n|j) = E_u V_\theta^n(z^n|j) =$

$\sum_{x^n\in\mathcal{X}^n} E_u(x^n|j)\sum_{s^n\in\mathcal{S}^n}\theta(s^n|x^n)V^n(z^n|x^n,s^n)$. Since we use again a stochastic encoder E_u, we average with respect to the channel input x^n and with respect to the channel states s^n.

In the following, we define a best channel to the eavesdropper. Throughout this work, we assume that such a best channel to the eavesdropper exists.

Definition 6 (Best Channel to the Eavesdropper): If there exists a $\theta^*\in\mathcal{P}(\mathcal{S}|\mathcal{X})$, such that for all $n\in\mathbb{N}$ the Markov chain

$$X^n \leftrightarrow Z^n_{\theta^{*,n}} \leftrightarrow Z^n_\theta, \quad \theta\in\mathcal{P}(\mathcal{S}^n|\mathcal{X}^n) \tag{9}$$

holds with $\theta^{*,n}(s^n|x^n)=\prod_{i=1}^n\theta_i^*(s_i|x_i)=\prod_{i=1}^n\theta^*(s_i|x_i)$, where Z^n_θ is the output of the channel V^n_θ, then we say that there exists a **best channel to the eavesdropper** and all channels V^n_θ are degraded with respect to the channel $V^n_{\theta^{*,n}}$.

Remark 5: In the case of a (semi-) blind adversary, i.e., if the channel input is not known non-causally at the jammer, it is sufficient to require a single-letter condition to hold. Hence, there exists a best channel to the eavesdropper if there exists a $\theta^*\in\mathcal{P}(\mathcal{S})$ such that for all other $\theta\in\mathcal{P}(\mathcal{S})$ the Markov chain

$$X \leftrightarrow Z_{\theta^*} \leftrightarrow Z_\theta, \tag{10}$$

holds.

Next, we will introduce the notions of strongly degraded, strongly noisier, and strongly less capable with independent states, respectively. Independent states mean that the states in the main and the eavesdropping channel can be chosen individually. If the AVC $\mathcal{W}$ from the transmitter to the legitimate receiver and the AVC $\mathcal{V}$ from the transmitter to the eavesdropper fulfill one of the before-mentioned criteria, the CR-assisted secrecy capacity formula can be simplified compared to the case that none of the above-mentioned criteria hold. We will prove the simplification of the CR-assisted secrecy capacity formula for the strongly less capable criterion, later.

Definition 7 (Strongly Degraded): If there exists a best channel θ^* to the eavesdropper according to Definition 6, then an AVWC is called **strongly degraded** (with independent states, see [18]) if the following Markov chain holds

$$X \leftrightarrow Y_\theta \leftrightarrow Z_{\theta^*}, \quad \forall\theta\in\mathcal{P}(\mathcal{S}|\mathcal{X}).$$

Definition 8 (Strongly Noisier With Independent States): The family of channels to the illegitimate receiver $\mathcal{V}=\{(V_s:\mathcal{X}\to\mathcal{P}(\mathcal{Z})):\ s\in\mathcal{S}\}$ is **strongly noisier** with independent states than the family of channels to the legitimate receiver $\mathcal{W}=\{(W_s:\mathcal{X}\to\mathcal{P}(\mathcal{Y})):\ s\in\mathcal{S}\}$ if there exists a best channel θ^* to the eavesdropper according to Definition 6 and if for every random variable A such that $A\leftrightarrow X\leftrightarrow(Y_\theta,Z_{\theta'})$ we have for all $\theta\in\mathcal{P}(\mathcal{S}|\mathcal{X})$

$$I(p_A;W_\theta)\geq I(p_A;V_{\theta^*}).$$

Definition 9 (Strongly Less Capable With Independent States): The family of channels to the illegitimate receiver $\mathcal{V}=\{(V_s:\mathcal{X}\to\mathcal{P}(\mathcal{Z})):\ s\in\mathcal{S}\}$ is **strongly less capable** with independent states than the family of channels to the legitimate receiver $\mathcal{W}=\{(W_s:\mathcal{X}\to\mathcal{P}(\mathcal{Y})):\ s\in\mathcal{S}\}$ if there exists a best channel θ^* to the eavesdropper according to Definition 6 and if for every $p\in\mathcal{P}(\mathcal{X})$ we have for all $\theta\in\mathcal{P}(\mathcal{S}|\mathcal{X})$

$$I(p;W_\theta)\geq I(p;V_{\theta^*}).$$

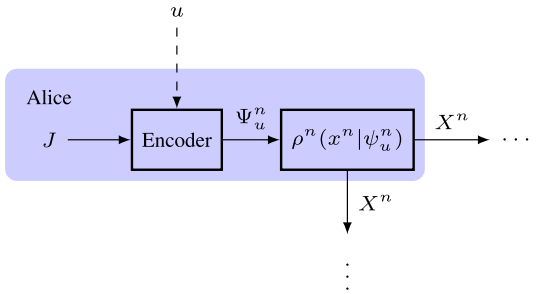

Fig. 2. Adapted system model with prefixing at Alice. With CR realization u, Alice encodes a secure message J into a codeword Ψ_u^n, of length n. The codeword serves as the input of a prefix channel $\rho(x^n|\psi_u^n)$, and is mapped to the channel input X^n. Other parts remain the same.

Remark 6: The requirement of an **i**ndependent and **i**dentically **d**istributed (i.i.d.) best channel to the eavesdropper is essential to formulate the strongly degraded, the strongly noisier, and the strongly less capable conditions as single-letter conditions. For the transmission link between the transmitter and the legitimate receiver, it has been shown in [35] that the jammer's best strategy is to perform a "memoryless attack". Hence, knowing the entire codesequence x^n non-causally is as good as knowing the codesymbol x_i non-causally at time i, while knowing the codeword x^n causally is as good as no knowledge at all [9]. For the transmission link between the transmitter and the eavesdropper, the i.i.d. structure of $\theta^{*,n}$ allows decomposing the multi-letter expressions into single-letter ones.

Remark 7: Just as in the stateless case [36], we have the following implication chain:

$$\text{Strongly Degraded} \rightarrow \text{Strongly Noisier} \rightarrow \text{Strongly Less Capable}.$$

Here, $X\rightarrow Y$ means X implies Y, but not vice versa.

III. Main Results

In the following, we state our main results. First, we present the secrecy capacity formulas for the general, and the strongly less capable cases, respectively, when the jammer has non-causal knowledge of the channel input. Then we provide the corresponding secrecy capacity formulas, when the jammer has no side information or only possesses knowledge of the messages.

A. Capacity Formulas for the General and the Less Capable Cases

Theorem 1: If there exists a best channel to the eavesdropper, the CR-assisted secrecy capacity for the AVWC with side information at the jammer $\widehat{\widehat{C}}_S^{\,\text{ran}}(\mathcal{W},\mathcal{V})$ is given by

$$\widehat{\widehat{C}}_S^{\,\text{ran}}(\mathcal{W},\mathcal{V}) = \max_{\substack{p_\Psi\in\mathcal{P}(\Psi),\\ \rho\in\mathcal{P}(\mathcal{X}|\Psi)}} \left(\min_{W\in\widehat{\widehat{\mathcal{W}}}} I(p_\Psi;\rho W) - \max_{V\in\widehat{\widehat{\mathcal{V}}}} I(p_\Psi;\rho V) \right), \tag{11}$$

with Ψ as a prefixing random variable and concatenated channels ρW and ρV, respectively.

Proof of Theorem 1: See Appendix F. ∎

Theorem 2: Let an AVWC $(\mathcal{W},\mathcal{V})$ be given. If for $(\mathcal{W},\mathcal{V})$, the channel $\mathcal{V}$ is strongly less capable with respect to the channel $\mathcal{W}$ and if there exists a best channel to the eavesdropper, then the CR-assisted code secrecy capacity $\widehat{\widehat{C}}_S^{\text{ran}}(\mathcal{W},\mathcal{V})$ is given by

$$\widehat{\widehat{C}}_S^{\text{ran}}(\mathcal{W},\mathcal{V}) = \max_{p_X\in\mathcal{P}(\mathcal{X})} \left(\min_{W\in\widehat{\widehat{\mathcal{W}}}} I(p_X;W) - \max_{V\in\widehat{\widehat{\mathcal{V}}}} I(p_X;V) \right) \tag{12}$$

Proof of Theorem 2: See Appendix G. ∎

The secrecy capacity $\widehat{\widehat{C}}_S^{\text{ran}}(\mathcal{W},\mathcal{V})$ depends on the row convex closures $\widehat{\widehat{\mathcal{W}}}$ and $\widehat{\widehat{\mathcal{V}}}$, like in [16] and [35].

B. Capacity Formulas Without Side Information at the Jammer, or Where the Jammer Only Knows the Messages

Corollary 1: Let an AVWC $(\mathcal{W},\mathcal{V})$ be given. If there exists a best channel to the eavesdropper and if the adversary is either blind or semi-blind, then the CR-assisted secrecy capacity under the maximum error criterion is given by

$$\widehat{C}_S^{\text{ran}}(\mathcal{W},\mathcal{V}) = \max_{\substack{p_\Psi\in\mathcal{P}(\Psi),\\ \rho\in\mathcal{P}(\mathcal{X}|\Psi)}} \left(\min_{W\in\widehat{\mathcal{W}}} I(p_\Psi;\rho W) - \max_{V\in\widehat{\mathcal{V}}} I(p_\Psi;\rho V) \right), \tag{13}$$

If for the AVWC $(\mathcal{W},\mathcal{V})$, the channel $\mathcal{V}$ is strongly less capable with respect to the channel $\mathcal{W}$, then the CR-assisted secrecy capacity under the maximum error criterion simplifies to

$$\widehat{C}_S^{\text{ran}}(\mathcal{W},\mathcal{V}) = \max_{p_X\in\mathcal{P}(\mathcal{X})} \left(\min_{W\in\widehat{\mathcal{W}}} I(p_X;W) - \max_{V\in\widehat{\mathcal{V}}} I(p_X;V) \right). \tag{14}$$

The result in (14) is an extension of [22, Corollary 1] to the strongly less capable case.

Proof of Corollary 1: By simple modifications (especially due to the fact that the adversary is (semi-) blind, it is sufficient to require the single-letter condition in Remark 5 to hold) in Lemma 13, the secrecy analysis, as well as in the converse proof, it is easy to see that the theorem holds. ∎

The secrecy capacity $\widehat{C}_S^{\text{ran}}(\mathcal{W},\mathcal{V})$ (in contrast to $\widehat{\widehat{C}}_S^{\text{ran}}(\mathcal{W},\mathcal{V})$) depends on the convex closures $\widehat{\mathcal{W}}$ and $\widehat{\mathcal{V}}$.

IV. Discussion

A. Input and State Constraints

The extension of the results to the case of input and state constraints is not straightforward. While the modifications in the sense of [29] might be possible and may lead to a single-letter CR-assisted secrecy capacity, the restrictions on the jammer's strategy are rigorous. In [29], the jammer is restricted to a type-constrained jamming strategy. In [37], the authors considered deterministic wiretap-codes for the AVWC with input and state peak constraints. The authors derived single-letter formulas for upper and lower bounds for the secrecy rate and CR-assisted secrecy rate under input and state peak constraints, in general, and derived capacity results for the strongly less noisy case. In [27], a general multi-letter formula for the achievable CR-assisted secrecy rate with input and state peak constraints is presented.

B. From Random to Deterministic - Not Elimination

In [10], Ahlswede proposes the Elimination of Correlation technique to reduce the amount of CR to only n^2. He then uses a prefix code to inform the receiver which realization of the randomness is used. This leads to the following dichotomy result: The deterministic code capacity (under the average error criterion) equals its CR-assisted capacity, or is equal to zero if the AVC is symmetrizable. Note that this technique cannot be used in our system model. If a prefix code were used to inform the receiver which deterministic code is used, the jammer would obtain this information as well, and we obtain once again the situation of the maximal error criterion for deterministic codes.

The authors of [16] present a technique to reduce the amount of CR. If a polynomial decay of the upper bound on the error probability is feasible, a polynomial amount of CR seems sufficient. The authors draw codewords not from the complete set of typical sequences but a "suitable" subset. This approach hints at the possibility of reducing the amount of CR in our scenario, too. However, there is a crucial difference. For WTCs, the secrecy capacity relies on auxiliary random variables, in general. One of these RVs can be interpreted as adding a "prefix" channel as part of the encoding process. In our case, this RV is represented by Ψ. In our achievability proof, apart from the requirements that codewords occur in multiple codebooks, indexed by the realization of the CR, and are bad only for few, it is important that codewords together with channel inputs and state sequences possess a Markov structure with high probability for all jamming strategies. This is the motivation behind Lemma 11 and Lemma 12. If the Markov structure holds, we can apply Lemma 13 when upper bounding the probability that an error bound λ is not met, Appendix F-C. However, since for all possible jamming strategies (and all messages pairs) the amount of CR realizations for which the Markov structure property does not hold has to be small compared to the total amount of CR, we have to choose an amount of CR which is larger than $\log|\mathcal{F}|$, which is exponentially growing with the blocklength n. At least for the general case, we do not see how we can circumvent this fact.

For the special cases that the AVC to the eavesdropper is strongly degraded, strongly noisier, or strongly less capable, the CR-assisted secrecy capacity does not rely on auxiliary RVs anymore. We think, in these cases, the approach of [16] can be used to upper bound the necessary amount of CR. The reduction of the amount of CR is meaningful since in practical implementations, CR might be expensive. Hence,

from a system design point of view, it makes sense to reduce the necessary amount of CR. In [35], the authors derived the CR capacity of an AVC with non-causal knowledge of the input at the adversary. The authors stated that the necessary amount of CR is between n^2 (the bound of a blind adversary) and $\exp\{n\epsilon\}$.

Unfortunately, this approach does not guarantee that the amount of CR can grow less than exponentially with respect to the blocklength. We assume that for the case of a blind adversary, Ahlswede's **R**obustification **T**echnique (RT) and ET work. Indeed, in [19], the authors applied exactly these methods, and limited the amount of common randomness to be polynomial (n^3). Furthermore, the authors derived a single-letter formula for the secrecy capacity for the AVWC, under the condition that there exists a best channel to the eavesdropper and a worst channel to the legitimate receiver, and under the condition that strong degradedness holds, [19, Remark 1]. For the general case, the authors derived a multi-letter formula for the AVWC, under the condition that there exists a best channel to the eavesdropper, [19, Theorem 5]. Both formulas are valid if the AVC to the legitimate receiver is not symmetrizable (for the case of the average error probability criterion).

In [21, Theorem 2], the authors presented results for the deterministic secrecy capacity for the AVWC in the presence of a blind adversary. The authors derived that the deterministic secrecy capacity equals its CR-assisted secrecy capacity (under the average error criterion) if the channel to the legitimate receiver is not symmetrizable. Interestingly, in this work the authors did not use the ET by Ahlswede. Limiting the amount of randomness is crucial, as well as deriving a formula for the deterministic case. However, for the reason mentioned above, the RT and ET are not applicable for the case of non-causal information of the channel input at the adversary. And for the blind adversary, results have been derived in [19] and [21] (for the case of the average error probability criterion).

However, deriving deterministic code results or a minimal amount of CR is not the intention of this work. Instead, we assume that there exists a sufficient amount of CR to compute fundamental results on the secrecy capacities for different knowledge scenarios at the jammer.

C. Justification for the Multi-Letter Requirements

We require a multi-letter condition to hold for the best channel to the eavesdropper in Definition 6 because of the following reason. Our secrecy analysis in the achievability part is based on typicality. This typicality approach requires that the underlying (single-letter) probability functions are repeated in an i.i.d. manner. Now, assume a single-letter best channel exists to the eavesdropper (which is essentially the adversary's worst case attack strategy). Repeating this strategy in an i.i.d. manner n times might be suboptimal for the adversary in general because the spaces spanned by the n-letter extension of the row convex closure and the row convex closure of the n-letter extension are not equivalent. In other words, the adversary's strategy space is larger when choosing the strategy from $\mathcal{P}(\mathcal{S}^n|\mathcal{X}^n)$ instead of $\mathcal{P}^n(\mathcal{S}|\mathcal{X})$. In Definition 6, we require explicitly that repeating the single-letter best channel is optimal for all blocklengths. To further illustrate this fact, we state the following.

Definition 10 (n-Letter Extension of $\widehat{\widetilde{\mathcal{W}}}$): The **$n$-letter extension of $\widehat{\widetilde{\mathcal{W}}}$** is defined as the set

$$\widetilde{\mathcal{W}}^n := \Bigg\{ W_{\hat{p}}^n(Y^n|X^n) : \sum_{s^n\in\mathcal{S}^n} \hat{p}(s^n|x^n) W^n(\cdot|x^n,s^n), \\ \hat{p}(s^n|x^n) \in \mathcal{P}(\mathcal{S}^n|\mathcal{X}^n), x^n \in \mathcal{X}^n \Bigg\} \tag{15}$$

Remark 8: Note that $\widetilde{\mathcal{W}}^n \neq \widehat{\widetilde{\mathcal{W}}}^n$. It can be shown that the operations of the Kronecker product and taking the row convex closure are not commutative. In other words, the row convex closure of a n-letter extension is not the n-letter extension of a row convex closure.

Example 1 (Continued): We have

$$W(\cdot|\cdot,S=0)\otimes W(\cdot|\cdot,S=0) = \left(\begin{smallmatrix} 1&0&0&0\\ 0&1&0&0\\ 0&0&1&0\\ 0&0&0&1 \end{smallmatrix}\right)$$
$$W(\cdot|\cdot,S=1)\otimes W(\cdot|\cdot,S=1) = \left(\begin{smallmatrix} 0&0&0&1\\ 0&0&1&0\\ 0&1&0&0\\ 1&0&0&0 \end{smallmatrix}\right)$$
$$W(\cdot|\cdot,S=0)\otimes W(\cdot|\cdot,S=1) = \left(\begin{smallmatrix} 0&1&0&0\\ 1&0&0&0\\ 0&0&0&1\\ 0&0&1&0 \end{smallmatrix}\right)$$
$$W(\cdot|\cdot,S=1)\otimes W(\cdot|\cdot,S=0) = \left(\begin{smallmatrix} 0&0&1&0\\ 0&0&0&1\\ 1&0&0&0\\ 0&1&0&0 \end{smallmatrix}\right)$$

Hence, when taking the row convex closure now, we obtain

$$\widetilde{\mathcal{W}}^2 = \Bigg\{ \left(\begin{smallmatrix} \alpha_1 & \alpha_2 & \alpha_3 & 1-\alpha_1-\alpha_2-\alpha_3\\ \beta_1 & \beta_2 & \beta_3 & 1-\beta_1-\beta_2-\beta_3\\ \gamma_1 & \gamma_2 & \gamma_3 & 1-\gamma_1-\gamma_2-\gamma_3\\ \delta_1 & \delta_2 & \delta_3 & 1-\delta_1-\delta_2-\delta_3 \end{smallmatrix}\right) : \\ \alpha_i,\beta_i,\gamma_i,\delta_i \in [0,1],\ i\in\{1,2,3\}, \\ \sum_{i=1}^3 \alpha_i = \sum_{i=1}^3 \beta_i = \sum_{i=1}^3 \gamma_i = 1 \Bigg\}$$

In contrast, when taking the row convex closure first, and then calculating the two letter extension, we obtain

$$\widehat{\widetilde{W}}_1(\cdot|\cdot)\otimes\widehat{\widetilde{W}}_2(\cdot|\cdot)) = \left(\begin{smallmatrix} \alpha_1 & 1-\alpha_1\\ 1-\beta_1 & \beta_1 \end{smallmatrix}\right)\otimes\left(\begin{smallmatrix} \alpha_2 & 1-\alpha_2\\ 1-\beta_2 & \beta_2 \end{smallmatrix}\right)$$
$$\widehat{\widetilde{\mathcal{W}}}^2 = \\ \Bigg\{ \left(\begin{smallmatrix} \alpha_1\alpha_2 & \alpha_1(1-\alpha_2) & (1-\alpha_1)\alpha_2 & (1-\alpha_1)(1-\alpha_2)\\ \alpha_1(1-\beta_2) & \alpha_1\beta_2 & (1-\alpha_1)(1-\beta_2) & (1-\alpha_1)\beta_2\\ (1-\beta_1)\alpha_2 & (1-\beta_1)(1-\alpha_2) & \beta_1\alpha_2 & \beta_1(1-\alpha_2)\\ (1-\beta_1)(1-\beta_2) & (1-\beta_1)\beta_2 & \beta_1(1-\beta_2) & \beta_1\beta_2 \end{smallmatrix}\right) : \\ \alpha_i,\beta_i\in[0,1],\ i\in\{1,2\} \Bigg\}.$$

It is easy to see that the row $\left[\begin{smallmatrix}\frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0\end{smallmatrix}\right]$ is achievable in $\widetilde{\mathcal{W}}^2$ but not in $\widehat{\widetilde{\mathcal{W}}}^2$.

D. Relation to the Secrecy Capacity Under Average Error Criterion

The average error criterion for AVCs has been considered for example by [17]. In the following, we provide the CR-assisted secrecy capacity formula under the average error criterion and set the capacity formulas into relation to each

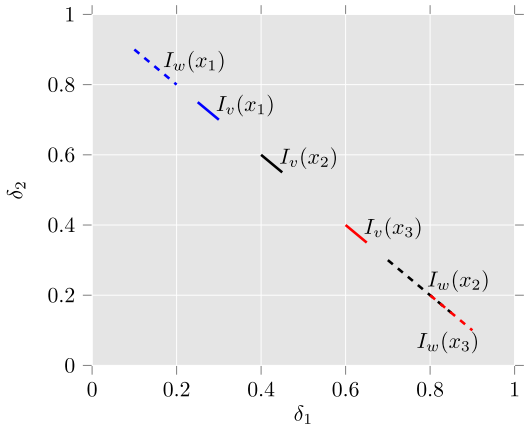

Fig. 3. Difference of capacities if the channel input is known or unknown at the jammer. Depicted are probability functions on $[0,1]$. Each conditional channel output probability is such a probability function. In order to use the same figure, we use both, $\delta_1 = W(y_1|\cdot,\cdot)$, $\delta_2 = W(y_2|\cdot,\cdot)$, and $\delta_1 = V(z_1|\cdot,\cdot)$, $\delta_2 = V(z_2|\cdot,\cdot)$. In both cases, we have $\delta_2 = 1-\delta_1$.

other. Let the achievable CR-assisted secrecy rates and the CR-assisted secrecy capacity $\widehat{C}^{\text{ran}}_{S,av}(\mathcal{W},\mathcal{V})$ under the average error criterion (averaging over the set of messages) be analogously defined as in Definition 4.

Corollary 2 (Common Randomness Assisted Secrecy Capacity Under the Average Error Criterion if the Family of Channels to the Illegitimate Receiver is Strongly Degraded, Strongly Noisier, or Strongly Less Capable With Independent States): If for an AVWC the family of channels to the illegitimate receiver $\mathcal{V}$ is strongly degraded, strongly noisier, or strongly less capable with independent states, then the CR-assisted secrecy capacity under the average error criterion for the standard AVWC is given by

$$\widehat{C}^{\text{ran}}_{S,av}(\mathcal{W},\mathcal{V}) = \max_{p_X}\left(\min_{W\in\widehat{\mathcal{W}}} I(p_X;W) - \max_{V\in\widehat{\mathcal{V}}} I(p_X;V)\right).$$

Corollary 3: Let an AVWC $(\mathcal{W},\mathcal{V})$ be given. If there exists a best channel to the eavesdropper, then

$$\widehat{C}^{\text{ran}}_{S,av}(\mathcal{W},\mathcal{V}) = \widehat{C}^{\text{ran}}_{S}(\mathcal{W},\mathcal{V}) \geq \widehat{\widehat{C}}^{\text{ran}}_{S}(\mathcal{W},\mathcal{V}). \tag{16}$$

Proof: The first equality in (16) follows because of Theorem 2 and Corollary 2. Furthermore, it is easy to see that $\widehat{\mathcal{W}} \subset \widehat{\widehat{\mathcal{W}}}$ and $\widehat{\mathcal{V}} \subset \widehat{\widehat{\mathcal{V}}}$. ∎

E. Example

To clarify the fundamental difference between the capacity formulas mentioned above, and to show that the inclusion can be strict, we provide an explicit example. First, we define $\mathcal{I}_{(\cdot)}(\cdot)$ as the convex hull of the row of channel matrices as follows.

Definition 11 ([10]): For a given $x\in\mathcal{X}$, let $\mathcal{I}_w(x)$ denote the convex hull of the set $\{W(\cdot|x,s): s\in\mathcal{S}\}$ of probability distributions on $\mathcal{Y}$, i.e., $\mathcal{I}_w(x) = \text{conv}\left(W(\cdot|x,s): s\in\mathcal{S}\right)$.

Example 2: We consider the following example. Let the channel matrices be given as follows.

$$W(\cdot|\cdot,s_1) = \begin{pmatrix} 0.1 & 0.9 \\ 0.7 & 0.3 \\ 0.8 & 0.2 \end{pmatrix}, \qquad W(\cdot|\cdot,s_2) = \begin{pmatrix} 0.2 & 0.8 \\ 0.85 & 0.15 \\ 0.9 & 0.1 \end{pmatrix}$$
$$V(\cdot|\cdot,s_1) = \begin{pmatrix} 0.25 & 0.75 \\ 0.4 & 0.6 \\ 0.6 & 0.4 \end{pmatrix}, \qquad V(\cdot|\cdot,s_2) = \begin{pmatrix} 0.3 & 0.7 \\ 0.45 & 0.55 \\ 0.65 & 0.35 \end{pmatrix}$$

From Fig. 3, it is easy to see that this AVWC fulfills the strongly less capable property for a blind adversary. The eavesdropping channel possesses for every convex (or row convex) closure a channel law, that is closer to a uniform distribution than the legitimate channel (i.e., is closer to the point $(\delta_1,\delta_2) = (\frac{1}{2},\frac{1}{2})$). We have

$$\begin{aligned}\widehat{W} &= \alpha W(\cdot|\cdot,s_1) + (1-\alpha)W(\cdot|\cdot,s_2) \\ &= \begin{pmatrix} 0.2-0.1\alpha & 0.8+0.1\alpha \\ 0.85-0.15\alpha & 0.15+0.15\alpha \\ 0.9-0.1\alpha & 0.1+0.1\alpha \end{pmatrix}, \\ \widehat{V} &= \beta V(\cdot|\cdot,s_1) + (1-\beta)V(\cdot|\cdot,s_2) \\ &= \begin{pmatrix} 0.3-0.05\beta & 0.7+0.05\beta \\ 0.45-0.05\beta & 0.55+0.05\beta \\ 0.65-0.05\beta & 0.35+0.05\beta \end{pmatrix}.\end{aligned}$$

The secrecy capacity $\widehat{C}^{\text{ran}}_{S}(\mathcal{W},\mathcal{V})$ of this AVWC can be calculated to $\widehat{C}^{\text{ran}}_{S}(\mathcal{W},\mathcal{V}) \approx 0.3$ bits per channel use, $p_X(0) = p_X(2) = 0.5$, $p_X(1) = 0$, $\alpha = 0.5$, $\beta \approx 1$. In contrast to that, one can easily see that the channels

$$\widehat{\widehat{W}} = \begin{pmatrix} 0.2 & 0.8 \\ 0.8 & 0.2 \\ 0.8 & 0.2 \end{pmatrix} \qquad \widehat{\widehat{V}} = \begin{pmatrix} 0.25 & 0.75 \\ 0.4 & 0.6 \\ 0.65 & 0.35 \end{pmatrix}$$

correspond to the worst and the best channels to Bob and Eve (for $n=1$), respectively, if the channel input is non-causally known at the jammer. In this case, that the formula for the secrecy capacity for the AVWC is evaluated with respect to the row convex closures, we obtain approximately 0.26 bits per channel use (which is strictly smaller than $\widehat{C}^{\text{ran}}_{S}(\mathcal{W},\mathcal{V})$), with input distribution $p_X(0) = p_X(1) = 0.5$, $p_X(2) = 0$. The second input symbol is used for the case with non-causal side information at the jammer instead of the third one as for the AVWC with blind adversary.

F. Summary

In this work, we derive a single-letter formula for the CR-assisted secrecy capacity under the maximum error criterion for an active attacker with non-causal side information of the codewords, provided there exists a best channel to the eavesdropper. Additionally, we provide a formula for the CR-assisted secrecy capacity for the case that the eavesdropping channel is strongly degraded, strongly noisier, or strongly less capable with respect to the main channel. We further allow that the messages might also be known at the jammer. We apply and extend methods of [16] and [22]. We show that the derived secrecy capacities depend on the row convex closures of the sets of channels to Bob and Eve for the general and the strongly degraded cases, respectively, if the input is non-causally known at the jammer and depend on the convex closures of the sets of channels if the channel input is not non-causally known at the jammer.

We compare our results to the CR-assisted secrecy capacity for the cases of maximum error criterion and blind adversary, maximum error criterion and semi-blind adversary, and the standard AVWC (average error probability and blind adversary). In the considered system model, the worst case occurs if the codewords (channel inputs) are non-causally known at the jammer. As we have shown, it does not matter if the jammer additionally knows the messages. The CR-assisted secrecy capacity is determined with respect to the

row convex closures of the channel sets. In contrast, if the adversary is blind or semi-blind, then the CR-assisted secrecy capacity under the average or maximum error criterion is determined with respect to the convex closure of the channel sets. We provided an example to illustrate this fundamental difference. It is quite obvious that optimizing over a larger set, here the row convex closure compared to the convex closure of the channel sets, may lead to a smaller CR-assisted secrecy capacity.

From a resource theory point of view, the necessary amount of CR is of interest. We do not upper bound the amount of CR. To ensure that codewords occur in sufficiently many codebooks in order to confuse the jammer, we give a lower bound on the amount of CR. This CR is known at the eavesdropper and hence cannot be used as key to achieve a secure transmission. Secrecy is achieved by wiretap coding.

Appendix A
Exchangeability of Order of Maximization

Lemma 2: Let the sequence $(a_{i,j})_{\substack{i\in\mathcal{A},\\ j\in\mathcal{B}}}$, $a_{i,j}\in\mathbb{R}$ be given, where $\mathcal{A},\mathcal{B}\subset\mathbb{N}$ are finite sets. Then

$$\max_{i\in\mathcal{A}}\max_{j\in\mathcal{B}}(a_{i,j})_{\substack{i\in\mathcal{A},\\ j\in\mathcal{B}}}=\max_{j\in\mathcal{B}}\max_{i\in\mathcal{A}}(a_{i,j})_{\substack{i\in\mathcal{A},\\ j\in\mathcal{B}}}.$$

Proof: Let $\mathcal{J}^*$ and $\mathcal{I}^*$ be given as

$$\mathcal{J}^*=\left\{\max_{i\in\mathcal{A}}(a_{i,j}):\ j\in\mathcal{B}\right\}$$
$$\mathcal{I}^*=\left\{\max_{j\in\mathcal{B}}(a_{i,j}):\ i\in\mathcal{A}\right\}$$

Then it is easy to see that

$$\max_{j\in\mathcal{B}}J^*=\max_{i\in\mathcal{A}}I^*$$

Intuitively, the result follows when imagining a matrix. If the global maximum is unique, then the operations of collecting the maximum in each column in the set $\mathcal{J}^*$ and then taking the maximal element of $\mathcal{J}^*$ is equivalent to collecting the maximum in each row in the set $\mathcal{I}^*$ and then taking the maximal element of $\mathcal{I}^*$.

If the global maximum is not unique, the result remains the same, but the indices $(i,j)\in\mathcal{A}\times\mathcal{B}$ might change. ∎

Appendix B
Variation Distance, Markov, Chernoff, and Chernoff-Hoeffding Bounds

Definition 12 (Variation Distance): The variation distance of two distributions P_1,P_2 on $\mathcal{X}$ is defined as

$$||P_1-P_2||_V=\sum_{x\in\mathcal{X}}|P_1(x)-P_2(x)|. \tag{17}$$

Lemma 3 ([34, Lemma 2.7]): If $||P_1-P_2||_V=\tau\le\frac{1}{2}$, then

$$|H(P_1)-H(P_2)|\le-\tau\log\frac{\tau}{|\mathcal{X}|}.$$

We give a reminder on Markov's inequality.

Lemma 4 (Markov's Inequality [38, Lemma 83]): Let X be a RV with mean $E[X]=\mu$ and let a be a positive number. Then

$$Pr\{X\ge a\}\le\frac{\mu}{a}.$$

Chernoff bounds are given as follows.

Lemma 5 (Chernoff Bounds, [39], [16, Lemma 2]): Let $X_1,X_2,\dots,X_n$ be i.i.d. RVs with values in $\{0,1\}$, with $Pr\{X_i=1\}=p$. For all $\epsilon\in(0,1)$ and $p_0<p<p_1$, the following bounds hold

$$Pr\left\{\frac{1}{n}\sum_{i=1}^{n}X_i>(1+\epsilon)p_1\right\}<\exp_e\left\{-\frac{\epsilon^2}{8}np_1\right\}, \tag{18}$$
$$Pr\left\{\frac{1}{n}\sum_{i=1}^{n}X_i<(1-\epsilon)p_0\right\}<\exp_e\left\{-\frac{3\epsilon^2}{8}np_0\right\}. \tag{19}$$

The Chernoff-Hoeffding bound is widely used in the proof. Therefore, it shall be stated here.

Lemma 6 (Chernoff-Hoeffding Bounds, [40, Theorem 1.1], [41]): Let $X_1,X_2,\dots,X_n$ be i.i.d. RVs with values in $[0,b]$, where b is a positive number. Further, let $E[X_i]=\mu$, and $0<\epsilon<\frac{1}{2}$. Then

$$Pr\left\{\frac{1}{n}\sum_{i=1}^{n}X_i\notin[(1\pm\epsilon)\mu]\right\}\le2\exp_e\left(-n\frac{\epsilon^2\mu}{3b}\right), \tag{20}$$

where $[(1\pm\epsilon)\mu]$ means the interval $[(1-\epsilon)\mu,(1+\epsilon)\mu]$.

Appendix C
Typical Sets

We summarize some known facts of typicality properties. Let $\delta>0$.

Lemma 7 (Properties of Typical Sets I, [34, Lemma 2.13, Problem 2.5]): Let $x^n\in\mathcal{T}^n_{p,\delta}$. Then for any $W:\mathcal{X}\to\mathcal{P}(\mathcal{Y})$

$$|\mathcal{T}^n_{pW,2|\mathcal{X}|\delta}|\le\exp\{n(H(pW)+f_1(\delta))\},$$
$$W^n(y^n|x^n)\le\exp\{-n(H(W|p)-f_2(\delta))\},\ \forall y^n\in\mathcal{T}^n_{W,\delta}(x^n),$$

for some functions $f_1(\delta),f_2(\delta)>0$ with $\lim_{\delta\to0}f_1(\delta)=0$ and $\lim_{\delta\to0}f_2(\delta)=0$.

Lemma 8 (Properties of Typical Sets II, [42, Lemma III.1.3]): For every $p\in\mathcal{P}(\mathcal{X})$, $W:\mathcal{X}\to\mathcal{P}(\mathcal{Y})$ and $x^n\in\mathcal{X}^n$

$$p^n(\mathcal{T}^n_{p,\delta})\ge1-(n+1)^{|\mathcal{X}|}\exp\{-nc\delta^2\},$$
$$W^n(\mathcal{T}^n_{W,\delta}(x^n)|x^n)\ge1-(n+1)^{|\mathcal{X}||\mathcal{Y}|}\exp\{-nc\delta^2\}.$$

with $c=\frac{1}{2\ln2}$. Furthermore, there exists an n_0 and a $c'>0$, depending on $|\mathcal{X}|,|\mathcal{Y}|$ and δ, such that for all $n>n_0$ for each $p\in\mathcal{P}(\mathcal{X})$ and $W:\mathcal{X}\to\mathcal{P}(\mathcal{Y})$

$$p^n(\mathcal{T}^n_{p,\delta})\ge1-\exp\{-nc'\delta^2\}, \tag{21}$$
$$W^n(\mathcal{T}^n_{W,\delta}(x^n)|x^n)\ge1-\exp\{-nc'\delta^2\}. \tag{22}$$

Lemma 9 (Properties of Typical Sets III, [34, Lemma 2.2]): Let $\mathcal{P}^n_0(\mathcal{S})$ be the set of all possible types of n-length sequences on $\mathcal{S}^n$. The cardinality of the set of all possible types of length n is upper bounded by

$$|\mathcal{P}^n_0(\mathcal{S})|\le(n+1)^{|\mathcal{S}|}.$$

Lemma 10 (Properties of Typical Sets IV, [43, Lemma 3] [31, Lemma 3]): Assume, the distributions $p, \bar{p} \in \mathcal{P}(\mathcal{X})$ and the two matrices $W, \overline{W} : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$ are given. For any positive integer n and sufficiently small $\delta > 0$,

$$(pW)^n(\mathcal{T}^n_{\overline{W},\delta}(\bar{x}^n)) \leq (n+1)^{|\mathcal{X}||\mathcal{Y}|} \exp\{-n(I(\bar{p};\overline{W}) - f_3(\delta))\},$$

for all $\bar{x}^n \in \mathcal{T}^n_{\bar{p},\delta}$ holds, with some $f_3(\delta) > 0$ and $\lim_{\delta\to 0} f_3(\delta) = 0$. Furthermore, there exist an n_0 and a $\nu > 0$, depending on $|\mathcal{X}|, |\mathcal{Y}|$ and δ, such that for all $n > n_0$,

$$(pW)^n(\mathcal{T}^n_{\overline{W},\delta}(\bar{x}^n)) \leq \exp\{-n(I(\bar{p};\overline{W}) - \nu))\}. \tag{23}$$

Lemma 11 (Properties of Typical Sets V, [14, Lemma 2]): Let the sequences $x^n \in \mathcal{X}^n$, $s^n \in \mathcal{S}^n$, and $\delta, \hat{\delta} > 0$ be given. Further, let (Ψ, X) be distributed according to $p_{\Psi,X} = p_\Psi \rho_{X|\Psi}$. Define the channel

$$\theta(s|x) := \frac{1}{N(x|x^n)} \sum_{i=1}^n \mathbb{1}(s_i = s, x_i = x).$$

Then,

$$Pr\left\{(\Psi^n, x^n, s^n) \notin \mathcal{T}^n_{p_\Psi \rho_{X|\Psi}\theta,\delta} | (\Psi^n, x^n) \in \mathcal{T}^n_{p_\Psi \rho_{X|\Psi},\hat{\delta}}\right\} \leq \exp\{-nh(\delta)\}, \tag{24}$$

where $h(\delta) \to 0$ as $\delta \to 0$. *Proof:* Follows for example by [34, Lemma 2.10, Lemma 2.12]. ∎

Lemma 12: Let $f : \mathcal{X}^n \to \mathcal{S}^n$ be fixed and let $(\Psi^n, X^n, f(X^n)) \in \Psi^n \times \mathcal{X}^n \times \mathcal{S}^n$ be distributed according to $p^n_\Psi \rho^n_{X|\Psi} \mathbb{1}_{f(X^n)}(X^n)$. Let A be defined as the following event.

$$A := \{\nexists \underline{\theta} \in \mathcal{P}_0(\mathcal{S}^n|\mathcal{X}^n) : \ (\Psi^n, X^n, f(X^n)) \in \mathcal{T}^n_{p_\Psi \rho_{X|\Psi}\underline{\theta},\delta}\}.$$

Then, we have for some $c' > 0$, depending on $|\Psi|$ and $|\mathcal{X}|$, and $\hat{\delta} > 0$

$$Pr\{A\} \leq \exp\{-nc'\hat{\delta}^2\} + (n+1)^{|\mathcal{X}||\mathcal{S}|} \exp\left\{-n \min_{\theta \in \mathcal{P}_0(\mathcal{S}^n|\mathcal{X}^n)} h_\theta(\delta)\right\},$$

where $h_\theta(\delta) \to 0$ if $\delta \to 0$.

Proof:

$$\begin{aligned}
Pr\{A\} &= Pr\left\{(\Psi^n, X^n) \notin \mathcal{T}^n_{p_\Psi \rho_{X|\Psi},\hat{\delta}}\right\} \\
&\quad Pr\left\{A|(\Psi^n, X^n) \notin \mathcal{T}^n_{p_\Psi \rho_{X|\Psi},\hat{\delta}}\right\} \\
&\quad + Pr\left\{(\Psi^n, X^n) \in \mathcal{T}^n_{p_\Psi \rho_{X|\Psi},\hat{\delta}}\right\} \\
&\quad Pr\left\{A|(\Psi^n, X^n) \in \mathcal{T}^n_{p_\Psi \rho_{X|\Psi},\hat{\delta}}\right\} \\
&\leq Pr\left\{(\Psi^n, X^n) \notin \mathcal{T}^n_{p_\Psi \rho_{X|\Psi},\hat{\delta}}\right\} \\
&\quad + Pr\left\{A|(\Psi^n, X^n) \in \mathcal{T}^n_{p_\Psi \rho_{X|\Psi},\hat{\delta}}\right\} \\
&\overset{(a)}{\leq} \exp\{-nc'\hat{\delta}^2\} + Pr\left\{A|(\Psi^n, X^n) \in \mathcal{T}^n_{p_\Psi \rho_{X|\Psi},\hat{\delta}}\right\} \\
&= \exp\{-nc'\hat{\delta}^2\} \\
&\quad + \sum_{x^n \in \mathcal{X}^n} Pr\left\{A|(\Psi^n, x^n) \in \mathcal{T}^n_{p_\Psi \rho_{X|\Psi},\hat{\delta}}\right\} \\
&\quad Pr\left\{X^n = x^n|(\Psi^n, x^n) \in \mathcal{T}^n_{p_\Psi \rho_{X|\Psi},\hat{\delta}}\right\} \\
&\overset{(b)}{=} \exp\{-nc'\hat{\delta}^2\} \\
&\quad + \sum_{\substack{\theta \in \mathcal{P}_0(\mathcal{S}^n|\mathcal{X}^n) \\ x^n : f(x^n) \in \mathcal{T}_\theta(x^n)}} Pr\Big\{(\Psi^n, x^n, f(x^n)) \notin \mathcal{T}^n_{p_\Psi \rho_{X|\Psi}\theta,\delta} | \\
&\quad (\Psi^n, x^n) \in \mathcal{T}^n_{p_\Psi \rho_{X|\Psi},\hat{\delta}}\Big\} \\
&\quad Pr\left\{X^n = x^n|(\Psi^n, x^n) \in \mathcal{T}^n_{p_\Psi \rho_{X|\Psi},\hat{\delta}}\right\} \\
&\overset{(c)}{\leq} \exp\{-nc'\hat{\delta}^2\} + \sum_{\theta \in \mathcal{P}_0(\mathcal{S}^n|\mathcal{X}^n)} \exp\{-nh_\theta(\delta)\} \\
&\quad \sum_{x^n : f(x^n) \in \mathcal{T}_\theta(x^n)} Pr\left\{X^n = x^n|(\Psi^n, x^n) \in \mathcal{T}^n_{p_\Psi \rho_{X|\Psi},\hat{\delta}}\right\} \\
&\overset{(d)}{\leq} \exp\{-nc'\hat{\delta}^2\} \\
&\quad + (n+1)^{|\mathcal{X}||\mathcal{S}|} \exp\left\{-n \min_{\theta \in \mathcal{P}_0(\mathcal{S}^n|\mathcal{X}^n)} h_\theta(\delta)\right\}.
\end{aligned}$$

(a) follows because of Lemma 8. (b) follows because each x^n invokes together with $f(x^n)$ a joint type $p\theta$, with $p \in \mathcal{P}_0(\mathcal{X}^n)$ and $\theta \in \mathcal{P}_0(\mathcal{S}^n|\mathcal{X}^n)$. So instead of summing over all $x^n \in \mathcal{X}^n$, we can iterate through all conditional types $\theta \in \mathcal{P}_0(\mathcal{S}^n|\mathcal{X}^n)$ and sum up over all x^n leading together with $f(x^n)$ to the conditional type $\theta \in \mathcal{P}_0(\mathcal{S}^n|\mathcal{X}^n)$. (c) follows because of Lemma 11, and (d) by type counting and bounding the last term in (c) by 1. ∎

APPENDIX D

Lemma 13: For any conditional type $\theta \in \mathcal{P}_0(\mathcal{S}^n|\mathcal{X}^n)$, define the probability measure $p_{\Psi XS}$ as

$$p_{\Psi XS}(\psi, x, s) = p_\Psi(\psi)\rho(x|\psi)\theta(s|x).$$

Let $\delta > 0$ and let $p_{\overline{\Psi XS}}$ be a type fulfilling $p_{\overline{\Psi}} = p_\Psi$ and

$$||p_{\Psi XS} - p_{\overline{\Psi XS}}||_V \leq \delta. \tag{25}$$

Moreover, let Ψ'^n be uniformly distributed on $\mathcal{T}^n_{p_\Psi}$. Then there exist an n_0 and a ν, depending on $|\mathcal{X}|, |\mathcal{Y}|, |\Psi|, |\mathcal{S}|$ and δ, such that for all $n > n_0$ we have for any $(x^n, s^n) \in \mathcal{T}^n_{p_{\overline{XS}}}$,

$$\begin{aligned}
&E\left[W^n\left(\left(\bigcup_{\underline{\theta} \in \mathcal{P}_0(\mathcal{S}^n|\mathcal{X}^n)} \mathcal{T}^n_{\rho W_{\underline{\theta}},\delta}(\Psi'^n)\right) \middle| x^n, s^n\right)\right] \\
&\leq \exp\left\{-n\left(\min_{\underline{\theta} \in \mathcal{P}_0(\mathcal{S}^n|\mathcal{X}^n)} I(p_\Psi; \rho W_{\underline{\theta}}) - \nu\right)\right\} \\
&\leq \exp\left\{-n\left(\min_{\underline{\theta} \in \mathcal{P}(\mathcal{S}|\mathcal{X})} I(p_\Psi; \rho W_{\underline{\theta}}) - \nu\right)\right\}.
\end{aligned}$$

Proof of Lemma 13: We divide the proof into two steps. First we provide an upper bound, and show then secondly that this upper bound holds for arbitrary sequences of the same type.

Let (Ψ^n, X^n, S^n) be uniformly distributed on $\mathcal{T}^n_{p_{\Psi XS}}$ and independent of Ψ'^n. First, we have

$$E\left[W^n\left(\left(\bigcup_{\underline{\theta} \in \mathcal{P}_0(\mathcal{S}^n|\mathcal{X}^n)} \mathcal{T}^n_{\rho W_{\underline{\theta}},\delta}(\Psi'^n)\right) \middle| X^n, S^n\right)\right]$$

$$
\begin{aligned}
&\overset{(a)}{\leq} \sum_{\underline{\theta}\in\mathcal{P}_0(\mathcal{S}^n|\mathcal{X}^n)} E\left[W^n\left(\mathcal{T}^n_{\rho W_{\underline{\theta}},\delta}(\Psi'^n)\middle|X^n,S^n\right)\right]\\
&\overset{(b)}{=} \sum_{\underline{\theta}\in\mathcal{P}_0(\mathcal{S}^n|\mathcal{X}^n)}\sum_{\psi'^n\in\Psi^n} p_{\Psi^n}(\psi'^n)\\
&\qquad \sum_{(\psi^n x^n s^n)\in\Psi^n\times\mathcal{X}^n\times\mathcal{S}^n} p^n_{\Psi XS}(\psi^n,x^n,s^n)\\
&\qquad W^n\left(\mathcal{T}^n_{\rho W_{\underline{\theta}},\delta}(\psi'^n)\middle|x^n,s^n\right)\\
&\overset{(c)}{=} \sum_{\underline{\theta}\in\mathcal{P}_0(\mathcal{S}^n|\mathcal{X}^n)}\sum_{\psi'^n\in\Psi^n} p_{\Psi^n}(\psi'^n)\\
&\qquad \sum_{(\psi^n x^n s^n)\in\Psi^n\times\mathcal{X}^n\times\mathcal{S}^n} p^n_{\Psi}(\psi^n)\rho^n(x^n|\psi^n)\\
&\qquad \theta^n(s^n|x^n)W^n\left(\mathcal{T}^n_{\rho W_{\underline{\theta}},\delta}(\psi'^n)\middle|x^n,s^n\right)\\
&\overset{(d)}{=} \sum_{\underline{\theta}\in\mathcal{P}_0(\mathcal{S}^n|\mathcal{X}^n)}\sum_{\psi'^n\in\Psi^n} p_{\Psi^n}(\psi'^n)\\
&\qquad (p_\Psi\rho W_\theta)^n\left(\mathcal{T}^n_{\rho W_{\underline{\theta}},\delta}(\psi'^n)\right)\\
&\overset{(e)}{\leq} \sum_{\underline{\theta}\in\mathcal{P}_0(\mathcal{S}^n|\mathcal{X}^n)}\exp\left\{-n\left(I(p_\Psi;\rho W_{\underline{\theta}})-\hat{\nu}\right)\right\}\sum_{\psi'^n\in\Psi^n} p_{\Psi^n}(\psi'^n)\\
&\overset{(f)}{\leq} (n+1)^{|\mathcal{X}||\mathcal{S}|}\\
&\qquad \exp\left\{-n\left(\min_{\underline{\theta}\in\mathcal{P}_0(\mathcal{S}^n|\mathcal{X}^n)} I(p_\Psi;\rho W_{\underline{\theta}})-\hat{\nu}\right)\right\}\\
&\overset{(g)}{\leq} \exp\left\{-n\left(\min_{\underline{\theta}\in\mathcal{P}_0(\mathcal{S}^n|\mathcal{X}^n)} I(p_\Psi;\rho W_{\underline{\theta}})-\nu\right)\right\}
\end{aligned}
$$

Here, (a) follows by the union bound. (b) follows by evaluating the expectation. (c) follows by assumption that $p_{\Psi XS}(\psi,x,s) = p_\Psi(\psi)\rho(x|\psi)\theta(s|x)$. (d) follows by expressing the probability function

$$
\begin{aligned}
&\sum_{(\psi^n x^n s^n)\in\Psi^n\times\mathcal{X}^n\times\mathcal{S}^n} p^n_{\Psi}(\psi^n)\rho^n(x^n|\psi^n)\theta^n(s^n|x^n)\\
&\qquad W^n\left(\left(\mathcal{T}^n_{\rho W_{\underline{\theta}},\delta}(\psi'^n)\right)\middle|x^n,s^n\right) \qquad (26)
\end{aligned}
$$

as the output probability function $(p_\Psi\rho W_\theta)^n\left(\mathcal{T}^n_{\rho W_{\underline{\theta}},\delta}(\psi'^n)\right)$. (e) follows by Lemma 10 with appropriate choice of $\hat{\nu}$, (f), and (g) follow by Lemma 9.

Next, assume that (Ψ^n, X^n, S^n) is uniformly distributed on $\mathcal{T}^n_{p_{\overline{\Psi XS}}}$ and independent of Ψ'^n. We will show that the above inequality also holds in this case up to small terms. Due to (25) and Lemma 3, we have

$$
\begin{aligned}
H(p_{\overline{\Psi XS}}) &\geq H(p_{\Psi XS}) + \delta\log\frac{\delta}{|\Psi||\mathcal{X}||\mathcal{S}|}\\
&=: H(p_{\Psi XS}) + \delta'.
\end{aligned}
$$

Furthermore, because of (25), we have $\mathcal{T}^n_{p_{\overline{\Psi XS}}} \subset \mathcal{T}^n_{p_{\Psi XS},\delta}$. Hence, for any nonnegative function $f(\psi^n,x^n,s^n)$, we have

$$
\begin{aligned}
&E[f(\Psi^n,X^n,S^n)]\\
&= \sum_{(\psi^n,x^n,s^n)\in\mathcal{T}^n_{p_{\overline{\Psi XS}}}} p^n_{\overline{\Psi XS}}(\psi^n,x^n,s^n)f(\psi^n,x^n,s^n)\\
&= \frac{1}{|\mathcal{T}^n_{p_{\overline{\Psi XS}}}|}\sum_{(\psi^n,x^n,s^n)\in\mathcal{T}^n_{p_{\overline{\Psi XS}}}} f(\psi^n,x^n,s^n)\\
&\leq (n+1)^{|\Psi||\mathcal{X}||\mathcal{S}|}\exp\{-nH(p_{\overline{\Psi XS}})\}\sum_{(\psi^n,x^n,s^n)\in\mathcal{T}^n_{p_{\overline{\Psi XS}}}} f(\psi^n,x^n,s^n)\\
&\leq (n+1)^{|\Psi||\mathcal{X}||\mathcal{S}|}\exp\{-n(H(p_{\Psi XS})-\delta')\}\\
&\qquad \sum_{(\psi^n,x^n,s^n)\in\mathcal{T}^n_{p_{\Psi XS},\delta}} f(\psi^n,x^n,s^n)\\
&\leq (n+1)^{|\Psi||\mathcal{X}||\mathcal{S}|}\exp\{n\delta''\}\\
&\qquad \sum_{(\psi^n,x^n,s^n)\in\mathcal{T}^n_{p_{\Psi XS},\delta}} p^n_\Psi(\psi^n)\rho^n(x^n|\psi^n)\theta^n(s^n|x^n)f(\psi^n,x^n,s^n)\\
&\leq (n+1)^{|\Psi||\mathcal{X}||\mathcal{S}|}\exp\{n\delta''\}\\
&\qquad \sum_{(\psi^n,x^n,s^n)\in\Psi^n\times\mathcal{X}^n\times\mathcal{S}^n} p^n_\Psi(\psi^n)\rho^n(x^n|\psi^n)\theta^n(s^n|x^n)f(\psi^n,x^n,s^n).
\end{aligned}
$$

With

$$
\begin{aligned}
f(\psi^n,x^n,s^n) = &\sum_{\underline{\theta}\in\mathcal{P}_0(\mathcal{S}^n|\mathcal{X}^n)}\sum_{\psi'^n\in\Psi^n} p_{\Psi^n}(\psi'^n)\\
&W^n\left(\left(\mathcal{T}^n_{\rho W_{\underline{\theta}},\delta}(\Psi'^n)\right)\middle|x^n,s^n\right),
\end{aligned}
$$

this shows

$$
\begin{aligned}
&E\left[W^n\left(\left(\bigcup_{\underline{\theta}\in\mathcal{P}_0(\mathcal{S}^n|\mathcal{X}^n)}\mathcal{T}^n_{\rho W_{\underline{\theta}},\delta}(\Psi'^n)\right)\middle|X^n,S^n\right)\right]\\
&\qquad \leq \exp\left\{-n\left(\min_{\underline{\theta}\in\mathcal{P}_0(\mathcal{S}^n|\mathcal{X}^n)} I(p_\Psi;\rho W_{\underline{\theta}})-\nu\right)\right\}.
\end{aligned}
$$

Secondly, for an arbitrary permutation of the index set $\{1,2,\ldots,n\}$ we have by definition

$$
\begin{aligned}
&\pi\left(\mathcal{T}^n_{\rho W_{\underline{\theta}},\delta}(\psi'^n)\right)\\
&:= \left\{\pi(y^n)\in\mathcal{Y}^n :\right.\\
&\qquad \left.\left|\frac{1}{n}N(a,b|\psi'^n,y^n)-\rho W_{\underline{\theta}}(b|a)\frac{1}{n}N(a|\psi'^n)\right|\leq\delta,\right.\\
&\qquad \left.\forall a\in\Psi, b\in\mathcal{Y}\right\}\\
&= \left\{y^n\in\mathcal{Y}^n :\right.\\
&\qquad \left.\left|\frac{1}{n}N(a,b|\psi'^n,\pi^{-1}(y^n))-\rho W_{\underline{\theta}}(b|a)\frac{1}{n}N(a|\psi'^n)\right|\leq\delta,\right.\\
&\qquad \left.\forall a\in\Psi, b\in\mathcal{Y}\right\}\\
&= \left\{y^n\in\mathcal{Y}^n :\right.\\
&\qquad \left.\left|\frac{1}{n}N(a,b|\pi(\psi'^n),y^n)-\rho W_{\underline{\theta}}(b|a)\frac{1}{n}N(a|\pi(\psi'^n))\right|\leq\delta,\right.\\
&\qquad \left.\forall a\in\Psi, b\in\mathcal{Y}\right\}\\
&=: \mathcal{T}^n_{\rho W_{\underline{\theta}},\delta}\left(\pi(\psi'^n)\right).
\end{aligned}
$$

Therefore, for a $(\tilde{x}^n, \tilde{s}^n)$ with $(\psi^n, \tilde{x}^n, \tilde{s}^n) \in \mathcal{T}^n_{p_{\Psi XS}}$ and an arbitrary permutation π, we have

$$\begin{aligned}
&E_{\Psi'^n}\left[W^n\left(\left(\bigcup_{\underline{\theta}\in\mathcal{P}_0(\mathcal{S}^n|\mathcal{X}^n)}\mathcal{T}^n_{\rho W_{\underline{\theta}},\delta}(\Psi'^n)\right)\middle|\tilde{x}^n,\tilde{s}^n\right)\right]\\
&\quad=\sum_{\psi'^n\in\mathcal{T}^n_P} p_{\Psi^n}(\psi'^n)\\
&\qquad W^n\left(\left(\bigcup_{\underline{\theta}\in\mathcal{P}_0(\mathcal{S}^n|\mathcal{X}^n)}\mathcal{T}^n_{\rho W_{\underline{\theta}},\delta}(\psi'^n)\right)\middle|\tilde{x}^n,\tilde{s}^n\right)\\
&\quad=\sum_{\psi'^n\in\mathcal{T}^n_P} p_{\Psi^n}(\psi'^n)\\
&\qquad W^n\left(\left(\bigcup_{\underline{\theta}\in\mathcal{P}_0(\mathcal{S}^n|\mathcal{X}^n)}\pi\left(\mathcal{T}^n_{\rho W_{\underline{\theta}},\delta}(\psi'^n)\right)\right)\middle|\pi(\tilde{x}^n,\tilde{s}^n)\right)\\
&\quad=\sum_{\psi'^n\in\mathcal{T}^n_P} p_{\Psi^n}(\psi'^n)\\
&\qquad W^n\left(\left(\bigcup_{\underline{\theta}\in\mathcal{P}_0(\mathcal{S}^n|\mathcal{X}^n)}\mathcal{T}^n_{\rho W_{\underline{\theta}},\delta}(\pi(\psi'^n))\right)\middle|\pi(\tilde{x}^n,\tilde{s}^n)\right)\\
&\quad\overset{(a)}{=}\sum_{\psi'^n\in\mathcal{T}^n_P} p_{\Psi^n}(\psi'^n)\\
&\qquad W^n\left(\left(\bigcup_{\underline{\theta}\in\mathcal{P}_0(\mathcal{S}^n|\mathcal{X}^n)}\mathcal{T}^n_{\rho W_{\underline{\theta}},\delta}(\psi'^n)\right)\middle|\pi(\tilde{x}^n,\tilde{s}^n)\right)\\
&\quad=E_{\Psi'^n}\left[W^n\left(\left(\bigcup_{\underline{\theta}\in\mathcal{P}_0(\mathcal{S}^n|\mathcal{X}^n)}\mathcal{T}^n_{\rho W_{\underline{\theta}},\delta}(\Psi'^n)\right)\middle|\pi(\tilde{x}^n,\tilde{s}^n)\right)\right],
\end{aligned}$$

where (a) follows because we sum up over all ψ'^n with the same type[5] (hence, $p_{\Psi^n}(\psi'^n)$ is identical for all ψ'^n of the same type).

Hence, we can rewrite the expectation as

$$\begin{aligned}
&E\left[W^n\left(\left(\bigcup_{\underline{\theta}\in\mathcal{P}_0(\mathcal{S}^n|\mathcal{X}^n)}\mathcal{T}^n_{\rho W_{\underline{\theta}},\delta}(\Psi'^n)\right)\middle|X^n,S^n\right)\right]\\
&\quad=\sum_{(\psi^n,x^n,s^n)\in\mathcal{T}^n_{p_{\Psi XS}}} p_{\Psi^n X^n S^n}(\psi^n,x^n,s^n)\\
&\qquad E_{\Psi'^n}\left[W\left(\left(\bigcup_{\underline{\theta}\in\mathcal{P}_0(\mathcal{S}^n|\mathcal{X}^n)}\mathcal{T}^n_{\rho W_{\underline{\theta}},\delta}(\Psi'^n)\right)\middle|x^n,s^n\right)\right]\\
&\quad=E\left[W\left(\left(\bigcup_{\underline{\theta}\in\mathcal{P}_0(\mathcal{S}^n|\mathcal{X}^n)}\mathcal{T}^n_{\rho W_{\underline{\theta}},\delta}(\Psi'^n)\right)\middle|\tilde{x}^n,\tilde{s}^n\right)\right],
\end{aligned}$$

for all $(\psi^n, \tilde{x}^n, \tilde{s}^n) \in \mathcal{T}^n_{p_{\Psi XS}}$. ∎

Appendix E
Proof of Lemma 1

Proof of Lemma 1: We consider both, the error probability and the information leakage. Let the maximum error probability and the information leakage, respectively, be given as

$$\begin{aligned}
&\hat{e}(\mathcal{K}_n) := \max_{f\in\mathcal{F}}\max_{j\in\mathcal{J}_n}\sum_{x^n\in\mathcal{X}^n} E(x^n|j)W^n(\mathcal{D}_j^c|x^n,f(x^n)),\\
&\max_{f\in\mathcal{F}}\max_{u\in\mathcal{U}_n} I(p_{J_n};E_uV_f^n).
\end{aligned}$$

Using the same (n, J_n) deterministic wiretap-code $\mathcal{K}_n$, as for the criteria above and considering now the maximization over $\theta \in \mathcal{P}(\mathcal{S}^n|\mathcal{X}^n)$ we can express the maximum error probability of transmitting one codeword as

$$\begin{aligned}
&\max_{j\in\mathcal{J}_n}\sum_{x^n\in\mathcal{X}^n} E(x^n|j)W_\theta^n(\mathcal{D}_j^c|x^n)\\
&\qquad=\max_{j\in\mathcal{J}_n}\sum_{x^n\in\mathcal{X}^n}\sum_{s^n\in\mathcal{S}^n} E(x^n|j)\theta(s^n|x^n)W^n(\mathcal{D}_j^c|x^n,s^n),
\end{aligned}$$

and hence we have

$$\begin{aligned}
&\max_{j\in\mathcal{J}_n}\sum_{x^n\in\mathcal{X}^n}\sum_{s^n\in\mathcal{S}^n} E(x^n|j)\theta(s^n|x^n)W^n(\mathcal{D}_j^c|x^n,s^n)\\
&\quad\le\max_{j\in\mathcal{J}_n}\sum_{x^n\in\mathcal{X}^n}\max_{s^n\in\mathcal{S}^n} E(x^n|j)W^n(\mathcal{D}_j^c|x^n,s^n)\\
&\quad=\max_{f\in\mathcal{F}}\max_{j\in\mathcal{J}_n}\sum_{x^n\in\mathcal{X}^n} E(x^n|j)W^n(\mathcal{D}_j^c|x^n,f(x^n))\\
&\quad=\hat{e}(\mathcal{K}_n)
\end{aligned}$$

Hence, even though the set of stochastic jamming strategies is larger than the set of deterministic jamming strategies, both will lead to the same error expression. Since

$$\begin{aligned}
E_uV_f^n(z^n|j) &= \sum_{x^n\in\mathcal{X}^n} E_u(x^n|j)V^n(z^n|x^n,f(x^n)),\\
V_\theta^n(z^n|x^n) &= \sum_{s^n\in\mathcal{S}^n}\theta(s^n|x^n)V^n(z^n|x^n,s^n),\\
E_uV_\theta^n(z^n|j) &= \sum_{x^n\in\mathcal{X}^n} E_u(x^n|j)\sum_{s^n\in\mathcal{S}^n}\theta(s^n|x^n)V^n(z^n|x^n,s^n),
\end{aligned}$$

for the leakage we can show that

$$\max_{f\in\mathcal{F}} I(p_{J_n};E_uV_f^n) = \max_{\theta\in\mathcal{P}(\mathcal{S}^n|\mathcal{X}^n)} I(p_{J_n};E_uV_{\theta^n}^n)$$

because the mutual information is convex in $V^n(z^n|x^n,s^n)$ for fixed input distribution. Hence, taking convex combinations of $V^n(z^n|x^n,s^n)$ does not increase the leakage term. Using Jensen's inequality and the fact that each value of $I(p_{J_n};E_uV_f^n)$ can also be achieved by $I(p_{J_n};E_uV_{\theta^n}^n)$, since the deterministic mappings $\mathcal{F}$ are a subset of the stochastic mappings $\mathcal{P}(\mathcal{S}^n|\mathcal{X}^n)$, $\mathcal{F} \subset \mathcal{P}(\mathcal{S}^n|\mathcal{X}^n)$, the equality is established, [18]. In other words, since the mutual information is a convex (row convex) function with respect to the conditional probability function of the output given the input for fixed input distribution, the optimal jamming strategy with respect to the secrecy constraint is achieved at the boundary of the probability polytope, i.e., is deterministic, [44, Proposition 2.4.1]. ∎

[5]Types are permutation invariant.

APPENDIX F
PROOF OF THEOREM 1

The extension from the standard AVWC to the case where the jammer additionally knows the channel input is not trivial. When using standard proof techniques from the AVWC, the jammer might be able to locate a channel input x^n to a specific deterministic wiretap-codebook $\mathcal{K}_n$. This automatically leads to the consideration of the deterministic code secrecy capacity of an AVWC under the maximum error criterion. Even without secrecy constraints, this problem remains unsolved, [10], [21]. To ensure that the confusion at the jammer with respect to the used codebook is sufficiently high, even if the channel input x^n is non-causally known, we fulfill an additional requirement in contrast to the standard AVWC. The used codewords x^n occur in multiple codebooks $\mathcal{K}_{n,\mathcal{U}_n}$, where $\mathcal{U}_n$ is the set of codebooks containing x^n as a codeword.

We use random coding arguments as in [16] and generate random sets of deterministic wiretap-codebooks. Note that we have to take into account that the jammer possesses non-causal knowledge about the channel input (and we allow knowledge of the messages, since we consider the maximum error), which results in a different error probability. When considering general WTCs, the capacity formulas depend on auxiliary RVs. One viewpoint to one of these auxiliary RVs is by introducing a prefix channel as part of the encoding process. For the prefixing, we follow [4, Lemma 4 and its proof], or [36, p.97, Addition of prefix channel] with slight modifications. In the original system model (Figure 1), the jammer knows the channel input X_u^n. If we concatenate a channel with the AVWC, and call the prefix variable Ψ_u^n, then the jammer does not know the channel input Ψ_u^n of the concatenated channel but an intermediate variable X^n, which is, in fact, the channel input of the original channel. However, we can adapt the codebook generation and decoding regions according to the concatenated channels ρW and ρV, respectively. For the secrecy analysis, we consider the mutual information $I(p_{J_n}; E_u V_{\theta^*}^n)$ and we have to show that the leakage to the eavesdropper vanishes asymptotically. Last, we show that the probability of obtaining codes for which both the decoding error probability and the leakage vanish asymptotically approaches one. For the converse, we modify the standard converse proof for the WTC.

For reasons that will become clear later in the proof, we choose an amount of CR that is lower bounded by

$$\begin{aligned}|\mathcal{U}_n| > \max\Bigg\{ & \frac{8}{\epsilon\tilde{p}} \log\left(|\mathcal{J}_n||\mathcal{L}_n||\mathcal{F}|\right), \\ & \exp\left\{n(H(X,\Psi) - R + \tilde{\delta})\right\} \\ & \left(\frac{\exp\{-nc'\delta'\}}{\lambda} \right. \\ & \left. + \frac{\exp\left\{-n\left(\min_{W\in\widehat{\mathcal{W}}} I(p_\Psi;\rho W) - R - \nu\right)\right\}}{\lambda} \right)^{-1} \Bigg\},\end{aligned} \tag{27}$$

for a $\tilde{p} > \exp\{-nc'\hat{\delta}^2\} + (n+1)^{|\mathcal{X}||\mathcal{S}|} \exp\left\{-n \min_{\theta\in\mathcal{P}_0(\mathcal{S}^n|\mathcal{X}^n)} h_\theta(\delta)\right\}$ (see Lemma 12).

A. Codebook Generation

We assume that for all $u \in \mathcal{U}_n$, $p_U(u) = \frac{1}{|\mathcal{U}_n|}$. Let $p \in \mathcal{P}(\Psi)$ be given. Partition the set of typical sequences $\mathcal{T}_{p,\delta}^n$ into disjoint subsets $C_{(j,l)}$ of size $|C_{(j,l)}| = \frac{|\mathcal{T}_{p,\delta}^n|}{|\mathcal{J}_n||\mathcal{L}_n|}$. Here $j \in \mathcal{J}_n = \{1,2,\dots,J_n\}$ and $l \in \mathcal{L}_n = \{1,2,\dots,L_n\}$ correspond to the secure and confusing messages, respectively. We have $J_n \cdot L_n = \exp\{nR\}$, and the transmission rate R will determined later. Let the random variable Ψ_{ujl}^n denote the codeword for the message pair $(j,l) \in \mathcal{J}_n \times \mathcal{L}_n$, if the CR has the realization $U = u$. The codewords Ψ_{ujl}^n and $\Psi_{u(jl)'}^n$ are independent of each other for all $(j,l) \neq (j,l)'$. Let $\hat{\chi} := \{\Psi_{ujl}^n : j \in \mathcal{J}_n, l \in \mathcal{L}_n, u \in \mathcal{U}_n\}$ be the family of RV, representing the random codewords. We start by generating a deterministic wiretap-code for each $u \in \mathcal{U}_n$ (still random in terms of random coding arguments). To indicate that each codebook at this point is a random variable, we add the argument $\hat{\chi}$. For each codebook $\mathcal{K}_{n,u}(\hat{\chi})$, we draw $J_n \cdot L_n$ codewords Ψ_{ujl}^n uniformly from the subsets $C_{(j,l)}$. For each Ψ_{ujl}^n we randomly choose X^n uniformly distributed over $\mathcal{T}_{\rho,\delta}^n(\Psi_{ujl}^n)$ as the channel input.

B. Decoding Regions

Let $\hat{\mathcal{D}}'_{ujl}(\hat{\chi})$ be given as

$$\hat{\mathcal{D}}'_{ujl}(\hat{\chi}) = \bigcup_{\underline{\theta}\in\mathcal{P}_0(\mathcal{S}^n|\mathcal{X}^n)} \mathcal{T}_{\rho W_{\underline{\theta}},\delta}^n(\Psi_{ujl}^n).$$

with[6] $(\rho W_{\underline{\theta}})(y|\psi) = \sum_{\substack{x\in\mathcal{X}\\ s\in\mathcal{S}}} \rho(x|\psi)\underline{\theta}(s|x)W)(y|x,s)$.

Then, we can define the decoding sets $\hat{\mathcal{D}}_{ujl}(\hat{\chi})$ as follows.

$$\hat{\mathcal{D}}_{ujl}(\hat{\chi}) = \hat{\mathcal{D}}'_{ujl}(\hat{\chi}) \bigcap \left(\bigcup_{\substack{(jl)'\in\mathcal{J}_n\times\mathcal{L}_n\\ (jl)'\neq(jl)}} \hat{\mathcal{D}}'_{u(jl)'}(\hat{\chi}) \right)^c. \tag{28}$$

For those sequences, which belong to multiple $\hat{\mathcal{D}}'_{u\cdot}(\hat{\chi})$ or to no $\hat{\mathcal{D}}'_{u\cdot}(\hat{\chi})$, a decoding error is declared.

C. Codebook Properties for Reliability

As already mentioned, we have to make sure, that every codeword occurs in multiple codebooks. By generating the codebooks $\mathcal{K}_{n,u}(\hat{\chi})$, $u \in \mathcal{U}_n$ as above, there are at most

$$\frac{|\mathcal{T}_{p,\delta}^n|}{J_n \cdot L_n} = \exp\{n(H(\Psi) - R + \epsilon_1(n))\}$$

non-overlapping codebooks in the worst case, where R corresponds to the code rate of a code with $J_n \cdot L_n$ messages. Intuitively, to ensure the occurrence of each codeword in k codebooks (on average), we should use an amount of CR which corresponds roughly to

$$|\mathcal{U}_n| \geq k \exp\{n(H(\Psi) - R + \epsilon_1(n))\}.$$

Later, we will derive a lower bound on the amount of CR, explicitly. We follow and extend the ideas of [16], [19],

[6] Note that $\underline{\theta}(s|x), x \in \mathcal{X}, s \in \mathcal{S}$ is a single-letter distribution on the set of all possible conditional types of s^n given x^n.

and [31]. Here, in contrast to the classical DMC, we have three error terms:

- given the received sequence Y^n, we do not find sequences Ψ^n_{ujl} and a channel input $X^n \in \mathcal{T}^n_{\rho,\delta}(\Psi^n_{ujl})$, such that Y^n is conditional typical given Ψ^n_{ujl} and $X^n \in \mathcal{T}^n_{\rho,\delta}(\Psi^n_{ujl})$,
- given the received sequence Y^n which is conditional typical given the codeword Ψ^n_{ujl} and the channel input $X^n \in \mathcal{T}^n_{\rho,\delta}(\Psi^n_{ujl})$, we find another codeword $\Psi^n_{u(jl)'}$ and channel input $X'^n \in \mathcal{T}^n_{\rho,\delta}(\Psi^n_{u(jl)'})$, such that Y^n is conditional typical given $\Psi^n_{u(jl)'}$ and $X'^n \in \mathcal{T}^n_{\rho,\delta}(\Psi^n_{u(jl)'})$,
- given the received sequence Y^n, there exist too many CR realizations u, such that for some messages $(j,l) \in \mathcal{J}_n \times \mathcal{L}_n$, the codeword $\Psi^n_{ujl} = \psi^n$, the channel input $X^n \in \mathcal{T}^n_{\rho,\delta}(\Psi^n_{ujl})$, $X^n = x^n$, and the state sequence $S^n = s^n$, the probability of $Y^n \in \hat{\mathcal{D}}^c_{ujl}(\hat{\chi})$ is lower bounded by some λ.

Since we apply random codes, we do actually not know which codebook realizations (in terms of random coding arguments) lead to a good error performance. But we know that the error probability vanishes averaged over a set of codebooks. Since the codewords occur in multiple codebooks, we have to take care of the situation that the codewords perform well in some codebooks, but not so well in others.

First, let us fix a pair $(j,l) \in \mathcal{J}_n \times \mathcal{L}_n$ and a $\underline{\theta} \in \mathcal{P}(\mathcal{S}|\mathcal{X})$. Arbitrarily choose a triple $(\psi^n, x^n, s^n) \in \mathcal{T}_{p_\Psi \times \rho \times \underline{\theta},\delta}$, with $\psi^n \in \mathcal{C}_{(j,l)}$, $x^n \in \mathcal{T}^n_{\rho,\delta'}(\psi^n)$ and $s^n \in \mathcal{T}^n_{\underline{\theta},\tilde{\delta}}(x^n)$. We have to show that if the sequence ψ^n is a codeword (occurring in multiple codebooks), then the state sequence is bad only for few codebooks, such that averaged over all codebooks, the error probability still vanishes. This has to hold for all pairs (j,l), sequences $\psi^n \in \mathcal{C}_{(j,l)}$, $x^n \in \mathcal{T}^n_{\rho,\delta}(\psi^n)$, and $s^n \in \mathcal{T}^n_{\underline{\theta},\tilde{\delta}}(x^n)$. We now can define the sets $\mathcal{U}(j,l,\psi^n,x^n,\hat{\chi})$ and $\mathcal{U}_0(j,l,\psi^n,x^n,s^n,\hat{\chi})$ as

$$\mathcal{U}(j,l,\psi^n,x^n,\hat{\chi}) := \{u : \Psi^n_{ujl} = \psi^n,\ X^n = x^n\},$$
$$\mathcal{U}_0(j,l,\psi^n,x^n,s^n,\hat{\chi}) := \Big\{u : \Psi^n_{ujl} = \psi^n,\ X^n = x^n, \text{ and } W^n(\hat{\mathcal{D}}^c_{ujl}(\hat{\chi})|x^n,s^n) > \lambda\Big\}.$$

Here, $\mathcal{U}(j,l,\psi^n,x^n,\hat{\chi})$ denotes the set of all codebooks, for which the sequence ψ^n is the codeword for the message pair (j,l) and x^n is the corresponding channel input, and $\mathcal{U}_0(j,l,\psi^n,x^n,s^n,\hat{\chi})$ is the set of all codebooks, for which the sequence ψ^n is the codeword for the message pair (j,l), x^n is the corresponding channel input, and the error bound λ is not met.

We can define the binary random variable $B(u,j,l,\psi^n,x^n,\hat{\chi})$ as

$$B(u,j,l,\psi^n,x^n,\hat{\chi}) = \begin{cases} 1 & \text{if} \quad u \in \mathcal{U}(j,l,\psi^n,x^n,\hat{\chi}) \\ 0 & \text{else.} \end{cases} \tag{29}$$

$$\begin{aligned} &Pr\{B(u,j,l,\psi^n,x^n,\hat{\chi}) = 1\} \\ &\quad = Pr\{\Psi^n_{ujl} = \psi^n\} Pr\{X^n = x^n | \Psi^n_{ujl} = \psi^n\} \\ &\quad = \frac{1}{|\mathcal{C}_{(j,l)}|} \frac{1}{|\mathcal{T}^n_{\rho,\delta}(\psi^n)|},\ \forall u \in \mathcal{U}_n,\ \forall (j,l) \in \mathcal{J}_n \times \mathcal{L}_n. \end{aligned} \tag{30}$$

It indicates whether the sequences ψ^n and x^n are the prefix variable and the channel input realizations for the codebook realization u and the message pair (j,l). By the Chernoff bound we obtain

$$\begin{aligned} &Pr\Big\{|\mathcal{U}(j,l,\psi^n,x^n,\hat{\chi})| \le \\ &\qquad (1-\epsilon_2)|\mathcal{U}_n| Pr\{B(u,j,l,\psi^n,x^n,\hat{\chi}) = 1\}\Big\} \\ &= Pr\Bigg\{ \sum_{u \in \mathcal{U}_n} B(u,j,l,\psi^n,x^n,\hat{\chi}) \le \\ &\qquad (1-\epsilon_2)|\mathcal{U}_n| Pr\{B(u,j,l,\psi^n,x^n,\hat{\chi}) = 1\}\Bigg\} \\ &\le \exp_e\left\{ -\frac{3\epsilon_2^2 |\mathcal{U}_n| J_n \cdot L_n \exp\{-n(H(X|\Psi)+\delta)\}}{8|\mathcal{T}^n_{p,\delta}|} \right\} \\ &\le \exp_e\left\{ -\frac{3}{8}\epsilon_2^2 |\mathcal{U}_n| \exp\{-n(H(X,\Psi) - R + \tilde{\delta})\} \right\}. \end{aligned}$$

Next, we will upper bound the probability that $|\mathcal{U}_0(j,l,\psi^n,x^n,s^n,\hat{\chi})|$ exceeds its expected value. We define the binary random variable $\tilde{B}(j,l,\psi^n,x^n,s^n,u,\lambda,\hat{\chi})$ as

$$\tilde{B}(j,l,\psi^n,x^n,s^n,u,\lambda,\hat{\chi}) = \begin{cases} 1 & \text{if} \quad u \in \mathcal{U}_0(j,l,\psi^n,x^n,s^n,\hat{\chi}) \\ 0 & \text{else.} \end{cases} \tag{31}$$

$$\begin{aligned} &Pr\left\{\tilde{B}(j,l,\psi^n,x^n,s^n,u,\lambda,\hat{\chi}) = 1\right\} \\ &= Pr\{B(u,j,l,\psi^n,x^n,\hat{\chi}) = 1\} \\ &\quad Pr\Big\{W^n(\hat{\mathcal{D}}^c_{ujl}(\hat{\chi})|x^n,s^n) > \lambda | B(u,j,l,\psi^n,x^n,\hat{\chi}) = 1\Big\}. \end{aligned} \tag{32}$$

It indicates whether the sequences ψ^n and x^n are the prefix variable and the channel input realizations for the codebook realization u and the message pair (j,l), and the error bound λ is not met.

We consider the case that the error bound is not met for a fixed $u \in \mathcal{U}_n$. By the Markov inequality Lemma 4 and by Lemma 13 we have

$$\begin{aligned} &Pr\big\{W^n(\hat{\mathcal{D}}^c_{ujl}(\hat{\chi})|x^n,s^n) > \lambda | B(u,j,l,\psi^n,x^n,\hat{\chi}) = 1\big\} \\ &\overset{(a)}{\le} \frac{E\left[W^n(\hat{\mathcal{D}}^c_{ujl}(\hat{\chi})|x^n,s^n) | B(u,j,l,\psi^n,x^n,\hat{\chi}) = 1\right]}{\lambda} \\ &\le \frac{1}{\lambda} E\Bigg[\Bigg(W^n\left(\hat{\mathcal{D}}'^c_{ujl}(\hat{\chi})|x^n,s^n\right) \\ &\qquad + W^n\Bigg(\bigcup_{\substack{(j,l)' \in \mathcal{J}_n \times \mathcal{L}_n \\ (j,l)' \neq (j,l)}} \hat{\mathcal{D}}'_{u(jl)'}(\hat{\chi})|x^n,s^n\Bigg)\Bigg) \\ &\qquad \Big| B(u,j,l,\psi^n,x^n,\hat{\chi}) = 1\Bigg] \end{aligned}$$

$$\overset{(b)}{\leq} \frac{1}{\lambda}\Bigg(\exp\{-nc'\delta'^2\} + \sum_{\substack{(j,l)'\in\mathcal{J}_n\times\mathcal{L}_n\\(j,l)'\neq(j,l)}} E\Big[\Big(W^n\Big(\hat{\mathcal{D}}'_{u(jl)'}(\hat{\chi})|x^n,s^n\Big)\Big) \Big| B(u,j,l,\psi^n,x^n,\hat{\chi})=1\Big]\Bigg)$$

$$\leq \frac{1}{\lambda}\Bigg(\exp\{-nc'\delta'^2\} + \sum_{\substack{(j,l)'\in\mathcal{J}_n\times\mathcal{L}_n\\(j,l)'\neq(j,l)}} E\Big[W^n\Big(\Big(\bigcup_{\underline{\theta}\in\mathcal{P}_0(\mathcal{S}^n|\mathcal{X}^n)}\mathcal{T}^n_{\rho W_{\underline{\theta}},\delta}(\Psi^n_{u(jl)'})\Big)\Big| x^n,s^n\Big) \Big| B(u,j,l,\psi^n,x^n,\hat{\chi})=1\Big]\Bigg)$$

$$\overset{(c)}{\leq} \frac{1}{\lambda}\Bigg(\exp\{-nc'\delta'^2\} + \sum_{\substack{(j,l)'\in\mathcal{J}_n\times\mathcal{L}_n\\(j,l)'\neq(j,l)}} \exp\Big\{-n\Big(\min_{W\in\widehat{\widehat{\mathcal{W}}}} I(p_\Psi;\rho W)-\nu\Big)\Big\}\Bigg)$$

$$\leq \frac{\exp\{-nc'\delta'^2\}}{\lambda} + \frac{\exp\Big\{-n\Big(\min_{W\in\widehat{\widehat{\mathcal{W}}}} I(p_\Psi;\rho W)-R-\nu\Big)\Big\}}{\lambda}.$$

Here, (a) follows by the Markov inequality (Lemma 4), (b) follows by Lemma 8 and the union bound, and (c) follows by Lemma 13 and the fact that $\Psi^n_{u(jl)'}$ and Ψ^n_{ujl} are independent of each other.

Then, identifying p_1 in Lemma 5 as

$$p_1 = \frac{\exp\{-nc'\delta'^2\}}{\lambda} + \frac{\exp\Big\{-n\Big(\min_{W\in\widehat{\widehat{\mathcal{W}}}} I(p_\Psi;\rho W)-R-\nu\Big)\Big\}}{\lambda},$$

we can bound the probability that $|\mathcal{U}_0(j,l,\psi^n,x^n,s^n,\hat{\chi})|$ exceeds a certain value as

$$Pr\big\{|\mathcal{U}_0(j,l,\psi^n,x^n,s^n,\hat{\chi})| \geq (1+\epsilon_2)|\mathcal{U}_n| Pr\{B(u,j,l,\psi^n,x^n,\hat{\chi})=1\}p_1\big\}$$

$$= Pr\Bigg\{\sum_{u\in\mathcal{U}_n}\tilde{B}(j,l,\psi^n,x^n,s^n,u,\lambda,\hat{\chi}) \geq (1+\epsilon_2)|\mathcal{U}_n| Pr\{B(u,j,l,\psi^n,x^n,\hat{\chi})=1\}p_1\Bigg\}$$

$$\leq \exp_e\Bigg\{-\frac{1}{8}\epsilon_2^2|\mathcal{U}_n|\exp\{-n(H(X,\Psi)-R+\tilde{\delta})\} \Bigg(\frac{\exp\{-nc'\delta'\}}{\lambda} + \frac{\exp\Big\{-n\Big(\min_{W\in\widehat{\widehat{\mathcal{W}}}} I(p_\Psi;\rho W)-R-\nu\Big)\Big\}}{\lambda}\Bigg)\Bigg\}.$$

Hence for all $|\mathcal{U}_n|$ fulfilling

$$|\mathcal{U}_n| > \exp\{n(H(X,\Psi)-R+\tilde{\delta})\} \Bigg(\frac{\exp\{-nc'\delta'\}}{\lambda} + \frac{\exp\Big\{-n\Big(\min_{W\in\widehat{\widehat{\mathcal{W}}}} I(p_\Psi;\rho W)-R-\nu\Big)\Big\}}{\lambda}\Bigg)^{-1}$$

the probabilities that codewords do not occur in at least $1-\epsilon_2$ times the expected number of codebooks and that codewords occur in more than $1+\epsilon_2$ times the expected number of codebooks for which the error bound is not met, vanish super exponentially fast.

The above described events have to hold for all $(j,l)\in\mathcal{J}_n\times\mathcal{L}_n$, $\psi^n\in\mathcal{C}_{(j,l)}$ $x^n\in\mathcal{T}^n_{\rho,\delta}(\psi^n)$ and $s^n\in\mathcal{S}^n$, for which there exists $\underline{\theta}\in\mathcal{P}_0(\mathcal{S}^n|\mathcal{X}^n)$: $(\psi^n,x^n,s^n)\in\mathcal{T}^n_{p_\Psi\rho_{X|\Psi}\underline{\theta},\delta}$. Hence,

$$Pr\Bigg\{\bigcap_{(j,l)\in\mathcal{J}_n\times\mathcal{L}_n}\ \bigcap_{\substack{\psi^n\in\mathcal{C}_{(j,l)}\\x^n\in\mathcal{T}^n_{\rho,\delta}(\psi^n)\\s^n\in\mathcal{S}^n\\\exists\underline{\theta}\in\mathcal{P}_0(\mathcal{S}^n|\mathcal{X}^n):(\psi^n,x^n,s^n)\in\mathcal{T}^n_{p_\Psi\rho_{X|\Psi}\underline{\theta},\delta}}}\Big\{|\mathcal{U}_0(j,l,\psi^n,x^n,s^n,\hat{\chi})|\leq (1+\epsilon_2)|\mathcal{U}_n|\cdot Pr\{B(u,j,l,\psi^n,x^n,\hat{\chi})=1\}p_1\Big\}\Bigg\}$$

$$= 1-Pr\Bigg\{\Bigg(\bigcap_{(j,l)\in\mathcal{J}_n\times\mathcal{L}_n}\ \bigcap_{\substack{\psi^n\in\mathcal{C}_{(j,l)}\\x^n\in\mathcal{T}^n_{\rho,\delta}(\psi^n)\\s^n\in\mathcal{S}^n\\\exists\underline{\theta}\in\mathcal{P}_0(\mathcal{S}^n|\mathcal{X}^n):(\psi^n,x^n,s^n)\in\mathcal{T}^n_{p_\Psi\rho_{X|\Psi}\underline{\theta},\delta}}}\Big\{|\mathcal{U}_0(j,l,\psi^n,x^n,s^n,\hat{\chi})|\leq (1+\epsilon_2)|\mathcal{U}_n|\cdot Pr\{B(u,j,l,\psi^n,x^n,\hat{\chi})=1\}p_1\Big\}\Bigg)^c\Bigg\}$$

$$\overset{(a)}{\geq} 1-|\mathcal{J}_n||\mathcal{L}_n|\frac{|\mathcal{T}^n_{p,\delta}|}{|\mathcal{J}_n||\mathcal{L}_n|}|\mathcal{T}^n_{\rho,\delta}||\mathcal{S}^n| \exp_e\Bigg\{-\frac{\epsilon_2^2|\mathcal{U}_n|\exp\{-n(H(X,\Psi)-R+\epsilon_1(n))\}}{8} \Bigg(\frac{\exp\{-nc'\delta'\}}{\lambda} + \frac{\exp\Big\{-n\Big(\min_{W\in\widehat{\widehat{\mathcal{W}}}} I(p_\Psi;\rho W)-R-\nu\Big)\Big\}}{\lambda}\Bigg)\Bigg\}$$

$$= 1-|\mathcal{T}^n_{p,\delta}||\mathcal{T}^n_{\rho,\delta}||\mathcal{S}^n| \exp_e\Bigg\{-\frac{\epsilon_2^2|\mathcal{U}_n|\exp\{-n(H(X,\Psi)-R+\epsilon_1(n))\}}{8} \Bigg(\frac{\exp\{-nc'\delta'\}}{\lambda} + \frac{\exp\Big\{-n\Big(\min_{W\in\widehat{\widehat{\mathcal{W}}}} I(p_\Psi;\rho W)-R-\nu\Big)\Big\}}{\lambda}\Bigg)\Bigg\}$$

and

$$Pr\Bigg\{\bigcap_{(j,l)\in\mathcal{J}_n\times\mathcal{L}_n}\ \bigcap_{\substack{\psi^n\in\mathcal{C}_{(j,l)}\\ x^n\in\mathcal{T}^n_{\rho,\delta}(\psi^n)\\ s^n\in\mathcal{S}^n\\ \exists\underline{\theta}\in\mathcal{P}_0(\mathcal{S}^n|\mathcal{X}^n):(\psi^n,x^n,s^n)\in\mathcal{T}^n_{p_\Psi\rho_{X|\Psi}\underline{\theta},\delta}}}\Big\{|\mathcal{U}(j,l,\psi^n,x^n,\hat{\chi})|\le (1-\epsilon_2)|\mathcal{U}_n|Pr\{B(u,j,l,\psi^n,x^n,\hat{\chi})=1\}\Big\}\Bigg\}$$

$$=1-Pr\Bigg\{\Bigg(\bigcap_{(j,l)\in\mathcal{J}_n\times\mathcal{L}_n}\ \bigcap_{\substack{\psi^n\in\mathcal{C}_{(j,l)}\\ x^n\in\mathcal{T}^n_{\rho,\delta}(\psi^n)\\ s^n\in\mathcal{S}^n\\ \exists\underline{\theta}\in\mathcal{P}_0(\mathcal{S}^n|\mathcal{X}^n):(\psi^n,x^n,s^n)\in\mathcal{T}^n_{p_\Psi\rho_{X|\Psi}\underline{\theta},\delta}}}\Big\{|\mathcal{U}(j,l,\psi^n,x^n,\hat{\chi})| \le (1-\epsilon_2)|\mathcal{U}_n|Pr\{B(u,j,l,\psi^n,x^n,\hat{\chi})=1\}\Big\}\Bigg)^c\Bigg\}$$

$$\overset{(b)}{\ge}1-|\mathcal{J}_n||\mathcal{L}_n|\frac{|\mathcal{T}^n_{p,\delta}|}{|\mathcal{J}_n||\mathcal{L}_n|}|\mathcal{T}^n_{\rho,\delta}||\mathcal{S}^n| \exp_e\left\{-\frac{3\epsilon_2^2|\mathcal{U}_n|J_n\cdot L_n}{8|\mathcal{T}^n_{p,\delta}||\mathcal{T}^n_{\rho,\delta}|}\right\}$$

$$=1-|\mathcal{T}^n_{p,\delta}||\mathcal{T}^n_{\rho,\delta}||\mathcal{S}^n|\exp_e\left\{-\frac{3\epsilon_2^2|\mathcal{U}_n|J_n\cdot L_n}{8|\mathcal{T}^n_{p,\delta}||\mathcal{T}^n_{\rho,\delta}|}\right\}.$$

Here, (a) and (b) follow by the union bound and summing over all $(j,l)\in\mathcal{J}_n\times\mathcal{L}_n$, $\psi^n\in\mathcal{C}_{(j,l)}$ $x^n\in\mathcal{T}^n_{\rho,\delta}(\psi^n)$ and $s^n\in\mathcal{S}^n$. Next, we will show that for all $(j,l)\in\mathcal{J}_n\times\mathcal{L}_n$ and all $f\in\mathcal{F}$ the amount of CR realizations for which Lemma 12 does not hold is small compared to the total amount of CR. First, fix a pair $(j,l)\in\mathcal{J}_n\times\mathcal{L}_n$ and a jamming strategy $f\in\mathcal{F}$ and let $\epsilon>0$. Furthermore, let $\tilde{p}>\exp\{-nc'\hat{\delta}^2\}+(n+1)^{|\mathcal{X}||\mathcal{S}|}\exp\left\{-n\min_{\theta\in\mathcal{P}_0(\mathcal{S}^n|\mathcal{X}^n)}h_\theta(\delta)\right\}$ (see Lemma 12). Then by the Chernoff bounds, Lemma 5, we can compute

$$Pr\Big\{\Big|\Big\{u\in\mathcal{U}_n:\ \nexists\theta\in\mathcal{P}_0(\mathcal{S}^n|\mathcal{X}^n)\ :(\Psi^n_{ujl},X^n,f(X^n))\in\mathcal{T}^n_{p_\Psi\rho_{X|\Psi}\underline{\theta},\delta}\Big\}\Big|>(1+\epsilon)|\mathcal{U}_n|\tilde{p}\Big\}<\exp_e\left\{-\frac{\epsilon|\mathcal{U}_n|\tilde{p}}{8}\right\}$$

Furthermore, the probability that this property holds for all $(j,l)\in\mathcal{J}_n\times\mathcal{L}_n$ and all $f\in\mathcal{F}$ can be lower bounded by the union bound to

$$Pr\Bigg\{\bigcap_{\substack{(j,l)\in\mathcal{J}_n\times\mathcal{L}_n\\ f\in\mathcal{F}}}\Big\{\Big|\{u\in\mathcal{U}_n:\ \nexists\theta\in\mathcal{P}_0(\mathcal{S}^n|\mathcal{X}^n)\ :(\Psi^n_{ujl},X^n,f(X^n))\in\mathcal{T}^n_{p_\Psi\rho_{X|\Psi}\underline{\theta},\delta}\}\Big|\le(1+\epsilon)|\mathcal{U}_n|\tilde{p}\Big\}\Bigg\}$$

$$=1-Pr\Bigg\{\Bigg(\bigcap_{\substack{(j,l)\in\mathcal{J}_n\times\mathcal{L}_n\\ f\in\mathcal{F}}}\Big\{\Big|\{u\in\mathcal{U}_n:\ \nexists\theta\in\mathcal{P}_0(\mathcal{S}^n|\mathcal{X}^n)\ :(\Psi^n_{ujl},X^n,f(X^n))\in\mathcal{T}^n_{p_\Psi\rho_{X|\Psi}\underline{\theta},\delta}\}\Big|\le(1+\epsilon)|\mathcal{U}_n|\tilde{p}\Big\}\Bigg)^c\Bigg\}$$

$$\ge 1-\bigcup_{\substack{(j,l)\in\mathcal{J}_n\times\mathcal{L}_n\\ f\in\mathcal{F}}}Pr\Big\{\Big|\{u\in\mathcal{U}_n:\ \nexists\theta\in\mathcal{P}_0(\mathcal{S}^n|\mathcal{X}^n)\ :(\Psi^n_{ujl},X^n,f(X^n))\in\mathcal{T}^n_{p_\Psi\rho_{X|\Psi}\underline{\theta},\delta}\}\Big|>(1+\epsilon)|\mathcal{U}_n|\tilde{p}\Big\}\Big\}$$

$$\ge 1-|\mathcal{J}_n||\mathcal{L}_n||\mathcal{F}|\exp_e\left\{-\frac{\epsilon|\mathcal{U}_n|\tilde{p}}{8}\right\},$$

which approaches 1 super exponentially fast in n by our choice of $|\mathcal{U}_n|$ in (27).

D. Codebook Realization

Now, let $\mathcal{K}^{\text{ran}}_n$ be a codebook realization of $\mathcal{K}^{\text{ran}}_n(\hat{\chi})$, fulfilling the aforementioned properties (codewords occur in sufficiently many (deterministic) codebooks, indexed by the realization of the CR, and are bad only for few, and codewords together with channel inputs and state sequences possess a Markov structure with high probability for all jamming strategies), with $\mathcal{D}'_{ujl}$ as

$$\hat{\mathcal{D}}'_{ujl}=\bigcup_{\underline{\theta}\in\mathcal{P}_0(\mathcal{S}^n|\mathcal{X}^n)}\mathcal{T}^n_{\rho W_{\underline{\theta}},\delta}(\psi^n_{ujl}).$$

with[7] $(\rho W_{\underline{\theta}})(y|\psi)=\sum_{\substack{x\in\mathcal{X}\\ s\in\mathcal{S}}}\rho(x\psi)\underline{\theta}(s|x)W)(y|x,s)$ and decoding sets $\mathcal{D}_{ujl}$, being as follows.

$$\mathcal{D}_{ujl}=\mathcal{D}'_{ujl}\bigcap\left(\bigcup_{\substack{(jl)'\in\mathcal{J}_n\times\mathcal{L}_n\\ (jl)\neq(jl)'}}\mathcal{D}'_{u(jl)'}\right)^c \tag{33}$$

For those sequences, which belong to multiple $\hat{\mathcal{D}}'_{u\cdot}$ or to no $\hat{\mathcal{D}}'_{u\cdot}$, a decoding error is declared.

E. Adaptation of the Error Criterion

We will modify the error criterion and require that both the secret message J and the confusing message L should be successfully decoded at Bob.

Hence, we have

$$\max_{\substack{j\in\mathcal{J}_n,\\ l\in\mathcal{L}_n,\\ f\in\mathcal{F}}}\sum_{u\in\mathcal{U}_n}p_U(u)\sum_{\psi^n\in\Psi^n}E_u(\psi^n|j)\sum_{x^n\in\mathcal{T}^n_{\rho,\delta}(\psi^n)}\frac{W^n(\mathcal{D}^c_{ujl}|x^n,f(x^n))}{|\mathcal{T}^n_{\rho,\delta}(\psi^n)|}$$

$$=\max_{\substack{j\in\mathcal{J}_n,\\ l\in\mathcal{L}_n,\\ f\in\mathcal{F}}}\sum_{\substack{\psi^n\in\Psi^n\\ x^n\in\mathcal{T}^n_{\rho,\delta}(\psi^n)\\ u\in\mathcal{U}_n}}\frac{p_U(u)E(\psi^n|j,l,u)W^n(\mathcal{D}^c_{ujl}|x^n,f(x^n))}{|\mathcal{T}^n_{\rho,\delta}(\psi^n)|}$$

$$=\max_{j\in\mathcal{J}_n}\max_{l\in\mathcal{L}_n}\max_{f\in\mathcal{F}}\sum_{\substack{\psi^n\in\Psi^n\\ x^n\in\mathcal{T}^n_{\rho,\delta}(\psi^n)\\ u\in\mathcal{U}_n:\\ \exists\underline{\theta}\in\mathcal{P}_0(\mathcal{S}^n|\mathcal{X}^n):(\psi^n,x^n,f(x^n))\in\mathcal{T}^n_{p_\Psi\rho_{X|\Psi}\underline{\theta},\delta}}}\frac{p_U(u)E(\psi^n|j,l,u)W^n(\mathcal{D}^c_{ujl}|x^n,f(x^n))}{|\mathcal{T}^n_{\rho,\delta}(\psi^n)|}$$

$$+\max_{j\in\mathcal{J}_n}\max_{l\in\mathcal{L}_n}\max_{f\in\mathcal{F}}\sum_{\substack{\psi^n\in\Psi^n\\ x^n\in\mathcal{T}^n_{\rho,\delta}(\psi^n)\\ u\in\mathcal{U}_n:\\ \nexists\underline{\theta}\in\mathcal{P}_0(\mathcal{S}^n|\mathcal{X}^n):(\psi^n,x^n,f(x^n))\in\mathcal{T}^n_{p_\Psi\rho_{X|\Psi}\underline{\theta},\delta}}}\frac{p_U(u)E(\psi^n|j,l,u)W^n(\mathcal{D}^c_{ujl}|x^n,f(x^n))}{|\mathcal{T}^n_{\rho,\delta}(\psi^n)|}$$

[7]Note that $\underline{\theta}(s|x),x\in\mathcal{X},s\in\mathcal{S}$ is a single-letter distribution on the set of all possible conditional types of s^n given x^n.

$$\le \max_{j\in\mathcal{J}_n}\max_{l\in\mathcal{L}_n}\max_{f\in\mathcal{F}}\sum_{\substack{\psi^n\in\Psi^n\\ x^n\in\mathcal{T}^n_{\rho,\delta}(\psi^n)\\ u\in\mathcal{U}_n:\\ \exists\underline{\theta}\in\mathcal{P}_0(\mathcal{S}^n|\mathcal{X}^n):(\psi^n,x^n,f(x^n))\in\mathcal{T}^n_{p_\Psi\rho_{X|\Psi}\underline{\theta},\delta}}}\frac{p_U(u)E(\psi^n|j,l,u)W^n(\mathcal{D}^c_{ujl}|x^n,f(x^n))}{|\mathcal{T}^n_{\rho,\delta}(\psi^n)|}$$

$$+\max_{j\in\mathcal{J}_n}\max_{l\in\mathcal{L}_n}\max_{f\in\mathcal{F}}\sum_{\substack{\psi^n\in\Psi^n\\ x^n\in\mathcal{T}^n_{\rho,\delta}(\psi^n)\\ u\in\mathcal{U}_n:\\ \exists\underline{\theta}\in\mathcal{P}_0(\mathcal{S}^n|\mathcal{X}^n):(\psi^n,x^n,f(x^n))\in\mathcal{T}^n_{p_\Psi\rho_{X|\Psi}\underline{\theta},\delta}}}\frac{E(\psi^n|j,l,u)}{|\mathcal{T}^n_{\rho,\delta}(\psi^n)|}\frac{1}{|\mathcal{U}_n|}$$

$$\left|\{u\in\mathcal{U}_n:\ \nexists\theta\in\mathcal{P}_0(\mathcal{S}^n|\mathcal{X}^n)\ :\ (\psi^n_{ujl},x^n,f(x^n))\in\mathcal{T}^n_{p_\Psi\times\rho_{X|\Psi}\times\underline{\theta},\delta}\}\right|$$

$$\overset{(a)}{\le}\max_{j\in\mathcal{J}_n}\max_{l\in\mathcal{L}_n}\max_{f\in\mathcal{F}}\sum_{\substack{\psi^n\in\Psi^n\\ x^n\in\mathcal{T}^n_{\rho,\delta}(\psi^n)\\ u\in\mathcal{U}_n:\\ \exists\underline{\theta}\in\mathcal{P}_0(\mathcal{S}^n|\mathcal{X}^n):(\psi^n,x^n,f(x^n))\in\mathcal{T}^n_{p_\Psi\rho_{X|\Psi}\underline{\theta},\delta}}}\frac{p_U(u)E(\psi^n|j,l,u)W^n(\mathcal{D}^c_{ujl}|x^n,f(x^n))}{|\mathcal{T}^n_{\rho,\delta}(\psi^n)|}$$

$$+\frac{(1+\epsilon)|\mathcal{U}_n|\tilde{p}}{|\mathcal{U}_n|}$$

$$=\max_{j\in\mathcal{J}_n}\max_{l\in\mathcal{L}_n}\max_{f\in\mathcal{F}}\sum_{\substack{\psi^n\in\Psi^n\\ x^n\in\mathcal{T}^n_{\rho,\delta}(\psi^n)\\ u\in\mathcal{U}_n:\\ \exists\underline{\theta}\in\mathcal{P}_0(\mathcal{S}^n|\mathcal{X}^n):(\psi^n,x^n,f(x^n))\in\mathcal{T}^n_{p_\Psi\rho_{X|\Psi}\underline{\theta},\delta}}}\frac{p_U(u)E(\psi^n|j,l,u)W^n(\mathcal{D}^c_{ujl}|x^n,f(x^n))}{|\mathcal{T}^n_{\rho,\delta}(\psi^n)|}$$

$$+(1+\epsilon)\tilde{p}$$

$$\le\max_{j\in\mathcal{J}_n}\max_{l\in\mathcal{L}_n}\max_{f\in\mathcal{F}}\sum_{\substack{\psi^n\in\Psi^n\\ x^n\in\mathcal{T}^n_{\rho,\delta}(\psi^n)\\ u\in\mathcal{U}_n:\\ \exists\underline{\theta}\in\mathcal{P}_0(\mathcal{S}^n|\mathcal{X}^n):(\psi^n,x^n,f(x^n))\in\mathcal{T}^n_{p_\Psi\rho_{X|\Psi}\underline{\theta},\delta}}}\Big(p_{UJL\Psi^nX^n}(u,j,l,\psi^n,x^n)$$

$$W^n(\mathcal{D}^c_{u,j,l}|x^n,f(x^n))\Big)+(1+\epsilon)\tilde{p}$$

$$=\max_{j\in\mathcal{J}_n}\max_{l\in\mathcal{L}_n}\max_{f\in\mathcal{F}}\sum_{\substack{\psi^n\in\Psi^n\\ x^n\in\mathcal{T}^n_{\rho,\delta}(\psi^n)\\ u\in\mathcal{U}_n:\\ \exists\underline{\theta}\in\mathcal{P}_0(\mathcal{S}^n|\mathcal{X}^n):(\psi^n,x^n,f(x^n))\in\mathcal{T}^n_{p_\Psi\rho_{X|\Psi}\underline{\theta},\delta}}}\Big(p_{U|JL\Psi^nX^n}(u|j,l,\psi^n,x^n)$$

$$p_{JL\Psi^nX^n}(j,l,\psi^n,x^n)W^n(\mathcal{D}^c_{ujl}|x^n,f(x^n))\Big)+(1+\epsilon)\tilde{p}$$

$$\le\max_{j\in\mathcal{J}_n}\max_{l\in\mathcal{L}_n}\max_{f\in\mathcal{F}}\sum_{\substack{\psi^n\in\Psi^n\\ x^n\in\mathcal{T}^n_{\rho,\delta}(\psi^n)\\ u\in\mathcal{U}_n:\\ \exists\underline{\theta}\in\mathcal{P}_0(\mathcal{S}^n|\mathcal{X}^n):(\psi^n,x^n,f(x^n))\in\mathcal{T}^n_{p_\Psi\rho_{X|\Psi}\underline{\theta},\delta}}}\Big(p_{U|\Psi^nX^nJL}(u|\psi^n,x^n,j,l)$$

$$W^n(\mathcal{D}^c_{ujl}|x^n,f(x^n))\Big)+(1+\epsilon)\tilde{p}$$

$$\le\max_{j\in\mathcal{J}_n}\max_{l\in\mathcal{L}_n}\max_{\substack{\psi^n\in\Psi^n\\ x^n\in\mathcal{T}^n_{\rho,\delta}(\psi^n)}}\sum_{u\in\mathcal{U}_n}\Big(p_{U|\Psi^nX^nJL}(u|\psi^n,x^n,j,l)$$

$$\exists\underline{\theta}\in\mathcal{P}_0(\mathcal{S}^n|\mathcal{X}^n):(\psi^n,x^n,s^n)\in\mathcal{T}^n_{p_\Psi\rho_{X|\Psi}\underline{\theta},\delta}\ s^n\in\mathcal{S}^n$$

$$W^n(\mathcal{D}^c_{ujl}|x^n,s^n)\Big)+(1+\epsilon)\tilde{p}$$

$$:=\hat{e}(\mathcal{K}^{\text{ran}}_n)$$

We first split the error probability into two terms with respect to those CR realizations for which the sequences $(\psi^n,x^n,f(x^n))$ fulfill the property of Lemma 12 or not. In the first term, there exists a $\underline{\theta}\in\mathcal{P}_0(\mathcal{S}^n|\mathcal{X}^n):(\psi^n,x^n,f(x^n))\in\mathcal{T}_{p_\Psi\rho_{X|\Psi}\underline{\theta},\delta}$, in the second term there does not exist such a $\underline{\theta}\in\mathcal{P}_0(\mathcal{S}^n|\mathcal{X}^n)$. Here, we have implicitly shown in Appendix F-C, that (a) follows with probability approaching 1, where Lemma 12 is applied.

Secondly, we consider the maximization over all terms (ψ^n,x^n,s^n). Our motivation to do so is to reduce the size of the space, over which should be optimized. The family $\mathcal{F}=\{f:\mathcal{X}^n\to\mathcal{S}^n\}$ consists of $|\mathcal{F}|=|\mathcal{S}^n|^{|\mathcal{X}^n|}$ elements, hence it grows doubly exponentially with n. By considering the maximum with respect to x^n, it is sufficient to consider the state sequence s^n maximizing the error probability. Hence, we can reduce the space size used for optimization to $\mathcal{X}^n\times\mathcal{S}^n$, which grows only exponentially in n.

F. Error Analysis

For the error probability we can overall conclude

$$\hat{e}(\mathcal{K}^{\text{ran}}_n)=(1+\epsilon)\tilde{p}$$

$$+\max_{j\in\mathcal{J}_n}\max_{l\in\mathcal{L}_n}\max_{\substack{\psi^n\in\Psi^n\\ x^n\in\mathcal{T}^n_{\rho,\delta}(\psi^n)}}\sum_{u\in\mathcal{U}_n}\Big(p_{U|\Psi^nX^nJL}(u|\psi^n,x^n,j,l)$$

$$\exists\underline{\theta}\in\mathcal{P}_0(\mathcal{S}^n|\mathcal{X}^n):(\psi^n,x^n,s^n)\in\mathcal{T}^n_{p_\Psi\rho_{X|\Psi}\underline{\theta},\delta}\ s^n\in\mathcal{S}^n$$

$$W^n(\mathcal{D}^c_{ujl}|x^n,s^n)\Big)$$

$$=(1+\epsilon)\tilde{p}+\max_{j\in\mathcal{J}_n}\max_{l\in\mathcal{L}_n}\max_{\substack{\psi^n\in\Psi^n\\ x^n\in\mathcal{T}^n_{\rho,\delta}(\psi^n)}}\Bigg($$

$$\sum_{u\in\mathcal{U}_0^c(j,l,\psi^n,x^n,s^n)}\Big(p_{U|\Psi^nX^nJL}(u|\psi^n,x^n,j,l)W^n(\mathcal{D}^c_{ujl}|x^n,s^n)\Big)$$

$$+\sum_{u\in\mathcal{U}_0(j,\psi^n,l,x^n,s^n)}\Big(p_{U|\Psi^nX^nJL}(u|\psi^n,x^n,j,l)W^n(\mathcal{D}^c_{ujl}|x^n,s^n)\Big)\Bigg)$$

$$\le(1+\epsilon)\tilde{p}+\lambda+\max_{j\in\mathcal{J}_n}\max_{l\in\mathcal{L}_n}\max_{\substack{\psi^n\in\Psi^n\\ x^n\in\mathcal{T}^n_{\rho,\delta}(\psi^n)}}\Bigg($$

$$\exists\underline{\theta}\in\mathcal{P}_0(\mathcal{S}^n|\mathcal{X}^n):(\psi^n,x^n,s^n)\in\mathcal{T}^n_{p_\Psi\rho_{X|\Psi}\underline{\theta},\delta}\ s^n\in\mathcal{S}^n$$

$$\sum_{u\in\mathcal{U}_0(j,\psi^n,l,x^n,s^n)}p_{U|\Psi^nX^nJL}(u|\psi^n,x^n,j,l)W^n(\mathcal{D}^c_{ujl}|x^n,s^n)\Big)$$

$$\le(1+\epsilon)\tilde{p}+\lambda+\max_{j\in\mathcal{J}_n}\max_{l\in\mathcal{L}_n}\max_{\substack{\psi^n\in\Psi^n\\ x^n\in\mathcal{T}^n_{\rho,\delta}(\psi^n)}}\Bigg($$

$$\exists\underline{\theta}\in\mathcal{P}_0(\mathcal{S}^n|\mathcal{X}^n):(\psi^n,x^n,s^n)\in\mathcal{T}^n_{p_\Psi\rho_{X|\Psi}\underline{\theta},\delta}\ s^n\in\mathcal{S}^n$$

$$\sum_{u\in\mathcal{U}_0(j,\psi^n,l,x^n,s^n)}\frac{p_{U\Psi^nX^nJL}(u,\psi^n,x^n,j,l)}{p_{\Psi^nX^nJL}(\psi^n,x^n,j,l)}\Big)$$

$$=(1+\epsilon)\tilde{p}+\lambda+\max_{j\in\mathcal{J}_n}\max_{l\in\mathcal{L}_n}\max_{\substack{\psi^n\in\Psi^n\\ x^n\in\mathcal{T}^n_{\rho,\delta}(\psi^n)}}\Bigg($$

$$\exists\underline{\theta}\in\mathcal{P}_0(\mathcal{S}^n|\mathcal{X}^n):(\psi^n,x^n,s^n)\in\mathcal{T}^n_{p_\Psi\rho_{X|\Psi}\underline{\theta},\delta}\ s^n\in\mathcal{S}^n$$

$$\left.\frac{\sum_{u\in\mathcal{U}_0(j,\psi^n,l,x^n,s^n)} p_{U\Psi^nX^nJL}(u,\psi^n,x^n,j,l)}{\sum_{u'\in\mathcal{U}(j,l,\psi^n,x^n)} p_{U\Psi^nX^nJL}(u',\psi^n,x^n,j,l)}\right)$$

$$= (1+\epsilon)\tilde{p} + \lambda + \max_{j\in\mathcal{J}_n}\max_{l\in\mathcal{L}_n}\max_{\substack{\psi^n\in\Psi^n\\ x^n\in\mathcal{T}^n_{\rho,\delta}(\psi^n)\\ s^n\in\mathcal{S}^n:\\ \exists\underline{\theta}\in\mathcal{P}_0(\mathcal{S}^n|\mathcal{X}^n):(\psi^n,x^n,s^n)\in\mathcal{T}^n_{p_{\Psi}\rho_{X|\Psi}\underline{\theta},\delta}}}$$

$$\frac{\sum_{u\in\mathcal{U}_0(j,\psi^n,l,x^n,s^n)} p_U(u)p_{\Psi^n|UJL}(\psi^n|u,j,l)p_{X^n|\Psi^n}(x^n|\psi^n)}{\sum_{u'\in\mathcal{U}(j,l,\psi^n,x^n)} p_U(u')p_{\Psi^n|UJL}(\psi^n|u',j,l)p_{X^n|\Psi^n}(x^n|\psi^n)}$$

$$= (1+\epsilon)\tilde{p} + \lambda + \max_{j\in\mathcal{J}_n}\max_{l\in\mathcal{L}_n}\max_{\substack{\psi^n\in\Psi^n\\ x^n\in\mathcal{T}^n_{\rho,\delta}(\psi^n)\\ s^n\in\mathcal{S}^n:\\ \exists\underline{\theta}\in\mathcal{P}_0(\mathcal{S}^n|\mathcal{X}^n):(\psi^n,x^n,s^n)\in\mathcal{T}^n_{p_{\Psi}\rho_{X|\Psi}\underline{\theta},\delta}}}$$

$$\frac{\sum_{u\in\mathcal{U}_0(j,\psi^n,l,x^n,s^n)} p_U(u)}{\sum_{u'\in\mathcal{U}(j,l,\psi^n,x^n)} p_U(u')}$$

$$= (1+\epsilon)\tilde{p} + \lambda + \max_{j\in\mathcal{J}_n}\max_{l\in\mathcal{L}_n}\max_{\substack{\psi^n\in\Psi^n\\ x^n\in\mathcal{T}^n_{\rho,\delta}(\psi^n)\\ s^n\in\mathcal{S}^n:\\ \exists\underline{\theta}\in\mathcal{P}_0(\mathcal{S}^n|\mathcal{X}^n):(\psi^n,x^n,s^n)\in\mathcal{T}^n_{p_{\Psi}\rho_{X|\Psi}\underline{\theta},\delta}}} \frac{|\mathcal{U}_0(j,l,\psi^n,x^n,s^n)|}{|\mathcal{U}(j,l,\psi^n,x^n)|}.$$

In Appendix F-C, we have implicitly shown, that the probability

$$Pr\left\{\frac{|\mathcal{U}_0(j,l,\psi^n,x^n,s^n)|}{|\mathcal{U}(j,l,\psi^n,x^n)|} \geq \frac{(1+\epsilon_2)|\mathcal{U}_n|Pr\{B(u,j,l,\psi^n,x^n,\hat{\chi})=1\}p_1}{(1-\epsilon_2)|\mathcal{U}_n|Pr\{B(u,j,l,\psi^n,x^n,\hat{\chi})=1\}}\right\}$$

vanishes super exponentially fast. Hence, with probability 1, we can upper bound $\hat{\hat{e}}(\mathcal{K}_n^{\text{ran}})$ as

$$\hat{\hat{e}}(\mathcal{K}_n^{\text{ran}}) \leq (1+\epsilon)\tilde{p} + \lambda + \max_{j\in\mathcal{J}_n}\max_{l\in\mathcal{L}_n}\max_{\substack{\psi^n\in\Psi^n\\ x^n\in\mathcal{T}^n_{\rho,\delta}(\psi^n)\\ s^n\in\mathcal{S}^n\\ \exists\underline{\theta}\in\mathcal{P}_0(\mathcal{S}^n|\mathcal{X}^n):(\psi^n,x^n,s^n)\in\mathcal{T}^n_{p_{\Psi}\rho_{X|\Psi}\underline{\theta},\delta}}}\left(\frac{(1+\epsilon_2)|\mathcal{U}_n|Pr\{B(u,j,l,\psi^n,x^n,\hat{\chi})=1\}p_1}{(1-\epsilon_2)|\mathcal{U}_n|Pr\{B(u,j,l,\psi^n,x^n,\hat{\chi})=1\}}\right)$$

$$= (1+\epsilon)\tilde{p} + \lambda + \frac{1+\epsilon_2}{1-\epsilon_2}\left(\frac{\exp\{-nc'\delta'\}}{\lambda} + \frac{\exp\left\{-n\left(\min_{W\in\widehat{\mathcal{W}}} I(p_\Psi;\rho W) - R - \nu\right)\right\}}{\lambda}\right)$$

We choose

$$R \leq \min_{W\in\widehat{\mathcal{W}}} I(p_\Psi;\rho W) - \nu$$

$$\lambda = \exp\{-n\frac{\tau}{2}\},$$

$$\tau < \min\left\{c'\delta',\ \min_{W\in\widehat{\mathcal{W}}} I(p_\Psi;\rho W) - R - \nu\right\}$$

and have shown an arbitrarily small error probability.

G. Codebook Properties for Secure Communication

We have to show that the leakage to the eavesdropper vanishes asymptotically. Therefore, we make use of the fact that there exists a best channel to the eavesdropper and the fact that the probability that the implied probability distributions are not in an ϵ region around the expected typical ones can be upper bounded using Chernoff bounds. Then we apply Lemma 3. If the variation distance of the channel output probability distribution and the conditional channel output probability distribution can be upper bounded, then the leakage can be upper bounded as well. To upper bound the variation distance, the triangle inequality will be used in combination with properties of typical sequences. Note that the existence of a best channel to the eavesdropper is crucial at this point to reduce the jammer's possible choices of jamming sequence from double exponentially many to exactly one, for the case of a best channel to the eavesdropper.

Notice that in contrast to the error analysis we do not average with respect to the CR when considering the leakage. In other words, the leakage has to vanish for all $u \in \mathcal{U}_n$, hence we will omit indexing on u. Operationally, that means the eavesdropper may have access to the CR. It is sufficient to consider the best channel to the eavesdropper, invoked by $\theta^{*,n} \in \mathcal{P}^n(\mathcal{S}|\mathcal{X})$[8], since fulfilling the secrecy requirement for the best channel to the eavesdropper implies that the secrecy requirement is fulfilled for all other channels to the eavesdropper by the data processing inequality, as well.

1) Relation to Total Variation Distance: For a fixed $u \in \mathcal{U}_n$, we have

$$I(p_{J_n}; E_uV^n_{\theta^{*,n}}) = H(p_{J_n}E_uV^n_{\theta^{*,n}}) - H(E_uV^n_{\theta^{*,n}}|p_{J_n})$$
$$(= H(Z^n_{\theta^{*,n}}) - H(Z^n_{\theta^{*,n}}|J))$$
$$= \frac{1}{J_n}\sum_{j\in\mathcal{J}_n}\left(H(p_{J_n}E_uV^n_{\theta^{*,n}}) - H(E_uV^n_{\theta^{*,n}}|j)\right)$$
$$= \frac{1}{J_n}\sum_{j\in\mathcal{J}_n}\left(H\Big(\frac{1}{J_n}\sum_{\substack{j\in\mathcal{J}_n,\\ \psi^n\in\Psi^n,\\ x^n\in\mathcal{T}_{\rho,\delta}(\psi^n)}} p_u(\psi^n|j)\rho(x^n|\psi^n)V_{\theta^{*,n}}(\cdot|x^n)\Big)\right.$$
$$\left. - H\Big(\sum_{\substack{\psi^n\in\Psi^n,\\ x^n\in\mathcal{T}_{\rho,\delta}(\psi^n)}} p_u(\psi^n|j)\rho(x^n|\psi^n)V_{\theta^{*,n}}(\cdot|x^n)\Big)\right)$$
$$= \frac{1}{J_n}\sum_{j\in\mathcal{J}_n}\left(H\big(\rho\bar{V}_{\theta^{*,n}}(\cdot)\big) - H\big(\rho\hat{V}_{\theta^{*,n}}(\cdot|j)\big)\right),$$

where we define

$$\frac{1}{J_n}\sum_{\substack{j\in\mathcal{J}_n,\\ \psi^n\in\Psi^n\\ x^n\in\mathcal{T}_{\rho,\delta}(\psi^n)}} p_u(\psi^n|j)\frac{1}{|\mathcal{T}^n_{\rho,\delta}(\psi^n)|}V_{\theta^{*,n}}(\cdot|x^n) = \rho\bar{V}_{\theta^{*,n}}(\cdot)$$

$$\sum_{\substack{\psi^n\in\Psi^n,\\ x^n\in\mathcal{T}_{\rho,\delta}(\psi^n)}} p_u(\psi^n|j)\frac{1}{|\mathcal{T}^n_{\rho,\delta}(\psi^n)|}V_{\theta^{*,n}}(\cdot|x^n) = \rho\hat{V}_{\theta^{*,n}}(\cdot|j).$$

Now, if we can show that

$$||\rho\bar{V}_{\theta^{*,n}}(\cdot) - \rho\hat{V}_{\theta^{*,n}}(\cdot|j)||_V \leq \epsilon_3 \leq \frac{1}{2}$$

[8]let $\theta^* \in \mathcal{P}(\mathcal{S}|\mathcal{X})$ be the single-letter best channel to the eavesdropper in this section, with $\theta^{*,n} = \prod_{i=1}^n \theta^*$

then we can apply Lemma 3 and obtain

$$I(p_{J_n}; E_u V^n_{\theta^{*,n}}) = \frac{1}{J_n}\sum_{j\in\mathcal{J}_n} |H(\rho\bar{V}_{\theta^{*,n}}(\cdot)) - H(\rho\hat{V}_{\theta^{*,n}}(\cdot|j))| \leq -\epsilon_3 \log \frac{\epsilon_3}{|\mathcal{Z}|^n}$$

We extend [22] to prove that the secrecy requirement is fulfilled. For some positive measure $\Omega(\cdot)$ on $\mathcal{Z}^n$ that will be defined later in this section, we have by the triangle inequality

$$||\rho\bar{V}_{\theta^{*,n}}(\cdot) - \rho\hat{V}_{\theta^{*,n}}(\cdot|j)||_V \leq ||\rho\hat{V}_{\theta^{*,n}}(\cdot|j) - \Omega(\cdot)||_V + ||\Omega(\cdot) - \rho\bar{V}_{\theta^{*,n}}(\cdot)||_V. \quad (34)$$

We will concentrate on the first term, since

$$||\Omega(\cdot) - \rho\bar{V}_{\theta^{*,n}}(\cdot)||_V = ||\frac{1}{J_n}\sum_{j\in\mathcal{J}_n}\left(\rho\hat{V}_{\theta^{*,n}}(\cdot|j) - \Omega(\cdot)\right)||_V \leq \frac{1}{J_n}\sum_{j\in\mathcal{J}_n}||\rho\hat{V}_{\theta^{*,n}}(\cdot|j) - \Omega(\cdot)||_V.$$

2) Construction of $\Omega(\cdot)$: We define the set $\varepsilon_1(\psi^n)$ and $\tilde{\Omega}(z^n)$ as

$$\varepsilon_1(\psi^n) = \mathcal{T}^n_{\rho V_{\theta^*},\delta}(\psi^n), \quad (35)$$

$$\tilde{\Omega}(z^n) = \mathbb{E}_{\Psi^n}[\rho V^n_{\theta^{*,n}}(z^n|\Psi^n)\mathbb{1}_{\varepsilon_1(\Psi^n)}(z^n)], \quad (36)$$

where we take the expectation over all $\psi^n \in \mathcal{T}^n_{p,\delta}$. Recall the definition of $\rho V^n_{\theta^{*,n}}(z^n|\psi^n)$ as

$$\rho V^n_{\theta^{*,n}}(z^n|\psi^n) = \left(\sum_{x^n\in\mathcal{T}^n_{\rho,\delta}(\psi^n)}\frac{1}{|\mathcal{T}_{\rho,\delta}(\psi^n)|}\sum_{s^n\in\mathcal{S}^n}\theta^{*,n}(s^n|x^n)V^n(z^n|x^n,s^n)\right)$$

Further, we define the set

$$\varepsilon_2 := \Big\{z^n \in \mathcal{T}_{Z_{\theta^{*,n}},2|\mathcal{X}||\Psi|\delta} : \tilde{\Omega}(z^n) \geq \exp\{-nc'\delta^2\}\exp\{-n(H(Z_{\theta^*}) + f_1(\delta))\}\Big\}, \quad (37)$$

with

$$|\mathcal{T}_{Z_{\theta^{*,n}},2|\mathcal{X}||\Psi|\delta}| \leq \exp\{n(H(Z_{\theta^*}) + f_1(\delta))\}, \quad \epsilon_n = \exp\{-nc'\delta^2\}.$$

where the cardinality bound on $\mathcal{T}_{Z_{\theta^{*,n}},2|\mathcal{X}||\Psi|\delta}$ and ϵ_n are motivated by Lemmas 7 and 8, respectively. We set

$$\Omega(z^n) = \tilde{\Omega}(z^n)\mathbb{1}_{\varepsilon_2}(z^n). \quad (38)$$

By definition, $\Omega(z^n) \geq \epsilon_n \exp\{-n(H(Z_{\theta^*}) + f_1(\delta))\}$, for all $z^n \in \varepsilon_2$, else $\Omega(z^n) = 0$. Note, that when summing up over all $z^n \in \varepsilon_2$ we get

$$\sum_{z^n\in\varepsilon_2}\Omega(z^n) = \Omega(\varepsilon_2) = \tilde{\Omega}(\varepsilon_2) = \tilde{\Omega}\left(\mathcal{T}_{Z_{\theta^{*,n}},2|\mathcal{X}||\Psi|\delta}\right) - \tilde{\Omega}\left(\mathcal{T}_{Z_{\theta^{*,n}},2|\mathcal{X}||\Psi|\delta}\setminus\varepsilon_2\right) \geq 1 - 2\epsilon_n,$$

where the inequality follows by the properties of typical sets and sequences, Lemma 8, i.e., by $\tilde{\Omega}\left(\mathcal{T}_{Z_{\theta^{*,n}},2|\mathcal{X}||\Psi|\delta}\right) \geq 1-\epsilon_n$, and $\tilde{\Omega}\left(\mathcal{T}_{Z_{\theta^{*,n}},2|\mathcal{X}||\Psi|\delta}\setminus\varepsilon_2\right) \leq \epsilon_n$. Similar to [22] we obtain a modification of $\rho V^n_{\theta^{*,n}}$ as

$$Q_{\theta^{*,n}}(z^n|\psi^n) := \rho V^n_{\theta^{*,n}}(z^n|\psi^n)\mathbb{1}_{\varepsilon_1(\psi^n)}(z^n)\mathbb{1}_{\varepsilon_2}(z^n), \quad (39)$$

and can define the event

$$\iota_1(j,z^n) := \left\{\frac{1}{L_n}\sum_{l=1}^{L_n} Q_{\theta^{*,n}}(\Psi^n_{jl}|z^n) \in [(1\pm\epsilon_n)\Omega(z^n)]\right\} \quad (40)$$

Lemma 14: For $\tau_a > 0$, the probability that $\iota_1(j,z^n)$ is not fulfilled can be upper bounded as

$$Pr\{\iota_1(j,z^n)^c\} \leq 2\exp_e\left\{-\frac{1}{3}\exp\{n\tau_a\}\right\} \quad (41)$$

Proof: We will apply a Chernoff-Hoeffding bound, Lemma 6.

$$Pr\left\{\frac{1}{L_n}\sum_{l=1}^{L_n} Q_{\theta^{*,n}}(z^n|\Psi^n_{jl}) \notin [(1\pm\epsilon_n)\Omega(z^n)]\right\} \leq 2\exp_e\left(-L_n\frac{\epsilon_n^2\Omega(z^n)}{3b_n}\right).$$

We can plug in the bounds for $Q_{\theta^{*,n}}(\Psi^n_{jl}, z^n)$ and $\Omega(z^n)$ induced by the restrictions to $\varepsilon_1(\psi^n)$ and ε_2, respectively,

$$Q_{\theta^{*,n}}(z^n|\Psi^n_{jl}) \leq \exp\{-n(H(Z_{\theta^*}|\Psi) - f_2(\delta))\}, \quad \Omega(z^n) \geq \epsilon_n\exp\{-n(H(Z_{\theta^*}) + f_1(\delta))\},$$

and obtain for the exponent

$$\begin{aligned} -L_n\frac{\epsilon_n^2\Omega(z^n)}{3b_n} &\leq -\frac{1}{3}L_n\epsilon_n^3\exp\{-n(H(Z_{\theta^*}) + f_1(\delta))\}\exp\{n(H(Z_{\theta^*}|\Psi) - f_2(\delta))\} \\ &= -\frac{1}{3}L_n\exp\Big\{-n(H(Z_{\theta^*}) - H(Z_{\theta^*}|\Psi) + f_1(\delta) + f_2(\delta)) + 3c'\delta^2\Big\} \\ &= -\frac{1}{3}L_n\exp\{-n(I(Z_{\theta^*};\Psi) + f_1(\delta) + f_2(\delta)) + 3c'\delta^2\}. \end{aligned}$$

If we choose L_n to be

$$L_n \geq \exp\{n(I(Z_{\theta^*};\Psi) + f_1(\delta) + f_2(\delta) + 3c'\delta^2 + \tau_a)\}, \quad \lim_{\delta\to 0} f_1(\delta) = \lim_{\delta\to 0} f_2(\delta) = \lim_{\delta\to 0} 3c'\delta^2 = 0,$$

then the probability that $\iota_1(j,z^n)$ is not fulfilled vanishes doubly exponentially fast. ∎

We define the event ι_0 as the event that $\iota_1(j,z^n)$ holds for all $j \in \mathcal{J}_n$, $z^n \in \mathcal{Z}^n$, and $u \in \mathcal{U}_n$

$$\iota_0 := \bigcap_{j\in\mathcal{J}_n}\bigcap_{z^n\in\mathcal{Z}^n}\bigcap_{u\in\mathcal{U}_n}\iota_1(j,z^n). \quad (42)$$

We can bound the probability of ι_0 from below as

$$\begin{aligned} Pr\{\iota_0\} &= 1 - Pr\{\iota_0^c\} \\ &= 1 - Pr\left\{\bigcup_{j\in\mathcal{J}_n}\bigcup_{z^n\in\mathcal{Z}^n}\bigcup_{u\in\mathcal{U}_n}\iota_1^c(j,z^n)\right\} \\ &\geq 1 - 2|\mathcal{J}_n||\mathcal{Z}|^n|\mathcal{U}_n|\exp_e\{-\frac{1}{3}\exp\{n\tau_a\}\}. \end{aligned}$$

Since $|\mathcal{J}_n|$, $|\mathcal{Z}|^n$, and $|\mathcal{U}_n|$ grow only exponentially fast in n, but $Pr\{\iota_1^c(j,z^n)\}$ vanishes doubly exponentially fast in n, the probability that ι_0 holds, approaches one.

3) Leakage Analysis: Let $\mathcal{K}_n^{\text{ran}}$ be a realization of the random CR-assisted code $\mathcal{K}_n^{\text{ran}}(\hat{\chi})$, fulfilling the required properties for guaranteeing secrecy. Furthermore, let ψ_{jl}^n be the codeword realization for the message pair $(j,l)\in\mathcal{J}_n\times\mathcal{L}_n$ for the CR-assisted code $\mathcal{K}_n^{\text{ran}}$ for a specific realization of $u\in\mathcal{U}_n$. Keep in mind that the leakage has to vanish for all $u\in\mathcal{U}_n$, and that we omit the indexing on u as before. We can bound the first term in equation (34) for any $j\in\mathcal{J}_n$ as

$$\begin{aligned} &\left\|\rho\hat{V}_{\theta^{*,n}}(\cdot|j)-\Omega(\cdot)\right\|_V \leq \\ &\left\|\frac{1}{L_n}\sum_{l=1}^{L_n}Q_{\theta^{*,n}}(\cdot|\psi_{jl}^n)-\Omega(\cdot)\right\|_V \qquad (43)\\ &+\left\|\frac{1}{L_n}\sum_{l=1}^{L_n}\rho V_{\theta^{*,n}}^n(\cdot|\psi_{jl}^n)\mathbb{1}_{\varepsilon_1(\psi_{jl}^n)}(\cdot)(\mathbb{1}_{\mathcal{Z}^n}(\cdot)-\mathbb{1}_{\varepsilon_2}(\cdot))\right\|_V \qquad (44)\\ &+\left\|\frac{1}{L_n}\sum_{l=1}^{L_n}\rho V_{\theta^{*,n}}^n(\cdot|\psi_{jl}^n)(\mathbb{1}_{\mathcal{Z}^n}(\cdot)-\mathbb{1}_{\varepsilon_1(\psi_{jl}^n)}(\cdot))\right\|_V. \qquad (45) \end{aligned}$$

In the following, we bound the right hand side of (43), and the terms in (44), (45), individually.

The right hand side of (43) can be bounded by the result of Lemma 14 to

$$\begin{aligned} &\left\|\frac{1}{L_n}\sum_{l=1}^{L_n}Q_{\theta^{*,n}}(\cdot|\psi_{jl}^n)-\Omega(\cdot)\right\|_V \\ &\quad=\sum_{z^n\in\mathcal{Z}^n}\left|\frac{1}{L_n}\sum_{l=1}^{L_n}Q_{\theta^{*,n}}(z^n|\psi_{jl}^n)-\Omega(z^n)\right| \\ &\quad\leq\sum_{z^n\in\mathcal{Z}^n}\epsilon_n\Omega(z^n) \\ &\quad\leq\epsilon_n \end{aligned}$$

For (44), we obtain

$$\begin{aligned} &\left\|\frac{1}{L_n}\sum_{l=1}^{L_n}\rho V_{\theta^{*,n}}^n(\cdot|\psi_{jl}^n)\mathbb{1}_{\varepsilon_1(\psi_{jl}^n)}(\cdot)(\mathbb{1}_{\mathcal{Z}^n}(\cdot)-\mathbb{1}_{\varepsilon_2}(\cdot))\right\|_V \\ &=\sum_{z^n\in\mathcal{Z}^n}\left|\frac{1}{L_n}\sum_{l=1}^{L_n}\rho V_{\theta^{*,n}}^n(z^n|\psi_{jl}^n)\mathbb{1}_{\varepsilon_1(\psi_{jl}^n)}(z^n)\right. \\ &\qquad\left.\left(\mathbb{1}_{\mathcal{Z}^n}(z^n)-\mathbb{1}_{\varepsilon_2}(z^n)\right)\right| \\ &=\frac{1}{L_n}\sum_{l=1}^{L_n}\sum_{z^n\in\mathcal{Z}^n}\rho V_{\theta^{*,n}}^n(z^n|\psi_{jl}^n)\mathbb{1}_{\varepsilon_1(\psi_{jl}^n)}(z^n)\mathbb{1}_{\mathcal{Z}^n}(z^n) \\ &\quad-\sum_{z^n\in\mathcal{Z}^n}\frac{1}{L_n}\sum_{l=1}^{L_n}\rho V_{\theta^{*,n}}^n(z^n|\psi_{jl}^n)\mathbb{1}_{\varepsilon_1(\psi_{jl}^n)}(z^n)\mathbb{1}_{\varepsilon_2}(z^n) \\ &\leq 1-\sum_{z^n\in\mathcal{Z}^n}\frac{1}{L_n}\sum_{l=1}^{L_n}Q_{\theta^{*,n}}(z^n|\psi_{j,l}^n) \\ &\leq 1-\sum_{z^n\in\mathcal{Z}^n}(1-\epsilon_n)\Omega(z^n) \\ &\leq 1-(1-\epsilon_n)(1-2\epsilon_n) \\ &\leq 3\epsilon_n-2\epsilon_n^2 \\ &\leq 3\epsilon_n. \end{aligned}$$

For (45), we obtain

$$\begin{aligned} &\left\|\frac{1}{L_n}\sum_{l=1}^{L_n}\rho V_{\theta^{*,n}}^n(\cdot|\psi_{jl}^n)(\mathbb{1}_{\mathcal{Z}^n}(\cdot)-\mathbb{1}_{\varepsilon_1(\psi_{jl}^n)}(\cdot))\right\|_V \\ &\overset{(a)}{=}\frac{1}{L_n}\sum_{l=1}^{L_n}\rho V_{\theta^{*,n}}^n(\varepsilon_1^c(\psi_{jl}^n)|\psi_{jl}^n) \\ &\overset{(b)}{=}\frac{1}{L_n}\sum_{l\in\mathcal{L}_n}\rho V_{\theta^{*,n}}^n(\mathcal{T}^c_{\rho V_{\theta^{*,n}},\delta}(\psi_{jl}^n)|\psi_{jl}^n) \\ &\overset{(c)}{\leq}\frac{1}{L_n}\sum_{l\in\mathcal{L}_n}\exp\{-nc'\delta^2\} \\ &\overset{(d)}{=}\epsilon_n. \end{aligned}$$

Here, (a) follows by summing up only over $z^n\in\varepsilon_1^c(\cdot)$. (b) follows by the definition of $\varepsilon_1(\psi_{jl}^n)$. (c) follows since the probability of not obtaining a conditional typical z^n can be upper bounded. (d) follows since the upper bound in (c) is valid for all ψ_{jl}^n.

Therefore, for (34) we obtain

$$\begin{aligned} &||\rho\bar{V}_{\theta^{*,n}}(Z^n)-\rho\hat{V}_{\theta^{*,n}}(Z^n|j)||_V\leq 10\epsilon_n \\ &I(p_{J_n};E_uV_{\theta^{*,n}}^n)\leq 10\ n\epsilon_n\log(|\mathcal{Z}|)-10\epsilon_n\log(10\epsilon_n), \end{aligned}$$

which vanishes as n goes to infinity because ϵ_n vanishes exponentially in n.

H. Existence of Codes Fulfilling Both the Error and the Secrecy Requirement

It remains to show that there exist codes fulfilling the error requirement and the secrecy requirement simultaneously.

Therefore, we define the following event.

$$\begin{aligned} \tilde{\iota} := \Bigg\{&\hat{\hat{e}}(\mathcal{K}_n^{\text{ran}})\leq \\ &(1+\epsilon)\tilde{p}+\lambda+\frac{1+\epsilon_2}{1-\epsilon_2}\Bigg(\frac{\exp\{-nc'\delta'\}}{\lambda} \\ &+\frac{\exp\left\{-n\left(\min_{W\in\widehat{\mathcal{W}}}I(p_\Psi;\rho W)-R-\nu\right)\right\}}{\lambda}\Bigg)\Bigg\} \end{aligned}$$

$$\hat{\iota} := \iota_0\cap\tilde{\iota}$$

Here, we can apply the union bound and obtain

$$\begin{aligned} Pr\{\hat{\iota}\} &= 1-Pr\{\hat{\iota}^c\} \\ &= 1-Pr\{\iota_0^c\cup\tilde{\iota}^c\} \\ &\geq 1-Pr\{\iota_0^c\}-Pr\{\tilde{\iota}^c\}, \end{aligned}$$

where both, $Pr\{\iota_0^c\}$ and $Pr\{\tilde{\iota}^c\}$ vanish super exponentially fast. Hence, there exist codes fulfilling the aforementioned criteria simultaneously. Finally, we get the achievable CR-assisted code secrecy rate as

$$\begin{aligned}\widehat{\widehat{R}}_S^{ran} &\leq \max_{\Psi\leftrightarrow X\leftrightarrow(Y,Z)} \Bigg(\min_{\theta\in\mathcal{P}(\mathcal{S}|\mathcal{X})} I(\Psi;Y_\theta) \\ &\qquad - \max_{\theta\in\mathcal{P}(\mathcal{S}|\mathcal{X})} I(\Psi;Z_\theta) \Bigg) \\ &= \max_{p_\Psi\in\mathcal{P}(\mathcal{V}),\rho\in\mathcal{P}(\mathcal{X}|\mathcal{V})} \Bigg(\min_{W\in\widehat{\widehat{\mathcal{W}}}} I(p_\Psi;\rho W) \\ &\qquad - \min_{V\in\widehat{\widehat{\mathcal{V}}}} I(P_\Psi;\rho V) \Bigg).\end{aligned}$$

I. Converse

What remains is to show the converse. We modify the standard converse of the WTC. As usual, we assumed strong secrecy in the achievability part and show in the converse, that even with weak secrecy the upper and lower bounds match.

Let $n\epsilon \geq \max_{u\in\mathcal{U}_n} I(J;Z_{\theta^*}^n|U=u)$. We consider a sequence $(\mathcal{K}_n^{\text{ran}})_{n=1}^\infty$ of $(n,J_n,\mathcal{U}_n,p_U)$ wiretap-codes for which $e(\mathcal{K}_n^{\text{ran}})\leq\hat{\epsilon}$ and for an $\epsilon,\hat{\epsilon}>0$, as $n\to\infty$.

$$\begin{aligned} nR_s &= H(J) \\ &\overset{(a)}{\leq} \min_{\theta\in\mathcal{P}(\mathcal{S}^n|\mathcal{X}^n)} I(J;Y_\theta^n|U) + 1 + \hat{\epsilon}H(J), \\ \to nR_s &\leq \frac{1}{1-\hat{\epsilon}}\Big(\min_{\theta\in\mathcal{P}(\mathcal{S}^n|\mathcal{X}^n)} I(J;Y_\theta^n|U) - I(J;Z_{\theta^*}^n|U) \\ &\qquad + \max_{u\in\mathcal{U}} I(J;Z_{\theta^*}^n|U=u) + 1\Big) \\ &\overset{(c)}{\leq} \frac{1}{1-\hat{\epsilon}}\Big(\min_{\theta\in\mathcal{P}(\mathcal{S}^n|\mathcal{X}^n)} I(J;Y_\theta^n|U) - I(J;Z_{\theta^*}^n|U) \\ &\qquad + n\epsilon + 1\Big) \\ &\overset{(d)}{=} \frac{1}{1-\hat{\epsilon}}\Big(\min_{\theta\in\mathcal{P}(\mathcal{S}^n|\mathcal{X}^n)} I(J,U;Y_\theta^n|U) - I(J,U;Z_{\theta^*}^n|U) \\ &\qquad + n\epsilon + 1\Big) \\ &\overset{(e)}{=} \frac{1}{1-\hat{\epsilon}}\Big(\min_{\theta\in\mathcal{P}(\mathcal{S}^n|\mathcal{X}^n)} I(\tilde{\Psi}^n;Y_\theta^n|U) - I(\tilde{\Psi}^n;Z_{\theta^*}^n|U) \\ &\qquad + n\epsilon + 1\Big) \\ &\overset{(f)}{\leq} \frac{1}{1-\hat{\epsilon}}\Big(\max_{u\in\mathcal{U}_n}\Big(\min_{\theta\in\mathcal{P}(\mathcal{S}^n|\mathcal{X}^n)} I(\tilde{\Psi}^n;Y_\theta^n|U=u) \\ &\qquad - I(\tilde{\Psi}^n;Z_{\theta^*}^n|U=u) + n\epsilon + 1\Big)\Big) \\ &\overset{(g)}{=} \frac{1}{1-\hat{\epsilon}}\Big(\min_{\theta\in\mathcal{P}(\mathcal{S}^n|\mathcal{X}^n)} I(\hat{\Psi}^n;Y_\theta^n) - I(\hat{\Psi}^n;Z_{\theta^*}^n) \\ &\qquad + n\epsilon + 1\Big) \\ &\overset{(h)}{\leq} \frac{1}{1-\hat{\epsilon}}\Big(\min_{\theta^n\in\mathcal{P}^n(\mathcal{S}|\mathcal{X})} I(\hat{\Psi}^n;Y_{\theta^n}^n) - I(\hat{\Psi}^n;Z_{\theta^*}^n) \\ &\qquad + n\epsilon + 1\Big) \\ &= \frac{1}{1-\hat{\epsilon}}\Big(\min_{\theta^n\in\mathcal{P}^n(\mathcal{S}|\mathcal{X})} \sum_{i=1}^n I(\hat{\Psi}^n;Y_{i,\theta_i}|Y_{\theta^{i-1}}^{i-1}) \\ &\qquad - \sum_{i=1}^n I(\hat{\Psi}^n;Z_{i,\theta_i^*}|Z_{i+1,\theta_{i+1}^{n,*}}^n) + n\epsilon + 1\Big) \\ &= \frac{1}{1-\hat{\epsilon}}\Big(\min_{\theta^n\in\mathcal{P}^n(\mathcal{S}|\mathcal{X})} \sum_{i=1}^n \Big(I(\hat{\Psi}^n, Z_{i+1,\theta_{i+1}^{n,*}}^n;Y_{i,\theta_i}|Y_{\theta^{i-1}}^{i-1}) \\ &\qquad - I(Z_{i+1,\theta_{i+1}^{n,*}}^n;Y_{i,\theta_i}|\hat{\Psi}^n,Y_{\theta^{i-1}}^{i-1})\Big) \\ &\qquad - \sum_{i=1}^n I(\hat{\Psi}^n;Z_{i,\theta_i^*}|Z_{i+1,\theta_{i+1}^{n,*}}^n) + n\epsilon + 1\Big) \\ &= \frac{1}{1-\hat{\epsilon}}\Big(\min_{\theta^n\in\mathcal{P}^n(\mathcal{S}|\mathcal{X})} \sum_{i=1}^n \Big(I(\hat{\Psi}^n, Z_{i+1,\theta_{i+1}^{n,*}}^n;Y_{i,\theta_i}|Y_{\theta^{i-1}}^{i-1}) \\ &\qquad - I(Z_{i+1,\theta_{i+1}^{n,*}}^n;Y_{i,\theta_i}|\hat{\Psi}^n,Y_{\theta^{i-1}}^{i-1})\Big) \\ &\qquad - \sum_{i=1}^n \Big(I(\hat{\Psi}^n,Y_{\theta^{i-1}}^{i-1};Z_{i,\theta_i^*}|Z_{i+1,\theta_{i+1}^{n,*}}^n) \\ &\qquad + I(Y_{\theta^{i-1}}^{i-1};Z_{i,\theta_i^*}|\hat{\Psi}^n,Z_{i+1,\theta_{i+1}^{n,*}}^n)\Big) \\ &\qquad + n\epsilon + 1\Big) \\ &\overset{(i)}{=} \frac{1}{1-\hat{\epsilon}}\Big(\min_{\theta^n\in\mathcal{P}^n(\mathcal{S}|\mathcal{X})} \sum_{i=1}^n I(\hat{\Psi}^n, Z_{i+1,\theta_{i+1}^{n,*}}^n;Y_{i,\theta_i}|Y_{\theta^{i-1}}^{i-1}) \\ &\qquad - \sum_{i=1}^n I(\hat{\Psi}^n,Y_{\theta^{i-1}}^{i-1};Z_{i,\theta_i^*}|Z_{i+1,\theta_{i+1}^{n,*}}^n) \\ &\qquad + n\epsilon + 1\Big) \\ &= \frac{1}{1-\hat{\epsilon}}\Big(\min_{\theta^n\in\mathcal{P}^n(\mathcal{S}|\mathcal{X})} \sum_{i=1}^n \Big(I(Z_{i+1,\theta_{i+1}^{n,*}}^n;Y_{i,\theta_i}|Y_{\theta^{i-1}}^{i-1}) \\ &\qquad + I(\hat{\Psi}^n;Y_{i,\theta_i}|Y_{\theta^{i-1}}^{i-1},Z_{i+1,\theta_{i+1}^{n,*}}^n) \\ &\qquad - I(Y_{\theta^{i-1}}^{i-1};Z_{i,\theta_i^*}|Z_{i+1,\theta_{i+1}^{n,*}}^n) \\ &\qquad - I(\hat{\Psi}^n;Z_{i,\theta_i^*}|Z_{i+1,\theta_{i+1}^{n,*}}^n,Y_{\theta^{i-1}}^{i-1})\Big) \\ &\qquad + n\epsilon + 1\Big) \\ &\overset{(j)}{=} \frac{1}{1-\hat{\epsilon}}\Big(\min_{\theta^n\in\mathcal{P}^n(\mathcal{S}|\mathcal{X})} \sum_{i=1}^n \Big(I(\hat{\Psi}^n;Y_{i,\theta_i}|Y_{\theta^{i-1}}^{i-1},Z_{i+1,\theta_{i+1}^{n,*}}^n) \\ &\qquad - I(\hat{\Psi}^n;Z_{i,\theta_i^*}|Z_{i+1,\theta_{i+1}^{n,*}}^n,Y_{\theta^{i-1}}^{i-1})\Big) \\ &\qquad + n\epsilon + 1\Big) \\ &\overset{(k)}{=} \frac{1}{1-\hat{\epsilon}}\Big(\min_{\theta^n\in\mathcal{P}^n(\mathcal{S}|\mathcal{X})} \sum_{i=1}^n \Big(I(\hat{\Psi}^n;Y_{i,\theta_i}|V_i) - I(\hat{\Psi}^n;Z_{i,\theta_i^*}|V_i)\Big) \\ &\qquad + n\epsilon + 1\Big) \\ &= \frac{1}{1-\hat{\epsilon}}\Big(\min_{\theta^n\in\mathcal{P}^n(\mathcal{S}|\mathcal{X})}\end{aligned}$$

$$\sum_{i=1}^{n}\Big(I(\hat{\Psi}^n,V_i;Y_{i,\theta_i}|V_i) - I(\hat{\Psi}^n,V_i;Z_{i,\theta_i^*}|V_i)\Big) + n\epsilon + 1\Big)$$

$$\overset{(l)}{=} \frac{1}{1-\hat{\epsilon}}\Big(\min_{\theta^n\in\mathcal{P}^n(\mathcal{S}|\mathcal{X})} \sum_{i=1}^{n}\Big(I(\Psi_i';Y_{i,\theta_i}|V_i) - I(\Psi_i';Z_{i,\theta_i^*}|V_i)\Big) + n\epsilon + 1\Big)$$

$$\overset{(m)}{=} \frac{1}{1-\hat{\epsilon}}\Big(\min_{\theta^n\in\mathcal{P}^n(\mathcal{S}|\mathcal{X})} n(I(\Psi_Q';Y_{Q,\theta_Q}|V_Q,Q) - I(\Psi_Q';Z_{Q,\theta_Q^*}|V_Q,Q)) + n\epsilon + 1\Big)$$

$$= \frac{1}{1-\hat{\epsilon}}\Big(\min_{\theta^n\in\mathcal{P}^n(\mathcal{S}|\mathcal{X})} n(I(\Psi';Y_\theta|V) - I(\Psi';Z_{\theta^*}|V)) + n\epsilon + 1\Big)$$

$$\leq \frac{1}{1-\hat{\epsilon}}\Big(\min_{\theta^n\in\mathcal{P}^n(\mathcal{S}|\mathcal{X})} n\max_{V=v}(I(\Psi';Y_\theta|V=v) - I(\Psi';Z_{\theta^*}|V=v)) + n\epsilon + 1\Big)$$

$$\leq \frac{1}{1-\hat{\epsilon}}\Big(\max_{\Psi\leftrightarrow X\leftrightarrow(Y_\theta,Z_{\theta^*})} \Big(n\min_{\theta\in\mathcal{P}(\mathcal{S}|\mathcal{X})} I(\Psi';Y_\theta) - nI(\Psi';Z_{\theta^*})\Big) + n\epsilon + 1\Big)$$

$$\Rightarrow R_s \leq \frac{1}{1-\hat{\epsilon}}\Big(\max_{\Psi'\leftrightarrow X\leftrightarrow(Y,Z)} \Big(\min_{\theta\in\mathcal{P}(\mathcal{S}|\mathcal{X})} I(\Psi';Y_\theta) - I(\Psi';Z_{\theta^*})\Big) + \frac{1}{n} + \epsilon\Big)$$

Here, (a) follows by Fano's inequality, where $\hat{\epsilon}$ approaches zero as $n\to\infty$, (b) follows by the definition of the leakage to the eavesdropper, (c) follows because the leakage to the eavesdropper vanishes with n. Now, (d) follows because J and U are independent, (e) by defining $\tilde{\Psi}=(J,U)$, (f) follows naturally. (g) follows because $\tilde{\Psi}\leftrightarrow X^n\leftrightarrow(Y_\theta^n,Z_{\theta^*}^n)$ forms a conditional Markov chain, given $u\in\mathcal{U}$. To see this we evaluate the following term.

$$\begin{aligned} &p_{\tilde{\Psi},X^n,Y_\theta^n,Z_{\theta^*}^n|U}(\cdot|u)\\ &\quad= p_{\tilde{\Psi}|U}(\cdot|u)p_{X^n|\tilde{\Psi},U}(\cdot|\cdot,u)p_{Y_\theta^n,Z_{\theta^*}^n|X^n,\tilde{\Psi},U}(\cdot|\cdot,u)\\ &\quad\overset{(n)}{=} p_{\tilde{\Psi}|U}(\cdot|u)p_{X^n|\tilde{\Psi}}(\cdot|\cdot)p_{Y_\theta^n,Z_{\theta^*}^n|X^n}(\cdot|\cdot)\end{aligned}$$

(n) follows because X^n and $(Y_\theta^n,Z_{\theta^*}^n)$ are connected through a memoryless channel. Remember that when upper bounding the capacity, only the marginals are of interest. Then, we can invoke the same marginals property and can describe the input output relation between X^n and $(Y_\theta^n,Z_{\theta^*}^n)$ by the channels $W_\theta^n(y^n|x^n)$, $V_{\theta^*}^n(z^n|x^n)$. Furthermore, we see that

$$p_{\hat{\Psi},X^n,Y_\theta^n,Z_{\theta^*}^n}(\cdot) = \max_{u\in\mathcal{U}_n} p_{\tilde{\Psi}|U}(\cdot|u)p_{X^n|\tilde{\Psi}}(\cdot|\cdot)p_{Y_\theta^n,Z_{\theta^*}^n|X^n}(\cdot|\cdot).$$

Finally, (h) follows since $\min_{\theta\in\mathcal{P}(\mathcal{S}^n|\mathcal{X}^n)} I(\tilde{\Psi}^n;Y_\theta^n) \leq \min_{\theta^n\in\mathcal{P}^n(\mathcal{S}|\mathcal{X})} I(\tilde{\Psi}^n;Y_{\theta^n}^n)$, with $\theta^n(s^n|x^n) = \prod_{i=1}^n\theta_i(s_i|x_i)$. (i) and (j) follow because of Csiszar's Sum Identity, because

$$\sum_{i=1}^{n} I(Z_{i+1,\theta_{i+1}^{n,*}}^n;Y_{i,\theta_i}|\tilde{\Psi}^n,Y_{\theta^{i-1}}^{i-1}) = \sum_{i=1}^{n} I(Y_{\theta^{i-1}}^{i-1};Z_{i,\theta_i^*}|\tilde{\Psi}^n,Z_{i+1,\theta_{i+1}^{n,*}}^n),$$

$$\sum_{i=1}^{n} I(Z_{i+1,\theta_{i+1}^{n,*}}^n;Y_{i,\theta_i}|Y_{\theta^{i-1}}^{i-1}) = \sum_{i=1}^{n} I(Y_{\theta^{i-1}}^{i-1};Z_{i,\theta_i^*}|Z_{i+1,\theta_{i+1}^{n,*}}^n).$$

(k) follows by identifying $V_i=(Z_{i+1,\theta_{i+1}^{n,*}}^n,Y_{\theta^{i-1}}^{i-1})$, (l) by identifying $\Psi_i'=(\tilde{\Psi}^n,V_i)$, and (m) follows by introducing a uniformly distributed time sharing variable Q.

Appendix G
Proof of Theorem 2

A. Achievability

We use the same approach as in [36], and have

$$\begin{aligned} I(X;Y_\theta) &\geq I(X;Z_{\theta^*}),\\ I(\Psi;Y_\theta) &= I(\Psi,X;Y_\theta) - I(X;Y_\theta|\Psi)\\ &= I(X;Y_\theta) + I(\Psi;Y_\theta|X) - I(X;Y_\theta|\Psi)\\ &= I(X;Y_\theta) - I(X;Y_\theta|\Psi),\\ I(\Psi;Z_{\theta^*}) &= I(X;Z_{\theta^*}) - I(X;Z_{\theta^*}|\Psi),\\ I(\Psi;Y_\theta) - I(\Psi;Z_{\theta^*}) &= I(X;Y_\theta) - I(X;Z_{\theta^*}) + I(X;Z_{\theta^*}|\Psi) - I(X;Y_\theta|\Psi),\end{aligned}$$

where we can upper bound

$$\begin{aligned} &I(X;Z_{\theta^*}|\Psi) - I(X;Y_\theta|\Psi)\\ &\quad\leq \max_{p_{\Psi X}}(I(X;Z_{\theta^*}|\Psi) - I(X;Y_\theta|\Psi))\\ &\quad= \max_{p_{\Psi X}}\Big(\sum_{\psi\in\Psi} p_\Psi(\psi)I(X;Z_{\theta^*}|\Psi=\psi) - I(X;Y_\theta|\Psi=\psi)\Big)\\ &\quad= \max_{p_X}(I(X;Z_{\theta^*}) - I(X;Y_\theta))\\ &\quad\leq 0.\end{aligned}$$

Hence, in total we obtain the following

$$\max_{p_\Psi,\rho_{X|\Psi}}(I(\Psi;Y_\theta) - I(\Psi;Z_{\theta^*})) \leq \max_{p_X}(I(X;Y_\theta) - I(X;Z_{\theta^*})),$$

with equality if we choose $\Psi=X$ as the channel input.

TABLE II
NOTATION, SYMBOLS AND MEANINGS

Symbols	Meaning
$\log(\cdot)$	Logarithm to base 2, $\log_2(\cdot)$, unless stated otherwise.
$\exp\{\cdot\}$, $\exp_e\{\cdot\}$	$2^{\{\cdot\}}$, $e^{\{\cdot\}}$.
X , x	The random variable X and its realization x.
$\mathcal{U}$	The set $\mathcal{U}$, sets are denoted by calligraphic letters.
$\lvert\mathcal{U}\rvert$	The cardinality of a set $\mathcal{U}$.
$\mathcal{P}(\mathcal{U})$	The set of all probability measures on a set $\mathcal{U}$.
$p^n(x^n)$	For $p \in \mathcal{P}(\mathcal{U})$ we define $p^n \in \mathcal{P}(\mathcal{U}^n)$ as $p^n(x^n) = \prod_i^n p(x_i)$.
pW , $pW(y)$	Induced output probability function by p_X and the channel $W(y\lvert x)$, $pW(y) = \sum_{x\in\mathcal{X}} p(x)W(y\lvert x)$.
$H(X)$, $H(p_X)$	Entropy of the RV X, written in terms of the involved RV or the involved probability function p_X.
$H(W\lvert p)$	The conditional Entropy of Y given X, $H(W\lvert p) = -\sum_{x,y} p(x)W(y\lvert x)\log W(y\lvert x)$.
$I(p;W)$, $I(X;Y)$	Mutual information between channel input and channel output, written in terms of the involved probability functions or the involved RV.
$N(a\lvert s^n)$	Number of occurrences of the symbol a in the sequence s^n.
$\mathcal{P}_0^n(\mathcal{S})$, $\mathcal{P}_0(\mathcal{S}^n)$	The set of all possible types of sequences of length n. We use the latter notation to emphasize that $p \in \mathcal{P}_0(\mathcal{S}^n)$ is a single-letter distribution defined by the empirical distribution of sequences of length n.
$\mathcal{T}_{p,\delta}^n \subset \mathcal{X}^n$	For a $p \in \mathcal{P}(\mathcal{X})$ and $\delta > 0$, this denotes the δ-typical set.
$\mathcal{T}_{W,\delta}^n(x^n) \subset \mathcal{Y}^n$	For a $W \in \mathcal{P}(\mathcal{Y}\lvert\mathcal{X})$ and a $\delta > 0$ this denotes the δ-conditionally typical set, given the sequence x^n.
$\mathcal{J}_n$, $\mathcal{L}_n$	Secure and confusing message sets.
$\Psi_{j,l,u}$, $\psi_{j,l,u}$	Codeword (RV and realization) for the messages $j \in \mathcal{J}_n$ and $l \in \mathcal{L}_n$ with CR realization $u \in \mathcal{U}_n$.
$\mathcal{X}, \mathcal{S}, \mathcal{Y}, \mathcal{Z}$	Channel input set, channel state set, channel output set at Bob, channel output set at Eve. All are finite sets.
$\rho^n(x^n\lvert\psi_{j,l,u}^n)$	Mapping from codeword to channel input.
$W^n(y^n\lvert x^n, s^n)$, $V^n(z^n\lvert x^n, s^n)$	DMCs from Alice to Bob and Alice to Eve, here s^n is the channel state, x^n is the channel input, and y^n and z^n are the received sequences at Bob and Eve, respectively.
$\mathcal{U}_n$	Common source of randomness, shared between Alice, Bob and Eve.
$\mathcal{W} = \{(W_s : \mathcal{X} \to \mathcal{P}(\mathcal{Y})) : \ s \in \mathcal{S}\}$	The family of channels to the legitimate receiver.
$\mathcal{V} = \{(V_s : \mathcal{X} \to \mathcal{P}(\mathcal{Z})) : \ s \in \mathcal{S}\}$	The family of channels to the illegitimate receiver.
$(\mathcal{W}, \mathcal{V})$	The AVWC.

TABLE II
(Continued.) NOTATION, SYMBOLS AND MEANINGS

$\mathcal{K}_n$	An (n, J_n) deterministic wiretap-code $\mathcal{K}_n$.
$E : \mathcal{J}_n \to \mathcal{P}(\mathcal{X}^n)$	A stochastic encoder for an (n, J_n) deterministic wiretap-code $\mathcal{K}_n$.
$\mathcal{D}_j$, $\mathcal{D}_{j,u}$, $\mathcal{D}_{jlu}$, $j \in \mathcal{J}_n$, $l \in \mathcal{L}_n$, $u \in \mathcal{U}_n$	Mutually disjoint decoding sets for an (n, J_n) deterministic wiretap-code $\mathcal{K}_n$, an $(n, J_n, \mathcal{U}_n, p_U)$ CR-assisted wiretap-code $\mathcal{K}_n^{\text{ran}}$, and an $(n, J_n, \mathcal{U}_n, p_U)$ CR-assisted wiretap-code $\mathcal{K}_n^{\text{ran}}$ with the requirement that confusing message should also be decoded at Bob.
$EW_{s^n}^n : \mathcal{J}_n \to \mathcal{P}(\mathcal{Y}^n)$	Channel from the secure messages to Bob, $EW_{s^n}^n(y^n\|j) = \sum_{x^n \in \mathcal{X}^n} E(x^n\|j) W^n(y^n\|x^n, s^n)$.
$e(\mathcal{K}_n)$	The maximum error probability for the AVWC for an (n, J_n) deterministic wiretap-code $\mathcal{K}_n$.
$\mathcal{F} : \mathcal{X}^n \to \mathcal{S}^n$	Set of all deterministic functions, mapping from the channel inputs to the channel states. Equivalently the set of all deterministic jamming strategies.
$\hat{e}(\mathcal{K}_n)$	Maximum error probability of (n, J_n) deterministic wiretap-code $\mathcal{K}_n$ for an AVWC if the jammer has non-causal knowledge about the channel input x^n.
$\mathcal{K}_n^{\text{ran}}$	An $(n, J_n, \mathcal{U}_n, p_U)$ CR-assisted wiretap-code $\mathcal{K}_n^{\text{ran}}$
$\mathcal{E} = \{(E_u : \mathcal{J}_n \to \mathcal{P}(\mathcal{X}^n)) : u \in \mathcal{U}_n\}$	Family of stochastic encoders for an $(n, J_n, \mathcal{U}_n, p_U)$ CR-assisted wiretap-code $\mathcal{K}_n^{\text{ran}}$.
$e(\mathcal{K}_n^{\text{ran}})$	The maximum error probability of an $(n, J_n, \mathcal{U}_n, p_U)$ CR-assisted wiretap-code $\mathcal{K}_n^{\text{ran}}$ averaged over all possible randomly chosen deterministic wiretap-codebooks.
$\hat{e}(\mathcal{K}_n^{\text{ran}})$	Maximum error probability of an $(n, J_n, \mathcal{U}_n, p_U)$ CR-assisted wiretap-code $\mathcal{K}_n^{\text{ran}}$ averaged over all possible randomly chosen deterministic wiretap-codebooks if the jammer has non-causal knowledge of the channel input x^n.
$\hat{\hat{e}}(\mathcal{K}_n^{\text{ran}})$	Upper bound of $\hat{e}(\mathcal{K}_n^{\text{ran}})$, results in the consideration of the maxima with respect to $\mathcal{J}_n$, $\mathcal{L}_n$, Ψ^n, $\mathcal{T}_{\rho,\delta}^n(\psi^n)$ and $\mathcal{S}^n$.
$\mathcal{F}'$	The family of all deterministic mappings $\mathcal{J}_n \times \mathcal{X}^n \to \mathcal{S}^n$
$\mathcal{F}''$	The family of all deterministic mappings $\mathcal{J}_n \to \mathcal{S}^n$
R_S	An achievable CR-assisted secrecy rate for the AVWC.
$\widehat{\widehat{R}}_S$	An achievable CR-assisted secrecy rate for the AVWC with non-causal knowledge of the channel input at the jammer.
$\widehat{C}_S^{\text{ran}}(\mathcal{W}, \mathcal{V})$	The CR-assisted secrecy capacity of the AVWC $(\mathcal{W}, \mathcal{V})$ with maximum error probability criterion, when the jammer has not non-causal knowledge about the channel input (or only knows the messages).
$\widehat{C}_{S,av}^{\text{ran}}(\mathcal{W}, \mathcal{V})$	The CR-assisted secrecy capacity of the AVWC $(\mathcal{W}, \mathcal{V})$ with average error probability criterion, when the jammer has not non-causal knowledge about the channel input (or only knows the messages).
$\widehat{\widehat{C}}_S^{\text{ran}}(\mathcal{W}, \mathcal{V})$	The CR-assisted secrecy capacity of the AVWC $(\mathcal{W}, \mathcal{V})$ with maximum error probability criterion if the jammer has non-causal knowledge of the channel input.
$\mathcal{P}(\mathcal{S}^n\|\mathcal{X}^n)$	The set of all stochastic jamming strategies.
$\widehat{\mathcal{W}}$	Convex closure of $\mathcal{W}$.
$\widehat{\widehat{\mathcal{W}}}$	Row convex closure of $\mathcal{W}$.

TABLE II
(Continued.) NOTATION, SYMBOLS AND MEANINGS

$\min_{W\in\widehat{\widehat{\mathcal{W}}}} I(p;W)$= $\min_{\theta\in\mathcal{P}(\mathcal{S}\|\mathcal{X})} I(p;W_\theta)$	Worst case mutual information.
$\theta^{*,n}\in\mathcal{P}^n(\mathcal{S}\|\mathcal{X})$, $V^n_{\theta^{*,n}}$	Best jamming strategy, leading to a best channel to the eavesdropper.
$\pi(\cdot)$	Permutation.
$C_{(j,l)}$, $j\in\mathcal{J}_n$, $l\in\mathcal{L}_n$	Disjoint subsets of the typical sequences $\mathcal{T}^n_{p,\delta}$ of size $\|C_{(j,l)}\|=\frac{\|\mathcal{T}^n_{p,\delta}\|}{\|\mathcal{J}_n\|\|\mathcal{L}_n\|}$.
$\hat{\chi}=\{\Psi^n_{ujl}: j\in\mathcal{J}_n, l\in\mathcal{L}_n, u\in\mathcal{U}_n\}$	The family of RV, representing random codewords. Also used as argument, when we use random coding arguments.
$\mathcal{K}^{\text{ran}}_n(\hat{\chi})$	Random $(n, J_n, \mathcal{U}_n, p_U)$ CR-assisted code.
$\mathcal{U}(j,l,\psi^n,x^n,\hat{\chi})$	The set of all codebooks, for which the sequence ψ^n is the codeword for the message pair (j,l) and x^n is the corresponding channel input.
$\mathcal{U}_0(j,l,\psi^n,x^n,s^n,\hat{\chi})$	The set of all codebooks, for which the sequence ψ^n is the codeword for the message pair (j,l), x^n is the corresponding channel input, and the error bound λ is not met.
$B(u,j,l,\psi^n,x^n,\hat{\chi})$	Binary RV, equals 1 if $u\in\mathcal{U}(j,l,\psi^n,x^n,\hat{\chi})$.
$\tilde{B}(j,l,\psi^n,x^n,s^n,u,\lambda,\hat{\chi})$	Binary RV, equals 1 if $u\in\mathcal{U}_0(j,l,\psi^n,x^n,s^n,\hat{\chi})$.
$\varepsilon_1(\psi^n)$	The set of typical output sequences z^n for which the conditional probability of obtaining the sequence z^n given the codeword ψ^n can be upper bounded in terms of the conditional entropy of Z_{θ^*} given Ψ.
$\tilde{\Omega}(z^n)$	Expectation (with respect to the codeword Ψ^n) of the conditional probability of obtaining the sequence z^n given the codeword Ψ^n. We consider only those summands in the expectation, for which the sequence z^n is in the set $\varepsilon_1(\psi^n)$.
ε_2	The set of typical output sequences z^n for which$\tilde{\Omega}(z^n)$ can be lower bounded in terms of the entropy of Z_{θ^*}.
$\Omega(z^n)$	Equals $\tilde{\Omega}(z^n)$, if z^n is element of ε_2, otherwise it equals zero. In other words, $\Omega(z^n)$ equals the expectation (with respect to the codeword Ψ^n) of the conditional probability of obtaining the sequence z^n given the codeword Ψ^n under the condition that the conditional probability of obtaining the sequence z^n given the codeword ψ^n can be upper bounded in terms of the conditional entropy of Z_{θ^*} given Ψ, and that this expectation can be lower bounded terms of the entropy of Z_{θ^*}.
$Q_{\theta^{*,n}}(z^n\|\psi^n)$	The conditional probability of the sequence z^n given ψ^n, under the condition that the sequence z^n belongs to $\varepsilon_1(\psi^n)$ and ε_2. Equals zero otherwise.
$\iota_1(j,z^n)$	Event that the expectation of $Q_{\theta^{*,n}}(z^n\|\Psi^n_{jl})$ with respect to the confusing messages L_n is in an ϵ_n-region of its expected value, $\Omega(z^n)$.
ι_0	Event that $\iota_1(j,z^n)$ holds for all $j\in\mathcal{J}_n$, $z^n\in\mathcal{Z}^n$, and $u\in\mathcal{U}_n$.
$\tilde{\iota}$	Event that a realization $\mathcal{K}^{\text{ran}}_n$ of a $\mathcal{K}^{\text{ran}}_n(\hat{\chi})$ fulfills the reliability constraint.
$\hat{\iota}$	Event that a realization $\mathcal{K}^{\text{ran}}_n$ of a $\mathcal{K}^{\text{ran}}_n(\hat{\chi})$ fulfills the reliability and secrecy constraints, simultaneously.
$f_{(\cdot)}(\delta)$	Function with $\lim_{\delta\to 0} f_{(\cdot)}(\delta)=0$.

APPENDIX H
NOMENCLATURE

See Table II.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their valuable time and suggestions. Their propositions

both enriched the discussion and the quality of our manuscript. They are grateful to them for sharing their thoughts on the amount of CR, deterministic codes, and multi-letter conditions.

REFERENCES

[1] C. R. Janda, E. A. Jorswieck, M. Wiese, and H. Boche, "Arbitrarily varying wiretap channels with and without non-causal side information at the jammer," in *Proc. IEEE Conf. Commun. Netw. Secur. (CNS)*, Jun. 2020, pp. 1–6.

[2] C. R. Janda, E. A. Jorswieck, M. Wiese, and H. Boche, "Arbitrarily varying wiretap channels with non-causal side information at the jammer," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2020, pp. 938–943.

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

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

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

[6] L. H. Ozarow and A. D. Wyner, "Wire-tap channel II," *AT T Bell Lab. Tech. J.*, vol. 63, no. 10, pp. 2135–2157, 1984.

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

[8] M. Nafea and A. Yener, "A new wiretap channel model and its strong secrecy capacity," *IEEE Trans. Inf. Theory*, vol. 64, no. 3, pp. 2077–2092, Mar. 2018.

[9] 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, Sep. 1960.

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

[11] T. Ericson, "Exponential error bounds for random codes in the arbitrarily varying channel," *IEEE Trans. Inf. Theory*, vol. IT-31, no. 1, pp. 42–48, Jan. 1985.

[12] A. D. Sarwate and M. Gastpar, "Channels with nosy 'noise,"' in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2007, pp. 996–1000.

[13] N. Cai, T. Chan, and A. Grant, "The arbitrarily varying channel when the jammer knows the channel input," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2010, pp. 295–299.

[14] A. D. Sarwate, "Coding against myopic adversaries," in *Proc. IEEE Inf. Theory Workshop (ITW)*, Aug. 2010, pp. 1–5.

[15] A. J. Budkuley, B. K. Dey, and V. M. Prabhakaran, "Communication in the presence of a state-aware adversary," *IEEE Trans. Inf. Theory*, vol. 63, no. 11, pp. 7396–7419, Nov. 2017.

[16] H. Boche, M. Cai, and N. Cai, "Message transmission over classical quantum channels with a jammer with side information: Message transmission capacity and resources," *IEEE Trans. Inf. Theory*, vol. 65, no. 5, pp. 2922–2943, May 2019.

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

[18] E. MolavianJazi, M. Bloch, and J. N. Laneman, "Arbitrary jamming can preclude secure communication," in *Proc. 47th Annu. Allerton Conf. Commun., Control, Comput.*, Sep. 2009, pp. 1069–1075.

[19] I. Bjelaković, H. Boche, and J. Sommerfeld, "Capacity results for arbitrarily varying wiretap channels," in *Information Theory, Combinatorics, and Search Theory: In Memory of Rudolf Ahlswede* (Lecture Notes in Computer Science), vol. 7777, H. Aydinian, F. Cicalese, and C. Deppe, Eds. Berlin, Germany: Springer, 2013, pp. 123–144.

[20] 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.

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

[22] 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.

[23] V. Aggarwal, L. Lai, A. R. Calderbank, and H. V. Poor, "Wiretap channel type II with an active eavesdropper," in *Proc. IEEE Int. Symp. Inf. Theory*, Feb. 2009, pp. 1944–1948.

[24] H. Boche and R. F. Wyrembelski, "Comparison of different attack classes in arbitrarily varying wiretap channels," in *Proc. IEEE Int. Work. Inf. Forensics Security*, Dec. 2012, pp. 270–275.

[25] A. S. Mansour, H. Boche, and R. F. Schaefer, "The secrecy capacity of the arbitrarily varying wiretap channel under list decoding," *Adv. Math. Commun.*, vol. 13, no. 1, pp. 11–39, 2019.

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

[27] 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.

[28] C. Wang, "On the capacity of the binary adversarial wiretap channel," in *Proc. 54th Annu. Allerton Conf. Commun., Control, Comput. (Allerton)*, Sep. 2016, pp. 363–369.

[29] Z. Goldfeld, P. Cuff, and H. H. Permuter, "Arbitrarily varying wiretap channels with type constrained states," *IEEE Trans. Inf. Theory*, vol. 62, no. 12, pp. 7216–7244, Dec. 2016.

[30] Y. Chen, D. He, C. Ying, and Y. Luo, "Strong secrecy of arbitrarily varying wiretap channels with constraints by stochastic code," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2021, pp. 843–848.

[31] 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.

[32] M. Tahmasbi, M. R. Bloch, and A. Yener, "Learning an adversary's actions for secret communication," *IEEE Trans. Inf. Theory*, vol. 66, no. 3, pp. 1607–1624, Mar. 2020.

[33] Y. Chen, D. He, and Y. Luo, "Strong secrecy of arbitrarily varying multiple access channels," *IEEE Trans. Inf. Forensics Security*, vol. 16, pp. 3662–3677, 2021.

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

[35] A. D. Sarwate and M. Gastpar, "Rateless codes for AVC models," *IEEE Trans. Inf. Theory*, vol. 56, no. 7, pp. 3105–3114, Jul. 2010.

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

[37] Y. Chen, D. He, C. Ying, and Y. Luo, "Strong secrecy of arbitrarily varying wiretap channel with constraints," *IEEE Trans. Inf. Theory*, vol. 68, no. 7, pp. 4700–4722, Jul. 2022.

[38] R. Ahlswede, *Storing Transmitting Data* (Foundations in Signal Processing, Communications and Networking), vol. 10, A. Ahlswede, I. Althöfer, C. Deppe, and U. Tamm, Eds. Cham, Switzerland: Springer, 2014.

[39] N. Cai, "Localized error correction in projective space," *IEEE Trans. Inf. Theory*, vol. 59, no. 6, pp. 3282–3294, Jun. 2013.

[40] D. P. Dubhashi and A. Panconesi, *Concentration of Measure for the Analysis of Randomized Algorithms*. Cambridge, U.K.: Cambridge Univ. Press, Oct. 2009.

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

[42] P. Shields, *The Ergodic Theory Discrete Sample Paths* (Graduate Studies in Mathematics), vol. 13. Providence, RI, USA: American Mathematical Society, Jul. 1996.

[43] R. F. Wyrembelski, I. Bjelaković, T. J. Oechtering, and H. Boche, "Optimal coding strategies for bidirectional broadcast channels under channel uncertainty," *IEEE Trans. Commun.*, vol. 58, no. 10, pp. 2984–2994, Oct. 2010.

[44] D. P. Bertsekas, *Convex Optimization Theory* (Athena Scientific Optimization and Computation Series). Nashua, NH, USA: Athena Scientific, 2009.

Carsten Rudolf Janda (Member, IEEE) received the Diploma degree in electrical engineering from the Technical University of Dresden in 2011. From 2013 to 2019, he was a Research Assistant with the Communications Theory Chair of the Communications Laboratory, TU Dresden. Since 2019, he has been a Research Assistant with the Information Theory and Communication Systems Department, Institute for Communications Technology, Technische Universität Braunschweig, Germany.

Moritz Wiese (Member, IEEE) received the Dipl.-Math. degree in mathematics from the University of Bonn, Germany, in 2007, and the Ph.D. degree from the Technical University of Munich, Münich, Germany, in 2013. From 2007 to 2010, he was a Research Assistant with Technische Universität Berlin, Berlin, Germany. From 2010 to 2014, he was with the Technical University of Munich as a Research Assistant. From 2014 to 2016, he was with the ACCESS Linnaeus Center, KTH Royal Institute of Technology, Stockholm, Sweden. He is currently with the Institute of Theoretical Information Technology, Technical University of Munich.

Eduard Axel Jorswieck (Fellow, IEEE) is currently the Managing Director of the Institute of Communications Technology, the Head of the Chair for Communications Systems, and a Full Professor at Technische Universität Braunschweig, Brunswick, Germany. From 2008 to 2019, he was the Head of the Chair of Communications Theory and a Full Professor at TU Dresden, Germany. His main research interests are in the broad area of communications, applied information theory, signal processing, and networking. He has published some 150 journal articles, 15 book chapters, three monographs, one book, and more than 300 conference papers on these topics. In 2006, he received the IEEE Signal Processing Society Best Paper Award. Since 2017, he serves as the Editor-in-Chief for the Springer *EURASIP Journal on Wireless Communications and Networking*. He currently serves as an Editor for IEEE TRANSACTIONS ON COMMUNICATIONS. He has served on the Editorial Boards for IEEE TRANSACTIONS ON SIGNAL PROCESSING, IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, IEEE SIGNAL PROCESSING LETTERS, and IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY.

Holger Boche (Fellow, IEEE) received the Dipl.-Ing. degree in electrical engineering, the Graduate degree in mathematics, and the Dr.-Ing. degree in electrical engineering from the Technische Universität Dresden, Dresden, Germany, in 1990, 1992, and 1994, respectively, the Ph.D. degree from Friedrich-Schiller Universität Jena, Jena, Germany, in 1997, and the Dr.rer.nat. degree in pure mathematics from the Technische Universität Berlin, Berlin, Germany. In 1997, he joined the Fraunhofer Institute for Telecommunications, Heinrich-Hertz-Institute (HHI), Berlin. From 2002 to 2010, he was a Full Professor of mobile communication networks with the Institute for Communications Systems, Technische Universität Berlin. In 2003, he became the Director of the Fraunhofer German-Sino Laboratory for Mobile Communications, Berlin, and in 2004, he became the Director of the Fraunhofer Institute for Telecommunications, HHI. He was a Visiting Professor with the ETH Zurich, Zurich, Switzerland, from 2004 to 2006 (winter), and with KTH Stockholm, Stockholm, Sweden, in 2005 (Summer). He is currently a Full Professor with the Institute of Theoretical Information Technology, Technische Universität München, Münich, Germany, where he joined in October 2010. Among his publications is the recent book *Information Theoretic Security and Privacy of Information Systems* (Cambridge University Press, 2017). Since 2014, he has been a member and an Honorary Fellow of the TUM Institute for Advanced Study, Münich, Germany. Since 2018, he has been a Founding Director of the Center for Quantum Engineering, Technische Universität München. Since 2021, he has been leading jointly with Frank Fitzek the BMBF Research Hub 6G-Life. He is a member of the IEEE Signal Processing Society SPCOM and SPTM Technical Committees. He was an elected member of the German Academy of Sciences (Leopoldina) in 2008 and the Berlin Brandenburg Academy of Sciences and Humanities in 2009. He was a recipient of the Research Award Technische Kommunikation from the Alcatel SEL Foundation in October 2003, the Innovation Award from the Vodafone Foundation in June 2006, and the Gottfried Wilhelm Leibniz Prize from the Deutsche Forschungsgemeinschaft (German Research Foundation) in 2008. He was a co-recipient of the 2006 IEEE Signal Processing Society Best Paper Award and a recipient of the 2007 IEEE Signal Processing Society Best Paper Award. He was the General Chair of the Symposium on Information Theoretic Approaches to Security and Privacy at IEEE GlobalSIP 2016.