A Channel Under Simultaneous Jamming and Eavesdropping Attack—Correlated Random Coding Capacities Under Strong Secrecy Criteria

Moritz Wiese, *Member, IEEE*, Janis Nötzel, and Holger Boche, *Fellow, IEEE*

***Abstract*— We give a complete characterization of the correlated random coding secrecy capacity of arbitrarily varying wiretap channels (AVWCs). We apply two alternative strong secrecy criteria, which both lead to the same multi-letter formula. The difference of these criteria lies in the treatment of correlated randomness; they coincide in the case of uncorrelated codes. On the basis of the derived formula, we show that the correlated random coding secrecy capacity is continuous as a function of the AVWC, in contrast to the discontinuous uncorrelated coding secrecy capacity. In the proof of the secrecy capacity formula for correlated random codes, we apply an auxiliary channel, which is compound from the sender to the intended receiver and arbitrarily varying from the sender to the eavesdropper.**

***Index Terms*— Arbitrarily varying channel, compound channel, correlated randomness, jamming, wiretap channel.**

I. Introduction

THIS paper brings together two areas of information theory: the arbitrarily varying channel (AVC) and the wiretap channel. This leads to the arbitrarily varying wiretap channel (AVWC): A sender would like to send information to a receiver through a noisy channel. Communication over this channel is subject to two difficulties. First, there is a second receiver, called an eavesdropper, which obtains its own noisy version of the channel inputs and should not be able to decode any information. Second, the state of the channels both to the intended receiver as well as to the eavesdropper can vary arbitrarily over time. Neither the sender nor the intended receiver know the true channel state. For a blocklength n, this means that the probability of the intended receiver obtaining the output sequence $y^n = (y_1, \dots, y_n)$ and the eavesdropper receiving $z^n = (z_1, \dots, z_n)$ given that $x^n = (x_1, \dots, x_n)$ was input to the channel is contained in the family

$$\Big\{U^n_{s^n}(y^n, z^n|x^n) = \prod_{i=1}^{n} U_{s_i}(y_i, z_i|x_i) : \\ s^n = (s_1, \dots, s_n) \in \mathcal{S}^n\Big\}. \tag{1}$$

Here, $\mathcal{S}$ is the finite state set and $\{U_s(\cdot, \cdot|\cdot) : s \in \mathcal{S}\}$ a family of stochastic matrices, which thus determines the AVWC.

One could regard the varying channel states as determined by nature. However, we will interpret them as the result of jamming from an intruder. So henceforth, we shall view the AVWC as a channel under two attacks at the same time: one passive (eavesdropping), one active (jamming).

The study of correlated random coding capacities in their own right instead of as mathematical tools applied in the proofs of uncorrelated coding capacity theorems is motivated by arbitarily varying channels (AVCs), which are AVWCs without the eavesdropper. By uncorrelated codes, we mean that sender and receiver have agreed on a procedure (f, ϕ) of data manipulation prior to transmission. Here, f is a possibly stochastic mapping from the messages to the channel inputs of a fixed blocklength, ϕ reverts channel outputs into messages. For transmission, each node separately executes its part of this procedure without relying on any further resources, in particular no common resources. What we call correlated random coding is usually called random coding and has been used as a mathematical tool ever since Shannon's 1948 paper [23]. Operationally, it means that sender and receiver agree on a family of deterministic codes $\{(f^\gamma, \phi^\gamma) : \gamma \in \Gamma\}$. Before communication, a random experiment following the distribution μ on Γ is performed. The outcome, say γ, is revealed to sender and intended receiver which then apply the deterministic code (f^γ, ϕ^γ).

It was already observed by Blackwell *et al.* [7] that whether correlated randomness is available to sender and receiver can be crucial when it comes to the AVC capacity. In fact, AVCs exhibit a dichotomy [1]: Their capacity for deterministic coding either equals their capacity for correlated random

Manuscript received October 30, 2014; revised April 3, 2016; accepted April 8, 2016. Date of publication May 10, 2016; date of current version June 14, 2016. J. Nötzel and H. Boche were supported by the Bundesministerium für Bildung und Forschung under Grant 01BQ1050. J. Nötzel was supported in part by the Generalitat de Catalunya through the CIRIT Project under Grant 2014 SGR 966, in part by the Deutsche Forschungsgemeinschaft under Grant NO 1129/1-1, and in part by the Ministerio de Economía y Competitividad under Grant FIS2013-40627-P. This paper was presented at the 2015 IEEE International Symposium on Information Theory.

M. Wiese is with the ACCESS Linnaeus Center, KTH Royal Institute of Technology, Stockholm SE-10044, Sweden (e-mail: moritzw@kth.se).

J. Nötzel was with the Lehrstuhl für Theoretische Informationstechnik, Technische Universität München, Munich 80333, Germany. He is now with the Física Teòrica: Informació i Fenòmens Quàntics Unit, Departament de Física, Universitat Autònoma de Barcelona, 08193 Bellaterra (Barcelona), Spain (e-mail: janis.notzel@uab.cat).

H. Boche is with the Lehrstuhl für Theoretische Informationstechnik, Technische Universität München, Munich 80333, Germany (e-mail: boche@tum.de).

Communicated by Y. Liang, Associate Editor for Shannon Theory.

Digital Object Identifier 10.1109/TIT.2016.2565482

coding or it equals zero. Csiszár and Narayan have identified the distinguishing property [12], called symmetrizability (a concept originally introduced by Ericson [15]). Without the use of correlated random coding, a symmetrizable AVC is useless; no message transmission is possible.

Thus one is led to regarding correlated randomness as an additional resource for communication. This resource can make communication possible where it is impossible without. Of course, it is important that the jammer has no access to this resource, i.e. that it does not know the outcome of the random experiment common to sender and receiver. In this paper, we will apply two strong secrecy criteria and show that the corresponding capacities for correlated random coding coincide. The first of these criteria is that

$$\max_{s^n} \sum_{\gamma} I(M \wedge Z^{\gamma}_{s^n})\mu(\gamma) \tag{2}$$

be small, where M is the message chosen uniformly at random and $Z^{\gamma}_{s^n}$ is the eavesdropper's output if the state sequence is s^n and the deterministic code $(f^{\gamma}, \phi^{\gamma})$ has been selected. This criterion was applied in [4] and [21]. The second, stronger one requires

$$\max_{s^n} \max_{\gamma} I(M \wedge Z^{\gamma}_{s^n}) \tag{3}$$

to be small. Both secrecy criteria assume that the eavesdropper knows the realization of the correlated randomness. This means that we have to assume the active and passive attacks to be uncoordinated in the sense that the eavesdropper does not inform the jammer about its knowledge of the correlated randomness.

We are not the first to study the capacity of the AVWC. A study of the Gaussian MIMO wiretap channel where the channel to the eavesdropper is arbitrarily varying has been done in [19] and [20]. Earlier approaches to the discrete AVWC as defined in (1) can be found in [4] and [21], which studied the secrecy capacity achieved by correlated random coding and used (2) as secrecy criterion. In both papers, closed-form secrecy capacity results could only be given after imposing additional conditions. In a recent preprint [17], single-letter lower and upper bounds for the AVWC with state constraints are derived which coincide in the case that the state is constrained to be typical with respect to a given probability distribution on the state alphabet.

The main result of this paper will be a complete characterization of the correlated random coding secrecy capacity under both criteria (2) and (3). The capacity formula we find is multi-letter. It was found in [4] for special AVWCs where there is a "best channel to the eavesdropper" and reduces to a single-letter formula under certain degradedness conditions as required in [21]. It is not clear whether a generally applicable single-letter formula exists at all. Still, the multi-letter formula allows for the approximate computation of the secrecy capacity up to a given complexity. However, this is not our main concern, so we do not provide any relation between complexity and approximation goodness.

With the help of the multi-letter formula, it can also be shown that the correlated random coding secrecy capacity is continuous in the channel. Thus small errors in the description of the family (1) do not have severe consequences for the capacity. If the capacity formula were not continuous, the channel would in general have to be estimated with infinite precision in order to meaningfully apply the capacity formula. The continuity of the correlated random coding secrecy capacity becomes even more remarkable as very simple examples with $|\mathcal{S}| = 2$ have been given in [9] which show that the uncorrelated coding secrecy capacity is a discontinuous function of the AVWC.

For the achievability part of the capacity theorem, we follow Ahlswede's strategy of deriving correlated random coding achievability results for AVCs from uncorrelated coding capacity results for compound channels. (In contrast to an AVC, a compound channel does not change its state during the transmission of a codeword.) This technique is known as the "robustification technique". Sender and receiver of an AVC randomly permute an uncorrelated code for a certain compound channel induced by the AVC and thus obtain a correlated random code with negligibly larger average error.

When applying the robustification technique to AVWCs, one has to take the secrecy criterion into account. As seen in [4], this requires a "best channel to the eavesdropper" if one assumes the channel to the eavesdropper to be compound as well. The central idea of our proof is to introduce the compound-arbitrarily varying wiretap channel (CAVWC). This channel is compound from sender to intended receiver and arbitrarily varying from sender to eavesdropper. We derive the uncorrelated coding secrecy capacity of this channel. After robustification, this also turns out to be the correlated random coding secrecy capacity of the AVWC.

We prove the achievability result for the CAVWC by random coding following Devetak [13]. This technique takes a resolvability approach to proving secrecy, cf. the discussion of resolvability and "capacity-based" approaches by Bloch and Laneman [8]. However, it does not follow an information spectrum approach like the techniques presented in [8]. To our knowledge, those techniques have not yet been shown to be able to handle arbitrarily varying channels. As the number of AVWC channel states grows exponentially with blocklength, very tight probability estimates have to be obtained from random coding. Devetak's method [13], originally in the language of quantum information theory, provides such estimates and was already applied in [4], [5], and [24] in a classical information theory setting.

In [10], an a priori upper bound on the amount of correlated randomness required to achieve the correlated random coding secrecy capacity was found. Such a bound is necessary for the converse of the correlated random coding secrecy capacity theorem for the AVWC. The reason for this is that the use of correlated randomness prohibits a straightforward application of the data processing inequality.

In a follow-up work [22] to this paper, we extend our analysis of the AVWC. We study the case when the eavesdropper has no knowledge of the correlated randomness and the case when there is no correlated randomness at all, i.e. the uncorrelated coding secrecy capacity.

Paper Outline: In Section II, we set the notation and give basic definitions. In Section III we define the AVWC and state the coding problem and the main result. Section IV discusses the main result of Section III. Section V introduces the CAVWC mentioned in the introduction, states the CAVWC coding problem and the corresponding secrecy capacity theorem. Section VI contains the proof of the achievability part of the coding theorem for the CAVWC. The achievability part of the correlated random coding theorem for the AVWC is derived from the achievability part of the coding theorem for the CAVWC in Section VII. Section VIII contains the converses. In Section IX, a short discussion concludes the paper. Several proofs are collected in the appendices.

II. Notation and Basic Definitions

Logarithms denoted by log are taken to the base 2; correspondingly, we set $\exp(x) = 2^x$. The cardinality of a finite set $\mathcal{A}$ is written $|\mathcal{A}|$. For a subset $\mathcal{E}$ of $\mathcal{A}$, we write $\mathcal{E}^c := \mathcal{A} \setminus \mathcal{E}$. The *indicator function* $\mathbb{1}_{\mathcal{E}}$ assumes the value 1 for arguments contained in $\mathcal{E}$ and 0 else. For n-tuples contained in $\mathcal{A}^n$, we write $x^n := (x_1, \dots, x_n) \in \mathcal{A}^n$.

The set of probability measures on the finite set $\mathcal{A}$ is denoted by $\mathcal{P}(\mathcal{A})$. For $P \in \mathcal{P}(\mathcal{A})$, we define the n-fold product measure $P^n \in \mathcal{P}(\mathcal{A}^n)$ by $P^n(x^n) := \prod_i P(x_i)$. We write stochastic matrices $\{W(b|a) : a \in \mathcal{A}, b \in \mathcal{B}\}$ with input alphabet $\mathcal{A}$ and output alphabet $\mathcal{B}$ as mappings $W : \mathcal{A} \longrightarrow \mathcal{P}(\mathcal{B})$. A nonnegative measure on $\mathcal{A}$ is a vector $(\mu(a))_{a\in\mathcal{A}}$ with $\mu(a) \geq 0$ for all $a \in \mathcal{A}$. Every probability measure is a nonnegative measure. The total variation distance of two nonnegative measures μ, ν on $\mathcal{A}$ is defined by $\|\mu - \nu\| := \sum_{a\in\mathcal{A}} |\mu(a) - \nu(a)|$.

If $\bar{X}, \bar{Y}$ are random variables, then we write the distribution of $\bar{X}$ as $P_{\bar{X}}$, the joint distribution of $\bar{X}$ and $\bar{Y}$ as $P_{\bar{X}\bar{Y}}$ and the conditional distribution of $\bar{X}$ given $\bar{Y}$ as $P_{\bar{X}|\bar{Y}}$.

For a sequence $x^n = (x_1, \dots, x_n) \in \mathcal{A}^n$ and $a \in \mathcal{A}$, the number $N(a|x^n)$ indicates the number of coordinates x_i of x^n with $x_i = a$. The type of x^n is the probability measure $q \in \mathcal{P}(\mathcal{A})$ defined by $q(a) := N(a|x^n)/n$. The set of all possible types of sequences of length n is denoted by $\mathcal{P}_0^n(\mathcal{A})$. For $\delta > 0$ and an $\mathcal{A}$-valued random variable $\bar{X}$, we define the typical set $\mathcal{T}^n_{\bar{X},\delta} \subset \mathcal{A}^n$ as the set of those $x^n \in \mathcal{A}^n$ satisfying the two conditions

$$\left|\frac{1}{n}N(a|x^n) - P_{\bar{X}}(a)\right| < \delta,$$
$$N(a|x^n) = 0 \quad \text{if } P_{\bar{X}}(a) = 0$$

for every $a \in \mathcal{A}$. For $\delta > 0$, an $\mathcal{A} \times \mathcal{B}$-valued random variable $(\bar{X}, \bar{Y})$ with joint distribution $P_{\bar{X}\bar{Y}}$ and an element x^n of $\mathcal{A}^n$, we define the conditionally typical set $\mathcal{T}^n_{\bar{Y}|\bar{X},\delta}(x^n)$ as the set of those $y^n \in \mathcal{B}^n$ satisfying the two conditions

$$\left|\frac{1}{n}N(a,b|x^n,y^n) - P_{\bar{Y}|\bar{X}}(b|a)\frac{1}{n}N(a|x^n)\right| < \delta,$$
$$N(a,b|x^n,y^n) = 0 \quad \text{if } P_{\bar{Y}|\bar{X}}(b|a) = 0$$

for all $a \in \mathcal{A}, b \in \mathcal{B}$.

III. Arbitrarily Varying Wiretap Channels

Let $\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{S}$ be finite sets. For every $s \in \mathcal{S}$, let a stochastic matrix $W_s : \mathcal{A} \to \mathcal{P}(\mathcal{B})$ and another stochastic matrix $V_s : \mathcal{A} \to \mathcal{P}(\mathcal{C})$ be given. For a number n and $x^n \in \mathcal{A}^n$, $y^n \in \mathcal{B}^n$, $s^n \in \mathcal{S}^n$, define

$$W^n_{s^n}(y^n|x^n) := \prod_{i=1}^n W_{s_i}(y_i|x_i).$$

We denote the family $\{W^n_{s^n} : s^n \in \mathcal{S}^n, n = 1, 2, \dots\}$ by $\mathfrak{W}$. In analogy to $W^n_{s^n}(y^n|x^n)$, we define $V^n_{s^n}(z^n|x^n)$ for $z^n \in \mathcal{C}^n$ and denote the corresponding family $\{V^n_{s^n} : s^n \in \mathcal{S}^n, n = 1, 2, \dots\}$ by $\mathfrak{V}$. We sometimes prefer to write $V^n(z^n|x^n, s^n)$ instead of $V^n_{s^n}(z^n|x^n)$. We call the pair $(\mathfrak{W}, \mathfrak{V})$ an *Arbitrarily Varying Wiretap Channel (AVWC)*. $\mathcal{S}$ is called the *state set* of $(\mathfrak{W}, \mathfrak{V})$.

Remark 1: One checks easily that the representation of an AVWC as a pair $(\mathfrak{W}, \mathfrak{V})$ is possible without losing generality. In general, any state $s \in \mathcal{S}$ together with an input $a \in \mathcal{A}$ will lead to a joint output distribution $U_s(\cdot, \cdot|a)$. But the performance of any of the codes defined below is measured with respect to the marginal output distributions $W_s(\cdot|a)$ and $V_s(\cdot|a)$. Thus for the purpose of this paper, all AVWCs with the same marginals $\mathfrak{W}$ and $\mathfrak{V}$ are equivalent.

An uncorrelated (n, J_n)-code $\mathcal{K}_n$ for the AVWC $(\mathfrak{W}, \mathfrak{V})$ consists of a stochastic encoder $E : \{1, \dots, J_n\} \to \mathcal{P}(\mathcal{A}^n)$ and a collection of mutually disjoint sets $\{\mathcal{D}_j \subset \mathcal{B}^n : 1 \leq j \leq J_n\}$. We abbreviate $\mathcal{J}_n := \{1, \dots, J_n\}$. Together with an AVWC $(\mathfrak{W}, \mathfrak{V})$, any uncorrelated (n, J_n)-code $\mathcal{K}_n$ defines a canonical family

$$\mathcal{F}(\mathcal{K}_n, \mathfrak{W}, \mathfrak{V}) := \{M^n, X^n, Y^n_{s^n}, Z^n_{s^n}, \hat{M}^n_{s^n} : s^n \in \mathcal{S}^n\} \tag{4}$$

of random variables, with M^n and $\hat{M}^n_{s^n}$ assuming values in $\mathcal{J}_n$, the values of X^n in $\mathcal{A}^n$, those of $Y^n_{s^n}$ in $\mathcal{B}^n$, those of $Z^n_{s^n}$ in $\mathcal{C}^n$, and such that for every $s^n \in \mathcal{S}^n$ the distribution of $(M^n, X^n, Y^n_{s^n}, Z^n_{s^n}, \hat{M}^n_{s^n})$ equals

$$\begin{aligned} &P_{M^n X^n Y^n_{s^n} Z^n_{s^n} \hat{M}^n_{s^n}}(j, x^n, y^n, z^n, \hat{j}) \\ &\quad = \frac{1}{J_n} E(x^n|j) W^n_{s^n}(y^n|x^n) V^n_{s^n}(z^n|x^n) \mathbb{1}_{\mathcal{D}_{\hat{j}}}(y^n). \end{aligned}$$

Recall that we incur no loss of generality by defining $Y^n_{s^n}$ and $Z^n_{s^n}$ to be independent conditional on X^n, as the joint distribution of $Y^n_{s^n}$ and $Z^n_{s^n}$ will never play any role (cf. Remark 1). The average error of $\mathcal{K}_n$ is given by

$$e(\mathcal{K}_n) := \max_{s^n \in \mathcal{S}^n} \mathbb{P}[M^n \neq \hat{M}^n_{s^n}].$$

Definition 2: A non-negative number R_S is an *achievable uncorrelated coding secrecy rate for the AVWC* $(\mathfrak{W}, \mathfrak{V})$ if there exists a sequence $(\mathcal{K}_n)_{n=1}^\infty$ of uncorrelated (n, J_n)-codes such that

$$\liminf_{n\to\infty} \frac{1}{n} \log J_n \geq R_S, \tag{5}$$

$$\lim_{n\to\infty} e(\mathcal{K}_n) = 0, \tag{6}$$

$$\lim_{n\to\infty} \max_{s^n\in\mathcal{S}^n} I(M^n \wedge Z^n_{s^n}) = 0. \tag{7}$$

The *uncorrelated coding secrecy capacity of* $(\mathfrak{W}, \mathfrak{V})$ is the supremum of all achievable secrecy rates R_S and is denoted by $C_S(\mathfrak{W}, \mathfrak{V})$.

We state this definition for reference. In this paper, we do not study $C_S(\mathfrak{W}, \mathfrak{V})$. This is done in [22]. Note the different roles the families $\mathfrak{W}$ and $\mathfrak{V}$ play. $\mathfrak{W}$ is an *Arbitrarily Varying Channel (AVC)* from a sender with alphabet $\mathcal{A}$ to a receiver with alphabet $\mathcal{B}$. Messages are supposed to be sent over this AVC in such a way that only a small, asymptotically negligible average error is incurred. This is reflected in condition (6). This communication is subject to an additional secrecy condition. An eavesdropper obtains a noisy version of the sender's channel inputs via the AVC $\mathfrak{V}$. Condition (7) guarantees secrecy no matter what the channel state is.

For given (n, J_n), we assume that the set of uncorrelated (n, J_n)-codes is indexed by the set Γ_n. That means that the set of all uncorrelated (n, J_n)-codes (with given channel input and output alphabets $\mathcal{A}$ and $\mathcal{B}$) has the form $\{\mathcal{K}_n(\gamma) : \gamma \in \Gamma_n\}$. For the uncorrelated (n, J_n)-code $\mathcal{K}_n(\gamma)$, with $\gamma \in \Gamma_n$, we write for the canonical family of random variables

$$\begin{aligned}&\mathcal{F}(\mathcal{K}_n(\gamma), \mathfrak{W}, \mathfrak{V})\\ &= \{M^n, X^n(\gamma), Y^n_{s^n}(\gamma), Z^n_{s^n}(\gamma), \hat{M}^n_{s^n}(\gamma) : s^n \in \mathcal{S}^n, \gamma \in \Gamma_n\}.\end{aligned}$$

A correlated random (n, J_n)-code $\mathcal{K}_n^{\mathrm{ran}}$ for the AVWC $(\mathfrak{W}, \mathfrak{V})$ then is given by a finitely supported[1] random variable G_n on Γ_n independent of all canonical families of random variables $\mathcal{F}(\mathcal{K}_n(\gamma), \mathfrak{W}, \mathfrak{V})$. In other words, G_n randomly chooses an uncorrelated (n, J_n)-code out of all possible ones and is independent of the message random variable, the randomness in the chosen stochastic encoder and the channel noise. The average error $e(\mathcal{K}_n^{\mathrm{ran}})$ is defined as

$$\begin{aligned}e(\mathcal{K}_n^{\mathrm{ran}}) &:= \max_{s^n \in \mathcal{S}^n} \mathbb{P}[M^n \neq \hat{M}^n_{s^n}(G_n)]\\ &= \max_{s^n \in \mathcal{S}^n} \sum_{\gamma \in \Gamma_n} \mathbb{P}[M^n \neq \hat{M}^n_{s^n}(\gamma)] P_{G_n}(\gamma),\end{aligned}$$

where $\sum_{\gamma \in \Gamma_n} a(\gamma) P_{G_n}(\gamma)$ is short for the finite sum $\sum_{\gamma \in \mathrm{supp}(G_n)} a(\gamma)\, P_{G_n}(\gamma)$.

In the case of correlated random codes, we consider two secrecy criteria, leading to two different notions of achievable rate.

Definition 3: A non-negative number R_S is called an *achievable correlated random coding mean secrecy rate for the AVWC* $(\mathfrak{W}, \mathfrak{V})$ if there exists a sequence $(\mathcal{K}_n^{\mathrm{ran}})_{n=1}^{\infty}$ of correlated random (n, J_n)-codes such that

$$\liminf_{n\to\infty} \frac{1}{n} \log J_n \geq R_S, \tag{8}$$

$$\lim_{n\to\infty} e(\mathcal{K}_n^{\mathrm{ran}}) = 0, \tag{9}$$

$$\lim_{n\to\infty} \max_{s^n \in \mathcal{S}^n} I(M^n \wedge Z^n_{s^n}(G_n) | G_n) = 0. \tag{10}$$

The supremum of all achievable secrecy rates for correlated random codes is called the *correlated random coding mean secrecy capacity of* $(\mathfrak{W}, \mathfrak{V})$ and denoted by $C^{\mathrm{mean}}_{S,\mathrm{ran}}(\mathfrak{W}, \mathfrak{V})$.

Definition 4: A non-negative number R_S is called an *achievable correlated random coding maximum secrecy rate for the AVWC* $(\mathfrak{W}, \mathfrak{V})$ if there exists a sequence $(\mathcal{K}_n^{\mathrm{ran}})_{n=1}^{\infty}$ of correlated random (n, J_n)-codes such that (8) and (9) hold and

$$\lim_{n\to\infty} \max_{s^n \in \mathcal{S}^n} \max_{\gamma \in \mathrm{supp}(G_n)} I(M^n \wedge Z^n_{s^n}(\gamma)) = 0. \tag{11}$$

The supremum of all achievable correlated random coding maximum secrecy rates is called the *correlated random coding maximum secrecy capacity of* $(\mathfrak{W}, \mathfrak{V})$ and denoted by $C^{\max}_{S,\mathrm{ran}}(\mathfrak{W}, \mathfrak{V})$.

Remark 5: It is immediately clear that $C^{\mathrm{mean}}_{S,\mathrm{ran}}(\mathfrak{W}, \mathfrak{V}) \geq C^{\max}_{S,\mathrm{ran}}(\mathfrak{W}, \mathfrak{V})$. Note that if G_n is deterministic, i.e., there is no correlated randomness, then both criteria (10) and (11) coincide with (7).

The secrecy capacities for correlated random codes are characterized by a multi-letter formula, extending the results of [4]. We set

$$\begin{aligned}&R^*_S(\mathfrak{W}, \mathfrak{V})\\ &:= \lim_{k\to\infty} \frac{1}{k} \sup_{\mathcal{Q}_k} \left(\min_{q \in \mathcal{P}(\mathcal{S})} I(\bar{U} \wedge \bar{Y}^k_q) - \max_{s^k \in \mathcal{S}^k} I(\bar{U} \wedge \bar{Z}^k_{s^k}) \right)\end{aligned} \tag{12}$$

where the set $\mathcal{Q}_k$ over which the supremum is taken contains those families of random variables

$$\{\bar{U}, \bar{X}^k, \bar{Y}^k_q, \bar{Z}^k_{s^k} : q \in \mathcal{P}(\mathcal{S}), s^k \in \mathcal{S}^k\}$$

which satisfy that $\bar{U}$ assumes values in some finite subset of the integers, the values of $\bar{X}^k$ lie in $\mathcal{A}^k$, those of $\bar{Y}^k_q$ in $\mathcal{B}^k$, those of $\bar{Z}^k_{s^k}$ in $\mathcal{C}^k$, and such that for every $q \in \mathcal{P}(\mathcal{S})$ and $s^k \in \mathcal{S}^k$,

$$\begin{aligned}&P_{\bar{U}\bar{X}^k\bar{Y}^k_q\bar{Z}^k_{s^k}}(u, x^k, y^k, z^k) = P_{\bar{U}}(u) P_{\bar{X}^k|\bar{U}}(x^k|u)\\ &\quad\times \left(\prod_{i=1}^{k} \left[\sum_{s \in \mathcal{S}} q(s) W_s(y_i|x_i) \right] \right) V^k_{s^k}(z^k|x^k).\end{aligned} \tag{13}$$

$P_{\bar{U}}$ and $P_{\bar{X}^k|\bar{U}}$ may be arbitrary probability distributions and stochastic matrices, respectively.

Theorem 6: For the AVWC $(\mathfrak{W}, \mathfrak{V})$, we have

$$C^{\mathrm{mean}}_{S,\mathrm{ran}}(\mathfrak{W}, \mathfrak{V}) = C^{\max}_{S,\mathrm{ran}}(\mathfrak{W}, \mathfrak{V}) = R^*_S(\mathfrak{W}, \mathfrak{V}).$$

The achievability part of Theorem 6 is proved in Section VII, its converse is proved in Section VIII.

Remark 7:

1) It is shown exactly as in [5], using Fekete's lemma [16] (a proof of which can be found in [11, Lemma 11.2]), that the limit on the right-hand side of (12) indeed exists and can in fact be replaced by a supremum.
2) For given k, the cardinality of $\mathcal{U}$ can be restricted to $|\mathcal{A}|^k$. This can be proved almost exactly as in the proof of [11, Th. 17.11]. The supremum in (12) then becomes a maximum.
3) If for $q \in \mathcal{P}(\mathcal{S})$ we define $W_q(b|a) := \sum_s q(s) W_s(b|a)$, the conditional probability of $\bar{Y}^k_q$ given $\bar{X}^k$ in (13) satisfies

$$P_{\bar{Y}^k_q|\bar{X}^k}(y^k|x^k) = \prod_{k=1}^{k} W_q(y_i|x_i) =: W^k_q(y^k|x^k).$$

The family $\{W^n_q : q \in \mathcal{P}(\mathcal{S}), n = 1, 2, \ldots\}$ is a memoryless channel which does not change its state during the transmission of a codeword. Such channels will appear later under the name of *compound channel.*

[1] "Finitely supported" means that the set $\mathrm{supp}(G_n) := \{\gamma \in \Gamma_n : P_{G_n}(\gamma) > 0\}$ called the *support of* G_n is finite.

4) In the proof of [22, Th. 1], it is exploited that $R_S^*(\mathfrak{W},\mathfrak{V})$ does not change if the sets over which the minimum and maximum are taken in (12) are replaced by different, but related ones. Note the equality $\max_{s^k\in\mathcal{S}^k} I(\bar{U}\wedge\bar{Z}^k_{s^k}) = \max_{\tilde{q}\in\mathcal{P}(\mathcal{S}^k)} I(\bar{U}\wedge\bar{Z}^k_{\tilde{q}})$, where $P_{\bar{Z}^k_{\tilde{q}}|\bar{X}^k}(z^k|x^k) = \sum_{s^k\in\mathcal{S}^k}\tilde{q}(s^k)V^k_{s^k}(z^k|x^k)$. This equality is due to the convexity of mutual information in the channel. On the other hand, if one analogously defines $P_{\bar{Y}^k_{\tilde{q}}|\bar{X}^k}(y^k|x^k) = \sum_{s^k\in\mathcal{S}^k}\tilde{q}(s^k)V^k_{s^k}(y^k|x^k)$ for $\tilde{q}\in\mathcal{P}(\mathcal{S}^k)$ and replaces the minimum over $q\in\mathcal{P}(\mathcal{S})$ by a minimum over $\tilde{q}\in\mathcal{P}(\mathcal{S}^k)$, then at first it is only clear that $\min_{\tilde{q}\in\mathcal{P}(\mathcal{S}^k)} I(\bar{U}\wedge\bar{Y}^k_{\tilde{q}}) \le \min_{q\in\mathcal{P}(\mathcal{S})} I(\bar{U}\wedge\bar{Y}^k_q)$. But one can actually prove that $R_S^*(\mathfrak{W},\mathfrak{V})$ equals

$$\lim_{k\to\infty}\frac{1}{k}\sup_{\tilde{\mathcal{Q}}_k}\left(\min_{\tilde{q}\in\mathcal{P}(\mathcal{S}^k)} I(\bar{U}\wedge\bar{Y}^k_{\tilde{q}}) - \max_{\tilde{q}\in\mathcal{P}(\mathcal{S}^k)} I(\bar{U}\wedge\bar{Z}^k_{\tilde{q}})\right), \tag{14}$$

where $\mathcal{Q}_k$ is naturally extended to the set $\tilde{\mathcal{Q}}_k$ which also contains the output random variables generated by the "states" $\tilde{q}$ as defined above. The equality is a consequence of the converse proof of Theorem 6, see Remark 17 in Section VIII.

5) Comparison of the right-hand side of (12) with the capacity expressions derived in [8] suggests that the terms $\min_{q\in\mathcal{P}(\mathcal{S})} I(\bar{U}\wedge\bar{Y}^k_q)$ are related to an inf-information rate for the AVC $\mathfrak{W}$ and $\max_{s^k\in\mathcal{S}^k} I(\bar{U}\wedge\bar{Z}^k_{s^k})$ to a sup-information rate for the AVC $\mathfrak{V}$, see also [18]. However, as AVCs have not yet been treated in the framework of information spectrum theory, this remains speculation for the time being.

IV. Discussion of Theorem 6

A. Multi-Letter vs. Single-Letter

The bound from Remark 7-2) on the size of $\mathcal{U}$ for fixed k does not give a general upper bound on the cardinality of the auxiliary alphabet $\mathcal{U}$. It could still be helpful in calculations of $R_S^*(\mathfrak{W},\mathfrak{V})$ if one knows from other arguments that there exists a k_0 such that, for $k\ge k_0$,

$$\frac{1}{k}\sup_{\mathcal{Q}_k}\left(\min_{q\in\mathcal{P}(\mathcal{S})} I(\bar{U}\wedge\bar{Y}^k_q) - \max_{s^k\in\mathcal{S}^k} I(\bar{U}\wedge\bar{Z}^k_{s^k})\right)$$

is sufficiently close to $R_S^*(\mathfrak{W},\mathfrak{V})$. From Remark 7-1) it follows that this approach would give a lower bound on the secrecy capacity. Note that it is not at all clear whether a single-letter characterization of $R_S^*(\mathfrak{W},\mathfrak{V})$ is available. In the case of the unavailability of a single-letter capacity expression, only approximate calculations of capacity are possible.

That the above multi-letter characterization can lead to further insights into the nature of AVWCs can be seen in Subsection IV-C, where the continuity of $R_S^*(\mathfrak{W},\mathfrak{V})$ in $(\mathfrak{W},\mathfrak{V})$ is shown. To show this a priori, i.e. without having the multi-letter expression for capacity, seems to be very hard. With the formula at hand, however, it can be done. For the uncorrelated coding secrecy capacity $C_S(\mathfrak{W},\mathfrak{V})$ (see Definition 2), a similar study of continuity is performed in [22], also on the basis of the multi-letter formula.

A single-letter formula for $C_{S,\mathrm{ran}}^{\mathrm{mean}}(\mathfrak{W},\mathfrak{V})$ has been given in [21] for AVWCs which satisfy certain conditions. We now present these conditions and show that if they are satisfied, the formula found in [21] coincides with $R_S^*(\mathfrak{W},\mathfrak{V})$, which then becomes single-letter.

The first condition of [21] is that $(\mathfrak{W},\mathfrak{V})$ be *strongly degraded with independent states*. This means

- that $\mathcal{S}=\mathcal{S}_1\times\mathcal{S}_2$ and that the families $\{W_{(s_1,s_2)}:(s_1,s_2)\in\mathcal{S}_1\times\mathcal{S}_2\}$ and $\{V_{(s_1,s_2)}:(s_1,s_2)\in\mathcal{S}_1\times\mathcal{S}_2\}$ of stochastic matrices determining $\mathfrak{W}$ and $\mathfrak{V}$ satisfy $W_{(s_1,s_2)}=W_{s_1}$ and $V_{(s_1,s_2)}=V_{s_2}$ for all (s_1,s_2); and
- that for every $q_1\in\mathcal{P}(\mathcal{S}_1)$ and $q_2\in\mathcal{P}(\mathcal{S}_2)$, the matrix V_{q_2} should be a degraded version of W_{q_1}, where

$$W_{q_1}(y|x) = \sum_{s_1\in\mathcal{S}_1} W_{s_1}(y|x)q_1(s_1),$$
$$V_{q_2}(z|x) = \sum_{s_2\in\mathcal{S}_2} V_{s_2}(z|x)q_2(s_2),$$

and V_{q_2} is a degraded version of W_{q_1} if there exists a stochastic matrix $T_{q_1q_2}:\mathcal{B}\to\mathcal{C}$ such that

$$V_{q_2}(z|x) = \sum_y T_{q_1q_2}(z|y)W_{q_1}(y|x). \tag{15}$$

(Observe: It is sufficient to require (15) to hold only for $s_2\in\mathcal{S}_2$ and $q_1\in\mathcal{P}(\mathcal{S}_1)$. The validity of (15) for all $q_1\in\mathcal{P}(\mathcal{S}_1)$ and $q_2\in\mathcal{P}(\mathcal{S}_2)$ then follows upon setting $T_{q_1q_2}(z|y) := \sum_{s_2} q_2(s_2)T_{q_1s_2}(z|y)$ for all $y\in\mathcal{B}$, $z\in\mathcal{C}$. Thus the function $(q_1,q_2)\mapsto T_{q_1q_2}$ can without loss of generality be assumed to be linear in q_2. This is not possible for q_1, as can be seen from analyzing Example 3 in [21].)

The second condition of [21] is essentially the *best channel to the eavesdropper* condition from [4], so we will henceforth call it this way. It requires that there exists an $s_*\in\mathcal{S}_2$ such that for all $s_2\in\mathcal{S}_2$, the channel V_{s_2} is a degraded version of V_{s_*}, with degradedness here defined analogously to (15). (The general definition of "best channel to the eavesdropper" in [4] and [21] does not require independent states.)

Corollary 1: If the AVWC $(\mathfrak{W},\mathfrak{V})$ is strongly degraded with independent states and has a best channel to the eavesdropper, then

$$R_S^*(\mathfrak{W},\mathfrak{V}) = \max_{\mathcal{Q}_1^*}\left(\min_{q_1\in\mathcal{P}(\mathcal{S}_1)} I(\bar{X}\wedge\bar{Y}_{q_1}) - \max_{s_2\in\mathcal{S}_2} I(\bar{X}\wedge\bar{Z}_{s_2})\right) \tag{16}$$

where the set $\mathcal{Q}_1^*$ over which the supremum is taken contains those families of random variables $\{\bar{X},\bar{Y}_{q_1},\bar{Z}_{s_2}\}$ which for every $q_1\in\mathcal{P}(\mathcal{S}_1)$ and $s_2\in\mathcal{S}_2$ satisfy

$$P_{\bar{X}\bar{Y}_{q_1}\bar{Z}_{s_2}}(x,y,z) = P_{\bar{X}}(x)W_{q_1}(y|x)V_{s_2}(z|x)$$

and where $\bar{X}$ is an arbitrary $\mathcal{A}$-valued random variable.

Proof: See Appendix A. ■

A single-letter capacity expression is given by [17] in the case that the state sequences are constrained to be typical with respect to a single fixed probability distribution on $\mathcal{P}(\mathcal{S})$.

B. The Amount of Correlated Randomness

Next we ask how many values the correlated randomness variable should attain with positive probability in order for $C^{\text{mean}}_{S,\text{ran}}(\mathfrak{W},\mathfrak{V})$ and $C^{\max}_{S,\text{ran}}(\mathfrak{W},\mathfrak{V})$ to be achievable. Note that the definitions allow every kind of correlated randomness as long as it is finitely supported. In the achievability proof of Theorem 6, we shall see that the uniform distribution on a set of cardinality $n!$ is sufficient, where n is the blocklength of the code. But even without referring to any achievability proof, it is possible to show *a priori* that the size of this set can still be reduced considerably.

The lemma which provides this reduction was proved in [4]. We slightly reformulate it here for our purposes. Its essence is that every secrecy rate $R_S < C^{\text{mean}}_{S,\text{ran}}(\mathfrak{W},\mathfrak{V})$ is achievable using an amount of correlated randomness which grows on the order of $n\log n$, given arbitrary upper bounds on the average error and the mutual information between message random variable and eavesdropper output. As its proof is based on nothing but the definition of achievable correlated random mean secrecy rate, this is an a priori result on the structure of optimal correlated random codes independent of Theorem 6 or any characterization of $C^{\text{mean}}_{S,\text{ran}}(\mathfrak{W},\mathfrak{V})$. In particular, it can be applied in the converse of Theorem 6.

Lemma 8 ([4], Lemma 6): Let $R_S < C^{\text{mean}}_{S,\text{ran}}(\mathfrak{W},\mathfrak{V})$. For every $\varepsilon > 0$ there exists a sequence $\mathcal{K}^{\text{ran}}_n$ of correlated random (n, J_n)-codes which for $n \ge n(R_S,\varepsilon)$ satisfies

$$\frac{1}{n}\log J_n \ge R_S - \varepsilon, \tag{17}$$

$$e(\mathcal{K}^{\text{ran}}_n) \le \varepsilon, \tag{18}$$

$$\max_{s^n\in\mathcal{S}^n} I(M^n \wedge Z^n_{s^n}(G_n)|G_n) \le \varepsilon, \tag{19}$$

and

$$|\text{supp}(G_n)| \le \frac{2n\log|\mathcal{A}|}{\varepsilon}(1+n\log|\mathcal{S}|)+1. \tag{20}$$

For completeness, we give the proof of this lemma in Appendix B. As every achievable correlated random maximum secrecy rate also is an achievable correlated random mean secrecy rate, the lemma carries over to the case of rates $R_S < C^{\max}_{S,\text{ran}}(\mathfrak{W},\mathfrak{V})$.

For AVCs, the first correlated randomness reduction result was presented by Ahlswede in [1]. A similar result has been found recently [10] for AVWCs where secrecy is measured in terms of the weak secrecy criterion or in terms of variation distance. However, the results in [1] and [10] both assume that the average error (and in [10] also the respective secrecy measures) decrease to zero at exponential speed, which is not required in Definitions 3 and 4.

C. Model Robustness and Continuity

Here we study the continuity of the correlated random coding secrecy capacity function in the channel. Continuity is an important property of a capacity function, a fact which is sometimes overlooked because it is usually simple to prove the continuity of single-letter formulas using the uniform continuity of mutual information. The question becomes non-trivial in the case of a multi-letter capacity formula like $R^*_S(\mathfrak{W},\mathfrak{V})$.

Suppose the capacity function were not continuous and assume that one estimates a channel which is close to a point of discontinuity. Then this channel has to be estimated to a precision which might be higher than achievable in the estimation process, or even higher than a computer can handle with reasonable effort. Otherwise, the capacity expression obtained from the formula is next to useless for this particular channel, as all of its values in the neighborhood of the estimated channel could be the correct one, and this range of possible values could take on arbitrary form. From this point of view, the lack of continuity of a capacity function is more dramatic than a lacking single-letter expression, because a multi-letter formula still allows an approximate calculation, whereas approximation is not possible if the capacity function is discontinuous.

We shall show that the capacity functions $C^{\text{mean}}_{S,\text{ran}}(\mathfrak{W},\mathfrak{V})$ and $C^{\max}_{S,\text{ran}}(\mathfrak{W},\mathfrak{V})$ are continuous. The argumentation relies on the fact that we have an explicit formula for these, as $C^{\text{mean}}_{S,\text{ran}}(\mathfrak{W},\mathfrak{V}) = C^{\max}_{S,\text{ran}}(\mathfrak{W},\mathfrak{V}) = R^*_S(\mathfrak{W},\mathfrak{V})$. It is thus an example of the usefulness of a multi-letter formula.

Of course, the set of AVWCs with given in- and output alphabets has to be equipped with a metric in order to be able to talk about the continuity of capacity in the channel. Let $(\mathfrak{W},\mathfrak{V})$ and $(\tilde{\mathfrak{W}},\tilde{\mathfrak{V}})$ be two AVWCs with input alphabet $\mathcal{A}$ and output alphabets $\mathcal{B},\mathcal{C}$ for the legitimate receiver and the eavesdropper, respectively. Denote the finite state space of $(\mathfrak{W},\mathfrak{V})$ by $\mathcal{S}$ and the finite state space of $(\tilde{\mathfrak{W}},\tilde{\mathfrak{V}})$ by $\tilde{\mathcal{S}}$. We measure the distance of $(\mathfrak{W},\mathfrak{V})$ and $(\tilde{\mathfrak{W}},\tilde{\mathfrak{V}})$ by what is called the *Hausdorff distance* of two sets.

For two stochastic matrices $W,\tilde{W}:\mathcal{A}\to\mathcal{B}$, we define

$$\|W-\tilde{W}\|_o := \max_{a\in\mathcal{A}}\|W(\cdot|a)-\tilde{W}(\cdot|a)\|.$$

We define four asymmetric distances

$$d_{B,1}(\mathfrak{W},\tilde{\mathfrak{W}}) := \max_{\tilde{s}\in\tilde{\mathcal{S}}}\min_{s\in\mathcal{S}}\|W_s-\tilde{W}_{\tilde{s}}\|_o,$$

$$d_{B,2}(\mathfrak{W},\tilde{\mathfrak{W}}) := \max_{s\in\mathcal{S}}\min_{\tilde{s}\in\tilde{\mathcal{S}}}\|W_s-\tilde{W}_{\tilde{s}}\|_o,$$

and analogously define $d_{E,1}(\mathfrak{V},\tilde{\mathfrak{V}}), d_{E,2}(\mathfrak{V},\tilde{\mathfrak{V}})$ by replacing $W_s,\tilde{W}_{\tilde{s}}$ in the above definitions by $V_s,\tilde{V}_{\tilde{s}}$. Then the Hausdorff distance between $(\mathfrak{W},\mathfrak{V})$ and $(\tilde{\mathfrak{W}},\tilde{\mathfrak{V}})$ is defined by

$$\begin{aligned} d((\mathfrak{W},\mathfrak{V}),(\tilde{\mathfrak{W}},\tilde{\mathfrak{V}})) := \max\{&d_{B,1}(\mathfrak{W},\tilde{\mathfrak{W}}), d_{E,1}(\mathfrak{V},\tilde{\mathfrak{V}}),\\ &d_{B,2}(\mathfrak{W},\tilde{\mathfrak{W}}), d_{E,2}(\mathfrak{V},\tilde{\mathfrak{V}})\}.\end{aligned}$$

One checks easily that this is an actual metric on the set of finite-state AVWCs with the corresponding alphabets $\mathcal{A},\mathcal{B},\mathcal{C}$.

Building on Theorem 6, we now state the central result concerning the continuity of the correlated random capacities.

Theorem 9: $R^*_S(\mathfrak{W},\mathfrak{V})$ is continuous in $(\mathfrak{W},\mathfrak{V})$ with respect to the metric d. Thus, $C^{\text{mean}}_{S,\text{ran}}(\mathfrak{W},\mathfrak{V})$ and $C^{\max}_{S,\text{ran}}(\mathfrak{W},\mathfrak{V})$ are continuous functions of the channel.

The proof of this theorem only requires minor changes compared to that of [9, Th. 2] where the continuity of the capacity of the corresponding compound wiretap channel is shown.

In contrast to the correlated random coding secrecy capacity, the uncorrelated coding secrecy capacity of AVWCs (see Definition 2) is known to be discontinuous. This was shown in [9] with a very simple example on small alphabets and a state set of no more than two elements. Hence the continuity of the correlated random coding secrecy capacity becomes even more remarkable, especially as the previous subsection IV-B has shown that only very little correlated randomness is required to cause such a qualitative change of capacity functions. The exact characterization of the discontinuity points of the uncorrelated coding secrecy capacity is more intricate. It is discussed in depth in [22].

V. The Compound-Arbitrarily Varying Wiretap Channel

To establish Theorem 6, we use Ahlswede's robustification technique [2]. It was developed to turn deterministic codes for compound channels into correlated random codes for AVCs. It has already been applied in [4] to compound and arbitrarily varying wiretap channels. The difference of this paper's approach is that the channel from sender to eavesdropper will always be arbitrarily varying. Therefore it is no longer necessary to assume the existence of a best channel to the eavesdropper.

We now formalize the idea of having a compound channel from $\mathcal{A}$ to $\mathcal{B}$ and an arbitrarily varying channel from $\mathcal{A}$ to $\mathcal{C}$. Let $\mathcal{R}$ be any set. For every $r \in \mathcal{R}$, let $W_r : \mathcal{X} \longrightarrow \mathcal{Y}$ be a stochastic matrix. Set $W_r^n(y^n|x^n) = \prod_{i=1}^n W_r(y_i|x_i)$. Note that here, in contrast to the AVC, the channel state remains constant over time. This defines a *compound channel* $\overline{\mathfrak{W}} := \{W_r^n : r \in \mathcal{R}, n = 1, 2, \ldots\}$. Together with the AVC $\mathfrak{V}$ from the previous section, we obtain the *compound-arbitrarily varying wiretap channel* (CAVWC) $(\overline{\mathfrak{W}}, \mathfrak{V})$.

We apply uncorrelated (n, J_n)-codes for message transmission over $(\overline{\mathfrak{W}}, \mathfrak{V})$. Together with $(\overline{\mathfrak{W}}, \mathfrak{V})$, every (n, J_n)-code defines a canonical family of random variables

$$\begin{aligned}&\mathcal{F}(\mathcal{K}_n, \overline{\mathfrak{W}}, \mathfrak{V})\\&\quad := \{(M^n, X^n, Y_r^n, Z_{s^n}^n, \hat{M}_r^n) : r \in \mathcal{R}, s^n \in \mathcal{S}^n\},\end{aligned} \tag{21}$$

where M^n and $\hat{M}_r^n$ assume values in $\mathcal{J}_n$, the values of X^n lie in $\mathcal{A}^n$, those of Y_r^n in $\mathcal{B}^n$ and those of $Z_{s^n}^n$ in $\mathcal{C}^n$ and where for any $r \in \mathcal{R}$ and $s^n \in \mathcal{S}^n$

$$\begin{aligned}&P_{M^n X^n Y_r^n Z_{s^n}^n \hat{M}_r^n}(j, x^n, y^n, z^n, \hat{j})\\&\quad = \frac{1}{J_n} E(x^n|j) W_r^n(y^n|x^n) V_{s^n}^n(z^n|x^n) \mathbb{1}_{\mathcal{D}_{\hat{j}}}(y^n).\end{aligned}$$

For the uncorrelated (n, J_n)-code $\mathcal{K}_n$, the average error is defined as

$$\bar{e}(\mathcal{K}_n) := \max_{r \in \mathcal{R}} \mathbb{P}[M^n \neq \hat{M}_r^n].$$

Definition 10: A nonnegative number R_S is called an *achievable secrecy rate for the CAVWC* $(\overline{\mathfrak{W}}, \mathfrak{V})$ if there exists a sequence $(\mathcal{K}_n)_{n=1}^\infty$ of uncorrelated (n, J_n)-codes such that

$$\begin{aligned}\liminf_{n\to\infty} \frac{1}{n} \log J_n &\geq R_S,\\ \lim_{n\to\infty} \bar{e}(\mathcal{K}_n) &= 0,\\ \lim_{n\to\infty} \max_{s^n \in \mathcal{S}^n} I(M^n \wedge Z_{s^n}^n) &= 0.\end{aligned} \tag{22}$$

The supremum of all achievable secrecy rates is called the *secrecy capacity of* $(\overline{\mathfrak{W}}, \mathfrak{V})$ and denoted by $C_S(\overline{\mathfrak{W}}, \mathfrak{V})$.

We are actually interested in a stronger, permutation invariant form of secrecy. This is because we mainly consider CAVWCs as an auxiliary channel model. We would like to exploit the achievability part of a coding theorem for CAVWCs to find rates that are achievable for the AVWC by correlated random codes. This can be done using Ahlswede's robustification technique, which requires an exponential decrease of the average error and "permutation invariance" of secrecy to be defined below.

For a permutation π contained in the symmetric group Π_n of permutations of $\{1, \ldots, n\}$, denote by E^π the stochastic encoder obtained from a stochastic encoder E via

$$E^\pi(x^n|j) := E(\pi^{-1}(x^n)|j). \tag{23}$$

Here, $\pi(x^n) = (x_{\pi(1)}, \ldots, x_{\pi(n)})$ for any $x^n \in \mathcal{A}^n$. The corresponding decoding sets are $\mathcal{D}_j^\pi := \{\pi(y^n) : y^n \in \mathcal{D}_j\}$. This family of codes together with $(\overline{\mathfrak{W}}, \mathfrak{V})$ induces a canonical *permutation invariant* family of random variables

$$\begin{aligned}&\mathcal{F}(\mathcal{K}_n, \overline{\mathfrak{W}}, \mathfrak{V}, \Pi_n)\\&\quad := \{(M^n, X^n(\pi), Y_r^n(\pi), Z_{s^n}^n(\pi), \hat{M}_r^n(\pi)) :\\&\qquad r \in \mathcal{R}, s^n \in \mathcal{S}^n, \pi \in \Pi_n\},\end{aligned} \tag{24}$$

where M^n and $\hat{M}_r^n(\pi)$ assume values in $\mathcal{J}_n$, the values of $X^n(\pi)$ lie in $\mathcal{A}^n$, those of $Y_r^n(\pi)$ in $\mathcal{B}^n$ and those of $Z_{s^n}^n(\pi)$ in $\mathcal{C}^n$ and where for any $r \in \mathcal{R}$ and $s^n \in \mathcal{S}^n$ and $\pi \in \Pi_n$

$$\begin{aligned}&P_{M^n X^n(\pi) Y_r^n(\pi) Z_{s^n}^n(\pi) \hat{M}_r^n(\pi)}(j, x^n, y^n, z^n, \hat{j})\\&\quad = \frac{1}{J_n} E^\pi(x^n|j) W_r^n(y^n|x^n) V_{s^n}^n(z^n|x^n) \mathbb{1}_{\mathcal{D}_{\hat{j}}^\pi}(y^n).\end{aligned}$$

For every permutation, we have $\mathbb{P}[M^n \neq \hat{M}^n(\pi)] = \mathbb{P}[M^n \neq \hat{M}^n(\mathrm{id})]$, where id denotes the identity permutation. Thus also in the permutation-invariant setting, we can still just write $\bar{e}(\mathcal{K}_n)$ for the average error of $\mathcal{K}_n$.

Definition 11: A nonnegative number R_S is called an *achievable permutation invariant secrecy rate for the CAVWC* $(\overline{\mathfrak{W}}, \mathfrak{V})$ if there exists a sequence $(\mathcal{K}_n)_{n=1}^\infty$ of uncorrelated (n, J_n)-codes and a $\beta > 0$ such that

$$\liminf_{n\to\infty} \frac{1}{n} \log J_n \geq R_S, \tag{25}$$

$$\liminf_{n\to\infty} -\frac{1}{n} \log \bar{e}(\mathcal{K}_n) \geq \beta, \tag{26}$$

$$\lim_{n\to\infty} \max_{s^n \in \mathcal{S}^n} \max_{\pi \in \Pi_n} I(M^n \wedge Z_{s^n}^n(\pi)) = 0. \tag{27}$$

The supremum of all achievable permutation invariant secrecy rates is called the *permutation invariant secrecy capacity of* $(\overline{\mathfrak{W}}, \mathfrak{V})$ and denoted by $C_S^{\pi\text{-inv}}(\overline{\mathfrak{W}}, \mathfrak{V})$.

Theorem 12: The permutation invariant secrecy capacity $C_S^{\pi\text{-inv}}(\overline{\mathfrak{W}}, \mathfrak{V})$ and the secrecy capacity $C_S(\overline{\mathfrak{W}}, \mathfrak{V})$ of the CAVWC $(\overline{\mathfrak{W}}, \mathfrak{V})$ both equal

$$\begin{aligned}&R_S^*(\overline{\mathfrak{W}}, \mathfrak{V})\\&\quad := \lim_{k\to\infty} \frac{1}{k} \sup_{\bar{\mathcal{Q}}_k} \left(\min_{r \in \mathcal{R}} I(\bar{U} \wedge \bar{Y}_r^k) - \max_{s^k \in \mathcal{S}^k} I(\bar{U} \wedge \bar{Z}_{s^k}^k)\right),\end{aligned}$$

where the set $\bar{\mathcal{Q}}_k$ contains those families of random variables

$$\{\bar{U}, \bar{X}^k, \bar{Y}_r^k, \bar{Z}^k_{s^k} : r \in \mathcal{R}, s^k \in \mathcal{S}^k\}$$

which satisfy that $\bar{U}$ assumes values in a finite subset of the integers, the values of $\bar{X}^k$ lie in $\mathcal{A}^k$, those of $\bar{Y}_r^k$ in $\mathcal{B}^k$, those of $\bar{Z}^k_{s^k}$ in $\mathcal{C}^k$, and such that for every $r \in \mathcal{R}$ and $s^k \in \mathcal{S}^k$,

$$\begin{aligned}P_{\bar{U}\bar{X}^k\bar{Y}_r^k\bar{Z}^k_{s^k}}&(u, x^k, y^k, z^k)\\ &= P_{\bar{U}}(u)P_{\bar{X}^k|\bar{U}}(x^k|u)W_r^k(y^k|x^k)V^k_{s^k}(z^k|x^k).\end{aligned}$$

$P_{\bar{U}}$ and $P_{\bar{X}|\bar{U}}$ may be arbitrary probability distributions and stochastic matrices, respectively.

The achievability part of the proof of Theorem 12 can be found in the next section. The converse is similar to, but simpler than that for Theorem 6, so we will not write it down explicitly. The converse for Theorem 6 can be found in Section VIII.

Remarks 7-1) and 7-2) apply for Theorem 12 as well. As in Remark 7-4), one could replace $\max_{s^k\in\mathcal{S}^k} I(\bar{U} \wedge \bar{Z}^k_{s^k})$ by $\max_{\tilde{q}\in\mathcal{P}(\mathcal{S}^k)} I(\bar{U} \wedge \bar{Z}^k_{s^k})$.

VI. Achievability Part of the Proof of Theorem 12

A. Reduction

We first reduce the claim of the achievability of $R_S^*(\overline{\mathfrak{W}}, \mathfrak{V})$ to a simpler achievability claim. This is done in three reduction steps. Each of these steps is relatively simple.

Reduction Step 1: As $C_S^{\pi\text{-inv}}(\overline{\mathfrak{W}}, \mathfrak{V}) \leq C_S(\overline{\mathfrak{W}}, \mathfrak{V})$, it is sufficient to show that $R_S^*(\overline{\mathfrak{W}}, \mathfrak{V})$ is an achievable permutation invariant secrecy rate for $(\overline{\mathfrak{W}}, \mathfrak{V})$.

Reduction Step 2: Call $R_S \geq 0$ an *achievable secrecy rate with exponentially decreasing error for the CAVWC* $(\overline{\mathfrak{W}}, \mathfrak{V})$ if there exists a sequence $(\mathcal{K}_n)_{n=1}^{\infty}$ of uncorrelated (n, J_n)-codes and a $\beta > 0$ such that (25) and (26) hold and (27) is replaced by the simpler condition

$$\lim_{n\to\infty} \max_{s^n\in S^n} I(M^n \wedge Z^n_{s^n}) = 0, \tag{28}$$

where M^n and the $Z^n_{s^n}$ are the corresponding elements of $\mathcal{F}(\mathcal{K}_n, \overline{\mathfrak{W}}, \mathfrak{V})$. To prove that $R_S^*(\overline{\mathfrak{W}}, \mathfrak{V})$ is an achievable permutation invariant secrecy rate for $(\overline{\mathfrak{W}}, \mathfrak{V})$, it is sufficient to prove that $R_S^*(\overline{\mathfrak{W}}, \mathfrak{V})$ is an achievable secrecy rate with exponentially decreasing error for $(\overline{\mathfrak{W}}, \mathfrak{V})$. This is due to the following lemma.

Lemma 13: Let $\mathcal{K}_n$ be an uncorrelated (n, J_n)-code with stochastic encoder E. Let M^n be the canonical message random variable and $\{Z^n_{s^n}(\pi) : s^n \in \mathcal{S}^n, \pi \in \Pi_n\}$ the family of canonical eavesdropper output random variables from $\mathcal{F}(\mathcal{K}_n, \overline{\mathfrak{W}}, \mathfrak{V}, \Pi_n)$. Let id be the identity permutation mapping each element of $\{1, \dots, n\}$ to itself. If there exists an $\varepsilon > 0$ such that

$$\max_{s^n} I(M^n \wedge Z^n_{s^n}(\text{id})) \leq \varepsilon, \tag{29}$$

then

$$\max_{\pi\in\Pi_n} \max_{s^n} I(M^n \wedge Z^n_{s^n}(\pi)) \leq \varepsilon. \tag{30}$$

Lemma 13 is proved in Appendix C and is based on the fact that $P_{M^n\pi(Z^n_{s^n}(\text{id}))} = P_{M^n Z^n_{\pi(s^n)}(\pi)}$.

Reduction Step 3: $R_S^*(\overline{\mathfrak{W}}, \mathfrak{V})$ is an achievable secrecy rate with exponentially decreasing error if, for every CAVWC $(\overline{\mathfrak{W}}, \mathfrak{V})$, the rate

$$R_S^{\dagger}(\overline{\mathfrak{W}}, \mathfrak{V}) := \max_{\bar{\mathcal{Q}}_1^{\dagger}} \left(\min_{r\in\mathcal{R}} I(\bar{X} \wedge \bar{Y}_r) - \max_{q\in\mathcal{P}(\mathcal{S})} I(\bar{X} \wedge \bar{Z}_q) \right) \tag{31}$$

is an achievable secrecy rate with exponentially decreasing error for $(\overline{\mathfrak{W}}, \mathfrak{V})$, where the set $\bar{\mathcal{Q}}_1^{\dagger}$ contains those families of random variables $\{\bar{X}, \bar{Y}_r, \bar{Z}_q : r \in \mathcal{R}, q \in \mathcal{P}(\mathcal{S})\}$ such that $\bar{X}$ is an arbitrary random variable assuming values in $\mathcal{A}$, the values of $\bar{Y}_r$ lie in $\mathcal{B}$, those of $\bar{Z}_q$ in $\mathcal{C}$, and

$$P_{\bar{X}\bar{Y}_r\bar{Z}_q}(x, y, z) = P_{\bar{X}}(x)W_r(y|x)\left(\sum_{s\in\mathcal{S}} q(s)V_s(z|x)\right).$$

This is proved using a standard channel prefixing argument, see Appendix D.

Having done these three reduction steps, it now remains to prove that $R_S^{\dagger}(\overline{\mathfrak{W}}, \mathfrak{V})$ is an achievable secrecy rate with exponentially decreasing error.

B. $R_S^{\dagger}(\overline{\mathfrak{W}}, \mathfrak{V})$ Is an Achievable Secrecy Rate With Exponentially Decreasing Error

Let $(\overline{\mathfrak{W}}, \mathfrak{V})$ be an AVWC. The proof that $R_S^{\dagger}(\overline{\mathfrak{W}}, \mathfrak{V})$ is an achievable secrecy rate with exponentially decreasing error for $(\overline{\mathfrak{W}}, \mathfrak{V})$ follows a random coding strategy. The random codewords are chosen as follows. Fix a blocklength n and a family $\{\bar{X}, \bar{Y}_r, \bar{Z}_q : r \in \mathcal{R}, q \in \mathcal{P}(\mathcal{S})\} \in \bar{\mathcal{Q}}_1^{\dagger}$ as in the definition of $R_S^{\dagger}(\overline{\mathfrak{W}}, \mathfrak{V})$. For arbitrary $\tau > 0$, set[2]

$$J_n := \left\lfloor \exp\left\{ n\left(\min_{r\in\mathcal{R}} I(\bar{X} \wedge \bar{Y}_r) - \max_{q\in\mathcal{P}(S)} I(\bar{X} \wedge \bar{Z}_q) - \tau \right)\right\} \right\rfloor, \tag{32}$$

and

$$L_n := \left\lfloor \exp\left\{ n\left(\max_{q\in\mathcal{P}(S)} I(\bar{X} \wedge \bar{Z}_q) + \frac{\tau}{4} \right)\right\} \right\rfloor \tag{33}$$

and define $\mathcal{J}_n = \{1, \dots, J_n\}$ and $\mathcal{L}_n := \{1, \dots, L_n\}$. Further, for some $\delta > 0$ to be chosen later, we define a family $\mathcal{X} := \{X_{jl} : j \in \mathcal{J}_n, l \in \mathcal{L}_n\}$ of random codewords in $\mathcal{X}^n$ with distribution

$$\mathbb{P}[X_{jl} = x^n] := P'(x^n) := \frac{P^n_{\bar{X}}(x^n)}{P^n_{\bar{X}}(\mathcal{T}^n_{\bar{X},\delta})} \mathbb{1}_{\mathcal{T}^n_{\bar{X},\delta}}(x^n).$$

Via $\mathcal{X}$, we obtain a randomly selected stochastic encoder

$$E^{\mathcal{X}}(x^n|j) := \frac{1}{L_n} \sum_{l=1}^{L_n} \mathbb{1}_{\{X_{jl}\}}(x^n). \tag{34}$$

[2] Recall that we use the convention $\exp(x) = 2^x$.

1) Reliability: With high probability, a realization of $E^{\mathcal{X}}$ determines an uncorrelated (n, J_n)-code denoted by $\mathcal{K}_n^{\mathcal{X}}$ for the compound channel $\overline{\mathfrak{W}}$ with exponentially small average error.

Lemma 14: For sufficiently small $\delta > 0$ there exists a $\tau_6 > 0$ such that, if n is sufficiently large, there exist decoding sets $\{\mathcal{D}_j^{\mathcal{X}} : j \in \mathcal{J}_n\}$ depending on $\mathcal{X}$ with the property that the (random) average error $\bar{e}(\mathcal{K}_n^{\mathcal{X}})$ of the random uncorrelated (n, J_n)-code $\mathcal{K}_n^{\mathcal{X}}$ with the stochastic encoder $E^{\mathcal{X}}$ and the decoding sets $\{\mathcal{D}_j^{\mathcal{X}} : j \in \mathcal{J}_n\}$ satisfies

$$\mathbb{P}\left\{\bar{e}(\mathcal{K}_n^{\mathcal{X}}) \leq 2^{-n\tau_6}\right\} \geq 1 - 2^{-n\tau_6}.$$

As the probability distribution of $\mathcal{X}$ is not completely standard, we include a proof of this lemma in Appendix F, although it does not differ much from the proof in [6]. The proof shows that the receiver can even decode the randomization index l in addition to the messages.

2) Secrecy: The uncorrelated (n, J_n)-code $\mathcal{K}_n^{\mathcal{X}}$ from Lemma 14 also satisfies the secrecy condition (28) with high probability. For the statement of the next lemma, recall that every realization of $\mathcal{X}$ together with the decoding sets $\{\mathcal{D}_j^{\mathcal{X}} : j \in \mathcal{J}_n\}$ from Lemma 14 gives rise to a canonical family of random variables $\mathcal{F}(\mathcal{K}_n^{\mathcal{X}}, \overline{\mathfrak{W}}, \mathfrak{V}) = \{M^n, X^n, Y_r^n, Z_{s^n}^n, \hat{M}_r^n : r \in \mathcal{R}, s^n \in \mathcal{S}^n\}$ as in (4). The dependence of these random variables on $\mathcal{X}$ is suppressed in the notation.

Lemma 15: For $\delta > 0$ sufficiently small, there exist $\tau_1, \tau_2 > 0$ such that if n is large enough, there exists a family $\{\Theta_{s^n} : s^n \in \mathcal{S}^n\}$ of finite measures on $\mathcal{C}^n$ such that the probability of the event

$$\iota_0 := \left\{\max_{j\in\mathcal{J}_n}\max_{s^n\in\mathcal{S}^n} \| P_{Z_{s^n}^n|M^n}(\,\cdot\,|j) - \Theta_{s^n}(\cdot)\| \leq 2^{-\tau_1 n}\right\}$$

is at least $1 - 2^{-\tau_2 n}$. (Note that $P_{Z_{s^n}^n|M^n}(\,\cdot\,|j)$ is a random variable depending on $\mathcal{X}$.)

This lemma is proved in Appendix G.

Corollary 2: For $\delta > 0$ small enough and for the τ_1, τ_2 from Lemma 15, if n is large enough, the probability of the event

$$\iota_0' := \left\{\max_{s^n\in\mathcal{S}^n} I(M^n \wedge Z_{s^n}^n) \leq 2^{-\frac{\tau_1}{2}n}\right\}$$

is at least $1 - 2^{-\tau_2 n}$. (Note again that the joint distribution of $Z_{s^n}^n$ and M^n is a random variable depending on $\mathcal{X}$.)

Corollary 2 immediately follows from Lemma 15 and the uniform continuity of mutual information in total variation distance [11, Lemma 2.7].

3) Synthesis of Reliability and Secrecy: Lemma 14 and Corollary 2 show that the probability that $\mathcal{K}_n^{\mathcal{X}}$ satisfies (25), (26) and (28) is positive if δ is sufficiently small and n sufficiently large, so a realization satisfying (25), (26) and (28) for $\beta = \tau_6 > 0$ and $R_S = R_S^{\dagger}(\overline{\mathfrak{W}}, \mathfrak{V}) - \tau$ must exist. Since $\tau > 0$ was arbitrary, this proves that $R_S^{\dagger}(\overline{\mathfrak{W}}, \mathfrak{V})$ is an achievable secrecy rate with exponentially decreasing error. As shown in Subsection VI-A, this implies that $R_S^{*}(\overline{\mathfrak{W}}, \mathfrak{V})$ is an achievable (permutation invariant) secrecy rate for $(\overline{\mathfrak{W}}, \mathfrak{V})$.

VII. Proof of the Achievability Part of Theorem 6

The proof that $R_S^*(\mathfrak{W}, \mathfrak{V})$ is a lower bound to $C_{S,\text{ran}}^{\max}(\mathfrak{W}, \mathfrak{V})$ and $C_{S,\text{ran}}^{\text{mean}}(\mathfrak{W}, \mathfrak{V})$ is based on the achievability part of Theorem 12 proved in the previous section. We start by defining a special family $\overline{\mathfrak{W}}$. For the family $\{W_s : s \in \mathcal{S}\}$ determining $\mathfrak{W}$ and every $q \in \mathcal{P}(\mathcal{S})$, set $W_q := \sum_{s\in\mathcal{S}} W_s q(s)$. We then define $\overline{\mathfrak{W}} := \{W_q^n : q \in \mathcal{P}(\mathcal{S}), n = 1, 2, \ldots\}$. This family together with $\mathfrak{V}$ defines the CAVWC $(\overline{\mathfrak{W}}, \mathfrak{V})$. Observe that for $R_S^*(\mathfrak{W}, \mathfrak{V})$ defined in (12), we have

$$R_S^*(\mathfrak{W}, \mathfrak{V}) = R_S^*(\overline{\mathfrak{W}}, \mathfrak{V}).$$

By Theorem 12 applied to the CAVWC $(\overline{\mathfrak{W}}, \mathfrak{V})$ defined above, there exists a $\beta > 0$ such that for every $0 < \varepsilon < \beta$ and sufficiently large n, there exists an uncorrelated (n, J_n)-code $\mathcal{K}_n$ satisfying

$$\begin{aligned}\frac{1}{n}\log J_n &\geq R_S^*(\overline{\mathfrak{W}}, \mathfrak{V}) - \varepsilon = R_S^*(\mathfrak{W}, \mathfrak{V}) - \varepsilon,\\ \bar{e}(\mathcal{K}_n) &= \max_{q\in\mathcal{P}(\mathcal{S})} \frac{1}{J_n}\sum_{j=1}^{J_n}\sum_{x^n\in\mathcal{A}^n} E(x^n|j)W_q^n(\mathcal{D}_j^c|x^n)\\ &\leq 2^{-n(\beta-\varepsilon)}, \end{aligned} \tag{35}$$

and

$$\max_{s^n\in S^n}\max_{\pi\in\mathcal{S}_n} I(M^n \wedge Z_{s^n}^n(\pi)) \leq \varepsilon. \tag{36}$$

The idea is to transform this uncorrelated (n, J_n)-code $\mathcal{K}_n$ into a correlated random (n, J_n)-code which has good reliability and secrecy properties for the AVWC $(\mathfrak{W}, \mathfrak{V})$. Central to this transformation is Ahlswede's robustification technique:

Lemma 16 [2]: If a function $f : \mathcal{S}^n \to [0, 1]$ satisfies

$$\sum_{s^n\in S^n} f(s^n)q(s_1)\cdots q(s_n) \geq 1 - \varepsilon' \tag{37}$$

for all $q \in \mathcal{P}_0^n(\mathcal{S})$ and some $\varepsilon' \in [0, 1]$, then

$$\frac{1}{n!}\sum_{\pi\in\Pi_n} f(\pi(s^n)) \geq 1 - 3(n+1)^{|\mathcal{S}|}\varepsilon'. \tag{38}$$

Define the function f by

$$f(s^n) := \frac{1}{J_n}\sum_{j\in\mathcal{J}_n}\sum_{x^n\in\mathcal{A}^n} E(x^n|j)W_{s^n}^n(\mathcal{D}_j|x^n).$$

It was already noted in Remark 7-3) that for any $q \in \mathcal{P}_0^n(\mathcal{S})$ and $x^n \in \mathcal{A}^n$ and $y^n \in \mathcal{B}^n$

$$\sum_{s^n} W_{s^n}^n(y^n|x^n)q(s_1)\cdots q(s_n) = W_q^n(y^n|x^n).$$

Thus by (35)

$$\begin{aligned}&\sum_{s^n\in\mathcal{S}^n} f(s^n)q(s_1)\cdots q(s_n)\\ &= \frac{1}{J_n}\sum_{j\in\mathcal{J}_n}\sum_{s^n\in\mathcal{S}^n}\sum_{x^n\in\mathcal{A}^n} E(x^n|j)W_{s^n}^n(\mathcal{D}_j|x^n)q(s_1)\cdots q(s_n)\\ &= \frac{1}{J_n}\sum_{j\in\mathcal{J}_n}\sum_{x^n\in\mathcal{A}^n} E(x^n|j)W_q^n(\mathcal{D}_j|x^n)\\ &\geq 1 - 2^{-n(\beta-\varepsilon)}.\end{aligned}$$

Now we derive a correlated random (n, J_n)-code $\mathcal{K}_n^{\mathrm{ran}}$ for message transmission over the AVWC $(\mathfrak{W}, \mathfrak{V})$ from $\mathcal{K}_n$. Let E^π be given by $E^\pi(x^n|j) := E(\pi^{-1}(x^n)|j)$ and let $\mathcal{D}_j^\pi := \{\pi(y^n) : y^n \in \mathcal{D}_j\}$. Further let G_n be uniformly distributed on this family indexed by Π_n. One has

$$\begin{aligned}
&1 - e(\mathcal{K}_n^{\mathrm{ran}})\\
&\quad = \frac{1}{n!}\sum_{\pi\in\Pi_n}\frac{1}{J_n}\sum_{j\in\mathcal{J}_n}\sum_{x^n} E^{\pi^{-1}}(x^n|j)W_{s^n}^n(\mathcal{D}_j^{\pi^{-1}}|x^n)\\
&\quad = \frac{1}{n!}\sum_{\pi\in\Pi_n}\frac{1}{J_n}\sum_{j\in\mathcal{J}_n}\sum_{x^n} E(\pi(x^n)|j)W_{s^n}^n(\mathcal{D}_j^{\pi^{-1}}|x^n)\\
&\quad = \frac{1}{n!}\sum_{\pi\in\Pi_n}\frac{1}{J_n}\sum_{j\in\mathcal{J}_n}\sum_{x^n} E(x^n|j)W_{s^n}^n(\mathcal{D}_j^{\pi^{-1}}|\pi^{-1}(x^n))\\
&\quad = \frac{1}{n!}\sum_{\pi\in\Pi_n}\frac{1}{J_n}\sum_{j\in\mathcal{J}_n}\sum_{x^n} E(x^n|j)W_{\pi(s^n)}^n(\mathcal{D}_j|x^n).
\end{aligned}$$

With $\varepsilon' = 2^{-n(\beta-\varepsilon)}$, Lemma 16 implies that the last term is lower-bounded by $1-(n+1)^{|S|}2^{-n(\beta-\varepsilon)} \geq 1 - 2^{-n(\beta-2\varepsilon)}$ for sufficiently large n. Thus $e(\mathcal{K}_n^{\mathrm{ran}})$ decreases exponentially in n, which settles the reliability properties of $\mathcal{K}_n^{\mathrm{ran}}$ for $(\mathfrak{W}, \mathfrak{V})$.

The secrecy properties (10) and (11) of $\mathcal{K}_n^{\mathrm{ran}}$ are immediate, as (36) implies

$$\frac{1}{n!}\sum_{\pi\in\Pi_n} I(M^n \wedge Z_{s^n}^n(\pi)) \leq \max_{\pi\in\Pi_n} I(M^n \wedge Z_{s^n}^n(\pi)) \leq \varepsilon$$

for every $s^n \in \mathcal{S}^n$. Hence $R_S^*(\overline{\mathfrak{W}}, \mathfrak{V}) = R_S^*(\mathfrak{W}, \mathfrak{V})$ is an achievable correlated random coding mean and maximum secrecy rate for the AVWC $(\mathfrak{W}, \mathfrak{V})$.

VIII. The Converses

We first prove the converse of Theorem 6. The converse of Theorem 12 is analogous with the simplifying exception that one does not have to deal with common randomness, so we will not write it down explicitly.

One unusual difficulty arises in the proof of the converse of Theorem 6. This difficulty consists in the fact that the common randomness prohibits a "naive" application of the data processing inequality. It is thus necessary to limit the amount of common randomness of an arbitrary correlated random code in order to overcome this difficulty. Recall that Lemma 8 is independent of Theorem 6, so it can be applied here to exactly this purpose of reducing correlated randomness.

Let $R_S < C_{S,\mathrm{ran}}^{\mathrm{mean}}(\mathfrak{W}, \mathfrak{V})$. From Lemma 8 we know that for every $\varepsilon > 0$ there is an $n(R_S, \varepsilon)$ such that for $n \geq n(R_S, \varepsilon)$ there is a correlated random (n, J_n)-code $\mathcal{K}_n^{\mathrm{ran}}$ satisfying

$$\frac{1}{n}\log J_n \geq R_S - \varepsilon, \tag{39}$$

$$e(\mathcal{K}_n^{\mathrm{ran}}) \leq \varepsilon, \tag{40}$$

$$\max_{s^n\in\mathcal{S}^n} I(M^n \wedge Z_{s^n}^n|G_n) \leq \varepsilon, \tag{41}$$

and

$$|\mathrm{supp}(G_n)| \leq \frac{2n\log|\mathcal{A}|}{\varepsilon}(1 + n\log|\mathcal{S}|) + 1. \tag{42}$$

Since the average error of $\mathcal{K}_n^{\mathrm{ran}}$ is affine in the channel, it does not change if one passes to the generalized channel state space $\mathcal{P}(\mathcal{S}^n)$. More precisely, for $\tilde{q} \in \mathcal{P}(\mathcal{S}^n)$ define

$$W_{\tilde{q}}^n(y^n|x^n) := \sum_{s^n\in\mathcal{S}^n}\tilde{q}(s^n)W_{s^n}^n(y^n|x^n).$$

Then

$$\begin{aligned}
&\max_{\tilde{q}\in\mathcal{P}(\mathcal{S}^n)}\frac{1}{J_n}\sum_{j=1}^{J_n}\sum_{\gamma\in\Gamma_n}\sum_{x^n\in\mathcal{A}^n}\Big(E^\gamma(x^n|j)\\
&\qquad\times W_{\tilde{q}}^n\big((\mathcal{D}_j^\gamma)^c|x^n\big)P_{G_n}(\gamma)\Big)\\
&\quad = \max_{s^n\in\mathcal{S}^n}\frac{1}{J_n}\sum_{j=1}^{J_n}\sum_{\gamma\in\Gamma_n}\sum_{x^n\in\mathcal{A}^n}\Big(E^\gamma(x^n|j)\\
&\qquad\times W_{s^n}^n\big((\mathcal{D}_j^\gamma)^c|x^n\big)P_{G_n}(\gamma)\Big)\\
&\quad \overset{(i)}{\leq} \varepsilon,
\end{aligned}\tag{43}$$

where (i) holds because the term on the left-hand side of the inequality sign equals $e(\mathcal{K}_n^{\mathrm{ran}})$ and due to (40). From (43), one infers that the average error of $\mathcal{K}_n^{\mathrm{ran}}$ for transmission over the compound channel $\overline{\mathfrak{W}}$ defined at the beginning of Section VII (cf. Remark 7) is upper-bounded by ε as well, i.e.

$$\begin{aligned}
&\max_{q\in\mathcal{P}(\mathcal{S})}\frac{1}{J_n}\sum_{j=1}^{J_n}\sum_{\gamma\in\Gamma_n}\sum_{x^n\in\mathcal{A}^n}E^\gamma(x^n|j)W_q^n\big((\mathcal{D}_j^\gamma)^c|x^n\big)P_{G_n}(\gamma)\\
&\quad\leq \varepsilon.
\end{aligned}\tag{44}$$

Due to Fano's inequality [11, Lemma 3.8], (44) implies for every $q \in \mathcal{P}(\mathcal{S})$

$$\begin{aligned}
&H(M^n|\hat{M}_q^n, G_n)\\
&\quad = \sum_{\gamma\in\mathrm{supp}(G_n)} H(M^n|\hat{M}_q^n, G_n = \gamma)P_{G_n}(\gamma)\\
&\quad \leq 1 + \sum_{\gamma\in\mathrm{supp}(G_n)}\mathbb{P}[M^n \neq \hat{M}_q^n|G_n = \gamma]P_{G_n}(\gamma)\log J_n\\
&\quad \leq 1 + \varepsilon\log J_n.
\end{aligned}$$

Here the $\hat{M}_q^n$ are the random variables from the canonical family $\mathcal{F}(\mathcal{K}_n^{\mathrm{ran}}, \overline{\mathfrak{W}}, \mathfrak{V})$ defined in (21). Hence the independence of M^n and G_n yields

$$\begin{aligned}
\log J_n &= H(M^n)\\
&= H(M^n|G_n)\\
&= I(M^n \wedge \hat{M}_q^n|G_n) + H(M^n|\hat{M}_q^n, G_n)\\
&\leq I(M^n \wedge \hat{M}_q^n|G_n) + 1 + \varepsilon\log J_n,
\end{aligned}$$

so by rearranging and taking (41) into account, we have for every $q \in \mathcal{P}(\mathcal{S})$ and $s^n \in \mathcal{S}^n$

$$\begin{aligned}
&(1-\varepsilon)\log J_n\\
&\quad\leq I(M^n \wedge \hat{M}_q^n|G_n) - I(M^n \wedge Z_{s^n}^n|G_n) + 1 + \varepsilon.
\end{aligned}\tag{45}$$

We have to get rid of G_n in some way. The only reasonable way to achieve this seems to be through the use of the convexity of the mutual information in the channel argument. But while this is a valid choice for the "secrecy term", it is certainly invalid for the "legal" term. This is due to the fact that G_n is independent of M^n, but not of $\hat{M}_q^n$ or Y_q^n. An application of the data processing inequality is thus only possible conditioned on G_n. Using the properties of entropy and conditional entropy and writing $K_n(R_S,\varepsilon) := |\mathrm{supp}(G_n)|$, we obtain

$$\begin{aligned} I(M^n \wedge \hat{M}_q^n | G_n) &\\ &= H(M^n) - H(M^n|Y_q^n, G_n)\\ &= H(M^n) - H(M^n, G_n|Y_q^n) + H(G_n|Y_q^n)\\ &\leq H(M^n) - H(M^n|Y_q^n) + H(G_n)\\ &\leq I(M^n \wedge Y_q^n) + \log K_n(R_S,\varepsilon). \end{aligned} \tag{46}$$

Thus if n is sufficiently large, we obtain

$$\begin{aligned} \frac{1}{n}\log J_n &\\ &\overset{(i)}{\leq} \frac{1}{n(1-\varepsilon)}\Big(\min_{q\in\mathcal{P}(\mathcal{S})} I(M^n \wedge \hat{M}_q^n|G_n)\\ &\qquad - \max_{s^n\in\mathcal{S}^n} I(M^n \wedge Z_{s^n}^n|G_n) + 1 + \varepsilon\Big)\\ &\overset{(ii)}{\leq} \frac{1}{n(1-\varepsilon)}\Big(\min_{q\in\mathcal{P}(\mathcal{S})} I(M^n \wedge Y_q^n) - \max_{s^n\in\mathcal{S}^n} I(M^n \wedge Z_{s^n}^n)\Big)\\ &\quad + \frac{\log K_n(R_S,\varepsilon)+1+\varepsilon}{n(1-\varepsilon)} \qquad (47)\\ &\overset{(iii)}{\leq} \frac{1}{n(1-\varepsilon)}\Big(\min_{q\in\mathcal{P}(\mathcal{S})} I(M^n \wedge Y_q^n) - \max_{s^n\in\mathcal{S}^n} I(M^n \wedge Z_{s^n}^n)\Big)\\ &\quad + \varepsilon. \qquad (48) \end{aligned}$$

Here, (i) is (45) and (ii) follows from (46). In (iii), the importance of Lemma 8 becomes evident: $K_n(R_S,\varepsilon)$ grows sub-exponentially in n, so for n sufficiently large, the second term of (47) is upper-bounded by ε.

If we set $\bar{U} := M^n$ and $\bar{X}^n := X^n$ and $\bar{Y}_q^n := Y_q^n$ and $\bar{Z}_{s^n}^n := Z_{s^n}^n$, we obtain the joint distributions

$$P_{\bar{U}\bar{X}^n\bar{Y}_q^n}(j,x^n,y^n) = \frac{1}{J_n}\sum_{\gamma\in\Gamma_n} P_{G_n}(\gamma)E^\gamma(x^n|j)W_q^n(y^n|x^n),$$

$$P_{\bar{U}\bar{X}^n\bar{Z}_{s^n}^n}(j,x^n,z^n) = \frac{1}{J_n}\sum_{\gamma\in\Gamma_n} P_{G_n}(\gamma)E^\gamma(x^n|j)V_{s^n}^n(z^n|x^n).$$

Making the additional assumption that $\bar{Y}_q^n$ and $\bar{Z}_{s^n}^n$ are conditionally independent given $\bar{X}^n$ for all pairs (q,s^n), which is possible without loss of generality, the family $\{\bar{U},\bar{X}^n,\bar{Y}_q^n,\bar{Z}_{s^n}^n : q\in\mathcal{P}(\mathcal{S}), s^n\in\mathcal{S}^n\}$ is contained in $\mathcal{Q}_n$ and thus has the form required in the definition of $R_S^*(\mathfrak{W},\mathfrak{V})$. Hence by (48), we obtain the inequality

$$\frac{1}{n}\log J_n \leq \frac{1}{1-\varepsilon}R_S^*(\mathfrak{W},\mathfrak{V}) + \varepsilon.$$

By (39) and as ε was arbitrary, we have $R_S \leq R_S^*(\mathfrak{W},\mathfrak{V})$, hence $C_{S,\mathrm{ran}}^{\mathrm{mean}}(\mathfrak{W},\mathfrak{V}) \leq R_S^*(\mathfrak{W},\mathfrak{V})$, and therefore also $C_{S,\mathrm{ran}}^{\max}(\mathfrak{W},\mathfrak{V}) \leq R_S^*(\mathfrak{W},\mathfrak{V})$. This completes the proof of the converse of Theorem 6.

Remark 17: If one applies Fano's inequality immediately after (43) and skips passing to (44), all steps of the converse go through in the same way, with the exception that the random variables Y_q^n have to be replaced by more general random variables $Y_{\tilde{q}}^n = \bar{Y}_{\tilde{q}}^n$ with conditional distribution $P_{\bar{Y}_{\tilde{q}}^n|\bar{X}^n}$ as defined in Remark 7-4). Therefore

$$R_S^*(\mathfrak{W},\mathfrak{V}) = C_{S,\mathrm{ran}}^{\mathrm{mean}}(\mathfrak{W},\mathfrak{V}) \leq (14).$$

Together with the simple fact $R_S^*(\mathfrak{W},\mathfrak{V}) \geq (14)$ already noticed in Remark 7-4), this proves the equality of $R_S^*(\mathfrak{W},\mathfrak{V})$ and 7-4) claimed in Remark 17.

The converse for Theorem 12 follows the same lines as that for Theorem 6. It is simpler as no common randomness has to be considered.

IX. Discussion

The main result of this paper is the correlated random coding secrecy capacity of the AVWC for the case where the eavesdropper is allowed access to the correlated randomness shared by sender and intended receiver. Applying Ahlswede's robustification technique, the main problem was solved via reduction to the secrecy capacity problem of the CAVWC, which is compound between the sender and the intended receiver and arbitrarily varying between the sender and the eavesdropper.

The secrecy capacity formula obtained in the main theorem is a multi-letter formula. Of course, this makes a direct computation impossible. On the other hand, it is not known whether a general, computable, single-letter formula exists at all. For a given AVWC, the value of the multi-letter formula can be approximated by restricting computation to a finite number of letters. An open problem not addressed in this paper is the goodness of finite-letter approximation.

However, the use of a capacity formula is much larger than just to calculate the capacity. It can be applied in the in-depth analysis of the channels in question. For example, using nothing but the capacity formula, it can be shown for discrete memoryless channels that the capacity of parallel channels is the sum of their capacities. For the AVWC, an analysis of the capacity formula shows that the correlated random coding secrecy capacity is continuous in the AVWC, which is impossible to derive a priori. This result is of great engineering importance because it ensures that small variations in the channel data cannot lead to completely different secrecy capacities. This is very reassuring, as lots of resources would otherwise have to be spent on channel estimation. In fact, the necessary precision of the channel estimate would grow without limits the closer the channel would be to a point of discontinuity of the secrecy capacity function.

Follow-up work on the AVWC correlated random coding secrecy capacity for the case that the eavesdropper has no knowledge of the correlated randomness as well as the AVWC uncorrelated coding secrecy capacity is presented in [22].

Appendix A
Proof of Corollary 1

It is obvious that the right-hand side of (16) is upper-bounded by $R_S^*(\mathfrak{W},\mathfrak{V})$, see Remark 7-1). Thus it remains to show the converse relation. Let k be a positive integer and let $\{\bar{U},\bar{X},\bar{Y}_{q_1}^k,\bar{Z}_{s^k}^k\} \in \mathcal{Q}_k$ be a family of random variables as in the definition of $R_S^*(\mathfrak{W},\mathfrak{V})$. The existence of a best channel to the eavesdropper guarantees that $I(\bar{U}\wedge\bar{Z}^k_{s_2^k})\le I(\bar{U}\wedge\bar{Z}^k_{s_*})$ for every $s_2^k\in\mathcal{S}_2^k$, where $P_{\bar{Z}^k_{s_*}|\bar{X}}(z^k|x^k)=\prod_{i=1}^k V_{s_*}(z_i|x_i)$. In particular, $I(\bar{U}\wedge\bar{Z}^k_{s_*})=\max_{s_2\in\mathcal{S}_2}I(\bar{U}\wedge\bar{Z}^k_{s_2^k})$. Therefore

$$\begin{aligned}
&\frac{1}{k}\Big(\min_{q_1\in\mathcal{P}(\mathcal{S}_1)}I(\bar{U}\wedge\bar{Y}^k_{q_1})-\max_{s_2^k\in\mathcal{S}_2^k}I(\bar{U}\wedge\bar{Z}^k_{s^k})\Big) &(49)\\
&=\frac{1}{k}\min_{q_1\in\mathcal{P}(\mathcal{S}_1)}\Big(I(\bar{U}\wedge\bar{Y}^k_{q_1})-I(\bar{U}\wedge\bar{Z}^k_{s_*})\Big)\\
&\overset{(i)}{\le}\frac{1}{k}\min_{q_1\in\mathcal{P}(\mathcal{S}_1)}I(\bar{U}\wedge\bar{Y}^k_{q_1}|\bar{Z}^k_{s_*}), &(50)
\end{aligned}$$

where strong degradedness was applied in (i). In a similar fashion as in the derivation of (23)-(26) in [21], one can rewrite (50) as $I(\bar{X}^*\wedge\bar{Y}^*_{q_1}|\bar{Z}^*_{s_*})$, where $\bar{X}^*$ is a random variable on $\mathcal{A}$ and the random variables $\bar{Y}^*_{q_1}$ and $\bar{Z}^*_{s_*}$ satisfy $P_{\bar{Y}^*_{q_1}|\bar{X}^*}=W_{q_1}$ and $P_{\bar{Z}^*_{s_*}|\bar{X}^*}=V_{s_*}$. Again using the strong degradedness of $(\mathfrak{W},\mathfrak{V})$ and the existence of a best channel to the eavesdropper and defining $\bar{Z}^*_{s_2}$ by its conditional distribution $P_{\bar{Z}^*_{s_2}|\bar{X}^*}=V_{s_2}$ for every $s_2\in\mathcal{S}_2$, one obtains

$$\begin{aligned}
&\min_{q_1\in\mathcal{P}(\mathcal{S}_1)}I(\bar{X}^*\wedge\bar{Y}^*_{q_1}|\bar{Z}^*_{s_*})\\
&\le\min_{q_1\in\mathcal{P}(\mathcal{S}_1)}\Big(I(\bar{X}^*\wedge\bar{Y}^*_{q_1})-I(\bar{X}^*\wedge\bar{Z}^*_{s_*})\Big)\\
&=\min_{q_1\in\mathcal{P}(\mathcal{S}_1)}I(\bar{X}^*\wedge\bar{Y}^*_{q_1})-\max_{s_2\in\mathcal{S}_2}I(\bar{X}^*\wedge\bar{Z}^*_{s_2}). &(51)
\end{aligned}$$

The family $\{\bar{X}^*,\bar{Y}^*_{q_1},\bar{Z}^*_{s_2}:q_1\in\mathcal{P}(\mathcal{S}_1),s_2\in\mathcal{S}_2\}$ is contained in $\mathcal{Q}_1^*$, so (51) is upper-bounded by the right-hand side of (16). Therefore (49) is also upper-bounded by the right-hand side of (16). This holds for every $\{\bar{U},\bar{X},\bar{Y}^k_{q_1},\bar{Z}^k_{s^k}\}\in\mathcal{Q}_k$, thus proving that (16) indeed is an equality. This proves Corollary 1.

Appendix B
Proof of Lemma 8

Let $R_S<C^{\mathrm{mean}}_{S,\mathrm{ran}}(\mathfrak{W},\mathfrak{V})$ and let $\varepsilon>0$. By Definition 3, there exists a sequence of correlated random (n,J_n)-codes $\tilde{\mathcal{K}}^{\mathrm{ran}}_n$ satisfying (8)-(10). In particular, for $\tilde{\varepsilon}:=\varepsilon/4$, there exists an $n(R_S,\tilde{\varepsilon})$ such that for $n\ge n(R_S,\tilde{\varepsilon})$

$$\begin{aligned}
\frac{1}{n}\log J_n&\ge R_S-\tilde{\varepsilon},\\
e(\tilde{\mathcal{K}}^{\mathrm{ran}}_n)&\le\tilde{\varepsilon}, &(52)\\
\max_{s^n\in\mathcal{S}^n}I(M^n\wedge Z^n_{s^n}(\tilde{G}_n)|\tilde{G}_n)&\le\tilde{\varepsilon}. &(53)
\end{aligned}$$

Here $\tilde{G}_n$ denotes the random variable on Γ_n according to which the realizations of $\tilde{\mathcal{K}}^{\mathrm{ran}}_n$ are chosen. For $\gamma\in\Gamma_n$, denote the average error incurred by the uncorrelated code $\mathcal{K}_n(\gamma)$ under channel state sequence $s^n\in\mathcal{S}^n$ by $e_{s^n}(\mathcal{K}^{\mathrm{ran}}_n)$. Let now

$$K=\left\lfloor\frac{4n\log|\mathcal{A}|}{\varepsilon}(1+n\log|\mathcal{S}|)+1\right\rfloor$$

(as n remains fixed throughout the proof, we suppress the dependence of K on n). Further let $\tilde{G}_n(1),\dots,\tilde{G}_n(K)$ be i.i.d. copies of $\tilde{G}_n$. For any $s^n\in\mathcal{S}^n$, one obtains (54)-(56) at the bottom of the page, where the union bound is applied in the inequality. Using the Markov inequality, (55) can be upper-bounded by

$$\begin{aligned}
&\exp\left(-\frac{K\varepsilon}{n\log|\mathcal{A}|}\right)\mathbb{E}\left[\exp\left(\sum_{k=1}^K\frac{e_{s^n}(\mathcal{K}_n(\tilde{G}_n(k))}{n\log|\mathcal{A}|}\right)\right]\\
&\overset{(i)}{\le}\exp\left(-\frac{K\varepsilon}{n\log|\mathcal{A}|}\right)\left(1+\mathbb{E}\left[\frac{e_{s^n}(\mathcal{K}_n(\tilde{G}_n))}{n\log|\mathcal{A}|}\right]\right)^K\\
&\overset{(ii)}{\le}\exp\left(-\frac{K\varepsilon}{n\log|\mathcal{A}|}\right)\left(1+\frac{\tilde{\varepsilon}}{n\log|\mathcal{A}|}\right)^K,
\end{aligned}$$

where (i) follows from the simple bound $e^t\le 1+t$ for $0\le t\le 1$ and (ii) is due to (52). The same sequence of arguments together with (53) gives the following upper bound for (56):

$$\begin{aligned}
&\exp\left(-\frac{K\varepsilon}{n\log|\mathcal{A}|}\right)\mathbb{E}\left[\exp\left(\sum_{k=1}^K\frac{I\big(M^n\wedge Z^n_{s^n}(\tilde{G}_n)|\tilde{G}_n\big)}{n\log|\mathcal{A}|}\right)\right]\\
&\le\exp\left(-\frac{K\varepsilon}{n\log|\mathcal{A}|}\right)\left(1+\mathbb{E}\left[\frac{I\big(M^n\wedge Z^n_{s^n}(\tilde{G}_n)|\tilde{G}_n\big)}{n\log|\mathcal{A}|}\right]\right)^K\\
&\le\exp\left(-\frac{K\varepsilon}{n\log|\mathcal{A}|}\right)\left(1+\frac{\tilde{\varepsilon}}{n\log|\mathcal{A}|}\right)^K.
\end{aligned}$$

Thus (54) is upper-bounded by

$$\begin{aligned}
&2\exp\left(-K\left(\frac{\varepsilon}{n\log|\mathcal{A}|}-\log\left(1+\frac{\tilde{\varepsilon}}{n\log|\mathcal{A}|}\right)\right)\right)\\
&\le 2\exp\left(-\frac{K\varepsilon}{2n\log|\mathcal{A}|}\right),
\end{aligned}$$

where the inequality comes from the simple bound $\log(1+t)\le 2t$. Therefore, again due to the union bound, (57) at the top of the next page holds. Due to the choice of K, the right-hand side of (57) is strictly smaller than 1. Thus there is a realization $\mathcal{K}_n(\gamma_1),\dots,\mathcal{K}_n(\gamma_K)$ of

$$\begin{aligned}
&\mathbb{P}\left\{\frac{1}{K}\sum_{k=1}^K e_{s^n}\big(\mathcal{K}_n(\tilde{G}_n(k)\big)\ge\varepsilon\text{ or }\frac{1}{K}\sum_{k=1}^K I\big(M^n\wedge Z^n_{s^n}(\tilde{G}_n(k))|\tilde{G}_n(k)\big)\ge\varepsilon\right\} &(54)\\
&\le\mathbb{P}\left\{\exp\left(\sum_{k=1}^K\frac{e_{s^n}\big(\mathcal{K}_n(\tilde{G}_n(k)\big)}{n\log|\mathcal{A}|}\right)\ge\exp\left(\frac{K\varepsilon}{n\log|\mathcal{A}|}\right)\right\} &(55)\\
&+\mathbb{P}\left\{\exp\left(\sum_{k=1}^K\frac{I\big(M^n\wedge Z^n_{s^n}(\tilde{G}_n(k))|\tilde{G}_n(k)\big)}{n\log|\mathcal{A}|}\right)\ge\exp\left(\frac{K\varepsilon}{n\log|\mathcal{A}|}\right)\right\}. &(56)
\end{aligned}$$

$$\mathbb{P}\left\{\frac{1}{K}\sum_{k=1}^{K} e_{s^n}\big(\mathcal{K}_n(\tilde{G}_n(k))\big) \geq \varepsilon \text{ or } \frac{1}{K}\sum_{k=1}^{K} I\big(M^n \wedge Z^n_{s^n}(\tilde{G}_n(k))|\tilde{G}_n(k)\big) \geq \varepsilon \text{ for some } s^n \in \mathcal{S}^n\right\}$$
$$\leq \exp\left(-\frac{K\varepsilon}{2n\log|\mathcal{A}|} + n\log|\mathcal{S}| + 1\right). \tag{57}$$

$\mathcal{K}_n(\tilde{G}_n(1)), \ldots, \mathcal{K}_n(\tilde{G}_n(K))$ such that if we define G_n to have the probability distribution

$$P_{G_n}(\gamma) = \frac{1}{K}\sum_{k=1}^{K}\delta_{\{\gamma_k\}}(\gamma),$$

the correlated random code $\mathcal{K}_n^{\mathrm{ran}}$ induced by G_n satisfies (17)-(20). The complicated form of $P_{G_n}$is necessary to account for the possibility that $\gamma_k = \gamma_{k'}$ for $k \neq k'$. If the indices are pairwise different, G_n is uniformly distributed on $\{\gamma_1, \ldots, \gamma_K\}$.

The above construction can be done for all $n \geq n(R_S, \varepsilon)$. Thus the proof of Lemma 8 is complete.

Appendix C
Proof of Lemma 13

Assume $\mathcal{K}_n$ satisfies (29) and has stochastic encoder E. Recall that E^π is defined by $E^\pi(x^n|j) := E(\pi^{-1}(x^n)|j)$. The random variables below are from the canonical permutation invariant family $\mathcal{F}(\mathcal{K}_n, \mathfrak{W}, \mathfrak{V}, \Pi_n)$.

Lemma 18: For every $\pi \in \Pi_n$, we have

$$P_{M_n\pi(Z^n_{s^n}(\mathrm{id}))} = P_{M_n Z^n_{\pi(s^n)}(\pi)}.$$

Proof: Let $j \in \mathcal{J}_n$ and $z^n \in \mathcal{C}^n$. Then

$$\begin{aligned}
&\mathbb{P}[M_n = j, \pi(Z^n_{s^n}(\mathrm{id})) = z^n] \\
&= \mathbb{P}[M_n = j, Z^n_{s^n}(\mathrm{id}) = \pi^{-1}(z^n)] \\
&= \frac{1}{J_n}\sum_{x^n} E(x^n|j)V^n_{s^n}(\pi^{-1}(z^n)|x^n) \\
&= \frac{1}{J_n}\sum_{x^n} E(\pi^{-1}(x^n)|j)V^n_{s^n}(\pi^{-1}(z^n)|\pi^{-1}(x^n)) \\
&= \frac{1}{J_n}\sum_{x^n} E^\pi(x^n|j)V^n_{\pi(s^n)}(z^n|x^n) \\
&= \mathbb{P}[M_n = j, Z^n_{\pi(s^n)}(\pi) = z^n].
\end{aligned}$$

■

Now assume that (29) holds. Then

$$\begin{aligned}
&\max_{\pi\in\Pi_n}\max_{s^n} I(M_n \wedge Z^n_{s^n}(\pi)) \\
&= \max_{\pi\in\Pi_n}\max_{s^n} I(M_n \wedge Z^n_{\pi(s^n)}(\pi)) \\
&\overset{(i)}{=} \max_{\pi\in\Pi_n}\max_{s^n} I(M_n \wedge \pi(Z^n_{s^n}(\mathrm{id}))) \\
&\overset{(ii)}{\leq} \max_{s^n} I(M_n \wedge Z^n_{s^n}(\mathrm{id})) \\
&\leq \varepsilon
\end{aligned}$$

where Lemma 18 was applied in (i) and the data processing inequality in (ii). Thus (29) implies (30).

Appendix D
Channel Prefixing

Assume that $R_S^\dagger(\tilde{\overline{\mathfrak{W}}}, \tilde{V})$ is achievable with exponentially decreasing error for $(\tilde{\overline{\mathfrak{W}}}, \tilde{V})$ for every CAVWC $(\tilde{\overline{\mathfrak{W}}}, \tilde{\mathfrak{V}})$. We have to show that then for a given CAVWC $(\overline{\mathfrak{W}}, \mathfrak{V})$, $R_S^*(\overline{\mathfrak{W}}, \mathfrak{V})$ also is an achievable rate with exponentially decreasing error for $(\overline{\mathfrak{W}}, \mathfrak{V})$. Choose a positive integer k, a finite subset $\mathcal{U}$ of the integers, and a stochastic matrix $T : \mathcal{U} \to \mathcal{P}(\mathcal{A}^k)$. For every $r \in \mathcal{R}$ and $s^k \in \mathcal{S}^k$, this induces stochastic matrices $\tilde{W}_r : \mathcal{U} \to \mathcal{P}(\mathcal{B}^k)$ and $\tilde{V}_{s^k} : \mathcal{U} \to \mathcal{P}(\mathcal{C}^k)$ defined by

$$\tilde{W}_r(y^k|u) := \sum_{x^k} T(x^k|u)W_r^k(y^k|x^k),$$
$$\tilde{V}_{s^k}(y^k|u) := \sum_{x^k} T(x^k|u)V_{s^k}^k(z^k|x^k).$$

This induces families

$$\tilde{\overline{\mathfrak{W}}} := \{\tilde{W}_r^n : r \in \mathcal{R}, n = 1, 2, \ldots\},$$
$$\tilde{\mathfrak{V}} := \{\tilde{V}^n_{s^{kn}} : s^{kn} \in (\mathcal{S}^k)^n, n = 1, 2, \ldots\},$$

and hence a CAVWC which we denote by $(\tilde{\overline{\mathfrak{W}}}, \tilde{\mathfrak{V}})$. The compound part $\tilde{\overline{\mathfrak{W}}}$ of this channel also has $\mathcal{R}$ as its state set, the state set of the eavesdropper channel $\tilde{\mathfrak{V}}$ equals $\mathcal{S}^k$. By assumption, $R_S^\dagger(\tilde{\overline{\mathfrak{W}}}, \tilde{\mathfrak{V}})$ is an achievable rate with exponentially decreasing error for $(\tilde{\overline{\mathfrak{W}}}, \tilde{\mathfrak{V}})$. Thus there exists a $\beta > 0$ such that for every $\varepsilon > 0$ and sufficiently large n, one obtains an (n, J_n)-code $\tilde{\mathcal{K}}_n$ for $(\tilde{\overline{\mathfrak{W}}}, \tilde{\mathfrak{V}})$ with canonical random family $\mathcal{F}(\tilde{\mathcal{K}}_n, \tilde{\overline{\mathfrak{W}}}, \tilde{\mathfrak{V}}) = \{\tilde{M}^n, \tilde{U}^n, \tilde{Y}_r^{kn}, \tilde{Z}^{kn}_{s^{kn}}, \tilde{\hat{M}}_r^n : r \in \mathcal{R}, s^{kn} \in (\mathcal{S}^k)^n\}$ satisfying

$$\frac{1}{n}\log J_n \geq R_S^\dagger(\tilde{\overline{\mathfrak{W}}}, \tilde{\mathfrak{V}}) - \varepsilon, \tag{58}$$
$$-\frac{1}{n}\log \bar{e}(\tilde{\mathcal{K}}_n) \geq \beta - \varepsilon, \tag{59}$$
$$\max_{s^{kn}\in(\mathcal{S}^k)^n} I(\tilde{M}^n \wedge \tilde{Z}^{kn}_{s^{kn}}) \leq \varepsilon. \tag{60}$$

Now define the stochastic encoder $E : \mathcal{J}_n \to \mathcal{P}(\mathcal{A}^{kn})$ through

$$E(x^{kn}|j) := \sum_{u^n\in\mathcal{U}^n} E^*(u^n|j)T^n(x^{kn}|u^n).$$

Together with the decoding sets $\mathcal{D}_j^*$ considered as sets $\mathcal{D}_j \subset \mathcal{B}^{kn}$, this defines an uncorrelated (kn, J_n)-code $\mathcal{K}_{kn}$ for the CAVWC $(\overline{\mathfrak{W}}, \mathfrak{V})$. Observe that, if $\mathcal{F}(\mathcal{K}_{kn}, \overline{\mathfrak{W}}, \mathfrak{V}) = \{M^n, X^n, Y_r^{kn}, Z^{kn}_{s^{kn}}, \hat{M}_r^n : r \in \mathcal{R}, s^{kn} \in \mathcal{S}^{kn}\}$ is the canonical random family of $\mathcal{K}_{kn}$, then for every $r \in \mathcal{R}_n$ and s^{kn} regarded either as an element of $\mathcal{S}^{kn}$ or $(\mathcal{S}^k)^n$, the joint probability of $(M^n, Y_r^{kn}, Z^{kn}_{s^{kn}}, \hat{M}_r^{kn})$ equals that of $(\tilde{M}^n, \tilde{Y}_r^{kn}, \tilde{Z}^{kn}_{s^{kn}}, \tilde{\hat{M}}_r^n)$.

It immediately follows that

$$\begin{aligned}\frac{1}{kn}\log J_n &\geq \frac{1}{k}R_S^{\dagger}(\tilde{\overline{\mathfrak{W}}},\tilde{\mathfrak{V}})-\frac{\varepsilon}{k},\\ -\frac{1}{kn}\log\bar{e}(\mathcal{K}_{kn}) &\geq \frac{\beta-\varepsilon}{k},\\ \max_{s^{kn}\in\mathcal{S}^{kn}} I(M^n\wedge Z^{kn}_{s^{kn}}) &\leq \varepsilon.\end{aligned}$$

Thus after optimization over T and k, it follows that $R_S^*(\overline{\mathfrak{W}},\mathfrak{V})$ is an achievable secrecy rate with exponentially decreasing error for $(\overline{\mathfrak{W}},\mathfrak{V})$.

Appendix E
Types and Typical Sequences

The proofs of Lemmas 14 and 15 require some facts about types and typical sequences. For reference, we include them here. $\mathcal{A},\mathcal{B}$ and $W,\tilde{W}$ are generic sets/stochastic matrices.

Lemma 19: Let $\bar{X}$ be an $\mathcal{A}$-valued random variable and let $x^n\in\mathcal{T}^n_{\bar{X},\delta}$. Further let $W:\mathcal{A}\longrightarrow\mathcal{P}(\mathcal{S})$. Then for any $\mathcal{B}$-valued random variable $\bar{Y}$ with $P_{\bar{Y}|\bar{X}}=W$ and all $y^n\in\mathcal{T}^n_{\bar{Y}|\bar{X},\delta}(x^n)$,

$$\begin{aligned}|\mathcal{T}^n_{\bar{Y},\delta}| &\leq \exp\{n(H(\bar{Y})+f_1(\delta))\},\\ W^n(y^n|x^n) &\leq \exp\{-n(H(\bar{Y}|\bar{X})-f_2(\delta))\}\end{aligned}$$

with universal $f_1(\delta),f_2(\delta)>0$ satisfying $\lim_{\delta\to 0}f_1(\delta)=\lim_{\delta\to 0}f_2(\delta)=0$.

Lemma 20: Let $\delta>0$. Let $(\bar{X},\bar{Y})$ assume values in $\mathcal{A}\times\mathcal{B}$ such that $P_{\bar{Y}|\bar{X}}=W$, for some $W:\mathcal{A}\longrightarrow\mathcal{P}(\mathcal{B})$, and let $x^n\in\mathcal{A}^n$. There exist a universal $c'>0$ and an $n_0=n_0(|\mathcal{A}|,|\mathcal{B}|,\delta)\geq 1$ such that for $n\geq n_0$

$$\begin{aligned}P^n_{\bar{X}}(\mathcal{T}^n_{\bar{X},\delta}) &\geq 1-2^{-nc'\delta^2},\\ W^n(\mathcal{T}^n_{\bar{Y}|\bar{X},\delta}(x^n)|x^n) &\geq 1-2^{-nc'\delta^2}.\end{aligned}$$

Lemma 21: The cardinality of $\mathcal{P}^n_0(\mathcal{S})$ is upper-bounded by $(n+1)^{|\mathcal{S}|}$.

The proofs of Lemmas 19-21 can be found in e.g. [11]. A proof of the next lemma can be found in [5].

Lemma 22: Let $(\bar{X},\bar{Y})$ and $(\bar{X}',\bar{Y}')$ two pairs of $\mathcal{A}\times\mathcal{B}$-valued random variables. Then for sufficiently small $\delta>0$ and any positive integer n,

$$\begin{aligned}&P^n_{\bar{Y}}(\mathcal{T}^n_{\bar{Y}'|\bar{X}',\delta}(x^n))\\ &\quad\leq (n+1)^{|\mathcal{A}||\mathcal{B}|}\exp\{-n(I(\bar{X}'\wedge\bar{Y}')-f_3(\delta))\} \qquad (61)\end{aligned}$$

for all $\tilde{x}^n\in\mathcal{T}^n_{\bar{X}',\delta}$ holds for a universal $f_3(\delta)>0$ with $\lim_{n\to\infty}f_3(\delta)=0$.

Note that the right-hand side of (61) does not depend on $(\bar{X},\bar{Y})$, so one might wonder how sharp this bound is. But we will apply the lemma in a case where $\bar{X}=\bar{X}'$ and where $P_{\bar{Y}|\bar{X}}$ and $P_{\bar{Y}'|\bar{X}'}$ may be close (see Appendix F). Thus it turns out to give the correct upper bound.

Appendix F
Proof of Lemma 14

The fact that the probability of $\bar{e}(\mathcal{K}^{\mathcal{X}}_n)$ being small is large is well-known in principle, cf. [11]. As our choice of codewords does not quite follow the standard approach and we use stochastic encoders, we present the proof nonetheless. We start with a lemma which assumes a finite state set for $\overline{\mathfrak{W}}$ and actually shows that the sender can also reliably decode the randomization index with high probability. Recall our definitions of J_n in (32), of L_n in (33) and of $\mathcal{X}$ at the beginning of Subsection VI-B. Also recall the positive δ,τ from those definitions.

Now we define for every finite $\mathcal{R}'\subset\mathcal{R}$ a random uncorrelated (n,J_nL_n)-code $\mathcal{K}^{\mathcal{X},\mathcal{R}'}_n$ as follows: We take $\mathcal{J}_n\times\mathcal{L}_n$ as its message set. The encoder $f^{\mathcal{X},\mathcal{R}'}$ of $\mathcal{K}^{\mathcal{X},\mathcal{R}'}_n$ maps every pair $(j,l)\in\mathcal{J}_n\times\mathcal{L}_n$ into the codeword X_{jl}. With

$$\tilde{\mathcal{D}}^{\mathcal{X},\mathcal{R}'}_{jl} := \bigcup_{r\in\mathcal{R}}\mathcal{T}^n_{\bar{Y}_r|\bar{X},\delta}(X_{jl}),$$

the decoding sets are defined as

$$\mathcal{D}^{\mathcal{X},\mathcal{R}'}_{jl} := \tilde{\mathcal{D}}^{\mathcal{X},\mathcal{R}'}_{jl}\cap\Bigl(\bigcup_{(j',l')\in\mathcal{J}_n\times\mathcal{L}_n\setminus\{(j,l)\}}\tilde{\mathcal{D}}^{\mathcal{X},\mathcal{R}'}_{j'l'}\Bigr)^c.$$

Obviously, the $\mathcal{D}^{\mathcal{X},\mathcal{R}'}_{jl}$ are pairwise disjoint ($(j,l)\in\mathcal{J}_n\times\mathcal{L}_n$).

When applied for transmission over the complete compound channel $\overline{\mathfrak{W}}$, the average error of $\mathcal{K}^{\mathcal{X},\mathcal{R}'}_n$ is denoted by $\bar{e}(\mathcal{K}^{\mathcal{X},\mathcal{R}'}_n)$ as usual. However, if the channel states are restricted to the set $\mathcal{R}'$, we denote the corresponding error by

$$\bar{e}_{\mathcal{R}'}(\mathcal{K}^{\mathcal{X},\mathcal{R}'}_n) := \max_{r\in\mathcal{R}'}\mathbb{P}[M^n\neq\hat{M}_r],$$

where M^n and $\hat{M}^n_r$ are the canonical random variables induced by $\mathcal{K}^{\mathcal{X},\mathcal{R}'}_n$ and $\overline{\mathfrak{W}}$. The dependence of $\hat{M}^n_r$ on $\mathcal{X}$ is suppressed in the notation. The next lemma deals with this restricted average error $\bar{e}_{\mathcal{R}'}(\mathcal{K}^{\mathcal{X},\mathcal{R}'}_n)$.

Lemma 23: Let $\mathcal{R}'\subset\mathcal{R}$ be finite. For sufficiently small $\delta>0$, there exists an $a(\mathcal{R}')=a(\mathcal{R}',\tau,\delta)>0$ such that

$$\mathbb{P}\left\{\bar{e}_{\mathcal{R}'}(\mathcal{K}^{\mathcal{X},\mathcal{R}'}_n)\leq 2^{-na(\mathcal{R}')}\right\}\geq 1-2^{-na(\mathcal{R}')}.$$

Proof: First, we bound the expectation of the average error under a fixed channel state $r\in\mathcal{R}'$, which is given by

$$\bar{e}_r(\mathcal{K}^{\mathcal{X},\mathcal{R}'}_n) := \frac{1}{J_nL_n}\sum_{j=1}^{J_n}\sum_{l=1}^{L_n}W^n_r((\mathcal{D}^{\mathcal{X},\mathcal{R}'}_{jl})^c|X_{jl}).$$

We have

$$\begin{aligned}\mathbb{E}\Bigl[\bar{e}_r(\mathcal{K}^{\mathcal{X},\mathcal{R}'}_n)\Bigr] &= \mathbb{E}\Bigl[W^n_r((\mathcal{D}^{\mathcal{X},\mathcal{R}'}_{11})^c|X_{11})\Bigr]\\ &\leq \mathbb{E}\Bigl[W^n_r((\tilde{\mathcal{D}}^{\mathcal{X},\mathcal{R}'}_{11})^c|X_{11})\Bigr] \qquad (62)\\ &\quad+\sum_{\substack{(j,l)\in\mathcal{J}_n\times\mathcal{L}_n:\\ (j,l)\neq(1,1)}}\mathbb{E}\Bigl[W^n_r(\tilde{\mathcal{D}}^{\mathcal{X},\mathcal{R}'}_{jl}|X_{11})\Bigr]. \qquad (63)\end{aligned}$$

For (62), we have

$$\mathbb{E}\Bigl[W^n_r((\tilde{\mathcal{D}}^{\mathcal{X},\mathcal{R}'}_{11})^c|X_{11})\Bigr]\leq\mathbb{E}\Bigl[W^n_r((\mathcal{T}^n_{\bar{Y}_r|\bar{X},\delta}(X_{11}))^c|X_{11})\Bigr],$$

which by Lemma 20 is upper-bounded by $2^{-nc'\delta^2}$ for n sufficiently large. Thus (62) is upper-bounded by the same number. For each of the terms in (63), we obtain

$$\mathbb{E}\Bigl[W^n_r(\tilde{\mathcal{D}}^{\mathcal{X},\mathcal{R}'}_{jl}|X_{11})\Bigr]\leq\sum_{r'\in\mathcal{R}'}\mathbb{E}\Bigl[W^n_r(\mathcal{T}^n_{\bar{Y}_{r'}|\bar{X},\delta}(X_{jl})|X_{11})\Bigr].$$

For sufficiently large n, the terms on the right-hand side can be written (recall that $(j,l) \neq (1,1)$)

$$\begin{aligned}
&\mathbb{E}\left[W_r^n(\mathcal{T}^n_{\bar{Y}_{r'}|\bar{X},\delta}(X_{jl})|X_{11})\right] \\
&\quad = \sum_{x^n,\tilde{x}^n \in \mathcal{T}^n_{\bar{X},\delta}} W_r^n(\mathcal{T}^n_{\bar{Y}_{r'}|\bar{X},\delta}(\tilde{x}^n)|x^n)P'(x^n)P'(\tilde{x}^n) \\
&\quad \overset{(i)}{\leq} (1-2^{-nc'\delta})^{-2} \sum_{\tilde{x}^n \in T^n_{\bar{X},\delta}} P^n_{\bar{Y}_r}(\mathcal{T}^n_{\bar{Y}_{r'}|\bar{X},\delta}(\tilde{x}^n))P^n_{\bar{X}}(\tilde{x}^n), \qquad (64)
\end{aligned}$$

where we used the definition of P' and Lemma 20 in (i). By Lemma 22,

$$P^n_{\bar{Y}_r}(\mathcal{T}^n_{\bar{Y}_{r'}|\bar{X},\delta}(\tilde{x}^n)) \leq (n+1)^{|\mathcal{A}||\mathcal{B}|}2^{-n(I(\bar{X}\wedge\bar{Y}_{r'})-f_3(\delta))}$$

with $f_3(\delta) \to 0$ as $\delta \to 0$. This immediately gives

$$(64) \leq (1-2^{-nc'\delta})^{-2}(n+1)^{|\mathcal{A}||\mathcal{B}|}2^{-n(I(\bar{X}\wedge\bar{Y}_{r'})-f_3(\delta))},$$

and we can upper-bound (63) by

$$|\mathcal{R}'|J_nL_n\exp\left\{-n\left(\min_{r'\in\mathcal{R}'} I(\bar{X}\wedge\bar{Y}_{r'})-2f_3(\delta)\right)\right\}.$$

If one chooses δ so small that $\tau \geq 4f_3(\delta) > 0$ and since $\mathcal{R}'$ is finite, this tends to 0 exponentially. Combining the bounds on (62) and (63), we thus obtain

$$\mathbb{E}\left[\bar{e}_r(\mathcal{K}_n^{\mathcal{X},\mathcal{R}'})\right] \leq 2^{-na'} \qquad (65)$$

for some appropriate $a'(\mathcal{R}') = a'(\mathcal{R}',\tau,\delta) > 0$.

Now we can complete the proof of the lemma. Using the Markov inequality and setting $a(\mathcal{R}') := a'(\mathcal{R}')/3$, we obtain from (65)

$$\begin{aligned}
&\mathbb{P}\left\{\bar{e}_{\mathcal{R}'}(\mathcal{K}_n^{\mathcal{X},\mathcal{R}'}) \leq 2^{-na(\mathcal{R}')}\right\} \\
&\quad = \mathbb{P}\left\{\max_{r\in\mathcal{R}'}\bar{e}_r(\mathcal{K}_n^{\mathcal{X},\mathcal{R}'}) \leq 2^{-na(\mathcal{R}')}\right\} \\
&\quad = \mathbb{P}\left[\bigcap_{r\in\mathcal{R}'}\left\{\bar{e}_r(\mathcal{K}_n^{\mathcal{X},\mathcal{R}'}) \leq 2^{-na(\mathcal{R}')}\right\}\right] \\
&\quad \geq 1-\sum_{r\in\mathcal{R}'}\mathbb{P}[\bar{e}_r(\mathcal{K}_n^{\mathcal{X},\mathcal{R}'}) > 2^{-na(\mathcal{R}')}] \\
&\quad \geq 1-2^{na(\mathcal{R}')}\sum_{r\in\mathcal{R}'}\mathbb{E}[\bar{e}_r(\mathcal{K}_n^{\mathcal{X},\mathcal{R}'})] \\
&\quad \geq 1-|\mathcal{R}'|2^{na(\mathcal{R}')}2^{-3na(\mathcal{R}')} \\
&\quad \geq 1-2^{-na(\mathcal{R}')}
\end{aligned}$$

for sufficiently large n. Thus the probability that $\bar{e}_{\mathcal{R}'}(\mathcal{K}_n^{\mathcal{X},\mathcal{R}'}) \leq 2^{-na(\mathcal{R}')}$ is lower-bounded by $1-2^{-na(\mathcal{R}')}$. This completes the proof. ■

We now invoke the approximation argument of [6, Lemma 4], from which we conclude that there exists a finite $\mathcal{R}' \subset \mathcal{R}$ such that

$$\mathbb{P}\left\{\bar{e}(\mathcal{K}_n^{\mathcal{X},\mathcal{R}'}) \leq 2^{-na(\mathcal{R}')/2}\right\} \geq 1-2^{-na(\mathcal{R}')/2}. \qquad (66)$$

Thus even though the code is designed for a finite state subset $\mathcal{R}'$, it also works well for transmission over the complete channel set $\overline{\mathfrak{W}}$. In particular, note that the index $l \in \mathcal{L}_n$ can still be decoded when a good realization of $\mathcal{K}_n^{\mathcal{X},\mathcal{R}'}$ is applied.

We choose any $\mathcal{R}'$ satisfying (66). Now recall the definition of $E^{\mathcal{X}}$. Together with the decoding sets

$$\mathcal{D}_j^{\mathcal{X}} := \bigcup_{l\in\mathcal{L}_n}\mathcal{D}_{jl}^{\mathcal{X},\mathcal{R}'} \qquad (j \in \mathcal{J}_n)$$

this defines a randomly chosen uncorrelated (n,J_n)-code $\mathcal{K}_n^{\mathcal{X}}$. Note that

$$\begin{aligned}
\bar{e}(\mathcal{K}_n^{\mathcal{X}}) &= \frac{1}{J_n}\sum_{j\in\mathcal{J}_n}\sum_{x^n}E^{\mathcal{X}}(x^n|j)W^n((\mathcal{D}_j^{\mathcal{X}})^c|x^n) \\
&= \frac{1}{J_nL_n}\sum_{j\in\mathcal{J}_n}\sum_{l\in\mathcal{L}_n}W^n((\mathcal{D}_j^{\mathcal{X}})^c|X_{jl}) \\
&\leq \frac{1}{J_nL_n}\sum_{j\in\mathcal{J}_n}\sum_{l\in\mathcal{L}_n}W^n((\mathcal{D}_{jl}^{\mathcal{X},\mathcal{R}'})^c|X_{jl}) \\
&= \bar{e}(\mathcal{K}_n^{\mathcal{X},\mathcal{R}'}).
\end{aligned}$$

This last term is exponentially small with high probability by (66), which proves Lemma 14 with $\tau_6 = a(\mathcal{R}')/2$.

Appendix G
Proof of Lemma 15

Recall the definitions of $J_n, L_n, \mathcal{X}$ from Subsection VI-B. Also recall the τ from the definitions of J_n and L_n and the δ from the definition of $\mathcal{X}$. Both τ and δ are arbitrary real numbers. These definitions will be valid throughout this appendix and for all lemmas stated here.

In the next two sections of this appendix, we will define events $\imath_1(j,z^n,s^n)$ and $\imath_2(j,s^n)$, for $j \in \mathcal{J}_n$, $z^n \in \mathcal{Z}^n$ and $s^n \in \mathcal{S}^n$. By Lemma 24 stated in the third section of this appendix, the $\imath_0$ defined in Lemma 15 satisfies

$$\imath_0 \supset \bigcap_{j,z^n,s^n}\imath_1(j,z^n,s^n) \cap \bigcap_{j,s^n}\imath_2(j,s^n). \qquad (67)$$

Further, Lemmas 25 and 26, also stated in the third section, will show that with an appropriate choice of δ, the probability of each of the events of the right-hand side of (67) is very close to 1. This is sufficient to show that $\mathbb{P}[\imath_0] > 1-2^{-\tau_2 n}$, as claimed in Lemma 15, which is done in the last section of this appendix.

A. Definition of $\imath_1(j,z^n,s^n)$

For some positive α to be chosen later, let $\varepsilon_n := 2^{-n\alpha}$. Fix $s^n \in \mathcal{S}^n$, and denote its type by $q \in \mathcal{P}_0^n(\mathcal{S})$. For $x^n \in \mathcal{A}^n$, define

$$\begin{aligned}
\mathcal{E}_1(x^n,s^n) := \Big\{&z^n \in \mathcal{T}^n_{\bar{Z}_q,4|\mathcal{A}||\mathcal{S}|\delta}: \\
&V^n_{s^n}(z^n|x^n) \leq \exp\big(-n(H(\bar{Z}_q|\bar{X})-f_2(3|\mathcal{S}|\delta))\big)\Big\},
\end{aligned}$$

where f_2 is the function from Lemma 19, and set

$$\tilde{\Theta}_{s^n}(z^n) := \mathbb{E}[V^n_{s^n}(z^n|X_{11})\mathbb{1}_{\mathcal{E}_1(X_{11},s^n)}(z^n)]. \qquad (68)$$

Further define

$$\mathcal{E}_2(s^n) := \{z^n \in \mathcal{T}^n_{\bar{Z}_q,4|\mathcal{A}||\mathcal{S}|\delta}: \tilde{\Theta}_{s^n}(z^n) \geq \varepsilon_n|\mathcal{T}^n_{\bar{Z}_q,4|\mathcal{A}||\mathcal{S}|\delta}|^{-1}\}$$

and set

$$\Theta_{s^n}(z^n) := \tilde{\Theta}_{s^n}(z^n)\mathbb{1}_{\mathcal{E}_2(s^n)}(z^n).$$

Note that by definition,

$$\Theta_{s^n}(z^n) \geq \varepsilon_n \exp\{-n(H(\bar{Z}_q) + f_1(4|\mathcal{A}||\mathcal{S}|\delta))\} \qquad (69)$$

for the function f_1 from Lemma 19 if $z^n \in \mathcal{T}^n_{\bar{Z}_q,4|\mathcal{A}||\mathcal{S}|\delta}$ and that $\Theta_{s^n}(z^n) = 0$ otherwise.

With the sets just defined, we obtain a modification of $V^n_{s^n}$ by defining

$$Q_{s^n,z^n}(x^n) := V^n_{s^n}(z^n|x^n)\mathbb{1}_{\mathcal{E}_1(x^n,s^n)}(z^n)\mathbb{1}_{\mathcal{E}_2(s^n)}(z^n).$$

Note that this is not an actual "channel" as in general $\sum_{z^n} Q_{s^n,z^n}(x^n) < 1$. Finally, we define

$$\imath_1(j, z^n, s^n) := \left\{\frac{1}{L_n}\sum_{l=1}^{L_n} Q_{s^n,z^n}(X_{jl}) \in [(1 \pm \varepsilon_n)\Theta_{s^n}(z^n)]\right\}$$

where $[(1 \pm \varepsilon_n)\Theta_{s^n}(z^n)]$ is short for $[(1 - \varepsilon_n)\Theta_{s^n}(z^n), (1 + \varepsilon_n)\Theta_{s^n}(z^n)]$.

B. Definition of $\imath_2(j, s^n)$

Let $q \in \mathcal{P}^n_0(\mathcal{S})$ be the type of s^n and let $\bar{S}_q$ be an $\mathcal{S}$-valued random variable with $P_{\bar{S}_q} = q$ and independent of the family $\{\bar{X}, \bar{Y}_r, \bar{Z}_q : r \in \mathcal{R}, q \in \mathcal{P}(\mathcal{S})\}$ which defines $J_n, L_n, \mathcal{X}$. Then we define

$$\imath_2(j, s^n) := \Big\{\left|\{l \in \mathcal{L}_n : s^n \in T^n_{\bar{S}_q|\bar{X},2\delta}(X_{jl})\}\right| \geq (1 - \varepsilon_n - 2^{-nc'\delta^2})L_n\Big\}.$$

C. Statement of Lemmas 24-26

Lemmas 24-26 will be proved in Appendix H.

Lemma 24: Assume a realization $\mathbf{x} := \{x_{jl} : j \in \mathcal{J}_n, l \in \mathcal{L}_n\}$ of $\mathcal{X}$ has the following properties: For all $j \in \mathcal{J}_n$ and $z^n \in \mathcal{C}^n$ and $q \in \mathcal{P}^n_0(\mathcal{S})$ and $s^n \in S^n$,

$$\frac{1}{L_n}\sum_{l=1}^{L_n} Q_{s^n,z^n}(x_{jl}) \in [(1 \pm \varepsilon_n)\Theta_{s^n}(z^n)], \qquad (70)$$

$$\frac{|\{l \in \mathcal{L}_n : s^n \in T^n_{\bar{S}_q,2\delta}(x_{jl})\}|}{L_n} \geq (1 - \varepsilon_n - 2^{-nc'\delta^2}). \qquad (71)$$

Then

$$\max_{j\in\mathcal{J}_n}\max_{s^n\in\mathcal{S}^n}\|P_{Z^n_{s^n}|M^n}(\,\cdot\,|j) - \Theta_{s^n}(\cdot)\| \leq 4(\varepsilon_n + 2^{-nc'\delta^2}).$$

In particular, (67) is true if the τ_1 in the definition of $\imath_0$ is set to $\tau_1 =: \min\{\alpha, c'\delta^2\}/2$.

Lemma 25: For sufficiently small $\alpha > 0$ and $\delta > 0$ there exists a $\tau_3 > 0$ such that for n large and every $j \in \mathcal{J}_n$, $z^n \in \mathcal{C}^n$ and $s^n \in \mathcal{S}^n$

$$\mathbb{P}\Big[\imath_1(j, z^n, s^n)^c\Big] \leq 2\exp\Big(-\exp\big\{n\tau_3\big\}\Big).$$

Lemma 26: For sufficiently small $\alpha > 0$ there exists a $\tau_5 > 0$ such that for n large and every $j \in \mathcal{J}_n$ and $s^n \in \mathcal{S}^n$,

$$\mathbb{P}[\imath_2(j, s^n)] \leq 2\exp\left(-\exp\left\{n\left(\max_{q\in\mathcal{P}(\mathcal{S})} I(\bar{X} \wedge \bar{Z}_q) + \tau_5\right)\right\}\right)$$

where the random variables $\bar{X}, \bar{Z}_q$ are those from the definition of $\imath_2(j, s^n)$.

D. Proof of Lemma 15

We choose α and δ so small that Lemmas 24-26 hold. Lemmas 25 and 26 show that the probability of the complement of each of the events $\imath_1(j, z^n, s^n)$ and $\imath_2(j, s^n)$ is upper-bounded by a term which tends to zero doubly-exponentially as the blocklength increases. Then

$$\begin{aligned}
&\mathbb{P}[\imath_0]\\
&= 1 - \mathbb{P}[\imath_0^c]\\
&\overset{(i)}{\geq} 1 - \mathbb{P}\left[\bigcup_{j,z^n,s^n}\imath_1(j, z^n, s^n)^c \cup \bigcup_{j,s^n}\imath_2(j, s^n)^c\right]\\
&\overset{(ii)}{\geq} 1 - 2J_n|\mathcal{C}|^n|\mathcal{S}|^n\exp\Big(-\exp\big\{n\tau_3\big\}\Big)\\
&\qquad - 2J_n|\mathcal{S}|^n\exp\left(-\exp\left\{n\left(\max_{q\in\mathcal{P}(\mathcal{S})} I(\bar{X} \wedge \bar{Z}_q) + \tau_5\right)\right\}\right)\\
&\overset{(iii)}{\geq} 1 - 2^{-n\tau_2},
\end{aligned}$$

where (i) is due to (67), which holds due to Lemma 24, (ii) is due to the union bound and (iii) holds because an appropriate $\tau_2 > 0$ can be found due to the doubly exponential decrease of the probabilities in Lemmas 25 and 26. Altogether, this proves Lemma 15.

APPENDIX H
PROOFS OF LEMMAS 24-26

E. Proof of Lemma 24

We first show two auxiliary results. Recall the convention that we sometimes write $V(c|a, s)$ instead of $V_s(c|a)$. We also use the same family of random variables $\{(\bar{X}, \bar{Y}_r, \bar{Z}_q, \bar{S}_q) : r \in \mathcal{R}, q \in \mathcal{P}(\mathcal{S})\}$ as in the definition of $\imath_2(j, s^n)$, i.e., the family $\{(\bar{X}, \bar{Y}_r, \bar{Z}_q) : r \in \mathcal{R}, q \in \mathcal{P}(\mathcal{S})\}$ is the family which defines J_n, L_n and $\mathcal{X}$, and every $\bar{S}_q$ is independent of this family, attains values in $\mathcal{S}$ and satisfies $P_{\bar{S}_q} = q$.

Lemma 27: Let $x^n \in \mathcal{T}^n_{\bar{X},\delta}$ and let s^n have type $q \in \mathcal{P}^n_0(\mathcal{S})$. Let the random variable $\underline{Z}_q$ satisfy $P_{\underline{Z}_q|\bar{X}\bar{S}_q}(\cdot|\cdot,\cdot) = V(\cdot|\cdot,\cdot)$. If $s^n \in \mathcal{T}^n_{\bar{S}_q,2\delta}(x^n)$, then $\mathcal{T}^n_{\underline{Z}_q|\bar{X}\bar{S}_q,\delta}(x^n, s^n) \subset \mathcal{E}_1(x^n, s^n)$.

Proof: For $x^n \in \mathcal{T}^n_{\bar{X},\delta}$, we have $\mathcal{T}^n_{\bar{Z}_q|\bar{X},3|\mathcal{S}|\delta}(x^n) \subset \mathcal{T}^n_{\bar{Z}_q,4|\mathcal{A}||\mathcal{S}|\delta}$. Thus due to Lemma 19, it suffices to show that if s^n has type q, then $\mathcal{T}^n_{\underline{Z}_q|\bar{X}\bar{S}_q,\delta}(x^n, s^n) \subset \mathcal{T}^n_{\bar{Z}_q|\bar{X},3|\mathcal{S}|\delta}(x^n)$. For $a \in \mathcal{A}$ and $c \in \mathcal{C}$, we calculate

$$\begin{aligned}
&\left|\frac{1}{n}N(c, a|z^n, x^n) - \sum_{s\in\mathcal{S}} q(s)V(c|a, s)\frac{1}{n}N(a|x^n)\right|\\
&\leq \sum_{s\in\mathcal{S}}\left|\frac{1}{n}N(c, a, s|z^n, x^n, s^n) - q(s)V(c|a, s)\frac{1}{n}N(a|x^n)\right|\\
&\leq \sum_{s\in\mathcal{S}}\left|\frac{1}{n}N(c, a, s|z^n, x^n, s^n) - V(c|a, s)\frac{1}{n}N(a, s|x^n, s^n)\right|\\
&\qquad + \sum_{s\in\mathcal{S}} V(c|a, s)\left|\frac{1}{n}N(a, s|x^n, s^n) - q(s)\frac{1}{n}N(a|x^n)\right|\\
&\leq |\mathcal{S}|(\delta + 2\delta) = 3|\mathcal{S}|\delta.
\end{aligned}$$

∎

Corollary 3: If n is sufficiently large, then every $s^n \in \mathcal{S}^n$ satisfies

$$\Theta_{s^n}(\mathcal{C}^n) \geq 1 - 2 \cdot 2^{-nc'\delta^2} - \varepsilon_n$$

Proof: Let s^n have type $q \in \mathcal{P}_0^n(\mathcal{S})$. By the definition of Θ_{s^n}, we have $\Theta_{s^n}(\mathcal{C}^n) = \Theta_{s^n}(\mathcal{E}_2(s^n))$. As the support of $\tilde{\Theta}_{s^n}$ is contained in $T^n_{\bar{Z}_q,4|\mathcal{A}||\mathcal{S}|\delta}$, we have $\Theta_{s^n}(\mathcal{E}_2(s^n)) \geq \tilde{\Theta}_{s^n}(\mathcal{T}^n_{\bar{Z}_q,4|\mathcal{A}||\mathcal{S}|\delta}) - \varepsilon_n = \tilde{\Theta}_{s^n}(\mathcal{C}^n) - \varepsilon_n$. By definition,

$$\begin{aligned}\tilde{\Theta}_{s^n}(\mathcal{C}^n) &= \mathbb{E}[V^n_{s^n}(\mathcal{E}_1(X_{11}, s^n)|X_{11})] \\ &\geq \mathbb{E}[V^n_{s^n}(\mathcal{E}_1(X_{11}, s^n)|X_{11})|s^n \in \mathcal{T}^n_{\bar{S}_q|\bar{X},2\delta}(X_{11})] \\ &\quad \times \mathbb{P}[s^n \in \mathcal{T}^n_{\bar{S}_q|\bar{X},2\delta}(X_{11})].\end{aligned}$$

For sufficiently large n

$$\begin{aligned}&\mathbb{E}[V^n_{s^n}(\mathcal{E}_1(X_{11}, s^n)|X_{11})|s^n \in \mathcal{T}^n_{\bar{S}_q|\bar{X},2\delta}(X_{11})] \\ &\quad \overset{(i)}{\geq} \mathbb{E}[V^n(\mathcal{T}^n_{\underline{Z}_q|\bar{X}\bar{S}_q,\delta}(X_{11}, s^n)|X_{11}, s^n)|s^n \in \mathcal{T}^n_{\bar{S}_q|\bar{X},2\delta}(X_{11})] \\ &\quad \overset{(ii)}{\geq} 1 - 2^{-nc'\delta^2},\end{aligned}$$

where we used Lemma 27 in (i) and Lemma 20 in (ii). Lemma 29 provides a lower bound on $\mathbb{P}[s^n \in \mathcal{T}^n_{\bar{S}_q|\bar{X},2\delta}(X_{11})]$, so altogether,

$$\begin{aligned}\Theta_{s^n}(\mathcal{C}^n) &\geq \tilde{\Theta}_{s^n}(\mathcal{C}^n) - \varepsilon_n \\ &\geq (1 - 2^{-nc'\delta^2})^2 - \varepsilon_n \\ &\geq 1 - 2 \cdot 2^{-nc'\delta^2} - \varepsilon_n. \qquad (72)\end{aligned}$$

■

Let $\mathbf{x} = \{x_{jl} : j \in \mathcal{J}_n, l \in \mathcal{L}_n\}$ be a realization of $\mathcal{X}$ satisfying (70) and (71). Let $\mathcal{K}_n$ be the corresponding uncorrelated (n, J_n)-code and $\mathcal{F}(\mathcal{K}_n, \overline{\mathfrak{W}}, \mathfrak{V}) = \{M^n, X^n, Y^n_r, Z^n_{s^n}, \hat{M}_r : r \in \mathcal{R}, s^n \in \mathcal{S}^n\}$ the canonical family of random variables associated with $\mathcal{K}_n$. For any $j \in \mathcal{J}_n$ and any s^n with type $q \in \mathcal{P}_0(\mathcal{S})$, we decompose the total variation distance as follows:

$$\begin{aligned}&\|P_{Z^n_{s^n}|M^n}(\,\cdot\,|j) - \Theta_{s^n}(\cdot)\| \\ &\leq \left\| \frac{1}{L_n} \sum_{l=1}^{L_n} Q_{s^n,\cdot}(x_{jl}) - \Theta_{s^n}(\cdot) \right\| \qquad (73) \\ &\quad + \left\| \frac{1}{L_n} \sum_{l=1}^{L_n} V^n_{s^n}(\,\cdot\,|x_{jl}) \mathbb{1}_{\mathcal{E}_1(x_{jl},s^n)}(\cdot)(\mathbb{1}_{\mathcal{C}^n}(\cdot) - \mathbb{1}_{\mathcal{E}_2(s^n)}(\cdot)) \right\| \qquad (74) \\ &\quad + \left\| \frac{1}{L_n} \sum_{l=1}^{L_n} V^n_{s^n}(\,\cdot\,|x_{jl})(\mathbb{1}_{\mathcal{C}^n}(\cdot) - \mathbb{1}_{\mathcal{E}_1(x_{jl},s^n)}(\cdot)) \right\|. \qquad (75)\end{aligned}$$

The term in (73) is upper-bounded by ε_n, because due to (70)

$$\begin{aligned}&\left\| \frac{1}{L_n} \sum_{l=1}^{L_n} Q_{s^n,\cdot}(x_{jl}) - \Theta_{s^n}(\cdot) \right\| \\ &\quad = \sum_{z^n} \left| \frac{1}{L_n} \sum_{l=1}^{L_n} Q_{s^n,z^n}(x_{jl}) - \Theta_{s^n}(z^n) \right| \\ &\quad \leq \varepsilon_n \sum_{z^n} \Theta_{s^n}(z^n) \\ &\quad \leq \varepsilon_n.\end{aligned}$$

Next, applying (70) in (i), we upper-bound (74) as

$$\begin{aligned}&\frac{1}{L_n} \sum_{l=1}^{L_n} \sum_{z^n} V_{s^n}(z^n|x_{jl}) \mathbb{1}_{\mathcal{E}_1(x_{jl},s^n)}(z^n) \\ &\qquad - \frac{1}{L_n} \sum_{l=1}^{L_n} \sum_{z^n} V_{s^n}(z^n|x_{jl}) \mathbb{1}_{\mathcal{E}_1(x_{jl},s^n)}(z^n) \mathbb{1}_{\mathcal{E}_2(s^n)}(z^n) \\ &\quad \leq 1 - \sum_{z^n} \frac{1}{L_n} \sum_{l=1}^{L_n} Q_{s^n,z^n}(x_{jl}) \\ &\quad \overset{(i)}{\leq} 1 - (1 - \varepsilon_n)\Theta_{s^n}(\mathcal{C}^n).\end{aligned}$$

Upon application of Corollary 3, we obtain that (74) can be upper-bounded by

$$1 - (1 - \varepsilon_n)(1 - 2 \cdot 2^{-nc'\delta^2} - \varepsilon_n) \leq 2(2^{-nc'\delta} + \varepsilon_n).$$

It remains to upper-bound (75). Recall the definition of $\underline{Z}_q$. We have

$$\begin{aligned}&\left\| \frac{1}{L_n} \sum_{l=1}^{L_n} V^n_{s^n}(\,\cdot\,|x_{jl})(\mathbb{1}_{\mathcal{C}^n}(\cdot) - \mathbb{1}_{\mathcal{E}_1(x_{jl},s^n)}(\cdot)) \right\| \\ &\quad = \frac{1}{L_n} \sum_{l=1}^{L_n} V^n_{s^n}(\mathcal{E}_1(x_{jl}, s^n)^c|x_{jl}) \\ &\quad = \frac{1}{L_n} \sum_{\substack{l \in \mathcal{L}_n: \\ \mathcal{T}^n_{\underline{Z}_q|\bar{X}\bar{S}_q,\delta}(x_{jl},s^n) \subset \mathcal{E}_1(x_{jl},s^n)}} V^n_{s^n}(\mathcal{E}_1(x_{jl}, s^n)^c|x_{jl}) \\ &\qquad + \frac{1}{L_n} \sum_{\substack{l \in \mathcal{L}_n: \\ \mathcal{T}^n_{\underline{Z}_q|\bar{X}\bar{S}_q,\delta}(x_{jl},s^n) \not\subseteq \mathcal{E}_1(x_{jl},s^n)}} V^n_{s^n}(\mathcal{E}_1(x_{jl}, s^n)^c|x_{jl}). \qquad (76)\end{aligned}$$

If $\mathcal{T}^n_{\underline{Z}_q|\bar{X}\bar{S}_q,\delta}(x_{jl}, s^n) \subset \mathcal{E}_1(x_{jl}, s^n)$, then by Lemma 20, we have

$$\begin{aligned}V^n_{s^n}(\mathcal{E}_1(x_{jl}, s^n)^c|x_{jl}) &\leq V^n(\mathcal{T}^n_{\underline{Z}_q|\bar{X}\bar{S}_q,\delta}(x_{jl}, s^n)^c|x_{jl}, s^n) \\ &\leq 2^{-nc'\delta^2}.\end{aligned}$$

By Lemma 27 and (71), the proportion of those $l \in \mathcal{L}_n$ for which $\mathcal{T}^n_{\underline{Z}_q|\bar{X}\bar{S}_q,\delta}(x_{jl}, s^n) \not\subseteq \mathcal{E}_1(x_{jl}, s^n)$ holds is upper-bounded by $\varepsilon_n + 2^{-nc'\delta^2}$. We can thus bound (76) by

$$2^{-nc'\delta^2} + \varepsilon_n + 2^{-nc'\delta} = \varepsilon_n + 2 \cdot 2^{-nc'\delta^2}.$$

Collecting the bounds on (73), (74) and (75) completes the proof of Lemma 24.

F. Proof of Lemma 25

Let $j \in \mathcal{J}_n, z^n \in \mathcal{C}^n, s^n \in \mathcal{S}^n$. We want to upper-bound the probability of the complement of $\iota_1(j, z^n, s^n)$, i.e., of the event that

$$\left\{ \frac{1}{L_n} \sum_{l=1}^{L_n} Q_{s^n,z^n}(X_{jl}) \notin [(1 \pm \varepsilon_n)\Theta_{s^n}(z^n)] \right\}.$$

The form of this event already suggests that a Chernoff bound may be the right method for the proof. Indeed, we will apply the following lemma.

Lemma 28: Let b be a positive number. Let $Z_1, \dots, Z_L$ be i.i.d. random variables with values in $[0, b]$ and expectation $\mathbb{E}Z_l = \nu$, and let $0 < \varepsilon < \frac{1}{2}$. Then

$$\mathbb{P}\left\{\frac{1}{L}\sum_{l=1}^{L} Z_i \notin [(1 \pm \varepsilon)\nu]\right\} \leq 2\exp\left(-L \cdot \frac{\varepsilon^2 \nu}{3b}\right).$$

Proof: The proof can be found in [14, Th. 1.1] and in [3]. ■

The claim of Lemma 25 follows from an application of Lemma 28. Due to the definition of $\mathcal{E}_1(x^n, s^n)$, the independent random variables $Q_{s^n,z^n}(X_{jl})$ are upper-bounded by $b_n := \exp\{-n(H(\bar{Z}_q|\bar{X}) - f_2(3|\mathcal{S}|\delta))\}$ and have mean $\Theta_{s^n}(z^n)$. Applying Lemma 28 gives

$$\begin{aligned}
&\mathbb{P}[\iota_1(j, z^n, s^n)^c] \\
&\quad = \mathbb{P}\left\{\frac{1}{L_n}\sum_{l=1}^{L_n} Q_{s^n,z^n}(X_{jl}) \notin [(1 \pm \varepsilon_n)\Theta_{s^n}(z^n)]\right\} \\
&\quad \leq 2\exp\left(-L_n \cdot \frac{\varepsilon_n^2 \Theta_{s^n}(z^n)}{3b_n}\right).
\end{aligned} \tag{77}$$

For the exponent on the right-hand side of (77), we obtain

$$\begin{aligned}
&-L_n \cdot \frac{\varepsilon_n^2 \Theta_{s^n}(z^n)}{3b_n} \\
&\overset{(i)}{=} -\left\lfloor \exp\left\{n\left(\max_{q\in\mathcal{P}(S)} I(\bar{X} \wedge \bar{Z}_q) + \frac{\tau}{4}\right)\right\}\right\rfloor \\
&\quad \times \frac{2^{-2\alpha n}\Theta_{s^n}(z^n)}{3} \exp\Big\{n\big(H(\bar{Z}_q|\bar{X}) - f_2(3|\mathcal{S}|\delta)\big)\Big\} \\
&\overset{(ii)}{\leq} -\exp\left\{n\left(\max_{q\in\mathcal{P}(S)} I(\bar{X} \wedge \bar{Z}_q) + \frac{\tau}{5}\right)\right\} \cdot \frac{2^{-3\alpha n}}{3} \\
&\quad \times \exp\Big\{-n\big(I(\bar{X} \wedge \bar{Z}_q) + f_1(4|\mathcal{A}||\mathcal{S}|\delta) + f_2(3|\mathcal{S}|\delta)\big)\Big\} \\
&\overset{(iii)}{\leq} -\exp\left\{n\left(\frac{\tau}{5} - 3\alpha - \frac{\log 3}{n} - f_4(|\mathcal{A}||\mathcal{S}|\delta)\right)\right\},
\end{aligned} \tag{78}$$

where in (i) we inserted the definitions of L_n, ε_n, b_n, in (ii) we used the bound $\Theta_{s^n}(z^n) \geq \varepsilon_n \exp\{-n(H(\bar{Z}_q)+f_1(4|\mathcal{A}||\mathcal{S}|\delta))\}$ from (69) and in (iii) we set $f_4(|\mathcal{A}||\mathcal{S}|\delta) := f_1(4|\mathcal{A}||\mathcal{S}|\delta) + f_2(3|\mathcal{S}|\delta)$. If we choose α, δ so small that

$$\tau_3 := \frac{\tau}{6} - 3\alpha - f_1(4|\mathcal{A}||\mathcal{S}|\delta) - f_2(3|\mathcal{S}|\delta) > 0,$$

then (78) decreases to $-\infty$ at exponential speed. Hence (77) gives a bound on $\mathbb{P}[\iota_1(j, z^n, s^n)]$ which decreases to zero at doubly-exponential speed. This completes the proof of Lemma 25.

G. Proof of Lemma 26

The proof of Lemma 26 also applies the Chernoff bound of Lemma 28. For the application of the Chernoff bound, we first need a lower bound on $\mathbb{E}[\mathbb{1}_{\mathcal{T}^n_{\bar{S}_q,2\delta}(X_{11})}] = \mathbb{P}[s^n \in \mathcal{T}^n_{\bar{S}_q,2\delta}(X_{11})]$.

Lemma 29: For sufficiently large n and every s^n of type q,

$$\mathbb{P}[s^n \in \mathcal{T}^n_{\bar{S}_q,2\delta}(X_{11})] \geq 1 - 2^{-nc'\delta^2}.$$

Proof: We first show

$$\mathcal{T}^n_{\bar{X}|\bar{S}_q,\delta/|S|}(s^n) \subset \{x^n \in \mathcal{T}^n_{\bar{X},\delta} : s^n \in \mathcal{T}^n_{\bar{S}_q,2\delta}(x^n)\}. \tag{79}$$

Let $x^n \in \mathcal{T}^n_{\bar{X}|\bar{S}_q,\delta/|\mathcal{S}|}(s^n)$. Clearly $\mathcal{T}^n_{\bar{X}|\bar{S}_q,\delta/|\mathcal{S}|}(s^n) \subset \mathcal{T}^n_{\bar{X},\delta}$. Then

$$\begin{aligned}
&\left|\frac{1}{n}N(s, a|s^n, x^n) - P_{\bar{S}_q|\bar{X}}(s|a)\frac{1}{n}N(a|x^n)\right| \\
&\quad = \left|\frac{1}{n}N(s, a|s^n, x^n) - \frac{1}{n}N(s|s^n)\frac{1}{n}N(a|x^n)\right| \\
&\quad \leq \left|\frac{1}{n}N(s, a|s^n, x^n) - P_{\bar{X}|\bar{S}_q}(a|s)\frac{1}{n}N(s|s^n)\right| \\
&\qquad + \frac{1}{n}N(s|s^n)\left|P_{\bar{X}}(a) - \frac{1}{n}N(a|x^n)\right| \\
&\quad \leq \frac{\delta}{|\mathcal{S}|} + \delta \leq 2\delta.
\end{aligned}$$

This proves (79). For n large, we can use this to continue with

$$\begin{aligned}
&\mathbb{P}[s^n \in \mathcal{T}^n_{\bar{S}_q|\bar{X},2\delta}(X_{11})] \\
&\quad \overset{(i)}{\geq} \mathbb{P}[\mathcal{T}^n_{\bar{X}|\bar{S}_q,\delta/|\mathcal{S}|}(s^n)] \\
&\quad = \sum_{x^n \in \mathcal{T}^n_{\bar{X}|\bar{S}_q,\delta/|\mathcal{S}|}(s^n)} p'(x^n) \\
&\quad \overset{(ii)}{\geq} \sum_{x^n \in \mathcal{T}^n_{\bar{X}|\bar{S}_q,\delta/|\mathcal{S}|}(s^n)} P^n_{\bar{X}}(x^n) \\
&\quad = P^n_{\bar{X}|\bar{S}_q}(\mathcal{T}^n_{\bar{X}|\bar{S}_q,\delta/|\mathcal{S}|}(s^n)|s^n) \\
&\quad \overset{(iii)}{\geq} 1 - 2^{-nc'\delta^2},
\end{aligned}$$

where we used (79) in (i), $\mathcal{T}^n_{\bar{X}|\bar{S}_q,\delta/|\mathcal{S}|}(s^n) \subset \mathcal{T}^n_{\bar{X},\delta}$ in (ii) and Lemma 20 in (iii). ■

Moving to the proof of Lemma 26, let $j \in \mathcal{J}_n$. The i.i.d. random variables $\mathbb{1}_{\mathcal{T}^n_{\bar{S}_q|\bar{X},2\delta}(X_{jl})}(s^n)$ ($l \in \mathcal{L}_n$) are upper-bounded by 1. Their expectation ν was lower-bounded in Lemma 29 by $1-2^{-nc'\delta^2}$. This and $1-\varepsilon_n-2^{-nc'\delta} \leq (1-\varepsilon_n)(1-2^{-nc'\delta})$ imply that $\iota_2(j, s^n)^c$ is contained in the event

$$\left\{\frac{1}{L_n}|\{l \in \mathcal{L}_n : s^n \in \mathcal{T}^n_{\bar{S}_q|\bar{X},2\delta}(X_{jl})\}| \leq (1-\varepsilon_n)\nu\right\}.$$

By Lemma 28,

$$\begin{aligned}
&\mathbb{P}\left\{\frac{1}{L_n}|\{l \in \mathcal{L}_n : s^n \in \mathcal{T}^n_{\bar{S}_q|\bar{X},2\delta}(X_{jl})\}| \leq (1-\varepsilon_n)\nu\right\} \\
&\quad \leq 2\exp\left(-L_n \cdot \frac{\varepsilon_n^2 \nu}{3}\right).
\end{aligned} \tag{80}$$

For the exponent on the right-hand side of (80), one obtains

$$\begin{aligned}
&-L_n \cdot \frac{\varepsilon_n^2 \nu}{3} \\
&\quad \overset{(i)}{=} -\left\lfloor \exp\left\{n\left(\max_{q\in\mathcal{P}(S)} I(\bar{X} \wedge \bar{Z}_q) + \frac{\tau}{4}\right)\right\}\right\rfloor \frac{2^{-2\alpha n}\nu}{3} \\
&\quad \overset{(ii)}{\leq} -\exp\left\{n\left(\max_{q\in\mathcal{P}(S)} I(\bar{X} \wedge \bar{Z}_q) + \frac{\tau}{5}\right)\right\} \\
&\qquad \times \frac{2^{-2\alpha n}(1 - 2^{-nc'\delta^2})}{3} \\
&\quad \leq -\exp\left\{n\left(\max_{q\in\mathcal{P}(S)} I(\bar{X} \wedge \bar{Z}_q) + \frac{\tau}{5} - 3\alpha\right)\right\}.
\end{aligned} \tag{81}$$

Here, (i) follows from the definitions of L_n and ε_n. In (ii), the bound $\nu \geq 1 - 2^{-nc'\delta^2}$ from Lemma 29 was used again. If α is so small that $\tau_5 := \tau/5 - 3\alpha > 0$, the right-hand side of (81) tends to $-\infty$ at exponential speed. Thus the right-hand side of (80) tends to 0 at doubly-exponential speed, and thus also $\mathbb{P}[\iota_2(j, s^n)^c]$. This completes the proof of Lemma 26.

References

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

[2] R. Ahlswede, "Arbitrarily varying channels with states sequence known to the sender," *IEEE Trans. Inf. Theory*, vol. IT-32, no. 5, pp. 621–629, Sep. 1986.

[3] 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.

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

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

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

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

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

[9] 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.

[10] H. Boche and R. F. Schaefer, "Arbitrarily varying wiretap channels with finite coordination resources," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC)*, Jun. 2014, pp. 746–751.

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

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

[13] I. Devetak, "The private classical capacity and quantum capacity of a quantum channel," *IEEE Trans. Inf. Theory*, vol. 51, no. 1, pp. 44–55, Jan. 2005.

[14] D. D. Dubhashi and A. Panconesi, *Concentration of Measure for the Analysis of Randomized Algorithms*. Cambridge, U.K.: Cambridge Univ. Press, 2012.

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

[16] M. Fekete, "Über die Verteilung der Wurzeln bei gewissen algebraischen Gleichungen mit ganzzahligen Koeffizienten," *Mathematische Zeitschrift*, vol. 17, no. 1, pp. 228–249, 1923.

[17] Z. Goldfeld, P. Cuff, and H. Permuter. (Jan. 2016). Arbitrarily varying wiretap channels with type constrained states. [Online]. Available: http://arxiv.org/abs/1601.03660

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

[19] X. He, A. Khisti, and A. Yener, "MIMO multiple access channel with an arbitrarily varying eavesdropper: Secrecy degrees of freedom," *IEEE Trans. Inf. Theory*, vol. 59, no. 8, pp. 4733–4745, Aug. 2013.

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

[21] 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.

[22] J. Nötzel, M. Wiese, and H. Boche, "The arbitrarily varying wiretap channel—Secret randomness, stability and super-activation," to appear *IEEE Trans. Inf. Theory*, Jan. 2015. [Online]. Available: http://arxiv.org/abs/1501.07439

[23] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 379–423, Jul./Oct. 1948.

[24] M. Wiese and H. Boche, "Strong secrecy for multiple access channels," in *Information Theory, Combinatorics, and Search Theory* (Lecture Notes in Computer Science), vol. 7777, H. Aydinian, F. Cicalese, and C. Deppe, Eds. Berlin, Germany: Springer, 2013, pp. 71–122.

Moritz Wiese (S'09–M'15) received the Dipl.-Math. degree in mathematics from the University of Bonn, Germany, in 2007. He obtained the PhD degree from Technische Universität München, Munich, Germany in 2013. From 2007 to 2010, he was a research assistant at Technische Universität Berlin, Germany, and from 2010 until 2014 at Technische Universität München. Since 2014 he has been with the ACCESS Linnaeus Center at KTH Royal Institute of Technology, Stockholm, Sweden.

Janis Nötzel received the Dipl. Phys. degree in physics from the Technische Universität Berlin, Germany, in 2007 and the PhD degree from Technische Universität München, Germany, in 2012. From 2008 to 2010 he was a research assistant at the Technische Universität Berlin, Germany, and from 2011 until 2015 at Technische Universität München. Since July 2015 he is a DFG research fellow at Universitat Autònoma de Barcelona, Spain.

Holger Boche (M'04–SM'07–F'11) received the Dipl.-Ing. and Dr.-Ing. degrees in electrical engineering from the Technische Universität Dresden, Dresden, Germany, in 1990 and 1994, respectively. He graduated in mathematics from the Technische Universität Dresden in 1992. From 1994 to 1997, he did Postgraduate studies in mathematics at the Friedrich-Schiller Universität Jena, Jena, Germany. He received his Dr. rer. nat. degree in pure mathematics from the Technische Universität Berlin, Berlin, Germany, in 1998. In 1997, he joined the Heinrich-Hertz-Institut (HHI) für Nachrichtentechnik Berlin, Berlin, Germany. Starting in 2002, he was a Full Professor for mobile communication networks with the Institute for Communications Systems, Technische Universität Berlin. In 2003, he became Director of the Fraunhofer German-Sino Lab for Mobile Communications, Berlin, Germany, and in 2004 he became the Director of the Fraunhofer Institute for Telecommunications (HHI), Berlin, Germany. Since October 2010 he has been with the Institute of Theoretical Information Technology and Full Professor at the Technische Universität München, Munich, Germany. Since 2014 he has been a member and honorary fellow of the TUM Institute for Advanced Study, Munich, Germany. He was a Visiting Professor with the ETH Zurich, Zurich, Switzerland, during the 2004 and 2006 Winter terms, and with KTH Stockholm, Stockholm, Sweden, during the 2005 Summer term. Prof. Boche is a Member of IEEE Signal Processing Society SPCOM and SPTM Technical Committee. He was elected a Member of the German Academy of Sciences (Leopoldina) in 2008 and of the Berlin Brandenburg Academy of Sciences and Humanities in 2009. He received 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 co-recipient of the 2006 IEEE Signal Processing Society Best Paper Award and recipient of the 2007 IEEE Signal Processing Society Best Paper Award. He is the General Chair of the Symposium on Information Theoretic Approaches to Security and Privacy at IEEE GlobalSIP 2016. Among his publications is the recent book Information Theoretic Security and Privacy of Information Systems (Cambridge University Press).