The Arbitrarily Varying Wiretap Channel – Secret Randomness, Stability and Super-Activation

J. Nötzel
Lehrstuhl für Theoretische Informationstechnik,
Technische Universität München
Email: janis.noetzel@tum.de

M. Wiese
ACCESS Linnaeus Center,
KTH Royal Institute of Technology,
Stockholm, Sweden
Email:moritzw@kth.se

H. Boche
Lehrstuhl für Theoretische Informationstechnik,
Technische Universität München
Email: boche@tum.de

***Abstract*—We study the arbitrarily varying wiretap channel (AVWC) under average error criterion when external common randomness (CR) can be used between the legitimate parties. We consider three scenarios: In the first one the CR is known to the eavesdropper, in the second it is not known to her and in the third there is no CR available. For the second scenario, we prove a complete coding theorem. For the third scenario it is known that the capacity function is discontinuous. We prove that it is nonetheless *stable* in the sense of being continuous around its positivity points. We characterize the points of discontinuity in terms of continuous functions. We then give a complete characterization of those pairs of AVWCs whose capacity can be super-activated in the unassisted third case - in terms of the capacity function describing the first case.**

I. Introduction

The AVWC is an information-theoretic model on the intersection between the two areas of secrecy and robust communication in information theory. In this model, a sender (Alice) would like to send messages to a legitimate receiver (Bob) over a noisy channel. Involved into the scenario are two other parties: a jammer (James) who can actively influence the channel and a second but illegitimate receiver (Eve). Alice's and Bob's goal is to achieve both reliable and secure communication: No matter what the input of James is, Bob should be able to decode Alice's messages with high probability (with respect to the average error criterion) while the mutual information between the messages and Eve's output should be close to zero. This secrecy criterion is known as strong secrecy.

We add the option of Alice and Bob having access to perfect copies of the outcomes of a random experiment $\mathcal{G}$ (a source of common randomness). In [17] we considered the case where Eve gets an exact copy of the outcomes received by Alice and Bob, whereas in this work we extend our study to the case where Eve remains completely ignorant.

The only party which has no access to $\mathcal{G}$ in all the scenarios we study is James. We label the capacities which we derive from the two scenarios the "CR assisted capacity" if Eve has information about $\mathcal{G}$ and "secret CR assisted capacity" if Eve has no information about it.

We use the label C_d for the capacity when no CR is available and C_r if CR may be used. The capacity when secret CR is available is labelled C_s. Throughout we assume that Eve cannot communicate to James. We study these models in an asymptotic scenario by letting the number n of channel uses go to infinity. The probabilistic law that governs the transmission over n channel uses is

$$w^{\otimes n}(y^n|x^n,s^n)v^{\otimes n}(z^n|x^n,s^n)=\prod_{i=1}^{n}w(y_i|x_i,s_i)v(z_i|x_i,s_i).$$

Here, $s^n=(s_1,\dots,s_n)$ are the inputs of James, $x^n=(x_1,\dots,x_n)$ those of Alice and $z^n=(z_1,\dots,z_n)$ the outputs of Eve, while $y^n=(y_1,\dots,y_n)$ are received by Bob. All letters are assumed to be taken from finite alphabets $\mathcal{S}$, $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{Z}$. The action of the channel is thus completely described by the pair $(\mathfrak{W},\mathfrak{V})=((w(\cdot|\cdot,s))_{s\in\mathcal{S}},(v(\cdot|\cdot,s))_{s\in\mathcal{S}})$ of collections of probability distributions at the receiver.

This model has two important restrictions: The case where $\mathfrak{V}$ does not convey any information about either one of its inputs is the arbitrarily varying channel (AVC). It fits into our framework by setting $\mathfrak{V}=\mathfrak{T}:=\{t\}$, where $t(z|x,s):=\frac{1}{|\mathcal{Z}|}$ for all z, x and s. This model has been introduced in [5].

In [2], it was proven that the deterministic capacity of an AVC (under average error probability criterion) is either zero or equals its random coding capacity. A formula for the latter was given. In our previous work [17] we found that this dichotomic behaviour extends to C_d and C_r.

It was proven in [11] that the deterministic message transmission capacity under average error of the AVC $\mathfrak{W}$ being zero is equivalent to $\mathfrak{W}$ being *symmetrizable*: There is a set $(u(\cdot|x))_{x\in\mathcal{X}}$ of probability distributions on $\mathcal{S}$ such that for every $x,x'\in\mathcal{X}$ we have

$$\sum_{s\in\mathcal{S}}u(s|x)w(y|x',s)=\sum_{s\in\mathcal{S}}u(s|x')w(y|x,s). \tag{1}$$

In this work, we prove the following: If $\mathfrak{W}$ is non-symmetrizable, then $C_d(\mathfrak{W},\mathfrak{V})=C_r(\mathfrak{W},\mathfrak{V})$ for all possible $\mathfrak{V}$. We do not attempt to give a necessary and sufficient condition for C_d to be positive - such a characterization is not even known for the usual wiretap channel. This model has been introduced in [18] and extended in [10]. The performance in the presence of a secret key (secret CR) was studied in [13]. We extend the latter work to AVWCs. In recent years there has been a growing interest in models which combine insufficient channel state information with secrecy requirements. Probably the earliest publications are [15] and [6]. Later important

developments are [3] and [4], among the recent ones is [12]. A surprising result that was discovered recently in [8] is that of super-activation of AVWCs. This effect was until then only known for information transmission capacities in quantum information theory.

The work [8] gave an explicit example of super-activation, but a deeper understanding of the effect was not achieved. We give a complete characterization of the effect in terms of C_r. In [17] it was proven that C_r is a continuous quantity, and while the statement may seem trivial at first sight, it becomes highly nontrivial when the following are taken into account: There is at least no obvious way to deduce this statement directly from the definition of capacity, without first proving a coding result. The latter route was taken in [17] where it was proven and stated in Theorem 8 that

$$C_r(\mathfrak{W},\mathfrak{V}) = \lim_{n\to\infty} \frac{1}{n} \max_{p\in\mathcal{P}(\mathcal{U}_n)} \max_{U\in C(\mathcal{U}_n,\mathcal{X}^n)} \left(\min_{q\in\mathcal{P}(\mathcal{S}^n)} I(p; W_q^n \circ U) - \max_{q\in\mathcal{P}(\mathcal{S}^n)} I(p; V_q^n \circ U) \right). \tag{2}$$

holds. Here, $\mathcal{U}_n$ denotes a finite set. For further notation see Section II. We are now able to go beyond this result: By utilizing an approach which is based on the work [11] we are able to prove continuity results even for C_d. As was argued in [7], the continuous dependence of the performance of a communication system on the relevant system parameters is of central importance.

In contrast to C_r, C_d is not a continuous function of the channel. This casts a flashlight on the importance of distributed resources in communication networks - in this case the use of small amounts of CR. One may now be tempted to think that the transmission of messages over AVWCs without the use of CR is an adventurous task. We prove here that such a perception is wrong: Our analysis shows that C_d is continuous around its positivity points, and we are able to give an exact characterization of the discontinuity points which relies purely on the computation of functions that are *continuous* themselves. Thus, positivity of C_d is *stable* under small variations of the system.

Concerning super-activation we give a characterization of pairs $(\mathfrak{W}_i,\mathfrak{V}_i)$ $(i=1,2)$ for which it is possible only in terms of C_r. We take the space for a short explanation here: For any two given AVWCs $(\mathfrak{W}_1,\mathfrak{V}_1)$ and $(\mathfrak{W}_2,\mathfrak{V}_2)$, we define $(\mathfrak{W}_1\otimes\mathfrak{W}_2,\mathfrak{V}_1\otimes\mathfrak{V}_2)$ to equal

$$((w_1(\cdot|\cdot,s)\otimes w_2(\cdot|\cdot,s'))_{s,s'\in\mathcal{S}}, (v_1(\cdot|\cdot,s)\otimes v_2(\cdot|\cdot,s'))_{s,s'\in\mathcal{S}}),$$

where two channels w,w' with input alphabets $\mathcal{A},\mathcal{A}'$ and output alphabets $\mathcal{B},\mathcal{B}'$ define $w\otimes w'$ with input alphabet $\mathcal{A}\times\mathcal{A}'$ and output alphabet $\mathcal{B}\times\mathcal{B}'$ via its transition probabilities $(w\otimes w')((b,b')|(a,a')) := w(b|a)w'(b'|a')$ for all possible in-and output letters. Since all state alphabets are assumed to be finite, there is no loss of generality in this definition. Then,

$$C_d(\mathfrak{W}_1\otimes\mathfrak{W}_2,\mathfrak{V}_1\otimes\mathfrak{V}_2) \geqslant C_d(\mathfrak{W}_1,\mathfrak{V}_1) + C_d(\mathfrak{W}_2,\mathfrak{V}_2)$$

follows trivially from the definition of $C^{(1)}$. In contrast, if

$$C_d(\mathfrak{W}_1\otimes\mathfrak{W}_2,\mathfrak{V}_1\otimes\mathfrak{V}_2) > C_d(\mathfrak{W}_1,\mathfrak{V}_1) + C_d(\mathfrak{W}_2,\mathfrak{V}_2)$$

holds, we speak of *super-additivity* and if even

$$C_d(\mathfrak{W}_1,\mathfrak{V}_1) = C_d(\mathfrak{W}_2,\mathfrak{V}_2) = 0, \tag{3}$$

$$C_d(\mathfrak{W}_1\otimes\mathfrak{W}_2,\mathfrak{V}_1\otimes\mathfrak{V}_2) > 0 \tag{4}$$

we speak of *super-activation*. It turns out via our Theorem 11 that the effect is connected to the super-activation of C_r, if the latter occurs.

As mentioned already, we also extend earlier research to the case where a linear amount of secret bits of CR can be used. Our restriction to positive rates of CR allows us to give an elegant formula for the corresponding capacity C_s as follows: For every $G>0$, it holds

$$C_s(\mathfrak{W},\mathfrak{V},G) = \min\{C_r(\mathfrak{W},\mathfrak{V})+G, C_r(\mathfrak{W},\mathfrak{T})\}. \tag{5}$$

This formula is generally 'hard to compute' in the sense that it requires one to calculate the limit in the formula (2) - as long as $G < C_r(\mathfrak{W},\mathfrak{T}) - C_r(\mathfrak{W},\mathfrak{V})$. If this condition is not met, then $C_s(\mathfrak{W},\mathfrak{V}) = C_r(\mathfrak{W},\mathfrak{T})$. Since the latter is the usual capacity of the AVC $\mathfrak{W}$, we conclude that if enough secret CR is available, the capacity of the system can be much more efficiently described - by a formula which does not require regularization anymore!

II. Notation and Definitions

All alphabets in this work are finite. Let $\mathcal{A}$ be a set. Its cardinality is denoted $|\mathcal{A}|$, $\mathcal{P}(\mathcal{A})$ denotes the set of probability distributions on it. The one-norm distance between $p,p'\in\mathcal{P}(\mathcal{A})$ is $\|p-p'\|_1 = \sum_{a\in\mathcal{A}}|p(a)-p'(a)|$. The expectation of $f:\mathcal{A}\to\mathbb{R}$ is written $\mathbb{E}f$ (the underlying distribution will always be clear from the context). Each $a^n\in\mathcal{A}^n$ defines an *n-type* $\bar{N}(\cdot|a^n)\in\mathcal{P}(\mathcal{A})$ via $N(a|a^n) := |\{i : a_i = a\}|$ and $\bar{N}(\cdot|a^n) := \frac{1}{n}N(\cdot|a^n)$. The set of all n-types is $\mathcal{P}_0^n(\mathcal{A})$. Each $p\in\mathcal{P}_0^n(\mathcal{A})$ defines the *typical set* $T_p := \{a^n : \bar{N}(\cdot|a^n) = p(\cdot)\}$. Given $\mathcal{A}'\subset\mathcal{A}$, the indicator function on $\mathcal{A}'$ is written $\mathbb{1}_{\mathcal{A}'}$. We set $[N] := \{1,\dots,N\}$. The set of channels mapping elements from $\mathcal{A}$ to $\mathcal{B}$ is denoted $C(\mathcal{A},\mathcal{B})$. Every channel w is uniquely represented by its set $\{w(b|a)\}_{a\in\mathcal{A},b\in\mathcal{B}}$ of transition probabilities. For $w,w'\in C(\mathcal{A},\mathcal{B})$ we set $\|w-w'\| := \max_{a\in\mathcal{A}}\|w(\cdot|a)-w'(\cdot|a)\|_1$. If $p\in\mathcal{P}(\mathcal{A})$ and $q\in\mathcal{P}(\mathcal{B})$, $p\otimes q\in\mathcal{P}(\mathcal{A}\times\mathcal{B})$ is defined by $(p\otimes q)(a,b) := p(a)q(b)$ and $p^{\otimes n}\in\mathcal{P}(\mathcal{A}^n)$ via $p^{\otimes n}(a^n) := \prod_{i=1}^n p(a_i)$. If $w\in C(\mathcal{A}\times\mathcal{B},\mathcal{C})$ and $p\in\mathcal{P}(\mathcal{B})$ then $w_p(c|a) := \sum_b p(b)w(c|a,b)$.

The Shannon entropy of $p\in\mathcal{P}(\mathcal{A})$ is denoted by $H(p)$, the relative entropy between two probability distributions $p,q\in\mathcal{P}(\mathcal{A})$ by $D(p\|q)$. Both exp and log are defined with respect to base 2. Every $p\in\mathcal{P}(\mathcal{A})$ and $w\in C(\mathcal{A},\mathcal{B})$ defines a random variable (A,B) via $\mathbb{P}((A,B)=(a,b)) = p(a)w(b|a)$ $\forall\ a\in\mathcal{A},\ b\in\mathcal{B}$). We set $I(p;w) := I(A;B)$. Throughout, $o:[0,1/2]\to\mathbb{R}_+$ is a symbol for any function satisfying $\lim_{x\to0}o(x)=0$. We now define codes, rates and capacities. Throughout, $(\mathfrak{W},\mathfrak{V})$ denotes an AVWC.

Definition 1. *A CR assisted code $\mathcal{K}_n$ for $(\mathfrak{W},\mathfrak{V})$ consists of: two natural numbers K and Γ, a set of encoders $\{E^\gamma\}_{\gamma\in\Gamma}\subset C([K],\mathcal{X}^n)$ and a collection $(D_k^\gamma)_{k,\gamma=1}^{K,\Gamma}$ of subsets D_k^γ of $\mathcal{Y}^n$ satisfying $D_k^\gamma\cap D_{k'}^\gamma=\varnothing$ for all $\gamma\in[\Gamma]$, whenever $k\neq k'$. Every such code defines the random variables $S_{s^n}:=(\mathfrak{K}_n,\mathfrak{K}'_n,\mathfrak{d}_n,\mathfrak{X}_n,\mathfrak{Y}_{s^n},\mathfrak{Z}_{s^n})$ $(s^n\in\mathcal{S}^n)$ via*

$$\mathbb{P}(S_{s^n}=(k,k',\gamma,x^n,y^n,z^n))$$
$$:=\frac{1}{\Gamma\cdot K}E^\gamma(x^n|k)\mathbb{1}_{D_{k'}^\gamma}(y^n)w^{\otimes n}(y^n|s^n,x^n)v^{\otimes n}(z^n|s^n,x^n).$$

The average error of $\mathcal{K}_n$ is

$$e(\mathcal{K}_n)=1-\min_{s^n\in\mathcal{S}^n}\frac{1}{K\Gamma}\sum_{k,\gamma=1}^{K,\Gamma}\sum_{x^n}E^\gamma(x^n|k)w^{\otimes n}(D_k^\gamma|s^n,x^n).$$

For every s^n, the average success probability is

$$d_{s^n}(\mathcal{K}_n)=\frac{1}{K\Gamma}\sum_{k,\gamma=1}^{K,\Gamma}\sum_{x^n}E^\gamma(x^n|k)w^{\otimes n}(D_k^\gamma|s^n,x^n).$$

Definition 2. *A secure CR assisted coding scheme for $(\mathfrak{W},\mathfrak{V})$ operating at rate R consists of a sequence $(\mathcal{K}_n)_{n\in\mathbb{N}}$ of CR assisted codes such that*

$$\lim_{n\to\infty}e(\mathcal{K}_n)=0,\qquad\liminf_{n\to\infty}\frac{1}{n}\log(K_n)=R,$$
$$\text{and}\qquad\limsup_{n\to\infty}\max_{s^n\in\mathcal{S}^n}I(\mathfrak{K}_n;\mathfrak{Z}_{s^n}|\mathfrak{d}_n)=0.$$

If $\Gamma_n=1$ for all $n\in\mathbb{N}$, $(\mathcal{K}_n)_{n\in\mathbb{N}}$ is called deterministic.

Definition 3. *A secure coding scheme $\mathcal{K}$ for $(\mathfrak{W},\mathfrak{V})$ operating at rate R and using an amount $G_{\mathfrak{K}}>0$ of secret CR consists of a sequence $(\mathcal{K}_n)_{n\in\mathbb{N}}$ of CR assisted codes satisfying*

$$\lim_{n\to\infty}\frac{1}{n}\log\Gamma_n=G_{\mathfrak{K}},\qquad\liminf_{n\to\infty}\frac{1}{n}\log(K_n)=R,$$
$$\lim_{n\to\infty}e(\mathcal{K}_n)=0,\qquad\limsup_{n\to\infty}\max_{s^n\in\mathcal{S}^n}I(\mathfrak{K}_n;\mathfrak{Z}_{s^n})=0.$$

Remark 4. *The amount of CR is only quantified when it is kept secret from Eve. The reason for this becomes clear from [17], where it is proven that already a logarithmic number of bits of CR is sufficient to achieve the full capacity C_r.*

Definition 5. *Let $(\mathfrak{W},\mathfrak{V})$ be an AVWC. Let $G>0$. $C_s(\mathfrak{W},\mathfrak{V},G)$ is the supremum over all $R\geqslant 0$ such that there is a secure coding scheme $\mathcal{K}$ for $(\mathfrak{W},\mathfrak{V})$ operating at rate R and using an amount G of secret CR.*
Further, $C_d(\mathfrak{W},\mathfrak{V})$ is the supremum over all $R\geqslant 0$ such that there is a secure deterministic coding scheme $\mathcal{K}$ at rate R.
Finally, $C_r(\mathfrak{W},\mathfrak{V})$ is the supremum over all $R\geqslant 0$ such that there exists a secure CR assisted coding scheme $\mathcal{K}$ at rate R.

Let M_{f} be the set of all finite sets of elements of $C(\mathcal{X},\mathcal{Y})$. Define $F:M_{\mathrm{f}}\to\mathbb{R}_+$ by setting, for each $\mathfrak{W}=(w(\cdot|\cdot,s))_{s\in\mathcal{S}}$,

$$F(\mathfrak{W}):=\max_u\min_{x\neq x'}\|\sum_{s\in\mathcal{S}}(u(s|x)w(\cdot|s,\hat{x})-u(s|\hat{x})w(\cdot|s,x))\|_1$$

Then '$F(\mathfrak{W})=0$' is equivalent to 'the AVC $\mathfrak{W}$ is symmetrizable'. As a metric on the set of AVWCs we use the Hausdorff-distance, defined by setting for AVCs $\mathfrak{W},\mathfrak{W}'$, $g(\mathfrak{W},\mathfrak{W}'):=\max_{s\in\mathcal{S}}\min_{\hat{s}\in\hat{\mathcal{S}}}\|w(\cdot|s,\cdot)-w'(\cdot|\hat{s},\cdot)\|$ and for AVWCs $(\mathfrak{W},\mathfrak{V})$, $(\mathfrak{W}',\mathfrak{V}')$ $d((\mathfrak{W},\mathfrak{V}),(\mathfrak{W}',\mathfrak{V}')):=$

$$\max\{g(\mathfrak{W},\mathfrak{W}'),g(\mathfrak{W}',\mathfrak{W}),g(\mathfrak{V},\mathfrak{V}'),g(\mathfrak{V}',\mathfrak{V})\}.$$

III. Main Results

Throughout this section, let $(\mathfrak{W},\mathfrak{V})$ denote an AVWC.

Theorem 6. *With $\mathfrak{T}$ as defined in the introduction, it holds*

$$C_s(\mathfrak{W},\mathfrak{V},G)=\min\{C_r(\mathfrak{W},\mathfrak{V})+G,C_r(\mathfrak{W},\mathfrak{T})\}.\quad(6)$$

Corollary 7. *For every $G>0$, the function $(\mathfrak{W},\mathfrak{V})\mapsto C_s(\mathfrak{W},\mathfrak{V},G)$ is continuous.*

Theorem 8 (Symmetrizability properties of C_d)**.**
1) *If $\mathfrak{W}$ is symmetrizable, then $C_d(\mathfrak{W},\mathfrak{V})=0$.*
2) *If $\mathfrak{W}$ is non-symmetrizable, then $C_d(\mathfrak{W},\mathfrak{V})=C_r(\mathfrak{W},\mathfrak{V})$.*

Theorem 9 (Stability of C_d)**.** *If $(\mathfrak{W},\mathfrak{V})$ satisfies $C_d(\mathfrak{W},\mathfrak{V})>0$ then there is $\epsilon>0$ such that for all $(\mathfrak{W}',\mathfrak{V}')$ with $d((\mathfrak{W},\mathfrak{V}),(\mathfrak{W}',\mathfrak{V}'))\leqslant\epsilon$ we have $C_d(\mathfrak{W}',\mathfrak{V}')>0$.*

Theorem 10 (Discontinuity properties of C_d)**.**
1) *C_d is discontinuous in the point $(\mathfrak{W},\mathfrak{V})$ if and only if it holds: $C_r(\mathfrak{W},\mathfrak{V})>0$ and $F(\mathfrak{W})=0$ but for all $\epsilon>0$ there is $\mathfrak{W}_\epsilon$ such that $d(\mathfrak{W},\mathfrak{W}_\epsilon)<\epsilon$ and $F(\mathfrak{W}_\epsilon)>0$.*
2) *If C_d is discontinuous in the point $(\mathfrak{W},\mathfrak{V})$ then it is discontinuous for all $\hat{\mathfrak{V}}$ for which $C_r(\mathfrak{W},\hat{\mathfrak{V}})>0$.*

It has been proven by explicit example in [9], Section V, that the function $(\mathfrak{W},\mathfrak{V})\mapsto C_d(\mathfrak{W},\mathfrak{V})$ can have discontinuity points.
Our next result is probably the most interesting in this work, since it sheds additional light on a new phenomenon: the super-activation of 'the' secrecy capacity of AVWCs:

Theorem 11 (Characterization of super-activation of C_d)**.** *Let $(\mathfrak{W}_i,\mathfrak{V}_i)_{i=1,2}$ be AVWCs.*
1) *If $C_d(\mathfrak{W}_1,\mathfrak{V}_1)=C_d(\mathfrak{W}_2,\mathfrak{V}_2)=0$, then $C_d(\mathfrak{W}_1\otimes\mathfrak{W}_2,\mathfrak{V}_1\otimes\mathfrak{V}_2)>0$ iff $\mathfrak{W}_1\otimes\mathfrak{W}_2$ is not symmetrizable and $C_r(\mathfrak{W}_1\otimes\mathfrak{W}_2,\mathfrak{V}_1\otimes\mathfrak{V}_2)>0$.*
2) *There exist AVWCs which exhibit the above behaviour.*
3) *If C_r shows super-activation for $(\mathfrak{W}_1,\mathfrak{V}_1)$ and $(\mathfrak{W}_2,\mathfrak{V}_2)$, then C_d shows super-activation for $(\mathfrak{W}_1,\mathfrak{V}_1)$ and $(\mathfrak{W}_2,\mathfrak{V}_2)$ if and only if at least one of $\mathfrak{W}_1$ or $\mathfrak{W}_2$ is non-symmetrizable.*
4) *If C_r shows no super-activation for $(\mathfrak{W}_1,\mathfrak{V}_1)$ and $(\mathfrak{W}_2,\mathfrak{V}_2)$ then super-activation of C_d can only happen if $\mathfrak{W}_1$ is non-symmetrizable and $\mathfrak{W}_2$ is symmetrizable and $C_r(\mathfrak{W}_1,\mathfrak{V}_1)=0$ and $C_r(\mathfrak{W}_2,\mathfrak{V}_2)>0$. The statement is independent of the specific labelling.*

While Theorem 11 offers a complete characterization, it does not give any explicit examples - fortunately an example has already been given in [8]. Theorem 11 now allows the systematic construction of such examples.

A. Technical definitions and facts

An important part of our results concerning $C_{\mathrm{d}}^{(1)}$ builds on the mathematical structure that was developed in [11].

The codes developed there are based on i.i.d. sampling of codewords which are all taken from one and the same set T_p. We use the same distribution for sampling our codewords. This ensures seamless connectivity to the results of [11].

Our notion of typicality is as follows: for arbitrary finite sets $\mathcal{A}, \mathcal{B}$, every $p \in \mathcal{P}(\mathcal{A})$, $\tilde{v} \in C(\mathcal{A}, \mathcal{B})$, $\delta > 0$ and $a^n \in \mathcal{A}^n$ we define $p_{AB} \in \mathcal{P}(\mathcal{A} \times \mathcal{B})$ via $p_{AB}(a,b) := \bar{N}(a|a^n)\tilde{v}(b|a)$ and

$$T^n_{p,\delta} := \{a^n \in \mathcal{A}^n : D(\bar{N}(\cdot|a^n)\|p) \leqslant \delta\}, \quad (7)$$

$$T_{\tilde{v},\delta}(a^n) := \{b^n : D(\bar{N}(\cdot|a^n, b^n)\|p_{AB}) \leqslant \delta\}. \quad (8)$$

We also set, for every $p \in \mathcal{P}(\mathcal{X})$ and AVWC $(\mathfrak{W}, \mathfrak{V})$

$$E(p) := \max_{q \in \mathcal{P}(\mathcal{S})} I(p; v_q) \text{ and } B(p) := \min_{q \in \mathcal{P}(\mathcal{S})} I(p; w_q). \quad (9)$$

Our most important tool is the Chernoff-Hoeffding bound:

Lemma 12. *Let $b > 0$, $\varepsilon \in [0, 1/2]$ and $Z_1, \ldots, Z_L$ be i.i.d. random variables with values in $[0, b]$ and $\nu := \mathbb{E}(Z_1)$. Then*

$$\mathbb{P}\left\{\frac{1}{L}\sum_{l=1}^{L} Z_i \notin [(1 \pm \varepsilon)\nu]\right\} \leqslant 2\exp\left(-L \cdot \frac{\varepsilon^2 \cdot \nu}{3 \cdot b}\right), \quad (10)$$

where $[(1 \pm \varepsilon)\nu]$ denotes the interval $[(1-\varepsilon)\nu, (1+\varepsilon)\nu]$.

IV. Proofs

We only give brief sketches of the main ideas behind our proofs here. The full proofs may be found in the accompanying paper [16].

Sketch of proof for the converse part of Theorem 6: Let $(\mathcal{K}_n)_{n\in\mathbb{N}}$ be a secure coding scheme which is assisted by secret CR and operating at rate $R \geqslant 0$ and satisfying $\limsup_{n\to\infty} \frac{1}{n}\log \Gamma_n = G \geqslant 0$. We have $H(\mathfrak{K}_n|\mathfrak{K}'_n, \mathfrak{b}_n) \approx 0$ from Fano's inequality and $\max_{s^n} I(\mathfrak{K}_n; \mathfrak{Z}_{s^n}) \approx 0$ by assumption. Thus $\log R \approx I(\mathfrak{K}_n; \mathfrak{K}'_n|\mathfrak{b}_n) - I(\mathfrak{K}_n; \mathfrak{Z}_{s^n})$. Since $H(\mathfrak{K}_n|\mathfrak{Z}_{s^n}) \leqslant H(\mathfrak{K}_n|\mathfrak{Z}_{s^n}, \mathfrak{b}_n) + H(\mathfrak{b}_n)$ we get

$$\forall s^n : \ nR \approx I(\mathfrak{K}_n; \mathfrak{K}'_n|\mathfrak{b}_n) - I(\mathfrak{K}_n; \mathfrak{Z}_{s^n}|\mathfrak{b}_n) + n \cdot G, \quad (11)$$

In [17] it was proven that there is (asymptotically) no loss in replacing $\mathfrak{b}_n$ with a $\mathfrak{b}'_n$ that has at most polynomial support. This and an application of the data processing inequality yields

$$\forall s^n \in \mathcal{S}^n, q \in \mathcal{P}(\mathcal{S}^n) : \ nR \approx I(\mathfrak{K}_n; \mathfrak{Z}^n_q) - I(\mathfrak{K}_n; \mathfrak{Z}_{s^n}) + nG.$$

This estimate allows us to derive the multi-letter converse. A second (single letter) upper bound arises if one forgets about the secrecy requirements. Taking the minimum over both upper bounds establishes the converse. ∎

An intermediate result:

For the direct part of Theorem 6 and statement 2) in Theorem 8 we make a random selection of codewords. Let $p \in \mathcal{P}^n_0(\mathcal{A})$, $q \in \mathcal{P}^n_0(\mathcal{S})$. Throughout, we will express asymptotic quantities as functions of the random variables (S, X, Z) defined via $\mathbb{P}((S, X, Z) = (s, x, z)) := p(x)q(s)v(z|x, s)$. The distribution of (S, X, Z) is labelled p_{SXZ}.

We define the i.i.d. random variables $\mathbf{X}_{kl\gamma}$ by setting $\mathbb{P}(\mathbf{X}_{kl\gamma} = x^n) := \frac{1}{|T_p|}\mathbb{1}_{T_p}(x^n)$ for all $k \in [K]$, $l \in [L]$ $\gamma \in [\Gamma]$, $x^n \in \mathcal{X}^n$, where $K, L, \Gamma \in \mathbb{N}$. Their realizations are written $\mathbf{x}_{kl\gamma}$. We write $\mathbf{x}$ for the realizations of the variable $\mathbf{X} = (\mathbf{X}_{kl\gamma})^{K,L,\Gamma}_{k,l,\gamma=1}$. An additional random variable X^n is distributed according to $\mathbb{P}(X^n = x^n) = \frac{1}{|T_p|}\mathbb{1}_{T_p}(x^n)$ as well. For any $\delta > 0$, $s^n \in \mathcal{S}^n$ and $z^n \in \mathcal{Z}^n$ we define $\Theta_{s^n,z^n} : \mathcal{X}^n \to [0, b]$ (where $b = 2^{-n(H(Z|X,S)-o(\delta))}$) by

$$M_{s^n,z^n} := \{x^n \in T_p : D(\bar{N}(\cdot|s^n, x^n, z^n)\|p_{SXZ}) \leqslant \delta\} \quad (12)$$

$$\Theta_{s^n,z^n}(x^n) := V^{\otimes n}(z^n|s^n, x^n)\mathbb{1}_{M(s^n,z^n)}(x^n). \quad (13)$$

The coding theorem for C_s needs a definition of decoder: For every $n \in \mathbb{N}$, set $\Xi_n := \mathcal{P}^n_0(\mathcal{S})$. For every x^n, define $\hat{D}_{x^n} := \bigcup_{\xi\in\Xi} T_{W_\xi,\delta}(x^n)$ and for a collection $\mathbf{x}_\gamma := (\mathbf{x}_{kl\gamma})^{K,L}_{k,l=1}$ of codewords set

$$D(\mathbf{x}_\gamma)_{kl} := \hat{D}_{\mathbf{x}_{kl\gamma}} \bigcap \Big(\bigcup_{k'\neq k}\bigcup_{l'\neq l} \hat{D}_{\mathbf{x}_{k'l'\gamma}}\Big)^{\complement}. \quad (14)$$

This defines the code $\mathcal{K}(\mathbf{x}_\gamma)$. We define the following events:

$$E_1 := \left\{\mathbf{x} | \forall k, s^n, z^n : \frac{1}{L\Gamma}\sum_{j=1}^{L,\Gamma} \Theta_{s^n,z^n}(\mathbf{x}_{kl\gamma}) \in \left[(1 \pm \epsilon)\mathbb{E}\Theta_{s^n,z^n}\right]\right\}$$

$$E_2 := \left\{\mathbf{x} : \min_{s^n} \frac{1}{\Gamma}\sum_{\gamma=1}^{\Gamma} d_{s^n}(\mathcal{K}_\gamma) \geqslant (1 - 2 \cdot 2^{-n\delta})\right\}.$$

where $\epsilon := 2^{-n\tau/4}$. We can then state that

Lemma 13. *Let $K, L, \Gamma \in \mathbb{N}$, $\tau > 0$ and $\beta > 0$. There are $\delta > 0$ and $N \in \mathbb{N}$ such that for all $n \geqslant N$ and $p \in \mathcal{P}^n_0(\mathcal{X})$:*
1) *If $\frac{1}{n}\log(L \cdot \Gamma) \geqslant E(p) + \tau$ and $\min_{x:p(x)>0} p(x) \geqslant \beta$, then $\mathbb{P}(E_1) \geqslant 1 - 2 \cdot |\mathcal{S} \times \mathcal{X} \times \mathcal{Z}|^n \cdot \exp(-2^{n\cdot\tau/6})$.*
2) *There is μ depending only on δ such that $\frac{1}{n}\log KL \leqslant B(p) - \mu$ implies $\mathbb{P}(E_2^{\complement}) \leqslant |\mathcal{S}|^n \exp(-\Gamma 2^{-n\delta})$.*
3) *It holds $\lim_{\tau\to 0}\delta(\tau) = 0$ and $\lim_{\delta\to 0}\mu(\delta) = 0$.*

Sketch of Proof: For every x^n, s^n and z^n, write

$$v^{\otimes n}(z^n|s^n, x^n) = 2^{-n(D(\bar{N}_{SXZ}\|p_{SXZ}) - D(\bar{N}_{XS}\|p_{XS}) + H(\hat{Z}|\hat{S},\hat{X}))}$$

where $\bar{N}_{SXZ}$ is the type of s^n, x^n, z^n and $\bar{N}_{SX}$ that of s^n, x^n and $\hat{S}\hat{X}\hat{Z}$ is distributed according to $\bar{N}_{SXZ}$. With $m(s^n, z^n) = 1$ if $M(s^n, z^n) \neq \varnothing$ and $m(s^n, z^n) = 0$ else it follows for all x^n, s^n and z^n

$$\Theta_{s^n,z^n}(x^n) \leqslant 2^{-n\cdot(H(Z|XS)-o(\delta))}, \quad (15)$$

$$\mathbb{E}\Theta_{s^n,z^n} \geqslant m(s^n, z^n) \cdot 2^{-n(H(Z|S)-o(\delta))}. \quad (16)$$

This implies 1) via the Chernoff bound Lemma 12. The proof of 2) is based on the use of Ahlswede's robustification technique [1] and a random choice of codewords: Associate to every $\mathbf{x}_\gamma = (\mathbf{x}_{kl\gamma})^{K,L}_{k,l=1}$ a code $\mathcal{K}(\mathbf{x}_\gamma)$ via (14). For every s^n and γ, standard arguments prove that for all large enough n it holds $\mathbb{E}d_{s^n}(\mathbf{X}_\gamma) \geqslant 1 - 2^{-n\cdot\delta/2}$. Then, 2) follows from Lemma 12. 3) is implicit in the proof of 1). ∎

Sketch of proof of the direct part of Theorem 6: Let $G > 0$ be given. Set $p := \arg\max_{p\in\mathcal{P}(\mathcal{X})}(B(p) - E(p))$, $G' := \max\{E(p), G\}$. Choose τ such that $G > \delta(\tau)$ with $\delta(\tau)$

from Lemma 13 and sequences $(K_n)_{n=1}^{\infty}$, $(L_n)_{n=1}^{\infty}$, $(\Gamma_n)_{n=1}^{\infty}$ of natural numbers such that

$$\lim_{n\to\infty} \tfrac{1}{n} \log \Gamma_n = G', \quad (17)$$

$$\lim_{n\to\infty} \tfrac{1}{n} \log L_n = E(p) - G' + 2\tau, \quad (18)$$

$$\lim_{n\to\infty} \tfrac{1}{n} \log K_n = B(p) - E(p) + G' - 2(\tau + \mu(\tau)). \quad (19)$$

Continuity of $B(\cdot) - E(\cdot)$ and Lemma 13 ensure that for all large enough n there is a type $p \in \mathcal{P}_0^n(\mathcal{X})$ and $\mathbf{x} \in E_1 \cap E_2$ such that the corresponding codes satisfy for all $s^n \in \mathcal{S}^n$:

$$\sum_{\gamma=1}^{\Gamma_n} \frac{1}{\Gamma_n} \sum_{k,l=1}^{K_n,L_n} \frac{1}{K_n L_n} w^{\otimes n}(D_{kl}^{\gamma}|s^n, \mathbf{x}_{kl\gamma}) \geqslant 1 - 2^{-n\cdot\nu(\tau)},$$

$$\max_{k\in[K_n]} \| \frac{1}{L_n\Gamma_n} \sum_{l,\gamma=1}^{L_n,\Gamma_n} v^{\otimes n}(\cdot|s^n, \mathbf{x}_{kl\gamma}) - r_{s^n} \|_1 \leqslant 2^{-n\cdot\nu(\tau)}$$

for some $\nu(\tau) > 0$ and $r_{s^n} := \mathbb{E}V_{s^n}(\cdot|X^n)$. This yields reliable communication. That we also get secure communication follows from [17, Lemma 20] via (20). Thus for each $\tau' > 0$,

$$\max_{p\in\mathcal{P}(\mathcal{X})} (\min_{q\in\mathcal{P}(\mathcal{S})} I(p; W_q) - \max_{q\in\mathcal{P}(\mathcal{S})} I(p; V_q)) + G' - \tau' \quad (20)$$

is an achievable rate. The remaining part of the proof uses standard blocking and prefixing arguments. ∎

For the proofs of Theorems 8, 9 and 10 we refer the reader to the accompanying paper [16].

Proof of Theorem 11: **1.** If $\mathfrak{W}_1 \otimes \mathfrak{W}_2$ is symmetrizable then $C_d(\mathfrak{W}_1 \otimes \mathfrak{W}_2, \mathfrak{V}_1 \otimes \mathfrak{V}_2) = 0$. If $\mathfrak{W}_1 \otimes \mathfrak{W}_2$ is not symmetrizable and $C_r(\mathfrak{W}_1 \otimes \mathfrak{W}_2, \mathfrak{V}_1 \otimes \mathfrak{V}_2) > 0$ then from Theorem 8, part 1, we know that $C_d(\mathfrak{W}_1 \otimes \mathfrak{W}_2, \mathfrak{V}_1 \otimes \mathfrak{V}_2) > 0$.
2. In [8, Section IV], an example of a pair $(\mathfrak{W}_i, \mathfrak{V}_i)_{i=1,2}$ was given where $\mathfrak{W}_1$ is symmetrizable, but $\mathfrak{W}_2$ is not. Thus $\mathfrak{W}_1 \otimes \mathfrak{W}_2$ is non-symmetrizable. Then Theorem 8, part 1, shows that $C_d(\mathfrak{W}_1 \otimes \mathfrak{W}_2, \mathfrak{V}_1 \otimes \mathfrak{V}_2) = C_r(\mathfrak{W}_1 \otimes \mathfrak{W}_2, \mathfrak{V}_1 \otimes \mathfrak{V}_2)$. In [8] it was further shown that $C_r(\mathfrak{W}_1, \mathfrak{V}_1) > 0$ and $C_d(\mathfrak{W}_2, \mathfrak{V}_2) = 0$.
3. By assumption, $C_r(\mathfrak{W}_i, \mathfrak{V}_i) = 0$ $(i = 1, 2)$ but $C_r(\mathfrak{W}_1 \otimes \mathfrak{V}_1, \mathfrak{W}_2 \otimes \mathfrak{V}_2) > 0$. The former implies $C_d(\mathfrak{W}_i, \mathfrak{V}_i) = 0$ $(i = 1, 2)$. If $\mathfrak{W}_1$ and $\mathfrak{W}_2$ were symmetrizable then $\mathfrak{W}_1 \otimes \mathfrak{W}_2$ would be symmetrizable, thus $C_d(\mathfrak{W}_1 \otimes \mathfrak{W}_2, \mathfrak{T}) = 0$, implying $C_d(\mathfrak{W}_1 \otimes \mathfrak{W}_2, \mathfrak{V}_1 \otimes \mathfrak{V}_2) = 0$. If either $\mathfrak{W}_1$ or $\mathfrak{W}_2$ are not symmetrizable then $\mathfrak{W}_1 \otimes \mathfrak{W}_2$ is not symmetrizable and this implies $C_d(\mathfrak{W}_1 \otimes \mathfrak{V}_1, \mathfrak{W}_2 \otimes \mathfrak{V}_2) = C_r(\mathfrak{W}_1 \otimes \mathfrak{V}_1, \mathfrak{W}_2 \otimes \mathfrak{V}_2) > 0$, by Theorem 8, part 1, and by assumption.
4. Let both $\mathfrak{W}_1$ and $\mathfrak{W}_2$ be symmetrizable. Then $\mathfrak{W}_1 \otimes \mathfrak{W}_2$ is symmetrizable. Since by assumption C_r shows no super-activation on the pair $(\mathfrak{W}_i, \mathfrak{V}_i)$ $(i = 1, 2)$ it follows that C_d cannot show super-activation as well. Thus at least one of the two AVCs has to be non-symmetrizable. Let w.l.o.g. this channel be $\mathfrak{W}_1$. If in addition $\mathfrak{W}_2$ would be non-symmetrizable, then $C_d(\mathfrak{W}_i, \mathfrak{V}_i) = C_r(\mathfrak{W}_i, \mathfrak{V}_i)$ would hold for $i = 1, 2$ and since $\mathfrak{W}_1 \otimes \mathfrak{W}_2$ would be non-symmetrizable as well, we would additionally have $C_d(\mathfrak{W}_1 \otimes \mathfrak{W}_2, \mathfrak{V}_1 \otimes \mathfrak{V}_2) = C_r(\mathfrak{W}_1 \otimes \mathfrak{W}_2, \mathfrak{V}_1 \otimes \mathfrak{V}_2)$. But since C_r shows no super-activation on the pair $(\mathfrak{W}_i, \mathfrak{V}_i)$ $(i = 1, 2)$ by assumption, this cannot be. Thus again w.l.o.g. $\mathfrak{W}_2$ is symmetrizable. If super-activation of C_d would take place, it would be necessary that $C_d(\mathfrak{W}_i, \mathfrak{V}_i) = 0$ $(i = 1, 2)$. But since $\mathfrak{W}_1$ is non-symmetrizable this requires that $C_r(\mathfrak{W}_1, \mathfrak{V}_1) = 0$ holds. If in addition $C_r(\mathfrak{W}_2, \mathfrak{V}_2) = 0$ would hold than C_d could not be super-activated since $C_{\mathrm{ran}}^{(1)}$ cannot be super-activated by assumption. Thus $C_{\mathrm{r}}^{(1)}(\mathfrak{W}_2, \mathfrak{V}_2) > 0$. ∎

ACKNOWLEDGEMENT

This work was supported by the DFG via grant BO 173420-1 (J.N.) and the BMBF via grant 16KIS0118 (J.N. and H.B.). H.B. thanks Dr. Plaga and Dr. Döbrich for discussions on attacks on quantum networks which motivated parts of this work. The corresponding problems for quantum channels are still open.

REFERENCES

[1] R. Ahlswede, "Coloring Hypergraphs: A New Approach to Multi-user Source Coding-II", *Journal of Combinatorics, Information & System Sciences* Vol. 5, No. 3, 220-268 (1980)
[2] R. Ahlswede, "Elimination of Correlation in Random Codes for Arbitrarily Varying Channels", *Z. Wahrscheinlichkeitstheorie verw. Gebiete* vol. 44, 159-175 (1978)
[3] I. Bjelaković, H. Boche, J. Somerfeld, "Secrecy Results for Compound Wiretap Channels", *Probl. Inf. Trans.*, vol. 49, no. 1, 73-98 (2013)
[4] I. Bjelaković, H. Boche, J. Somerfeld, "Capacity Results for Arbitrarily Varying Wiretap Channels", *LNCS* vol. 7777, 123-144 (2013)
[5] D. Blackwell, L. Breiman, A.J. Thomasian, "The capacities of certain channel classes under random coding", *Ann. Math. Stat.* 31, 558-567 (1960)
[6] M. Bloch, J.N. Laneman, "On the secrecy capacity of arbitrary wiretap channel," *Forty-Sixth Annual Allerton Conference, Allerton House, Illinois, USA* (2008)
[7] H. Boche, J. Nötzel, "Positivity, discontinuity, finite resources, and nonzero error for arbitrarily varying quantum channels", *J. Math. Phys.*, 55, 122201 (2014)
[8] H. Boche, R. Schaefer, "Capacity Results and Super-Activation for Wiretap Channels With Active Wiretappers", *IEEE Trans. Inf. Forensic Secur.*, vol. 8, no. 9, 1482-1496 (2013)
[9] H. Boche, R.F. Schaefer, H. Vincent Poor, "On the Continuity of the Secrecy Capacity of Compound and Arbitrarily Varying Wiretap Channels", *arXiv:1409.4752* (2014)
[10] I. Csiszar and J. Körner, "Broadcast channels with confidential messages", *IEEE Trans. Inf. Theory*, vol. 24, no. 3, 339-348 (1978)
[11] I. Csiszár, P. Narayan, "The Capacity of the Arbitrarily Varying Channel Revisited: Positivity, Constraints", *IEEE Trans. Inf. Theory* vol. 34, no. 2, 181-193 (1988)
[12] X. He, A. Khisti, A. Yener, "Mimo multiple access channel with an arbitrarily varying eavesdropper: Secrecy degrees of freedom", *IEEE Trans. Inf. Theory*, vol. 59, no. 8, 4733-4745 (2013)
[13] W. Kang, N. Liu, "Wiretap Channel with Shared Key", *IEEE Inf. Theory Workshop - ITW 2010 Dublin* (2010)
[14] J. Kiefer, J. Wolfowitz, "Channels with arbitrarily varying channel probability functions", *Information and Control* 5, 44-54 (1962)
[15] Y. Liang, G. Kramer, H. Poor, and S. Shamai, "Compound wiretap channels", *EURASIP Journal on Wireless Communications and Networking* (2008)
[16] J. Nötzel, M. Wiese, H. Boche, "The Arbitrarily Varying Wiretap Channel – Secret Randomness, Stability and Super-Activation", *to be on the arXiv soon under exactly this title*
[17] M. Wiese, J. Nötzel, H. Boche, "The Arbitrarily Varying Wiretap Channel – deterministic and correlated random coding capacities under the strong secrecy criterion", preprint, arXiv:1410.8078 (2014)
[18] A. Wyner, "The wire-tap channel," *The Bell System Tech. J.*, vol. 54, no. 8, pp. 1355–1387, (1975)