

# The Arbitrarily Varying Channel with Colored Gaussian Noise

Uzi Pereg<sup>1</sup> and Yossef Steinberg<sup>2</sup>

<sup>1</sup> Institute for Communications Engineering, Technical University of Munich

<sup>2</sup> Department of Electrical Engineering, Technion

Email: uzi.pereg@tum.de, ysteinbe@ee.technion.ac.il

## Abstract

We address the arbitrarily varying channel (AVC) with colored Gaussian noise. The work consists of three parts. First, we study the *general* discrete AVC with fixed parameters, where the channel depends on two state sequences, one arbitrary and the other fixed and known. This model can be viewed as a combination of the AVC and the time-varying channel. We determine both the deterministic code capacity and the random code capacity. Super-additivity is demonstrated, showing that the deterministic code capacity can be strictly larger than the weighted sum of the parametric capacities.

In the second part, we consider the arbitrarily varying Gaussian product channel (AVGPC). Hughes and Narayan characterized the random code capacity through min-max optimization leading to a “double” water filling solution. Here, we establish the deterministic code capacity and also discuss the game-theoretic meaning and the connection between double water filling and Nash equilibrium. As in the case of the standard Gaussian AVC, the deterministic code capacity is discontinuous in the input constraint, and depends on which of the input or state constraint is higher. As opposed to Shannon’s classic water filling solution, it is observed that deterministic coding using independent scalar codes is suboptimal for the AVGPC.

Finally, we establish the capacity of the AVC with colored Gaussian noise, where double water filling is performed in the frequency domain. The analysis relies on our preceding results, on the AVC with fixed parameters and the AVGPC.

## Index Terms

Arbitrarily varying channel, water filling, colored Gaussian noise, time varying channel, Gaussian product channel, deterministic code, random code.

## I. INTRODUCTION

A channel with colored Gaussian noise was first studied by Shannon [94], introducing the water filling optimal power allocation. This channel is the spectral counterpart of the Gaussian product channel (see *e.g.* [27, Section 9.5]). Those results led to useful algorithms for DSL and OFDM systems, and were generalized to multiple-input multiple output (MIMO) wireless communication systems as well (see *e.g.* [99, 38, 12, 11, 93, 41]). Furthermore, for some networks, water filling is performed in multiple stages [26, 111, 113, 114, 71, 105]. A limit formula for the capacity of the general time-varying channel (TVC) is given in [102] (see also [29, 47, 3, 33, 10, 76, 87, 112]). Another relevant setting is that of a finite-state channel, where the state evolves as a Markov chain [110, 74, 14, 73, 46, 100, 98]. In practice, there is often uncertainty regarding channel statistics, due to a variety of causes such as fading in wireless communication [95, 92, 1, 80, 42, 25, 59, 57], memory faults in storage [68, 51, 69, 66], malicious attacks on identification systems [45, 62], and cyber-physical warfare [97, 72, 104]. The arbitrarily varying channel (AVC) is an appropriate model to describe such a situation [16, 73].

Blackwell *et al.* [16] determined the random code capacity of the general AVC, *i.e.* the capacity achieved with shared randomness between the encoder and the decoder. It was also demonstrated in [16] that the random code capacity is not necessarily achievable using deterministic codes. A well-known result by Ahlswede [5] is the dichotomy property of the AVC, *i.e.* the deterministic code capacity, also referred to as ‘capacity’, either equals the random code capacity or else, it is zero. Subsequently, Ericson [37] and Csiszár and Narayan [30] have established a simple single-letter condition, namely non-symmetrizability, which is both necessary and sufficient for the capacity to be positive. Schaefer *et al.* [91] demonstrated the super-additivity phenomenon, *i.e.* when the capacity of a product of orthogonal AVCs is strictly larger than the sum of the capacities of the components. Csiszár and Narayan [31, 30] also considered the AVC when input and state constraints are imposed on the user and the jammer, respectively, due to their power limitations. Not only the constrained setting provokes serious technical difficulties analytically, but also, as shown in [30], constraints have a significant effect on the behavior of the capacity. Specifically, it is shown in [30] that dichotomy in the sense of [5] no longer holds when state constraints are imposed on the jammer. That is, the deterministic code capacity of the general AVC can be lower than the random code capacity, and yet non-zero.

The Gaussian AVC is specified by the relation  $\mathbf{Y} = \mathbf{X} + \mathbf{S} + \mathbf{Z}$ , where  $\mathbf{X}$  and  $\mathbf{Y}$  are the input and output sequences, respectively;  $\mathbf{S}$  is a state sequence of unknown joint distribution  $F_S$ , not necessarily independent nor stationary; and the noise

sequence  $\mathbf{Z}$  is i.i.d.  $\sim \mathcal{N}(0, \sigma^2)$ . The state sequence can be thought of as if generated by an adversary, or a *jammer*, who randomizes the channel states arbitrarily in an attempt to disrupt communication. It is also possible for  $\mathbf{S}$  to be a deterministic unknown state sequence. It is assumed that the user and the jammer have power limitations, and are subject to input and state constraints,  $\frac{1}{n} \sum_{i=1}^n X_i^2 \leq \Omega$  and  $\frac{1}{n} \sum_{i=1}^n S_i^2 \leq \Lambda$ , respectively, where  $n$  is the transmission length. In [60], Hughes and Narayan showed that the random code capacity is given by  $C_1^* = \frac{1}{2} \log(1 + \frac{\Omega}{\sigma^2 + \Lambda})$ . Subsequently, Csiszár and Narayan [32] showed that the deterministic code capacity is given by

$$C_1 = \begin{cases} C_1^* & \text{if } \Lambda < \Omega, \\ 0 & \text{if } \Lambda \geq \Omega. \end{cases} \quad (1)$$

It is noted in [32] that this result is *not* a straightforward consequence of the elegant Elimination Technique [5], used by Ahlswede to establish dichotomy for the AVC without constraints. Hosseinigoki and Kosut [57] determined the capacity in multiple side information scenarios for the Gaussian AVC with fast fading. Hughes and Narayan [61] determined the random code capacity of the arbitrarily varying Gaussian product channel (AVGPC), and showed that it is obtained as a “double” water filling solution to an optimization min-max problem, maximizing over input power allocation and minimizing over state power allocation. In the solution, the jammer performs water filling first, attempting to whiten the overall noise as much as possible, and then the user performs water filling taking into account the total interference power, contributed by both the channel noise and the jamming signal [61]. The Gaussian AVC is also considered in [4, 101, 70, 88, 90, 56, 59].

Extensive research has been conducted on other AVC models as well, of which we name a few. Recently, the arbitrarily varying wiretap channel has been extensively studied, as *e.g.* in [77, 17, 9, 18, 19, 78, 48, 2], including input and state constraints in [13, 64, 40]. The capacity region of the arbitrarily varying multiple access channel (MAC) with and without constraints is characterized in [85, 63, 7, 8]; capacity bounds for the arbitrarily varying broadcast channel are derived in [63, 52]; and for the arbitrarily varying relay channel in [83, 81]. Additional results on arbitrarily varying multi-user channels and constraints are derived *e.g.* in [108, 24, 50, 106, 84, 65]. Transmission of an arbitrarily varying Wyner-Ziv source over a Gel’fand-Pinsker channel is considered in [109, 107], and related problems were recently presented in [24, 22, 21]. Various Gaussian AVC networks are studied *e.g.* in [89, 49, 23, 54, 55, 82, 83, 85, 58].

In this paper, we address the AVC with colored Gaussian noise. The body of this manuscript consists of three parts, of which the first and the second can also be viewed as milestones on our path to the main result. First, we study the *general* discrete AVC with fixed parameters. This model is a combination of the TVC and the AVC, as the channel depends on two state sequences, one arbitrary and the other fixed. We determine both the deterministic code capacity and the random code capacity. Deterministic code super-additivity is demonstrated, showing that the capacity can be strictly larger than the weighted sum of the parametric capacities. In the second part of this paper, we establish the deterministic code capacity of the AVGPC, where there is *white* Gaussian noise and no parameters. We also give observations and discuss the game-theoretic interpretation of Hughes and Narayan’s random code characterization [61], and the connection between the double water filling solution and the idea of Nash equilibrium in game theory. We further examine the connection between the AVGPC and the product MAC [26, 71] (without a state), pointing out the similarities and differences between the models, results, and interpretation. As in the case of the standard Gaussian AVC, the deterministic code capacity is discontinuous in the input constraint, and depends on which of the input or state constraint is higher. As opposed to Shannon’s classic water filling solution [94], it is observed that deterministic coding using independent scalar codes is suboptimal for the AVGPC. Finally, we establish the capacity of the AVC with colored Gaussian noise, where double water filling is performed in the frequency domain.

While the results on the AVC with fixed parameters and on the AVGPC stand in their own right, they also play a key role in our proof of the main capacity theorem for the AVC with colored Gaussian noise. In the random code analysis for the AVC with fixed parameters, we modify Ahlswede’s Robustification Technique (RT) [6]. Essentially, the RT uses a reliable code for the compound channel to construct a random code for the AVC applying random permutations to the codeword symbols. A straightforward application of Ahlswede’s RT does not work here, since the user cannot apply permutations to the parameter sequence. Hence, we give a modified RT which is restricted to permutations that do not affect the parameter sequence, *i.e.* such that the parameter sequence is an eigenvector of all of our permutation matrices. The second part of the paper builds on identifying the symmetrizing jamming strategies and minimal symmetrizability costs for the AVGPC. At last, we use the results on the AVC with fixed parameters and the AVGPC in our proof of the capacity theorem for the AVC with colored Gaussian noise. By orthogonalization of the noise covariance, the AVC with colored Gaussian noise is transformed into an AVC with fixed parameters, which are determined by the spectral representation of the noise covariance matrix. This in turn yields double water-filling optimization in analogy to the AVGPC.

## II. CHANNELS WITH FIXED PARAMETERS

In this section we consider the AVC with fixed parameters. The results in this section will be used to analyze the AVC with colored Gaussian noise.

### A. Notation

We use the following notation. Calligraphic letters  $\mathcal{X}, \mathcal{S}, \mathcal{T}, \mathcal{Y}, \dots$  are used for finite sets. Lowercase letters  $x, s, t, y, \dots$  stand for constants and values of random variables, and uppercase letters  $X, S, T, Y, \dots$  stand for random variables. The distribution of a random variable  $X$  is specified by a probability mass function (pmf)  $P_X(x) = p(x)$  over a finite set  $\mathcal{X}$ . The set of all pmfs over  $\mathcal{X}$  is denoted by  $\mathcal{P}(\mathcal{X})$ . The set of all probability kernels  $p(x|t)$  is denoted by  $\mathcal{P}(\mathcal{X}|\mathcal{T})$ . We use  $x^j = (x_1, x_2, \dots, x_j)$  to denote a sequence of letters from  $\mathcal{X}$ . A random sequence  $X^n$  and its distribution  $P_{X^n}(x^n) = p(x^n)$  are defined accordingly. For a pair of integers  $i$  and  $j$ ,  $1 \leq i \leq j$ , we define the discrete interval  $[i : j] = \{i, i+1, \dots, j\}$ .

The type  $\hat{P}_{x^n}$  of a given sequence  $x^n$  is defined as the empirical distribution  $\hat{P}_{x^n}(a) = N(a|x^n)/n$  for  $a \in \mathcal{X}$ , where  $N(a|x^n)$  is the number of occurrences of the symbol  $a$  in the sequence  $x^n$ . A type class is denoted by  $\mathcal{T}^n(\hat{P}) = \{x^n : \hat{P}_{x^n} = \hat{P}\}$ . Similarly, define the joint type  $\hat{P}_{x^n, y^n}(a, b) = N(a, b|x^n, y^n)/n$  for  $a \in \mathcal{X}$ ,  $b \in \mathcal{Y}$ , where  $N(a, b|x^n, y^n)$  is the number of occurrences of the symbol pair  $(a, b)$  in the sequence  $(x_i, y_i)_{i=1}^n$ . Then, a conditional type is defined as  $\hat{P}_{x^n|y^n}(a, b) = \hat{P}_{x^n, y^n}(a, b)/\hat{P}_{y^n}(b)$ . Furthermore, we define the  $\delta$ -typical set  $\mathcal{A}_\delta^{(n)}(p)$  with respect to a distribution  $p(x)$  by

$$\mathcal{A}_\delta^{(n)}(p) \triangleq \left\{ x^n \in \mathcal{X}^n : \forall a \in \mathcal{X}, \begin{cases} |p(a) - \hat{P}_{x^n}(a)| \leq \delta \text{ if } p(a) > 0, \text{ and} \\ \hat{P}_{x^n}(a) = 0 \text{ if } p(a) = 0 \end{cases} \right\}. \quad (2)$$

The distribution of a real random variable  $Z \in \mathbb{R}$  is represented by a cumulative distribution function (cdf)  $F_Z(z) = \Pr(Z \leq z)$  over the real line, or alternatively, the probability density function (pdf)  $f_Z(z)$ , when it exists. The notation  $\mathbf{z} = (z_1, z_2, \dots, z_n)$  is used when it is understood from the context that the length of the sequence is  $n$ , and the  $\ell^2$ -norm of  $\mathbf{z}$  is denoted by  $\|\mathbf{z}\|$ . The trace of a matrix  $A \in \mathbb{R}^{m \times n}$  is denoted by  $\text{tr}(A)$ .

### B. Channel Description

A state-dependent discrete memoryless channel (DMC) with parameters  $(\mathcal{X} \times \mathcal{S} \times \mathcal{T}, W_{Y|X, S, T}, \mathcal{Y})$  consists of finite input alphabet  $\mathcal{X}$ , state alphabet  $\mathcal{S}$ , parameters alphabet  $\mathcal{T}$ , output alphabet  $\mathcal{Y}$ , and a conditional pmf  $W_{Y|X, S, T}$  over  $\mathcal{Y}$ . The channel is without feedback, and it is memoryless when conditioned on the state and parameter sequences, *i.e.*

$$W_{Y^n|X^n, S^n, T^n}(y^n|x^n, s^n, t^n) = \prod_{i=1}^n W_{Y|X, S, T}(y_i|x_i, s_i, t_i). \quad (3)$$

The AVC with fixed parameters is a DMC  $W_{Y|X, S, T}$  where the parameter sequence is fixed, while the state sequence has an unknown distribution, not necessarily independent nor stationary. That is, the parameter sequence is given by

$$T^n = \theta^n, \quad (4)$$

where  $\theta_1, \theta_2, \dots$  is a given sequence of letters from  $\mathcal{T}$ , known to the encoder, decoder, and jammer. Whereas, the state sequence  $S^n \sim q(s^n|\theta^n)$  with an unknown joint pmf  $q(s^n|\theta^n)$  over  $\mathcal{S}^n$ . In particular,  $q(s^n|\theta^n)$  could give mass 1 to some state sequence  $s^n$ . The AVC with fixed parameters is denoted by  $\mathcal{W} = \{W_{Y|X, S, T}, \theta^\infty\}$ , where  $\theta^\infty$  is a short notation for the sequence  $(\theta_i)_{i=1}^\infty$ .

The compound channel with fixed parameters is used as a tool in the analysis. Different models of compound channels are described in the literature [29]. Here, the compound channel with fixed parameters is a DMC  $W_{Y|X, S, T}$  where the state has a conditional product distribution  $q(s|t)$  that is not known in exact, but rather belongs to a family of conditional distributions  $\mathcal{Q}$ , with  $\mathcal{Q} \subseteq \mathcal{P}(\mathcal{S}|\mathcal{T})$ . That is,

$$S^n \sim \prod_{i=1}^n q(s_i|\theta_i) \quad (5)$$

with an unknown conditional pmf  $q(s|t) \in \mathcal{Q}$ . We note that this differs from the classical definition of the compound channel, as in [29], where the state is fixed throughout the transmission.

*Remark 1.* Note that the special case of a channel  $W_{Y|X, S, T=t}$ , with a *constant* parameter  $\theta_i = t$  for  $i = 1, 2, \dots$ , reduces to the standard state-dependent DMC. Thereby, the AVC  $\mathcal{W}_t = \{W_{Y|X, S, T=t}\}$  with a constant parameter can be regarded as the traditional AVC, as introduced by Blackwell *et al.* [16]. On the other hand, the special case of a channel  $W_{Y|X, S, T} = W_{Y|X, T}$ , which does not depend on the state  $S$ , reduces to a TVC [102].

*Remark 2.* The AVC with colored Gaussian noise does *not* fit the description above. Nevertheless, the fixed parameters model is a crucial tool for our final goal, *i.e.* to determine the capacity of the AVC with colored Gaussian noise.

### C. Coding

We introduce some preliminary definitions.

*Definition 1 (Code).* A  $(2^{nR}, n)$  code for the AVC  $\mathcal{W}$  with fixed parameters consists of the following; a message set  $[1 : 2^{nR}]$ , where  $2^{nR}$  is assumed to be an integer, an encoding function  $f : [1 : 2^{nR}] \times \mathcal{T}^n \rightarrow \mathcal{X}^n$ , and a decoding function  $g : \mathcal{Y}^n \times \mathcal{T}^n \rightarrow [1 : 2^{nR}]$ .

Given a message  $m \in [1 : 2^{nR}]$  and a parameter sequence  $\theta^n$ , the encoder transmits the codeword  $x^n = f(m, \theta^n)$ . The decoder receives the channel output  $y^n$ , and finds an estimate of the message  $\hat{m} = g(y^n, \theta^n)$ . We denote the code by  $\mathcal{C} = (f(\cdot, \cdot), g(\cdot, \cdot))$ .

We proceed now to coding schemes when using stochastic-encoder stochastic-decoder pairs with common randomness.

*Definition 2 (Random code).* A  $(2^{nR}, n)$  random code for the AVC  $\mathcal{W}$  with fixed parameters consists of a collection of  $(2^{nR}, n)$  codes  $\{\mathcal{C}_\gamma = (f_\gamma, g_\gamma)\}_{\gamma \in \Gamma}$ , along with a probability distribution  $\mu(\gamma)$  over the code collection  $\Gamma$ . We denote such a code by  $\mathcal{C}^\Gamma = (\mu, \Gamma, \{\mathcal{C}_\gamma\}_{\gamma \in \Gamma})$ .

### D. Input and State Constraints

Next, we consider input constraints and state constraint, imposed on the encoder and the jammer, respectively. We note that the constraints specifications are known to both the user and the jammer in this model. Let  $\phi : \mathcal{X} \rightarrow [0, \infty)$ ,  $k = 1, 2$ , and  $l : \mathcal{S} \rightarrow [0, \infty)$  be some given bounded functions, and define

$$\phi^n(x^n) = \frac{1}{n} \sum_{i=1}^n \phi(x_i), \quad (6)$$

$$l^n(s^n) = \frac{1}{n} \sum_{i=1}^n l(s_i). \quad (7)$$

Let  $\Omega > 0$  and  $\Lambda > 0$ . Below, we specify the input constraint  $\Omega$  and state constraint  $\Lambda$ , corresponding to the functions  $\phi^n(x^n)$  and  $l^n(s^n)$ , respectively. It is assumed that for some  $a \in \mathcal{X}$  and  $b \in \mathcal{S}$ ,  $\phi(a) = l(b) = 0$ .

As the parameter sequence  $\theta^\infty \equiv (\theta_i)_{i=1}^\infty$  is fixed and known to the encoder, the decoder and the jammer, the input and state constraints below are specified for a particular sequence. Given an input constraint  $\Omega$ , the encoding function needs to satisfy

$$\phi^n(f(m, \theta^n)) \leq \Omega, \text{ for all } m \in [1 : 2^{nR}]. \quad (8)$$

That is, the input sequence satisfies  $\phi^n(X^n) \leq \Omega$  with probability 1.

Moving to the state constraint  $\Lambda$ , we have different definitions for the AVC and for the compound channel. The compound channel has a constraint on average, where the state sequence satisfies  $\mathbb{E}_q l^n(S^n) \leq \Lambda$ , while the AVC has an almost-surely constraint,  $l^n(S^n) \leq \Lambda$  with probability (w.p.) 1. Explicitly, we say that a compound channel is under a state constraint  $\Lambda$  if  $\mathcal{Q} \subseteq \overline{\mathcal{P}}_\Lambda(\mathcal{S}|\theta^\infty)$ , where

$$\overline{\mathcal{P}}_\Lambda(\mathcal{S}|\theta^\infty) \triangleq \bigcap_{n=1}^{\infty} \left\{ q(s|t) : \frac{1}{n} \sum_{i=1}^n \sum_{s \in \mathcal{S}} q(s|\theta_i) l(s) \leq \Lambda \right\}. \quad (9)$$

As for the AVC  $\mathcal{W}$ , it is now assumed that the joint distribution of the state sequence is limited to  $q(s^n|\theta^n) \in \mathcal{P}_\Lambda(\mathcal{S}^n|\theta^n)$ , where

$$\mathcal{P}_\Lambda(\mathcal{S}^n|\theta^n) \triangleq \{q(s^n|\theta^n) \in \mathcal{P}(\mathcal{S}^n|\mathcal{T}^n) : q(s^n|\theta^n) = 0 \text{ if } l^n(s^n) > \Lambda\}. \quad (10)$$

This includes the case of a deterministic unknown state sequence, *i.e.* when  $q$  gives probability 1 to a particular  $s^n \in \mathcal{S}^n$  with  $l^n(s^n) \leq \Lambda$ .

### E. Capacity Under Constraints

We move to the definition of an achievable rate and the capacity of the AVC  $\mathcal{W}$  with fixed parameters under input and state constraints. Codes over the AVC  $\mathcal{W}$  with fixed parameters are defined as in Definition 1, with the additional constraint (8) on the codebook.

Define the conditional probability of error of a code  $\mathcal{C}$  given a state sequence  $s^n \in \mathcal{S}^n$  by

$$P_e^{(n)}(\mathcal{C}|s^n, \theta^n) \triangleq \frac{1}{2^{nR}} \sum_{m=1}^{2^{nR}} \sum_{y^n: g(y^n, \theta^n) \neq m} W_{Y^n|X^n, S^n, T^n}(y^n|f(m, \theta^n), s^n, \theta^n). \quad (11a)$$

Now, define the average probability of error of  $\mathcal{C}$  for some distribution  $q(s^n|\theta^n) \in \mathcal{P}(\mathcal{S}^n)$ ,

$$P_e^{(n)}(q, \theta^n, \mathcal{C}) \triangleq \sum_{s^n \in \mathcal{S}^n} q(s^n|\theta^n) P_e^{(n)}(\mathcal{C}|s^n, \theta^n). \quad (11b)$$

*Definition 3* (Achievable rate and capacity under constraints). A code  $\mathcal{C} = (f, g)$  is called a  $(2^{nR}, n, \varepsilon)$  code for the AVC  $\mathcal{W}$  with fixed parameters under input constraint  $\Omega$  and state constraint  $\Lambda$ , when (8) is satisfied and

$$P_e^{(n)}(q, \theta^n, \mathcal{C}) \leq \varepsilon, \quad \text{for all } q \in \mathcal{P}_\Lambda(\mathcal{S}^n|\theta^n), \quad (12)$$

or, equivalently,  $P_e^{(n)}(\mathcal{C}|s^n, \theta^n) \leq \varepsilon$  for all  $s^n \in \mathcal{S}^n$  with  $l^n(s^n) \leq \Lambda$ .

We say that a rate  $R \geq 0$  is achievable under constraints if for every  $\varepsilon > 0$  and sufficiently large  $n$ , there exists a  $(2^{nR}, n, \varepsilon)$  code for the AVC  $\mathcal{W}$  with fixed parameters under input constraint  $\Omega$  and state constraint  $\Lambda$ . The operational capacity is defined as the supremum of achievable rates, and it is denoted by  $\mathbb{C}(\mathcal{W})$ . We use the term ‘capacity’ referring to this operational meaning, and in some places we call it the deterministic code capacity in order to emphasize that achievability is measured with respect to deterministic codes.

Analogously to the deterministic case, a  $(2^{nR}, n, \varepsilon)$  random code  $\mathcal{C}^\Gamma$  satisfies the requirements

$$\sum_{\gamma} \mu(\gamma) \phi^n(f(m, \theta^n)) \leq \Omega, \quad \text{for all } m \in [1 : 2^{nR}], \quad (13a)$$

and

$$P_e^{(n)}(q, \mathcal{C}^\Gamma) \triangleq \sum_{\gamma \in \Gamma} \mu(\gamma) P_e^{(n)}(q, \theta^n, \mathcal{C}_\gamma) \leq \varepsilon, \quad \text{for all } q \in \mathcal{P}_\Lambda(\mathcal{S}^n|\theta^n). \quad (13b)$$

The capacity region achieved by random codes is then denoted by  $\mathbb{C}^*(\mathcal{W})$ , and it is referred to as the *random code capacity*.

The definitions above are naturally extended to the compound channel with fixed parameters, under input constraints  $\Omega$  and state constraint  $\Lambda$ , by limiting the requirements (8), (12) and (13) to conditionally memoryless state distributions  $q \in \mathcal{Q}$ . The respective deterministic code capacity  $\mathbb{C}(\mathcal{W}^\mathcal{Q})$  and random code capacity  $\mathbb{C}^*(\mathcal{W}^\mathcal{Q})$  are defined accordingly.

### III. MAIN RESULTS – CHANNELS WITH FIXED PARAMETERS

In this section, we establish the random code capacity of the AVC with fixed parameters. To this end, we first give an auxiliary result on the compound channel.

#### A. The Compound Channel with Fixed Parameters

We begin with the capacity theorem for the compound channel  $\mathcal{W}^\mathcal{Q} = \{W_{Y|X,S,T}, \mathcal{Q}, \theta^\infty\}$ . This is an auxiliary result, obtained by a simple extension of [29, Exercise 6.8]. A similar result appears in [74] as well. Given a parameter sequence  $\theta^n$  of a fixed length, define

$$C_n(\mathcal{W}^\mathcal{Q}) = \max_{p(x|t): \mathbb{E}\phi(X) \leq \Omega} \inf_{q(s|t) \in \mathcal{Q}} I_q(X; Y|T), \quad (14)$$

with  $(T, S, X) \sim P_T(t)p(x|t)q(s|t)$ , where  $P_T$  is the type of the parameter sequence  $\theta^n$ .

*Lemma 1.* The capacity of the compound channel  $\mathcal{W}^\mathcal{Q}$  with fixed parameters, under input constraint  $\Omega$  and state constraint  $\Lambda$ , is given by

$$\mathbb{C}(\mathcal{W}^\mathcal{Q}) = \liminf_{n \rightarrow \infty} C_n(\mathcal{W}^\mathcal{Q}), \quad (15)$$

and it is identical to the random code capacity, *i.e.*  $\mathbb{C}^*(\mathcal{W}^\mathcal{Q}) = \mathbb{C}(\mathcal{W}^\mathcal{Q})$ .

The proof of Lemma 1 is given in Appendix A.

#### B. The AVC with Fixed Parameters – Random Code Capacity

We determine the random code capacity of the AVC with fixed parameters,  $\mathcal{W} = \{W_{Y|X,S,T}, \theta^\infty\}$ , under input constraint  $\Omega$  and state constraint  $\Lambda$ . The random code derivation is based on our result on the compound channel with fixed parameters and a variation of Ahlswede’s Robustification Technique (RT). Define

$$C_n^*(\mathcal{W}) \triangleq C_n(\mathcal{W}^\mathcal{Q}) \Big|_{\mathcal{Q} = \overline{\mathcal{P}}_\Lambda(\mathcal{S}|\theta^\infty)}. \quad (16)$$

We begin with a lemma, based on Ahlswede’s RT [6] (see also [82, Lemma 9]). We modify it here to include the parameter sequence  $\theta^n$  and the constraint on the family of conditional state distributions  $q(s|t)$ .

*Lemma 2* (Modified RT). Let  $h : \mathcal{S}^n \times \mathcal{T}^n \rightarrow [0, 1]$  be a given function. If, for some fixed  $\alpha_n \in (0, 1)$ , and for all  $q^n(s^n|\theta^n) = \prod_{i=1}^n q(s_i|\theta_i)$ , with  $q \in \overline{\mathcal{P}}_\Lambda(\mathcal{S}|\theta^\infty)$ ,

$$\sum_{s^n \in \mathcal{S}^n} q^n(s^n|\theta^n) h(s^n, \theta^n) \leq \alpha_n, \quad (17)$$

then,

$$\frac{1}{|\Pi(\theta^n)|} \sum_{\pi \in \Pi(\theta^n)} h(\pi s^n, \theta^n) \leq \beta_n, \quad \text{for all } s^n \in \mathcal{S}^n \text{ such that } l^n(s^n) \leq \Lambda, \quad (18)$$

where  $\Pi(\theta^n)$  is the set of all  $n$ -tuple permutations  $\pi : \mathcal{S}^n \rightarrow \mathcal{S}^n$  such that  $\pi\theta^n = \theta^n$ , and  $\beta_n = (n+1)^{|\mathcal{S}||\mathcal{T}|} \alpha_n$ .

Originally, Ahlswede's RT is stated so that (17) holds for any  $q(s) \in \mathcal{P}(\mathcal{S})$ , without state constraint (see [6]), and without conditioning on the parameter sequence  $\theta^n$ . We give the proof of Lemma 2 in Appendix B. Next, we give our random code capacity theorem.

*Theorem 3.* The random code capacity of the AVC  $\mathcal{W}$  with fixed parameters, under input constraint  $\Omega$  and state constraint  $\Lambda$ , is given by

$$\mathbb{C}^*(\mathcal{W}) = \liminf_{n \rightarrow \infty} \mathbb{C}_n^*(\mathcal{W}). \quad (19)$$

The proof of Theorem 3 is given in Appendix C. The proof is based on our extension of Ahlswede's RT above. Essentially, we use a reliable code for the compound channel to construct a random code for the AVC by applying random permutations to the codeword symbols. However, here, we only use permutations that do not affect the parameter sequence  $\theta^n$ . The result above plays a central role in the proof of the capacity theorem in Section V, where the AVC with colored Gaussian noise is considered.

We also give an equivalent formulation in terms of the random code capacity of the traditional AVC. As mentioned in Remark 1, the case of an AVC  $\{W_{Y|X,S,T=t}\}$  with a constant parameter  $\theta_i = t$  reduces to the traditional AVC under input and state constraints. For this channel, Csiszár and Narayan [31] showed that the random code capacity is given by

$$\mathbb{C}_t^*(\Omega, \Lambda) \triangleq \min_{q(s) : \mathbb{E}l(S) \leq \Lambda} \max_{p(x) : \mathbb{E}\phi(X) \leq \Omega} I_q(X; Y|T=t) = \max_{p(x) : \mathbb{E}\phi(X) \leq \Omega} \min_{q(s) : \mathbb{E}l(S) \leq \Lambda} I_q(X; Y|T=t) \quad (20)$$

where the last equality is due to the minimax theorem [96]. Then, define

$$\mathbb{R}_n^*(\mathcal{W}) \triangleq \min_{\lambda_1, \dots, \lambda_n : \frac{1}{n} \sum_{i=1}^n \lambda_i \leq \Lambda} \max_{\omega_1, \dots, \omega_n : \frac{1}{n} \sum_{i=1}^n \omega_i \leq \Omega} \frac{1}{n} \sum_{i=1}^n \mathbb{C}_{\theta_i}^*(\omega_i, \lambda_i), \quad (21)$$

*Lemma 4.*

$$\mathbb{R}_n^*(\mathcal{W}) = \mathbb{C}_n^*(\mathcal{W}). \quad (22)$$

The proof of Lemma 4 is given in Appendix D. Theorem 3 and Lemma 4 yield the following consequence.

*Corollary 5.* The random code capacity of the AVC  $\mathcal{W}$  with fixed parameters, under input constraint  $\Omega$  and state constraint  $\Lambda$ , is given by

$$\mathbb{C}^*(\mathcal{W}) = \liminf_{n \rightarrow \infty} \mathbb{R}_n^*(\mathcal{W}). \quad (23)$$

The corollary will also be useful in our analysis of the AVC with colored Gaussian noise.

### C. The AVC with Fixed Parameters – Deterministic Code Capacity

We move to the deterministic code capacity of the AVC with fixed parameters,  $\mathcal{W} = \{W_{Y|X,S,T}, \theta^\infty\}$ , under input constraint  $\Omega$  and state constraint  $\Lambda$ .

1) *Capacity Theorem:* Before we state the capacity theorem, we give a few definitions. We begin with symmetrizability of a channel *without* parameters.

*Definition 4* (see [30]). A state-dependent DMC  $V_{Y|X,S}$  is said to be *symmetrizable* if for some conditional distribution  $J(s|x)$ ,

$$\sum_{s \in \mathcal{S}} V_{Y|X,S}(y|x_1, s) J(s|x_2) = \sum_{s \in \mathcal{S}} V_{Y|X,S}(y|x_2, s) J(s|x_1), \quad \forall x_1, x_2 \in \mathcal{X}, y \in \mathcal{Y}. \quad (24)$$

Equivalently, the channel  $\tilde{V}(y|x_1, x_2) = \sum_{s \in \mathcal{S}} V_{Y|X,S}(y|x_1, s) J(s|x_2)$  is symmetric, i.e.  $\tilde{V}(y|x_1, x_2) = \tilde{V}(y|x_2, x_1)$ , for all  $x_1, x_2 \in \mathcal{X}$  and  $y \in \mathcal{Y}$ . We say that such a  $J : \mathcal{X} \rightarrow \mathcal{S}$  symmetrizes  $V_{Y|X,S}$ .

Intuitively, symmetrizability identifies a poor channel, where the jammer can impinge the communication scheme by randomizing the state sequence  $S^n$  according to  $J^n(s^n|x_2^n) = \prod_{i=1}^n J(s_i|x_{2,i})$ , for some codeword  $x_2^n$ . Suppose that the transmitted codeword is  $x_1^n$ . The codeword  $x_2^n$  can be thought of as an impostor sent by the jammer. Now, since the ‘‘average channel’’  $\tilde{V}$  is symmetric with respect to  $x_1^n$  and  $x_2^n$ , the two codewords appear to the receiver as equally likely. Indeed, by [37], if the AVC  $\{V_{Y|X,S}\}$  without parameters and free of constraints is symmetrizable, then its capacity is zero.

We will assume that either the channels  $W_{Y|X,S}(\cdot|\cdot, \cdot, \theta_i)$  are all symmetrizable, or the number of non-symmetrizable channels grows linearly with  $n$ . That is,

$$\text{either } |\mathcal{I}(n)| = 0 \text{ or } |\mathcal{I}(n)| = \Omega(n), \quad (25a)$$

where

$$\mathcal{I}(n) = \{i \in [1 : n] : W_{Y|X,S}(\cdot|\cdot, \cdot, \theta_i) \text{ is non-symmetrizable}\}. \quad (25b)$$

The asymptotic notation  $f(n) = \Omega(n)$  means that there exist  $n_0 > 0$  and  $0 < \alpha \leq 1$  such that  $f(n) \geq \alpha n$  for all  $n \geq n_0$ . An intuitive explanation for this assumption is given in Remark 3 below. Next, we define a symmetrizability cost and threshold for the AVC with fixed parameters. For every  $n$  and  $p(x|t)$  with

$$\frac{1}{n} \sum_{i=1}^n p(x|\theta_i) \phi(x) \leq \Omega, \quad (26)$$

define the *minimal symmetrizability cost* by

$$\tilde{\Lambda}_n(p) \triangleq \min \frac{1}{n} \sum_{i=1}^n \sum_{x \in \mathcal{X}} \sum_{s \in \mathcal{S}} p(x|\theta_i) J_{\theta_i}(s|x) l(s) = \min \sum_{t \in \mathcal{T}} \sum_{x \in \mathcal{X}} \sum_{s \in \mathcal{S}} P_T(t) p(x|t) J_t(s|x) l(s), \quad (27)$$

where the minimization is over the conditional distributions  $J_t(s|x)$  that symmetrize  $W_{Y|X,S,T}(\cdot|\cdot, \cdot, t)$ , for  $t \in \mathcal{T}$  (see Definition 4). We use the convention that a minimum value over an empty set is  $+\infty$ . Note that the last equality in (27) holds since  $P_T$  is defined as the type of the parameter sequence  $\theta^n$ , hence averaging over time is the same as averaging according to  $P_T$ . In addition, define the *symmetrizability threshold*

$$L_n^* \triangleq \max_{p(x|t) : \frac{1}{n} \sum_{i=1}^n p(x|\theta_i) \phi(x) \leq \Omega} \tilde{\Lambda}_n(p). \quad (28)$$

Intuitively,  $\tilde{\Lambda}_n(p)$  is the minimal average state cost which the jammer has to pay to symmetrize the channel at each time instance, for a given conditional input distribution  $p(x|t)$ . If this minimal state cost violates the state constraint  $\Lambda$ , then the jammer is prohibited from symmetrizing the channel. Indeed, we will show that if there exists an input distribution  $p(x|t)$  with  $\frac{1}{n} \sum_{i=1}^n p(x|\theta_i) \phi(x) \leq \Omega$  and  $\tilde{\Lambda}_n(p) > \Lambda$  for large  $n$ , then the deterministic code capacity is positive. The symmetrizability threshold  $L_n^*$  is the worst symmetrizability cost from the jammer’s perspective.

Our capacity result is stated below. Let

$$C_n(\mathcal{W}) \triangleq \begin{cases} \min_{q(s|t) : \mathbb{E}_q l(S) \leq \Lambda} \max_{\substack{p(x|t) : \mathbb{E} \phi(X) \leq \Omega, \\ \tilde{\Lambda}_n(p) \geq \Lambda}} I_q(X; Y|T) & \text{if } L_n^* > \Lambda, \\ 0 & \text{if } L_n^* \leq \Lambda \end{cases}, \quad (29)$$

with  $(T, S, X) \sim P_T(t)p(x|t)q(s|t)$ , where  $P_T$  is the type of the parameter sequence  $\theta^n$  with a fixed length  $n$ .

**Theorem 6.** Assume that  $L_n^* \neq \Lambda$  for sufficiently large  $n$  and that (25) holds. The capacity of an AVC  $\mathcal{W}$  with fixed parameters, under input constraint  $\Omega$  and state constraint  $\Lambda$ , is given by

$$\mathbb{C}(\mathcal{W}) = \liminf_{n \rightarrow \infty} C_n(\mathcal{W}). \quad (30)$$

In particular, if the channels  $W_{Y|X,S,T}(\cdot|\cdot, \cdot, t)$ ,  $t \in \mathcal{T}$ , are non-symmetrizable, then  $\mathbb{C}(\mathcal{W}) = \mathbb{C}^*(\mathcal{W}) = \liminf_{n \rightarrow \infty} C_n^*(\mathcal{W})$ . That is, the deterministic code capacity coincides with the random code capacity.

The proof of Theorem 6 is given in Appendix G. The theorem will also play a central role in the proof of the capacity theorem in Section V.

**Remark 3.** Observe that the second part of the theorem implies that for the case where there are no constraints, i.e.  $\Omega = \phi_{max}$  and  $\Lambda = l_{max}$ , non-symmetrizability is a sufficient condition for positive capacity. Specifically, according to the definition of  $\tilde{\Lambda}_n(p)$ ,  $L_n^*$  in (27)-(28), if some of the channels  $W_{Y|X,S,T}(\cdot|\cdot, \cdot, \theta_i)$  are non-symmetrizable, then the symmetrizability threshold is  $L_n^* = \infty$ , hence the capacity is positive. Intuitively, if the number of such channels is constant, i.e.  $|\mathcal{I}(n)| = c$  for all  $n$ , it seems that this assignment of  $L_n^*$  does not make sense, since the user cannot achieve positive rates by coding over a negligible fraction of the block. Yet, our assumption in (25) excludes this scenario. In particular, if  $|\mathcal{I}(n)|$  is non-zero, then we assume that  $|\mathcal{I}(n)|$  grows linearly in  $n$ , in which case positive rates can be achieved by coding over the part of the block that lies

within  $\mathcal{I}(n)$ . Furthermore, without constraints, we may replace the linear growth assumption with a poly-logarithmic one, i.e.  $|\mathcal{I}(n)| = \Omega((\log n)^a)$ , with  $a > 1$ . Indeed, based on Ahlswede's elimination technique [5], the random code capacity can be achieved with a code collection of polynomial size,  $|\Gamma| = n^2$ . Therefore, without state constraints, the random element  $\gamma \in \Gamma$  can be reliably sent to the receiver over the sub-block  $\mathcal{I}(n)$ , at rate  $\rho_n = \frac{\log |\Gamma|}{(\log n)^a} = 2(\log n)^{-(a-1)}$ , which tends to zero as  $n \rightarrow \infty$ , hence the decrease in the overall rate is negligible as well. We deduce that if  $|\mathcal{I}(n)| = \Omega((\log n)^a)$ , then the deterministic code capacity of the AVC with fixed parameters without constraints is the same as the random code capacity, i.e.

$$\mathbb{C}(\mathcal{W}) = \mathbb{C}^*(\mathcal{W}) = \liminf_{n \rightarrow \infty} \min_{q(s|t)} \max_{p(x|t)} I_q(X; Y|T). \quad (31)$$

*Remark 4.* Even in the case where there are no parameters, the boundary case where  $L_n^* = \Lambda$  is an open problem. Although, for the traditional AVC, it is conjectured in [30] that the capacity is zero in this case. Similarly, we conjecture that the capacity of the AVC with fixed parameters is given by  $\mathbb{C}(\mathcal{W}) = \liminf_{n \rightarrow \infty} \mathbb{C}_n(\mathcal{W})$  for all values of  $\{L_n^*\}_{n \geq 1}$ , provided that (25) holds. There are special cases where we know that this holds, given in the corollary below. The corollary is based on the remark following Theorem 3 in [30].

*Corollary 7.* Let  $\mathcal{W}$  be an AVC with fixed parameters such that all channels  $W_{Y|X,S,T}(\cdot|\cdot, \cdot, t)$ ,  $t \in \mathcal{T}$ , are symmetrizable. If the minimum in (27) is attained by a 0-1 law, for every  $n$  and  $p(x|t)$  with  $\frac{1}{n} \sum_{i=1}^n p(x|\theta_i) \phi(x) \leq \Omega$ , then

$$\mathbb{C}(\mathcal{W}) = \liminf_{n \rightarrow \infty} \mathbb{C}_n(\mathcal{W}). \quad (32)$$

The proof of Corollary 7 is given in Appendix H. In particular, we note that the condition of 0-1 law in Corollary 7 holds when the output  $Y$  is a deterministic function of  $X$ ,  $S$ , and  $T$ . As opposed to Theorem 6, the statement in Corollary 7 holds for all values of  $\{L_n^*\}_{n \geq 1}$ .

2) *Decoding Rule:* We specify the decoding rule and state the corresponding properties, which are used in the analysis. To specify the decoding rule, we define the decoding sets  $\mathcal{D}(m) \subseteq \mathcal{Y}^n \times \mathcal{T}^n$ , for  $m \in [1 : 2^{nR}]$ , such that  $g(y^n, \theta^n) = m$  iff  $(y^n, \theta^n) \in \mathcal{D}(m)$ .

*Definition 5 (Decoder).* Given the codebook  $\{f(m, \theta^n)\}_{m \in [1:2^{nR}]}$ , declare that  $(y^n, \theta^n) \in \mathcal{D}(m)$  if there exists  $s^n \in \mathcal{S}^n$  with  $l^n(s^n) \leq \Lambda$  such that the following hold.

1) For  $(T, X, S, Y)$  that is distributed according to the joint type  $\hat{P}_{\theta^n, f(m, \theta^n), s^n, y^n}$ , we have that

$$D(P_{T,X,S,Y} || P_T \times P_{X|T} \times P_{S|T} \times W_{Y|X,S,T}) \leq \eta. \quad (33)$$

2) For every  $\tilde{m} \neq m$  such that for some  $\tilde{s}^n \in \mathcal{S}^n$  with  $l^n(\tilde{s}^n) \leq \Lambda$ ,

$$D(P_{T,\tilde{X},\tilde{S},Y} || P_T \times P_{\tilde{X}|T} \times P_{\tilde{S}|T} \times W_{Y|X,S,T}) \leq \eta, \quad (34)$$

where  $(T, \tilde{X}, \tilde{S}, Y) \sim \hat{P}_{\theta^n, f(\tilde{m}, \theta^n), \tilde{s}^n, y^n}$ , we have that

$$I(X, Y; \tilde{X} | S, T) \leq \eta. \quad (35)$$

We note that in Definition 5, the variables  $T, X, \tilde{X}, S, \tilde{S}, Y$  are dummy random variables, distributed according to the joint type of  $(\theta^n, f(m, \theta^n), f(\tilde{m}, \theta^n), s^n, \tilde{s}^n, y^n)$ , where  $f(m, \theta^n)$  is a ‘‘tested’’ codeword,  $f(\tilde{m}, \theta^n)$  is a competing codeword,  $s^n$  is a ‘‘tested’’ state sequence,  $\tilde{s}^n$  is a competing state sequence, and  $y^n$  is the received sequence. None of the sequences are random here. We may have that the conditional type  $P_{Y|X,S,T}$  differs from the actual channel  $W_{Y|X,S,T}$ . Therefore, the divergences and mutual informations in Definition 5 could be positive.

For the definition above to be proper, the decoding sets need to be disjoint, as stated in the following lemma.

*Lemma 8 (Decoding Disambiguity).* Suppose that in each codebook, all codewords have the same conditional type, i.e.  $\hat{P}_{f(m, \theta^n) | \theta^n} = p$  for all  $m \in [1 : 2^{nR}]$ . Assume (25) holds, that for some  $\delta_0, \delta_1 > 0$ ,  $P_T(t) \geq \delta_0$ ,  $p(x|t) \geq \delta_1$ ,  $\forall x \in \mathcal{X}$ ,  $t \in \mathcal{T}$ , and also

$$\tilde{\Lambda}_n(p) > \Lambda. \quad (36)$$

Then, for sufficiently small  $\eta > 0$ ,

$$\mathcal{D}(m) \cap \mathcal{D}(\tilde{m}) = \emptyset, \text{ for all } m \neq \tilde{m}. \quad (37)$$

The proof of Lemma 8 is given in Appendix E.

3) *Codebook Generation*: We now extend Csiszár and Narayan's lemma for the codebook generation [30].

*Lemma 9* (Codebooks Generation). For every  $\varepsilon > 0$ , sufficiently large  $n$ , rate  $R \geq \varepsilon$  and conditional type  $p(x|t)$ , there exist a set of codewords  $\{x^n(m, \theta^n)\}_{m \in [1:2^{nR}]}$  of conditional type  $p$ , such that for every  $a^n \in \mathcal{X}^n$  and  $s^n \in \mathcal{S}^n$  with  $l^n(s^n) \leq \Lambda$ , and every joint type  $P_{T,X,\tilde{X},S}$  with  $P_{X|T} = P_{\tilde{X}|T} = p$ , the following hold.

$$|\{\tilde{m} : (\theta^n, a^n, x^n(\tilde{m}, \theta^n), s^n) \in \mathcal{T}^n(P_{T,X,\tilde{X},S})\}| \leq 2^{n([R-I(\tilde{X};X,S|T)]_+ + \varepsilon)}, \quad (38)$$

$$|\{m : (\theta^n, x^n(m, \theta^n), s^n) \in \mathcal{T}^n(P_{T,X,S})\}| \leq 2^{n(R - \frac{\varepsilon}{2})}, \text{ if } I(X; S|T) > \varepsilon, \quad (39)$$

and

$$\begin{aligned} & |\{m : (\theta^n, x^n(m, \theta^n), x^n(\tilde{m}, \theta^n), s^n) \in \mathcal{T}^n(P_{T,X,\tilde{X},S}), \text{ for some } \tilde{m} \neq m\}| \\ & \leq 2^{n(R - \frac{\varepsilon}{2})}, \text{ if } I(X; \tilde{X}, S|T) - [R - I(\tilde{X}; S|T)]_+ > \varepsilon. \end{aligned} \quad (40)$$

The proof of Lemma 9 is given in Appendix F.

#### D. Super-Additivity

We also give an equivalent formulation with a sum over  $i \in [1 : n]$ . Here, as opposed to the previous section, the formula *cannot* be expressed in terms of the capacities of the constant-parameter AVCs  $\{W_{Y|X,S,T=\theta_i}\}$ . Considering the AVC without constraints, Schaefer *et al.* [91] showed that the capacity of any product AVC that is composed of a symmetrizable channel and a non-symmetrizable channel is larger than the sum of the individual capacities (see Theorem 6 in [91]). Similarly, we give an example at the end of this section where the capacity of the AVC *with fixed parameters* is larger than the weighted sum of the capacities of the constant-parameter AVCs  $\{W_{Y|X,S,T=\theta_i}\}$ . This phenomenon can be viewed as an instance of the super-additivity property in [91].

We begin with constant-parameter definitions, *i.e.* for a fixed  $T = t$ . For every input distribution  $p(x)$  with  $\mathbb{E}\phi(X) \leq \Omega$ , define the constant-parameter minimal symmetrizability cost by

$$\tilde{\Lambda}(p, t) \triangleq \min_{x \in \mathcal{X}} \sum_{s \in \mathcal{S}} p(x) J(s|x) l(s), \quad (41)$$

where the minimization is over the distributions  $J(s|x)$  that symmetrize  $W_{Y|X,S,T}(\cdot, \cdot, t)$ , where  $t \in \mathcal{T}$  is fixed (see Definition 4). Then, we can write the minimal symmetrizability cost defined in (27) as

$$\tilde{\Lambda}_n(p(\cdot|\cdot)) = \frac{1}{n} \sum_{i=1}^n \tilde{\Lambda}(p(\cdot|\theta_i), \theta_i). \quad (42)$$

Let

$$R_n(\mathcal{W}) \triangleq \begin{cases} \min_{\lambda_1, \dots, \lambda_n : \frac{1}{n} \sum_{i=1}^n \lambda_i \leq \Lambda} \max_{\omega_1, \dots, \omega_n, \tilde{\lambda}_1, \dots, \tilde{\lambda}_n : \frac{1}{n} \sum_{i=1}^n \omega_i \leq \Omega, \frac{1}{n} \sum_{i=1}^n \tilde{\lambda}_i \geq \Lambda} \frac{1}{n} \sum_{i=1}^n C_{\theta_i}(\omega_i, \tilde{\lambda}_i, \lambda_i) & \text{if } L_n^* > \Lambda, \\ 0 & \text{if } L_n^* \leq \Lambda \end{cases}, \quad (43)$$

where

$$C_t(\Omega, \Delta, \Lambda) \triangleq \min_{q(s) : \mathbb{E}_q l(S) \leq \Lambda} \max_{p(x) : \mathbb{E} \phi(X) \leq \Omega, \tilde{\Lambda}(p, t) \geq \Delta} I_q(X; Y|T = t) \quad (44)$$

We note that based on Csiszár and Narayan's result in [30], the capacity of the constant-parameter AVC  $\{W_{Y|X,S,T=t}\}$  is given by  $C_t(\Omega, \Delta, \Lambda)$  with  $\Delta = \Lambda$ .

*Lemma 10.*

$$R_n(\mathcal{W}) = C_n(\mathcal{W}). \quad (45)$$

The proof of Lemma 10 is given in Appendix I. Theorem 6, Corollary 7, and Lemma 10 yield the following consequence.

*Corollary 11.* The deterministic code capacity of the AVC  $\mathcal{W}$  with fixed parameters, under input constraint  $\Omega$  and state constraint  $\Lambda$ , is given by

$$\mathbb{C}(\mathcal{W}) = \liminf_{n \rightarrow \infty} R_n(\mathcal{W}), \text{ if } L_n^* \neq \Lambda \text{ for sufficiently large } n \text{ and (25) holds.} \quad (46)$$

Furthermore, if the minimum in (41) is attained by a 0-1 law, for every  $p(x)$  with  $\mathbb{E}\phi(X) \leq \Omega$ , and for all  $t \in \mathcal{T}$ , then

$$\mathbb{C}(\mathcal{W}) = \liminf_{n \rightarrow \infty} R_n(\mathcal{W}), \quad (47)$$

for all values of  $\{L_n\}_{n \geq 1}$ .

The corollary will also be useful in our analysis of the AVC with colored Gaussian noise.

*Example 1.* Consider the arbitrarily varying binary symmetric channel (BSC) with fixed parameters,

$$Y = X + S + Z_T \quad \text{mod } 2 \quad (48)$$

with  $\mathcal{X} = \mathcal{S} = \mathcal{T} = \{0, 1\}$ , where  $Z_t \sim \text{Bernoulli}(\varepsilon_t)$ , for  $t = 0, 1$ ,  $\varepsilon_0 < \varepsilon_1 < \frac{1}{2}$ . Consider a parameter sequence with an empirical distribution  $P_T(0) = P_T(1) = \frac{1}{2}$ , say  $\theta_{2i} = 0$  and  $\theta_{2i-1} = 1$  for  $i = 1, 2, \dots$ . Suppose that the user and the jammer are subject to input constraint  $\Omega$  and state constraint  $\Lambda$ , respectively, with Hamming weight cost functions, i.e.  $\phi(x) = x$  and  $l(s) = s$ .

For the constant-parameter AVC, we have by Definition 4 that  $W_{Y|X,S,T=t}$  is symmetrized by any symmetric distribution, i.e. with  $J(s|1) = 1 - J(s|0)$ . Denoting  $\zeta = J(1|1) = 1 - J(1|0)$ , we have that

$$\tilde{\Lambda}(P_X, t) = \min_{0 \leq \zeta \leq 1} [(1 - \zeta)P_X(0) + \zeta P_X(1)] = \min(P_X(0), P_X(1)). \quad (49)$$

Based on the analysis by Csiszár and Narayan [30, Example 1], the capacity of the constant-parameter AVC under input constraint  $\omega$  and state constraint  $\lambda$  is given by

$$\tilde{\mathbb{C}}_t(\omega, \lambda) = \begin{cases} 0 & \text{if } \omega < \lambda < \frac{1}{2} \\ h(\omega * \lambda * \varepsilon_t) - h(\omega * \lambda * \varepsilon_t) & \text{if } \lambda < \omega < \frac{1}{2} \\ 1 - h(\omega * \lambda * \varepsilon_t) & \text{if } \lambda < \frac{1}{2} \leq \omega \\ 0 & \text{if } \lambda \geq \frac{1}{2} \end{cases} \quad (50)$$

where  $h(x) = -x \log x - (1 - x) \log x$  is the binary entropy function and  $a * b = (1 - a)b + a(1 - b)$ .

Suppose that

$$\varepsilon_0 = \frac{1}{4}, \quad \varepsilon_1 = \frac{5}{12}, \quad \Omega = \frac{5}{16}, \quad \Lambda = \frac{1}{4}. \quad (51)$$

For those values, we have that

$$L_n^* = \max_{P_{X|T}: \frac{1}{2}\mathbb{E}(X|T=0) + \frac{1}{2}\mathbb{E}(X|T=1) \leq \Omega} \left[ \frac{1}{2}P_{X|T}(1|0) + \frac{1}{2}P_{X|T}(1|1) \right] = \Omega = \frac{5}{16}. \quad (52)$$

Thus, by Corollary 11, the capacity is given by

$$\mathbb{C}(\mathcal{W}) = h\left(\frac{5}{16} * \frac{7}{16}\right) - h\left(\frac{7}{16}\right) = \frac{1}{2} (h(\omega_0 * \lambda_0 * \varepsilon_0) - h(\lambda_0 * \varepsilon_0)) + \frac{1}{2} (h(\omega_1 * \lambda_1 * \varepsilon_1) - h(\lambda_1 * \varepsilon_1)) \quad (53)$$

with  $\omega_0 = \omega_1 = \frac{5}{16}$ ,  $\lambda_0 = \frac{3}{8}$  and  $\lambda_1 = \frac{1}{8}$ . Whereas, using two separate codes for  $W_{Y|X,S,T=0}$  and  $W_{Y|X,S,T=1}$  independently, the rate achieved is

$$\frac{1}{2}\tilde{\mathbb{C}}_0(\omega_0, \lambda_0) + \frac{1}{2}\tilde{\mathbb{C}}_1(\omega_1, \lambda_1) = 0 + \frac{1}{2} (h(\omega_1 * \lambda_1 * \varepsilon_1) - h(\lambda_1 * \varepsilon_1)) < \mathbb{C}(\mathcal{W}). \quad (54)$$

This can be viewed as an instance of the more general phenomenon of super-additivity, that holds for any product AVC which is composed of a symmetrizable AVC and a non-symmetrizable AVC [91, Theorem 6].

### E. Example: Channel with Fadings

To illustrate our results, we give another example.

*Example 2.* Consider an arbitrarily varying fading channel,

$$Y_i = \theta_i X_i + S_i + Z_i, \quad (55)$$

with a Gaussian noise sequence  $Z^n$  that is i.i.d.  $\sim \mathcal{N}(0, \sigma^2)$ , where  $\theta_1, \theta_2, \dots$  is a sequence of fixed fading coefficients. Recently, Hosseinigoki and Kosut [57] considered this channel with a random memoryless sequence of fading coefficients. Yet, we assume that the fading coefficients are fixed, and belong to a finite set  $\mathcal{T}$ . Intuitively, the jammer would like to confuse the decoder by sending a state sequence that simulates the sequence  $\theta^n X^n \equiv (\theta_i X_i)_{i=1}^n$ . Indeed, as seen below, the deterministic code capacity is positive only if there exists an input distribution such that  $\frac{1}{n} \sum_{i=1}^n \theta_i^2 \mathbb{E}X_i^2 > \Lambda$ , in which case the jammer cannot simulate  $\theta^n X^n$  without violating the state constraint.

Although we previously assumed that the alphabets are finite, our results can be extended to the continuous case as well, using standard discretization techniques [15, 5] [36, Section 3.4.1]. By Theorem 3, the random code capacity is given by

$$\mathbb{C}^*(\mathcal{W}) = \liminf_{n \rightarrow \infty} \mathbb{C}_n^*(\mathcal{W}). \quad (56)$$

Then, we show that

$$\mathbb{C}_n^*(\mathcal{W}) = \min_{\lambda(t): \mathbb{E}\lambda(T) \leq \Lambda} \max_{\omega(t): \mathbb{E}\omega(T) \leq \Omega} \mathbb{E} \left[ \frac{1}{2} \log \left( 1 + \frac{T^2 \omega(T)}{\lambda(T) + \sigma^2} \right) \right], \quad (57)$$

with expectation over  $T \sim P_T$ , where  $P_T$  is the type of the sequence  $\theta^n$ .

As for the deterministic code capacity, we show that the minimum in (27) is attained by a 0-1 law that gives probability 1 to  $s = \theta_i^2 x$ , hence we can determine the capacity using Corollary 7. We show that the minimal symmetrizability cost is given by

$$\tilde{\Lambda}_n(F_{X|T}) = \frac{1}{n} \sum_{i=1}^n \theta_i^2 \mathbb{E}[X^2 | T = \theta_i] = \mathbb{E}(T^2 X^2), \quad (58)$$

and deduce that the capacity of the AVC with fixed fading coefficients is given by

$$\mathbb{C}(\mathcal{W}) = \liminf_{n \rightarrow \infty} \mathbb{C}_n(\mathcal{W}), \quad (59)$$

with

$$\mathbb{C}_n(\mathcal{W}) \triangleq \begin{cases} \min_{\lambda(t): \mathbb{E}\lambda(T) \leq \Lambda} \max_{\substack{\omega(t): \mathbb{E}\omega(T) \leq \Omega, \\ \mathbb{E}(T^2 \omega(T)) \geq \Lambda}} \mathbb{E} \left[ \frac{1}{2} \log \left( 1 + \frac{T^2 \omega(T)}{\lambda(T) + \sigma^2} \right) \right] & \text{if } \max_{\omega(t): \mathbb{E}\omega(T) \leq \Omega} \mathbb{E}(T^2 \omega(T)) > \Lambda, \\ 0 & \text{if } \max_{\omega(t): \mathbb{E}\omega(T) \leq \Omega} \mathbb{E}(T^2 \omega(T)) \leq \Lambda \end{cases}. \quad (60)$$

The derivation is given in Appendix J. We note that the last expression has the same form as the capacity formula established by Hosseinigoki and Kosut [57] for a random memoryless sequence of fading coefficients.

Next, we extend the result above to continuous fading coefficients, where  $\mathcal{T} = [-t_0, t_0] \subset \mathbb{R}$ . First, we observe that the formulas above can also be written as

$$\mathbb{C}_n^*(\mathcal{W}) = \min_{\substack{\lambda_1, \dots, \lambda_n: \\ \frac{1}{n} \sum_{i=1}^n \lambda_i \leq \Lambda}} \max_{\substack{\omega_1, \dots, \omega_n: \\ \frac{1}{n} \sum_{i=1}^n \omega_i \leq \Omega}} \frac{1}{n} \sum_{i=1}^n \frac{1}{2} \log \left( 1 + \frac{\theta_i^2 \omega_i}{\lambda_i + \sigma^2} \right), \quad (61)$$

and

$$\mathbb{C}_n(\mathcal{W}) = \begin{cases} \min_{\substack{\lambda_1, \dots, \lambda_n: \\ \frac{1}{n} \sum_{i=1}^n \lambda_i \leq \Lambda}} \max_{\substack{\omega_1, \dots, \omega_n: \\ \frac{1}{n} \sum_{i=1}^n \omega_i \leq \Omega, \\ \frac{1}{n} \sum_{i=1}^n \theta_i^2 \omega_i \geq \Lambda}} \frac{1}{n} \sum_{i=1}^n \frac{1}{2} \log \left( 1 + \frac{\theta_i^2 \omega_i}{\lambda_i + \sigma^2} \right) & \text{if } \max_{\substack{\omega_1, \dots, \omega_n: \\ \frac{1}{n} \sum_{i=1}^n \omega_i \leq \Omega}} \frac{1}{n} \sum_{i=1}^n \theta_i^2 \omega_i > \Lambda, \\ 0 & \text{otherwise.} \end{cases} \quad (62)$$

This follows from the same considerations as in the proofs of Lemma 4 and Lemma 10. Now, if the fading coefficients are continuous, then one may perform the discretization procedure in [36, Section 3.4.1]. Hence, the deterministic and random code capacities in the continuous case are also given by the limit infimum of the formulas (61) and (62), respectively.

#### IV. THE ARBITRARILY VARYING GAUSSIAN PRODUCT CHANNEL

From this point on, we consider Gaussian AVCs, without parameters. In this section, we consider the Gaussian product channel. Our results on the AVC with colored Gaussian noise, in the next section, are based on the capacity theorems of the AVC with fixed parameters, in the previous section, and on the analysis in the current section.

##### A. Channel Description

The state-dependent Gaussian product channel consists of a set of  $d$  parallel channels,

$$Y_j = X_j + S_j + Z_j, \quad j \in [1 : d], \quad (63)$$

where  $j$  is the channel index,  $d$  is the dimension (number of channels), and  $Z^d$  is a Gaussian vector with zero mean and covariance matrix  $K_Z$ . Let  $\mathbf{X}_j = (X_{j,i})_{i=1}^n$ ,  $\mathbf{S}_j = (S_{j,i})_{i=1}^n$  and  $\mathbf{Z}_j = (Z_{j,i})_{i=1}^n$  denote the input, state and noise sequences associated with the  $j$ th channel, respectively, where  $i \in [1 : n]$  is the time index, and let  $\mathbf{X}^d = (\mathbf{X}_j)_{j=1}^d$ ,  $\mathbf{S}^d = (\mathbf{S}_j)_{j=1}^d$  and  $\mathbf{Z}^d = (\mathbf{Z}_j)_{j=1}^d$ . The corresponding output of the product channel is the vector sequence  $\mathbf{Y}^d = \mathbf{X}^d + \mathbf{S}^d + \mathbf{Z}^d$ .

The Gaussian arbitrarily varying product channel (AVGPC) is a state-dependent Gaussian product channel with  $d$  state sequences  $(\mathbf{S}_1, \dots, \mathbf{S}_d)$  of unknown distribution, not necessarily independent nor stationary. That is,  $(\mathbf{S}_1, \dots, \mathbf{S}_d) \sim F_{\mathbf{S}_1, \dots, \mathbf{S}_d}$ ,

where  $F_{\mathbf{s}_1, \dots, \mathbf{s}_d}$  is an unknown joint cumulative distribution function (cdf) over  $\mathbb{R}^{nd}$ . In particular,  $F_{\mathbf{s}_1, \dots, \mathbf{s}_d}$  could give probability mass 1 to a particular sequence of state vectors  $(\mathbf{s}_1, \dots, \mathbf{s}_d) \in \mathbb{R}^{nd}$ . The channel is subject to input constraint  $\Omega > 0$  and state constraint  $\Lambda > 0$ ,

$$\begin{aligned} \sum_{j=1}^d \|\mathbf{X}_j\|^2 &\leq n\Omega \quad \text{w.p. } 1, \\ \sum_{j=1}^d \|\mathbf{S}_j\|^2 &\leq n\Lambda \quad \text{w.p. } 1. \end{aligned} \quad (64)$$

## B. Coding

We introduce preliminary definitions for the AVGPC.

*Definition 6 (Code).* A  $(2^{nR}, n)$  code for the AVGPC consists of the following; a message set  $[1 : 2^{nR}]$ , where it is assumed throughout that  $2^{nR}$  is an integer, a sequence of  $d$  encoding functions  $\mathbf{f}_j : [1 : 2^{nR}] \rightarrow \mathbb{R}^n$ , for  $j \in [1 : d]$ , such that

$$\sum_{j=1}^d \|\mathbf{f}_j(m)\|^2 \leq n\Omega, \quad \text{for } m \in [1 : 2^{nR}], \quad (65)$$

and a decoding function  $g : \mathbb{R}^{nd} \rightarrow [1 : 2^{nR}]$ . Given a message  $m \in [1 : 2^{nR}]$ , the encoder transmits  $\mathbf{x}_j = \mathbf{f}_j(m)$ , for  $j \in [1 : d]$ . The codeword is then given by  $\mathbf{x}^d = \mathbf{f}^d(m) \triangleq (\mathbf{f}_1(m), \mathbf{f}_2(m), \dots, \mathbf{f}_d(m))$ . The decoder receives the channel outputs  $\mathbf{y}^d = (\mathbf{y}_1, \dots, \mathbf{y}_d)$ , and finds an estimate of the message  $\hat{m} = g(\mathbf{y}^d)$ . We denote the code by  $\mathcal{C} = (\mathbf{f}^d, g)$ .

Define the conditional probability of error of a code  $\mathcal{C}$  given the sequence  $\mathbf{s}^d = (\mathbf{s}_1, \dots, \mathbf{s}_d)$  by

$$P_{e|\mathbf{s}^d}^{(n)}(\mathcal{C}) \triangleq \frac{1}{2^{nR}} \sum_{m=1}^{2^{nR}} \int_{\mathbf{y}^d \in \mathbb{R}^{nd} : g(\mathbf{y}^d) \neq m} d\mathbf{y}^d \cdot f_{\mathbf{Y}^d | m, \mathbf{s}^d}(\mathbf{y}^d), \quad (66)$$

where  $f_{\mathbf{Y}^d | m, \mathbf{s}^d}(\mathbf{y}^d) = \prod_{i=1}^n f_{Z^d}(y_i^d - \mathbf{f}_i^d(m) - \mathbf{s}_i^d)$ , with

$$f_{Z^d}(z^d) = \frac{1}{\sqrt{(2\pi)^d |K_Z|}} e^{-\frac{1}{2} z^d K_Z^{-1} (z^d)^T}. \quad (67)$$

A code  $\mathcal{C} = (\mathbf{f}^d, g)$  is called a  $(2^{nR}, n, \varepsilon)$  code for the AVGPC if

$$P_{e|\mathbf{s}^d}^{(n)}(\mathcal{C}) \leq \varepsilon, \quad \text{for all } \mathbf{s}^d \in \mathbb{R}^{nd} \text{ with } \sum_{j=1}^d \|\mathbf{s}_j\|^2 \leq n\Lambda. \quad (68)$$

We say that a rate  $R$  is achievable if for every  $\varepsilon > 0$  and sufficiently large  $n$ , there exists a  $(2^{nR}, n, \varepsilon)$  code for the AVGPC. The operational capacity is defined as the supremum of all achievable rates, and it is denoted by  $\mathbb{C}(K_Z)$ . We use the term ‘capacity’ referring to this operational meaning, and in some places we call it the deterministic code capacity to emphasize that achievability is measured with respect to deterministic codes.

We proceed now to coding schemes when using stochastic-encoder stochastic-decoder pairs with common randomness.

*Definition 7 (Random code).* A  $(2^{nR}, n)$  random code for the AVGPC consists of a collection of  $(2^{nR}, n)$  codes  $\{\mathcal{C}_\gamma = (\mathbf{f}_\gamma^d, g_\gamma)\}_{\gamma \in \Gamma}$ , along with a pmf  $\mu(\gamma)$  over the code collection  $\Gamma$ . We denote such a code by  $\mathcal{C}^\Gamma = (\mu, \Gamma, \{\mathcal{C}_\gamma\}_{\gamma \in \Gamma})$ . Analogously to the deterministic case, a  $(2^{nR}, n, \varepsilon)$  random code for the AVGPC satisfies

$$\sum_{\gamma \in \Gamma} \mu(\gamma) \sum_{j=1}^d \|\mathbf{f}_{\gamma, j}(m)\|^2 \leq n\Omega, \quad \text{for all } m \in [1 : 2^{nR}], \quad (69)$$

and

$$P_{e|\mathbf{s}^d}^{(n)}(\mathcal{C}^\Gamma) \triangleq \sum_{\gamma \in \Gamma} \mu(\gamma) P_{e|\mathbf{s}^d}^{(n)}(\mathcal{C}_\gamma) \leq \varepsilon \quad \text{for all } \mathbf{s}^d \in \mathbb{R}^{nd} \text{ with } \sum_{j=1}^d \|\mathbf{s}_j\|^2 \leq n\Lambda. \quad (70)$$

The capacity achieved by random codes is denoted by  $\mathbb{C}^*(K_Z)$ , and it is referred to as the *random code capacity*.

### C. Related Work

Consider the AVGPC with parallel Gaussian channels, where the covariance matrix of the additive noise is

$$\Sigma = \text{diag}\{\sigma_1^2, \dots, \sigma_d^2\}, \quad (71)$$

*i.e.*  $Z_1, \dots, Z_d$  are independent and  $Z_j \sim \mathcal{N}(0, \sigma_j^2)$ . Denote the random code capacity of the AVGPC with parallel channels by  $\mathbb{C}^*(\Sigma)$ . Hughes and Narayan [61] have shown that the solution for the random code capacity is given by “double” water filling, where the jammer performs water filling first, attempting to whiten the overall noise as much as possible, and then the user performs water filling taking into account the total noise power, which is contributed by both the channel and the jammer. The formal definitions are given below. Let

$$N_j^* = [\beta - \sigma_j^2]_+, \quad j \in [1 : d] \quad (72)$$

with  $[t]_+ = \max\{0, t\}$ , where  $\beta \geq 0$  is chosen to satisfy

$$\sum_{j=1}^d [\beta - \sigma_j^2]_+ = \Lambda. \quad (73)$$

Next, let

$$P_j^* = [\alpha - (N_j^* + \sigma_j^2)]_+, \quad j \in [1 : d], \quad (74)$$

where  $\alpha \geq 0$  is chosen to satisfy

$$\sum_{j=1}^d [\alpha - (N_j^* + \sigma_j^2)]_+ = \Omega. \quad (75)$$

We can now define Hughes and Narayan’s capacity formula [61],

$$\mathbb{C}^*(\Sigma) \triangleq \sum_{j=1}^d \frac{1}{2} \log \left( 1 + \frac{P_j^*}{N_j^* + \sigma_j^2} \right). \quad (76)$$

*Theorem 12* (see [61]). The random code capacity of the AVGPC is given by

$$\mathbb{C}^*(\Sigma) = \mathbb{C}^*(\Sigma). \quad (77)$$

### D. Observations on The Water Filling Game

We give further observations on the results by Hughes and Narayan [61], which will be useful in the sequel.

1) *Game Theoretic Interpretation:* By [61, Theorem 3], the random code capacity is the solution of the following optimization problem,

$$\min \max \sum_{j=1}^d \frac{1}{2} \log \left( 1 + \frac{P_j}{N_j + \sigma^2} \right), \quad (78)$$

where the minimization is over the simplex  $\mathcal{F}_{\text{state}} = \{(N_1, \dots, N_d) : \sum_{j=1}^d N_j \leq \Lambda\}$ , and the maximization is over the simplex  $\mathcal{F}_{\text{input}} = \{(P_1, \dots, P_d) : \sum_{j=1}^d P_j \leq \Omega\}$ .

The optimization problem is thus interpreted as a two-player zero-sum simultaneous game, played by the user and the jammer, where  $\mathcal{F}_{\text{input}}$  and  $\mathcal{F}_{\text{state}}$  are the respective action sets. The payoff function  $v : \mathcal{F}_{\text{input}} \times \mathcal{F}_{\text{state}} \rightarrow \mathbb{R}$  is defined such that, given a profile  $(P_1, \dots, P_d, N_1, \dots, N_d)$ ,

$$v(P_1, \dots, P_d, N_1, \dots, N_d) \triangleq \sum_{j=1}^d \frac{1}{2} \log \left( 1 + \frac{P_j}{N_j + \sigma^2} \right). \quad (79)$$

We have defined a game with pure strategies, *i.e.* the players’ actions are deterministic. In the communication model, the optimal coding and jamming scheme are random in general, yet the capacity can be achieved with deterministic power allocations, as in the game.

The optimal power allocation has a water filling analogy (see *e.g.* [27, Section 9.4]), where the jammer pours water of volume  $\Lambda$  to a vessel, and then the encoder pours more water of volume  $\Omega$ . The shape of the bottom of the vessel is determined by the noise variances  $\sigma_1^2, \dots, \sigma_d^2$ . The jammer brings the water level to  $\beta$ , and then the encoder brings the water level to  $\alpha$ . Water filling for the AVGPC is illustrated in Figure 1, for  $\Omega = 13$ ,  $\Lambda = 8$ ,  $d = 10$ ,  $(\sigma_j^2)_{j=1}^{10} = (5, 8, 3, 1.5, 2.5, 1.8, 3.2, 9, 4.5, 5.5)$ . The light shade “fluid” is the jammer’s water filling and the dark shade “fluid” is the transmitter’s. The resulting “water levels” are

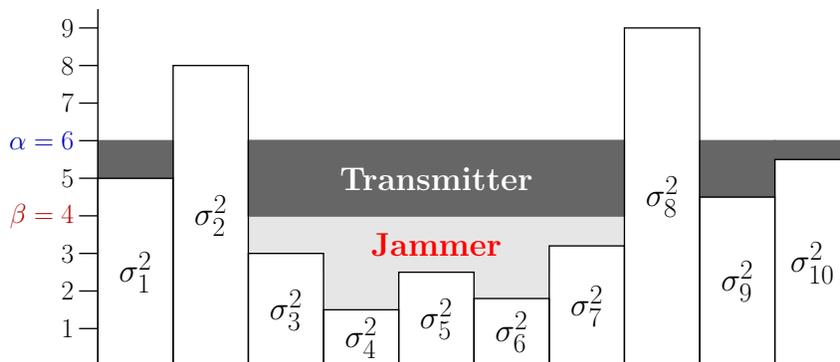


Fig. 1. Water filling for the AVGPC, for  $\Omega = 13$ ,  $\Lambda = 8$ ,  $d = 10$ ,  $(\sigma_j^2)_{j=1}^{10} = (5, 8, 3, 1.5, 2.5, 1.8, 3.2, 9, 4.5, 5.5)$ . The light shade “fluid” is the jammer’s water filling and the dark shade “fluid” is the transmitter’s. The resulting “water levels” are  $\beta = 4$  and  $\alpha = 6$ , hence  $(N_j^*)_{j=1}^{10} = (0, 0, 1, 2.5, 1.5, 2.2, 0.8, 0, 0, 0)$  and  $(P_j^*)_{j=1}^{10} = (1, 0, 2, 2, 2, 2, 2, 1.5, 0.5)$ .

$\beta = 4$  and  $\alpha = 6$ . Then, substituting into (72) and (74) yields the power allocations  $(N_j^*)_{j=1}^{10} = (0, 0, 1, 2.5, 1.5, 2.2, 0.8, 0, 0, 0)$  for the jammer and  $(P_j^*)_{j=1}^{10} = (1, 0, 2, 2, 2, 2, 2, 1.5, 0.5)$  for the transmitter.

One can easily prove the following properties of the random code capacity characterization.

**Lemma 13.** The quantities defined by (72)-(76) satisfy

$$\begin{aligned}
 & 1) \alpha > \beta & 2) N_j^* > 0 \Rightarrow P_j^* > 0 \forall j \in [1 : d] \\
 & 3) P_j^* + N_j^* + \sigma_j^2 = \max(\alpha, \sigma_j^2) & 4) C^*(\Sigma) = \sum_{j=1}^d \frac{1}{2} \log \frac{\max(\alpha, \sigma_j^2)}{\max(\beta, \sigma_j^2)}.
 \end{aligned} \tag{80}$$

For completeness, we give the proof of Lemma 13 is given in Appendix K. Based on the water filling analogy of the power allocation above, part 1 of Lemma 13 is natural, since  $\beta$  is interpreted as the water level after the jammer pours his share, and  $\alpha$  is interpreted as the water level after the user pours *additional* water after that (see Figure 1). Part 3 and part 4 are not surprising either since, as can be seen in Figure 1, the variance of the combined interference  $(Z_j + S_j)$  is  $\max(\beta, \sigma_j^2)$  and the variance of the channel output  $Y_j$  is  $\max(\alpha, \sigma_j^2)$ .

Observe that an equivalent statement of part 2 is the following. If the user discards a channel, *i.e.* assigns  $P_j^* = 0$  to the  $j$ th channel, then the jammer does not invest power in this channel either, *i.e.*  $N_j^* = 0$ . This claim is also intuitive, and from a game theoretic perspective, it is an aspect of the jammer’s rationality, as explained below. As mentioned above the optimization problem is interpreted as a two-player zero-sum simultaneous game between the user and the jammer. The value of such a game is attained by a pair of strategies which forms a Nash equilibrium [103] (see also [79][75, Theorem 3.1.4]). That is, if the user and the jammer were to agree to use the power allocation strategies  $(P_j^*)_{j=1}^d$  and  $(N_j^*)_{j=1}^d$ , then neither player could profit by deviating from his original strategy, provided that the other player respects the agreement. Now, suppose that for some  $j \in [1 : d]$ ,  $P_j^* = 0$  and  $N_j^* > 0$ . Then, the jammer is wasting energy, and can surely profit from diverging this energy to some other channel  $j'$  with  $P_{j'}^* > 0$ . Thus, such strategy profile is irrational and cannot be a Nash equilibrium.

For a general AVC, a coding scheme which assumes that the jammer is using his optimal strategy would typically fail. The code needs to be robust standing against any state sequence that satisfies the state constraint. For example, consider a scalar Gaussian AVC [60], specified by  $\mathbf{Y} = \mathbf{X} + \mathbf{S} + \mathbf{Z}$ , under input constraint  $\|\mathbf{X}\|^2 \leq n\Omega$  and state constraint  $\|\mathbf{S}\|^2 \leq n\Lambda$ , where the noise sequence  $\mathbf{Z}$  is i.i.d.  $\sim \mathcal{N}(0, \sigma^2)$ . Suppose that the receiver is using joint typicality decoding for a Gaussian channel  $\mathbf{Y} = \mathbf{X} + \mathbf{V}$ , where  $\mathbf{V}$  is i.i.d.  $\sim \mathcal{N}(0, \Lambda + \sigma^2)$  (see [27, Section 9.1]), corresponding to the optimal jamming strategy. Then, the jammer can fail the decoder by selecting a state sequence such that  $\|\mathbf{S}\|^2 = \frac{n\Lambda}{2}$ , for instance. As a result, there is a high probability that the square norm of the output sequence is below  $n(\Lambda + \sigma^2 - \delta)$ , for small  $\delta > 0$ , in which case the decoder cannot establish joint typicality and declares an error. The same principle holds in our problem. The user cannot assume that the jammer is using his optimal power allocation, and a reliable code must be robust standing against any power allocation of the jammer.

2) *Multiple Access Channel Analogy:* Water filling in two (or more) stages appears in other settings in the literature, *e.g.* [26, 71, 111, 113]. Consider a Gaussian product multiple access channel (MAC), where  $Y_j = X_{1,j} + X_{2,j} + Z_j$ ,  $j \in [1 : d]$ , under the input constraints  $\|\mathbf{X}_1^d\|^2 \leq n\Omega$  and  $\|\mathbf{X}_2^d\|^2 \leq n\Lambda$ . This can be viewed as a different variation of the AVGPC where a second transmitter replaces the jammer. By [26], a corner point of the capacity region can be achieved by applying water filling to the total power in the first step, and then to the power of User 2 in the second step. Specifically, by [26, Section

III.B.], the optimal power allocations  $(P_j^*)_{j=1}^d$  and  $(N_j^*)_{j=1}^d$ , for Encoder 1 and Encoder 2, respectively, which achieve a corner point of the capacity region, satisfy

$$P_j^* + N_j^* = [\alpha - \sigma_j^2]_+, \quad j \in [1 : d], \quad (81)$$

such that  $\sum_{j=1}^d (P_j^* + N_j^*) = \Omega + \Lambda$ , and

$$N_j^* = [\beta - \sigma_j^2]_+, \quad j \in [1 : d], \quad (82)$$

such that  $\sum_{j=1}^d N_j^* = \Lambda$ . Following part 3 of Lemma 13, it can be seen that the strategy above is equivalent to (72)-(75). The total power allocation in (81) seems natural in order to maximize the sum rate. Though, our presentation in (72)-(75) is intuitive for the Gaussian product MAC as well. Indeed, using successive cancellation decoding, the receiver estimates the transmission of User 1 while treating the transmission of User 2 as noise, and then subtracts the estimated sequence from the received sequence to decode the transmission of User 2. Hence, decoding for User 1 is analogous to the decoder in our problem. Nevertheless, in the next section, we show that the deterministic code capacity in our adversarial problem has a different behavior.

Another water filling game is described by Lai and El Gamal in [71], who considered the flat fading MAC  $Y = h_1 X_1 + h_2 X_2 + Z$  with selfish users, where the fading coefficients are continuous random variables, distributed according to  $(h_1, h_2) \sim \mu$ . Suppose that the users are subject to average input constraints,  $\mathbb{E}_\mu \|\mathbf{X}_1\|^2 \leq n\Omega$  and  $\mathbb{E}_\mu \|\mathbf{X}_2\|^2 \leq n\Lambda$ . As shown in [71], a maximum sum-rate point on the capacity region boundary is achieved if the users perform water filling treating each other's transmission as noise. It is further shown that opportunistic communication is optimal, where User 1 only transmits if his water level times fading coefficient is at least as high as that of User 2, and vice versa. That is, the power allocations of the users are given by

$$\begin{aligned} P_{h_1, h_2}^* &= \begin{cases} [\beta_1 - \sigma^2/h_1]_+ & \text{if } \beta_1 h_1 \geq \beta_2 h_2, \\ 0 & \text{otherwise} \end{cases}, \\ N_{h_1, h_2}^* &= \begin{cases} [\beta_2 - \sigma_j^2/h_2]_+ & \text{if } \beta_1 h_1 \leq \beta_2 h_2, \\ 0 & \text{otherwise} \end{cases}, \end{aligned} \quad (83)$$

where  $\beta_1$  and  $\beta_2$  are chosen such that  $\mathbb{E} P_{h_1, h_2}^* = \Omega$  and  $\mathbb{E} N_{h_1, h_2}^* = \Lambda$ . This threshold operation resembles the result in the next section, on the deterministic code capacity of the AVGPC, except that the phase transition of the AVGPC depends only on the ‘‘water volumes’’  $\Omega$  and  $\Lambda$  (see Subsection IV-F).

### E. Results

We give our result on the AVGPC with parallel Gaussian channels, where the covariance matrix of the additive noise is  $\Sigma = \text{diag}\{\sigma_1^2, \dots, \sigma_d^2\}$ , i.e.  $Z_1, \dots, Z_d$  are independent and  $Z_j \sim \mathcal{N}(0, \sigma_j^2)$ . The deterministic code capacity of the AVGPC with parallel channels is denoted by  $\mathbb{C}(\Sigma)$ .

We establish the capacity of the AVGPC. Based on Csiszár and Narayan's result in [30], the deterministic code capacity of an AVC under input and state constraints is given in terms of channel symmetrizability and the minimal state cost for the jammer to symmetrize the channel (see also [73] [82, Definition 5 and Theorem 5]). By [30, Definition 2], a AVGPC is symmetrized by a conditional pdf  $\varphi(s^d|x^d)$  if

$$\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \varphi(s^d|x_2^d) f_{Z^d}(y^d - x_1^d - s^d) ds^d = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \varphi(s^d|x_1^d) f_{Z^d}(y^d - x_2^d - s^d) ds^d, \quad \forall x_1^d, x_2^d, y^d \in \mathbb{R}^d, \quad (84)$$

where  $f_{Z^d}(z^d) = \prod_{j=1}^d \frac{1}{\sqrt{2\pi\sigma_j^2}} e^{-z_j^2/2\sigma_j^2}$ . In particular, observe that (84) holds for  $\varphi(s^d|x^d) = \delta(s^d - x^d)$ , where  $\delta(\cdot)$  is the Dirac delta function. In other words, the channel is symmetrized by a distribution  $\varphi(s^d|x^d)$  which gives probability 1 to  $S^d = x^d$ . For the AVGPC, the minimal state cost for the jammer to symmetrize the channel, for an input distribution  $f_{X^d}$ , is given by

$$\tilde{\Lambda}(F_{X^d}) = \min \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f_{X^d}(x^d) \varphi(s^d|x^d) \|s^d\|^2 ds^d dx^d, \quad (85)$$

where the minimization is over all conditional pdfs  $\varphi(s^d|x^d)$  that symmetrize the channel, that is, satisfy (84). The following lemma states that the minimal state cost for symmetrizability is the same as the input power. The lemma will be used in the achievability proof of the capacity theorem.

*Lemma 14.* For a zero mean Gaussian vector  $X^d \sim \mathcal{N}(\mathbf{0}, K_X)$ ,

$$\tilde{\Lambda}(F_{X^d}) = \text{tr}(K_X). \quad (86)$$

The proof of Lemma 14 is given in Appendix L. The proof builds on our observation that (84) holds if and only if  $\varphi(s^d|x^d) = \varphi(s^d - x^d|0)$ . This in turn leads to the conclusion that the minimum in (85) is attained by  $\varphi_{x^d}(s^d) = \delta(s^d - x^d)$ . Moving to the capacity theorem, define

$$\mathbb{C}(\Sigma) = \begin{cases} \mathbb{C}^*(\Sigma) & \text{if } \Omega > \Lambda, \\ 0 & \text{otherwise.} \end{cases} \quad (87)$$

*Theorem 15.* The deterministic code capacity of the AVGPC is given by

$$\mathbb{C}(\Sigma) = \mathbb{C}(\Sigma). \quad (88)$$

The proof of Theorem 15 is given in Appendix M. Considering the scalar case, Csiszár and Narayan showed the direct part by providing a coding scheme for the Gaussian AVC [32]. While the receiver in their coding scheme uses simple minimum-distance decoding, the analysis is fairly complicated. Here, on the other hand, we treat the AVGPC using a much simpler approach. To prove direct part, we consider the optimization problem based on the capacity formula of the general AVC under input and state constraints, which is given in terms of symmetrizing state distributions. We use Lemma 14 to show that if  $\Omega > \Lambda$ , then the transmitter’s water filling strategy in (74) guarantees that  $\tilde{\Lambda}(F_{x^d}) > \Lambda$ . Intuitively, this means that the jammer cannot symmetrize the channel without violating the state constraint. In this scenario, the random code capacity can be achieved with deterministic codes as well.

#### F. Discussion

We give a couple of remarks on our result in Theorem 15. As in the case of the Gaussian scalar AVC [32], the capacity is discontinuous in the input constraint, and has a phase transition behavior, depending on whether  $\Omega > \Lambda$  or  $\Omega \leq \Lambda$ . We give an intuitive explanation below. For the classic Gaussian AVC, reliable communication requires the power of the transmitted signal to be higher than the power of the jamming signal, otherwise the jammer can confuse the receiver by making the state sequence  $\mathbf{S}$  “look like” the input sequence  $\mathbf{X}$  [32]. At a first glance at our problem, one might have expected that the input power  $P_j$  of the  $j$ th channel also needs to be higher than the jamming power  $N_j$ , in order for the output  $\mathbf{Y}_j$  to be useful. This is not the case. Since the decoder has the vector of outputs  $(\mathbf{Y}_1, \dots, \mathbf{Y}_d)$ , even if  $\mathbf{S}_j$  looks like  $\mathbf{X}_j$ , the receiver could still gain information from  $\mathbf{Y}_j$  as the other outputs may “break the symmetry”.

Based on Shannon’s classic water filling result [94], the capacity of the Gaussian product channel,  $Y_j = X_j + V_j$ ,  $j \in [1 : d]$ , can be achieved by combining  $d$  independent encoder-decoder pairs, where the  $j$ th pair is associated with a capacity achieving code for the scalar Gaussian channel under input constraint  $P_j^*$ . However, based on Csiszár and Narayan’s result on the Gaussian single AVC [32], the capacity of the  $j$ th AVC,  $Y_j = X_j + S_j + Z_j$ , is zero under input constraint  $P_j^*$  and state constraint  $N_j^*$  for  $P_j^* \leq N_j^*$ . This means that, in contrast to the Shannon’s Gaussian product channel [94], using  $d$  independent encoder-decoder pairs over the AVGPC is suboptimal in general. This can be viewed as a constrained version of the super-additivity phenomenon in [91].

## V. MAIN RESULTS – AVC WITH COLORED GAUSSIAN NOISE

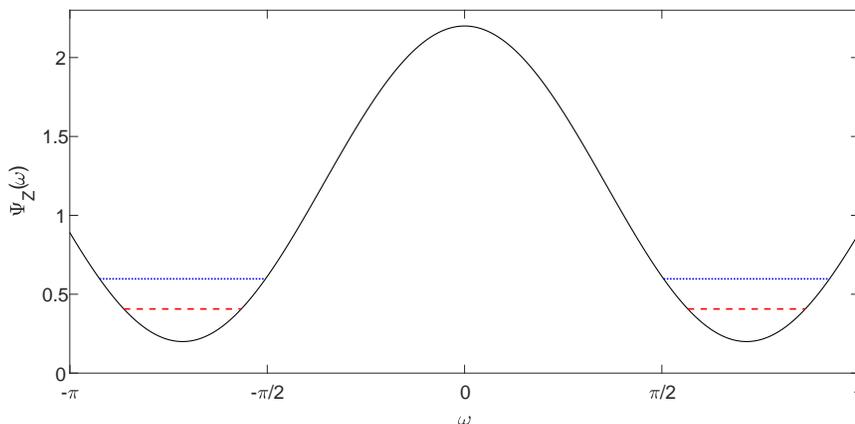


Fig. 2. Water filling in the frequency domain for the AVC with colored Gaussian noise. The curve depicts the power spectral density  $\Psi_Z(\omega)$  of the noise process  $Z^n$ . The red dashed line indicates the “water level”  $\beta$  which corresponds to the jammer’s water filling, and the blue dotted line indicates the “water level”  $\alpha$  which corresponds to the transmitter’s water filling.

We consider an AVC with colored Gaussian noise, *i.e.*

$$\mathbf{Y} = \mathbf{X} + \mathbf{Z} + \mathbf{S}, \quad (89)$$

where  $\mathbf{Z}$  is a zero mean stationary Gaussian process, with power spectral density  $\Psi_Z(\omega)$ . Assume that the power spectral density is bounded and integrable. We denote the random code capacity and the deterministic code capacity of this channel by  $\mathbb{C}^*(\Psi_Z)$  and  $\mathbb{C}(\Psi_Z)$ , respectively.

We show that the optimal power allocations of the user and the jammer are given by “double” water filling in the frequency domain. Define

$$b^*(\omega) = [\beta - \Psi_Z(\omega)]_+, \quad -\pi \leq \omega \leq \pi, \quad (90)$$

where  $\beta \geq 0$  is chosen to satisfy

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} [\beta - \Psi_Z(\omega)]_+ d\omega = \Lambda. \quad (91)$$

Next, define

$$a^*(\omega) = [\alpha - (b^*(\omega) + \Psi_Z(\omega))]_+, \quad -\pi \leq \omega \leq \pi, \quad (92)$$

where  $\alpha \geq 0$  is chosen to satisfy

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} [\alpha - (b^*(\omega) + \Psi_Z(\omega))]_+ d\omega = \Omega. \quad (93)$$

Now, let

$$\mathbb{C}^*(\Psi_Z) \triangleq \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{2} \log \left( 1 + \frac{a^*(\omega)}{b^*(\omega) + \Psi_Z(\omega)} \right) d\omega. \quad (94)$$

*Theorem 16.* The random code capacity of the AVC with colored Gaussian noise is given by

$$\mathbb{C}^*(\Psi_Z) = \mathbb{C}^*(\Psi_Z), \quad (95)$$

and the deterministic code capacity is given by

$$\mathbb{C}(\Psi_Z) = \begin{cases} \mathbb{C}^*(\Psi_Z) & \text{if } \Omega > \Lambda, \\ 0 & \text{otherwise.} \end{cases} \quad (96)$$

The proof of Theorem 16 is given in Appendix N, combining our previous results on the AVC with fixed parameters and the AVGPC. Despite the common belief that the characterization for a channel with colored Gaussian noise easily follows from the results for the product channel setting, the analysis is more involved. While standard orthogonalization transforms the channel into an equivalent one with statistically independent noise instances, the noise in the transformed channel is not necessarily white. As the noise variance may change over time, we observe that the transformed channel is in fact an AVC with fixed parameters which represent the sequence of noise variances. Using Corollary 5 and Corollary 11, we obtain deterministic and random capacity formulas that are analogous to those of the AVGPC, and use Toeplitz matrix properties to express the formulas as integrals in the frequency domain.

The optimal power allocation has a water filling analogy in the frequency domain (see *e.g.* [27, Section 9.5]), where the jammer pours water of volume  $\Lambda$  on top of the power spectral density  $\Psi_Z(\omega)$ , and then the encoder pours more water of volume  $\Omega$ . The jammer brings the water level to  $\beta$ , and then the encoder brings the water level to  $\alpha$ . The process is illustrated in Figure 2.

#### APPENDIX A PROOF OF THEOREM 1

Consider the compound channel  $\mathcal{W}^{\mathcal{Q}}$  with fixed parameters under input constraint  $\Omega$  and state constraint  $\Lambda$ .

### A. Achievability Proof

To show achievability, we construct a code based on conditional typicality decoding with respect to a channel state type, which is “close” to one of the state distributions in  $\mathcal{Q}$ .

Denote the type of the parameter sequence by  $P_T = \hat{P}_{\theta^n}$ . Define a set  $\hat{\mathcal{Q}}_n$  of conditional state types,

$$\hat{\mathcal{Q}}_n = \left\{ \hat{P}_{s^n|\theta^n} : (\theta^n, s^n) \in \mathcal{A}_{\delta_1}^{(n)}(P_T \times q), \text{ for some } q \in \mathcal{Q} \right\}, \quad (97)$$

with  $(P_T \times q)(t, s) = P_T(t)q(s|t)$ , and

$$\delta_1 \triangleq \frac{\delta}{2 \cdot |\mathcal{S}|}, \quad (98)$$

where  $\delta > 0$  is arbitrarily small. In words,  $\hat{\mathcal{Q}}_n$  is the set of conditional types  $q'(s|t)$ , given a parameter sequence  $\theta^n$ , such that the joint type is  $\delta_1$ -close to  $P_T(t)q(s|t)$ , for some conditional state distribution  $q(s|t)$  in  $\mathcal{Q}$ . We note that the sets  $\mathcal{Q}$  and  $\hat{\mathcal{Q}}_n$  could be disjoint, since  $\mathcal{Q}$  is not limited to conditional empirical distributions. Nevertheless, for a fixed  $\delta > 0$  and sufficiently large  $n$ , every  $q \in \mathcal{Q}$  can be approximated by some  $q' \in \hat{\mathcal{Q}}_n$ . Indeed, for sufficiently large  $n$ , there exists a joint type  $P'_T(t)q'(s|t)$  such that  $|P'_T(t)q'(s|t) - P_T(t)q(s|t)| \leq \delta_1/|\mathcal{S}|$ , hence  $|P'_T(t) - P_T(t)| \leq \delta_1$  and  $|P_T(t)q'(s|t) - P_T(t)q(s|t)| \leq \delta_1 q'(s|t) \leq \delta_1$ . Now, a code is constructed as follows.

*Codebook Generation:* Fix  $P_{X|T}$  such that  $\mathbb{E}\phi(X) \leq \Omega - \varepsilon$ , where

$$\mathbb{E}\phi(X) = \sum_{t \in \mathcal{T}} P_T(t) \mathbb{E}(\phi(X)|T=t) = \frac{1}{n} \sum_{i=1}^n \sum_{x \in \mathcal{X}} P_{X|T}(x|\theta_i) \phi(x). \quad (99)$$

Generate  $2^{nR}$  independent sequences at random,  $x^n(m, \theta^n) \sim \prod_{i=1}^n P_{X|T}(x_i|\theta_i)$ , for  $m \in [1 : 2^{nR}]$ .

*Encoding:* To send a message  $m$ , if  $\phi^n(x^n(m, \theta^n)) \leq \Omega$ , transmit  $x^n(m, \theta^n)$ . Otherwise, transmit an idle sequence  $x^n = (a, a, \dots, a)$  with  $\phi(a) = 0$ .

*Decoding:* Find a unique  $\hat{m} \in [1 : 2^{nR}]$  for which there exists  $q \in \hat{\mathcal{Q}}_n$  such that  $(\theta^n, x^n(\hat{m}, \theta^n), y^n) \in \mathcal{A}_{\delta}^{(n)}(P_T P_{X,Y|T}^q)$ , where

$$P_{X,Y|T}^q(x, y|t) = P_{X|T}(x|t) \sum_{s \in \mathcal{S}} q(s|t) W_{Y|X,S,T}(y|x, s, t). \quad (100)$$

If there is none, or more than one such  $\hat{m}$ , declare an error. We note that using the set of types  $\hat{\mathcal{Q}}_n$  instead of the original set of state distributions  $\mathcal{Q}$  alleviates the analysis, since  $\mathcal{Q}$  is not necessarily finite nor countable.

*Analysis of Probability of Error:* Assume without loss of generality that the user sent  $M = 1$ . By the union of events bound, we have that  $\Pr(\hat{M} \neq 1) \leq \Pr(\mathcal{E}_1) + \Pr(\mathcal{E}_2 | \mathcal{E}_1^c) + \Pr(\mathcal{E}_3 | \mathcal{E}_1^c)$ , where

$$\begin{aligned} \mathcal{E}_1 &= \{(\theta^n, X^n(1, \theta^n)) \notin \mathcal{A}_{\delta}^{(n)}(P_T P_{X|T})\}, \\ \mathcal{E}_2 &= \{(\theta^n, X^n(1, \theta^n), Y^n) \notin \mathcal{A}_{\delta}^{(n)}(P_T P_{X|T} P_{Y|X,T}^q) \text{ for all } q' \in \hat{\mathcal{Q}}_n\}, \\ \mathcal{E}_3 &= \{(\theta^n, X^n(m, \theta^n), Y^n) \in \mathcal{A}_{\delta}^{(n)}(P_T P_{X|T} P_{Y|X,T}^q) \text{ for some } m \neq 1, q' \in \hat{\mathcal{Q}}_n\}. \end{aligned} \quad (101)$$

The first term tends to zero exponentially by the law of large numbers and Chernoff's bound (see e.g. [67, Theorem 1.2]). Now, suppose that the event  $\mathcal{E}_1^c$  occurs. Then, for sufficiently small  $\delta$ , we have that  $\phi^n(X^n(1, \theta^n)) \leq \Omega$ , since  $\mathbb{E}\phi(X) \leq \Omega - \varepsilon$ . Hence,  $X^n(1, \theta^n)$  is the channel input.

Next, we claim that the second error event implies that  $(\theta^n, X^n(1, \theta^n), Y^n) \notin \mathcal{A}_{\delta/2}^{(n)}(P_T P_{X|T} P_{Y|X,T}^q)$ , where  $q(s|t)$  is the *actual* state distribution chosen by the jammer. Assume to the contrary that  $\mathcal{E}_2$  holds, but  $(\theta^n, X^n(1, \theta^n), Y^n) \in \mathcal{A}_{\delta/2}^{(n)}(P_T P_{X|T} P_{Y|X,T}^q)$ . For sufficiently large  $n$ , there exists a conditional type  $q' \in \hat{\mathcal{Q}}_n$  that approximates  $q$  in the sense that  $|P_T(t)q'(s|t) - P_T(t)q(s|t)| \leq \delta_1$  for all  $s \in \mathcal{S}$  and  $t \in \mathcal{T}$ , hence

$$|P_T(t)P_{Y|X,T}^{q'}(y|x, t) - P_T(t)P_{Y|X,T}^q(y|x, t)| \leq |\mathcal{S}| \cdot \delta_1 = \frac{\delta}{2}, \quad (102)$$

for all  $x \in \mathcal{X}$ ,  $t \in \mathcal{T}$ ,  $y \in \mathcal{Y}$  (see (98)-(100)). To show  $\delta$ -typicality with respect to  $q'(s|t)$ , we observe that

$$\begin{aligned}
& \left| \hat{P}_{\theta^n, X^n(1, \theta^n), Y^n}(t, x, y) - P_T(t)P_{X|T}(x|t)P_{Y|X,T}^{q'}(y|x, t) \right| \\
&= \left| \hat{P}_{\theta^n, X^n(1, \theta^n), Y^n}(t, x, y) - P_T(t)P_{X|T}(x|t)P_{Y|X,T}^q(y|x, t) + P_T(t)P_{X|T}(x|t)P_{Y|X,T}^q(y|x, t) \right. \\
&\quad \left. - P_T(t)P_{X|T}(x|t)P_{Y|X,T}^{q'}(y|x, t) \right| \\
&\leq \left| \hat{P}_{\theta^n, X^n(1, \theta^n), Y^n}(t, x, y) - P_T(t)P_{X|T}(x|t)P_{Y|X,T}^q(y|x, t) \right| \\
&\quad + \left| P_T(t)P_{X|T}(x|t)P_{Y|X,T}^q(y|x, t) - P_T(t)P_{X|T}(x|t)P_{Y|X,T}^{q'}(y|x, t) \right| \\
&\leq \frac{\delta}{2} + \frac{\delta}{2}P_{X|T}(x|t) \leq \delta,
\end{aligned} \tag{103}$$

where the first inequality is due to the triangle inequality, and the second inequality follows from (102) and the assumption that  $(\theta^n, X^n(1, \theta^n), Y^n) \in \mathcal{A}_{\delta/2}^{(n)}(P_T P_{X|T} P_{Y|X,T}^q)$ . It follows that  $(\theta^n, X^n(1, \theta^n), Y^n) \in \mathcal{A}_{\delta}^{(n)}(P_T P_{X|T} P_{Y|X,T}^{q'})$ , and  $\mathcal{E}_2$  does not hold. Thus,

$$\Pr(\mathcal{E}_2 \mid \mathcal{E}_1^c) \leq \Pr\left((\theta^n, X^n(1, \theta^n), Y^n) \notin \mathcal{A}_{\delta/2}^{(n)}(P_T P_{X|T} P_{Y|X,T}^q)\right). \tag{104}$$

This tends to zero exponentially as  $n \rightarrow \infty$  by the law of large numbers and Chernoff's bound (see e.g. [67, Theorem 1.2]).

Moving to the third error event, as the number of type classes in  $\mathcal{S}^n$  is bounded by  $(n+1)^{|\mathcal{S}|}$ , we have that

$$\Pr(\mathcal{E}_3 \mid \mathcal{E}_1^c) \leq (n+1)^{|\mathcal{S}|} \cdot \sup_{q' \in \hat{\mathcal{Q}}_n} \Pr\left((\theta^n, X^n(m, \theta^n), Y^n) \in \mathcal{A}_{\delta}^{(n)}(P_T P_{X|T} P_{Y|X,T}^{q'}) \text{ for some } m \neq 1\right). \tag{105}$$

For every  $m \neq 1$ ,  $X^n(m, \theta^n)$  is independent of  $Y^n$ , hence

$$\begin{aligned}
& \Pr\left((\theta^n, X^n(m), Y^n) \in \mathcal{A}_{\delta}^{(n)}(P_T P_{X|T} P_{Y|X,T}^q)\right) \\
&= \sum_{x^n \in \mathcal{X}^n} P_{X^n|T^n}(x^n|\theta^n) \sum_{y^n : (\theta^n, x^n, y^n) \in \mathcal{A}_{\delta}^{(n)}(P_T P_{X|T} P_{Y|X,T}^q)} P_{Y^n|T^n}^q(y^n|\theta^n).
\end{aligned} \tag{106}$$

Let  $(\theta^n, x^n, y^n) \in \mathcal{A}_{\delta}^{(n)}(P_T P_{X|T} P_{Y|X,T}^q)$ . Then,  $(\theta^n, y^n) \in \mathcal{A}_{\delta/2}^{(n)}(P_T P_{Y|T}^q)$  with  $\delta/2 \triangleq |\mathcal{X}| \cdot \delta$ . By Lemmas 2.6-2.7 in [29],

$$P_{Y^n|T^n}^q(y^n|\theta^n) = 2^{-n(H(\hat{P}_{y^n|t|\theta^n}) + D(\hat{P}_{y^n|\theta^n} \| P_{Y|T}))} \leq 2^{-nH(\hat{P}_{y^n|\theta^n})} \leq 2^{-n(H_{q'}(Y|T) - \varepsilon_1(\delta))}, \tag{107}$$

where  $\varepsilon_1(\delta) \rightarrow 0$  as  $\delta \rightarrow 0$ . Therefore, by (105)–(107),

$$\begin{aligned}
\Pr(\mathcal{E}_3) &\leq (n+1)^{|\mathcal{S}|} \cdot \sup_{q' \in \hat{\mathcal{Q}}_n} 2^{nR} \\
&\quad \sum_{x^n \in \mathcal{X}^n} P_{X^n|T^n}(x^n|\theta^n) \cdot |\{y^n : (\theta^n, x^n, y^n) \in \mathcal{A}_{\delta}^{(n)}(P_T P_{X|T} P_{Y|X,T}^q)\}| \cdot 2^{-n(H_{q'}(Y|T) - \varepsilon_1(\delta))} \\
&\leq \sup_{q' \in \hat{\mathcal{Q}}_n} (n+1)^{|\mathcal{S}|} 2^{-n[I_{q'}(X;Y|T) - R - \varepsilon_2(\delta)]},
\end{aligned} \tag{108}$$

with  $\varepsilon_2(\delta) \rightarrow 0$  as  $\delta \rightarrow 0$ , where the last inequality is due to [29, Lemma 2.13]. The RHS of (108) tends to zero exponentially as  $n \rightarrow \infty$ , provided that  $R < I_{q'}(X;Y|T) - \varepsilon_2(\delta)$ . The probability of error, averaged over the class of codebooks, exponentially decays to zero as  $n \rightarrow \infty$ . Therefore, there must exist a  $(2^{nR}, n, e^{-an})$  deterministic code, for a sufficiently large  $n$ . This completes the proof of the direct part.

### B. Converse Proof

Since the deterministic code capacity is always bounded by the random code capacity, we consider a sequence of  $(2^{nR}, n, \alpha_n)$  random codes, where  $\alpha_n \rightarrow 0$  as  $n \rightarrow \infty$ . Then, let  $X^n = f_{\gamma}^n(M, \theta^n)$  be the channel input sequence, and  $Y^n$  be the corresponding output sequence, where  $\gamma \in \Gamma$  is the random element shared between the encoders and the decoder. For every  $q \in \mathcal{Q}$ , we have by Fano's inequality that  $H_q(M|Y^n, T^n = \theta^n, \gamma) \leq n\varepsilon_n$ , hence

$$\begin{aligned}
nR &= H(M|T^n = \theta^n, \gamma) = I_q(M; Y^n|T^n = \theta^n, \gamma) + H(M|Y^n, T^n = \theta^n, \gamma) \\
&\leq I_q(M, \gamma; Y^n|T^n = \theta^n) + n\varepsilon_n = I_q(M, \gamma, X^n; Y^n|T^n = \theta^n) + n\varepsilon_n \\
&= I_q(X^n; Y^n|T^n = \theta^n) + n\varepsilon_n,
\end{aligned} \tag{109}$$

where  $\varepsilon_n \rightarrow 0$  as  $n \rightarrow \infty$ . The third equality holds since  $X^n$  is a deterministic function of  $(M, \gamma, \theta^n)$ , and the last equality since  $(M, \gamma) \ominus (X^n, T^n) \ominus Y^n$  form a Markov chain. It follows that

$$R - \varepsilon_n \leq \frac{1}{n} \sum_{i=1}^n I_q(X_i; Y_i | T_i = \theta_i) = I_q(X; Y | T, K) \leq I_q(X, K; Y | T) \quad (110)$$

for all  $q \in \mathcal{Q}$ , with  $X \equiv X_K$ ,  $Y \equiv Y_K$ ,  $T \equiv T_K = \theta_K$ , where the random variable  $K$  is uniformly distributed over  $[1 : n]$ , and  $\varepsilon_n \rightarrow 0$  as  $n \rightarrow \infty$ . Observe that the random variable  $T$  is distributed according to

$$P_T(t) = \Pr(\theta_K = t) = \sum_{i: \theta_i = t} \Pr(K = i) = \frac{1}{n} \cdot N(t | \theta^n) = \hat{P}_{\theta^n}(t), \quad (111)$$

where  $N(t | \theta^n)$  is the number of occurrences of the symbol  $t \in \mathcal{T}$  in the sequence  $\theta^n$ . Since  $K \ominus (T, X) \ominus Y$  form a Markov chain, we have that

$$R - \varepsilon_n \leq \inf_{q \in \mathcal{Q}} I_q(K, X; Y | T) = \inf_{q \in \mathcal{Q}} I_q(X; Y | T). \quad (112)$$

□

## APPENDIX B PROOF OF LEMMA 2

We state the proof of our modified version of Ahlswede's RT [6]. The proof follows the lines of [6, Subsection IV-B], which we modify here to include a constraint on the family of state distributions  $q(s)$  and the parameter sequence  $\theta^n$ . Let  $\tilde{s}^n \in \mathcal{S}^n$  such that  $l^n(\tilde{s}^n) \leq \Lambda$ . Denote the conditional type of  $\tilde{s}^n \in \mathcal{S}^n$  given  $\theta^n$  by  $\hat{q}(s | t)$ . Observe that  $\hat{q} \in \overline{\mathcal{P}}_\Lambda(\mathcal{S} | \theta^\infty)$  (see (9)), since  $\frac{1}{n} \sum_{i=1}^n \sum_{s \in \mathcal{S}} q(s | \theta_i) l(s) = l^n(\tilde{s}^n)$ .

Given a permutation  $\pi \in \Pi(\theta^n)$ ,

$$\sum_{s^n \in \mathcal{S}^n} q^n(s^n | \theta^n) h(s^n, \theta^n) = \sum_{s^n \in \mathcal{S}^n} q^n(\pi s^n | \theta^n) h(\pi s^n, \theta^n) = \sum_{s^n \in \mathcal{S}^n} q^n(\pi s^n | \pi \theta^n) h(\pi s^n, \pi \theta^n) = \sum_{s^n \in \mathcal{S}^n} q^n(s^n | \theta^n) h(\pi s^n, \pi \theta^n), \quad (113)$$

where the first equality holds since  $\pi$  is a bijection, the second equality holds since  $\pi \theta^n = \theta^n$  for every  $\pi \in \Pi(\theta^n)$ , and the last equality holds due to the product form of the conditional distribution  $q^n(s^n | t^n) = \prod_{i=1}^n q(s_i | t_i)$ . Hence, taking  $q = \hat{q}$ ,

$$\sum_{s^n \in \mathcal{S}^n} \hat{q}^n(s^n | \theta^n) h(s^n, \theta^n) = \frac{1}{|\Pi(\theta^n)|} \sum_{\pi \in \Pi(\theta^n)} \sum_{s^n \in \mathcal{S}^n} \hat{q}^n(s^n | \theta^n) h(\pi s^n, \pi \theta^n), \quad (114)$$

and by (17),

$$\sum_{s^n \in \mathcal{S}^n} \hat{q}^n(s^n | \theta^n) \left[ \frac{1}{|\Pi(\theta^n)|} \sum_{\pi \in \Pi(\theta^n)} h(\pi s^n, \pi \theta^n) \right] \leq \alpha_n. \quad (115)$$

Thus,

$$\sum_{s^n: \hat{P}_{s^n | \theta^n} = \hat{q}} \hat{q}^n(s^n | \theta^n) \left[ \frac{1}{|\Pi(\theta^n)|} \sum_{\pi \in \Pi(\theta^n)} h(\pi s^n, \pi \theta^n) \right] \leq \alpha_n. \quad (116)$$

As the expression in the square brackets is identical for all sequences  $s^n$  of conditional type  $\hat{q}$ , we have that

$$\left[ \frac{1}{|\Pi(\theta^n)|} \sum_{\pi \in \Pi(\theta^n)} h(\pi \tilde{s}^n, \pi \theta^n) \right] \cdot \sum_{s^n: \hat{P}_{s^n | \theta^n} = \hat{q}} \hat{q}^n(s^n | \theta^n) \leq \alpha_n. \quad (117)$$

The second sum is the probability of the conditional type class of  $\hat{q}$ , hence

$$\sum_{s^n: \hat{P}_{s^n | \theta^n} = \hat{q}} \hat{q}^n(s^n | \theta^n) \geq \frac{1}{(n+1)^{|\mathcal{S}| |\mathcal{T}|}}, \quad (118)$$

by [27, Theorem 11.1.4]. The proof follows from (117) and (118). □

## APPENDIX C PROOF OF THEOREM 3

Consider the AVC  $\mathcal{W}$  with fixed parameters under input constraint  $\Omega$  and state constraint  $\Lambda$ .

### A. Achievability Proof

To prove the random code capacity theorem for the AVC with fixed parameters, we use our result on the compound channel along with our modified Robustification Technique (RT), *i.e.* Lemma 2.

Let  $R < C^*$ . At first, we consider the compound channel under input constraint  $\Omega$ , with  $\mathcal{Q} = \overline{\mathcal{P}}_\Lambda(\mathcal{S}|\theta^\infty)$ . According to Lemma 1, for some  $\delta > 0$  and sufficiently large  $n$ , there exists a  $(2^{nR}, n)$  code  $\mathcal{C} = (f(m, \theta^n), g(y^n, \theta^n))$  for the compound channel  $\mathcal{W}^{\overline{\mathcal{P}}_\Lambda(\mathcal{S}|\theta^\infty)}$  with fixed parameters such that

$$\phi^n(f(m, \theta^n)) \leq \Omega, \text{ for all } m \in [1 : 2^{nR}], \quad (119)$$

and

$$P_e^{(n)}(q, \theta^n, \mathcal{C}) = \sum_{s^n \in \mathcal{S}^n} q(s^n|\theta^n) P_e^{(n)}(\mathcal{C}|s^n, \theta^n) \leq e^{-2\delta n}, \quad (120)$$

for all product state distributions  $q(s^n|\theta^n) = \prod_{i=1}^n q(s_i|\theta_i)$ , with  $q \in \overline{\mathcal{P}}_\Lambda(\mathcal{S}|\theta^\infty)$ .

Therefore, by Lemma 2, taking  $h_0(s^n, \theta^n) = P_e^{(n)}(\mathcal{C}|s^n, \theta^n)$  and  $\alpha_n = e^{-2\delta n}$ , we have that for a sufficiently large  $n$ ,

$$\frac{1}{|\Pi(\theta^n)|} \sum_{\pi \in \Pi(\theta^n)} P_e^{(n)}(\mathcal{C}|\pi s^n, \theta^n) \leq (n+1)|\mathcal{S}| e^{-2\delta n} \leq e^{-\delta n}, \quad (121)$$

for all  $s^n \in \mathcal{S}^n$  with  $l^n(s^n) \leq \Lambda$ , where the sum is over the set of all  $n$ -tuple permutations such that  $\pi\theta^n = \theta^n$ .

On the other hand, for every  $\pi \in \Pi(\theta^n)$ ,

$$\begin{aligned} P_e^{(n)}(\mathcal{C}|\pi s^n, \theta^n) &\stackrel{(a)}{=} \frac{1}{2^{nR}} \sum_{m=1}^{2^{nR}} \sum_{y^n: g(y^n, \theta^n) \neq m} W_{Y^n|X^n, S^n, T^n}(y^n|f(m, \theta^n), \pi s^n, \theta^n) \\ &\stackrel{(b)}{=} \frac{1}{2^{nR}} \sum_{m=1}^{2^{nR}} \sum_{y^n: g(\pi y^n, \theta^n) \neq m} W_{Y^n|X^n, S^n, T^n}(\pi y^n|f(m, \theta^n), \pi s^n, \theta^n) \\ &\stackrel{(c)}{=} \frac{1}{2^{nR}} \sum_{m=1}^{2^{nR}} \sum_{y^n: g(\pi y^n, \theta^n) \neq m} W_{Y^n|X^n, S^n, T^n}(y^n|\pi^{-1}f(m, \theta^n), s^n, \pi^{-1}\theta^n), \end{aligned} \quad (122)$$

where (a) is obtained by plugging  $\pi s^n$  in (11a); in (b) we substitute  $\pi y^n$  instead of  $y^n$ ; and (c) holds because the channel is memoryless. Since  $\pi\theta^n = \theta^n$  for every  $\pi \in \Pi(\theta^n)$ , it follows that

$$P_e^{(n)}(\mathcal{C}|\pi s^n, \theta^n) = \frac{1}{2^{nR}} \sum_{m=1}^{2^{nR}} \sum_{y^n: g(\pi y^n, \theta^n) \neq m} W_{Y^n|X^n, S^n, T^n}(y^n|\pi^{-1}f(m, \theta^n), s^n, \theta^n). \quad (123)$$

Then, consider the  $(2^{nR}, n)$  random code  $\mathcal{C}^{\Pi(\theta^n)}$ , specified by

$$f_\pi^n(m, \theta^n) = \pi^{-1}f(m, \theta^n), \quad g_\pi(y^n, \theta^n) = g(\pi y^n, \theta^n), \quad (124)$$

with a uniform distribution  $\mu(\pi) = \frac{1}{|\Pi(\theta^n)|}$  for  $\pi \in \Pi(\theta^n)$ . As the inputs cost is additive (see (6)), the permutation does not affect the costs of the codewords, hence the random code satisfies the input constraint  $\Omega$ . From (123), we see that  $P_e^{(n)}(\mathcal{C}^{\Pi(\theta^n)}|s^n, \theta^n) = \sum_{\pi \in \Pi(\theta^n)} \mu(\pi) \cdot P_e^{(n)}(\mathcal{C}|\pi s^n, \theta^n)$ , for all  $s^n \in \mathcal{S}^n$  with  $l^n(s^n) \leq \Lambda$ . Therefore, together with (121), we have that the probability of error of the random code  $\mathcal{C}^{\Pi(\theta^n)}$  is bounded by  $P_e^{(n)}(q, \theta^n, \mathcal{C}^{\Pi(\theta^n)}) \leq e^{-\delta n}$ , for every  $q(s^n|\theta^n) \in \mathcal{P}_\Lambda(\mathcal{S}^n|\theta^n)$ . It follows that  $\mathcal{C}^{\Pi(\theta^n)}$  is a  $(2^{nR}, n, e^{-\delta n})$  random code for the AVC  $\mathcal{W}$  with fixed parameters under input constraint  $\Omega$  and state constraint  $\Lambda$ .  $\square$

### B. Converse Proof

Assume to the contrary that there exists an achievable rate pair

$$R > C(\mathcal{W}^{\mathcal{Q}}) \Big|_{\mathcal{Q}=\overline{\mathcal{P}}_{\Lambda-\delta}(\mathcal{S}|\theta^\infty)}, \quad (125)$$

using random codes over the AVC  $\mathcal{W}$  under input constraint  $\Omega$  and state constraint  $\Lambda$ , where  $\delta > 0$  is arbitrarily small. That is, for every  $\varepsilon > 0$  and sufficiently large  $n$ , there exists a  $(2^{nR}, n)$  random code  $\mathcal{C}^\Gamma = (\mu, \Gamma, \{\mathcal{C}_\gamma\}_{\gamma \in \Gamma})$  for the AVC  $\mathcal{W}$ , such that  $\sum_{\gamma \in \Gamma} \mu(\gamma) \phi^n(f_\gamma(m, \theta^n)) \leq \Omega$ , and

$$P_e^{(n)}(q, \theta^n, \mathcal{C}^\Gamma) \leq \varepsilon, \quad (126)$$

for all  $m \in [1 : 2^{nR}]$  and  $q(s^n|\theta^n) \in \mathcal{P}_\Lambda(\mathcal{S}^n|\theta^n)$ . In particular, for distributions  $q(\cdot|\theta^n)$  that give mass 1 to some sequence  $s^n \in \mathcal{S}^n$  with  $l^n(s^n) \leq \Lambda$ , we have that  $P_e^{(n)}(\mathcal{C}^\Gamma|s^n, \theta^n) \leq \varepsilon$ .

Consider using the random code  $\mathcal{C}^\Gamma$  over the compound channel  $\mathcal{W}^{\overline{\mathcal{P}}_{\Lambda-\delta}(S)}$  with fixed parameters under input constraint  $\Omega$ . Let  $\overline{q}(s|t) \in \overline{\mathcal{P}}_{\Lambda-\delta}(S)$  be a given state distribution. Then, define a sequence of conditionally independent random variables  $\overline{S}_1, \dots, \overline{S}_n \sim \overline{q}(s|t)$ . Letting  $\overline{q}^n(s^n|\theta^n) \triangleq \prod_{i=1}^n \overline{q}(s_i|\theta_i)$ , the probability of error is bounded by

$$P_e^{(n)}(\overline{q}, \theta^n, \mathcal{C}^\Gamma) \leq \sum_{s^n: l^n(s^n) \leq \Lambda} \overline{q}^n(s^n|\theta^n) P_e^{(n)}(\mathcal{C}^\Gamma | s^n, \theta^n) + \Pr\left(l^n(\overline{S}^n) > \Lambda\right). \quad (127)$$

The first sum is bounded by (126), and the second term vanishes by the law of large numbers, since  $\overline{q} \in \overline{\mathcal{P}}_{\Lambda-\delta}(S|\theta^\infty)$ . It follows that the random code  $\mathcal{C}^\Gamma$  achieves a rate  $R$  as in (125) over the compound channel  $\mathcal{W}^{\overline{\mathcal{P}}_{\Lambda-\delta}(S)}$  with fixed parameters under input constraint  $\Omega$ , for an arbitrarily small  $\delta > 0$ , in contradiction to Lemma 1. We deduce that the assumption is false, and  $\mathbb{C}^*(\mathcal{W}) \leq \mathbb{C}(\mathcal{W}^\Omega)|_{\mathcal{Q}=\overline{\mathcal{P}}_{\Lambda}(S|\theta^\infty)} = \mathbb{C}_n^*(\mathcal{W})$ .  $\square$

#### APPENDIX D PROOF OF LEMMA 4

To prove that  $\mathbb{R}_n^*(\mathcal{W}) = \mathbb{C}_n^*(\mathcal{W})$ , we begin with the property in the lemma below.

*Lemma 17.* Let  $\omega_i^*, \lambda_i^*, i \in [1:n]$ , be the parameters that achieve the saddle point in (21), i.e.

$$\mathbb{R}_n^*(\mathcal{W}) = \frac{1}{n} \sum_{i=1}^n \mathbb{C}_{\theta_i}(\omega_i^*, \lambda_i^*). \quad (128)$$

Then, for every  $i, j \in [1:n]$  such that  $\theta_i = \theta_j$ , we have that  $\omega_i^* = \omega_j^*$  and  $\lambda_i^* = \lambda_j^*$ .

*Proof of Lemma 17.* For every  $i \in [1:n]$ , let  $p_i, q_i$  denote input and state distributions such that  $\mathbb{E}\phi(X_i) \leq \omega_i^*$ ,  $\mathbb{E}l(S_i) \leq \lambda_i^*$  for  $X_i \sim p_i, S_i \sim q_i$ . Now, suppose that  $\theta_i = \theta_j = t$ , and define

$$p'(x) = \frac{1}{2}[p_i(x) + p_j(x)], \quad q'(s) = \frac{1}{2}[q_i(s) + q_j(s)]. \quad (129)$$

Then,  $\mathbb{E}\phi(X') = \frac{1}{2}[\mathbb{E}\phi(X_i) + \mathbb{E}\phi(X_j)]$  and  $\mathbb{E}l(S') = \frac{1}{2}[\mathbb{E}l(S_i) + \mathbb{E}l(S_j)]$  for  $X' \sim p', S' \sim q'$ . Furthermore, since the mutual information is concave- $\cap$  in the input distribution and convex- $\cup$  in the state distribution, we have that

$$\begin{aligned} \frac{1}{2} [I_{q'}(X_i; Y_i | T_i = t) + I_{q'}(X_j; Y_j | T_j = t)] &\leq I_{q'}(X'; Y' | T' = t) \\ \frac{1}{2} [I_{q_i}(X'; Y' | T' = t) + I_{q_j}(X'; Y' | T' = t)] &\geq I_{q'}(X'; Y' | T' = t). \end{aligned} \quad (130)$$

Therefore, the saddle point distributions must satisfy  $p_i = p_j = p'$  and  $q_i = q_j = q'$ , hence  $\omega_i^* = \omega_j^*$  and  $\lambda_i^* = \lambda_j^*$ .  $\square$

Next, it can be inferred from Lemma 17 that

$$\begin{aligned} \mathbb{R}_n^*(\mathcal{W}) &= \min_{\substack{(\lambda_t)_{t \in \mathcal{T}}: \\ \sum_{t \in \mathcal{T}} P_T(t) \lambda_t \leq \Lambda}} \max_{\substack{(\omega_t)_{t \in \mathcal{T}}: \\ \sum_{t \in \mathcal{T}} P_T(t) \omega_t \leq \Omega}} \sum_{t \in \mathcal{T}} P_T(t) \mathbb{C}_t(\omega_t, \lambda_t) \\ &= \min_{\substack{(\lambda_t)_{t \in \mathcal{T}}, q(s|t): \\ \mathbb{E}_q[l(S)|T=t] \leq \lambda_t \\ \sum_{t \in \mathcal{T}} P_T(t) \lambda_t \leq \Lambda}} \max_{\substack{(\omega_t)_{t \in \mathcal{T}}, p(x|t): \\ \mathbb{E}[\phi(X)|T=t] \leq \omega_t \\ \sum_{t \in \mathcal{T}} P_T(t) \omega_t \leq \Omega}} I_q(X; Y | T) \\ &= \min_{q(s|t): \mathbb{E}_q l(S) \leq \Lambda} \max_{p(x|t): \mathbb{E}\phi(X) \leq \Omega} I_q(X; Y | T) = \mathbb{C}_n^*(\mathcal{W}), \end{aligned} \quad (131)$$

where  $P_T$  is the type of the parameter sequence  $\theta^n$ . The second equality follows from the definition of  $\mathbb{C}_t^*(\omega_t, \lambda_t)$  in (20), using the minimax theorem [96] to switch between the order of the minimum and maximum. In the third line, we eliminate the slack variables  $\lambda_i$  and  $\omega_i$  replacing  $\mathbb{E}_q l(S_i)$  and  $\mathbb{E}\phi(X_i)$ , respectively. The last equality holds by the definition of  $\mathbb{C}_n^*(\mathcal{W})$  in (16).  $\square$

#### APPENDIX E PROOF OF LEMMA 8

Consider the AVC  $\mathcal{W}$  with fixed parameters under input constraint  $\Omega$  and state constraint  $\Lambda$ . Let  $\theta^n$  be sequence of fixed parameters for a given blocklength. Recall that  $T$  is a random variable that is distributed as the type of  $\theta^n$ . We extend the proof in [30]. First, we give an auxiliary lemma, which we also used in [85].

*Lemma 18* (See [30] [85, Lemma 11]). For every pair of conditional state distributions  $Q(s|x, t)$  and  $Q'(s|x, t)$  such that

$$\max_{t, x, s} \left\{ \sum_{t, x, s} P_T(t) p(x|t) Q(s|x, t) l(s), \sum_{t, x, s} P_T(t) p(x|t) Q'(s|x, t) l(s) \right\} < \tilde{\Lambda}_n(p), \quad (132)$$

there exists  $\xi > 0$  such that

$$\max_{x, \tilde{x}, y} \left| \sum_{t,s} P_T(t) Q(s|\tilde{x}, t) W_{Y|X,S,T}(y|x, s, t) - \sum_{t,s} P_T(t) Q'(s|x, t) W_{Y|X,S,T}(y|\tilde{x}, s, t) \right| \geq \xi. \quad (133)$$

*Proof of Lemma 18.* Assume to the contrary that the LHS in (133) is zero, and define

$$Q_A(s|x, t) = \frac{1}{2} (Q(s|x, t) + Q'(s|x, t)). \quad (134)$$

Using the symmetry between  $Q$  and  $Q'$ , we have that

$$\begin{aligned} 0 &= \max_{x, \tilde{x}, y} \left| \sum_{t \in \mathcal{T}} \sum_{s \in \mathcal{S}} P_T(t) Q(s|\tilde{x}, t) W_{Y|X,S,T}(y|x, s, t) - \sum_{t \in \mathcal{T}} \sum_{s \in \mathcal{S}} P_T(t) Q'(s|x, t) W_{Y|X,S,T}(y|\tilde{x}, s, t) \right| \\ &= \frac{1}{2} \max_{x, \tilde{x}, y} \left| \sum_{t \in \mathcal{T}_n} \sum_{s \in \mathcal{S}} P_T(t) Q(s|\tilde{x}, t) W_{Y|X,S,T}(y|x, s, t) - \sum_{t \in \mathcal{T}_n} \sum_{s \in \mathcal{S}} P_T(t) Q'(s|x, t) W_{Y|X,S,T}(y|\tilde{x}, s, t) \right| \\ &\quad + \frac{1}{2} \max_{x, \tilde{x}, y} \left| \sum_{t \in \mathcal{T}_n} \sum_{s \in \mathcal{S}} P_T(t) Q'(s|\tilde{x}, t) W_{Y|X,S,T}(y|x, s, t) - \sum_{t \in \mathcal{T}_n} \sum_{s \in \mathcal{S}} P_T(t) Q(s|x, t) W_{Y|X,S,T}(y|\tilde{x}, s, t) \right| \\ &\geq \max_{x, \tilde{x}, y} \left| \sum_{t \in \mathcal{T}_n} \sum_{s \in \mathcal{S}} P_T(t) Q_A(s|x, t) W_{Y|X,S,T}(y|\tilde{x}, s, t) - \sum_{t \in \mathcal{T}_n} \sum_{s \in \mathcal{S}} P_T(t) Q_A(s|\tilde{x}, t) W_{Y|X,S,T}(y|x, s, t) \right|. \end{aligned} \quad (135)$$

Since we have assumed that  $P_T(t) > \delta_0$  for all  $t \in \mathcal{T}$ , it follows that

$$\sum_{s \in \mathcal{S}} Q_A(s|x, t) W_{Y|X,S,T}(y|\tilde{x}, s, t) = \sum_{s \in \mathcal{S}} Q_A(s|\tilde{x}, t) W_{Y|X,S,T}(y|x, s, t), \quad (136)$$

for all  $t \in \mathcal{T}$ ,  $x, \tilde{x} \in \mathcal{X}$  and  $y \in \mathcal{Y}$ . In other words,  $Q_A(\cdot|\cdot, t)$  symmetrizes the channel  $W_{Y|X,S,T}(\cdot|\cdot, \cdot, t)$  for all  $t \in \mathcal{T}$ . Therefore, by the definition of  $\tilde{\Lambda}_n(p)$  in (27), we have that

$$\sum_{t,x,s} P_T(t) p(x|t) Q_A(s|x, t) l(s) = \frac{1}{n} \sum_{i=1}^n \sum_{x,s} p(x|\theta_i) Q_A(s|x, \theta_i) l(s) \geq \tilde{\Lambda}_n(p) \quad (137)$$

in contradiction to (132). The equality above holds because  $T$  is distributed as the type of the parameter sequence  $\theta^n$ , hence averaging over time is the same as averaging according to  $P_T$ . It follows that the LHS of (133) must be positive. This completes the proof of the auxiliary Lemma.  $\square$

We move to the main part of the proof. To show that (37) holds for sufficiently small  $\eta$ , assume to the contrary that there exists  $y^n$  such that  $(y^n, \theta^n)$  is in  $\mathcal{D}(m) \cap \mathcal{D}(\tilde{m}) \neq \emptyset$ . By the assumption in the lemma, the codewords  $\{f(m, \theta^n)\}_{m \in [1:2^{nR}]}$  have the same conditional type. In particular,  $P_{\tilde{X}|T} = P_{X|T} = p$ .

By Condition 1) of the decoding rule,

$$\begin{aligned} &D(P_{T,X,S,Y} \| P_T \times P_{X|T} \times P_{S|T} \times W_{Y|X,S,T}) \\ &= \sum_{t,x,s,y} P_{T,X,S,Y}(t, x, s, y) \log \frac{P_{T,X,S,Y}(t, x, s, y)}{P_T(t) p(x|t) P_{S|T}(s|t) W_{Y|X,S,T}(y|x, s, t)} \leq \eta, \end{aligned} \quad (138)$$

and by Condition 2) of the decoding rule,

$$I(X, Y; \tilde{X}|S, T) = \sum_{t,x,\tilde{x},s,y} P_{T,X,\tilde{X},S,Y}(t, x, \tilde{x}, s, y) \log \frac{P_{\tilde{X}|X,S,T,Y}(\tilde{x}|x, s, t, y)}{P_{\tilde{X}|S,T}(\tilde{x}|s, t)} \leq \eta, \quad (139)$$

where  $T, X, \tilde{X}, S, Y$  are distributed according to the joint type of  $\theta^n, f^n(m, \theta^n), f^n(\tilde{m}, \theta^n), s^n$ , and  $y^n$ . Adding (138) and (139) yields

$$\sum_{t,x,\tilde{x},s,y} P_{T,X,\tilde{X},S,Y}(t, x, \tilde{x}, s, y) \log \frac{P_{T,X,\tilde{X},S,Y}(t, x, \tilde{x}, s, y)}{P_T(t) p(x|t) P_{\tilde{X}|S,T}(\tilde{x}|s, t) W_{Y|X,S,T}(y|x, s, t)} \leq 2\eta. \quad (140)$$

That is,  $D(P_{T,X,\tilde{X},S,Y} \| P_T \times p \times p \times P_{S|\tilde{X},T} \times W_{Y|X,S,T}) \leq 2\eta$ . Therefore, by the log-sum inequality (see e.g. [27, Theorem 2.7.1]),

$$\begin{aligned} &D(P_{T,X,\tilde{X},Y} \| P_T \times p \times p \times V_{Y|X,\tilde{X},T}) \\ &\leq D(P_{T,X,\tilde{X},S,Y} \| P_T \times p \times p \times P_{S|\tilde{X},T} \times W_{Y|X,S,T}) \leq 2\eta, \end{aligned} \quad (141)$$

where  $V_{Y|X,\tilde{X},T}(y|x,\tilde{x},t) = \sum_{s \in \mathcal{S}} W_{Y|X,S,T}(y|x,s,t)P_{S|\tilde{X},T}(s|\tilde{x},t)$ . Then, by Pinsker's inequality (see *e.g.* [29, Problem 3.18]),

$$\sum_{t,x,\tilde{x},y} |P_{T,X,\tilde{X},Y}(t,x,\tilde{x},y) - P_T(t)p(x|t)p(\tilde{x}|t)V_{Y|X,\tilde{X},T}(y|x,\tilde{x},t)| \leq c\sqrt{2\eta}, \quad (142)$$

where  $c > 0$  is a constant. By the same arguments, (34) implies that

$$\sum_{t,x,\tilde{x},y} |P_{T,X,\tilde{X},Y}(t,x,\tilde{x},s) - P_T(t)p(x|t)p(\tilde{x}|t)V'_{Y|X,\tilde{X},T}(y|x,\tilde{x},t)| \leq c\sqrt{2\eta}, \quad (143)$$

where  $V'_{Y|X,\tilde{X},T}(y|x,\tilde{x},t) = \sum_{s \in \mathcal{S}} W_{Y|X,S,T}(y|\tilde{x},s,t)P_{\tilde{S}|X,T}(s|x,t)$ . Now, observe that inserting the sum over  $t \in \mathcal{T}$  into the absolute value maintains the inequality, by the triangle inequality. Furthermore, since  $p(x|t) > \delta_1$ , for  $x \in \mathcal{X}$ ,  $t \in \mathcal{T}$ , we have that

$$\max_{x,\tilde{x},y} \left| \sum_{t \in \mathcal{T}_n} P_T(t)V_{Y|X,\tilde{X},T}(y|x,\tilde{x},t) - \sum_{t \in \mathcal{T}_n} P_T(t)V'_{Y|X,\tilde{X},T}(y|x,\tilde{x},t) \right| \leq \frac{2c\sqrt{2\eta}}{\delta_1^2}, \quad (144)$$

Equivalently, the above can be expressed as

$$\max_{x,\tilde{x},y} \left| \sum_{t,s} P_T(t)P_{S|\tilde{X},T}(s|\tilde{x},t)W_{Y|X,S,T}(y|x,s,t) - \sum_{t,s} P_T(t)P_{\tilde{S}|X,T}(s|x,t)W_{Y|X,S,T}(y|\tilde{x},s,t) \right| \leq \frac{2c\sqrt{2\eta}}{\delta_1^2}, \quad (145)$$

Now, we show that the state distributions  $Q = P_{S|\tilde{X},T}$  and  $Q' = P_{\tilde{S}|X,T}$  satisfy the conditions of Lemma 18. Indeed,

$$\begin{aligned} & \max \left\{ \sum_{t,\tilde{x},s} P_T(t)p(\tilde{x}|t)Q(s|\tilde{x})l(s), \sum_{t,x,s} P_T(t)p(x|t)Q'(s|x)l(s) \right\} \\ &= \max \left\{ \sum_{t,\tilde{x},s} P_T(t)p(\tilde{x}|t)P_{S|\tilde{X},T}(s|\tilde{x},t)l(s), \sum_{t,x,s} P_T(t)p(x|t)P_{\tilde{S}|X,T}(s|x,t)l(s) \right\} \\ &= \max \left\{ \sum_s P_S(s)l(s), \sum_s P_{\tilde{S}}(s)l(s) \right\} \\ &= \max \{l^n(s^n), l^n(\tilde{s}^n)\} \leq \Lambda < \tilde{\Lambda}_n(p), \end{aligned} \quad (146)$$

where the last inequality is due to (36). Thus, there exists  $\xi > 0$  such that (133) holds with  $Q = P_{S|\tilde{X},T}$  and  $Q' = P_{\tilde{S}|X,T}$ , which contradicts (145), if  $\eta$  is sufficiently small such that  $\frac{2c\sqrt{2\eta}}{\delta_1^2} < \xi$ .  $\square$

## APPENDIX F PROOF OF LEMMA 9

Let  $Z^n(m, \theta^n)$ ,  $m \in [1 : 2^{nR}]$ , be statistically independent sequences, uniformly distributed over the conditional type class  $\mathcal{T}^n(p)$ . Fix  $a^n \in \mathcal{X}^n$  and  $s^n \in \mathcal{S}^n$ , and consider a joint type  $P_{T,X,\tilde{X},S}$ , such that  $P_{X|T} = P_{\tilde{X}|T} = p$ . We intend to show that  $\{Z^n(m, \theta^n)\}$  satisfy each of the desired properties with double exponential high probability  $(1 - e^{-2^{\text{E}n}})$ ,  $\text{E} > 0$ , implying that there exists a deterministic codebook that satisfies (38)-(40) simultaneously. We begin with the following large deviations result by Csisár and Narayan [30].

*Lemma 19* (see [30, Lemma A1]). Let  $\alpha, \beta \in [0, 1]$ , and consider a sequence of random vectors  $U^n(m)$ , and functions  $\varphi_m : \mathcal{X}^{nm} \rightarrow [0, 1]$ , for  $m \in [1 : M]$ . If

$$\mathbb{E}(\varphi_m(U^n(1) \dots, U^n(m)) | U^n(1) \dots, U^n(m-1)) \leq \alpha \quad \text{a.s., for } m \in [1 : M], \quad (147)$$

then

$$\Pr \left( \sum_{m=1}^M \varphi_m(U^n(1) \dots, U^n(m)) > M\beta \right) \leq \exp\{-M(\beta - \alpha \log e)\}. \quad (148)$$

To show that (38) holds, consider the indicator

$$\varphi_m(Z^n(1, \theta^n), \dots, Z^n(m, \theta^n)) = \begin{cases} 1 & \text{if } (\theta^n, Z^n(m, \theta^n), Z^n(\tilde{m}, \theta^n), s^n) \in \mathcal{T}^n(P_{T,X,\tilde{X},S}) \\ & \text{for some } \tilde{m} < m \\ 0 & \text{otherwise} \end{cases} \quad (149)$$

By standard type class considerations (see *e.g.* [67, Theorem 1.3]), we have that

$$\mathbb{E}[\varphi_m(Z^n(1, \theta^n), \dots, Z^n(m, \theta^n)) | Z^n(1, \theta^n), \dots, Z^n(m-1, \theta^n)] \leq 2^{-n(I(\tilde{X};T,X,S) - \frac{\epsilon}{4} - R)} \leq 2^{-n(I(\tilde{X};X,S|T) - \frac{\epsilon}{4} - R)}, \quad (150)$$

where the last inequality holds since  $I(\tilde{X}; T, X, S) \geq I(\tilde{X}; X, S|T)$ .

Next, we use Lemma 19, and plug

$$\begin{aligned} M &= 2^{nR}, \quad U^n(m) = Z^n(m, \theta^n), \\ \alpha &= 2^{-n(I(\tilde{X}; X, S|T) - \frac{\varepsilon}{4} - R)}, \quad \beta = 2^n \left( [R - I(\tilde{X}; X, S|T)]_+ - R + \varepsilon \right). \end{aligned} \quad (151)$$

For sufficiently large  $n$ , we have that  $M(\beta - \alpha \log e) \geq 2^{n\varepsilon/2}$ . Hence, by Lemma 19,

$$\Pr \left( \sum_{m=1}^{2^{nR}} \varphi_m(Z^n(1, \theta^n), \dots, Z^n(2^{nR}, \theta^n)) > 2^n \left( [R - I(\tilde{X}; X, S|T)]_+ + \varepsilon \right) \right) \leq e^{-2^{n\varepsilon/2}}. \quad (152)$$

By the symmetry between  $m$  and  $\tilde{m}$  in the derivation above, the double exponential decay of the probability in (152) implies that there exists a codebook that satisfies (38).

Similarly, to show (39), we replace the indicator of the type  $P_{X, \tilde{X}, S|T}$  in (149) by an indicator of the type  $P_{\tilde{X}, S|T}$ , and rewrite (150) with  $I(\tilde{X}; S|T)$ , to obtain

$$\Pr \left( |\{\tilde{m} : (\theta^n, Z^n(\tilde{m}, \theta^n), s^n) \in \mathcal{T}^n(P_{T, \tilde{X}, S})\}| > 2^n \left( [R - I(\tilde{X}; S|T)]_+ + \varepsilon_1 \right) \right) < e^{-2^{n\varepsilon_1/2}}, \quad (153)$$

where  $\varepsilon_1 > 0$  is arbitrarily small. If  $I(\tilde{X}; S|T) > \varepsilon$  and  $R \geq \varepsilon$ , then choosing  $\varepsilon_1 = \frac{\varepsilon}{2}$ , we have that

$$\left[ R - I(\tilde{X}; S|T) \right]_+ + \varepsilon_1 \leq R - \frac{\varepsilon}{2}, \quad (154)$$

hence,

$$\Pr \left( |\{\tilde{m} : (\theta^n, Z^n(\tilde{m}, \theta^n), s^n) \in \mathcal{T}^n(P_{T, \tilde{X}, S})\}| > 2^{n(R - \frac{\varepsilon}{2})} \right) < e^{-2^{n\varepsilon/4}}. \quad (155)$$

It remains to show that (40) holds. Assume that

$$I(X; \tilde{X}, S|T) - \left[ R - I(\tilde{X}; S|T) \right]_+ > \varepsilon. \quad (156)$$

Let  $\mathcal{J}_m$  denote the set of indices  $\tilde{m} < m$  such that  $(\theta^n, Z^n(\tilde{m}, \theta^n), s^n) \in \mathcal{T}^n(P_{T, \tilde{X}, S})$ , provided that their number does not exceed  $2^n \left( [R - I(\tilde{X}; S|T)]_+ + \frac{\varepsilon}{8} \right)$ ; else, let  $\mathcal{J}_m = \emptyset$ . Also, let

$$\psi_m(Z^n(1, \theta^n), \dots, Z^n(m, \theta^n)) = \begin{cases} 1 & \text{if } (\theta^n, Z^n(m, \theta^n), Z^n(\tilde{m}, \theta^n), s^n) \in \mathcal{T}^n(P_{T, X, \tilde{X}, S}) \\ & \text{for some } \tilde{m} \in \mathcal{J}_m, \\ 0 & \text{otherwise.} \end{cases} \quad (157)$$

Then, choosing  $\varepsilon_1 = \frac{\varepsilon}{8}$  in (153) yields

$$\begin{aligned} \Pr \left( \sum_{m=1}^{2^{nR}} \psi_m(Z^n(1, \theta^n), \dots, Z^n(m, \theta^n)) \neq |\{m : \right. \\ \left. (\theta^n, Z^n(m, \theta^n), Z^n(\tilde{m}, \theta^n), s^n) \in \mathcal{T}^n(P_{T, X, \tilde{X}, S}) \text{ for some } \tilde{m} < m\}| \right) < e^{-2^{n\varepsilon/16}}. \end{aligned} \quad (158)$$

Therefore, instead of bounding the set of messages, it is sufficient to consider the sum  $\sum \psi_m(Z^n(1, \theta^n), \dots, Z^n(m, \theta^n))$ . Furthermore, by standard type class considerations (see *e.g.* [67, Theorem 1.3]), we have that

$$\begin{aligned} \mathbb{E} \left( \psi_m(Z^n(1, \theta^n), \dots, Z^n(m, \theta^n)) \mid Z^n(1, \theta^n), \dots, Z^n(m-1, \theta^n) \right) &\leq |\mathcal{J}_m| \cdot 2^{-n(I(X; \tilde{X}, S|T) - \frac{\varepsilon}{8})} \\ &\leq 2^n \left( [R - I(\tilde{X}; S|T)]_+ - I(X; \tilde{X}, S|T) + \frac{\varepsilon}{4} \right) < 2^{-3n\varepsilon/4}, \end{aligned} \quad (159)$$

where the last inequality is due to (156). Thus, by Lemma 19,

$$\Pr \left( \sum_{m=1}^{2^{nR}} \psi_m(Z^n(1, \theta^n), \dots, Z^n(m, \theta^n)) > 2^{n(R - \frac{\varepsilon}{2})} \right) < e^{-2^{n(R - \frac{3\varepsilon}{4})}} \leq e^{-2^{n\varepsilon/4}}, \quad (160)$$

as we have assumed that  $R \geq \varepsilon$ . Equations (158) and (160) imply that the property in (40) holds with double exponential probability  $1 - e^{-2^{E_1 n}}$ , where  $E_1 > 0$ .  $\square$

APPENDIX G  
PROOF OF THEOREM 6

A. Achievability Proof

Suppose that  $L_n^* > \Lambda$  for sufficiently large  $n$ . Let  $\varepsilon > 0$  be chosen later, and let  $P_{X|T}$  be a conditional type over  $\mathcal{X}$ , for which  $P_{X|T}(x|t) > 0 \forall x \in \mathcal{X}, t \in \mathcal{T}$ , and  $\mathbb{E}\phi(X) \leq \Omega$ , with

$$\tilde{\Lambda}_n(P_{X|T}) > \Lambda. \quad (161)$$

As explained below, we may assume without loss of generality that for some  $\delta_0 > 0$  that does not depend on  $n$ , we have that  $P_T(t) > \delta_0$  for all  $t \in \mathcal{T}$ . Indeed, following our assumption in (25), the asymptotic capacity formula  $\liminf C_n(\mathcal{W})$  does not change when we remove parameter values  $t \in \mathcal{T}$  such that  $P_T(t) \rightarrow 0$ . Hence, coding can be limited to the rest of the block with negligible rate decrease, thus removing those parameters from consideration. Then, choose  $\eta > 0$  to be sufficiently small such that Lemma 8 guarantees that the decoder in Definition 5 is well defined. Now, Lemma 9 assures that there is a codebook  $\{x^n(m, \theta^n)\}_{m \in [1:2^{nR}]}$  of conditional type  $p$  that satisfies (38)-(40). Consider the following coding scheme.

*Encoding:* To send  $m \in [1:2^{nR}]$ , transmit  $x^n(m, \theta^n)$ .

*Decoding:* Find a unique message  $\hat{m}$  such that  $(y^n, \theta^n)$  belongs to  $\mathcal{D}(\hat{m})$ , as in Definition 5. If there is none, declare an error. Lemma 8 guarantees that there cannot be two messages for which this holds.

*Analysis of Probability of Error:* Fix  $s^n \in \mathcal{S}^n$  with  $l^n(s^n) \leq \Lambda$ , let  $q = P_{S|T}$  denote the conditional type of  $s^n$  given  $\theta^n$ , and let  $M$  denote the transmitted message. Consider the error events

$$\mathcal{E}_1 = \{D(P_{T,X,S,Y} \| P_T \times P_{X|T} \times P_{S|T} \times W_{Y|X,S}) > \eta\} \quad (162)$$

$$\mathcal{E}_2 = \{\text{Condition 2) of the decoding rule is violated}\} \quad (163)$$

and

$$\mathcal{F}_1 = \{I_q(X; S|T) > \varepsilon\}, \quad (164)$$

$$\mathcal{F}_2 = \{I_q(X; \tilde{X}, S|T) > [R - I(\tilde{X}; S|T)]_+ + \varepsilon, \text{ for some } \tilde{m} \neq M\}, \quad (165)$$

where  $(T, X, \tilde{X}, S)$  are dummy random variables, which are distributed as the joint type of  $(\theta^n, x^n(M, \theta^n), x^n(\tilde{m}, \theta^n), s^n)$ . By the union of events bound,

$$P_e^{(n)}(\mathcal{C}|s^n, \theta^n) \leq \Pr(\mathcal{F}_1) + \Pr(\mathcal{F}_2) + \Pr(\mathcal{E}_1 \cap \mathcal{F}_1^c) + \Pr(\mathcal{E}_2 \cap \mathcal{F}_2^c), \quad (166)$$

where the conditioning on  $S^n = s^n$  and  $T^n = \theta^n$  is omitted for convenience of notation. Based on Lemma 9, the probabilities of the events  $\mathcal{F}_1$  and  $\mathcal{F}_2$  tend to zero as  $n \rightarrow \infty$ , by (39) and (40), respectively.

Now, suppose that Condition 1) of the decoding rule is violated. Observe that the event  $\mathcal{E}_1 \cap \mathcal{F}_1^c$  implies that

$$\begin{aligned} & D(P_{T,X,S,Y} \| P_{T,X,S} \times W_{Y|X,S,T}) \\ &= D(P_{T,X,S,Y} \| P_T \times P_{X|T} \times P_{S|T} \times W_{Y|X,S,T}) - I(X; S|T) > \eta - \varepsilon. \end{aligned} \quad (167)$$

Then, by standard large deviations considerations (see *e.g.* [27, pp. 362–364]),

$$\begin{aligned} \Pr(\mathcal{E}_1 \cap \mathcal{F}_1^c) &\leq \max_{P_{T,X,S,Y} : \mathcal{E}_1 \cap \mathcal{F}_1^c \text{ holds}} 2^{-n(D(P_{T,X,S,Y} \| P_{T,X,S} \times W_{Y|X,S,T}) - \varepsilon)} \\ &< 2^{-n(\eta - 2\varepsilon)}, \end{aligned} \quad (168)$$

which tends to zero as  $n \rightarrow \infty$ , for sufficiently small  $\varepsilon > 0$ , with  $\varepsilon < \frac{1}{2}\eta$ .

Moving to Condition 2) of the decoding rule, let  $\mathcal{D}_2$  denote the set of joint types  $P_{T,X,\tilde{X},S}$  such that

$$D(P_{T,X,S,Y} \| P_T P_{X|T} \times P_{S|T} \times W_{Y|X,S,T}) \leq \eta, \quad (169)$$

$$D(P_{\tilde{X},\tilde{S},Y} \| P_{\tilde{X}} \times P_{\tilde{S}|T} \times W_{Y|X,S,T}) \leq \eta, \text{ for some } \tilde{S} \sim \tilde{q}(s|t), \quad (170)$$

$$I_q(X, Y; \tilde{X}|S, T) > \eta. \quad (171)$$

Then, by standard type class considerations (see *e.g.* [67, Theorem 1.3]),

$$\begin{aligned} \Pr(\mathcal{E}_2 \cap \mathcal{F}_2^c | M = m) &\leq \sum_{\substack{P_{T,X,\tilde{X},S} \in \mathcal{D}_2 : \\ \mathcal{F}_2^c \text{ holds}}} |\{\tilde{m} : (\theta^n, x^n(m, \theta^n), x^n(\tilde{m}, \theta^n), s^n) \in \mathcal{T}^n(P_{T,X,\tilde{X},S})\}| \\ &\times 2^{-n(I_q(\tilde{X}; Y|X,S,T) - \varepsilon)}, \end{aligned} \quad (172)$$

for every given  $m \in [1 : 2^{nR}]$ . Hence, by (38),

$$\Pr(\mathcal{E}_2 \cap \mathcal{F}_2^c) \leq \sum_{\substack{P_{T,X,\tilde{X},S} \in \mathcal{D}_2 : \\ \mathcal{F}_2^c \text{ holds}}} 2^{-n(I_q(\tilde{X}; Y|X,S,T) - [R - I_q(\tilde{X}; X, S|T)]_+ - 2\varepsilon)}. \quad (173)$$

To further bound  $\Pr(\mathcal{E}_2 \cap \mathcal{F}_2^c)$ , consider the following cases. Suppose that  $R \leq I_q(\tilde{X}; S|T)$ . Then, given  $\mathcal{F}_2^c$ , we have that

$$I_q(X; \tilde{X}|S, T) \leq I_q(X; \tilde{X}, S|T) \leq \varepsilon. \quad (174)$$

By (171), it then follows that

$$\begin{aligned} I_q(\tilde{X}; Y|X, S, T) &= I_q(\tilde{X}; X, Y|S, T) - I_q(\tilde{X}; X|S, T) \\ &\geq \eta - \varepsilon. \end{aligned} \quad (175)$$

Returning to (173), we note that since the number of types is polynomial in  $n$ , the cardinality of the set of types  $\mathcal{D}_2$  can be bounded by  $2^{n\varepsilon}$ , for sufficiently large  $n$ . Hence, by (173) and (175), we have that  $\Pr(\mathcal{E}_2 \cap \mathcal{F}_2^c) \leq 2^{-n(\eta - 4\varepsilon)}$ , which tends to zero as  $n \rightarrow \infty$ , for  $\varepsilon < \frac{1}{4}\eta$ .

Otherwise, if  $R > I_q(\tilde{X}; S|T)$ , then given  $\mathcal{F}_2^c$ ,

$$\begin{aligned} R &> I_q(X; \tilde{X}, S|T) + I(\tilde{X}; S|T) - \varepsilon \\ &= I_q(\tilde{X}; X, S|T) + I(X; S|T) - \varepsilon \\ &\geq I_q(\tilde{X}; X, S|T) - \varepsilon. \end{aligned} \quad (176)$$

Thus,

$$\left[ R - I_q(\tilde{X}; X, S|T) \right]_+ \leq R - I_q(\tilde{X}; X, S|T) + \varepsilon. \quad (177)$$

Hence, by (173) we have that

$$\begin{aligned} \Pr(\mathcal{E}_2 \cap \mathcal{F}_2^c) &\leq \sum_{\substack{P_{T,X,\tilde{X},S} \in \mathcal{D}_2 \\ \mathcal{F}_2^c \text{ holds}}} 2^{-n(I(\tilde{X}; X, S, Y|T) - R - 3\varepsilon)} \\ &\leq \sum_{\substack{P_{T,X,\tilde{X},S} \in \mathcal{D}_2 : \\ \mathcal{F}_2^c \text{ holds}}} 2^{-n(I_q(\tilde{X}; Y|T) - R - 3\varepsilon)}. \end{aligned} \quad (178)$$

For  $P_{T,X,\tilde{X},S} \in \mathcal{D}_2$ , we have by (170) that  $P_{T,\tilde{X},\tilde{S},Y}$  is arbitrarily close to some  $P_{T,X,\tilde{S},\tilde{Y}}$ , where

$$P_{T,\tilde{X},\tilde{S},\tilde{Y}}(x, s, y) = P_T(t)P_{X|T}(x|t)\tilde{q}(s|t)W_{Y|X,S,T}(y|x, s, t), \quad (179)$$

if  $\eta > 0$  is sufficiently small. In which case,

$$I_q(\tilde{X}; Y|T) \geq I_{\tilde{q}}(X; Y|T) - \delta, \quad (180)$$

where  $\delta > 0$  is arbitrarily small. Therefore, provided that

$$R < \min_{q(s|t) : \mathbb{E}_q l(S) \leq \Lambda} I_q(X; Y|T) - \delta - 5\varepsilon, \quad (181)$$

we have that  $\Pr(\mathcal{E}_2 \cap \mathcal{F}_2^c) \leq 2^{-n(I_q(\tilde{X}; Y|T) - R - 4\varepsilon)}$  tends to zero as  $n \rightarrow \infty$ .  $\square$

### B. Converse Proof

We will use the following lemma, based on the observations of Ericson [37].

*Lemma 20.* Consider the AVC with fixed parameters free of state constraints, and let  $\mathcal{C} = (f, g)$  be a  $(2^{nR}, n)$  deterministic code. Suppose that the channels  $W_{Y|X,S,T}(\cdot, \cdot, \theta_i)$  are symmetrizable for all  $i \in [1 : n]$ , and let  $J_t(s|x)$ ,  $t \in \mathcal{T}$ , be a set of conditional state distributions that satisfy (24). If  $R > 0$ , then

$$P_e^{(n)}(\tilde{q}, \theta^n, \mathcal{C}) \geq \frac{1}{4}, \quad (182)$$

for

$$\tilde{q}(s^n|\theta^n) = \frac{1}{2^{nR}} \sum_{m=1}^{2^{nR}} J_{\theta^n}(s^n|f^n(m, \theta^n)), \quad (183)$$

where  $J_{\theta^n}(s^n|x^n) = \prod_{i=1}^n J_{\theta_i}(s_i|x_i)$ .

For completeness, we give the proof below.

*Proof of Lemma 20.* Denote the codebook size by  $M = 2^{nR}$ , and the codewords by  $x^n(m, \theta^n) = f^n(m, \theta^n)$ .

Under the conditions of the lemma,

$$\begin{aligned} P_e^{(n)}(\tilde{q}, \theta^n, \mathcal{C}) &= \sum_{s^n \in \mathcal{S}^n} q(s^n|\theta^n) \frac{1}{M} \sum_{m=1}^M \sum_{y^n: g(y^n, \theta^n) \neq m} W^n(y^n|x^n(m, \theta^n), s^n, \theta^n) \\ &= \frac{1}{M^2} \sum_{\tilde{m}=1}^{2^{nR}} \sum_{s^n \in \mathcal{S}^n} J_{\theta^n}(s^n|x^n(\tilde{m}, \theta^n)) \sum_{m=1}^M \sum_{y^n: g(y^n, \theta^n) \neq m} W^n(y^n|x^n(m, \theta^n), s^n, \theta^n) \end{aligned} \quad (184)$$

where we have defined  $W^n \equiv W_{Y^n|X^n, S^n, T^n}$  for short notation. By switching between the summation indices  $m$  and  $\tilde{m}$ , we obtain

$$\begin{aligned} P_e^{(n)}(\tilde{q}, \theta^n, \mathcal{C}) &= \frac{1}{2M^2} \sum_{m, \tilde{m}} \sum_{y^n: g(y^n, \theta^n) \neq m} \sum_{s^n \in \mathcal{S}^n} W^n(y^n|x^n(m, \theta^n), s^n, \theta^n) J_{\theta^n}(s^n|x^n(\tilde{m}, \theta^n)) \\ &\quad + \frac{1}{2M^2} \sum_{m, \tilde{m}} \sum_{y^n: g(y^n, \theta^n) \neq \tilde{m}} \sum_{s^n \in \mathcal{S}^n} W^n(y^n|x^n(\tilde{m}, \theta^n), s^n, \theta^n) J_{\theta^n}(s^n|x^n(m, \theta^n)). \end{aligned} \quad (185)$$

Now, as the channel is memoryless,

$$\begin{aligned} \sum_{s^n \in \mathcal{S}^n} W^n(y^n|x^n(\tilde{m}, \theta^n), s^n, \theta^n) J_{\theta^n}(s^n|x^n(m, \theta^n)) &= \prod_{i=1}^n \sum_{s_i \in \mathcal{S}} W_{Y_i|X_i, S_i, T_i}(y_i|x_i(\tilde{m}, \theta^n), s_i, \theta_i) J_{\theta_i}(s_i|x_i(m, \theta^n)) \\ &= \prod_{i=1}^n \sum_{s_i \in \mathcal{S}} W_{Y_i|X_i, S_i, T_i}(y_i|x_i(m, \theta^n), s_i, \theta_i) J_{\theta_i}(s_i|x_i(\tilde{m}, \theta^n)) \\ &= \sum_{s^n \in \mathcal{S}^n} W^n(y^n|x^n(m, \theta^n), s^n, \theta^n) J_{\theta^n}(s^n|x^n(\tilde{m}, \theta^n)), \end{aligned} \quad (186)$$

where the second equality is due to (24). Therefore,

$$\begin{aligned} P_e^{(n)}(\tilde{q}, \theta^n, \mathcal{C}) &\geq \frac{1}{2M^2} \sum_{\tilde{m} \neq m} \sum_{s^n \in \mathcal{S}^n} \left[ \sum_{y^n: g(y^n, \theta^n) \neq m} W^n(y^n|x^n(m, \theta^n), s^n, \theta^n) J_{\theta^n}(s^n|x^n(\tilde{m}, \theta^n)) \right. \\ &\quad \left. + \sum_{y^n: g(y^n, \theta^n) \neq \tilde{m}} W^n(y^n|x^n(m, \theta^n), s^n, \theta^n) J_{\theta^n}(s^n|x^n(\tilde{m}, \theta^n)) \right] \\ &\geq \frac{1}{2M^2} \sum_{\tilde{m} \neq m} \sum_{s^n \in \mathcal{S}^n} \sum_{y^n \in \mathcal{Y}^n} W^n(y^n|x^n(m, \theta^n), s^n, \theta^n) J_{\theta^n}(s^n|x^n(\tilde{m}, \theta^n)) \\ &= \frac{M(M-1)}{2M^2} = \frac{1}{2} \left(1 - \frac{1}{M}\right). \end{aligned} \quad (187)$$

Assuming the sum rate is positive, we have that  $M \geq 2$ , hence  $P_e^{(n)}(\tilde{q}, \theta^n, \mathcal{C}) \geq \frac{1}{4}$ .  $\square$

Now, we are in position to prove the converse part of Theorem 6. Consider a sequence of  $(2^{nR}, n, \alpha_n)$  deterministic codes  $\mathcal{C}_n$  over the AVC with fixed parameters under input constraint  $\Omega$  and state constraint  $\Lambda$ , where  $\alpha_n \rightarrow 0$  as  $n \rightarrow \infty$ . In particular, the conditional probability of error given a state sequence  $s^n$  is bounded by

$$P_e^{(n)}(\mathcal{C}_n|s^n, \theta^n) \leq \alpha_n, \text{ for } s^n \in \mathcal{S}^n \text{ with } l^n(s^n) \leq \Lambda. \quad (188)$$

Let  $X^n = f(M, \theta^n)$  be the channel input sequence, and let  $Y^n$  be the corresponding output.

Consider using the same code over the compound channel with fixed parameters, *i.e.* where the jammer selects a state sequence at random according to a product distribution,  $\bar{S}^n \sim \prod_{i=1}^n q(\bar{s}_i|\theta_i)$ , under the *average* state constraint  $\frac{1}{n} \sum_{i=1}^n \mathbb{E}_q l(S_i) \leq \Lambda - \delta$ . Here, there is no state constraint with probability 1, as the jammer may select a sequence  $\bar{S}^n$  with  $l^n(\bar{S}^n) > \Lambda$ . Yet, the probability of error is bounded by

$$P_e^{(n)}(\bar{q}, \theta^n, \mathcal{C}_n) \leq \sum_{s^n: l^n(s^n) \leq \Lambda} \bar{q}^n(s^n|\theta^n) P_e^{(n)}(\mathcal{C}_n^\Gamma|s^n, \theta^n) + \Pr(l^n(\bar{S}^n) > \Lambda). \quad (189)$$

The first sum is bounded by (188), and the second term vanishes by the law of large numbers, since  $\bar{q} \in \overline{\mathcal{P}}_{\Lambda-\delta}(\mathcal{S}|\theta^\infty)$ . It follows that the code sequence of the constrained AVC achieves the same rate  $R$  over the compound channel  $W_{Y|X,\bar{S},T}$ . As in Appendix A, Fano's inequality implies that for every jamming strategy  $\bar{q}^n(s^n|\theta^n)$ ,

$$R \leq \min_{\bar{q}(s|t): \mathbb{E}_q l(S) \leq \Lambda} I_{\bar{q}}(X; Y|T) + \varepsilon_n, \quad (190)$$

with  $X \triangleq X_K$ ,  $T \equiv \theta_K$ ,  $Y \triangleq Y_K$ , where  $K$  is uniformly distributed over  $[1 : n]$ . Hence,  $T$  is distributed according to the type of the parameter sequence  $\theta^n$  (see (111)).

Returning to the original AVC, suppose that  $L_n^* > \Lambda$ . It remains to show that  $R > 0$  implies that  $\tilde{\Lambda}_n(P_{X|T}) \geq \Lambda$ . If the channels  $W_{Y|X,S,T}(\cdot|\cdot, \cdot, \theta_i)$  is non-symmetrizable for some  $i \in [1 : n]$ , then  $\tilde{\Lambda}_n(P_{X|T}) = +\infty$ , and there is nothing to show. Hence, consider the case where  $W_{Y|X,S,T}(\cdot|\cdot, \cdot, \theta_i)$  are symmetrizable for all  $i \in [1 : n]$ . Assume to the contrary that  $R > 0$  and  $\tilde{\Lambda}_n(P_{X|T}) < \Lambda$ . Hence, there exist conditional state distributions  $J_{\theta_i}(s|x)$  that symmetrize  $W_{Y|X,S,T}(\cdot|\cdot, \cdot, \theta_i)$ , such that

$$\tilde{\Lambda}_n(P_{X|T}) = \frac{1}{n} \sum_{i=1}^n \sum_{x,s} P_{X|T}(x|\theta_i) J_{\theta_i}(s|x) l(s) < \Lambda. \quad (191)$$

Now, consider the following jamming strategy. First, the jammer selects a codeword  $\tilde{X}^n$  from the codebook uniformly at random. Then, the jammer selects a sequence  $\tilde{S}^n$  at random, according to the conditional distribution

$$\Pr(\tilde{S}^n = s^n | \tilde{X}^n = x^n) = J_{\theta^n}(s^n|x^n) \triangleq \prod_{i=1}^n J_{\theta_i}(s_i|x_i). \quad (192)$$

At last, if  $l^n(\tilde{S}^n) \leq \Lambda$ , the jammer chooses the state sequence to be  $S^n = \tilde{S}^n$ . Otherwise, the jammer chooses  $S^n$  to be some sequence of zero cost. Such jamming strategy satisfies the state constraint  $\Lambda$  with probability 1.

To contradict our assumption that  $\tilde{\Lambda}(P_{X|T}) < \Lambda$ , we first show that  $\mathbb{E} l^n(\tilde{S}^n) = \tilde{\Lambda}(P_{X|T})$ . Observe that for every  $x^n \in \mathcal{X}^n$ ,

$$\mathbb{E} \left( l^n(\tilde{S}^n) | \tilde{X}^n = x^n \right) = \frac{1}{n} \sum_{i=1}^n \sum_{s \in \mathcal{S}} l(s) J_{\theta_i}(s|x_i). \quad (193)$$

Since  $\tilde{X}^n$  is distributed as  $X^n$ , we obtain

$$\mathbb{E} l^n(\tilde{S}^n) = \sum_{s \in \mathcal{S}} l(s) \cdot \frac{1}{n} \sum_{i=1}^n \mathbb{E} J_{\theta_i}(s|X_i) = \frac{1}{n} \sum_{i=1}^n \sum_{x,s} P_{X|T}(x|\theta_i) J_{\theta_i}(s|x) l(s) = \tilde{\Lambda}_n(P_{X|T}) < \Lambda. \quad (194)$$

Thus, by Chebyshev's inequality we have that for sufficiently large  $n$ ,

$$\Pr \left( l^n(\tilde{S}^n) > \Lambda \right) \leq \delta_0, \quad (195)$$

where  $\delta_0 > 0$  is arbitrarily small. Now, on the one hand, the probability of error is bounded by

$$\begin{aligned} P_e^{(n)}(q, \theta^n, \mathcal{C}_n) &\geq \Pr \left( g(Y^n, \theta^n) \neq M, l^n(\tilde{S}^n) \leq \Lambda \right) \\ &= \sum_{s^n: l^n(s^n) \leq \Lambda} \tilde{q}(s^n|\theta^n) P_e^{(n)}(\mathcal{C}_n|s^n, \theta^n), \end{aligned} \quad (196)$$

where  $\tilde{q}(s^n|\theta^n)$  is as defined in (183). On the other hand, the sequence  $\tilde{S}^n$  can be thought of as the state sequence of an AVC without a state constraint, hence, by Lemma 20,

$$\begin{aligned} \frac{1}{4} &\leq P_e^{(n)}(\tilde{q}, \theta^n, \mathcal{C}_n) \leq \sum_{s^n: l^n(s^n) \leq \Lambda} \tilde{q}(s^n|\theta^n) P_e^{(n)}(\mathcal{C}_n|s^n, \theta^n) + \Pr \left( l^n(\tilde{S}^n) > \Lambda \right) \\ &\leq \sum_{s^n: l^n(s^n) \leq \Lambda} \tilde{q}(s^n|\theta^n) P_e^{(n)}(\mathcal{C}_n|s^n, \theta^n) + \delta_0. \end{aligned} \quad (197)$$

Thus, by (196)-(197), the probability of error is bounded by  $P_e^{(n)}(q, \theta^n, \mathcal{C}_n) \geq \frac{1}{4} - \delta_0$ . As this cannot be the case for a code with vanishing probability of error, we deduce that the assumption is false, *i.e.*  $R > 0$  implies that  $\tilde{\Lambda}_n(P_{X|T}) \geq \Lambda$ .

If  $L_n^* < \Lambda$ , then  $\tilde{\Lambda}_n(P_{X|T}) < \Lambda$  for all  $P_{X|T}$  with  $\mathbb{E}\phi(X) \leq \Omega$ , and a positive rate cannot be achieved. This completes the converse proof.  $\square$

APPENDIX H  
PROOF OF COROLLARY 7

Assume that the AVC  $\mathcal{W}$  with fixed parameters satisfies the conditions of Corollary 7. Looking into the converse proof above, the following addition suffices. We show that for every code  $\mathcal{C}_n$  as in the converse proof above,  $\tilde{\Lambda}_n(P_{X|T}) = \Lambda$  implies that  $R = 0$ . Since there is only a polynomial number of types, we may consider  $P_{X|T}(x|t)$  to be the conditional type of  $f^n(m, \theta^n)$  given  $\theta^n$ , for all  $m \in [1 : 2^{nR}]$  (see [29, Problem 6.19]).

Suppose that  $\tilde{\Lambda}_n(P_{X|T}) = \Lambda$ , assume to the contrary that  $R > 0$ , and let  $J_i(s|x)$  be distributions that achieve the minimum in (27), *i.e.*

$$\tilde{\Lambda}_n(p) = \frac{1}{n} \sum_{i=1}^n \sum_{x,s} P_{X|T}(x|\theta_i) J_i(s|x) l(s) = \Lambda. \quad (198)$$

Based on the condition of the corollary, we may assume that  $J_i(s|x)$  is a 0-1 law, *i.e.*

$$J_i(s|x) = \begin{cases} 1 & \text{if } s = G_i(x), \\ 0 & \text{otherwise} \end{cases}, \quad (199)$$

for some deterministic function  $G_i : \mathcal{X} \rightarrow \mathcal{S}$ .

Recall that we have defined  $X = X_K$ ,  $Y = Y_K$  in the converse proof, where  $K$  is a uniformly distributed variable over  $[1 : n]$ . Thus, by (198),

$$\mathbb{E}(G_K(X)) = \frac{1}{n} \sum_{i=1}^n \sum_{x,s} p(x|\theta_i) J_i(s|x) l(s) = \Lambda. \quad (200)$$

Now, consider the following jamming strategy. First, the jammer selects a codeword  $\tilde{X}^n$  from the codebook uniformly at random. Then, given  $\tilde{X}^n = x^n$ , the jammer chooses the state sequence  $S^n = (G_i(x_i))_{i=1}^n$ . Observe that

$$l^n(S^n) = \frac{1}{n} \sum_{i=1}^n l(G_i(x_i)) = \mathbb{E}(G_K(X)) = \Lambda, \quad (201)$$

where the last equality is due to (200). Thus, the state sequence satisfies the state constraint. Now, observe that the jamming strategy  $S^n = (G(\tilde{X}_i))_{i=1}^n$  is equivalent to  $S^n \sim \tilde{q}(s^n|\theta^n)$  as in (183). Thus, by Lemma 20, we have that  $P_e^{(n)}(\tilde{q}, \mathcal{C}_n) \geq \frac{1}{4}$ , hence a positive rate cannot be achieved.  $\square$

APPENDIX I  
PROOF OF LEMMA 10

Suppose that  $L_n^* > \Lambda$ . The proof is similar to that of Lemma 4. We begin with the property in the lemma below.

*Lemma 21.* Let  $\omega_i^*$ ,  $\lambda_i^*$ ,  $\tilde{\lambda}_i^*$ ,  $i \in [1 : n]$ , be the parameters that achieve the saddle point in (43), *i.e.*

$$R_n(\mathcal{W}) = \frac{1}{n} \sum_{i=1}^n C_{\theta_i}(\omega_i^*, \lambda_i^*, \tilde{\lambda}_i^*). \quad (202)$$

Then, for every  $i, j \in [1 : n]$  such that  $\theta_i = \theta_j$ , we have that  $\omega_i^* = \omega_j^*$ ,  $\tilde{\lambda}_i^* = \tilde{\lambda}_j^*$ , and  $\lambda_i^* = \lambda_j^*$ .

*Proof of Lemma 21.* For every  $i \in [1 : n]$ , let  $p_i, q_i$  denote input and state distributions such that  $\mathbb{E}\phi(X_i) \leq \omega_i^*$ ,  $\tilde{\Lambda}_{\theta_i}(p_i) \geq \tilde{\lambda}_i^*$ ,  $\mathbb{E}l(S_i) \leq \lambda_i^*$  for  $X_i \sim p_i$ ,  $S_i \sim q_i$ . Now, suppose that  $\theta_i = \theta_j = t$ , and define

$$p'(x) = \frac{1}{2}[p_i(x) + p_j(x)], \quad q'(s) = \frac{1}{2}[q_i(s) + q_j(s)]. \quad (203)$$

Then,  $\mathbb{E}\phi(X') = \frac{1}{2}[\mathbb{E}\phi(X_i) + \mathbb{E}\phi(X_j)]$ ,  $\Lambda_t(p') = \frac{1}{2}[\Lambda_t(p_i) + \Lambda_t(p_j)]$ , and  $\mathbb{E}l(S') = \frac{1}{2}[\mathbb{E}l(S_i) + \mathbb{E}l(S_j)]$  for  $X' \sim p'$ ,  $S' \sim q'$ . Furthermore, since the mutual information is concave- $\cap$  in the input distribution and convex- $\cup$  in the state distribution, we have that

$$\begin{aligned} \frac{1}{2} [I_{q'}(X_i; Y_i|T_i = t) + I_{q'}(X_j; Y_j|T_j = t)] &\leq I_{q'}(X'; Y'|T' = t) \\ \frac{1}{2} [I_{q_i}(X'; Y'|T' = t) + I_{q_j}(X'; Y'|T' = t)] &\geq I_{q'}(X'; Y'|T' = t). \end{aligned} \quad (204)$$

Therefore, the saddle point distributions must satisfy  $p_i = p_j = p'$  and  $q_i = q_j = q'$ , hence  $\omega_i^* = \omega_j^*$ ,  $\tilde{\lambda}_i^* = \tilde{\lambda}_j^*$ , and  $\lambda_i^* = \lambda_j^*$ .  $\square$

Next, it can be inferred from Lemma 21 that

$$\begin{aligned}
R_n(\mathcal{W}) &= \min_{\substack{(\lambda_t)_{t \in \mathcal{T}}: \\ \sum_{t \in \mathcal{T}} P_T(t) \lambda_t \leq \Lambda}} \max_{\substack{(\omega_t)_{t \in \mathcal{T}}, (\tilde{\lambda}_t)_{t \in \mathcal{T}}: \\ \sum_{t \in \mathcal{T}} P_T(t) \omega_t \leq \Omega \\ \sum_{t \in \mathcal{T}} P_T(t) \tilde{\lambda}_t \geq \Lambda}} \sum_{t \in \mathcal{T}} P_T(t) C_t(\omega_t, \lambda_t) \\
&= \min_{\substack{(\lambda_t)_{t \in \mathcal{T}}, q(s|t): \\ \mathbb{E}_q[l(S)|T=t] \leq \lambda_t \\ \sum_{t \in \mathcal{T}} P_T(t) \lambda_t \leq \Lambda}} \max_{\substack{(\omega_t)_{t \in \mathcal{T}}, (\tilde{\lambda}_t)_{t \in \mathcal{T}}, p(x|t): \\ \mathbb{E}[\phi(X)|T=t] \leq \omega_t, \tilde{\Lambda}(p, t) \geq \tilde{\lambda}_t \\ \sum_{t \in \mathcal{T}} P_T(t) \omega_t \leq \Omega, \sum_{t \in \mathcal{T}} P_T(t) \tilde{\lambda}_t \geq \Lambda}} I_q(X; Y|T) \\
&= \min_{q(s|t): \mathbb{E}_q l(S) \leq \Lambda} \max_{\substack{p(x|t): \mathbb{E} \phi(X) \leq \Omega, \\ \tilde{\Lambda}_n(p) \geq \Lambda}} I_q(X; Y|T) = C_n(\mathcal{W}), \tag{205}
\end{aligned}$$

where  $P_T$  is the type of the parameter sequence  $\theta^n$ . The second equality follows from the definition of  $C_t(\omega_t, \lambda_t, \tilde{\lambda}_t)$  in (44), using the minimax theorem [96] to switch between the order of the minimum and maximum. In the third line, we eliminate the slack variables  $\lambda_i$ ,  $\omega_i$ , and  $\tilde{\lambda}_i$ , replacing  $\mathbb{E}_q l(S_i)$ ,  $\mathbb{E} \phi(X_i)$ , and  $\tilde{\Lambda}(p, \theta_i)$ , respectively. The last equality holds by the definition of  $C_n(\mathcal{W})$  in (29).  $\square$

## APPENDIX J ANALYSIS OF EXAMPLE 2

Consider the fading AVC in Example 2. To show the direct part with random codes, set the conditional input distribution  $X \sim \mathcal{N}(0, \omega(t))$  given  $T = t$  in (21). Then, for every  $t \in \mathcal{T}$ ,

$$I_q(X; Y|T = t) \geq \frac{1}{2} \log \left( 1 + \frac{t^2 \omega(t)}{\lambda'(t) + \sigma^2} \right), \tag{206}$$

where we have denoted  $\lambda'(t) \triangleq \mathbb{E}(S^2|T = t)$ . The last inequality holds since Gaussian noise is known to be the worst additive noise under variance constraint [34, Lemma II.2]. The direct part follows. As for the converse part, consider a jamming scheme where the state is drawn according to the conditional distribution  $S \sim \mathcal{N}(0, \lambda(t))$  given  $T = t$ . Then, the proof follows from Shannon's classic result on the Gaussian channel  $Y = tX + V$  with  $V \sim \mathcal{N}(0, \lambda(t) + \sigma^2)$ .

We move to the deterministic code capacity. By Definition 4, the constant-parameter channel  $W_{Y|X, S, T=t}$  is symmetrized by a conditional pdf  $\varphi(s|x)$  if

$$\int_{-\infty}^{\infty} \varphi(s|x_2) f_Z(y - tx_1 - s) ds = \int_{-\infty}^{\infty} \varphi(s|x_1) f_Z(y - tx_2 - s) ds, \quad \forall x_1, x_2, y \in \mathbb{R}, \tag{207}$$

where  $f_Z(z) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-z^2/2\sigma^2}$ . Equivalently, the constant-parameter channel is symmetrized by  $\varphi_x(s) \equiv \varphi(s|x)$  if

$$\int_{-\infty}^{\infty} \varphi_0(s) f_Z(y - tx - s) ds = \int_{-\infty}^{\infty} \varphi_x(s) f_Z(y - s) ds, \tag{208}$$

for all  $x, y \in \mathbb{R}$ . By substituting  $z = y - tx - s$  in the LHS, and  $\bar{z} = y - s$  in the RHS, we have

$$\int_{-\infty}^{\infty} \varphi_0(y - tx - z) f_Z(z) dz = \int_{-\infty}^{\infty} \varphi_x(y - \bar{z}) f_Z(\bar{z}) d\bar{z}. \tag{209}$$

For every  $x \in \mathbb{R}$ , define the random variable  $\bar{S}(x) \sim \varphi_x$ . We note that the RHS is the convolution of the pdfs of the random variables  $Z$  and  $\bar{S}(x)$ , while the LHS is the convolution of the pdfs of the random variables  $Z$  and  $\bar{S}(0) + x$ . This is not surprising since the channel output  $Y$  is a sum of independent random variables, and thus the pdf of  $Y$  is a convolution of pdfs. It follows that  $\varphi_0(y - tx) = \varphi_x(y)$ , and by plugging  $s$  instead of  $y$ , we have that  $\varphi_x$  symmetrizes the constant-parameter channel  $W_{Y|X, S, T=t}$  if and only if

$$\varphi_x(s) = \varphi_0(s - tx). \tag{210}$$

Then, the corresponding state cost satisfies

$$\begin{aligned}
\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X|T}(x|t) \varphi_x(s) s^2 dx ds &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X|T}(x|t) \varphi_0(s - tx) s^2 ds dx \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X|T}(x|t) \varphi_0(a) (a + tx)^2 da dx \\
&= \int_{-\infty}^{\infty} \left[ \int_{-\infty}^{\infty} (tx + a)^2 f_{X|T}(x|t) dx \right] \varphi_0(a) da \tag{211}
\end{aligned}$$

where the second equality follows by the integral substitution of  $a = s - tx$ . Observe that the bracketed integral can be expressed as

$$\int_{-\infty}^{\infty} (tx + a)^2 f_{X|T}(x|t) dx = \mathbb{E}[(tX + a)^2 | T = t] = t^2 \mathbb{E}[X^2 | T = t] + a^2. \quad (212)$$

Thus, by (211),

$$\begin{aligned} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X|T}(x|t) \varphi_x(s) s^2 dx ds &= t^2 \mathbb{E}[X^2 | T = t] + \int_{-\infty}^{\infty} a^2 \varphi_0(a) da \\ &\geq t^2 \mathbb{E}[X^2 | T = t]. \end{aligned} \quad (213)$$

Note that the last inequality holds for any  $\varphi_x$  which symmetrizes the channel, and in particular for  $\hat{\varphi}_x(s) = \delta(s - tx)$ , where  $\delta(\cdot)$  is the Dirac delta function. In addition, since  $\hat{\varphi}_0$  gives probability 1 to  $S = 0$ , we have that (213) holds with equality for  $\hat{\varphi}_x$ , and thus,

$$\tilde{\Lambda}(F_{X|T}) = \frac{1}{n} \sum_{i=1}^n t^2 \mathbb{E}[X^2 | T = t] = \sum_{t \in \mathcal{T}} P_T(t) t^2 \mathbb{E}[X^2 | T = t] = \mathbb{E}(T^2 \omega(T)), \quad (214)$$

with  $\omega(t) \equiv \mathbb{E}[X^2 | T = t]$ . Hence,

$$L_n^* = \max_{\omega(t) : \mathbb{E}\omega(T) \leq \Omega} \mathbb{E}(T^2 \omega(T)). \quad (215)$$

Having shown that the minimum in (27) is attained by a 0-1 law, we have by Corollary 7 that the capacity of the fading AVC is  $\mathbb{C}(\mathcal{W}) = \liminf C_n(\mathcal{W})$ , with

$$C_n(\mathcal{W}) = \begin{cases} \min_{F_{S|T} : \mathbb{E}S^2 \leq \Lambda} \max_{\substack{F_{X|T} : \mathbb{E}X^2 \leq \Omega, \\ \mathbb{E}(T^2 X^2) \geq \Lambda}} I_q(X; Y | T) & \text{if } \max_{\omega(t) : \mathbb{E}\omega(T) \leq \Omega} \mathbb{E}(T^2 \omega(T)) > \Lambda, \\ 0 & \text{if } \max_{\omega(t) : \mathbb{E}\omega(T) \leq \Omega} \mathbb{E}(T^2 \omega(T)) \leq \Lambda \end{cases}. \quad (216)$$

To show the direct part, we only need to consider the case where  $\max_{\omega(t) : \mathbb{E}\omega(T) \leq \Omega} \mathbb{E}(T^2 \omega(T)) > \Lambda$ . Then, set the conditional input distribution  $X \sim \mathcal{N}(0, \omega(t))$  given  $T = t$  in (216). As in the direct part with random codes,

$$I_q(X; Y | T = t) \geq \frac{1}{2} \log \left( 1 + \frac{t^2 \omega(t)}{\lambda'(t) + \sigma^2} \right), \quad (217)$$

with  $\lambda'(t) \triangleq \mathbb{E}(S^2 | T = t)$ , since Gaussian noise is the worst additive noise under variance constraint [34, Lemma II.2]. The direct part follows. As for the converse part, for the conditional distribution  $S \sim \mathcal{N}(0, \lambda(t))$  given  $T = t$ , we have that

$$I_q(X; Y | T = t) \leq \frac{1}{2} \log \left( 1 + \frac{t^2 \omega'(t)}{\lambda(t) + \sigma^2} \right), \quad (218)$$

with  $\omega'(t) \triangleq \mathbb{E}(X^2 | T = t)$ , since the Gaussian distribution maximizes the differential entropy. The proof follows.  $\square$

## APPENDIX K PROOF OF LEMMA 13

### Part 1

Since  $\sum_{j'=1}^d P_{j'}^* = \Omega > 0$ , there must be some  $j \in [1 : d]$  such that  $P_j^* = \alpha - (N_j^* + \sigma_j^2) > 0$ , thus  $\alpha > N_j^* + \sigma_j^2$ . If  $N_j^* = 0$ , then it follows that  $\beta \leq \sigma_j^2$ , hence

$$\alpha > N_j^* + \sigma_j^2 = \sigma_j^2 \geq \beta. \quad (219)$$

Otherwise,  $N_j^* = \beta - \sigma_j^2 > 0$ , thus by the assumption  $P_j^* > 0$ , we have that

$$0 < P_j^* = \alpha - (N_j^* + \sigma_j^2) = \alpha - \beta. \quad (220)$$

### Part 2

Assume to the contrary that  $N_j^* = \beta - \sigma_j^2 > 0$  and  $P_j^* = 0$ . The assumption  $P_j^* = 0$  implies that  $\alpha \leq N_j^* + \sigma_j^2 = \beta$ , in contradiction to part 1 of the Lemma. Hence, the assumption is false, and  $N_j^* > 0$  implies that  $P_j^* > 0$ .

Part 3 and Part 4

By the definition of  $N_j^*$  in (72), we have that  $N_j^* + \sigma_j^2 = \max(\beta, \sigma_j^2)$  for all  $j \in [1 : d]$ . Thus,

$$P_j^* + N_j^* + \sigma_j^2 = \max(\beta, \sigma_j^2) + [\alpha - \max(\beta, \sigma_j^2)]_+ = \max(\alpha, \beta, \sigma_j^2) = \max(\alpha, \sigma_j^2), \quad (221)$$

where the last equality is due to part 1. Part 4 immediately follows.  $\square$

APPENDIX L  
PROOF OF LEMMA 14

Let  $X^d$  be a zero mean random vector with the covariance matrix  $K_X$ . Observe that by (84), the AVGPC is symmetrized by a conditional pdf  $\varphi_{x^d}(s^d) = \varphi(s^d|x^d)$  if

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \varphi_0(s^d) f_{Z^d}(y^d - x^d - s^d) ds^d = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \varphi_{x^d}(s^d) f_{Z^d}(y^d - s^d) ds^d, \quad (222)$$

for all  $x^d, y^d \in \mathbb{R}^d$ . By substituting  $z^d = y^d - x^d - s^d$  in the LHS, and  $\bar{z}^d = y^d - s^d$  in the RHS, this is equivalent to

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \varphi_0(y^d - x^d - z^d) f_{Z^d}(z^d) dz^d = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \varphi_{x^d}(y^d - \bar{z}^d) f_{Z^d}(\bar{z}^d) d\bar{z}^d. \quad (223)$$

For every  $x^d \in \mathbb{R}^d$ , define the random vector  $\bar{S}^d(x^d) \sim \varphi_{x^d}$ . We note that the RHS is the convolution of the pdfs of the random vectors  $Z^d$  and  $\bar{S}^d(x^d)$ , while the LHS is the convolution of the pdfs of the random vectors  $Z^d$  and  $\bar{S}^d(0) + x^d$ . This is not surprising since the channel output  $Y^d$  is a sum of independent random vectors, and thus the pdf of  $Y^d$  is a convolution of pdfs. It follows that  $\varphi_0(y^d - x^d) = \varphi_{x^d}(y^d)$ , and by plugging  $s^d$  instead of  $y^d$ , we have that  $\varphi_{x^d}$  symmetrizes the AVGPC if and only if

$$\varphi_{x^d}(s^d) = \varphi_0(s^d - x^d). \quad (224)$$

Then, the corresponding state cost satisfies

$$\begin{aligned} & \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X^d}(x^d) \varphi_{x^d}(s^d) \|s^d\|^2 dx^d ds^d \\ &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X^d}(x^d) \varphi_0(s^d - x^d) \|s^d\|^2 ds^d dx^d \\ &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X^d}(x^d) \varphi_0(a^d) \|a^d + x^d\|^2 da^d dx^d \\ &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left[ \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \|x^d + a^d\|^2 f_{X^d}(x^d) dx^d \right] \varphi_0(a^d) da^d \end{aligned} \quad (225)$$

where the second equality follows by the integral substitution of  $a^d = s^d - x^d$ . Observe that the bracketed integral can be expressed as

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \|x^d + a^d\|^2 f_{X^d}(x^d) dx^d = \mathbb{E} \|X^d + a^d\|^2 = \text{tr}(K_X) + \|a^d\|^2. \quad (226)$$

Thus, by (225),

$$\begin{aligned} & \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X^d}(x^d) \varphi_{x^d}(s^d) \|s^d\|^2 dx^d ds^d \\ &= \text{tr}(K_X) + \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \|a^d\|^2 \varphi_0(a^d) da^d \\ &\geq \text{tr}(K_X). \end{aligned} \quad (227)$$

Note that the last inequality holds for any  $\varphi_{x^d}$  which symmetrizes the channel. Now, observe that (224) holds for  $\hat{\varphi}_{x^d}(s^d) = \delta(s^d - x^d)$ , where  $\delta(\cdot)$  is the Dirac delta function, hence  $\hat{\varphi}_{x^d}$  symmetrizes the channel. In addition, since  $\hat{\varphi}_0$  gives probability 1 to  $S^d = 0$ , we have that (227) holds with equality for  $\hat{\varphi}_{x^d}$ , and thus,  $\tilde{\Lambda}(F_{X^d}) = \text{tr}(K_X)$ .  $\square$

APPENDIX M  
PROOF OF THEOREM 15

Consider the AVGPC under input constraint  $\Omega$  and state constraint  $\Lambda$ .

### Achievability Proof

Assume that  $\Omega > \Lambda$ . We show that  $\mathbb{C}(\Sigma) \geq \mathbb{C}(\Sigma) = \mathbb{C}^*(\Sigma)$ . By [28, Theorem 3], if there exists an input distribution  $F_{X^d}$  such that  $\tilde{\Lambda}(F_{X^d}) > \Lambda$ , then the capacity is given by

$$\mathbb{C}(\Sigma) = \max_{\substack{F_{X^d} : \sum_{j=1}^d P_j \leq \Omega \\ \tilde{\Lambda}(F_{X^d}) \geq \Lambda}} \min_{F_{S^d} : \sum_{j=1}^d N_j \leq \Lambda} I(X^d; Y^d), \quad (228)$$

where  $P_j = \mathbb{E}X_j^2$  and  $N_j = \mathbb{E}S_j^2$ .

Consider the input distribution  $F_{X^d}$  of a Gaussian vector  $X^d \sim \mathcal{N}(\mathbf{0}, K_X)$ , where the covariance matrix is given by  $K_X = \text{diag}(P_1^*, \dots, P_d^*)$ . By Lemma 14, we have that

$$\tilde{\Lambda}(F_{X^d}) = \text{tr}(K_X) = \sum_{j=1}^d P_j^* = \Omega. \quad (229)$$

Having assumed that  $\Omega > \Lambda$ , it follows that  $\tilde{\Lambda}(F_{X^d}) > \Lambda$ , hence (228) applies. Then, setting  $X^d \sim \mathcal{N}(\mathbf{0}, K_X)$  yields

$$\mathbb{C}(\Sigma) \geq \min_{F_{S^d} : \sum_{j=1}^d N_j \leq \Lambda} I(X^d; Y^d) \quad (230)$$

$$\geq \min_{F_{S^d} : \sum_{j=1}^d N_j \leq \Lambda} \sum_{j=1}^d I(X_j; Y_j) \quad (231)$$

$$\geq \min_{F_{S^d} : \sum_{j=1}^d N_j \leq \Lambda} \sum_{j=1}^d \frac{1}{2} \log \left( 1 + \frac{P_j^*}{N_j + \sigma_j^2} \right), \quad (232)$$

where the second inequality holds as  $X_1, \dots, X_d$  are independent and since conditioning reduces entropy, and the last inequality holds since Gaussian noise is known to be the worst additive noise under variance constraint [34, Lemma II.2].

From this point, we use the considerations given in [61]. To prove the direct part, it remains to show that the assignment of  $N_j = N_j^*$ , for  $j \in [1 : d]$ , is optimal in the RHS of (232), where  $N_j^*$  are as defined in (72)-(73). An assignment of  $N_1, \dots, N_d$  is optimal if and only if it satisfies the KKT optimality conditions [20, Section 5.5.3],

$$\sum_{j'=1}^d N_{j'} = \Lambda, \quad N_j \geq 0, \quad (233)$$

$$\frac{P_j^*}{(N_j + \sigma_j^2) \cdot (N_j + \sigma_j^2 + P_j^*)} \leq \theta, \quad (234)$$

$$\left( \theta - \frac{P_j^*}{(N_j + \sigma_j^2) \cdot (N_j + \sigma_j^2 + P_j^*)} \right) N_j = 0, \quad (235)$$

for  $j \in [1 : d]$ , where  $\theta > 0$  is a Lagrange multiplier.

We claim that the conditions are met by

$$\theta = \theta^* \triangleq \frac{\alpha - \beta}{\alpha\beta}, \quad \text{and } N_j = N_j^*, \quad \text{for } j \in [1 : d]. \quad (236)$$

Condition (233) is met by the definition of  $N_j^*$ ,  $j \in [1 : d]$ , in (72)-(73). Let  $j \in [1 : d]$  be a given channel index. We consider the following cases. Suppose that  $N_j^* = 0$ . Then, Condition (235) is clearly satisfied. Now, if  $P_j^* = 0$ , then Condition (234) is satisfied since  $\alpha > \beta$  by part 1 of Lemma 13. Otherwise,  $0 < P_j^* = \alpha - (N_j^* + \sigma_j^2) = \alpha - \sigma_j^2$ , and then

$$\frac{P_j^*}{(N_j + \sigma_j^2) \cdot (N_j + \sigma_j^2 + P_j^*)} = \frac{\alpha - \sigma_j^2}{\sigma_j^2 \alpha} \leq \frac{\alpha - \beta}{\alpha\beta} = \theta^*, \quad (237)$$

where the last inequality holds since  $N_j^* = 0$  only if  $\beta \leq \sigma_j^2$ . Thus, Condition (234) is satisfied.

Next, suppose that  $N_j^* > 0$ , hence  $N_j^* + \sigma_j^2 = \beta$ . By part 2 of Lemma 13, this implies that  $P_j^* > 0$ , i.e.  $P_j^* = \alpha - (N_j^* + \sigma_j^2) = \alpha - \beta$ . Thus,

$$\frac{P_j^*}{(N_j + \sigma_j^2) \cdot (N_j + \sigma_j^2 + P_j^*)} = \frac{\alpha - \beta}{\beta \cdot \alpha} = \theta^*, \quad (238)$$

and thus Condition (234) is satisfied with equality, and Condition (235) is satisfied as well.

As the KKT conditions are satisfied under (236), we deduce that the assignment of  $N_j = N_j^*$ ,  $j \in [1 : d]$ , minimizes the RHS of (232). Together with (232), this implies that  $\mathbb{C}(\Sigma) \geq \mathbb{C}^*(\Sigma)$  for  $\Omega > \Lambda$ .

### Converse Proof

We use a similar technique as in [32] (see also [37, 16]). In general, the deterministic code capacity is bounded by the random code capacity, hence  $\mathbb{C}(\Sigma) \leq \mathbb{C}^*(\Sigma) = \mathbb{C}^\star(\Sigma)$ , by Theorem 12. It remains to show that if  $\Omega \leq \Lambda$ , then the capacity is zero. Suppose that  $\Omega \leq \Lambda$ , and assume to the contrary that there exists an achievable rate  $R > 0$ . Then, there exists a sequence of  $(2^{nR}, n, \varepsilon_n)$  codes  $\mathcal{C}_n = (\mathbf{f}^d, g)$  for the AVGPC such that  $\varepsilon_n \rightarrow 0$  as  $n \rightarrow \infty$ , where the size of the message set is at least 2, i.e.  $M \triangleq 2^{nR} \geq 2$ .

Consider a jammer who chooses the state sequence from the codebook uniformly at random, i.e.  $\mathbf{S}^d = \mathbf{f}^d(M')$ , where  $M'$  is uniformly distributed over  $[1 : M]$ . This choice meets the state constraint, since the square norm of the state sequence is  $\|\mathbf{S}^d\|^2 \leq \Omega \leq \Lambda$ . The average probability of error is then bounded by

$$P_e^{(n)}(F_{\mathbf{S}^d}, \mathcal{C}) = \frac{1}{M^2} \sum_{m=1}^M \sum_{m'=1}^M \int_{\mathcal{D}_e(m, m')} f_{\mathbf{Z}^d}(\mathbf{z}^d) d\mathbf{z}^d, \quad (239)$$

where  $f_{\mathbf{Z}^d}(\mathbf{z}^d) = \prod_{j=1}^d \frac{1}{(2\pi\sigma_j^2)^{n/2}} e^{-\|\mathbf{z}^d\|^2/2\sigma_j^2}$ , and

$$\mathcal{D}_e(m, m') = \{\mathbf{z}^d : g(\mathbf{f}^d(m) + \mathbf{f}^d(m') + \mathbf{z}^d) \neq m\}. \quad (240)$$

By interchanging the summation variables  $m$  and  $m'$ , we now have that

$$\begin{aligned} P_e^{(n)}(F_{\mathbf{S}^d}, \mathcal{C}) &= \frac{1}{2M^2} \sum_{m, m'} \int_{\mathcal{D}_e(m, m')} f_{\mathbf{Z}^d}(\mathbf{z}^d) d\mathbf{z}^d + \frac{1}{2M^2} \sum_{m, m'} \int_{\mathcal{D}_e(m', m)} f_{\mathbf{Z}^d}(\mathbf{z}^d) d\mathbf{z}^d \\ &\geq \frac{1}{2M^2} \sum_{m, m' : m \neq m'} \int_{\mathcal{D}_e(m, m') \cup \mathcal{D}_e(m, m')} f_{\mathbf{Z}^d}(\mathbf{z}^d) d\mathbf{z}^d. \end{aligned} \quad (241)$$

Next, observe that for  $m \neq m'$ ,  $\mathcal{D}_e(m, m') \cup \mathcal{D}_e(m, m') = \mathbb{R}^{nd}$ , and thus the probability of error is lower bounded by

$$P_e^{(n)}(F_{\mathbf{S}^d}, \mathcal{C}) \geq \frac{M(M-1)}{2M^2} \geq \frac{1}{4}, \quad (242)$$

where the last inequality holds since  $M \geq 2$ . Hence, the assumption is false and a positive rate cannot be achieved when  $\Omega \leq \Lambda$ . This completes the proof of the converse part.  $\square$

### APPENDIX N PROOF OF THEOREM 16

Consider the AVC with colored Gaussian noise. First, we show that the problem can be transformed into that of an AVC with fixed parameters. Then, we derive a limit expression for the random code capacity, and prove the capacity characterization in Theorem 16 using the Toeplitz matrix properties in the auxiliary lemma below. To derive the deterministic code capacity, we use similar symmetrizability and optimization arguments as in our proofs for the Gaussian product channel.

*Lemma 22.* [35, Section 2.3] (see also [43, 53] [39, Section 8.5]) Let  $\Psi_Z(\omega)$  be the power spectral density of a zero mean stationary process  $\{Z_i\}_{i=1}^\infty$ . Assume that  $\Psi_Z : [-\pi, \pi] \rightarrow [0, \nu]$  is bounded and integrable, for some  $\nu > 0$ , and denote the auto-correlation function by

$$r_Z(\ell) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \Psi_Z(\omega) e^{j\omega\ell} d\omega, \quad \ell = 0, 1, 2, \dots \quad (243)$$

with  $j = \sqrt{-1}$ . For a sequence  $\mathbf{Z}$  of length  $n$ , let  $\sigma_1^2, \dots, \sigma_n^2$  denote the eigenvalues of the  $n \times n$  covariance matrix  $K_Z$ , where  $K_Z(i, j) = r_Z(|i-j|)$  for  $i, j \in [1 : n]$ . Then, for every real, monotone non-increasing, and bounded function  $G : [0, \nu] \rightarrow [0, \eta]$ ,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n G(\sigma_i^2) = \frac{1}{2\pi} \int_{-\pi}^{\pi} G(\Psi_Z(\omega)) d\omega \quad (244)$$

if the integral exists.

#### A. Transformation to AVC with Fixed Parameters

Let  $K_Z$  denote the  $n \times n$  covariance matrix of the noise sequence  $\mathbf{Z}$ . Consider the eigen decomposition of the covariance matrix  $K_Z$ , and denote the eigenvector and eigenvalue matrices by  $Q$  and  $\Sigma$ , respectively, i.e.

$$K_Z = Q\Sigma Q^T, \quad \text{where } QQ^T = I \text{ and } \Sigma = \text{diag}\{\sigma_1^2, \dots, \sigma_n^2\}. \quad (245)$$

We claim that the capacity of the AVC with colored Gaussian noise is the same as the capacity of the following AVC,

$$\mathbf{Y}' = \mathbf{X}' + \mathbf{Z}' + \mathbf{S}', \quad (246)$$

where  $\mathbf{X}' = Q^T \mathbf{X}$ ,  $\mathbf{Z}' = Q^T \mathbf{Z}$ , and  $\mathbf{S}' = Q^T \mathbf{S}$ . Since  $Q$  is a unitary matrix, *i.e.*  $Q^{-1} = Q^T$ , the input and state constraints remain the same, as  $\|\mathbf{X}'\|^2 = (\mathbf{X}')^T \mathbf{X}' = \mathbf{X}^T Q Q^T \mathbf{X} = \mathbf{X}^T \mathbf{X} = \|\mathbf{X}\|^2 \leq n\Omega$ , and similarly,  $\|\mathbf{S}'\|^2 = \|\mathbf{S}\|^2 \leq n\Lambda$ . Furthermore, the noise covariance matrix is now

$$K_{Z'} = Q^T K_Z Q = \Sigma = \text{diag}\{\sigma_1^2, \dots, \sigma_n^2\}. \quad (247)$$

This transformation can be thought of as a linear system, which is *not* time invariant. Hence, the noise of the transformed channel is a Gaussian process, but it is non-stationary. Thereby, the input-output relation above specifies a time varying channel,  $\{F_{Y_1, \dots, Y_n | X_1, \dots, X_n, S_1, \dots, S_n}\}_{n=1}^{\infty}$ . From operational perspective, if there exists a  $(2^{nR}, n, \varepsilon)$  code  $\mathcal{C} = (\mathbf{f}, g)$  for the original AVC with colored Gaussian noise, then the code  $\mathcal{C}' = (\mathbf{f}', g')$ , given by  $\mathbf{f}'(m) = Q^T \mathbf{f}(m)$  and  $g'(\mathbf{y}') = g(Q\mathbf{y}')$ , is a  $(2^{nR}, n, \varepsilon)$  code for the transformed AVC in (246). Similarly, if there exists a  $(2^{nR}, n, \varepsilon)$  code  $\mathcal{C}' = (\mathbf{f}', g')$  for the transformed AVC, then the code  $\mathcal{C} = (\mathbf{f}, g)$ , given by  $\mathbf{f}(m) = Q\mathbf{f}'(m)$  and  $g(\mathbf{y}) = g'(Q^T \mathbf{y})$ , is a  $(2^{nR}, n, \varepsilon)$  code for the original AVC. Thus, the original AVC and the transformed AVC have the same operational capacity.

Therefore, we can assume without loss of generality that the noise sequence has independent components  $Z_i \sim \mathcal{N}(0, \sigma_i^2)$ ,  $i \in [1 : n]$ . Assume, at first, that  $\sigma_i^2 \in \mathcal{T}$  for  $i \in [1 : n]$ , with some set  $\mathcal{T}$  of finite size, which does not grow with  $n$ , and that  $\sigma_i^2 > \delta$ , where  $\delta > 0$  is arbitrarily small. Hence, observe that the channel in (246) is equivalent to a channel  $W_{Y'' | X'', S'', T''}$  with fixed parameters, specified by

$$Y'' = X'' + S'' + Z_t'', \text{ where } Z_t'' \sim \mathcal{N}(0, t^2) \quad (248)$$

with the parameter sequence  $\sigma_1, \sigma_2, \dots$ . It is left to determine the random code capacity and deterministic code capacity of the Gaussian AVC with fixed parameters in (248). Although we previously assumed in Sections II and III that the input, state, and output alphabets are finite, our results can be extended to the continuous case as well, using standard discretization techniques [15, 5] [36, Section 3.4.1].

Now, consider the double water filling allocation,

$$b_i^* = [\beta' - \sigma_i^2]_+, \quad (249)$$

$$a_i^* = [\alpha' - (b_i^* + \sigma_i^2)]_+, \quad (250)$$

for  $i \in [1 : n]$ , where  $\beta' > 0$  and  $\alpha' > 0$  are chosen to satisfy  $\frac{1}{n} \sum_{i=1}^n [\beta' - \sigma_i^2]_+ = \Lambda$  and  $\frac{1}{n} \sum_{i=1}^n [\alpha' - (b_i^* + \sigma_i^2)]_+ = \Omega$ , respectively. Define

$$\mathbb{C}_n^*(K_Z) \triangleq \frac{1}{2n} \sum_{i=1}^n \log \left( 1 + \frac{a_i^*}{b_i^* + \sigma_i^2} \right). \quad (251)$$

## B. Random Code Capacity

Now that we have shown that the problem reduces to that of an AVC with fixed parameters, we have by Corollary 5 that the random code capacity is given by

$$\mathbb{C}^*(\Psi_Z) = \liminf_{n \rightarrow \infty} \max_{\substack{P_1, \dots, P_n : \\ \frac{1}{n} \sum_{i=1}^n P_i \leq \Omega}} \min_{\substack{N_1, \dots, N_n : \\ \frac{1}{n} \sum_{i=1}^n N_i \leq \Lambda}} \frac{1}{n} \sum_{i=1}^n \mathbb{C}_{\sigma_i}^*(P_i, N_i), \quad (252)$$

where  $\mathbb{C}_{\sigma}^*(P, N)$  is the random code capacity of the traditional AVC under input constraint  $P$  and state constraint  $N$ . Hughes and Narayan [60] showed that the random code capacity of such a channel, where the noise sequence is i.i.d.  $\sim \mathcal{N}(0, \sigma^2)$ , is given by

$$\mathbb{C}_{\sigma}^*(P, N) = \frac{1}{2} \log \left( 1 + \frac{P}{N + \sigma^2} \right). \quad (253)$$

Hence, for the AVC with colored Gaussian noise,

$$\mathbb{C}^*(\Psi_Z) = \liminf_{n \rightarrow \infty} \min_{\substack{N_1, \dots, N_n : \\ \frac{1}{n} \sum_{i=1}^n N_i \leq \Lambda}} \max_{\substack{P_1, \dots, P_n : \\ \frac{1}{n} \sum_{i=1}^n P_i \leq \Omega}} \frac{1}{2n} \sum_{i=1}^n \log \left( 1 + \frac{P_i}{N_i + \sigma_i^2} \right). \quad (254)$$

Next, observe that this is the same min-max optimization as for the AVGPC in (78), due to [61], with  $d \leftarrow n$ ,  $\Omega \leftarrow (n\Omega)$ ,  $\Lambda \leftarrow (n\Lambda)$ . Therefore, by Theorem 12 [61] and (254),

$$\mathbb{C}^*(\Psi_Z) = \liminf_{n \rightarrow \infty} \mathbb{C}_n^*(K_Z). \quad (255)$$

Given a bounded power spectral density  $\Psi_Z : [-\pi, \pi] \rightarrow [0, \nu]$ , define a function  $G : [0, \nu] \rightarrow [0, \eta]$  by

$$G(x) = \frac{1}{2} \log \left( 1 + \frac{[\alpha' - [\beta' + x]_+]_+}{[\beta' - x]_+ + x} \right) = \begin{cases} \frac{1}{2} \log \left( \frac{\alpha'}{\beta'} \right) & \text{if } x < \beta' \\ \frac{1}{2} \log \left( \frac{\alpha'}{x} \right) & \text{if } \beta' \leq x < \alpha' \\ 0 & \text{if } x \geq \alpha' \end{cases} \quad (256)$$

and observe that

$$C_n^*(K_Z) = \frac{1}{n} \sum_{i=1}^n G(\sigma_i^2). \quad (257)$$

As  $G(x)$  is non-increasing and bounded by  $\eta = \frac{1}{2} \log[1 + \Omega/\delta]$ , we have by Lemma 22 that

$$\liminf_{n \rightarrow \infty} C_n^*(K_Z) = \frac{1}{2\pi} \int_{-\pi}^{\pi} G(\Psi_Z(\omega)) d\omega. \quad (258)$$

Observing that the function defined in (256) is also continuous, while  $\Psi_Z(\omega)$  is bounded and integrable, it follows that the integral exists [86, Theorem 6.11]. Plugging (256) into the RHS of (258), we obtain

$$\liminf_{n \rightarrow \infty} C_n^*(K_Z) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{2} \log \left( 1 + \frac{[\alpha - [\beta + \Psi_Z(\omega)]_+]_+}{[\beta - \Psi_Z(\omega)]_+ + \Psi_Z(\omega)} \right) d\omega \quad (259)$$

where  $\beta$  and  $\alpha$  satisfy (91) and (93), respectively. Since the covariance matrix of the stationary noise process is Toeplitz (see e.g. [43]), the density of eigenvalues on the real line tends to the power spectral density [44]. Given that the power spectral density is bounded and integrable, we have that the sequence of eigenvalues  $\sigma_1^2, \sigma_2^2, \dots$  is summable [43, Theorem 4.2], and thus, bounded as well. Hence, we can remove the assumption that the set of noise variances has finite cardinality, by quantization of the variances. The random code characterization now follows from (255) and (259).

### C. Deterministic Code Capacity

Moving to the deterministic code capacity, observe that for a constant-parameter Gaussian AVC, where the noise sequence is i.i.d.  $\sim \mathcal{N}(0, \sigma^2)$ , we have that  $\tilde{\Lambda}(F_X, \sigma) = \mathbb{E}X^2$ , by Lemma 14, taking  $d = 1$ . Therefore, for the Gaussian AVC with a parameter sequence  $\sigma_1^2, \dots, \sigma_n^2$ ,

$$L_n^* = \max_{F_{X|T} : \frac{1}{n} \sum_{i=1}^n \mathbb{E}[X^2|T=\sigma_i] \leq \Omega} \frac{1}{n} \sum_{i=1}^n \tilde{\Lambda}(F_{X|T=\sigma_i}, \sigma_i) = \max_{F_{X|T} : \frac{1}{n} \sum_{i=1}^n \mathbb{E}[X^2|T=\sigma_i] \leq \Omega} \frac{1}{n} \sum_{i=1}^n \mathbb{E}[X_i^2|T = \sigma_i] = \Omega, \quad (260)$$

where the first equality holds by the definition of  $L_n^*$  in (28) and by (42). It can further be seen from the proof of Lemma 14 in Appendix L that the Gaussian channel  $Y = X + S + Z_\sigma$  is symmetrized by a distribution  $\varphi(s|x)$  that gives probability 1 to  $S = x$ , and that the minimum in the formula of  $\tilde{\Lambda}(F_X, \sigma)$  in (41) is attained with this distribution.

Therefore, by Corollary 11, the capacity of the AVC with colored Gaussian noise is given by the limit inferior of

$$R_n(\mathcal{W}) = \begin{cases} \min_{\substack{N_1, \dots, N_n : \\ \frac{1}{n} \sum_{i=1}^n N_i \leq \Lambda}} \max_{\substack{P_1, \dots, P_n, \tilde{\lambda}_1, \dots, \tilde{\lambda}_n : \\ \frac{1}{n} \sum_{i=1}^n P_i \leq \Omega, \frac{1}{n} \sum_{i=1}^n \tilde{\lambda}_i \geq \Lambda}} \frac{1}{n} \sum_{i=1}^n C_{\sigma_i}(P_i, \tilde{\lambda}_i, N_i) & \text{if } L_n^* > \Lambda, \\ 0 & \text{if } L_n^* \leq \Lambda \end{cases} \quad (261)$$

where

$$C_\sigma(P, \Delta, N) = \min_{F_{S''} : \mathbb{E}S''^2 \leq N} \max_{\substack{F_{X''} : \mathbb{E}X''^2 \leq P, \\ \tilde{\Lambda}_\sigma(F_{X''}, \sigma) \geq \Delta}} I_q(X''; Y''|T'' = \sigma). \quad (262)$$

Consider the direct part. Suppose that  $\Omega > \Lambda$ , hence  $L_n^* > \Lambda$  (see (260)), and set  $P_i = \tilde{\lambda}_i = a_i^*$  for  $i \in [1 : n]$ . This choice of parameters satisfies the optimization constraints in (261), as  $\sum_{i=1}^n P_i = \Omega$ , and also  $\sum_{i=1}^n \tilde{\lambda}_i = \Omega > \Lambda$ . Therefore,

$$\begin{aligned} R_n(\mathcal{W}) &\geq \min_{\substack{N_1, \dots, N_n : \\ \frac{1}{n} \sum_{i=1}^n N_i \leq \Lambda}} \frac{1}{n} \sum_{i=1}^n C_{\sigma_i}(a_i^*, a_i^*, \lambda_i) = \min_{\substack{N_1, \dots, N_n, F_{S''m} : \\ \mathbb{E}S''^2 \leq N_i, \frac{1}{n} \sum_{i=1}^n N_i \leq \Lambda}} \frac{1}{n} \sum_{i=1}^n I_q(X_i''; Y_i''|T_i'' = \sigma_i), \\ &\geq \min_{N_1, \dots, N_n : \sum_{i=1}^n N_i \leq n\Lambda} \frac{1}{n} \sum_{i=1}^n \frac{1}{2} \log \left( 1 + \frac{a_i^*}{N_i + \sigma_i^2} \right) \end{aligned} \quad (263)$$

where the last inequality holds since Gaussian noise is known to be the worst additive noise under variance constraint [34, Lemma II.2]. Next, observe that this is the same minimization as in (232), in the proof of the direct part for the AVGPC, with

$d \leftarrow n$ ,  $\Omega \leftarrow (n\Omega)$ ,  $\Lambda \leftarrow (n\Lambda)$  (see proof of Theorem 15 in Appendix M). Therefore, the minimum is attained with  $N_i = b_i^*$ , and the RHS of (255) is achievable with deterministic codes as well, provided that  $\Omega > \Lambda$ .

The converse part is straightforward. Since the deterministic code capacity is always bounded by the random code capacity, we have that  $\mathbb{C}(\Psi_Z) \leq \mathbb{C}^*(\Psi_Z) = \mathbb{C}^*(\Psi_Z)$ . If  $\Omega \leq \Lambda$ , then  $L_n^* \leq \Lambda$  by (260), hence  $\mathbb{C}(K_Z) = \liminf R_n(\mathcal{W}) = 0$  by the second part of Corollary 11.  $\square$

#### REFERENCES

- [1] A. Abdul Salam, R. Sherif, S. Al-Araji, K. Mezher, and Q. Nasir. Novel approach for modeling wireless fading channels using a finite state markov chain. *ETRI J.*, 39(5):718–728, October 2017.
- [2] A. Ahlswede, I. Althöfer, C. Deppe, and U. Tamm. *Probabilistic methods and distributed information*. Springer, 2019.
- [3] R. Ahlswede. The weak capacity of averaged channels. *J. Prob. Theory and Related Areas*, 11(1):61–73, 1968.
- [4] R. Ahlswede. The capacity of a channel with arbitrarily varying additive gaussian channel probability functions. In *Trans. 6th Prague Conf. Inform. Theory, Statist. Decision Func., Random Processes*, Prague, Czech Republic, Sep 1971.
- [5] R. Ahlswede. Elimination of correlation in random codes for arbitrarily varying channels. *Z. Wahrscheinlichkeitstheorie Verw. Gebiete*, 44(2):159–175, Jun 1978.
- [6] R. Ahlswede. Arbitrarily varying channels with states sequence known to the sender. *IEEE Trans. Inform. Theory*, 32(5):621–629, Sep 1986.
- [7] R. Ahlswede and N. Cai. *Arbitrarily varying multiple-access channels*. Universität Bielefeld., 1996.
- [8] R. Ahlswede and N. Cai. Arbitrarily varying multiple-access channels. i. ericson’s symmetrizability is adequate, gubner’s conjecture is true. *IEEE Trans. Inform. Theory*, 45(2):742–749, Mar 1999. ISSN 0018-9448.
- [9] H. Aydinian, F. Cicalese, and C. Deppe. *Information Theory, Combinatorics, and Search Theory*. Springer, 2013.
- [10] S. Barbarossa and A. Scaglione. On the capacity of linear time-varying channels. In *Proc. IEEE Int’l Conf. Acoust., Speech, Signal Process (ICASSP’1999)*, volume 5, pages 2627–2630, Phoenix, AZ, USA, March 1999.
- [11] E. Biglieri, J. Proakis, and S. Shamai. Fading channels: information-theoretic and communications aspects. *IEEE Trans. Inform. Theory*, 44(6):2619–2692, Oct 1998.
- [12] E. Biglieri, R. Calderbank, A. Constantinides, A. Goldsmith, A. Paulraj, and H. V. Poor. *MIMO wireless communications*. Cambridge university press, 2007.
- [13] I. Bjelaković, H. Boche, and J. Sommerfeld. Capacity results for arbitrarily varying wiretap channels. In *Information Theory, Combinatorics, and Search Theory*, pages 123–144. Springer, 2013.
- [14] D. Blackwell, L. Breiman, and A. J. Thomasian. Proof of shannon’s transmission theorem for finite-state indecomposable channels. *Ann. Math. Stat.*, 29(4):1209–1220, 1958.
- [15] D. Blackwell, L. Breiman, and A. J. Thomasian. The capacity of a class of channels. *Ann. Math. Statist.*, 30(4):1229–1241, Dec 1959.
- [16] D. Blackwell, L. Breiman, and A. J. Thomasian. The capacities of certain channel classes under random coding. *Ann. Math. Statist.*, 31(3):558–567, Sep 1960.
- [17] H. Boche and R. F. Schaefer. Capacity results and super-activation for wiretap channels with active wiretappers. *IEEE Trans. Inform. Theory*, 8(9):1482–1496, Aug 2013.
- [18] H. Boche, R. F. Schaefer, and H. V. Poor. On arbitrarily varying wiretap channels for different classes of secrecy measures. In *Proc. IEEE Int’l Symp. Inform. Theory (ISIT’2014)*, pages 2376–2380, Honolulu, Hawaii, Jun 2014.
- [19] H. Boche, R. F. Schaefer, and H. V. Poor. On the continuity of the secrecy capacity of compound and arbitrarily varying wiretap channels. 10(12):2531–2546, Dec 2015.
- [20] S. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge university press, 2004.
- [21] A. J. Budkuley and S. Jaggi. Communication over an arbitrarily varying channel under a state-myopic encoder. [arXiv:1804.10221](https://arxiv.org/pdf/1804.10221), Apr 2018. URL <https://arxiv.org/pdf/1804.10221.pdf>.
- [22] A. J. Budkuley and S. Jaggi. Communication over an arbitrarily varying channel under a state-myopic encoder. In *Proc. IEEE Int’l Symp. Inform. Theory (ISIT’2018)*, Vail, Colorado, Jun 2018.
- [23] A. J. Budkuley, B. K. Dey, and V. M. Prabhakaran. Dirty paper arbitrarily varying channel with a state-aware adversary. In *Proc. IEEE Inform. Theory Workshop(ITW’2015)*, pages 94–98, Jeju, South Korea, Oct 2015.
- [24] A. J. Budkuley, B. K. Dey, and V. M. Prabhakaran. Communication in the presence of a state-aware adversary. *IEEE Trans. Inform. Theory*, 63(11):7396–7419, Nov 2017.
- [25] G. Caire and S. Shamai. On the capacity of some channels with channel state information. *IEEE Trans. Inform. Theory*, 45(6):2007–2019, Sep 1999.
- [26] R. S. Cheng and S. Verdú. Gaussian multiaccess channels with isi: capacity region and multiuser water-filling. *IEEE Trans. Inform. Theory*, 39(3):773–785, May 1993.
- [27] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley, 2 edition, 2006.
- [28] I. Csiszár. Arbitrarily varying channels with general alphabets and states. *IEEE Trans. Inform. Theory*, 38(6):1725–1742, Nov 1992.

- [29] I. Csiszár and J. Körner. *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Cambridge University Press, 2 edition, 2011.
- [30] I. Csiszár and P. Narayan. The capacity of the arbitrarily varying channel revisited: positivity, constraints. *IEEE Trans. Inform. Theory*, 34(2):181–193, Mar 1988.
- [31] I. Csiszár and P. Narayan. Arbitrarily varying channels with constrained inputs and states. *IEEE Trans. Inform. Theory*, 34(1):27–34, Jan 1988.
- [32] I. Csiszár and P. Narayan. Capacity of the gaussian arbitrarily varying channel. *IEEE Transactions on Information Theory*, 37(1):18–26, Jan 1991.
- [33] A. Das and P. Narayan. Capacities of time-varying multiple-access channels with side information. *IEEE Trans. Inform. Theory*, 48(1):4–25, Jan 2002.
- [34] S. N. Diggavi and T. M. Cover. The worst additive noise under a covariance constraint. *IEEE Trans. Inform. Theory*, 47(7):3072–3081, Nov 2001.
- [35] P. M. Ebert. Error bounds for parallel communication channels. 1966.
- [36] A. El Gamal and Y. Kim. *Network Information Theory*. Cambridge University Press, 2011.
- [37] T. Ericson. Exponential error bounds for random codes in the arbitrarily varying channel. *IEEE Trans. Inform. Theory*, 31(1):42–48, Jan 1985.
- [38] G. J. Foschini. Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas. *Bell Labs Tech. J*, 1(2):41–59, 1996.
- [39] R. G. Gallager. *Information theory and reliable communication*, volume 2. Springer, 1968.
- [40] Z. Goldfeld, P. Cuff, and H. H. Permuter. Arbitrarily varying wiretap channels with type constrained states. *IEEE Trans. Inform. Theory*, 62(12):7216–7244, Dec 2016.
- [41] A. Goldsmith, S. A. Jafar, N. Jindal, and S. Vishwanath. Capacity limits of mimo channels. *IEEE J. Selected Areas Comm.*, 21(5):684–702, 2003.
- [42] A. J. Goldsmith and P. P. Varaiya. Capacity of fading channels with channel side information. *IEEE Trans. Inform. Theory*, 43(6):1986–1992, Nov 1997.
- [43] R. M. Gray. Toeplitz and circulant matrices: A review. *Foundations and Trends<sup>®</sup> in Communications and Information Theory*, 2(3):155–239, 2006.
- [44] U. Grenander and G. Szegö. *Toeplitz forms and their applications*, volume 321. Univ. California Press, 2001.
- [45] O. Gungor, C. E. Koksal, and H. E. Gamal. An information theoretic approach to rf fingerprinting. In *(ACSSC'2013)*, pages 61–65, Nov 2013.
- [46] G. Han. A randomized algorithm for the capacity of finite-state channels. *IEEE Trans. Inform. Theory*, 61(7):3651–3669, July 2015.
- [47] T. S. Han. *Information-spectrum methods in information theory*, volume 50. Springer Science & Business Media, 2013.
- [48] D. He and Y. Luo. Arbitrarily varying wiretap channel with state sequence known or unknown at the receiver. [arXiv:1701.02043](https://arxiv.org/abs/1701.02043), Dec 2017.
- [49] X. He and A. Yener. Gaussian two-way wiretap channel with an arbitrarily varying eavesdropper. In *Proc. Global Commun. Conf. (GLOBECOM'2011)*, pages 854–858, Houston, TX, USA, Dec 2011.
- [50] X. He, A. Khisti, and A. Yener. MIMO multiple access channel with an arbitrarily varying eavesdropper: Secrecy degrees of freedom. *IEEE Trans. Inform. Theory*, 59(8):4733–4745, Aug 2013.
- [51] C. Heegard and A. E. Gamal. On the capacity of computer memory with defects. *IEEE Trans. Inform. Theory*, 29(5):731–739, Sep 1983.
- [52] E. Hof and S. I. Bross. On the deterministic-code capacity of the two-user discrete memoryless arbitrarily varying general broadcast channel with degraded message sets. *IEEE Trans. Inform. Theory*, 52(11):5023–5044, Nov 2006.
- [53] J. L. Holsinger. Digital communication over fixed time-continuous channels with memory-with special application to telephone channels. 1964.
- [54] F. Hosseiniogoki and O. Kosut. The gaussian interference channel in the presence of a malicious jammer. In *Proc. Allerton Conf. Commun., Control, Computing*, pages 679–686, Monticello, IL, USA, Sep. 2016.
- [55] F. Hosseiniogoki and O. Kosut. The gaussian interference channel in the presence of malicious jammers. [arXiv:1712.04133](https://arxiv.org/abs/1712.04133), December 2017. URL <https://arxiv.org/pdf/1712.04133.pdf>.
- [56] F. Hosseiniogoki and O. Kosut. Capacity of the gaussian arbitrarily-varying channel with list decoding. In *Proc. IEEE Int'l Symp. Inform. Theory (ISIT'2009)*, pages 471–475, Vail, CO, USA, June 2018.
- [57] F. Hosseiniogoki and O. Kosut. Capacity of gaussian arbitrarily-varying fading channels. In *Proc. Ann. Conf. Inform. Sciences Syst. (CISS'2019)*, pages 1–6, March 2019.
- [58] F. Hosseiniogoki and O. Kosut. Packing lemmas for gaussian jamming networks. In *Talk given in Inform. Theory Appl. Workshop (ITA'2019)*, San Diego, California, February 2019.
- [59] F. Hosseiniogoki and O. Kosut. List-decoding capacity of the gaussian arbitrarily-varying channel. *Entropy*, 21(6):575, 2019.
- [60] B. Hughes and P. Narayan. Gaussian arbitrarily varying channels. *IEEE Trans. Inform. Theory*, 33(2):267–284, March

1987.

- [61] B. Hughes and P. Narayan. The capacity of a vector gaussian arbitrarily varying channel. *IEEE Trans. Inform. Theory*, 34(5):995–1003, Sep. 1988.
- [62] T. Ignatenko and F. M. J. Willems. Biometric security from an information-theoretical perspective. *Foundations and Trends® in Communications and Information Theory*, 7(2–3):135–316, 2012.
- [63] J. H. Jahn. Coding of arbitrarily varying multiuser channels. *IEEE Trans. Inform. Theory*, 27(2):212–226, Mar 1981.
- [64] C. R. Janda, M. Wiese, J. Nötzel, H. Boche, and E. A. Jorswieck. Wiretap-channels under constrained active and passive attacks. In *(CNS'2015)*, pages 16–21, Jun 2015.
- [65] T. Keresztfalvi and A. Lapidoth. Semi-robust communications over a broadcast channel. *IEEE Trans. Inform. Theory*, 65(8):5043–5049, Aug 2019.
- [66] Y. Kim and B. V. K. V. Kumar. Writing on dirty flash memory. In *Proc. Allerton Conf. Commun., Control, Computing*, pages 513–520, Monticello, IL, USA, Sept 2014.
- [67] G. Kramer. Topics in multi-user information theory. *Foundations and Trends in Communications and Information Theory*, 4(4–5):265–444, 2008.
- [68] A. V. Kuznetsov and B. S. Tsybakov. Coding in a memory with defective cells. *Problemy peredachi informatsii*, 10(2): 52–60, 1974.
- [69] A. V. Kuzntsov and A. J. H. Vinck. On the general defective channel with informed encoder and capacities of some constrained memories. *IEEE Trans. Inform. Theory*, 40(6):1866–1871, Nov 1994.
- [70] R. J. La and V. Anantharam. A game-theoretic look at the gaussian multiaccess channel. *DIMACS Series Discrete Math. Theor. Comp. Science*, 66:87–106, 2004.
- [71] L. Lai and H. E. Gamal. The water-filling game in fading multiple-access channels. *IEEE Trans. Inform. Theory*, 54(5):2110–2122, May 2008.
- [72] R. Langner. Stuxnet: Dissecting a cyberwarfare weapon. *IEEE Security Privacy*, 9(3):49–51, May 2011.
- [73] A. Lapidoth and P. Narayan. Reliable communication under channel uncertainty. *IEEE Trans. Inform. Theory*, 44(6): 2148–2177, Oct 1998.
- [74] A. Lapidoth and I. E. Telatar. The compound channel capacity of a class of finite-state channels. *IEEE Trans. Inform. Theory*, 44(3):973–983, May 1998.
- [75] K. Leyton-Brown and Y. Shoham. *Essentials of game theory: A concise multidisciplinary introduction*. Morgan & Claypool, 2008.
- [76] M. Martone. Blind adaptive detection of ds/cdma signals on time-varying multipath channels with antenna arrays using high-order statistics. *IEEE Trans. Comm.*, 48(9):1590–1600, Sep. 2000.
- [77] E. MolavianJazi, M. Bloch, and J. N. Laneman. Arbitrary jamming can preclude secure communication. In *Proc. Allerton Conf. Commun., Control, Computing*, pages 1069–1075, Monticello, IL, USA, Sep 2009.
- [78] J. Nötzel, M. Wiese, and H. Boche. The arbitrarily varying wiretap channel — secret randomness, stability, and super-activation. *IEEE Trans. Inform. Theory*, 62(6):3504–3531, Jun 2016.
- [79] G. Owen. *Game Theory*. Emerald Group Publishing, 4 edition, 2013.
- [80] L. H. Ozarow, S. Shamai, and A. D. Wyner. Information theoretic considerations for cellular mobile radio. *IEEE Trans. Vehicular Tech.*, 43(2):359–378, 1994.
- [81] U. Pereg and Y. Steinberg. The arbitrarily varying gaussian relay channel with sender frequency division. In *Proc. Allerton Conf. Commun., Control, Computing*, pages 1097–1103, Monticello, IL, USA, Oct 2018.
- [82] U. Pereg and Y. Steinberg. The arbitrarily varying channel under constraints with side information at the encoder. *IEEE Trans. Inform. Theory*, 65(2):861–887, Feb 2019.
- [83] U. Pereg and Y. Steinberg. The arbitrarily varying relay channel. *20th Anniversary of Entropy - Recent Advances in Entropy and Information-Theoretic Concepts and Their Applications*, 21(5):516, 2019.
- [84] U. Pereg and Y. Steinberg. The arbitrarily varying broadcast channel with causal side information at the encoder. *accepted to IEEE Trans. Inform. Theory*, 2019. doi: 10.1109/TIT.2019.2927696.
- [85] U. Pereg and Y. Steinberg. The capacity region of the arbitrarily varying mac: with and without constraints. *submitted to IEEE Trans. Inform. Theory*, 2019.
- [86] W. Rudin. *Principles of Mathematical Analysis*. McGraw-Hill, 3 edition, 1976.
- [87] H. Saggarr, G. Pottie, and B. Daneshrad. On maximizing the average capacity with interference alignment in a time varying channel. In *2016 Information Theory and Applications Workshop (ITA)*, pages 1–5, Jan 2016.
- [88] A. D. Sarwate and M. Gastpar. Randomization bounds on gaussian arbitrarily varying channels. In *Proc. IEEE Int’l Symp. Inform. Theory (ISIT’2006)*, pages 2161–2165, Seattle, WA, USA, July 2006.
- [89] A. D. Sarwate and M. Gastpar. Arbitrarily dirty paper coding and applications. In *Proc. IEEE Int’l Symp. Inform. Theory (ISIT’2008)*, pages 925–929, Toronto, ON, Canada, July 2008.
- [90] A. D. Sarwate and M. Gastpar. Relaxing the gaussian avc. [arXiv:1209.2755](https://arxiv.org/abs/1209.2755), 2012.
- [91] R. F. Schaefer, H. Boche, and H. V. Poor. Super-activation as a unique feature of arbitrarily varying wiretap channels. In *2016 IEEE International Symposium on Information Theory (ISIT)*, pages 3077–3081, July 2016.

- [92] S. Shamai and A. Steiner. A broadcast approach for a single-user slowly fading mimo channel. *IEEE Trans. Inform. Theory*, 49(10):2617–2635, Oct 2003.
- [93] S. Shamai, L. H. Ozarow, and A. D. Wyner. Information rates for a discrete-time gaussian channel with intersymbol interference and stationary inputs. *IEEE Trans. Inform. Theory*, 37(6):1527–1539, Nov 1991.
- [94] C. E. Shannon. Communication in the presence of noise. *Proc. IRE*, 37(1):10–21, Jan 1949.
- [95] M. K. Simon and M. S. Alouini. *Digital communication over fading channels*, volume 95. John Wiley & Sons, 2005.
- [96] M. Sion. On general minimax theorems. *Pacific J. Math*, 8(1):171–176, Mar 1958.
- [97] J. Slay and M. Miller. Lessons learned from the maroochy water breach. In E. Goetz and S. Sheno, editors, *Critical Infrastructure Protection*, pages 73–82, Boston, MA, 2008. Springer US.
- [98] R. M. Sundaram, A. K. Das, D. Jalihal, and V. Ramaiyan. Optimal frame synchronization over a finite state markov channel. In *Proc. IEEE Int’l Symp. Inform. Theory (ISIT’2017)*, pages 486–490, Aachen, Germany, June 2017.
- [99] E. Telatar. Capacity of multi-antenna gaussian channels. *Euro. Trans. Telecomm.*, 10(6):585–595, 1999.
- [100] P. J. Thomas and A. W. Eckford. Capacity of a simple intercellular signal transduction channel. *IEEE Trans. Inform. Theory*, 62(12):7358–7382, Dec 2016.
- [101] T. G. Thomas and B. Hughes. Exponential error bounds for random codes on gaussian arbitrarily varying channels. *IEEE Trans. Inform. Theory*, 37(3):643–649, May 1991.
- [102] S. Verdú and T. S. Han. A general formula for channel capacity. *IEEE Trans. Inform. Theory*, 40(4):1147–1157, July 1994.
- [103] J. von Neumann and O. Morgenstern. *Theory of Games and Economic Behavior*. Princeton university press, 1944.
- [104] A. S. Vora and A. A. Kulkarni. Minimax theorems for finite blocklength lossy joint source-channel coding over an avc. [arXiv:1907.05324](https://arxiv.org/abs/1907.05324), 2019.
- [105] Z. Wang, V. Aggarwal, and X. Wang. Iterative dynamic water-filling for fading multiple-access channels with energy harvesting. *IEEE Trans. Inform. Theory*, 33(3):382–395, March 2015.
- [106] M. Wiese and H. Boche. The arbitrarily varying multiple-access channel with conferencing encoders. *IEEE Trans. Inform. Theory*, 59(3):1405–1416, March 2013.
- [107] A. Winshtok. *Source and Channel Coding Problems in the Presence of Arbitrarily Varying Side Information*. M.sc. thesis, Technion - Israel Institute of Technology, Haifa, Mar 2007.
- [108] A. Winshtok and Y. Steinberg. The arbitrarily varying degraded broadcast channel with states known at the encoder. In *Proc. IEEE Int’l Symp. Inform. Theory (ISIT’2006)*, pages 2156–2160, Seattle, Washington, Jul 2006.
- [109] A. Winshtok and Y. Steinberg. Joint source-channel coding for arbitrarily varying wyner-ziv source and gel’fand-pinsker channel. In *Proc. Allerton Conf. Commun., Control, Computing*, pages 1064–1070, Monticello, IL, USA, Sep 2006.
- [110] J. Wolfowitz. *Coding theorems of information theory*, volume 31. Springer-Verlag Berlin Heidelberg, 3 edition, 2012.
- [111] C. Y. Wong, R. S. Cheng, K. B. Lataief, and R. D. Murch. Multiuser ofdm with adaptive subcarrier, bit, and power allocation. *IEEE Trans. Inform. Theory*, 17(10):1747–1758, Oct 1999.
- [112] X. Wang and M. T. Orchard. On reducing the rate of retransmission in time-varying channels. *IEEE Trans. Comm.*, 51(6):900–910, June 2003.
- [113] W. Yu and J. M. Cioffi. Fdma capacity of gaussian multiple-access channels with isi. *IEEE Trans. Inform. Theory*, 50(1):102–111, Jan 2002.
- [114] W. Yu, W. Rhee, S. Boyd, and J. M. Cioffi. Iterative water-filling for gaussian vector multiple-access channels. *IEEE Trans. Inform. Theory*, 50(1):145–152, Jan 2004.