The Arbitrarily Varying Channel With Colored Gaussian Noise

Uzi Pereg®, Member, IEEE, and Yossef Steinberg®, Fellow, IEEE

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, a model that combines 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. 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 [100], introducing the water filling optimal power allocation. This channel is the spectral counterpart of the Gaussian product channel (see e.g. [29, 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. [11], [12], [41], [45], [99], [106]). Furthermore, for some networks, water filling is performed in multiple stages [28], [75], [113], [119], [121], [122]. A limit formula for the capacity of the general time-varying channel (TVC) is given in [110] (see also [3], [10], [31], [51], [80], [93], [120]). Another relevant setting is that of a finite-state channel, where the state evolves as a Markov chain [14], [50], [77], [78], [105], [107], [118]. In practice, there is often uncertainty regarding channel statistics, due to a variety of causes such as fading in wireless communication [1], [27], [46], [61], [63], [86], [98], [101], memory faults in storage [55], [70], [72], [73], malicious attacks on identification systems [49], [66], cyber-physical warfare [76], [103], [112]. The arbitrarily varying channel (AVC) is an appropriate model to describe such a situation [16], [77].

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 [40] and Csiszár and Narayan [32] have established a simple single-letter condition, namely nonsymmetrizability, which is both necessary and sufficient for the capacity to be positive. Schaefer et al. [97] 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 [32], [33] 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 [32], constraints have a significant effect on the behavior of the capacity. Specifically, it is shown in [32] 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 $Y = X + S + Z$, where $X$ and $Y$ are the input and output sequences, respectively; $S$ is a state sequence of unknown joint distribution $F_S$, not necessarily independent nor stationary; and the noise sequence $Z$ 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 $S$ 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 [64], Hughes and Narayan showed that the random code capacity is given by \( C^*_1 = \frac{1}{n} \log (1 + \frac{\Omega}{\Lambda}) \). Subsequently, Csiszár and Narayan [34] showed that the deterministic code capacity is given by

\[
C_1 = \begin{cases} 
C^*_1 & \text{if } \Lambda < \Omega, \\
0 & \text{if } \Lambda \geq \Omega. 
\end{cases}
\]  

As noted in [34], this result is not a straightforward consequence of the elegant Elimination Technique [5], by Ahlswede to establish dichotomy for the AVC without constraints. Hosseinigoki and Kosut [61] determined the capacity in multiple side information scenarios for the Gaussian AVC with fast fading. Hughes and Narayan [65] 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 [65]. The Gaussian AVC is also considered in [4], [60], [63], [74], [94], [96], [108]. An analogous behavior was recently demonstrated for blockchain attacks [36], [109].

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 [2], [9], [17]–[19], [52], [82], [84], including input and state constraints in [13], [43], [68]. The capacity region of the arbitrarily varying multiple access channel (MAC) with and without constraints is characterized in [7], [8], [67], [91]; capacity bounds for the arbitrarily varying broadcast channel are derived in [56], [67]; and for the arbitrarily varying relay channel in [87], [89]. Additional results on arbitrarily varying multi-user channels and constraints are derived e.g. in [26], [54], [69], [90], [114], [116]. Transmission of an arbitrarily varying Wyner-Ziv source over a Gel’fand-Pinsker channel is considered in [115], [117], and related problems were recently presented in [23], [24], [26]. Various Gaussian AVC networks are studied e.g. in [25], [53], [58], [59], [62], [88], [89], [91], [95].

In this paper, we address the AVC with colored Gaussian noise. The practical implications of our results are not just for the evaluation of the optimal throughput, but also for developing optimal communication strategies for the transmitter and the jammer. For example, for an OFDM system, the water filling solution leads to an optimal usage of bandwidth [44]. Based on our results, the optimal transmission and jamming strategies are obtained as a double water filling solution as long as the user has strictly larger transmission power than the jammer, i.e. when \( \Omega > \Lambda \). Otherwise, the optimal jamming strategy is to simulate the legitimate transmission and use the same spectrum allocation as the transmitter does.

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 [65], 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 [28], [75] (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 [100], 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. As demonstrated in Figure 1, the power spectral density of the noise process is pictured as the bottom of a vessel. First, the jammer pours water of volume \( \Lambda \) into the vessel, and then the encoder pours more water of volume \( \Omega \). The deterministic code capacity is given by the double water filling solution only if the user has strictly larger transmission power than the jammer, i.e. \( \Omega > \Lambda \), and it is zero otherwise.

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 code-word 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 Fig. 1. Water filling in the frequency domain for the AVC with colored Gaussian noise and fixed fading is given in Subsection III-D. The set of all probability kernels \( P_{X^n} \) is denoted by \( \mathcal{P} = \{ P^n \} \). Similarly, define the joint type \( P_{x^n,y^n}(a,b) \) as the number of occurrences of the symbol pair \((a,b)\) in the sequence \( (x^n, y^n) \). A conditional type is defined as \( P_{x^n|y^n}(a,b) = P_{x^n,y^n}(a,b)/P_{y^n}(b) \), and a conditional type class as \( \mathcal{T}^n(P|y^n) = \{ x^n : P_{x^n|y^n} = P \} \). Furthermore, we define the \( \delta \)-typical set \( \mathcal{A}^n_\delta(p) \) with respect to a distribution \( p(x) \) by

\[
\mathcal{A}^n_\delta(p) = \left\{ x^n \in \mathcal{X}^n : \forall a \in \mathcal{X}, \quad \left| p(a) - P_{x^n}(a) \right| \leq \delta \text{ if } p(a) > 0, \text{ and } \right.
\]

\[
\left. P_{x^n}(a) = 0 \text{ if } p(a) = 0 \right\}.
\]

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 \( Z = (z_1, z_2, \ldots, z_n) \) is used when it is understood from the context that the length of the sequence is \( n \), and the \( \ell^2 \)-norm of \( z \) is denoted by \( \| z \| \).

### B. Channel Description

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

\[
W_{Y^n|X^n,S^n,T^n}(y^n|x^n, s^n, t^n) = \prod_{i=1}^{n} W_{Y_i|X_i,S_i,T_i}(y_i|x_i, s_i, t_i).
\]

The AVC with fixed parameters is a DMC \( W_{Y^n|X^n,S^n,T^n} \) where the state sequence has an unknown distribution, not necessarily independent nor stationary, while the parameter sequence is fixed. That is, the state sequence \( S^n \sim q(s^n|\theta^n) \) with an unknown joint pmf \( q(s^n|\theta^n) \) over \( S^n \). In particular, \( q(s^n|\theta^n) \) could give mass 1 to some state sequence \( s^n \). Whereas, the parameter sequence is fixed as

\[
T^n \equiv \theta^n,
\]
where \( \theta_1, \theta_2, \ldots \) is a given sequence of letters from \( \mathcal{T} \), known to the encoder, decoder, and jammer. We stress that as opposed to the state sequence \( S^n \), the parameter sequence \( T^n \equiv \theta^n \) is not arbitrary nor random; \( \theta^n \) is a specific (deterministic) sequence that is set a priori, and this specific sequence is an integral part of the channel model. The AVC with fixed parameters is denoted by \( \mathcal{W} = \{ W_{Y|X,S,T}, \theta^n \} \), where \( \theta^n \) is a short notation for the sequence \( \{ \theta_i \}_{i=1}^n \).

The compound channel with fixed parameters is used as a tool in the analysis. Different models of compound channels are described in the literature [31]. 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 \( Q_n \), with \( Q_n \subseteq \mathcal{P}(S|T) \). That is,

\[
S^n \sim \prod_{i=1}^{n} q(s_i|\theta_i) \tag{5}
\]

with an unknown conditional pmf \( q(s|t) \in Q_n \). Note that this differs from the classical definition of the compound channel, as in [31], where the state is does not vary throughout the transmission.

**Remark 1:** Note that the special case of a channel \( W_{Y|X,S,T=\ell} \), with a constant parameter sequence \( \theta_i = t \) for \( i = 1, 2, \ldots \), reduces to the standard state-dependent DMC. Thereby, the AVC \( \mathcal{W}_k = \{ W_{Y|X,S,T=\ell} \} \) with a constant parameter value 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 time-varying channel (TVC) [110].

**Remark 2:** It may be tempting to mistaken the channel for an AVC with partial side information at the encoder and the decoder. However, this scenario is very different from ours, because in our setting \( T^n \equiv \theta^n \) where \( \theta_1, \theta_2, \ldots \) is a specific sequence. Therefore, the channel cannot be regarded as an AVC with a composite state sequence \( S^n = (S_t, T^n)_{i=1}^n \). We will come back to this in the discussion section (see Subsection VI-A).

**Remark 3:** 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. We will see that orthogonalization of the stationary noise process will transform the AVC with colored Gaussian noise into an AVC with fixed parameters, which are determined by the spectral representation of the 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 \( f : [1 : 2^nR] \times T^n \rightarrow \mathcal{X}^n \), and a decoding function \( g : Y^n \times T^n \rightarrow [1 : 2^nR] \).

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 \( \{ E_{\gamma} = (F_{\gamma}, \theta^n) \} \), along with a probability distribution \( \mu(\gamma) \) over the code collection \( \Gamma \). We denote such a code by \( \Phi^\Gamma = (\mu, \Gamma, \{ E_{\gamma} \} \gamma \in \Gamma) \).

### 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 \( \phi : \mathcal{X} \rightarrow [0, \infty) \) and \( l : S \rightarrow [0, \infty) \) be some given bounded functions, and define

\[
\phi^n(x^n) = \frac{1}{n} \sum_{i=1}^{n} \phi(x_i), \quad l^n(s^n) = \frac{1}{n} \sum_{i=1}^{n} l(s_i). \tag{6}
\]

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 \( b \in S \), \( \phi(a) = l(b) = 0 \).

As the parameter sequence \( \theta^n \equiv (\theta_i)_{i=1}^\infty \) 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

\[
\phi^n(f(m, \theta^n)) \leq \Omega, \quad \text{for all } m \in [1 : 2^nR]. \tag{7}
\]

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 \( \mathbb{E} l^n(S^n) \leq \Lambda \), 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 \( Q_n \subseteq \mathcal{P}(S|\theta^n) \), where

\[
\mathcal{P}(S|\theta^n) \triangleq \left\{ q(s|t) : \frac{1}{n} \sum_{i=1}^{n} \sum_{s \in S} q(s|\theta_i)l(s) \leq \Lambda \right\}. \tag{8}
\]

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

\[
\mathcal{P}(S^n|\theta^n) \triangleq \left\{ q(s^n|\theta^n) \in \mathcal{P}(S^n|T^n) : q(s^n|\theta^n) = 0 \text{ if } l^n(s^n) > \Lambda \right\}. \tag{9}
\]

This includes the case of a deterministic unknown state sequence, i.e. when \( q \) gives probability 1 to a particular \( s^n \in S^n \) 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 (7) on the codebook.

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

$$P_e^n(\mathcal{C}|s^n, \theta^n) \triangleq \frac{1}{2^nR} \sum_{m=1}^{2^nR} \sum_{y^n: g(y^n, s^n) \neq m} W_{Y^n|X^n, s^n, T^n}(y^n|f(m, \theta^n), s^n, \theta^n). \quad (10a)$$

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

$$P_e^n(q, \theta^n, \mathcal{C}) \triangleq \sum_{s^n \in S^n} q(s^n|\theta^n)P_e^n(\mathcal{C}|s^n, \theta^n). \quad (10b)$$

**Definition 3 (Achievable Rate and Capacity Under Constraints):** A code $\mathcal{C}$ is a $(2^nR, n, \varepsilon)$ code for the AVC $\mathcal{W}$ with fixed parameters under input constraint $\Omega$ and state constraint $\Lambda$, if it satisfies

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

or, equivalently,

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

for all $q \in \mathcal{P}_\Omega(S^n|\theta^n), \quad (11)$$

and state constraint $\Lambda$.

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 $C(\mathcal{W})$. 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 $\mathcal{C}^T$ satisfies the requirements

$$\sum_{\gamma} \mu(\gamma)\phi^n(f_{\gamma}(m, \theta^n)) \leq \Omega, \quad (12a)$$

and state constraint $\Lambda$. The operational capacity is defined as the supremum of achievable rates, and it is denoted by $C(\mathcal{W})$. 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.

**III. MAIN RESULTS – AVC WITH FIXED PARAMETERS**

In this section, we determine the random code capacity and deterministic code capacity of the AVC with fixed parameters. At this point, the parameter notations $T^n$, $\theta^n$, $T$ and $t$ can be confusing. In order to help the reader follow the results, we give a brief recap. Recall that $W_{Y^n|X^n, S^n, T^n}$ is a state-dependent channel, hence it depends on the state sequence $S^n$ and the parameter sequence $T^n$. Now, $\theta^n$ is a specific sequence in $T^n$ that is set by the channel model, hence $\theta_1, \theta_2, \ldots$ are fixed parameters. We write $T^n \equiv \theta^n$ to express that the parameter sequence of the channel is fixed to be $\theta^n$, and this is why the channel is called an AVC with fixed parameters. In the capacity characterization, we use a dummy random variable $T \sim \hat{P}_T^n$, where $\hat{P}_T^n$ is the empirical distribution of the parameter sequence $\theta^n$. Thus, $T$ has a particular distribution which is also preset in the channel model, and depends on the blocklength $n$ as well. We use $t \in \mathcal{T}$ to denote some parameter value. As mentioned in Remark 1, the case where the fixed parameters are identical, i.e., $\theta_i = t$ for all $i \in [1 : n]$, reduces to the traditional AVC without parameters. We sometimes refer to this special case as a constant-parameter AVC.

A. Random Code Capacity

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

$$C_1^*(\mathcal{W}) \triangleq \min_{q(s|t)} \max_{p(x|t)} I_q(X; Y|T), \quad (13)$$

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

$$\mathbb{E}\phi(X) = \sum_{t \in \mathcal{T}} P_T^n(t) \sum_{x \in \mathcal{X}} p(x|t)\phi(x)$$

$$= \frac{1}{n} \sum_{i=1}^{n} \sum_{x \in \mathcal{X}} p(x|\theta_i)\phi(x) \quad (14)$$

$$\mathbb{E}l(S) = \sum_{t \in \mathcal{T}} P_T^n(t) \sum_{s \in \mathcal{S}} q(s|t)l(s)$$

$$= \frac{1}{n} \sum_{i=1}^{n} \sum_{s \in \mathcal{S}} q(s|\theta_i)l(s) \quad (15)$$

The conditional mutual information in this case is given by

$$I_q(X; Y|T) = \sum_{t,s,x,y} P_T^n(t)q(s|t)p(x|t)W_{Y^n|X^n, S^n, T^n}(y|x, s, t) \times \log \sum_{s', y} q(s'|t)W_{Y^n|X^n, S^n, T^n}(y|x, s', t) \quad (16)$$

Next, we give our random code capacity theorem.

**Theorem 1:** The random code capacity of the AVC $\mathcal{W}$ with fixed parameters, under input constraint $\Omega$ and state constraint $\Lambda$, is given by

$$C_1^*(\mathcal{W}) = \lim_{n \to \infty} C_1^*(\mathcal{W}). \quad (17)$$
The proof of Theorem 1 is given in the appendix in multiple stages:

- In Appendix A, we give the capacity theorem for the compound channel with fixed parameters. This is an auxiliary result, obtained by a simple extension of [31, Exercise 6.8]. The statement of the capacity result is given in Lemma A.1, followed by the proof.
- In Appendix B, we present a variation of the robustification technique (RT), modifying Ahlswede’s original combinatorial lemma [6].
- As can be seen in Lemma B.1, we modify the technique by restricting the permutation set such that the sequence of fixed parameters is an eigenvector of the permutation matrix. This turns out to be useful in the last part of the random code capacity derivation, where the RT is used as a tool in the achievable proof.
- In Appendix C, we give the proof of Theorem 1 using the auxiliary result and tools described 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 value \( \theta_t = t \) reduces to the traditional AVC under input and state constraints. For this channel, Csiszár and Narayan [33] showed that the random code capacity is given by

\[
C_R^*(\Omega, \Lambda) \triangleq \max_\{q(x): E(I(S) \leq \Lambda) \} \min_\{p(x): E(\theta(X) \leq \Omega)\} E(I_q(X; Y|T = t) = t).
\]

(18)

Then, define

\[
R_n^*(W) \triangleq \max_\{\lambda_1, \ldots, \lambda_n: \sum_{i=1}^n \lambda_i \leq \Lambda \} \min_\{\omega_1, \ldots, \omega_n: \sum_{i=1}^n \omega_i \leq \Omega\} \frac{1}{n} \sum_{i=1}^n C_{\theta_i}^*(\omega_i, \lambda_i),
\]

(19)

Lemma 2: Let \( C^*_n(W) \) and \( R_n^*(W) \) be as in (13) and (19), respectively. Then,

\[
R_n^*(W) = C^*_n(W).
\]

(20)

The proof of Lemma 2 is given in Appendix D. Theorem 1 and Lemma 2 yield the following consequence.

Corollary 3: The random code capacity of the AVC \( W \) with fixed parameters, under input constraint \( \Omega \) and state constraint \( \Lambda \), is given by

\[
C^*(W) = \lim_{n \to \infty} R_n^*(W).
\]

(21)

Corollary 3 shows that the random code capacity of the AVC with fixed parameters is given by the average of the capacities of the individual AVCs \( \{W_Y|X,S,T(\cdot, \cdot, \theta_i)\} \) over time.

This is not surprising given the characterization of the TVC [110] [31, Problem 6.8]. On the other hand, we will shortly see that the deterministic code capacity does not admit such property (see Corollary 7 and Example 1 in Subsection III-C below). The corollary will also be useful in our analysis of the AVC with colored Gaussian noise, since in the continuous case the type of the parameter sequence is not well defined.

B. Deterministic Code Capacity

We move to the deterministic code capacity of the AVC with fixed parameters, \( W = \{W_Y|X,S,T, \theta^n\} \), under input constraint \( \Omega \) and state constraint \( \Lambda \). Before we state the capacity theorem, we give a few definitions. We begin with symmetrizability of a channel without parameters.

Definition 4 (see [32]): A state-dependent DMC \( V_Y|X,S \) is said to be symmetrizable if for some conditional distribution \( J(s|x) \),

\[
\sum_{s \in S} V_{Y|X,S}(y|x_1, s)J(s|x_2) = \sum_{s \in S} V_{Y|X,S}(y|x_2, s)J(s|x_1),
\]

\[
\forall x_1, x_2 \in \mathcal{X}, y \in \mathcal{Y}.
\]

(22)

Equivalently, the channel \( \tilde{V}(y|x_1, x_2) = \sum_{s \in S} V_{Y|X,S}(y|x_1, s)J(s|x_2) \) 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 \( V : \mathcal{X} \to \mathcal{S} \) 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 \( J^n(s^n|x^n_2) = \prod_{i=1}^n J(s_i|x_2,i) \), 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” \( \bar{V} \) is symmetric with respect to \( x_1^n \) and \( x_2^n \), the two codewords appear to the receiver as equally likely. Indeed, by [40], 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 \( W_Y|X,S(\cdot, \cdot, \theta_i) \) are all symmetrizable, or the number of non-symmetrizable channels grows linearly with \( n \). That is, for sufficiently large \( n \),

\[
\text{either } |\mathcal{I}(n)| = 0 \quad \text{or} \quad |\mathcal{I}(n)| = \Omega(n),
\]

(23a)

where

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

(23b)

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 4 below. Next, we define a symmetrizability cost and threshold for the AVC with fixed parameters. For every \( n \) and \( p(x|t) \) with

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

(24)
define the minimal symmetrizability cost by

\[
\tilde{\Lambda}_n(p) \triangleq \min_{\frac{1}{n}\sum_{i=1}^{n}\sum_{x,s \in S} p(x|\theta_i)J_{\theta_i}(s|x)l(s)} \frac{1}{n}\sum_{i=1}^{n}\sum_{x,s \in S} p(x|\theta_i)J_{\theta_i}(s|x)l(s),
\]

(25)

where the minimization is over the conditional distributions \(J_{\theta_i}(s|x)\) that symmetrize \(W_{Y|X,S,T}(\cdot|\cdot,t)\), for \(t \in T\) (see Definition 4). We use the convention that a minimum value over an empty set is \(+\infty\). Note that the last equality in (25) holds since \(P_T^{(n)}\) is defined as the type of the parameter sequence \(\theta^n\), hence averaging over time is the same as averaging according to \(P_T^{(n)}\). In addition, define the symmetrizability threshold

\[
L_n^* \triangleq \max_{p(x|t)} \frac{1}{\tilde{\Lambda}_n(p)} \sum_{i=1}^{n}\sum_{x,s \in S} p(x|\theta_i) J_{\theta_i}(s|x)l(s).
\]

(26)

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 \(\frac{1}{n}\sum_{i=1}^{n} p_n(x|\theta_i) \leq \Lambda\) 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

\[
C_n(W) \triangleq \begin{cases} 
\min_{q(s|t)} \max_{p(x|t)} I_q(X;Y|T) & \text{if } L_n^* > \Lambda, \\
0 & \text{if } L_n^* \leq \Lambda 
\end{cases}
\]

(27)

with \((T,S,X) \sim P_T(t)p(x|t)q(s|t)\), where \(P_T\) is the type of the parameter sequence \(\theta^n\) for a given block length \(n\). We use the convention that a maximum value over an empty set is 0.

**Theorem 4:** Assume that \(L_n^* \neq \Lambda\) for sufficiently large \(n\) and that (23) holds. The capacity of an AVC \(W\) with fixed parameters, under input constraint \(\Omega\) and state constraint \(\Lambda\), is given by

\[
C(W) = \liminf_{n \to \infty} C_n(W).
\]

(28)

In particular, if the channels \(W_{Y|X,S,T}(\cdot|\cdot,t)\), \(t \in T\), are non-symmetrizable, then \(C(W) = C^*(W) = \liminf_{n \to \infty} C_n(W)\).

That is, the deterministic code capacity coincides with the random code capacity.

The proof of the capacity theorem above consists of two stages:

- In Appendix E, we present the deterministic decoding rule and codebook generation for the AVC with fixed parameters and establish key properties that will be used in the main proof of the capacity theorem. Our derivation extends and modifies the methods of Csiszár and Narayan [32], who considered the AVC without parameters. Specifically, in Appendix E-A, we specify a two-step decoding rule and prove decoding disambiguity (see Lemma E.1). That is, we show that two different messages cannot satisfy the decoding rule simultaneously. In Appendix E-B, we show the existence of a deterministic codebook that satisfies three inequalities that are crucial for the achievability proof. Roughly speaking, Lemma E.3 bounds the number of “bad” messages with unwanted correlations between the codewords and an arbitrary state sequence.

- In Appendix F, we give the proof of Theorem 4, where the achievability proof utilizes the tools and properties that were established in the preceding appendix, and the converse proof extends observations due Ericson [40].

The theorem above will also play a central role in the proof of the capacity theorem in Section V.

**Remark 4:** Observe that the second part of the theorem implies that for the case where there are no constraints, i.e., \(\Omega = \phi_{max}\) and \(\Lambda = \lambda_{max}\), non-symmetrizability of the channels \(W_{Y|X,S,T}(\cdot|\cdot,t)\) is a sufficient condition for positive capacity. Specifically, according to the definition of \(\tilde{\Lambda}_n(p)\), \(L_n^*\) in (25)-(26), if some of the channels \(W_{Y|X,S,T}(\cdot|\cdot,t)\) 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. \(|I(n)| = c\) 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 (23) excludes this scenario. In particular, if \(|I(n)|\) is non-zero, then we assume that \(|I(n)|\) grows linearly in \(n\), in which case positive rates can be achieved by coding over the part of the block that lies within \(I(n)\). Furthermore, without constraints, we may replace the linear growth assumption with a poly-logarithmic one, i.e. \(|I(n)| = \Omega((\log n)^a)\), 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, \(|T| = n^2\). Therefore, without state constraints, the random element \(\gamma \in \Gamma\) can be reliably sent to the receiver over the sub-block \(I(n)\), at rate \(\rho_n = \frac{\log |I|}{(\log n)^a} = 2(\log n)^{-(a-1)}\), which tends to zero as \(n \to \infty\), hence the decrease in the overall rate is negligible as well. We deduce that if \(|I(n)| = \Omega((\log n)^a)\), then the deterministic code capacity of the AVC with fixed parameters without constraints is the same as the random code capacity, i.e.

\[
C(W) = C^*(W) = \liminf_{n \to \infty} \min_{q(s|t)} \max_{p(x|t)} I_q(X;Y|T).
\]

(29)

The question of whether there exists a (simple) necessary and sufficient condition for positive capacity remains open in general, i.e. without the assumption in (23). The condition that not all channels \(W_{Y|X,S,T}(\cdot|\cdot,t)\) are symmetrizable is necessary, but insufficient, while non-symmetrizability of all channels \(W_{Y|X,S,T}(\cdot|\cdot,t)\) is a sufficient condition but it is not necessary.

**Remark 5:** Even in the case where there are no parameters, the boundary case where \(L_n^* = \Lambda\) is an open problem. Although, it is conjectured in [32] that the capacity is zero in
this case. Similarly, we conjecture that the capacity of the AVC with fixed parameters is given by $C(W) = \liminf_{n \to \infty} C_n(W)$ for all values of $\{L_n^*\}_{n \geq 1}$, provided that (23) 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 [32].

The following corollary considers a special class of channels, for which the capacity characterization in Theorem 4 can be extended to all values of $L_n^*$ and $\Lambda$.

**Corollary 5:** Let $W$ be an AVC with fixed parameters such that all channels $W_{Y|X,S,T}=\{\cdot,\cdot,\cdot, t\}$, $t \in T$, are symmetrizable. If the minimum in (25) 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

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

The proof of Corollary 5 is given in Appendix G. In particular, we note that the condition of 0-1 law in Corollary 5 holds when the output $Y$ is a deterministic function of $X$, $S$, and $T$. As opposed to Theorem 4, the statement in Corollary 5 holds for all values of $\{L_n^*\}_{n \geq 1}$. As we will see later on, the Gaussian AVC satisfies the condition of the corollary above.

### C. Super-Additivity

We also give an equivalent formulation with a sum over $i \in [1 : n]$. As opposed to the random code characterization in Subsection III-A, the capacity formula with deterministic codes cannot be expressed by an average of the capacities of the individual AVCs $\{W_{Y|X,S,T=\theta_i}\}$. Considering the AVC without constraints, Schaefer et al. [97] 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 [97]). 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 [97].

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

$$\bar{\lambda}(p, t) \triangleq \min_{x \in \mathcal{X}} \sum_{s \in \mathcal{S}} p(x) J(s|x) l(s), \quad (31)$$

where the minimization is over the distributions $J(s|x)$ that symmetrize $W_{Y|X,S,T}(\cdot,\cdot,\cdot, t)$ for a given $t \in T$. (see Definition 4). Then, we can write the minimal symmetrizability cost defined in (25) as

$$\bar{\lambda}_n(p(\cdot)) = \frac{1}{n} \sum_{i=1}^n \bar{\lambda}(p(\cdot|\theta_i), \theta_i). \quad (32)$$

Let

$$R_n(W) \triangleq \begin{cases} 
\min_{\lambda_1,\ldots,\lambda_n: \frac{1}{n} \sum_{i=1}^n \lambda_i \leq \Lambda} \max_{\omega_1,\ldots,\omega_n: \frac{1}{n} \sum_{i=1}^n \omega_i \leq \Omega} \frac{1}{n} \sum_{i=1}^n C_n(\omega_i, \lambda_i) & \text{if } L_n^* > \Lambda, \\
0 & \text{if } L_n^* \leq \Lambda 
\end{cases} \quad (33)$$

where

$$C_n(\Omega, \Delta, \Lambda) \triangleq \min_{q(s) : \mathbb{E}(S) \leq \Delta} \max_{p(x) : \mathbb{E} \phi(X) \leq \Omega} I_q(X; Y|T = t) \quad (34)$$

We note that based on Csiszár and Narayan’s result in [32], the capacity of the constant-parameter AVC $\{W_{Y|X,S,T=\theta_i}\}$ is given by $C_n(\Omega, \Delta, \Lambda)$ with $\Delta = \Lambda$. That is, in our formula, there is an additional degree of freedom compared to the characterization for the traditional AVC in [32]. We will see in Example 1 below that the additional degree of freedom $\Delta$ can be associated with the super-additivity property.

**Lemma 6:** Let $C_n(W)$ and $R_n(W)$ be as in (27) and (33) respectively. Then,

$$R_n(W) = C_n(W). \quad (35)$$

The proof of Lemma 6 is given in Appendix H. Theorem 4, Corollary 5, and Lemma 6 yield the following consequence.

**Corollary 7:** The deterministic code capacity of the AVC $W$ with fixed parameters, under input constraint $\Omega$ and state constraint $\Lambda$, is given by

$$C(W) = \liminf_{n \to \infty} R_n(W), \text{ if } L_n^* \neq \Lambda \text{ for sufficiently large } n \text{ and (23) holds.} \quad (36)$$

Furthermore, if the minimum in (31) is attained by a 0-1 law, for every $p(x)$ with $\mathbb{E}\phi(X) \leq \Omega$, and for all $t \in T$, then

$$C(W) = \liminf_{n \to \infty} R_n(W), \quad (37)$$

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,

$$Y = X + S + Z_T \mod 2 \quad (38)$$

with $X = S = T = \{0, 1\}$, where $Z_t \sim \text{Bernoulli}(\varepsilon_t)$, for $t = 0, 1$, $\varepsilon_0 < \varepsilon_1 < \frac{1}{2}$, and $T^n \equiv \theta^n$ is a given sequence of fixed parameters. 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 $l(s) = s$. One may view this as a channel that alternates between two AVCs, $Y = X + S + Z_0$ and $Y = X + S + Z_1$, with addition modulo 2.

Recall from Remark 1, that the special case where the parameter sequence is constant, i.e., $\theta_i = t$ for all $i$, reduces to the traditional AVC without parameters. That is, if $\theta_i = 0$ for
all $i$, then we obtain the traditional arbitrarily varying BSC that was considered in [32, Example 1]. The same reduction also applies if the parameter sequence is given by $\theta_i = 1$ for all $i$. Then, consider a given $t \in \{0, 1\}$. For such constant-parameter AVC, where $\theta_i = t$ for all $i$, 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 $\zeta = J(1|1) = 1 - J(0|0)$, we have

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

(39)

Let $\overline{C}_t(\omega, \lambda)$ denote the deterministic code capacity of the constant-parameter AVC $\{W_{Y|X,S,T=t}\}$, i.e. with $\theta_i = t$ for all $i$, under input constraint $\omega$ and state constraint $\lambda$. Then, based on the analysis by Csiszár and Narayan [32, Example 1] for the traditional AVC,

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

(40)

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

Next, we consider a sequence of fixed parameters that varies in time. Let $\theta^0$ be a parameter sequence with half zeros and half ones, say $\theta_{2i} = 0$ and $\theta_{2i-1} = 1$ for $i = 1, 2, \ldots$. In the limit of $n \to \infty$, the empirical distribution of the parameter sequence $(P_T^{(n)}(0), P_T^{(n)}(1))$ converges to $(\frac{1}{2}, \frac{1}{2})$. Hence, it is sufficient to consider even $n$, in which case $P_T^{(n)}(0) = P_T^{(n)}(1) = \frac{1}{2}$. Suppose that

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

(41)

We derive the deterministic code capacity using Corollary 7. By (32) and (39), the minimal symmetrizability cost satisfies $\Delta_0(P_{X|T}) \leq \frac{1}{2} \sum_{\varepsilon_0, \varepsilon_1} P_{X|T}(1|t)$ with equality for $P_{X|T}(1|t) \leq \frac{1}{2}$, $t \in \{0, 1\}$. Thus, the symmetrizability threshold defined in (26) is given by

$$
L_n^* = \max_{P_{X|T}: \frac{1}{2} \sum_{\varepsilon_0, \varepsilon_1} E[\phi(X)|T=\theta_t] \leq \Omega} \tilde{\Lambda}_n(P_{X|T})
$$

$$
= \max_{P_{X|T}: \frac{1}{2} P_{X|T}(1|0) + \frac{1}{2} P_{X|T}(1|1) \leq \Omega} \left[ \frac{1}{2} P_{X|T}(1|0) + P_{X|T}(1|1) \right]
$$

$$
= \Omega = \frac{5}{16}
$$

(42)

where the second equality holds since $E[\phi(X)|T = t] = P_{X|T}(1|t)$ for $t \in \{0, 1\}$. Thus, $L_n^* > \Lambda$.

By Corollary 7, the deterministic code capacity is given by

$$
C(W) = \lim \inf_{n \to \infty} R_n(W),
$$

with

$$
C(W) = \min_{\lambda_0, \lambda_1: \frac{1}{2} \lambda_0 + \lambda_1 \leq \Lambda} \max_{\omega, \lambda_0, \lambda_1: \varepsilon_0, \varepsilon_1, \omega_0, \omega_1, \lambda_0, \lambda_1: \varepsilon_0, \varepsilon_1, \omega_0, \omega_1, \lambda_0, \lambda_1: \frac{1}{2} \sum_{\varepsilon_t \in \{0, 1\}} C_t(\omega_t, \delta_t, \lambda_t)
$$

(43)

(see (33)-(34)). Hence, we need to compute $C_t(\omega, \delta, \lambda)$ first. Taking $X \sim \text{Bernoulli}(p)$ and $S \sim \text{Bernoulli}(q)$, where $0 \leq p, q \leq 1$, we have

$$
C_t(\omega, \delta, \lambda) = \min_{0 \leq \omega \leq \lambda} \max_{\delta \geq \min(p, 1-p)} (h(p \ast q \ast \varepsilon_t) - h(q \ast \varepsilon_t))
$$

$$
= \begin{cases} 
(h(\bar{\omega}_{\delta, \lambda} \ast \varepsilon_t) - h(\lambda \ast \varepsilon_t)) & \text{if } \lambda \leq \frac{1}{2}, \bar{\omega}_{\delta, \lambda} \leq \frac{1}{2} \\
1 - h(\lambda \ast \varepsilon_t) & \text{if } \lambda \leq \frac{1}{2}, \bar{\omega}_{\delta, \lambda} \geq \frac{1}{2} \\
0 & \text{if } \lambda \geq \frac{1}{2}
\end{cases}
$$

(44)

with $\bar{\omega}_{\delta, \lambda} \equiv \min(\omega, \delta)$, for $t \in \{0, 1\}$.

For every convex-concave function $F(x, y)$, the saddle point of $F$ is in $(x^*, y^*)$ if and only if $F(x^*, y^*) \leq F(x, y^*) \forall x, y$. As the function $f(p, q) = h(p \ast q \ast \varepsilon) - h(q \ast \varepsilon)$ is convex in $q$ and concave in $p$, the saddle-point of the functional in the RHS of (43) is attained for

$$
\lambda_0 \ast \varepsilon_0 = \lambda_1 \ast \varepsilon_1, \quad \frac{1}{2}(\lambda_0 + \lambda_1) = \Lambda
$$

(45a)

$$
\omega_0 = \omega_1 = \delta_0 = \delta_1 = \Omega
$$

(45b)

with $\Omega$ and $\Lambda$ as in (41). This solution can be thought of as the binary version of double water filling (see Figure 1 and Section IV). Namely, the jammer generates a binary state sequence $S^n$ such that the total interference is white, i.e. $S^n + Z_{\theta_t} \sim \text{Bernoulli}(\lambda_{\theta_t} \ast \varepsilon_{\theta_t})$ is an i.i.d. sequence. Given white interference, the optimal input is white as well. We deduce from Corollary 7 that the deterministic code capacity can be expressed as

$$
C(W) = R_n(W)
$$

$$
= \frac{1}{2} (h(a_0 \ast b_0 \ast \varepsilon_0) - h(b_0 \ast \varepsilon_0)) + \frac{1}{2} (h(a_1 \ast b_1 \ast \varepsilon_1) - h(b_1 \ast \varepsilon_1))
$$

(46)

with $a_0 = a_1 = \frac{5}{16}, b_0 = \frac{3}{8}$ and $b_1 = \frac{1}{8}$.

On the other hand, if one would decide to use two separate codes for $W_{Y|X,S,T=0}$ and $W_{Y|X,S,T=1}$ independently, then he or she could only achieve the rate

$$
\frac{1}{2} C_0(a_0, b_0) + \frac{1}{2} C_1(a_1, b_1)
$$

$$
= 0 + \frac{1}{2} (h(a_1 \ast b_1 \ast \varepsilon_1) - h(b_1 \ast \varepsilon_1))
$$

$$
< C(W).
$$

(47)

That is, using two separate codes for the channel described above is suboptimal, because the deterministic code capacity of the arbitrarily varying BSC with fixed parameters is strictly larger than the average of the capacities of the individual AVCs, $\{W_{Y|X,S,T=0}\}$ and $\{W_{Y|X,S,T=1}\}$. 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 [97, Theorem 6].
D. Example: Channel With Fadings

To illustrate our results, we give another example.

Example 2: Consider an arbitrarily varying fading channel,

\[ Y_i = \theta_i X_i + S_i + Z_i, \]

with a Gaussian noise sequence \( Z^n \) that is i.i.d. \( \sim \mathcal{N}(0, \sigma^2) \), where \( \theta_1, \theta_2, \ldots \) is a sequence of fixed fading coefficients. Recently, Hosseinigoki and Kosut [61] considered this channel with a random memoryless sequence of fading coefficients. Yet, we assume that the fading coefficients are fixed, and establish by Hosseinigoki and Kosut [61] for a random memoryless sequence of fading coefficients.

We show that the capacity of the AVC with fixed fading coefficients is given by

\[ C_\nu^*(W) = \liminf_{n \to \infty} C_n^*(W), \]

Then, we show that

\[ C_n^*(W) = \min_{\lambda(t) : E(X) \leq \Lambda} \max_{\omega(t) : E(\theta T) \leq \Omega} \mathbb{E} \left[ \frac{1}{2} \log \left( 1 + \frac{T^2 \omega(T)}{\lambda(T) + \sigma^2} \right) \right], \]

where \( T \sim P_T^{(n)} \) is a dummy random variable which is distributed according to the type of the sequence \( \theta^n \), namely \( P_T^{(n)} = P_{\theta^n} \).

As for the deterministic code capacity, we show that the minimum in (25) is attained by a 0-1 law that gives probability 1 to \( s = \theta_i \) for some \( i \), hence we can determine the capacity using Corollary 5. We show that the minimal symmetrizability cost is given by

\[ \Lambda_n(F_{X|T}) = \frac{1}{n} \sum_{i=1}^n \theta_i^2 \mathbb{E}[X^2 | T = \theta_i] = \mathbb{E}(T^2 X^2), \]

with \( T \sim P_T^{(n)} \) and \( X \sim F_{X|T} \) as defined above, and we deduce that the capacity of the AVC with fixed fading coefficients is given by

\[ C(W) = \liminf_{n \to \infty} C_n(W), \]

with

\[ C_n(W) \triangleq \left\{ \begin{array}{ll}
\min_{\lambda(t) : E(X) \leq \Lambda} & \max_{\omega(t) : E(\theta T) \leq \Omega} \mathbb{E} \left[ \frac{1}{2} \log \left( 1 + \frac{T^2 \omega(T)}{\lambda(T) + \sigma^2} \right) \right] \\
\text{if } & \max_{\omega(t) : E\theta T) \leq \Omega} \mathbb{E}(T^2 \omega(T)) > \Lambda, \\
0 & \text{if } \max_{\omega(t) : E(\theta T) \leq \Omega} \mathbb{E}(T^2 \omega(T)) \leq \Lambda.
\end{array} \right. \]
sequences \((S_1, \ldots, S_d)\) of unknown distribution, not necessarily independent nor stationary. That is, \((S_1, \ldots, S_d) \sim F_{S_1, \ldots, S_d}\), where \(F_{S_1, \ldots, S_d}\) is an unknown joint cumulative distribution function (cdf) over \(\mathbb{R}^{nd}\). In particular, \(F_{S_1, \ldots, S_d}\) could give probability mass 1 to a particular sequence of state vectors \((s_1, \ldots, s_d) \in \mathbb{R}^{nd}\). The channel is subject to input constraint \(\Omega > 0\) and state constraint \(\Lambda > 0\),

\[
\sum_{j=1}^{d} \|X_j\|^2 \leq n\Omega \quad \text{w.p. } 1,
\]

\[
\sum_{j=1}^{d} \|S_j\|^2 \leq n\Lambda \quad \text{w.p. } 1. \tag{57}
\]

### B. Coding

We introduce preliminary definitions for the AVGPC.

**Definition 5 (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 \(f_j : [1 : 2^nR] \rightarrow \mathbb{R}^s\), for \(j \in [1 : d]\), such that

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

and a decoding function \(g : \mathbb{R}^{nd} \rightarrow [1 : 2^nR]\). Given a message \(m \in [1 : 2^nR]\), the encoder transmits \(x_j = f_j(m)\), for \(j \in [1 : d]\). The codeword is then given by \(X^d = f^d(m) \triangleq (f_1(m), f_2(m), \ldots, f_d(m))\). The decoder receives the channel outputs \(y^d = (y_1, \ldots, y_d)\), and finds an estimate of the message \(\hat{m} = g(y^d)\). We denote the code by \(\mathcal{C} = (f^d, g)\).

Define the conditional probability of error of a code \(\mathcal{C}\) given the sequence \(s^d = (s_1, \ldots, s_d)\) by

\[
P_{e|s^d}(\mathcal{C}) \equiv \frac{1}{2^nR} \sum_{m=1}^{2^nR} \int_{y \in \mathbb{R}^{nd} : g(y^d) \neq m} d\text{y}^d \cdot f_{\mathcal{C}}(m, s^d(y^d)), \tag{59}
\]

where \(f_{\mathcal{C}}(m, s^d(y^d)) = \prod_{i=1}^{n} f_{Z_i}(y_i^d - f_i(m) - s_i^d)\), with

\[
f_{Z_i}(z_i^d) = \frac{1}{\sqrt{(2\pi)^d|K_Z|}}e^{-\frac{1}{2}z_i^dK_Z^{-1}z_i^d}. \tag{60}
\]

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

\[
P_{e|s^d}(\mathcal{C}) \leq \varepsilon, \quad \text{for all } s^d \in \mathbb{R}^{nd} \text{ with } \sum_{j=1}^{d} \|s_j\|^2 \leq n\Lambda. \tag{61}
\]

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 \(C(\Sigma)\). 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 6 (Random Code):** A \((2^nR, n)\) random code for the AVGPC consists of a collection of \((2^nR, n)\) codes \(\{\mathcal{C}_\gamma = (f^d, g_\gamma)\}_{\gamma \in \Gamma}\), 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

\[
\sum_{\gamma \in \Gamma} \mu(\gamma) \sum_{j=1}^{d} ||f_{\mathcal{C}_\gamma}(j)||^2 \leq n\Omega, \quad \text{for all } m \in [1 : 2^nR], \tag{62}
\]

and

\[
P_{e|s^d}(\mathcal{C}^\gamma) \equiv \sum_{\gamma \in \Gamma} \mu(\gamma)P_{e|s^d}(\mathcal{C}_\gamma) \leq \varepsilon
\]

for all \(s^d \in \mathbb{R}^{nd}\) with \(\sum_{j=1}^{d} \|s_j\|^2 \leq n\Lambda. \tag{63}
\]

The capacity achieved by random codes is denoted by \(C^\ast(K_Z)\), 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

\[
\Sigma = \text{diag}(\sigma^2_1, \ldots, \sigma^2_d), \tag{64}
\]

i.e. \(Z_1, \ldots, Z_d\) are independent and \(Z_j \sim N(0, \sigma^2_j)\). Denote the random code capacity of the AVGPC with parallel channels by \(C^\ast(\Sigma)\). Hughes and Narayan [65] 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

\[
N^*_j = [\beta - \sigma^2_j]^+, \quad j \in [1 : d]
\]

with \([t]^+ = \max\{0, t\}\), where \(\beta > 0\) is chosen to satisfy

\[
\sum_{j=1}^{d} [\beta - \sigma^2_j]^+ = \Lambda. \tag{66}
\]

Next, let

\[
P^*_j = [\alpha - (N^*_j + \sigma^2_j)]^+, \quad j \in [1 : d], \tag{67}
\]

where \(\alpha > 0\) is chosen to satisfy

\[
\sum_{j=1}^{d} [\alpha - (N^*_j + \sigma^2_j)]^+ = \Omega. \tag{68}
\]

We can now define Hughes and Narayan’s capacity formula [65].

\[
C^\ast(\Sigma) \triangleq \sum_{j=1}^{d} \frac{1}{2} \log \left(1 + \frac{P^*_j}{N^*_j + \sigma^2_j} \right). \tag{69}
\]

**Theorem 8 (see [65]):** The random code capacity of the AVGPC is given by

\[
C^\ast(\Sigma) = C^\ast(\Sigma). \tag{70}
\]
Fig. 2. 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 $(N_j^*_{\beta})_{j=1}^{10} = (0, 0, 1, 2.5, 1.5, 2.2, 0.8, 0, 0, 0, 0)$ and $(P_j^*)_{j=1}^{10} = (1, 0, 2, 2, 2, 1.5, 0.5)$.

D. Observations on the Water Filling Game

We give further observations on the results by Hughes and Narayan [65], which will be useful in the sequel. By [65, Theorem 3], the random code capacity is the solution of the following optimization problem,

$$\min_{\beta} \max_{\alpha} \sum_{j=1}^{d} \frac{1}{2} \log \left( 1 + \frac{P_j}{N_j + \sigma_j^2} \right), \quad (71)$$

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

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

$$v(P_1, \ldots, P_d, N_1, \ldots, N_d) \triangleq \sum_{j=1}^{d} \frac{1}{2} \log \left( 1 + \frac{P_j}{N_j + \sigma_j^2} \right). \quad (72)$$

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. [29, 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 $\sigma_1^2, \ldots, \sigma_d^2$. 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 2, 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 $\alpha = 6$. Then, substituting into (65) and (67) yields the power allocations $(N_j^*_{\beta})_{j=1}^{10} = (0, 0, 1, 2.5, 1.5, 2.2, 0.8, 0, 0, 0, 0)$ for the jammer and $(P_j^*)_{j=1}^{10} = (1, 0, 2, 2, 2, 1.5, 0.5)$ for the transmitter.

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

Lemma 9: The quantities defined by (65)-(69) satisfy

$$1) \alpha > \beta$$
$$2) N_j^* > 0 \Rightarrow P_j^* > 0 \forall j \in [1 : d]$$
$$3) P_j^* + N_j^* + \sigma_j^2 = \max(\alpha, \sigma_j^2)$$
$$4) C(\Sigma) = \sum_{j=1}^{d} \frac{1}{2} \log \frac{\max(\alpha, \sigma_j^2)}{\max(\beta, \sigma_j^2)}. \quad (73)$$

For completeness, we give the proof of Lemma 9 in Appendix J. Based on the water filling analogy of the power allocation above, part 1 of Lemma 9 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 2). Part 3 and part 4 are not surprising either since, as can be seen in Figure 2, 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 [111] (see also [85] [79, 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 diverting this energy to some other channel $j'$ with $P_j^* > 0$. Thus, such strategy profile is irrational and cannot be a Nash equilibrium.
In the discussion section, we will consider the analogy between water filling games for the AVGPC and for the multiple access channel (see Subsection VI-C).

E. Results

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

We determine the capacity of the AVGPC. Based on Csiszár and Narayan’s result in [32], 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 [77] [88, Definition 5 and Theorem 5]). By [32, Definition 2], a AVGPC is symmetrized by a conditional pdf \( \varphi(s^d|x^d) \) if

\[
\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \varphi(s^d|x^d) f_{Z^d}(y^d - x^d - s^d) ds^d = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \varphi(s^d|x^d) f_{Z^d}(y^d - x^d - s^d) ds^d, \\
\forall x^d_1, x^d_2, y^d \in \mathbb{R}^d,
\]

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 (74) 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

\[
\tilde{\Lambda}(F_{X^d}) = \min \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X^d}(x^d) \varphi(s^d|x^d) \|s^d\|^2 ds^d dx^d,
\]

where the minimization is over all conditional pdfs \( \varphi(s^d|x^d) \) that symmetrize the channel, that is, satisfy (74). 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 10: For a zero mean Gaussian vector \( X^d \sim N(0, K_X) \),

\[
\tilde{\Lambda}(F_{X^d}) = \text{tr}(K_X).
\]

The proof of Lemma 10 is given in Appendix K. The proof builds on our observation that (74) 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 (75) is attained by \( \varphi(s^d|x^d) = \delta(s^d - x^d) \), where \( \delta(\cdot) \) denotes the Dirac delta function. Moving to the capacity theorem, define

\[
\mathcal{C}(\Sigma) = \begin{cases} 
\mathcal{C}^*(\Sigma) & \text{if } \Omega > \Lambda, \\
0 & \text{otherwise}.
\end{cases}
\]

Theorem 11: The deterministic code capacity of the AVGPC is given by

\[
\mathcal{C}(\Sigma) = \mathcal{C}(\Sigma).
\]

The proof of Theorem 11 is given in Appendix L. We note that the result above can also be obtained by combining the results of Hughes and Narayan [64] with the observations of Csiszár [30]. The analysis in the proof of Theorem 11 is independent of our results in Section III, as the AVGPC does not have fixed parameters. Considering the scalar case, Csiszár and Narayan showed the direct part by providing a coding scheme for the Gaussian AVC [34]. 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 simpler approach. To prove the 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 10 to show that if \( \Omega > \Lambda \), then the transmitter’s water filling strategy in (67) 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. In the discussion section, we will compare between the derivations and also discuss the disadvantages of our approach compared to the direct derivation in [34] (see Subsection VI-B).

F. Discussion on the Gaussian Product Channel

We give a couple of remarks on our result in Theorem 11. As in the case of the Gaussian scalar AVC [34], 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 \( S \) “look like” the input sequence \( X \) [34]. 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 \( Y_j \) to be useful. This is not the case. Since the decoder has the vector of outputs \( (Y_1, \ldots, Y_d) \), even if \( S_j \) looks like \( X_j \), the receiver could still gain information from \( Y_j \) as the other outputs may “break the symmetry”.

Based on Shannon’s classic water filling result [100], 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 [34], 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 Shannon’s Gaussian product channel [100], 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 [97].

V. MAIN RESULTS – AVC WITH COLORED GAUSSIAN NOISE

We consider an AVC with colored Gaussian noise, i.e.
\[ Y = X + Z + S, \]  
where \( Z \) 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 \( C^r(\Psi_Z) \) and \( C(\Psi_Z) \), 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
\[ b^*(\omega) = \left[ \beta - \Psi_Z(\omega) \right]_+, \quad -\pi \leq \omega \leq \pi, \]  
where \( \beta \geq 0 \) is chosen to satisfy
\[ \frac{1}{2\pi} \int_{-\pi}^{\pi} \left[ \beta - \Psi_Z(\omega) \right]_+ d\omega = \Lambda. \]  
Next, define
\[ a^*(\omega) = \left[ \alpha - (b^*(\omega) + \Psi_Z(\omega)) \right]_+, \quad -\pi \leq \omega \leq \pi, \]  
where \( \alpha \geq 0 \) is chosen to satisfy
\[ \frac{1}{2\pi} \int_{-\pi}^{\pi} \left[ \alpha - (b^*(\omega) + \Psi_Z(\omega)) \right]_+ d\omega = \Omega. \]

Now, let
\[ C^r(\Psi_Z) \equiv \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{2} \log \left( 1 + \frac{a^*(\omega)}{b^*(\omega) + \Psi_Z(\omega)} \right) d\omega. \]

**Theorem 12:** The random code capacity of the AVC with colored Gaussian noise is given by
\[ C^r(\Psi_Z) = C^r(\Psi_Z), \]  
and the deterministic code capacity is given by
\[ C(\Psi_Z) = \begin{cases} C^r(\Psi_Z) & \text{if } \Omega > \Lambda, \\ 0 & \text{otherwise.} \end{cases} \]  

The proof of Theorem 12 is given in Appendix M, 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 3 and Corollary 7, 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. [29, 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 1 and Example 3 below.

Our observations in the previous section can be also be extended to the AVC with colored Gaussian noise. In particular, the optimal jamming strategy is given by the water filling solution only if the user has strictly larger transmission power that the jammer, i.e. when \( \Omega > \Lambda \). Otherwise, the optimal jamming strategy is to simulate the input and use the same power allocation as the encoder uses. We further observed that as opposed to Shannon’s classic result [100], using \( d \) independent codes for the parallel channels of an AVGPC may result in a strictly lower rate than the deterministic code capacity (see Subsection IV-F). Similarly, for the AVC with colored Gaussian noise, using independent deterministic codes for orthogonal frequency bands is suboptimal in general.

**Example 3:** Consider an AVC with colored Gaussian noise, where the power spectral density of the noise process is given by
\[ \Psi_Z(\omega) = \frac{3}{1.25 + \cos(\omega)}, \quad -\pi \leq \omega \leq \pi. \]

This spectral representation corresponds to a noise process with an exponential auto-correlation function \( r_Z(k) = 4 \left( \frac{1}{2} \right)^{|k|} \), where the auto-correlation function is defined as \( r_Z(k) \equiv r_Z(i, i + k) \equiv E(Z_i Z_{i+k}) \).

Suppose that the input and state constraints are \( \Omega = 7 \) and \( \Lambda = 5 \). The double water filling procedure is illustrated in Figure 3 along with the optimal power spectral densities. The curve in Figure 3a depicts the power spectral density \( \Psi_Z(\omega) \) of the Gaussian noise process. First, the jammer pours water of volume \( \Lambda \) on top of \( \Psi_Z(\omega) \), and then the encoder pours more water of volume \( \Omega \), where \( \Omega \) and \( \Lambda \) are the input and state constraints, respectively. The water levels after the first and the second stages of the procedure are \( \beta \) and \( \alpha \), respectively (see (81) and (83)). By Theorem 12, the deterministic code and random code capacities are the same as \( \Omega > \Lambda \). By numerical calculation of the water filling solution, we obtain the water levels \( \beta = 3.1518 \) and \( \alpha = 4.8322 \). Hence, the optimal strategy for the encoder and the jammer is to use the power spectral densities \( \Psi_X(\omega) = a^*(\omega) \) and \( \Psi_S(\omega) = b^*(\omega) \), where
\[ b^*(\omega) = \begin{cases} 3.1518 - \frac{3}{1.25 + \cos(\omega)} & \text{if } |\omega| \leq 1.8736 \\ 0 & \text{otherwise.} \end{cases} \]  

and
\[ a^*(\omega) = \begin{cases} 1.6805 & \text{if } |\omega| \leq 1.8736 \\ 4.8324 - \frac{3}{1.25 + \cos(\omega)} & \text{if } 1.8736 \leq |\omega| \leq 2.2513 \\ 0 & \text{otherwise.} \end{cases} \]
Fig. 3. Double water filling in the frequency domain for the AVC with colored Gaussian noise in Example 3, with the input and state constraints $\Omega = 7$ and $\Lambda = 5$. This results in the capacity value $C(\Psi_Z) = C^\star(\Psi_Z) \approx 25.1327$ bits per transmission (80).

Now, suppose that the input and state constraints are $\Omega = 5$ and $\Lambda = 7$. Then, using random codes, the optimal power spectral densities are

$$a^\star(\omega) = \begin{cases} 1.1671 & \text{if } |\omega| \leq 2.0169 \\ 4.8322 - \frac{3}{1.25 + \cos(\omega)} & \text{if } 2.0169 \leq |\omega| \leq 2.2513 \\ 0 & \text{otherwise} \end{cases}$$

(92)

As the formula (84) changes slowly in the signal to noise ratio, it turns out that the random code capacity has roughly the same value, i.e. $C^\star(\Psi_Z) \approx 25.1327$ bits per transmission.

VI. DISCUSSION AND SUMMARY

In this section, we compare our results and analysis with previous work and methods that are commonly used in the literature. Specifically, we compare between the fixed-parameter model and the AVC with partial side information; we discuss the advantages and disadvantages of the discretization approach that was suggested in a comment by Csiszár and Narayan [34] compared with the direct derivation that was used in the same paper; and we portray the analogy between water filling games for the AVC and for the multiple access channel.

A. Fixed Parameters Vs. Side Information

The AVC with fixed parameters

$$W = \{ W_{Y|X,S,T} | \theta^\infty \}$$

(93)

is defined in Subsection II-B as a state-dependent channel that depends on the state-parameter pair $S = (S,T)$, where the sequence $S^n$ is arbitrary as for the traditional AVC, and the sequence $T^n = \theta^n$ is fixed as for the time-varying channel (see Remark 1). Having a parameter sequence that is known to the encoder, the decoder, and the jammer, it may be tempting to mistake the channel for an AVC with partial side information (PSI), i.e. when $(S^n, T^n)$ are arbitrary yet the sequence $T^n$ is available to the encoder and the decoder as side information [83]. However, this scenario is very different from ours, because in our setting $T^n$ is not an arbitrary sequence, namely $T^n = \theta^n$ where $\theta_1, \theta_2, \ldots$ is a specific sequence, and this specific sequence is an integral part of the channel model.

The characterization for an AVC with partial side information is significantly different as well. In particular, by Theorem 1, the random code capacity of the AVC with fixed parameters is given by

$$C^\star(W) = \liminf_{n \to \infty} \frac{1}{n} \max \left\{ I_q(X;Y|T) : q(s|t) \right\}$$

with partial side information (PSI), i.e. when $(S^n, T^n)$ are arbitrary yet the sequence $T^n$ is available to the encoder and the decoder as side information [83]. However, this scenario is very different from ours, because in our setting $T^n$ is not an arbitrary sequence, namely $T^n = \theta^n$ where $\theta_1, \theta_2, \ldots$ is a specific sequence, and this specific sequence is an integral part of the channel model.

The characterization for an AVC with partial side information is significantly different as well. In particular, by Theorem 1, the random code capacity of the AVC with fixed parameters is given by

$$C^\star(W) = \liminf_{n \to \infty} \frac{1}{n} \max \left\{ I_q(X;Y|T) : q(s|t) \right\}$$

with partial side information (PSI), i.e. when $(S^n, T^n)$ are arbitrary yet the sequence $T^n$ is available to the encoder and the decoder as side information [83]. However, this scenario is very different from ours, because in our setting $T^n$ is not an arbitrary sequence, namely $T^n = \theta^n$ where $\theta_1, \theta_2, \ldots$ is a specific sequence, and this specific sequence is an integral part of the channel model.

The characterization for an AVC with partial side information is significantly different as well. In particular, by Theorem 1, the random code capacity of the AVC with fixed parameters is given by

$$C^\star(W) = \liminf_{n \to \infty} \frac{1}{n} \max \left\{ I_q(X;Y|T) : q(s|t) \right\}$$

with partial side information (PSI), i.e. when $(S^n, T^n)$ are arbitrary yet the sequence $T^n$ is available to the encoder and the decoder as side information [83]. However, this scenario is very different from ours, because in our setting $T^n$ is not an arbitrary sequence, namely $T^n = \theta^n$ where $\theta_1, \theta_2, \ldots$ is a specific sequence, and this specific sequence is an integral part of the channel model.
other hand, the random code capacity of the AVC with partial side information is given by
\[
C^*(V_{PS1}) = \min_{q(t)} \min_{q(s|t)} \max_{p(x|t) : \mathbb{E}_p(x) \leq \Omega} I_q(X; Y|T)
\]
(96)
with \((T, S, X) \sim q(t)q(s|t)p(x|t)\). Note that in this case, we have minimization over the distribution of \(T\) (cf. (95) and (96)). Therefore, in general, the random code capacity of the AVC with partial side information can be lower than that of the AVC with fixed parameters.

**B. Discretization Vs. Direct Derivation**

We have repeatedly used the fact that the result for the Gaussian AVC can be derived from the discrete result using discretization techniques. This approach differs from the one taken by Csiszár and Narayan [34] and by Hughes and Narayan [64, 65] using minimum-distance decoding with a geometric interpretation. This seems to have led to a belief that it is not possible to use the discrete result for this purpose. Notwithstanding, Csiszár and Narayan write in their 1991 paper on the Gaussian AVC [34, page 19]:

“Suitable approximation arguments would enable a derivation of our theorem directly from the results of [C&N ’88]. Instead, we prefer to present a more transparent and direct proof, which will also serve to keep this paper self-contained.”

where [C&N ’88] refers to their paper on the discrete AVC [32]. In fact, the paper [34] was concluded with the observation that extending the techniques may become unmanageable and thus require a recourse to discretization (see Section III therein). Later, in [30], Csiszár considered the general AVC with general alphabets using the measure-theoretic formalism, and demonstrated the discretization procedure in detail. Nonetheless, it should be noted that when comparing between the derivations in our work and in [34], the approach in [34] has the remarkable advantage of showing that the capacity can be achieved with a minimum-distance decoder, and this has significant implications in practice, whereas the implementation of our discretized decoder would be highly inefficient.

**C. Water Filling Game for Multiple Access Channels**

We consider the analogy between the water filling game in Subsection IV-D and water filling results for the multiple access channel (MAC). Water filling in two (or more) stages appears in other settings in the literature, e.g. [28], [75], [119], [121]. Consider a Gaussian product MAC, where \(Y_j = X_{1,j} + X_{2,j} + Z_j, j \in \{1 \ldots d\}\), under the input constraints \(\|X_{1,j}\|^2 \leq n\Omega\) and \(\|X_{2,j}\|^2 \leq n\Lambda,\) with \(Z_j \sim \mathcal{N}(0,\sigma_j^2)\). Intuitively, the AVGPC can be thought of as a byzantine variation of the MAC, where the second transmitter is replaced by a jammer. By [28], 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 [28, 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
\[
P_j^* + N_j^* = \left[\alpha - \sigma_j^2\right]_+, j \in \{1 \ldots d\},
\]
(97)
such that \(\sum_{j=1}^{d}(P_j^* + N_j^*) = \Omega + \Lambda\), and
\[
N_j^* = \left[\beta - \sigma_j^2\right]_+, j \in \{1 \ldots d\},
\]
(98)
such that \(\sum_{j=1}^{d}N_j^* = \Lambda\). Following part 3 of Lemma 9, it can be seen that the strategy above is equivalent to (65)-(68). The total power allocation in (97) seems natural in order to maximize the sum rate. Though, our presentation in (65)-(68) 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, we have seen in Section IV-E that the deterministic code capacity in our adversarial problem has a different behavior. Specifically, the deterministic code characterization depends on who has more power, the transmitter or the jammer.

Another water filling game is described by Lai and El Gamal in [75], 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, \(\mathbb{E}_\mu \|X_1\|^2 \leq n\Omega\) and \(\mathbb{E}_\mu \|X_2\|^2 \leq n\Lambda\). As shown in [75], 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
\[
P_{h_1, h_2}^* = \begin{cases} 
[\beta_1 - \sigma_1^2/h_1]_+ & \text{if } \beta_1 \frac{h_1}{h_2} \geq \beta_2 \\
0 & \text{otherwise}
\end{cases}
\]
\[
N_{h_1, h_2}^* = \begin{cases} 
[\beta_2 - \sigma_2^2/h_2]_+ & \text{if } \beta_1 \frac{h_1}{h_2} \leq \beta_2 \\
0 & \text{otherwise}
\end{cases}
\]
(99)
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 Section IV-E, 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).

Game-theoretic interpretations for the general AVC and for other communication settings can also be found in [16, 20, 31, 81, 88, 104] and references therein.

**D. Summary**

To summarize, the main purpose of this work is to establish the capacity of the AVC with colored Gaussian noise. However, orthogonalizing the AVC with a stationary noise process
transforms the channel noise into a non-stationary process. This led us to consider three different channel models, that are independent of each other and cannot be presented as special cases of one another.

First, we studied 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. The sequence of fixed parameters \( \theta_1, \theta_2, \ldots \) is not arbitrary nor random, but a specific sequence which is an integral part of the model. We determined both the deterministic code capacity and the random code capacity. We showed that the random code capacity of the AVC with fixed parameters is given by the average of the capacities of the individual AVCs \( \{W_{Y|X,S,T|\cdot,\cdot,\theta_i}\} \) over time. This is not surprising given the characterization of the TVC [110] [31, Problem 6.8]. On the other hand, we demonstrated that the deterministic code capacity is super-additive, i.e. the deterministic code capacity can be strictly larger than the time-average of the capacities for each parameter value in separate.

In the second part of this paper, we derived the deterministic code capacity of the AVGPC, where there is white Gaussian noise and no parameters. We also discussed the game-theoretic interpretation of Hughes and Narayan’s random code characterization [65], and the connection between the double water filling solution and the idea of Nash equilibrium in game theory. We further examined the connection between the AVGPC and the product MAC [28], [75] (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 [100], it is observed that deterministic coding using independent scalar codes is suboptimal for the AVGPC.

At last, we determined the random code and deterministic code capacities of the AVC with colored Gaussian noise, where double water filling is performed in the frequency domain. As demonstrated in Figure 1, the power spectral density of the noise process is pictured as the bottom of a vessel. First, the jammer pours water of volume \( \Omega \) into the vessel, and then the encoder pours more water of volume \( \Omega \), where \( \Omega \) and \( \Omega \) are the power constraints on the transmitter and the jammer. Our main result in this paper is that the deterministic code capacity of the AVC with colored Gaussian noise is given by

\[
C(\Psi_Z) = \begin{cases} 
\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{2} \log \max\left(\alpha, \frac{\max(\alpha, \Psi_Z(\omega))}{\max(\beta, \Psi_Z(\omega))}\right) d\omega & \text{if } \Lambda < \Psi, \\
0 & \text{if } \Lambda \geq \Psi.
\end{cases}
\]  

where \( \beta \) and \( \alpha \) are the water levels after the first and the second stages of the water filling procedure (see Figure 1). Hence, the optimal power allocation is given by the double water filling solution only if the user has strictly larger transmission power than the jammer, i.e. \( \Omega > \Lambda \), and it is zero otherwise. The practical implications of our results are not just for the evaluation of the optimal throughput, but also for developing optimal communication strategies for the transmitter and for the jammer. For example, for an OFDM system, the water filling solution leads to an optimal usage of bandwidth [44]. Based on our results, the optimal jamming strategy is given by the water filling solution only if the user has strictly larger transmission power that the jammer, i.e. when \( \Omega > \Lambda \). Otherwise, the optimal jamming strategy is to simulate the input and use the same power allocation as the encoder uses.

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.

VII. Analysis

The proofs are given in Appendices A-M, which are organized as follows. Appendices A-I consist of the proofs for the results in Section III on the AVC with fixed parameters. In Appendix A, we characterize the capacity of the compound channel with fixed parameters. This is an auxiliary result used in the random code analysis for the AVC with fixed parameters. In Appendix B, we present our modification for the combinatorial robustification technique, which is also used as a tool in the random code analysis. In Appendices C and D, we prove Theorem 1 and Lemma 2, respectively, establishing the random code capacity theorem and its equivalent formulation for the AVC with fixed parameters. In Appendices E and F, we give the deterministic coding schemes and their properties, and then prove the deterministic code capacity theorem, Theorem 4. In Appendices G and H, we prove Corollary 5 and Lemma 6. In Appendix I, we give the analysis for Example 2, deriving the capacity of the AVC with white Gaussian noise and fixed fading coefficients.

Appendices J-L contain the proofs for the results on the AVGPC in Section IV. In Appendix J, we show the properties of the water filling game as stated in Lemma 9. In Appendix K, we calculate the minimal symmetrizability cost and prove Lemma 10, and in Appendix L, we prove the deterministic code capacity theorem, Theorem 11.
Appendix M provides the proof for the main result of this work, Theorem 12, where we determine the random code capacity and deterministic code capacity of the AVC with colored Gaussian noise using the results mentioned above.

**APPENDIX A**

**THE COMPOUND CHANNEL WITH FIXED PARAMETERS**

In this appendix, we give the capacity theorem for the compound channel. This is an auxiliary result, obtained by a simple extension of [31, Exercise 6.8]. A similar result appears in [78] as well.

Consider the compound channel \( \mathcal{W} \mathcal{Q}^n \) with fixed parameters under input constraint \( \Omega \) and state constraint \( \Lambda \). Recall that for the compound channel, the state sequence is distributed according to \( S^n \sim \prod_{i=1}^{n} q(s_i | \theta_i) \), where the conditional state distribution \( q(s_i | \theta_i) \) belongs to a given set \( \mathcal{Q} \), with a given sequence of fixed parameters \( \theta^0 \) (see Subsection II-B). Furthermore, under a state constraint \( \Lambda \), the set of state distributions \( \mathcal{Q} \) is a subset of \( \mathcal{P}_\Lambda(S(\theta^0)) \) (see (8)). Given a parameter sequence \( \theta^n \) of a given block length \( n \), define

\[
C_n(\mathcal{W} \mathcal{Q}^n) = \max_{p(x|t) : \mathcal{E}(x) \subseteq \Omega} \inf_{q(s|t) \in \mathcal{Q}} I_q(X;Y|T), \tag{101}
\]

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

**Lemma A.1:** The capacity of the compound channel \( \mathcal{W} \mathcal{Q}^n \) with fixed parameters, under input constraint \( \Omega \) and state constraint \( \Lambda \), i.e., with \( \mathcal{Q} \subseteq \mathcal{P}_\Lambda(S(\theta^0)) \), is given by

\[
C(\mathcal{W} \mathcal{Q}^n) = \lim_{n \to \infty} C_n(\mathcal{W} \mathcal{Q}^n), \tag{102}
\]

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

The proof is given below.

**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 \( P_T^n = \hat{P}_\theta^n \).

Define a set \( \mathcal{Q}_n \) of conditional state types, \( \mathcal{Q}_n = \{ \hat{P}_{s^n}^{(n)}(\theta^n, s^n) : (\theta^n, s^n) \in A_{\delta_1}(P^n_T \times q) \text{ for some } q \in \mathcal{Q} \} \),

\[
\hat{P}_{s^n}^{(n)}(\theta^n, s^n) = P_T^n(t)q(s|t), \text{ and } \delta_1 = \frac{\delta}{2|S|}, \tag{104}
\]

where \( \delta > 0 \) is arbitrarily small. In words, \( \hat{Q}_n \) 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 \( P_T^n(t)q(s|t) \), for some conditional state distribution \( q(s|t) \) in \( \mathcal{Q}_n \). We note that the sets \( \mathcal{Q}_n \) and \( \mathcal{Q}_n \) could be disjoint, since \( \mathcal{Q}_n \) is not limited to conditional empirical distributions. Nevertheless, for a given \( \delta > 0 \) and sufficiently large \( n \), every \( q \in \mathcal{Q}_n \) can be approximated by some \( q' \in \mathcal{Q}_n \). Indeed, for sufficiently large \( n \), there exists a joint type \( P_T^n(t)q'(s|t) \) such that \(|P_T^n(t)q'(s|t) - P_T^n(t)q(s|t)| \leq \delta_1|/|S| \), hence \( |P_T^n(t) - P_T^n(t)| \leq \delta_1 \) and \( |P_T^n(t)q'(s|t) - P_T^n(t)q(s|t)| \leq \delta_1 \). Note that \( P_T^n(t) \) is an auxiliary for the argument above, but we will no longer use it. Now, a code is constructed as follows.

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

\[
\mathbb{E}(\phi(X) = \sum_{t \in T} P_T^n(t)\mathbb{E}(\phi(X)|T = t) = \frac{1}{n} \sum_{t \in T} \sum_{i=1}^{n} P_{X|T}(x_i|x_i)\phi(x). \tag{105}
\]

As mentioned in Subsection II-D, we assume that there exists \( a \in \mathcal{X} \) with \( \phi(a) = 0 \), and thus we can always find such \( P_{X|T} \). Generate \( 2^nR \) independent sequences at random, \( x^n(m, \theta^n) \sim \prod_{i=1}^{n} P_{X|T}(x_i|x_i) \), for \( m \in [1 : 2^nR] \).

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

**Decoding:** Find a unique \( \hat{m} \in [1 : 2^nR] \), for which there exists \( q \in \hat{Q}_n \) such that \( (\theta^n, x^n(m, \theta^n), y^n) \in A_{\delta_1}(P_T^n P_{X|T} y^n) \), where

\[
P_{X,Y|T}^q(x, y|t) = \sum_{s \in S} q(s|t)W_{Y|X,S,T}(y|x, s, t). \tag{106}
\]

If there is none, or more than one such \( \hat{m} \), declare an error. We note that using the set of types \( \mathcal{Q}_n \) instead of the original set of state distributions \( \mathcal{Q}_n \) alleviates the analysis, since \( \mathcal{Q}_n \) 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 \( \Pr \left( \hat{M} \neq 1 \right) \leq \Pr (\mathcal{E}_1) + \Pr (\mathcal{E}_2 | \mathcal{E}_1 | \mathcal{E}_1 | \mathcal{E}_1) + \Pr (\mathcal{E}_3 | \mathcal{E}_2 )\), where

\[
\mathcal{E}_1 = \{ (\theta^n, X^n(1, \theta^n)) \notin A_{\delta_1}(P_T^n) P_{X|T} \},
\mathcal{E}_2 = \{ (\theta^n, X^n(1, \theta^n), Y^n) \notin A_{\delta_1}(P_T^n P_{X|T} P_{Y|X,T}^q) \}
\text{ for all } q' \in \hat{Q}_n \},
\mathcal{E}_3 = \{ (\theta^n, X^n(m, \theta^n), Y^n) \in A_{\delta_1}(P_T^n P_{X|T} P_{Y|X,T}^q) \}
\text{ for some } m \neq 1, q' \in \hat{Q}_n \}. \tag{107}
\]

The first term tends to zero exponentially by the law of large numbers and Chernoff’s bound (see e.g. [71, Theorem 1.2]). Now, suppose that the event \( \mathcal{E}_1 \) occurs. Then, for sufficiently small \( \delta \), we have that \( \phi(X^n(1, \theta^n)) \leq \Omega \), since \( \mathbb{E}(\phi(X) \leq \Omega - \epsilon \). Hence, \( X^n(1, \theta^n) \) is the channel input.

Next, we claim that the second error event implies that \( (\theta^n, X^n(1, \theta^n), Y^n) \notin A_{\delta_1}(P_T^n P_{X|T} y^n) \), where \( q(s|t) \) is the actual state distribution chosen by the jammer. Assume to the contrary that \( \mathcal{E}_2 \) holds, but \( (\theta^n, X^n(1, \theta^n), Y^n) \in A_{\delta_1}(P_T^n P_{X|T} P_{Y|X,T}^q) \). For sufficiently large \( n \), there exists a conditional type \( q' \in \hat{Q}_n \) that approximates \( q \) in the sense that \( |P_T^n(t)q'(s|t) - P_T^n(t)q(s|t)| \leq \delta_1 \) for all \( s \in S \)
for all \( x \in \mathcal{X}, t \in \mathcal{T}, y \in \mathcal{Y} \) (see (104)-(106)). To show \( \delta \)-typicality with respect to \( q'(s|t) \), we observe that

\[
|\hat{P}_{\theta^n,X^n(1,\theta^n),Y^n}(x,y) - P_T^n(t)P_{X^n|T}(x,t)P_{Y^n|X,T}(y|x,t)| \\
\leq |S| \cdot \delta_1 = \frac{\delta}{2},
\]

for some \( \delta > 0 \). This tends to zero exponentially as \( n \to \infty \) by the law of large numbers and Chernoff's bound (see e.g. [71, Theorem 1.2]).

Thus, \( \mathcal{E}_2 \) does not hold. Therefore, the RHS of (114) tends to zero exponentially as \( n \to \infty \), 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 \( n \to \infty \). Therefore, there must exist a \((2^R,n,e^{-an})\) 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 \((2^R,n,\alpha_n)\) random codes, where \( \alpha_n \to 0 \) as \( n \to \infty \).

Then, let \( X^n = I_\gamma(M, \theta^n) \) be the channel input sequence, and \( Y^n \) be the corresponding output sequence, where \( \gamma \in \Gamma \) is the random element shared between the encoder and the decoder. For every \( q \in \mathcal{Q}_n \), we have by Fano’s inequality that

\[
\mathcal{H}_q(M|Y^n, T^n = \theta^n, \gamma) \leq n \varepsilon_n,
\]

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

\[
R = \mathcal{H}(M|T^n = \theta^n, \gamma) = I_q(M;Y^n|T^n = \theta^n, \gamma) + \mathcal{H}(M|Y^n, T^n = \theta^n, \gamma)
\]

for all \( q \in \mathcal{Q}_n \), 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 $\varepsilon_n \to 0$ as $n \to \infty$. Observe that the random variable $T$ is distributed according to

$$P_T(t) = \Pr(\theta_K = t) = \sum_{i : \theta_i = t} \Pr(K = i)$$

$$= \frac{1}{n} \cdot N(t|\theta^n) = \hat{P}_\theta(t),$$

(117)

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

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

(118)

\section*{APPENDIX B

MODIFIED ROBUSTIFICATION TECHNIQUE}

We give a lemma that is based on Ahlswede’s RT [6] (see also [88, Lemma 9]). We modify it here to include the parameter sequence $\theta^n$ and the constraint on the family of conditional state distributions $q(s|t)$. Recall that we have defined $\mathcal{P}_\Lambda(S|\theta^n)$ in (8) as the set of all conditional state distributions $q(s|t)$ that satisfy $\frac{1}{n} \sum_{i=1}^n \sum_{s \in S} q(s|\theta_j)(l(s)) \leq \Lambda$.

The RT lemma provides a useful combinatorial property, which is given in terms of permutations on $n$-tuple sequences (see Definition 1.3.1 in [22]). Note that permutations can also be represented by $n \times n$ matrices [22, p. 33].

\textbf{Lemma B.1 (Modified RT):} Let $\Lambda \geq 0$ and $l : S \to [0, \infty)$. In addition, let $h : S^n \times T^n \to [0, 1]$ be a given function. If, for some $\alpha_n \in (0, 1)$, and for all $q^n(s^n|\theta^n) = \prod_{i=1}^n q(s_i|\theta_i)$, with $q \in \mathcal{P}_\Lambda(S|\theta^n)$,

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

(119)

then,

$$\frac{1}{\Pi(\theta^n)} \sum_{\pi \in \Pi(\theta^n)} h(\pi s^n, \theta^n) \leq \beta_n,$$

(120)

for all $s^n \in S^n$ such that $l^n(s^n) \leq \Lambda$,

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

Originally, Ahlswede’s RT is stated so that (119) holds for any $q(s) \in \mathcal{P}(S)$, without state constraint (see [6]), and without conditioning on the parameter sequence $\theta^n$. We give the proof of Lemma B.1 in Appendix B.

\textbf{Proof of Lemma B.1:} 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 $\bar{s}^n \in S^n$ such that $l^n(\bar{s}^n) \leq \Lambda$. Denote the conditional type of $\bar{s}^n \in S^n$ given $\theta^n$ by $\bar{q}(s|t)$. Observe that $\bar{q} \in \mathcal{P}_\Lambda(S|\theta^n)$ (see (8)), since

$$\frac{1}{n} \sum_{i=1}^n \sum_{s \in S} \bar{q}(s|\theta_i)(l(s)) = l^n(\bar{s}^n).$$

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

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

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

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

(121)

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 $q^n(s^n|\theta^n) = \prod_{i=1}^n q(s_i|\theta_i)$. Hence, taking $q = \bar{q}$,

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

(122)

and by (119),

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

(123)

Thus,

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

(124)

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

$$\left[ \frac{1}{\Pi(\theta^n)} \sum_{\pi \in \Pi(\theta^n)} h(\pi s^n, \pi \theta^n) \right] \cdot \sum_{s^n : P_{s^n|\theta^n} = \bar{q}} \bar{q}^n(s^n|\theta^n) \leq \alpha_n,$$

(125)

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

$$\sum_{s^n : P_{s^n|\theta^n} = \bar{q}} \bar{q}^n(s^n|\theta^n) \geq \frac{1}{(n + 1)|S|/\alpha_n},$$

(126)

by [29, Theorem 11.1.4]. The proof follows from (125) and (126).

\section*{APPENDIX C

PROOF OF THEOREM 1}

Consider the AVC $W$ 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 in Lemma A.1, along with our modified Robustification Technique (RT), i.e., Lemma B.1.

At first, consider the compound channel with fixed parameters under input constraint $\Omega$, with the set of conditional state distributions $Q_n = \overline{\mathcal{P}}_A(S^\theta_n)$ (see (8)). The characterization for the compound channel is given in Appendix A in terms of $C_n(WQ_n)$ as defined in (101). Then, observe that

$$Q_n = \overline{\mathcal{P}}_A(S^\theta_n) = \left\{ q(s|t) : \frac{1}{n} \sum_{i=1}^{n} \sum_{s \in S} q(s|t) I(s) \leq \Lambda \right\} = \left\{ q(s|t) : \mathbb{E}(S) \leq \Lambda \right\}$$

(127)

for $(T,S) \sim P_T^{(n)}(t)q(s|t)$, where $P_T^{(n)}$ is the type of the parameter sequence $\theta^n$. Hence,

$$C_n(WQ_n)\big|_{Q_n=\overline{\mathcal{P}}_A(S^\theta_n)} = \max_{p(x|t):\mathbb{E}(X)\leq \Omega} \min_{q(s|t):\mathbb{E}(S)\leq \Lambda} I_q(X;Y|T)$$

$$= \min_{q(s|t):\mathbb{E}(S)\leq \Lambda} \max_{p(x|t):\mathbb{E}(X)\leq \Omega} I_q(X;Y|T)$$

$$= C^*_n(W)$$

(128)

The first equality holds by (101) and (127), as $\overline{\mathcal{P}}_A(S^\theta_n)$ is a compact convex set; the second equality follows from the minimax theorem [102] since $I_q(X;Y|T)$ is concave in $p(x|t)$ and convex in $q(s|t)$; and the last equality holds by the definition of $C^*_n(W)$ in (13).

Then, let

$$R < C^*_n(W) = C_n(W\overline{\mathcal{P}}_A(S^\theta_n))$$

(129)

where $n > 0$ is arbitrarily large. According to Lemma A.1, for some $\epsilon > 0$ and sufficiently large $n$, there exists a $(2^{nR}, n)$ code $\mathcal{C} = (\pi, g(y^n, \theta^n))$ for the compound channel $W\overline{\mathcal{P}}_A(S^\theta_n)$ with fixed parameters such that

$$\phi^n(f(m, \theta^n)) \leq \Omega$$

for all $m \in [1:2^{nR}]$, (130)

and

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

(131)

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

Therefore, by Lemma B.1, 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 $n$,

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

(132)

for all $s^n \in S^n$ 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 $\pi \in \Pi(\theta^n)$,

$$P_e^{(n)}(\mathcal{C}|\pi s^n, \theta^n)$$

(133)

with $W^n = W_{Y^n|X^n,S^n,T^n}$ for short notation, where (a) is obtained by plugging $\pi s^n$ in (10a); 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

$$P_e^{(n)}(\mathcal{C}|\pi s^n, \theta^n) = \frac{1}{2^n} \sum_{m=1}^{2^n} \sum_{m=1}^{2^n} W^n(y^n|f(m, \theta^n), \pi s^n, \theta^n)$$

(134)

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

$$f^n(m, \theta^n) = \pi^{-1}f(m, \theta^n), g_\mathcal{C}(y^n, \theta^n) = g(\pi y^n, \theta^n),$$

(135)

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 (134), we see that

$$P_e^{(n)}(\mathcal{C}^{\Pi(\theta^n)}) | s^n, \theta^n) = \sum_{\pi \in \Pi(\theta^n)} \mu(\pi) P_e^{(n)}(\mathcal{C}|\pi s^n, \theta^n)$$

for all $s^n \in S^n$ with $l^n(s^n) \leq \Lambda$. Therefore, together with (132), 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 $q(s^n|\theta^n) \in \mathcal{P}_A(S^n|\theta^n)$. It follows that $\mathcal{C}^{\Pi(\theta^n)}$ is a $(2^{nR}, n, e^{-\delta n})$ random code for the AVC $W$ with fixed parameters under input constraint $\Omega$ and state constraint $\Lambda$.

B. Converse Proof

Assume to the contrary that there exists an achievable rate pair

$$R > C(WQ_n)\big|_{Q_n=\overline{\mathcal{P}}_A(S^\theta_n)}$$

(136)

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

$$\sum_{\gamma \in \Gamma} \mu(\gamma) \phi^n(f_\gamma(m, \theta^n)) \leq \Omega,$$

and

$$P_e^{(n)}(q, \theta^n, \mathcal{C}) \leq \epsilon,$$

(137)

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


Consider using the random code $C^{\Gamma}$ over the compound channel $\mathcal{W}^{\mathcal{P}_{\lambda,\delta}(S)}$ with fixed parameters under input constraint $\Omega$. Let $\mathcal{T}(s|t) \in \mathcal{P}_{\lambda,\delta}(S)$ be a given state distribution. Then, define a sequence of conditionally independent random variables $S_1, \ldots, S_n \sim \mathcal{T}(s|t)$. Letting $\mathcal{T}(s^n|\theta^n) \triangleq \prod_{i=1}^n \mathcal{T}(s_i|\theta_i)$, the probability of error is bounded by

$$
P_e^{(n)}(\mathcal{T}; \theta^n, C^{\Gamma}) \leq \sum_{s^n: t^n(s^n) \leq \Lambda} \mathcal{T}(s^n|\theta^n)P_e^{(n)}(C^{\Gamma}|s^n, \theta^n) + \Pr \left( t^n(\mathcal{T}) > \Lambda \right).$$

The first sum is bounded by (137), and the second term vanishes by the law of large numbers, since $\mathcal{T} \in \mathcal{P}_{\lambda,\delta}(S|\theta^n)$. It follows that the random code $C^{\Gamma}$ achieves a rate $R$ as in (136) over the compound channel $\mathcal{W}^{\mathcal{P}_{\lambda,\delta}(S)}$ with fixed parameters under input constraint $\Omega$, for an arbitrarily small $\delta > 0$, in contradiction to Lemma A.1. We deduce that the assumption is false, and $C^\ast(\mathcal{W}) \leq C(\mathcal{W}^{\mathcal{Q}_n})|_{\mathcal{Q}_n = \mathcal{P}_{\lambda}(S|\theta^n)} = C^\ast$.

\section*{APPENDIX D}

\section*{Proof of Lemma 2}

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

\textbf{Lemma D.1:} Let $\omega^i_*, \lambda^i_*, i \in [1:n]$, be the parameters that achieve the saddle point in (19), i.e.

$$R^\ast_n(\mathcal{W}) = \frac{1}{n} \sum_{i=1}^n C_{\theta_i}(\omega^i_*, \lambda^i_*).$$

Then, for every $i, j \in [1:n]$ such that $\theta_i = \theta_j$, we have that $\omega^i_* = \omega^j_*$ and $\lambda^i_* = \lambda^j_*$.

\textit{Proof of Lemma D.1:} For every $i \in [1:n]$, let $p_i, q_i$ denote input and state distributions such that $E(\theta_i) \leq \omega^i_*$, $E(\theta_i) \leq \lambda^i_*$ for $X_i \sim p_i$, $S_i \sim q_i$. Now, suppose that $\theta_i = \theta_j$, and define

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

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

$$I_q(X_i; Y_i|T_i = t) + I_q(X_j; Y_j|T_j = t) \leq I_q(X'; Y'|T' = t)$$

and

$$I_{q_i}(X_i'; Y_i'|T_i = t) + I_{q_j}(X_j'; Y_j'|T_j = t) \geq I_{q_i}(X_i'; Y_i'|T_i = t).$$

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 D.1 that $R^\ast_n(\mathcal{W})$

$$= \min_{(\lambda^i, \omega^i) \in \mathcal{T}} \max_{\omega^i \in \mathcal{T}} \sum_{t \in \mathcal{T}} P^\ast_n(t)C(\omega^i, \lambda^i)$$

$$= \min_{(\lambda^i, \omega^i) \in \mathcal{T}} \max_{\omega^i \in \mathcal{T}} \bigg(\sum_{t \in \mathcal{T}} P^\ast_n(t)C(\omega^i, \lambda^i)\bigg)$$

$$= \min_{\omega^i \in \mathcal{T}} \max_{\omega^i \in \mathcal{T}} \bigg(\sum_{t \in \mathcal{T}} P^\ast_n(t)C(\omega^i, \lambda^i)\bigg)$$

$$= \sum_{t \in \mathcal{T}} P^\ast_n(t)C(\omega^i, \lambda^i)$$

where $P^\ast_n(t)$ is the type of the parameter sequence $\theta^n$. The second equality follows from the definition of $C^\ast_n(\omega^i, \lambda^i)$ in (18), using the minimax theorem [102] 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 $\mathbb{E}(S_i)$ and $\mathbb{E}(X_i)$, respectively. The last equality holds by the definition of $C^\ast_n(\mathcal{W})$ in (13). $\square$

\section*{APPENDIX E}

\section*{Deterministic Coding Scheme}

We present the deterministic decoding rule and codebook generation for the AVC with fixed parameters and establish key properties that will be used in the main proof of the capacity theorem. Consider the AVC $\mathcal{W}$ with fixed parameters under input constraint $\Omega$ and state constraint $\Lambda$. Let $\theta^n$ be a sequence of fixed parameters for a given blocklength, recall that $T$ is a random variable that is distributed as the type of $\theta^n$, and define the subset $T_n \subseteq T$ of parameter values that appear in the sequence $\theta^n$ at least once, i.e.

$$T_n \triangleq \{ t \in T : P^\ast_n(t) > 0 \}.$$  

To simplify the notation, we will omit the super-script and write $P_T$ instead of $P^\ast_T$.

\subsection*{A. Decoder}

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 $D(m) \subseteq \mathcal{Y}^m \times \mathcal{T}^m$, for $m \in [1:2^nR]$, such that $g(y^n, \theta^n) = m$ iff $(y^n, \theta^n) \in D(m)$.

\textbf{Definition 7 (Decoder):} Given the codebook $\{f(m, \theta^n)\}_{m \in [1:2^nR]}$, declare that $(y^n, \theta^n) \in D(m)$ if there exists $s^n \in S^n$ with $l^n(s^n) \leq \Lambda$ such that the following hold. 1)

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

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

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

$$D(P_{T,\tilde{s}^n,S,Y}|P_{T} \times P_{X|T} \times P_{S|T} \times W_{Y|X,S,T}) \leq \eta.$$
where \((T, \bar{X}, \bar{S}, Y) \sim \hat{P}_{\theta_n, \tilde{m}(\theta_n), \tilde{y}_n}\), we have that
\[
I(X; Y; \bar{X}|S, T) \leq \eta. \tag{146}
\]

We note that in Definition 7, the variables \(T, X, \bar{X}, Y, S, \bar{S}, Y\) are dummy random variables, distributed according to the joint type of \((\theta^n, f(m, \theta^n), f(\tilde{m}, \theta^n), \tilde{s}^n, \tilde{\bar{s}}^n, \tilde{y}^n)\), 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 7 could be positive.

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

**Lemma E.1 (Decoding Disambiguity):** Suppose that in each codebook, all codewords have the same conditional type, i.e., \(\hat{P}_{(m, \theta^n)}(a^n) = p\) for all \(m \in [1:2^n]\). Assume that (23) holds, and assume that for some \(\delta > 0\), \(p(x|t) \geq \delta \forall x \in \mathcal{X}, t \in \mathcal{T}\), and also
\[
\tilde{\Lambda}_n(p) > \Lambda. \tag{147}
\]

Then, for sufficiently small \(\eta > 0\),
\[
\mathcal{D}(m) \cap \mathcal{D}(\tilde{m}) = \emptyset, \text{ for all } m \neq \tilde{m}. \tag{148}
\]

Lemma E.1 is proved below by extending the derivation in [32].

Before we give the proof of Lemma E.1, we give an auxiliary lemma, which resembles Lemma 11 in [91].

**Lemma E.2 (See [32]):** For a given block length \(n\), let \(P_T\) denote the type of the fixed parameter sequence \(\theta^n\), and let \(p(x|t)\) be some conditional input distribution. Then, for every pair of conditional state distributions \(Q(s|x, t)\) and \(Q'(s|x, t)\) such that
\[
\max \left\{ P_T(t)p(x|t)Q(s|x, t)l(s), P_T(t)p(x|t)Q'(s|x, t)l(s) \right\} < \tilde{\Lambda}_n(p), \tag{149}
\]

there exists \(\xi \equiv \xi(n, \theta^n, p, Q, Q') > 0\) such that
\[
\max_{x, \bar{x}, t} \sum_{t \in \mathcal{T}} P_T(t) \left| \sum_{s \in \mathcal{S}} Q(s|\bar{x}, t)W_{Y|X,S,T}(y|x, s, t) - \sum_{s \in \mathcal{S}} Q'(s|\bar{x}, t)W_{Y|X,S,T}(y|\bar{x}, s, t) \right| \geq \xi. \tag{150}
\]

**Proof of Lemma E.2**

Assume to the contrary that the LHS in (150) is zero, and define
\[
Q_A(s|x, t) = \frac{1}{2}(Q(s|x, t) + Q'(s|x, t)). \tag{151}
\]

Using the symmetry between \(Q\) and \(Q'\), and the fact that \(P_T(t) = 0\) for \(t \notin \mathcal{T}_n\), we have that
\[
0 = \max_{x, \bar{x}, \bar{y}, t} \sum_{t \in \mathcal{T}} P_T(t) \left| \sum_{s \in \mathcal{S}} Q(s|x, t)W_{Y|X,S,T}(y|x, s, t) - \sum_{s \in \mathcal{S}} Q'(s|x, t)W_{Y|X,S,T}(y|\bar{x}, s, t) \right| \tag{152}
\]
\[
= \frac{1}{2} \max_{x, \bar{x}, \bar{y}, t} \sum_{t \in \mathcal{T}_n} P_T(t) \left| \sum_{s \in \mathcal{S}} Q(s|x, t)W_{Y|X,S,T}(y|x, s, t) - \sum_{s \in \mathcal{S}} Q'(s|x, t)W_{Y|X,S,T}(y|\bar{x}, s, t) \right| \tag{153}
\]
\[
\geq \max_{x, \bar{x}, \bar{y}, t} \sum_{t \in \mathcal{T}_n} P_T(t) \left| \sum_{s \in \mathcal{S}} Q_A(s|x, t)W_{Y|X,S,T}(y|x, s, t) - \sum_{s \in \mathcal{S}} Q_A(s|x, t)W_{Y|X,S,T}(y|\bar{x}, s, t) \right|. \tag{154}
\]

Since \(P_T(t) > 0\) for all \(t \in \mathcal{T}_n\), it follows that
\[
\sum_{s \in \mathcal{S}} Q_A(s|x, t)W_{Y|X,S,T}(y|\bar{x}, s, t) = \sum_{s \in \mathcal{S}} Q_A(s|x, t)W_{Y|X,S,T}(y|x, s, t), \tag{155}
\]
for all \(t \in \mathcal{T}_n\), \(x, \bar{x} \in \mathcal{X}\) and \(y \in \mathcal{Y}\). In other words, \(Q_A(\cdot|\cdot, t)\) symmetrizes the channel \(W_{Y|X,S,T}(\cdot|\cdot, t)\) for all \(t \in \mathcal{T}_n\). Therefore, by the definition of \(\Lambda_n(p)\) in (25), we have that
\[
\sum_{t \in \mathcal{T}_n} P_T(t)p(x|t)Q_A(s|x, t)l(s) = \sum_{t \in \mathcal{T}_n} P_T(t)p(x|t)Q_A(s|x, t)l(s) \geq \Lambda_n(p) \tag{156}
\]
in contradiction to (149). 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 (150) must be positive. This completes the proof of the auxiliary Lemma.

**Proof of Lemma E.1**

We move to the main part of the proof of the decoding ambiguity as stated in Lemma E.1. To show that (148) 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 \(\{f(m, \theta^n)\}_{m \in [1:2^n]}\) have the same conditional type. In particular, \(P_{\bar{X}|T} = P_{X|T} = p\).
By Condition 1) of the decoding rule, in (144),
\[
D(P_{T,X,S,Y}||P_T \times P_{X|T} \times P_{S|T} \times W_{Y|X,S,T}) = \sum_{t,x,s,y} P_{T,X,S,Y}(t,x,s,y) \cdot \log \frac{P_{T,X,S,Y}(t,x,s,y)}{P_T(t)p(x|t)P_{X|T}(x|s,t)W_{Y|X,S,T}(y|x,s,t)} \leq \eta, \tag{155}
\]
and by Condition 2) of the decoding rule, in (146),
\[
I(X,Y;\tilde{X}|S,T) = \sum_{t,x,\tilde{x},s,y} P_{T,X,\tilde{X},S,Y}(t,x,\tilde{x},s,y) \cdot \log \frac{P_{T,X,\tilde{X},S,Y}(t,x,\tilde{x},s,y)}{P_{\tilde{X}|S,T}(\tilde{x}|s,t)W_{Y|X,S,T}(y|x,s,t)} \leq \eta, \tag{156}
\]
where \( T,X,\tilde{X},S,Y \) are distributed according to the joint type of \( \theta^n, f^n(m,\theta^n), \tilde{m}^n, \), \( s^n \), and \( y^n \). Adding (155) and (156) yields
\[
\sum_{t,x,\tilde{x},s,y} P_{T,X,\tilde{X},S,Y}(t,x,\tilde{x},s,y) \cdot \log \frac{P_{T,X,\tilde{X},S,Y}(t,x,\tilde{x},s,y)}{P_T(t)p(x|t)P_{\tilde{X}|S,T}(\tilde{x}|s,t)W_{Y|X,S,T}(y|x,s,t)} \leq 2\eta. \tag{157}
\]
That is,
\[
D(P_{T,X,\tilde{X},S,Y}||P_T \times p \times p \times P_{\tilde{X}|S,T} \times W_{Y|X,S,T}) \leq 2\eta. \tag{158}
\]
where \( V_{Y|X,\tilde{X},T}(y|x,\tilde{x},t) = \sum_{s \in S} W_{Y|X,S,T}(y|x,s,t) P_{S|\tilde{X},T}(s|\tilde{x},t) \), then, by Pinsker's inequality (see e.g. [29, Theorem 2.7.1]),
\[
\sum_{t,x,\tilde{x},y} P_T(t)P_{\tilde{X},Y|T}(x,\tilde{x},y|t) - p(x|t)p(\tilde{x}|t)V_{Y|X,\tilde{X},T}(y|x,\tilde{x},t) \leq c \sqrt{2\eta}, \tag{159}
\]
where \( c > 0 \) is a constant. By the same arguments, (145) implies that
\[
\sum_{t,x,\tilde{x},y} P_T(t)P_{\tilde{X},Y|T}(x,\tilde{x},y|t) - p(x|t)p(\tilde{x}|t)V'_{Y|X,\tilde{X},T}(y|x,\tilde{x},t) \leq c \sqrt{2\eta}, \tag{160}
\]
where
\[
V'_{Y|X,\tilde{X},T}(y|x,\tilde{x},t) = \sum_{s \in S} W_{Y|X,S,T}(y|x,s,t) P_{S|\tilde{X},T}(s|\tilde{x},t). \tag{161}
\]
Now, since \( p(x|t) > \delta \), for \( x \in X, t \in T_n \), we have that
\[
\max_{x,\tilde{x},y} \sum_{t \in T_n} P_T(t)\left|V_{Y|X,\tilde{X},T}(y|x,\tilde{x},t) - V'_{Y|X,\tilde{X},T}(y|x,\tilde{x},t)\right| \leq \frac{2c\sqrt{2\eta}}{\delta^2}, \tag{162}
\]
Equivalently, the above can be expressed as
\[
\max_{x,\tilde{x},s} \sum_{t \in T_n} P_T(t) \sum_{s \in S} P_{S|\tilde{X},T}(s|\tilde{x},t)W_{Y|X,S,T}(y|x,s,t) - \sum_{s \in S} P_{S|X,T}(s|x,t)W_{Y|X,S,T}(y|x,s,t) \leq \frac{2c\sqrt{2\eta}}{\delta^2}, \tag{163}
\]
Now, we show that the state distributions \( Q = P_{S|\tilde{X},T} \) and \( Q' = P_{S|X,T} \) satisfy the conditions of Lemma E.2. Indeed,
\[
\max_{x,\tilde{x},s} \left\{ \sum_{t \in T_n} P_T(t)p(\tilde{x}|t)Q(s|\tilde{x})l(s) \right\} = \max_{x,\tilde{x},s} \left\{ \sum_{t \in T_n} P_T(t)p(\tilde{x}|t)P_{S|\tilde{X},T}(s|\tilde{x},t)l(s) \right\} \leq \max \left\{ \sum_{s} P_S(s)l(s), \sum_{s} P_{S|X,T}(s|x,t)l(s) \right\} \leq \max \left\{ l^n(s^n), l^n(\tilde{s}^n) \right\} \leq \Lambda < \tilde{\Lambda}_n(p), \tag{164}
\]
where the last inequality is due to (147). Thus, there exists \( \xi > 0 \) such that (150) holds with \( Q = P_{S|\tilde{X},T} \) and \( Q' = P_{S|X,T} \), which contradicts (162), if \( \eta \) is sufficiently small such that \( \frac{2c\sqrt{2\eta}}{\delta^2} < \xi \).

\( \square \)

\[ \text{B. Codebook Generation} \]

We now extend Csiszár and Narayan’s lemma for the codebook generation [32].

**Lemma E.3 (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 \( \{x^n(m, \theta^n)\}_{m \in [1:2^n]} \) of conditional type \( p \), such that for every \( a^n \in X^n \) and \( s^n \in S^n \) 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.

\[
|\{m : (\theta^n, a^n, x^n(m, \theta^n), s^n) \in T^n(P_{T,X,\tilde{X},S})\}| \leq 2^{n \left[ R - I(X;S|T) + \varepsilon \right]}, \tag{165}
\]

and

\[
|\{m : (\theta^n, a^n, x^n(m, \theta^n), s^n) \in T^n(P_{T,X,\tilde{X},S})\}| \leq 2^{n \left[ R - I(\tilde{X};S|T) \right] + \varepsilon}. \tag{166}
\]

The proof is given below.
Proof of Lemma E.3

Let \( Z^n(m, \theta^n), m \in [1 : 2^n R] \), be statistically independent sequences, uniformly distributed over the conditional type class \( T^n(p|\theta^n) \), where \( \theta^n \) is the sequence of fixed parameters. Fix \( a^n \in \mathcal{X}^n \) and \( s^n \in \mathcal{S}^n \), and consider a joint type \( P_{X,\mathcal{X},\mathcal{S}|T} \), such that \( P_{X|T} = P_{\mathcal{X}|T} = P \). We intend to show that \( \{Z^n(m, \theta^n)\} \) satisfy each of the desired properties with double exponential high probability \( (1 - e^{-e^{2^n}}) \). \( E > 0 \), implying that there exists a deterministic codebook that satisfies (164)-(166) simultaneously. We begin with the following large deviations result by Csiszar and Narayan [32].

Lemma E.4 (see [32, Lemma A1]): Let \( \alpha, \beta \in [0, 1] \), and consider a sequence of random vectors \( U^n(m) \), and functions \( \varphi_m : \mathcal{X}^n \rightarrow [0, 1] \), for \( m \in [1 : M] \). If

\[
\begin{align*}
\mathbb{E} \left[ \varphi_m(U^n(1) \ldots U^n(m))|U^n(1) \ldots U^n(m-1)\right] &\leq \alpha \\
&\text{a.s., for } m \in [1 : M],
\end{align*}
\]

then

\[
\Pr \left( \sum_{m=1}^{M} \varphi_m(U^n(1) \ldots U^n(m)) > M\beta \right) \leq \exp\left(-M(\beta - \alpha \log e)\right). \tag{168}
\]

To show that (164) holds, consider the indicator

\[
\varphi_m(Z^n(1, \theta^n), \ldots, Z^n(m, \theta^n)) =
\begin{cases}
1 & \text{if } (a^n, Z^n(m, \theta^n), s^n) \in T^n(P_{X,\mathcal{X},\mathcal{S}|T}|\theta^n) \\
0 & \text{otherwise}.
\end{cases}
\tag{169}
\]

Observe that by the definition of a conditional type class (see Subsection II-A), the condition \( (\theta^n, a^n, x^n(\mathcal{X}, \theta^n), s^n) \in T^n(P_{X,\mathcal{X},\mathcal{S}|T}|\theta^n) \) in (164) can also be written as \( (a^n, x^n(\mathcal{X}, \theta^n), s^n) \in T^n(P_{X,\mathcal{X},\mathcal{S}|T}|\theta^n) \). By standard type class considerations (see e.g. [71, Theorem 1.3]), we have that

\[
\mathbb{E} \left[ \varphi_m(Z^n(1, \theta^n), \ldots, Z^n(m, \theta^n))|Z^n(1, \theta^n), \ldots, Z^n(m-1, \theta^n)\right] \leq 2^{-n(I(\mathcal{X}, \mathcal{S}|T) - \frac{1}{2} - R)} , \tag{170}
\]

Next, we use Lemma E.4, and plug

\[
M = 2^n R, \quad U^n(m) = Z^n(m, \theta^n) ,
\]

\[
\alpha = 2^{-n(I(\mathcal{X}, \mathcal{S}|T) - \frac{1}{2} - R)} , \quad \beta = 2^{-n([R-I(\mathcal{X}, \mathcal{S}|T)]_+ + \varepsilon)} .
\tag{171}
\]

For sufficiently large \( n \), we have that \( M(\beta - \alpha \log e) \geq 2^{n\varepsilon/2} \). Hence, by Lemma E.4,

\[
\Pr \left( \sum_{m=1}^{2^n R} \varphi_m(Z^n(1, \theta^n), \ldots, Z^n(2^n R, \theta^n)) \right) > 2^n([R-I(\mathcal{X}, \mathcal{S}|T)]_+ + \varepsilon) \leq e^{-2^{n\varepsilon/2}} . \tag{172}
\]

The double exponential decay of the probability in (172) implies that there exists a codebook that satisfies (164).

Similarly, to show (165), we replace the indicator of the type \( P_{X,\mathcal{X},\mathcal{S}|T} \) in (169) by an indicator of the type \( P_{X,S|T} \), and rewrite (170) with \( I(\mathcal{X}; S|T) \), to obtain

\[
\Pr \left( \left| \{m : (Z^n(m, \theta^n), s^n) \in T^n(P_{X,S|T}|\theta^n)\} \right| > \epsilon \right) > 2^n([R-I(\mathcal{X}; S|T)]_+ + \varepsilon) \leq e^{-2^{n\varepsilon/2}} , \tag{173}
\]

where \( \varepsilon > 0 \) is arbitrarily small. Observe that the condition \( (\theta^n, x^n(\mathcal{X}, \theta^n), s^n) \in T^n(P_{X,\mathcal{X},\mathcal{S}|T}|\theta^n) \) can also be written as \( (x^n(\mathcal{X}, \theta^n), s^n) \in T^n(P_{X,S|T}|\theta^n) \). If \( I(\mathcal{X}; S|T) > \epsilon \) and \( R \geq \varepsilon \), then choosing \( \varepsilon = \frac{\varepsilon}{2} \), we have that

\[
\left| [R-I(\mathcal{X}; S|T)]_+ + \varepsilon \right| \leq R - \varepsilon , \tag{175}
\]

hence,

\[
\Pr \left( \left| \{m : (Z^n(m, \theta^n), s^n) \in T^n(P_{X,S|T}|\theta^n)\} \right| > 2^n(R - \varepsilon) \right) < e^{-2^{n\varepsilon/4}} . \tag{176}
\]

It remains to show that (166) holds. Assume that

\[
I(\mathcal{X}; \mathcal{X}, S|T) = [R-I(\mathcal{X}; S|T)]_+ > \epsilon . \tag{177}
\]

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

\[
\psi_m(Z^n(1, \theta^n), \ldots, Z^n(m, \theta^n)) =
\begin{cases}
1 & \text{if } (Z^n(\tilde{m}, \theta^n), Z^n(\tilde{m}, \theta^n), s^n) \in T^n(P_{X,\mathcal{X},\mathcal{S}|T}|\theta^n) \\
0 & \text{otherwise}.
\end{cases}
\tag{178}
\]

As before, note that the condition \( (\tilde{\theta}^n, x^n(\tilde{m}, \theta^n), s^n) \in T^n(P_{X,\mathcal{X},\mathcal{S}|T}|\theta^n) \) in (166) can also be written as \( (x^n(\tilde{m}, \theta^n), s^n) \in T^n(P_{X,\mathcal{X},\mathcal{S}|T}|\theta^n) \). Then, choosing \( \varepsilon_1 = \frac{\varepsilon}{2} \) in (174) yields

\[
\Pr \left( \sum_{m=1}^{2^n R} \psi_m(Z^n(1, \theta^n), \ldots, Z^n(m, \theta^n)) \right) \leq \{m : (Z^n(m, \theta^n), s^n) \in T^n(P_{X,\mathcal{X},\mathcal{S}|T}|\theta^n) \text{ for some } \tilde{m} \leq m \} \right) < e^{-2^{n\varepsilon/16}} . \tag{179}
\]

Therefore, instead of bounding the set of messages, it is sufficient to consider the sum \( \sum \psi_m(Z^n(1, \theta^n), \ldots, Z^n(m, \theta^n)) \). Furthermore, by standard type class considerations (see e.g. [71, Theorem 1.3]), we have that

\[
\mathbb{E} \left[ \psi_m(Z^n(1, \theta^n), \ldots, Z^n(m, \theta^n))\right]Z^n(1, \theta^n), \ldots, Z^n(m-1, \theta^n) \leq \epsilon \leq |J_m| \cdot 2^{-n(I(\mathcal{X}, \mathcal{X}, S|T)-\frac{1}{2})} \leq 2^n([R-I(\mathcal{X}, \mathcal{X}, S|T)]_+ - I(\mathcal{X}, \mathcal{X}, S|T)+\frac{1}{2}) < 2^{-3\varepsilon/4},
\tag{180}
\]
where the last inequality is due to (177). Thus, by Lemma E.4,
\begin{equation}
\Pr \left( \sum_{m=1}^{2^{nR}} \psi_m \left( Z^n(1, \theta^n), \ldots, Z^n(m, \theta^n) \right) > 2^{n(R - \frac{\epsilon}{4})} \right) < e^{-2^{n(R - \frac{\epsilon}{4})}} \leq e^{-2^{n+1}},
\end{equation}
(181)
as we have assumed that $R > \epsilon$. Equations (179) and (181) imply that the property in (166) holds with double exponential probability $1 - e^{-2^{n+1}}$, where $E_1 > 0$.

**APPENDIX F**

**PROOF OF THEOREM 4**

**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 $X$, for which $P_{X|T}(x|t) > 0 \forall x \in X, t \in T$, and $E\phi(X) \leq \Omega$, with
\begin{equation}
\tilde{\Lambda}(P_{X|T}) > \Lambda.
\end{equation}
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 $P_T^{(n)}(t) > \delta_0$ for all $t \in T$. Indeed, following our assumption in (23), the asymptotic capacity formula $\liminf C_n(W)$ does not change when we remove parameter values $t \in T$ such that $P_T^{(n)}(t) \rightarrow 0$. 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 E.1 guarantees that the decoder in Definition 7 is well defined. Now, Lemma E.3 assures that there is a codebook \{x^n(m, \theta^n)\} in $[1 : 2^{nR}]$ of conditional type $p$ that satisfies (164)-(166). 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 $(\hat{y}, \theta^n)$ belongs to $D(\hat{m})$, as in Definition 7. If there is none, declare an error. Lemma E.1 guarantees that there cannot be two messages for which this holds.

**Analysis of Probability of Error:** Fix $s^n \in S^n$ with $l^n(s^n) \leq \Lambda$, let $q = P_{S|T}$ denote the conditional type of $s^n$ given $\theta^n$, and let $M$ and $Y^n$ denote the transmitted message and the channel output, respectively. To simplify notation, we denote the type of the parameter sequence $\theta^n$ by $P_T$ instead of $P_T^{(n)}$. Since we are looking at a given parameter sequence $\theta^n$, state sequence $s^n$, and encoder-decoder mappings, the only sources of randomness at this point are the uniform distribution of the message and the channel transition. Thus, all of the probabilistic events that will be considered in the derivation below can be regarded as subsets of $[1 : 2^{nR}] \times \mathbb{Y}^n$.

Consider the error events
\begin{equation}
\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 \}.
\end{equation}
(183)
\begin{equation}
\mathcal{E}_2 = \{ \text{Condition 2 of the decoding rule is violated} \}.
\end{equation}
(184)
and
\begin{equation}
\mathcal{F}_1 = \{ I_q(X; S|T) > \varepsilon \},
\end{equation}
(185)
\begin{equation}
\mathcal{F}_2 = \{ I_q(X; \tilde{X}, S|T) > \left[ R - I(\tilde{X}; S|T) \right] + \varepsilon, \text{ for some } \tilde{m} \neq M \},
\end{equation}
(186)
where $(T, X, \tilde{X}, S, Y)$ are dummy random variables, which are distributed as the joint type of $(\theta^n, x^n(M, \theta^n), x^n(\tilde{m}, \theta^n), s^n, Y^n)$. Notice that this type is random, hence the divergence $D(P_{T,X,S,Y} || P_T \times P_{X|T} \times P_{S|T} \times W_{Y|X,S})$ is a random variable, and $\mathcal{E}_1$ is the event that this random variable is larger than $\eta$. In a similar manner, $\mathcal{F}_1$ and $\mathcal{F}_2$ are probabilistic events in terms of the random variables $I_q(X; S|T)$, $I_q(X; \tilde{X}, S|T)$, and $I(\tilde{X}; S|T)$. As for the decoding error, the event $\mathcal{E}_2$ means that there exists $\tilde{m} \neq M$, with $D(P_{T,X,S,Y} || P_T \times P_{X|T} \times P_{S|T} \times W_{Y|X,S}) \leq \eta$ for some $\tilde{s}^n \in S^n$ where $l^n(\tilde{s}) \leq \Lambda$, such that $I(X, Y; \tilde{X}|S) > \eta$ (see Definition 7). Hence, the events $\mathcal{E}_1$ and $\mathcal{E}_2$ depend on the transmitted message $M$ and on the channel output $Y^n$, while the events $\mathcal{F}_1$ and $\mathcal{F}_2$ depend only on $M$.

By the union of events bound,
\begin{equation}
P_c^{(n)}(\emptyset | 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)
\end{equation}
(187)
where the conditioning on $S^n = s^n$ and $T^n = \theta^n$ is omitted for convenience of notation. Based on Lemma E.3, the probabilities of the events $\mathcal{F}_1$ and $\mathcal{F}_2$ tend to zero as $n \rightarrow \infty$, by (165) and (166), 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{align}
D(P_{T,X,S,Y} || P_T \times P_{X|T} \times P_{S|T} \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{align}
(188)
Then, by standard large deviations considerations (see e.g. [29, pp. 362–364]),
\begin{equation}
\Pr(\mathcal{E}_1 \cap \mathcal{F}_1^c) \leq 2^{-n \min_{P_{T,X,S,Y} : \mathcal{E}_1 \cap \mathcal{F}_1^c \text{ holds}} \left( D(P_{T,X,S,Y} || P_T \times P_{X|T} \times P_{S|T} \times W_{Y|X,S,T}) - \varepsilon \right)}
\end{equation}
\begin{equation}
< 2^{-n(\eta - 2\varepsilon)},
\end{equation}
(189)
where the minimum is over all joint distributions of the random variables $T, X, S, Y$ that satisfy $D(P_{T,X,S,Y} || P_T \times P_{X|T} \times P_{S|T} \times W_{Y|X,S,T}) > \eta$ and $I_q(X; S|T) \leq \varepsilon$. The last bound tends to zero as $n \rightarrow \infty$, for sufficiently small $\varepsilon > 0$, such that $\varepsilon < \frac{1}{16}$.

Moving to Condition 2) of the decoding rule, let $D_2$ denote the set of joint types $P_{T,X,\tilde{X},S}$ such that
\begin{align}
D(P_{T,X,S,Y} || P_T \times P_{X|T} \times P_{S|T} \times W_{Y|X,S,T}) & \leq \eta, \\
D(P_{X,\tilde{X},S,Y} || P_{X,\tilde{X}} \times P_{S|T} \times W_{Y|X,S,T}) & \leq \eta,
\end{align}
(190)
for some $S \sim \tilde{q}(s|t)$,
\begin{equation}
I_q(X, Y; \tilde{X}|S, T) > \eta.
\end{equation}
(192)
Then, by standard type class considerations (see e.g. [71, Theorem 1.3]),
\[
\Pr \left( \mathcal{E}_2 \cap \mathcal{F}_2^c \mid M = m \right) \\
\leq \sum_{P_{T,X,Y,S} \in \mathcal{D}_2 : \mathcal{F}_2^c \text{ holds}} \left| \{ \tilde{m} : (\theta^m, x^n(m, \theta^m), x^n(\tilde{m}, \theta^m), s^n) \in \mathcal{T} \} \right| \cdot 2^{-n(T(\tilde{x};Y|X,S,T) - \varepsilon)},
\]
(193)
for every given \( m \in [1:2^nR] \). We note that \( m \) determines whether the condition \( \mathcal{F}_2^c \) holds with certainty. Hence, if \( m \) satisfies \( \mathcal{F}_2^c \) then \( \Pr(\mathcal{E}_2 \cap \mathcal{F}_2^c \mid M = m) = \Pr(\mathcal{E}_2 \mid M = m) \), otherwise \( \Pr(\mathcal{E}_2 \cap \mathcal{F}_2^c \mid M = m) = 0 \). Hence, by (164),
\[
\Pr(\mathcal{E}_2 \cap \mathcal{F}_2^c) \leq \sum_{P_{T,X,Y,S} \in \mathcal{D}_2 : \mathcal{F}_2^c \text{ holds}} 2^{-n\left(\mathcal{I}_q(\tilde{x};Y|X,S,T) - [R - \mathcal{I}_q(\tilde{x};X,S,T)]_+ - 2\varepsilon\right)},
\]
(194)
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
\[
I_q(\tilde{x};X,S,T) \leq I_q(\tilde{x};X,S|T) \leq \varepsilon.
\]
(195)
By (192), it then follows that
\[
I_q(\tilde{x};Y|X,S,T) = I_q(\tilde{x};X,Y|S,T) - I_q(\tilde{x};X|S,T) \\
\geq \eta - \varepsilon.
\]
(196)
Returning to (194), 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 (194) and (196), we have that \( \Pr(\mathcal{E}_2 \cap \mathcal{F}_2^c) \leq 2^{-n(\eta - 4\varepsilon)} \), which tends to zero as \( n \to \infty \), for \( \varepsilon < \frac{1}{4}\eta \).

Otherwise, if \( R > I_q(\tilde{x};S|T) \), then given \( \mathcal{F}_2^c \),
\[
R > I_q(\tilde{x};X,S,T) + I(\tilde{x};X|S,T) - \varepsilon \\
= I_q(\tilde{x};X,S,T) + I(\tilde{x};X|S,T) - \varepsilon \\
\geq I_q(\tilde{x};X,S|T) - \varepsilon.
\]
(197)
Thus,
\[
\left[ R - I_q(\tilde{x};X,S|T) \right]_+ \leq R - I_q(\tilde{x};X,S|T) + \varepsilon.
\]
(198)
Hence, by (194) we have that
\[
\Pr(\mathcal{E}_2 \cap \mathcal{F}_2^c) \leq \sum_{P_{T,X,Y,S} \in \mathcal{D}_2 : \mathcal{F}_2^c \text{ holds}} 2^{-n\left(\mathcal{I}_q(\tilde{x};Y|X,S,T) - R - 3\varepsilon\right)},
\]
(199)
For \( P_{T,X,Y,S} \in \mathcal{D}_2 \), we have by (191) that \( P_{T,X,Y,S} \) is arbitrarily close to some \( P_{T,X,Y,S}^{\tilde{m}} \), where
\[
P_{T,X,Y,S}^{\tilde{m}}(x,s,y) = P_{T}(t)P_{X|T}(x|t)\tilde{q}(s|t)W_{Y|X,S,T}(y|x,s,t),
\]
(200)
if \( \eta > 0 \) is sufficiently small. In which case,
\[
I_q(\tilde{x};Y|T) \geq I_q(X;Y|T) - \delta,
\]
(201)
where \( \delta > 0 \) is arbitrarily small. Therefore, provided that
\[
R \leq \min_{q(s|t) : \mathbb{E}(\tilde{S}) \leq \lambda} I_q(X;Y|T) - \delta - 5\varepsilon,
\]
(202)
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 \( n \to \infty \).

B. Converse Proof

We will use the following lemma, based on the observations of Ericson [40].

\textbf{Lemma F1:} 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 \( W_{Y|X,S,T}(\cdot, \cdot, \cdot, \cdot) \) are symmetrizable for all \( i \in [1:n] \), and let \( J_i(s|x) \), \( t \in T \), be a set of conditional state distributions that satisfy (22). If \( R > 0 \), then
\[
P^{(n)}_{\varepsilon}(\tilde{q}, \theta^n, \mathcal{C}) = \frac{1}{4^n}.
\]
for \( \tilde{q}(s^n|\theta^n) = \frac{1}{2^nR} \sum_{m=1}^{2^nR} J_m^n(s^n|f(m, \theta^n)) \),
(203)
where \( J_m^n(s^n|x^n) = \prod_{i=1}^{n} J_i(s_i|x_i) \).

For completeness, we give the proof below.

\textbf{Proof of Lemma F1:} Denote the codebook size by \( M = 2^nR \), and the codewords by \( x^n(m, \theta^n) = f(m, \theta^n) \). Under the conditions of the lemma,
\[
P^{(n)}_{\varepsilon}(\tilde{q}, \theta^n, \mathcal{C})
\]
\[
= \sum_{s^n \in \mathcal{S}^n} q(s^n|\theta^n) \frac{1}{M} \sum_{m=1}^{M} \sum_{y^n} W_{Y^n}(y^n|x^n(m, \theta^n), s^n, \theta^n)
\]
\[
= \sum_{s^n \in \mathcal{S}^n} \frac{1}{M} \sum_{m=1}^{M} \sum_{y^n} J_m^n(s^n|x^n(m, \theta^n)) W_{Y^n}(y^n|x^n(m, \theta^n), s^n, \theta^n)
\]
(204)
where have defined \( W_{Y^n} = W_{Y^n|X^n,S^n,T^n} \) for short notation. By switching between the summation indices \( m \) and \( \tilde{m} \), we obtain
\[
P^{(n)}_{\varepsilon}(\tilde{q}, \theta^n, \mathcal{C})
\]
\[
= \frac{1}{2M^2} \sum_{m, \tilde{m}} \sum_{y^n(x^n(\tilde{m}, \theta^n))} W_{Y^n}(y^n|x^n(\tilde{m}, \theta^n), s^n, \theta^n) J_m^n(s^n|x^n(m, \theta^n))
\]
\[
+ \frac{1}{2M^2} \sum_{m, \tilde{m}} \sum_{y^n(x^n(m, \theta^n))} W_{Y^n}(y^n|x^n(\tilde{m}, \theta^n), s^n, \theta^n) J_m^n(s^n|x^n(m, \theta^n)).
\]
(205)
Now, as the channel is memoryless,
\[
\sum_{s^n \in S^n} W^n(y^n | x^n(\tilde{m}, \theta^n), s^n, \theta^n) J_{\theta^n}(s^n | x^n(\tilde{m}, \theta^n)) \\
= \prod_{i=1}^{n} \sum_{s_i \in S} W_{Y_i | X_i, S_i, T_i}(y_i | x_i(\tilde{m}, \theta^n), s_i, \theta_i) \\
\times J_{\theta_i}(s_i | x_i(m, \theta^n)) \\
= \prod_{i=1}^{n} \sum_{s_i \in S} W_{Y_i | X_i, S_i, T_i}(y_i | x_i(m, \theta^n), s_i, \theta_i) \\
\times J_{\theta_i}(s_i | x_i(\tilde{m}, \theta^n)) \\
= \sum_{s^n \in S^n} W^n(y^n | x^n(m, \theta^n), s^n, \theta^n) J_{\theta^n}(s^n | x^n(\tilde{m}, \theta^n)),
\]
(206)

where the second equality is due to (22). Therefore,
\[
P_e^n(q, \theta^n, \epsilon) \\
\geq \frac{1}{2M^2} \sum_{\tilde{m} \neq m} \sum_{s^n \in S^n} \left[ \\
\sum_{y^n : g(y^n, \theta^n) \neq m} W^n(y^n | x^n(m, \theta^n), s^n, \theta^n) J_{\theta^n}(s^n | x^n(\tilde{m}, \theta^n)) + \\
\sum_{y^n : g(y^n, \theta^n) \neq \tilde{m}} W^n(y^n | x^n(m, \theta^n), s^n, \theta^n) \times J_{\theta^n}(s^n | x^n(\tilde{m}, \theta^n)) \right] \\
\geq \frac{1}{2M^2} \sum_{\tilde{m} \neq m} \sum_{s^n \in S^n} \sum_{y^n \in Y^n} W^n(y^n | x^n(m, \theta^n), s^n, \theta^n) \times J_{\theta^n}(s^n | x^n(\tilde{m}, \theta^n)) \\
= \frac{M(M - 1)}{2M^2} = \frac{1}{2} \left(1 - \frac{1}{M}\right).
\]
(207)

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

Now, we are in position to prove the converse part of Theorem 4. Consider a sequence of \((2^mR, n, \alpha_n)\) deterministic codes \(\mathcal{C}_n\) over the AVC with fixed parameters under input constraint \(\Omega\) and state constraint \(\Lambda\), where \(\alpha_n \to 0\) as \(n \to \infty\). In particular, the conditional probability of error given a state sequence \(s^n\) is bounded by
\[
P_e^n(\mathcal{C}_n | s^n, \theta^n) \leq \alpha_n, \quad \text{for } s^n \in S^n \text{ with } l^n(s^n) \leq \Lambda.
\]
(208)

Let \(X^n = I(M, \theta^n)\) 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 \(\bar{S}^n = \prod_{i=1}^{n} q(S_i | \theta_i)\), under the average state constraint \(\frac{1}{n} \sum_{i=1}^{n} E(I(S_i)) \leq \Lambda - \delta\). 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
\[
P_e^n(q, \theta^n, \mathcal{C}_n) \leq \sum_{s^n : l^n(s^n) \leq \Lambda} q^n(s^n | \theta^n) P_e^n(q^n | \mathcal{C}_n) + \Pr(l^n(\bar{S}^n) > \Lambda).
\]
(209)

The first sum is bounded by (208), and the second term vanishes by the law of large numbers, since \(q \in \mathcal{F}_{\Lambda-\delta}(S | \theta^n)\).

It follows that the code sequence of the constrained AVC achieves the same rate \(R\) over the compound channel \(W_{Y | X, S, T}\). As in Appendix A, Fano’s inequality implies that for every jamming strategy \(\mathcal{T}(s^n | \theta^n)\),
\[
R \leq \min_{\mathcal{F}(\mathcal{S}) \subseteq \Lambda} I(X; Y | T) + \epsilon_n,
\]
(210)

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 (117)).

Returning to the original AVC, suppose that \(L_n^* > \Lambda\). It remains to show that \(R > 0\) implies that \(\Lambda_n(P_{X|T}) \geq \Lambda\). If the channels \(W_{Y | X, S, T}(\cdot | s, \theta_i)\) is non-symmetrizable for some \(i \in [1 : n]\), then \(\Lambda_n(P_{X|T}) = +\infty\), and there is nothing to show. Hence, consider the case where \(W_{Y | X, S, T}(\cdot | s, \theta_i)\) are symmetrizable for all \(i \in [1 : n]\). Assume to the contrary that \(R > 0\) and \(\Lambda_n(P_{X|T}) < \Lambda\). Hence, there exist conditional state distributions \(J_{\theta_i}(s | x)\) that symmetrize \(W_{Y | X, S, T}(\cdot | s, \theta_i)\), such that
\[
\bar{\Lambda}_n(P_{X|T}) = \frac{1}{n} \sum_{i=1}^{n} P_{X|T}(x | \theta_i) J_{\theta_i}(s | x) l(s) < \Lambda.
\]
(211)

Now, consider the following jamming strategy. First, the jammer selects a codeword \(X^n\) from the codebook uniformly at random. Then, the jammer selects a sequence \(\bar{S}^n\) at random, according to the conditional distribution
\[
\Pr(\bar{S}^n = s^n \mid \bar{X} = x^n) = J_{\theta^n}(s^n | x^n) \equiv \prod_{i=1}^{n} J_{\theta_i}(s_i | x_i).
\]
(212)

At last, if \(l^n(\bar{S}^n) \leq \Lambda\), the jammer chooses the state sequence to be \(S^n = \bar{S}^n\). 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 \(\Lambda(P_{X|T}) < \Lambda\), we first show that \(E(l^n(S^n)) = \bar{\Lambda}(P_{X|T})\). Observe that for every \(x^n \in X^n\),
\[
E(l^n(S^n) | \bar{X} = x^n) = \frac{1}{n} \sum_{i=1}^{n} l(s) J_{\theta_i}(s | x_i).
\]
(213)

Since \(\bar{X}^n\) is distributed as \(X^n\), we obtain
\[
E(l^n(\bar{S}^n)) = \sum_{s \in S} l(s) \cdot \frac{1}{n} \sum_{i=1}^{n} E J_{\theta_i}(s | x_i)
\]
\[
= \frac{1}{n} \sum_{i=1}^{n} \sum_{s \in S} P_{X|T}(x | \theta_i) J_{\theta_i}(s | x) l(s)
\]
\[
= \bar{\Lambda}_n(P_{X|T})
\]
\[
< \Lambda.
\]
(214)

Thus, by Chebyshev’s inequality we have that for sufficiently large \(n\),
\[
\Pr\left(l^n(\bar{S}^n) > \Lambda\right) \leq \delta_0,
\]
(215)
where \( \delta_0 > 0 \) is arbitrarily small. Now, on the one hand, the probability of error is bounded by
\[
P_e^{(n)}(q, \theta^n, \mathcal{C}_n) \geq \Pr\left( g(Y^n, \theta^n) \neq M, l^n(\tilde{S}^n) \leq \Lambda \right) = \sum_{s^n : l^n(s^n) \leq \Lambda} \tilde{q}(s^n|\theta^n) P_e^{(n)}(\mathcal{C}_n | s^n, \theta^n),
\]
(216)

where \( \tilde{q}(s^n|\theta^n) \) is as defined in (203). 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 F.1,
\[
\frac{1}{4} \leq P_e^{(n)}(q, \theta^n, \mathcal{C}_n) \leq \sum_{s^n : l^n(s^n) \leq \Lambda} \tilde{q}(s^n|\theta^n) P_e^{(n)}(\mathcal{C}_n | s^n, \theta^n) + \Pr\left( l^n(\tilde{S}^n) > \Lambda \right) \leq \sum_{s^n : l^n(s^n) \leq \Lambda} \tilde{q}(s^n|\theta^n) P_e^{(n)}(\mathcal{C}_n | s^n, \theta^n) + \delta_0.
\]
(217)

Thus, by (216)-(217), 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 \( L^*_n(\mathcal{P}_X|\mathcal{T}) \geq \Lambda \).

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

**APPENDIX G**

**PROOF OF COROLLARY 5**

Assume that the AVC \( \mathcal{W} \) with fixed parameters satisfies the conditions of Corollary 5. 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, \( \tilde{L}_n(\mathcal{P}_X|\mathcal{T}) = \Lambda \) implies that \( R = 0 \). Since there is only a polynomial number of types, we can consider \( \mathcal{P}_X|\mathcal{T}(x|t) \) to be the conditional type of \( l(m, \theta^n) \) given \( \theta^n \), for all \( m \in [1 : 2^nR] \) (see [31, Problem 6.19]).

Suppose that \( \tilde{L}_n(\mathcal{P}_X|\mathcal{T}) = \Lambda \), assume to the contrary that \( R > 0 \), and let \( J_i(s|x), i \in [1 : n] \), be the set of distributions that achieves the minimum in (25), i.e.
\[
\tilde{L}_n(p) = \frac{1}{n} \sum_{i=1}^{n} \sum_{x,s} \mathcal{P}_X|\mathcal{T}(x|\theta_i) J_i(s|x) l(s) = \Lambda.
\]
(218)

Based on the condition of the corollary, we may assume that \( J_i(s|x) \) is a 0-1 law, i.e.
\[
J_i(s|x) = \begin{cases} 1 & \text{if } s = G_i(x), \\ 0 & \text{otherwise} \end{cases},
\]
(219)

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 (218),
\[
\mathbb{E}l(G_K(X)) = \frac{1}{n} \sum_{i=1}^{n} \sum_{x,s} p(x|\theta_i) J_i(s|x) l(s) = \Lambda.
\]
(220)

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
\[
l^n(S^n) = \frac{1}{n} \sum_{i=1}^{n} l(G_i(x_i)) = \mathbb{E}l(G_K(X)) = \Lambda,
\]
(221)

where the last equality is due to (220). Thus, the state sequence satisfies the state constraint. Now, observe that the jamming strategy \( S^n = (G_i(\tilde{X}_i))_{i=1}^{n} \) is equivalent to \( S^n \sim \tilde{q}(s^n|\theta^n) \) as in (203). Thus, by Lemma F.1, we have that \( P_e^{(n)}(\tilde{q}, \mathcal{C}_n) \geq \frac{1}{4} \), hence a positive rate cannot be achieved. \( \square \)

**APPENDIX H**

**PROOF OF LEMMA 6**

Suppose that \( L^*_n > \Lambda \). The proof is similar to that of Lemma 2. We begin with the property in the lemma below.

**Lemma H.1:** Let \( \omega^*_i, \lambda^*_i, \tilde{\lambda}^*_i, i \in [1 : n] \), be the parameters that achieve the saddle point in (33), i.e.
\[
R_n(\mathcal{W}) = \frac{1}{n} \sum_{i=1}^{n} C_{i}(\omega^*_i, \lambda^*_i, \tilde{\lambda}^*_i).
\]
(222)

Then, for every \( i, j \in [1 : n] \) such that \( \theta_i = \theta_j \), we have that \( \omega^*_i = \omega^*_j, \lambda^*_i = \lambda^*_j, \) and \( \lambda^*_i = \lambda^*_j \).

**Proof of Lemma H.1:** For every \( i \in [1 : n] \), let \( p_i, q_i \) denote input and state distributions such that \( \mathbb{E}\phi(X_i) \leq \omega^*_i, \mathbb{E}_i(p_i) \geq \lambda^*_i, \mathbb{E}l(S_i) \leq \lambda^*_i \) for \( X_i \sim p_i, S_i \sim q_i \). Now, suppose that \( \theta_i = \theta_j = t \), and define
\[
p^t(x) = \frac{1}{2} [p_i(x) + p_j(x)], \quad q^t(s) = \frac{1}{2} [q_i(s) + q_j(s)].
\]
(223)

Then, \( \mathbb{E}\phi(X') = \frac{1}{2} [\mathbb{E}\phi(X_i) + \mathbb{E}\phi(X_j)], \mathbb{E}_t(p') = \frac{1}{2} [\mathbb{E}_i(p_i) + \mathbb{E}_j(p_j)], \) 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
\[
\frac{1}{2} [I_{q_i} (X_i; Y_i | T_i = t) + I_{q_j} (X_j; Y_j | T_j = t)] \\
\leq I_{q^t} (X'; Y'| T' = t),
\]
\[
\frac{1}{2} [I_{q_i} (X_i'; Y_i'| T' = t) + I_{q_j} (X_j; Y_j'| T = t)] \\
\geq I_{q^t} (X'; Y'| T' = t).
\]
(224)

Therefore, the saddle point distributions must satisfy \( p_i = p_j = p' \) and \( q_i = q_j = q' \), hence \( \omega^*_i = \omega^*_j, \lambda^*_i = \lambda^*_j \) and \( \lambda^*_i = \lambda^*_j \). \( \square \)
Next, it can be inferred from Lemma H.1 that

$$R_n(W) = \min_{(\lambda_i)_{i \in \mathbb{T}}: \sum t_i P_T^n(t) \lambda_i \leq \Lambda} \max_{(\omega_i)_{i \in \mathbb{T}}: \sum t_i P_T^n(t) \omega_i \leq \Omega} \sum_{t \in \mathbb{T}} P_T^n(t) C(t, \omega_t, \lambda_t)$$

where $P_T^n(t)$ is the type of the parameter sequence $\theta^n$. The second equality follows from the definition of $C_t(\omega_t, \lambda_t, \lambda_i)$ in (34), using the minimax theorem [102] to switch between the order of the minimum and maximum. In the third line, we eliminate the slack variables $\lambda_i$, $\omega_i$, and $\lambda_i$, replacing $E(S_i)$, $E(\phi(X_i))$, and $\Lambda(p, \theta_i)$, respectively. The last equality holds by the definition of $C_n(W)$ in (27).

**APPENDIX I
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 (19). Then, for every $t \in \mathbb{T}$,

$$I_q(X; Y|T = t) \geq \frac{1}{2} \log \left( 1 + \frac{t^2 \omega(t)}{\lambda(t) + \sigma^2} \right)$$

where we have denoted $\lambda(t) \equiv E(S^2|T = t)$. The last inequality holds since Gaussian noise is known to be the worst additive noise under variance constraint [37, 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 $W_{Y|X,S,T=1}$ is symmetrized by a conditional pdf $\phi(s|x)$ if

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

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 $\varphi_x(s) \equiv \phi(s|x)$ if

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

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

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

(229)

For every $x \in \mathbb{R}$, define the random variable $S(x) \sim \varphi_x$. We note that the RHS is the convolution of the pdfs of the random variables $Z$ and $S(x)$, while the LHS is the convolution of the pdfs of the random variables $Z$ and $S(0) + tx$. 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 $W_{Y|X,S,T=1}$ if and only if $\varphi_x(s) = \varphi_0(s - tx)$.

(230)

Then, the corresponding state cost satisfies

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X|T}(x|t) \varphi_x(s) s^2 dx ds = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X|T}(x|t) \varphi_0(s - tx)s^2 dx ds$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X|T}(x|t) \varphi_0(a + tx)^2 dx da$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (tx + a)^2 f_{X|T}(x|t) dx \varphi_0(a) da$$

(231)

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

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

(232)

Thus, by (231),

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X|T}(x|t) \varphi_x(s)^2 ds^2 dx ds = t^2 E[X^2|T = t]^2 + \int_{-\infty}^{\infty} a^2 \varphi_0(a) da$$

$$\geq t^2 E[X^2|T = t].$$

(233)

Note that the last inequality holds for any $\varphi_x$ which symmetrizes the channel, and in particular for $\varphi_x(s) = \delta(s - tx)$, where $\delta(\cdot)$ is the Dirac delta function. In addition, since $\varphi_0$ gives probability 1 to $S = 0$, we have that (233) holds with equality for $\varphi_x$, and thus,

$$\tilde{L}_n = \max_{\omega(t) : E(\omega(T)) \leq \Omega} E(T^2 \omega(T)),$$

(235)
Having shown that the minimum in (25) is attained by a 0-1 law, we have by Corollary 5 that the capacity of the fading AVC is \( C(W) = \lim_{n \to \infty} C_n(W) \), with

\[
C_n(W) = \begin{cases} 
\min_{F_{X|T} : \mathbb{E}Z^2 \leq \Lambda} \max_{I_q(X; Y|T)} & \\
\max_{\omega(t)} \mathbb{E}(T^2 \omega(T)) > \Lambda, & \text{if } \omega(t) \in \mathbb{E}(T^2 \omega(T)) \leq \Omega, \\
0 & \text{if } \omega(t) \in \mathbb{E}(T^2 \omega(T)) \leq \Lambda.
\end{cases}
\]

(236)

To show the direct part, we only need to consider the case where \( \max_{\omega(t)} \mathbb{E}(T^2 \omega(T)) > \Lambda \). Then, set the conditional input distribution \( X \sim N(0, \omega(t)) \) given \( T = t \) in (236).

As in the direct part with random codes,

\[
I_q(X; Y|T = t) \geq \frac{1}{2} \log \left( 1 + \frac{t^2 \omega(t)}{N(t) + \sigma^2} \right),
\]

(237)

with \( \lambda(t) \not\in \mathbb{E}(S^2|T = t) \), since Gaussian noise is the worst additive noise under variance constraint [37, Lemma II.2]. The direct part follows. As for the converse part, for the conditional distribution \( S \sim N(0, \lambda(t)) \) given \( T = t \), we have that

\[
I_q(X; Y|T = t) \leq \frac{1}{2} \log \left( 1 + \frac{t^2 \omega(t)}{\lambda(t) + \sigma^2} \right),
\]

(238)

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

**Appendix J**

**Proof of Lemma 9**

**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

\[
\alpha > N_j^* + \sigma_j^2 = \sigma_j^2 \geq \beta.
\]

(239)

Otherwise, \( N_j^* = \beta - \sigma_j^2 > 0 \), thus by the assumption \( P_j^* > 0 \), we have that \( 0 < P_j^* = \alpha - (N_j^* + \sigma_j^2) = \alpha - \beta \). \( \alpha \geq \beta \).

(240)

**Part 2**

Assume to the contrary that \( N_j^* = N_j^* + \sigma_j^2 = \beta \), 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 (65), we have that \( N_j^* + \sigma_j^2 = \max(\alpha, \beta, \sigma_j^2) \) for all \( j \in [1 : d] \). Thus,

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

\[
= \max(\alpha, \beta, \sigma_j^2) + \max(\alpha, \beta, \sigma_j^2),
\]

(241)

where the last equality is due to part 1. Part 4 immediately follows. \( \square \)
hence $\hat{\varphi}_{x^d}$ symmetrizes the channel. In addition, since $\varphi_0$ gives probability 1 to $S^d = 0$, we have that (247) holds with equality for $\varphi_{x^d}$, and thus, $\Lambda(F_{X^d}) = \text{tr}(K_X)$.

\[\square\]

APPENDIX L
PROOF OF THEOREM 11

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

Achievability Proof

Assume that $\Omega > \Lambda$. We show that $C(\Sigma) \geq C(\Sigma) = C^*(\Sigma)$. By [30, Theorem 3], if there exists an input distribution $F_{X^d}$ such that $\Lambda(F_{X^d}) > \Lambda$, then the capacity is given by

$$C(\Sigma) = \max_{F_{X^d}: \sum_{j=1}^{d} p_j \leq \Omega} \min_{N_{j}} \frac{I(X^d; Y^d)}{N_{j} \leq \Lambda},$$

(248)

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 $X^d \sim N(0, K_X)$, where the covariance matrix is given by $K_X = \text{diag}(P_1^*, \ldots, P_d^*)$. By Lemma 10, we have that

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

(249)

Having assumed that $\Omega > \Lambda$, it follows that $\tilde{\Lambda}(F_{X^d}) > \Lambda$, hence (248) applies. Then, setting $X^d \sim N(0, K_X)$ yields

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

(250)

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

(251)

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

(252)

where the second inequality holds as $X_1, \ldots, X_d$ 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 [37, Lemma II.2].

From this point, we use the considerations given in [65]. 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 (252), where $N_j^*$ are as defined in (65)-(66). An assignment of $N_1, \ldots, N_d$ is optimal if and only if it satisfies the KKT optimality conditions [21, Section 5.5.3],

$$\sum_{j=1}^{d} N_j = \Lambda, N_j \geq 0,$$

(253)

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

(254)

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

(255)

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

We claim that the conditions are met by

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

(256)

Condition (253) is met by the definition of $N_j^*$, $j \in [1 : d]$, in (65)-(66). Let $j \in [1 : d]$ be a given channel index. We consider the following cases. Suppose that $N_j^* = 0$. Then, Condition (255) is clearly satisfied. Now, if $P_j^* = 0$, then Condition (254) is satisfied since $\alpha > \beta$ by part 1 of Lemma 9. Otherwise, $0 < P_j^* = \alpha - (N_j^* + \sigma_j^2) = \alpha - \sigma_j^2$, and then

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

(257)

where the last inequality holds since $N_j^* = 0$ only if $\beta \leq \sigma_j^2$. Thus, Condition (254) is satisfied.

Next, suppose that $N_j^* > 0$, hence $N_j^* + \sigma_j^2 = \beta$. By part 2 of Lemma 9, this implies that $P_j^* > 0$, i.e. $P_j^* = \alpha - (N_j^* + \sigma_j^2) = \alpha - \beta$. Thus,

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

(258)

and thus Condition (254) is satisfied with equality, and Condition (255) is satisfied as well.

As the KKT conditions are satisfied under (256), we deduce that the assignment of $N_j = N_j^*$, $j \in [1 : d]$, minimizes the RHS of (252). Together with (252), this implies that $C(\Sigma) \geq C^*(\Sigma)$ for $\Omega > \Lambda$.

Converse Proof

We use a similar technique as in [34] (see also [16], [40]). In general, the deterministic code capacity is bounded by the random code capacity, hence $C(\Sigma) \leq C^*(\Sigma) = C^*(\Sigma)$, by Theorem 8. 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 $C_n = \{f^d, g\}$ 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. $M = 2^{2nR} \geq 2$.

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

$$P_e^{(n)}(F_{S^d}, C') = \frac{1}{M^2} \sum_{m=1}^{M} \sum_{m'=1}^{M} \int_{D_e(m, m')} f_{Z}(z^d)dz^d,$$

(259)

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

$$D_e(m, m') = \{z^d : g(f^d(m) + f^d(m') + z^d) \neq m\}.$$

(260)
By interchanging the summation variables \( m \) and \( m' \), we now have that
\[
P_e(n)(F_{S^d}, \mathcal{C}) = \frac{1}{2M^2} \sum_{m,m'} \int_{\mathcal{D}_c(m,m')} f_{Z_d}(z^d) dz^d
\]
\[
+ \frac{1}{2M^2} \sum_{m,m'} \int_{\mathcal{D}_c(m',m)} f_{Z_d}(z^d) dz^d
\]
\[
\geq \frac{1}{2M^2} \sum_{m':m\neq m'} \int_{\mathcal{D}_c(m',m)} f_{Z_d}(z^d) dz^d. \quad (261)
\]
Next, observe that for \( m \neq m' \), \( \mathcal{D}_c(m,m') \cup \mathcal{D}_c(m',m) = \mathbb{R}^{nd} \), and thus the probability of error is lower bounded by
\[
P_e(n)(F_{S^d}, \mathcal{C}) \geq \frac{M(M-1)}{2M^2} \geq \frac{1}{4}, \quad (262)
\]
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.

**APPENDIX M
PROOF OF THEOREM 12**

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 12 using the Toeplitz matrix properties in the auxiliary lemma below. To derive the deterministic code capacity, we use similar symmetrization and optimization arguments as in our proofs for the Gaussian product channel.

**Lemma M.1:** [38, Section 2.3] (see also [47], [57], [42, Section 8.5]) Let \( \Psi_Z(\omega) \) be the power spectral density of a zero mean stationary process \( \{Z_n\}_{n=1}^{\infty} \). Assume that \( \Psi_Z : [-\pi, \pi] \rightarrow [0, \nu] \) is bounded and integrable, for some \( \nu > 0 \), and denote the auto-correlation function by
\[
r_Z(\ell) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \Psi_Z(\omega) e^{j\omega \ell} d\omega, \quad \ell = 0, 1, 2, \ldots \quad (263)
\]
with \( j = \sqrt{-1} \). For a sequence \( Z \) of length \( n \), let \( \sigma_1^2, \ldots, \sigma_n^2 \) 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 \( G : [0, \nu] \rightarrow [0, \eta] \),
\[
\lim_{n \rightarrow \infty} n \sum_{i=1}^{n} G(\sigma_i^2) = \frac{1}{2\pi} \int_{-\pi}^{\pi} G(\Psi_Z(\omega)) d\omega \quad (264)
\]
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 \( Z \). Consider the eigen decomposition of the covariance matrix \( K_Z \), and denote the eigenvector and eigenvalue matrices by \( Q \) and \( \Sigma \), respectively, i.e.
\[
K_Z = Q \Sigma Q^T, \quad \text{where} \quad QQ^T = I \quad \text{and} \quad \Sigma = \text{diag}\{\sigma_1^2, \ldots, \sigma_n^2\}. \quad (265)
\]
We claim that the capacity of the AVC with colored Gaussian noise is the same as the capacity of the following AVC,
\[
Y' = X' + Z' + S', \quad (266)
\]
where \( X' = QT X, \ Z' = QT Z, \) and \( S' = QT S \). Since \( Q \) is a unitary matrix, i.e. \( Q^{-1} = Q^T \), the input and state constraints remain the same, as \( \|X\|^2 = (X^T X = X^T Q^T Q X = X^T X \leq n\Omega \), and similarly, \( \|S\|^2 = \|S\|^2 \leq n\Lambda \). Furthermore, the noise covariance matrix is now
\[
K_{Z'} = QT K_Q Q = \text{diag}\{\sigma_1^2, \ldots, \sigma_n^2\}. \quad (267)
\]
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, \( \{F_{Y_i} \}_{n=1}^{\infty} \). From an operational perspective, if there exists a \( (2nR, n, \epsilon) \) code \( \mathcal{C}' = (f, g') \) for the original AVC with colored Gaussian noise, then the code \( \mathcal{C}' = (f', g') \), given by \( f'(m) = Q^T f(m) \) and \( g'(y') = g(Qy') \), is a \( (2nR, n, \epsilon) \) code for the transformed AVC in (266). Similarly, if there exists a \( (2nR, n, \epsilon) \) code \( \mathcal{C}' = (f', g') \) for the transformed AVC, then the code \( \mathcal{C} = (f, g) \), given by \( f(m) = Q^T f(m) \) and \( g(y) = g'(Qy) \), is a \( (2nR, n, \epsilon) \) 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 \( \sigma_i^2 \in T \) for \( i \in [1 : n] \), with some set \( T \) 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 (266) is equivalent to a channel \( W_{Y''|X'', S'', T''} \) with fixed parameters, specified by
\[
Y'' = X'' + S'' + Z_i'', \quad \text{where} \quad Z_i'' \sim \mathcal{N}(0, t^2) \quad (268)
\]
with the parameter sequence \( \sigma_1, \sigma_2, \ldots \). It is left to determine the random code capacity and deterministic code capacity of the Gaussian AVC with fixed parameters in (268). 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 [5], [15] [39, Section 3.4.1].

Now, consider the double water filling allocation,
\[
b_i^* = \left[ \beta' - \sigma_i^2 \right]_+, \quad (269)
\]
\[
a_i^* = \left[ \alpha' - (b_i^* + \sigma_i^2) \right]_+, \quad (270)
\]
for \( i \in [1 : n] \), where \( \beta' > 0 \) and \( \alpha' > 0 \) are chosen to satisfy
\[
\sum_{i=1}^{n} [\beta' - \sigma_i^2]_+ = \Lambda \quad \text{and} \quad \sum_{i=1}^{n} [\alpha' - (b_i^* + \sigma_i^2)]_+ = \Omega, \quad \text{respectively. Define}
\]
\[
C_{n}(K_Z) \triangleq \frac{1}{2n} \sum_{i=1}^{n} \log \left( 1 + \frac{a_i^2}{b_i^* + \sigma_i^2} \right). \quad (271)
\]
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 3 that the random code capacity is given by

$$C^*(\Psi_Z) = \liminf_{n \to \infty} \max_{\frac{1}{2} \sum_{i=1}^{N} P_i \leq \Omega} \min_{\frac{1}{2} \sum_{i=1}^{N} N_i \leq \Lambda} \frac{1}{n} \sum_{i=1}^{n} C_8(P_i, N_i),$$  \hspace{1cm} (272)

where $C_8^*(P, N)$ is the random code capacity of the traditional AVC under input constraint $P$ and state constraint $N$. Hughes and Narayan [64] 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

$$C_8^*(P, N) = \frac{1}{2} \log\left(1 + \frac{P}{N + \sigma^2}\right).$$  \hspace{1cm} (273)

Hence, for the AVC with colored Gaussian noise,

$$C^*(\Psi_Z) = \liminf_{n \to \infty} C_n^*(K_Z).$$  \hspace{1cm} (275)

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

$$G(x) = \frac{1}{2} \log\left(1 + \frac{\alpha' - \beta' + x}{|\beta' - x|}\right).$$

Next, observe that this is the same min-max optimization as for the AVGPC in (71), due to [65, with $d \leftarrow n$, $\Omega \leftarrow (n\Omega)$, $\Lambda \leftarrow (n\Lambda)$. Therefore, by Theorem 8 [65] and (274),

$$C^*(\Psi_Z) = \liminf_{n \to \infty} C_n^*(K_Z).$$

and observe that

$$C_n^*(K_Z) = \frac{1}{n} \sum_{i=1}^{n} G(\sigma_i^2).$$  \hspace{1cm} (277)

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

$$\liminf_{n \to \infty} C_n^*(K_Z) = \frac{1}{2\pi} \int_{-\pi}^{\pi} G(\Psi_Z(\omega)) \, d\omega.$$  \hspace{1cm} (278)

Observing that the function defined in (276) is also continuous, while $\Psi_Z(\omega)$ is bounded and integrable, it follows that the integral exists [92, Theorem 6.11]. Plugging (276) into the RHS of (278), we obtain

$$\liminf_{n \to \infty} C_n^*(K_Z) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{2} \log\left(1 + \frac{\alpha - |\beta + \Psi_Z(\omega)|_+}{|\beta - \Psi_Z(\omega)|_+ + \Psi_Z(\omega)}\right) \, d\omega$$  \hspace{1cm} (279)

where $\beta$ and $\alpha$ satisfy (81) and (83), respectively. Since the covariance matrix of the stationary noise process is Toeplitz (see e.g. [47]), the density of eigenvalues on the real line tends to the power spectral density [48]. Given that the power spectral density is bounded and integrable, we have that the sequence of eigenvalues $\sigma_1^2, \sigma_2^2, \ldots$ is summable [47, 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 (275) and (279).

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 $\tilde{A}(F_X, \sigma) = \mathbb{E}X^2$, by Lemma 10, taking $d = 1$. Therefore, for the Gaussian AVC with a parameter sequence $\sigma_1^2, \ldots, \sigma_n^2$,}

$$L_n^* = \min_{F_X : \frac{1}{n} \sum_{i=1}^{n} \sigma_i^2 \leq \Omega} \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[X_i^2 | T = \sigma_i] = \Omega,$$

where the first equality holds by the definition of $L_n^*$ in (26) and by (32). It can further be seen from the proof of Lemma 10 in Appendix K 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 $\tilde{A}(F_X, \sigma)$ in (31) is attained with this distribution.

Therefore, by Corollary 7, the capacity of the AVC with colored Gaussian noise is given by the limit inferior of

$$R_n(W) = \left\{ \begin{array}{ll} \min_{\frac{1}{2} \sum_{i=1}^{N} P_i \leq \Omega} \max_{\frac{1}{2} \sum_{i=1}^{N} N_i \leq \Lambda} \frac{1}{n} \sum_{i=1}^{n} C_8(P_i, \tilde{\lambda}_i, N_i) & \text{if } L_n^* > \Lambda, \\ 0 & \text{if } L_n^* \leq \Lambda \end{array} \right.$$  \hspace{1cm} (281)

where

$$C_8(P, \Delta, N) = \min_{F_{\mathcal{X}|\mathcal{Y}}} \max_{\mathcal{X}^2 \leq \Delta} \frac{1}{n} \sum_{i=1}^{n} I_q(X_i^n; Y_i^n | T = \sigma).$$  \hspace{1cm} (282)

Consider the direct part. Suppose that $\Omega > \Lambda$, hence $L_n^* > \Lambda$ (see (280)), and set $P_i = \lambda_i = \sigma_i^2$ for $i \in [1 : n]$. This choice of parameters satisfies the optimization constraints in (281), as
where the last inequality holds since Gaussian noise is known to be the worst additive noise under variance constraint [37, Lemma II.2]. Next, observe that this is the same minimization as in (252), in the proof of the direct part for the AVCGP, with $d \rightarrow \omega$, $\Omega \rightarrow (n \Omega)$, $\Lambda \rightarrow (n \Lambda)$ (see proof of Theorem 11 in Appendix L). Therefore, the minimum is attained with $N_i = b_i^*$, and the RHS of (275) 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 $\mathcal{C}(\Psi_Z) \leq \mathcal{C}(\Psi_Z) = \mathcal{C}(\Psi_Z)$. If $\Omega \leq \Lambda$, then $L_n^* \leq \Lambda$ by (280), hence $\mathcal{C}(\Psi_Z) = \liminf R_n(\mathcal{W}) = 0$ by the second part of Corollary 7.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their constructive comments and suggestions.

REFERENCES

[1] A. A. Salam, R. Sheriff, S. Al-Araj, K. Mezher, and Q. Nasir, "Novel approach for modeling wireless fading channels using a finite state Markov chain," *ETRI J.*, vol. 39, no. 5, pp. 718–728, Oct. 2017.

[2] A. Ahlswede, I. Althöfer, C. Deppe, and U. Tamm, *Probabilistic Methods and Distributed Information*. Cham, Switzerland: Springer, 2019.

[3] R. Ahlswede, "The weak capacity of averaged channels," *J. Probab. Theory Rel. Areas*, vol. 11, no. 1, pp. 61–73, 1968.

[4] R. Ahlswede, "The capacity of a channel with arbitrarily varying additive Gaussian channel probability functions," in *Proc. 6th Prague Conf. Inf. Theory, Statist. Decis. Functions, Random Processes*, Prague, Czech Republic, Sep. 1971, pp. 1–9.

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

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

[7] R. Ahlswede and N. Cai, "Arbitrarily varying multiple-access channels," Univ. Bielefeld, Bielefeld, Germany, Tech. Rep. SFB 343, 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. Inf. Theory*, vol. 45, no. 2, pp. 742–749, Mar. 1999.

[9] H. Aydinian, F. Cicales and C. Deppe, *Information Theory, Combinatorics, and Search Theory*. Berlin, Germany: Springer, 2013.

[10] S. Barbarossa and A. Scaglione, "On the capacity of linear time-varying channels," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Phoenix, AZ, USA, vol. 5, Mar. 1999, pp. 2627–2630.

[11] E. Biglieri, J. Proakis, and S. Shamai (Shitz), "Fading channels: Information-theoretic and communications aspects," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2619–2692, Oct. 1998.

[12] E. Biglieri, R. Calderbank, A. Constantinescu, A. Goldsmith, A. Paulraj, and H. V. Poor, *MIMO Wireless Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2007.

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

[14] D. Blackwell, L. Breiman, and A. J. Thomasian, "Proof of Shannon’s transmission theorem for finite-state indecomposable channels," *Ann. Math. Statist.*, vol. 29, no. 4, pp. 1209–1220, Dec. 1958.

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

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

[17] I. Boche and R. F. Schaefer, "Capacity results and super-activation for wiretap channels with active wiretappers," *IEEE Trans. Inf. Forensics Security*, vol. 8, no. 9, pp. 1482–1496, Sep. 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. Symp. Inf. Theory*, Honolulu, HI, USA, Jun. 2014, pp. 2376–2380.

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

[20] J. M. Borden, D. M. Mason, and R. J. McEliece, "Some information theoretic saddlepoints," *SIAM J. Control Optim.*, vol. 23, no. 1, pp. 129–143, Jan. 1985.

[21] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[22] R. A. Brualdi and D. Cvetkovic, *A Combinatorial Approach to Matrix Theory and Its Applications*. Boca Raton, FL, USA: CRC Press, 2008.

[23] A. J. Budkuley and S. Jaggi, "Communication over an arbitrarily varying channel under a state-myopic encoder," Apr. 2018, arXiv:1804.10221. [Online]. Available: http://arxiv.org/abs/1804.10221.

[24] A. J. Budkuley and S. Jaggi, "Communication over an arbitrarily varying channel under a state-myopic encoder," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Vail, CO, USA, Jun. 2018, pp. 616–620.

[25] A. J. Budkuley, B. K. Dey, and V. M. Prabhakaran, "Dirty paper arbitrarily varying channel with a state-aware adversary," in *Proc. IEEE Inf. Theory Workshop (ITW)*, Jeju, South Korea, Oct. 2015, pp. 94–98.

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

[27] W. Caire and S. Shamai (Shitz), "On the capacity of some channels with channel state information," *IEEE Trans. Inf. Theory*, vol. 45, no. 6, pp. 2007–2019, Sep. 1999.

[28] R. S. Cheng and S. Verdú, "Gaussian multiaccess channels with ISI: Capacity region and multiuser water-filling," *IEEE Trans. Inf. Theory*, vol. 39, no. 3, pp. 773–785, May 1993.

[29] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Hoboken, NJ, USA: Wiley, 2006.

[30] I. Csiszár, "Arbitrarily varying channels with general alphabets and states," *IEEE Trans. Inf. Theory*, vol. 38, no. 6, pp. 1725–1742, Nov. 1992.

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

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

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

[34] I. Csiszár and P. Narayan, "Capacity of the Gaussian arbitrarily varying channel," *IEEE Trans. Inf. Theory*, vol. 37, no. 1, pp. 18–26, Jan. 1991.

[35] A. Das and P. Narayan, "Capacities of time-varying multiple-access channels with side information," *IEEE Trans. Inf. Theory*, vol. 48, no. 1, pp. 4–25, Jan. 2002.

[36] A. Dembo et al., "Everything is a race and Nakamoto always wins," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.*, 2020, pp. 859–878. [Online]. Available: https://arxiv.org/abs/2005.10484.

[37] S. N. Diggavi and T. M. Cover, "The worst additive noise under a covariance constraint," *IEEE Trans. Inf. Theory*, vol. 47, no. 7, pp. 3072–3081, Nov. 2001.

[38] P. M. Ebert, "Error bounds for parallel communication channels," MIT Res. Lab. Elect., Cambridge, MA, USA, Tech. Rep. 448, 1966.
