entropy

MDPI

Article

List-Decoding Capacity of the Gaussian Arbitrarily-Varying Channel †

Fatemeh Hosseinigoki * and Oliver Kosut

School of Electrical, Computer and Energy Engineering, Arizona State University, Tempe, AZ 85287, USA; okosut@asu.edu

* Correspondence: fhossei1@asu.edu

† This paper is an extended version of our paper published in the 2018 IEEE International Symposium on Information Theory (ISIT), Vail, CO, USA, 17–22 June 2018.

Received: 11 April 2019; Accepted: 4 June 2019; Published: 7 June 2019

Abstract: In this paper, we determine the capacity of the Gaussian arbitrarily-varying channel with a (possibly stochastic) encoder and a deterministic list-decoder under the average probability of error criterion. We assume that both the legitimate and the adversarial signals are restricted by their power constraints. We also assume that there is no path between the adversary and the legitimate user but the adversary knows the legitimate user's code. We show that for any list size L, the capacity is equivalent to the capacity of a point-to-point Gaussian channel with noise variance increased by the adversary power, if the adversary has less power than L times the transmitter power; otherwise, the capacity is zero. In the converse proof, we show that if the adversary has enough power, then the decoder can be confounded by the adversarial superposition of several codewords while satisfying its power constraint with positive probability. The achievability proof benefits from a novel variant of the Csiszár-Narayan method for the arbitrarily-varying channel.

Keywords: gaussian arbitrarily-varying channel; list-decoding; stochastic encoder; capacity

1. Introduction

An arbitrarily-varying channel (AVC) represents a memoryless channel including unknown parameters that are changing arbitrarily from channel use to channel use. Because these parameters (state) can change arbitrarily, we consider this to be a model for an active adversarial jammer. This adversary sends its signal to restrain the legitimate transmitter and receiver from maintaining reliable communication. In wireless channels, these unpleasant adversaries can easily enter channels, so it is of great importance to study the adversary's effect on the channel capacity. The capacity of the AVC depends on the coding method (such as randomized coding, stochastic encoder or deterministic coding), the performance criterion (such as average or maximum error probability) and the amount of adversary's knowledge about the transmitted signal (omniscient, myopic or oblivious adversary). Table 1 provides a summary of various models appearing in the literature, including the one considered here.

Blackwell et al. introduced AVC in Reference [1] and under the average error probability criterion they found the capacity of the discrete memoryless AVC to be given by a min-max expression over the mutual information of input and output. They employed *randomized coding* that is, common randomness between the encoder and the decoder and assumed the jammer to be *oblivious* that is, the jammer does not have any information about the transmitted signal except the code. In Reference [2], it is shown that the min-max expression is equivalent to the corresponding max-min one. Further, in Reference [3], the authors examined that this capacity remains the same even for the maximum error probability criterion, again provided access to common randomness. The case without common

randomness between the transmitter and the receiver is referred to as the deterministic code setting. Ahlswede in a notable paper [4] characterized the deterministic capacity of a discrete AVC under the average probability of error through a dichotomy theorem. He proved that the capacity either corresponds to the AVC randomized code capacity or else it equals zero but he did not state any necessary or sufficient condition for which of the two cases prevails. Ericson, in Reference [5], found the necessary condition for the positive alternative by defining *symmetrizability*. A symmetrizable AVC is an AVC in which the adversary can mimic the transmitted signal in order to prevent the decoder from distinguishing between the true message and an adversarial imitation. Thus, he showed that if the deterministic code capacity of an AVC is positive then the channel should be nonsymmetrizable. Later, in Reference [6], a sufficient condition was provided by Csiszár and Narayen stating that if the AVC is nonsymmetrizable then the deterministic code capacity is not zero. Therefore, considering both conditions, the deterministic code capacity of an AVC is positive if and only if the channel is nonsymmetrizable.

The capacity of discrete AVC is also investigated under input and state (or adversarial signal) constraints. Restricted by peak or average input and state cost constraints, the random code capacity of discrete AVC is studied in Reference [7] using the average probability of error as the performance criterion. Furthermore, the second part of Csiszár and Narayan work in Reference [6] focuses on the deterministic code capacity of AVC under input and state constraints for the same performance criterion. They proved that in this case if the capacity is positive then it is less than or equal to the corresponding random code capacity. In particular, with input and state constraints, the capacity can be positive but strictly less than the random code capacity. Note that this does not occur without cost constraints. Csiszár, in Reference [8], extended this result to general input and output alphabets and state sets rather than only finite alphabets and state sets.

There is a wide variety of research on different versions of AVCs under various adversaries model, including References [9–11]. Sarwate, in [9], considered a myopic adversary in which there is a discrete memoryless channel (DMC) between the legitimate user and the jammer and the jammer chooses its signal based on this noisy version of the user's codeword. He found the capacity by minimizing over all DMCs that the jammer can applied by its worst strategies. In Reference [10], single letter characterizations of capacity is obtained in the presence of a delayed adversary which can observe the transmitted signal after a delay. By assuming randomization at the encoder, the capacity is corresponding to the randomized code capacity. B. K. Dey et al., in Reference [11], obtained upper bounds on the capacity of binary channel in the presence of a causal adversary for both maximal and average error probabilities.

On the other hand, AVCs are also studied in network settings throughout [12–18], such as multiple-access and broadcast channels. The authors in Reference [12–14] investigated the capacity of arbitrarily varying multiple-access channels (AVMACs). In Reference [13], it is proved that symmetrizability, which is defined for the two-user AVC, is a sufficient condition for an AVMAC to have empty interior for its deterministic code capacity region. Moreover, throughout [15–18], the capacity of arbitrarily varying wiretap channels (AVWCs) are determined.

Table 1. References' Summary.

Reference	Discrete/Gaussian	Shared Randomness	Cost Constraints	Max/Avg	List Decoding	Adversarial Model	Capacity or Notes
[1]	Discrete	✓	✗	Avg	✗	Oblivious	Blackwell et al. introduced AVC.
[3]	Discrete	✓	✗	Max and Avg	✗	Oblivious	Capacity remains the same for Max. error Probability.
[4]	Discrete	✗	✗	Max and Avg	✗	Oblivious	Studied deterministic capacity in a dichotomy theorem.
[5]	Discrete	✓	✗	Avg	✗	Oblivious	Found the necessary condition for the positive deterministic capacity by defining symmetrizability.
[6]	Discrete	✗	✓	Avg	✗	Oblivious	$\begin{cases} \max\limits_{\substack{P:g(P)\le\Gamma \\ \Lambda_0(P)\ge\Lambda}} \min\limits_{\substack{Y:P_{YXS}\in\mathscr{C}_0 \text{ for} \\ \text{some } S, \text{ with } P(X)=P}} I(X;Y), & \max\limits_{P:g(P)\le\Gamma} \Lambda_0(P) > \Lambda \\ 0, & \max\limits_{P:g(P)\le\Gamma} \Lambda_0(P) < \Lambda. \end{cases}$
[6]	Discrete	✗	✗	Avg	✗	Oblivious	$\max\limits_{P} \min\limits_{\substack{Y:P_{YXS}\in\mathscr{C}_0 \text{ for} \\ \text{some } S, \text{ with } P(X)=P}} I(X;Y)$ if and only if $C>0$
[7]	Discrete	✓	✓	Avg	✗	Oblivious	$\max\limits_{X:\mathbb{E}g(X)\le\Gamma} \min\limits_{S:\mathbb{E}S\le\Lambda} I(X;Y_{X,S})$
[8]	Discrete	✗	✓	Avg	✗	Oblivious	Capacity with general alphabets and states.
[9]	Discrete	✓	✗	Max	✗	Myopic	$\max\limits_{P(x)} \min\limits_{V\in W} I(P,V)$ W is a set of transition matrices from X to Y
[10]	Discrete	✓	✓	Max	✗	Delay	$\max\limits_{P(x)} \min\limits_{Q\in\mathcal{Q}} I(P,W_Q)$ $\mathcal{Q}$ is a set of transition matrices from X to Y
[11]	Discrete	✗	✓	Max and Avg	✗	Causal	Upper bounds on the capacity
[19]	Discrete	✗	✗	Avg	✓	Oblivious	$\begin{cases} \max\limits_{P} \min\limits_{\substack{P_{YXS}:P_{YXS}\in\mathscr{C}_0 \text{ for} \\ \text{some } S, \text{ with } P(X)=P}} I(X;Y), & L > M \\ 0, & L \le M. \end{cases}$
[20]	Discrete	✓	✓	Max	✓	Oblivious	$\max\limits_{P\in\mathcal{P}(\mathcal{X})} \min\limits_{U\in\mathcal{U}(P,\Lambda)} I(P,\sum W(y,x,s)U(s\|x))$
[20]	Discrete	✗	✓	Avg	✓	Oblivious	Upper and lower bounds on the capacity
[21]	Gaussian	✓	✓	Max	✗	Oblivious	$C\left(\frac{P}{\Lambda+\sigma^2}\right)$
[22]	Gaussian	✗	✓	Avg	✗	Oblivious	$\begin{cases} C\left(\frac{P}{\Lambda+\sigma^2}\right), & P > \Lambda \\ 0, & P \le \Lambda. \end{cases}$
This Paper	Gaussian	✗	✓	Avg	✓	Oblivious	$\begin{cases} C\left(\frac{P}{\Lambda+\sigma^2}\right), & L > \frac{\Lambda}{P} \\ 0, & L < \frac{\Lambda}{P}. \end{cases}$

This paper focuses on the Gaussian AVC (GAVC), wherein all alphabets are continuous rather than discrete. Initially, Ahlswede, in Reference [23], studied the capacity of a GAVC in which the adversary chooses the noise variance rather than an additive signal. Hughes and Narayan in Reference [21] determined the randomized code capacity of GAVC under the peak power input and state constraints. They further extended their result in Reference [24] for a vector GAVC. The deterministic code capacity of the GAVC, for the average probability of error, was found in Reference [22]. The authors showed that if the adversary's power is greater than the legitimate transmitter's power, then symmetrizability occurs causing the capacity to drop to zero. Note that for a discrete AVC with no cost constraint non-symmetrizability makes the deterministic capacity positive and equal to the randomized capacity [6] (Theorem 1). It is further proved in [6] (Theorem 3) that under input and state cost constraints, non-symmetrizability only results in positive deterministic capacity but it is sometimes strictly less than the randomized capacity. In the Gaussian case, even though there are input and state cost constraints, the behavior is like that of a discrete AVC with no cost constraint, in that if the channel is non-symmetrizable, then its deterministic capacity is positive and equal to the randomized capacity [22].

For the first time, in Reference [19], Hughes showed that using list-decoding, in which the decoder can decode to a small (and bounded) list rather than a unique message estimate, causes positive capacity for most symmetrizable discrete-memoryless AVCs. Intuitively, list-decoding combats the symmetrizing attack by allowing the list to contain the true message as well as the counterfeit(s) generated by the adversary; thus, the receiver can successfully decode even if it cannot specify the correct message. Furthermore, the authors in Reference [20] extended the list-decoding result to the discrete-memoryless AVCs with state constraints. They determined upper and lower bounds on the capacity by introducing two notions of symmetrizability for this channel.

The capacity of AVMAC was also studied with list-decoding in References [25,26]. Sirin Nitinawarat in Reference [25] introduced symmetrizability of an AVMAC and showed that the capacity region for deterministic codes with fixed list-size is empty if the list size is less than the symmetrizability Ω. He obtained that the capacity corresponds to the random code capacity if the list size is greater than $(1+\Omega)^2$. H. Boche and R.F. Schaefer in Reference [26] obtained the list capacity region of AVMAC with conferencing encoders which is proved for large list size to be equal to the common randomness assisted capacity region. Moreover, in Reference [27], the authors found the deterministic code and random code capacity regions of arbitrarily varying broadcast channel with receiver side information. By defining a concept of symmetrizability for the channel, they characterized deterministic list codes capacity region as either identical to the random code capacity region or empty.

In this paper, we characterize the capacity of GAVC in Reference [22], using list-decoding for any list size and almost all power values of the transmitter and adversary, a similar result to that of Hughes in Reference [19] which obtained the list capacity for the discrete-memoryless AVC, for which the capacity was determined in Reference [6]. We assume that the encoder may be stochastic—that is, the encoder has access to private randomness—and a deterministic list decoder with constant list size L. Under the average probability of error criterion and without common randomness, we obtain the capacity of GAVC with list decoding to be equal to the corresponding randomized code capacity if the list size is greater than the power ratio of the jammer to the legitimate user; otherwise, the capacity is zero. Generally, our problem is a generalized version of the multiple packing of spherical caps problem in Reference [28] with Gaussian noise; although, they assumed maximal probability of error as the performance criterion. Their upper bound, which is only calculated for the noiseless case, depends on the list size L even in the asymptotic case. However, we only have list size in our symmetrizability conditions rather than the capacity itself.

In our converse proof (in Section 4), the adversary focuses on two possible strategies, one of which is simply sending Gaussian noise which causes the channel to act as a standard Gaussian channel with increased noise variance. The second strategy for the adversary is to transmit the superposition of some random (counterfeit) codewords, which is shown to be possible with positive probability if its

power is large enough. In our achievability proof (in Section 5), we employ Gaussian codewords with a particular version of minimum distance list-decoding based on typicality. We extend the scheme of Reference [22] to show that with high probability a random Gaussian codebook has desirable properties to make the probability of error zero. However, our achievability proof originates from the idea of Reference Reference [6] based on typical sets, rather than the geometric approach of Reference [22]. This scheme allows us for simpler characterizations of codebook constraints. It is worth mentioning that we prove the achievability for the deterministic encoder since it suffices to achieve a rate even by a deterministic code, that is any deterministic code is a realistic value of a stochastic code. Our converse and achievability proofs apply for any list size; our previous work [29] provided proofs only for list size $L = 2$.

Notations: Upper case letters denote random variables while lower case letters specify a deterministic value or the realization of a random variable. Bold letters denotes n-length vectors. We indicate inner product and 2-norm by $\langle \cdot, \cdot \rangle$ and $\| \cdot \|$, respectively. We use $| \cdot |^+$ and $\mathbb{E}[\cdot]$ to denote the positive-part function and the expectation, respectively. Also, for an integer N, notation $[N]$ represents the set $\{1, 2, 3, \ldots, N\}$. Notation $\mathbf{I}_n$ and $\mathbf{1}_n$ stand for the identity matrix of size n and a vector of size n with n ones elements, respectively. For a vector $\mathbf{v}$, we use superscript $\mathbf{v}^T$ to denote its transpose. Both $\log(\cdot)$ and $\exp(\cdot)$ functions has base 2, so we define the Gaussian channel capacity function $C(x) = \frac{1}{2}\log(1 + x)$.

2. Problem Statement

A GAVC is a modified standard point-to-point Gaussian channel in the presence of an additive arbitrarily chosen adversary signal. Both transmitter and receiver do not know anything about the adversary signal and the adversary do not have any information about the transmitted signal except the codebooks. The received signal is given by

$$\mathbf{Y} = \mathbf{x} + \mathbf{s} + \mathbf{V} \tag{1}$$

where the n-length vector $\mathbf{x}$ is the legitimate transmitter's signal, $\mathbf{s}$ represents the independent adversary signal and noise $\mathbf{V}$ is a sequence of n-length i.i.d. zero mean Gaussian random variables with variance σ^2, independent of $\mathbf{x}$ and $\mathbf{s}$.

We have the assumption of peak power constraints for the transmitter and adversary signals respectively as $\|\mathbf{x}\|^2 \leq nP$ and $\|\mathbf{s}\|^2 \leq n\Lambda$. In addition, the transmitter and receiver are assumed to know the three parameters P, Λ and σ^2.

An (n, N, N_r, L) stochastic list code for the GAVC is given by:

- Message set $\mathcal{M} = [N]$ and encoder private randomness set $\mathcal{K} = [N_r]$,
- Stochastic encoding function $\mathbf{x}(M, K) : \mathcal{M} \times \mathcal{K} \to \mathbb{R}^n$,
- List decoding function $\phi(\mathbf{y}) : \mathbb{R}^n \to \mathcal{D}_L = \{\mathcal{L} : \mathcal{L} \subset [N], |\mathcal{L}| \leq L\}$,

where the rate of the code is $R = \frac{1}{n}\log(N/L)$. The transmitter encodes the message M and its private randomness K to $\mathbf{x}(M, K)$ where M and K are chosen uniformly respectively from sets $\mathcal{M}$ and $\mathcal{K}$. At the receiver, signal $\mathbf{Y}$ is decoded by a deterministic function ϕ to the set $\mathcal{D}_L$ which is the set of all subsets of $[N]$ with cardinality at most L. In other words, L is the size of the list decoder.

First, define the probability of error $e(\mathbf{s}, i)$ for a specific message $i \in [N]$ in the presence of a specific adversary signal $\mathbf{s} \in \mathbb{R}^n$ as the probability that $i \notin \phi(\mathbf{y})$. Therefore, the average probability of error for $\mathbf{s}$ is given by

$$\bar{e}(\mathbf{s}) = \frac{1}{N}\sum_{i=1}^{N} e(\mathbf{s}, i). \tag{2}$$

Finally, the overall probability of error $P_e^{(n)}$ is obtained by maximizing over all possible choices of jammers' sequences $\mathbf{s}$ satisfying peak power constraint $\|\mathbf{s}\|^2 \leq n\Lambda$. Suppose r is rate of private

randomness. Given L and r, rate R is *achievable* if there exists a sequence of $\left(n, L2^{nR}, 2^{nr}, L\right)$ codes such that $\lim_{n\to\infty} P_e^{(n)} = 0$. The list-code capacity $\mathscr{C}(L, r)$ is the supremum of all achievable rates given L and r.

3. Main Results

Theorem 1. *The list-code capacity of GAVC is given by*

$$\mathscr{C}(L, r) = \begin{cases} C\left(\frac{P}{\Lambda+\sigma^2}\right), & L > \frac{\Lambda}{P} \\ 0, & L < \frac{\Lambda}{P}. \end{cases} \tag{3}$$

Note that the capacity for $\Lambda = LP$ is unsolved.

Remark 1. *Note that this result holds for all r, including $r = 0$ which corresponds to a deterministic encoder. That is, the capacity does not depend on the amount of private randomness.*

Remark 2. *The condition on the ratio $\frac{\Lambda}{P}$ determines whether it is possible for the adversary to launch a symmetrizing attack, wherein it transmits a superposition of codewords. Since each codeword has power P, the most codewords that the adversary can superpose while obeying its power constraint of Λ is the $\lfloor \frac{\Lambda}{P} \rfloor$. Thus, if the allowable list size is greater than $\frac{\Lambda}{P}$, then even under this attack the decoder can output a list made up of the true message and the superposed codewords selected by the adversary. Of course, the decoder does not know which is which but it can still guarantee that the true message is in the list. Thus, the worst the adversary can do is to act as an extra additive Gaussian noise with variance Λ, so the capacity is equal to the capacity of a standard Gaussian channel with increased noise variance as in $C(\frac{P}{\Lambda+\sigma^2})$. However, if the allowable list size is less than $\frac{\Lambda}{P}$, then there are too many possibilities for the decoder to decode correctly, so the capacity is zero. Note that none of this depends on the channel noise, so σ^2 does not come into play in the condition on L.*

Remark 3. *For the achievability proof, we make no assumptions about what the adversary does. However, for the converse proof, it is allowable to weaken the adversary by making certain assumptions about its behavior, because doing so can only increase the achievable rates. Since the converse is an upper bound on achievable rates, weakening the adversary in this manner still yields a valid upper bound.*

In our proofs in Sections 4 and 5, without loss of generality we restrict ourselves to the transmitter's power $P = 1$ which can be done by scaling the output signal.

4. Converse Proof

Without loss of generality, suppose $P = 1$. For the first case where $\Lambda < L$, we assume that we have a code $(n, L2^{nR}, 2^{nr}, L)$ with vanishing probability of error. Since these codes must function for arbitrary jamming signals, we may derive an outer bound by assuming the adversary transmits Gaussian noise with variance $\Lambda - \gamma$ for any $\gamma > 0$ or **0** if Gaussian realization has power greater than Λ. By the law of large numbers, with high probability the resulting channel is equivalent to a standard Gaussian channel with noise variance $\sigma^2 + \Lambda - \gamma$. Thus, since γ can be chosen arbitrarily small, from the capacity of a non-adversarial Gaussian channel,

$$\mathscr{C}(L, r) \leq C\left(\frac{1}{\sigma^2 + \Lambda}\right). \tag{4}$$

Now, assume the symmetrizable case where $\Lambda > L$. In order to show $\mathscr{C}(L, r) = 0$, first consider a sequence of stochastic codebooks and probability of error $P_e^{(n)}$. We claim that if $R > 0$ and the jammer has the following strategy, then $P_e^{(n)}$ is bounded away from zero for sufficiently large n: The jammer randomly and uniformly chooses L messages $M_1, \ldots, M_L$ from $[L2^{nR}]$ and also L private keys

$K_1, \ldots, K_L$ from $[2^{nr}]$ where M_i and K_i are independent. Note that the jammer knows the transmitted codebook. The jammer then constructs

$$\mathbf{Z} = \mathbf{x}(M_1, K_1) + \ldots + \mathbf{x}(M_L, K_L) - L\boldsymbol{\mu} \tag{5}$$

where $\boldsymbol{\mu} \in \mathbb{R}^n$ is a constant vector that we will choose later. The jammer transmits $\mathbf{S} = \mathbf{Z}$ if $\|\mathbf{Z}\|^2 \leq n\Lambda$ or else transmits $\mathbf{S} = \mathbf{0}$. In the former case, the received signal is

$$\mathbf{Y} = \mathbf{x}(M_0, K_0) + \mathbf{x}(M_1, K_1) + \ldots + \mathbf{x}(M_L, K_L) - L\boldsymbol{\mu} + \mathbf{V} \tag{6}$$

$$= \boldsymbol{\mu} + \sum_{i=0}^{L} (\mathbf{x}(M_i, K_i) - \boldsymbol{\mu}) + \mathbf{V} \tag{7}$$

where M_0 is the true message. If $M_0, M_1, \ldots, M_L$ are all different and for all sets $D \subset \{0, 1, \ldots, L\}$ with $|D| = L$,

$$\left\| \sum_{i \in D} \mathbf{x}(M_i, K_i) - L\boldsymbol{\mu} \right\|^2 \leq n\Lambda, \tag{8}$$

then from the decoder's perspective, any L of the $L+1$ messages might have been forged by the adversary. Therefore, the list decoder with list size at most L has a probability of error at least $\frac{1}{L+1}$; that is, the probability that the decoder chooses L from the $L+1$ messages that does not contain the true message M_0. That is,

$$P_e^{(n)} \geq \frac{1}{L+1} \mathbb{P}\Bigg(\left\| \sum_{i \in D} \mathbf{x}(M_i, K_i) - L\boldsymbol{\mu} \right\|^2 \leq n\Lambda \text{ for all } D \subset \{0, 1, \ldots, L\} \text{ with } |D| = L, \text{ and } M_0, M_1, \ldots, M_L \text{ are distinct} \Bigg) \tag{9}$$

$$\geq \frac{1}{L+1} \left[\mathbb{P}\left(\left\| \sum_{i \in D} \mathbf{X}_i - L\boldsymbol{\mu} \right\|^2 \leq n\Lambda \text{ for all } D \subset \{0, 1, \ldots, L\} \text{ with } |D| = L \right) - \left(1 - \frac{L2^{nR} - 1}{L2^{nR}} \cdot \frac{L2^{nR} - 2}{L2^{nR}} \cdots \frac{L2^{nR} - L}{L2^{nR}} \right) \right] \tag{10}$$

where $\mathbf{X}_i = \mathbf{x}(M_i, K_i)$ and the second term in (10) shows the probability of messages $M_0, \ldots, M_L$ not being distinct which tends to zero as $n \to \infty$. Note that $\mathbf{X}_0, \mathbf{X}_1, \ldots, \mathbf{X}_L$ are independent and each distributed as a transmitted sequence from the code. We proceed to show that there exists a choice of $\boldsymbol{\mu}$ based only on the codebook such that (10) is bounded away from zero for sufficiently large n if $R > 0$.

Let

$$\alpha^\star = \inf \left\{ \alpha : \liminf_{n \to \infty} \max_{\boldsymbol{\mu} \in \mathbb{R}^n} \mathbb{P}(\|\mathbf{X}_0 - \boldsymbol{\mu}\|^2 \leq n\alpha) > 0 \right\}. \tag{11}$$

Note that $\alpha^\star \leq 1$, since by the power constraint we always have $\mathbb{P}(\|\mathbf{X}_0\|^2 \leq n) = 1$. Fix any $\delta > 0$ and let $\alpha = \alpha^\star + \delta/2$. Let

$$\gamma = \liminf_{n \to \infty} \max_{\boldsymbol{\mu} \in \mathbb{R}^n} \mathbb{P}(\|\mathbf{X}_0 - \boldsymbol{\mu}\|^2 \leq n\alpha). \tag{12}$$

Since $\alpha > \alpha^\star$ we have $\gamma > 0$. Thus for n sufficiently large, there exists $\boldsymbol{\mu} \in \mathbb{R}^n$ such that

$$\mathbb{P}(\|\mathbf{X}_0 - \boldsymbol{\mu}\|^2 \leq n\alpha) \geq \gamma - \delta. \tag{13}$$

This $\boldsymbol{\mu}$ is the one to be used by the jammer in (5). Let $\mathcal{B}_n$ be the set of all $\mathbf{x}$ that satisfy $\|\mathbf{x}-\boldsymbol{\mu}\|^2 \leq n\alpha$. Note that $\mathbb{P}(\mathbf{X}_0 \in \mathcal{B}_n) \geq \gamma - \delta$.

Since $\alpha - \delta < \alpha^\star$, by the definition of $\alpha^\star$,

$$\liminf_{n\to\infty} \max_{\boldsymbol{\mu}' \in \mathbb{R}} \mathbb{P}(\|\mathbf{X}_0 - \boldsymbol{\mu}'\|^2 \leq n(\alpha-\delta)) = 0. \tag{14}$$

Specifically, there exists n sufficiently large so that for all $\boldsymbol{\mu}' \in \mathbb{R}^n$,

$$\mathbb{P}(\|\mathbf{X}_0 - \boldsymbol{\mu}'\|^2 \leq n(\alpha-\delta)) \leq \delta. \tag{15}$$

Fix any $\mathbf{x}_1 \in \mathcal{B}_n$ and consider those $\mathbf{x} \in \mathcal{B}_n$ such that

$$\langle \mathbf{x}-\boldsymbol{\mu}, \mathbf{x}_1 - \boldsymbol{\mu}\rangle > n\sqrt{\delta\alpha} \tag{16}$$

which implies

$$\begin{aligned}
\left\|\boldsymbol{\mu} + \sqrt{\delta\alpha^{-1}}(\mathbf{x}_1-\boldsymbol{\mu}) - \mathbf{x}\right\|^2 &= \|\mathbf{x}-\boldsymbol{\mu}\|^2 + \delta\alpha^{-1}\|\mathbf{x}_1-\boldsymbol{\mu}\|^2 - 2\sqrt{\delta\alpha^{-1}}\langle \mathbf{x}-\boldsymbol{\mu}, \mathbf{x}_1-\boldsymbol{\mu}\rangle && (17)\\
&< n\alpha + n\delta - n2\delta && (18)\\
&= n(\alpha-\delta). && (19)
\end{aligned}$$

Thus, we obtain the following by applying (15) with $\boldsymbol{\mu}' = \boldsymbol{\mu} + \sqrt{\delta\alpha^{-1}}(\mathbf{x}_1-\boldsymbol{\mu})$ as

$$\mathbb{P}(\langle \mathbf{X}_0-\boldsymbol{\mu}, \mathbf{X}_1-\boldsymbol{\mu}\rangle > n\sqrt{\delta\alpha},\, \mathbf{X}_0, \mathbf{X}_1 \in \mathcal{B}_n) \leq \max_{\mathbf{x}_1 \in \mathcal{B}_n} \mathbb{P}(\langle \mathbf{X}_0-\boldsymbol{\mu}, \mathbf{x}_1-\boldsymbol{\mu}\rangle > n\sqrt{\delta\alpha},\, \mathbf{X}_0 \in \mathcal{B}_n) \leq \delta. \tag{20}$$

Moreover, if $\mathbf{x}_1, \ldots, \mathbf{x}_L \in \mathcal{B}_n$ satisfy $\langle \mathbf{x}_i-\boldsymbol{\mu}, \mathbf{x}_j-\boldsymbol{\mu}\rangle \leq n\sqrt{\delta\alpha}$ for all $i \neq j \in \{1,\ldots,L\}$, then

$$\begin{aligned}
\|\mathbf{x}_1 + \ldots + \mathbf{x}_L - L\boldsymbol{\mu}\|^2 &= \sum_{i=0}^{L-1} \|\mathbf{x}_i - \boldsymbol{\mu}\|^2 + 2\sum_{i=0}^{L-2}\sum_{j=i+1}^{L-1} \langle \mathbf{x}_i-\boldsymbol{\mu}, \mathbf{x}_j-\boldsymbol{\mu}\rangle && (21)\\
&\leq n\big(L\alpha + L(L-1)\sqrt{\delta\alpha}\big) && (22)\\
&\leq n\big(L + L\delta + L(L-1)\sqrt{\delta\alpha}\big) && (23)\\
&< n\Lambda && (24)
\end{aligned}$$

where (23) holds since $\alpha < \alpha^\star + \delta \leq 1 + \delta$ and (24) holds for sufficiently small δ and by assumption $\Lambda > L$. Now we have

$$\begin{aligned}
&\mathbb{P}\left(\left\|\sum_{i\in D} \mathbf{X}_i - L\boldsymbol{\mu}\right\|^2 \leq n\Lambda \text{ for all distinct set } D \subset \{0,1,\ldots,L\} \text{ with } |D| = L\right)\\
&\geq \mathbb{P}\left(\langle \mathbf{X}_i-\boldsymbol{\mu}, \mathbf{X}_j-\boldsymbol{\mu}\rangle \leq n\sqrt{\delta\alpha} \text{ for all } i,j \in \{0,1,\ldots,L\}, i\neq j, \mathbf{X}_0, \mathbf{X}_1, \ldots, \mathbf{X}_L \in \mathcal{B}_n\right) && (25)\\
&\geq \mathbb{P}(\mathbf{X}_0 \in \mathcal{B}_n)^{(L+1)} - \frac{L(L+1)}{2}\,\mathbb{P}(\langle \mathbf{X}_0-\boldsymbol{\mu}, \mathbf{X}_1-\boldsymbol{\mu}\rangle > n\sqrt{\delta\alpha}, \mathbf{X}_0, \mathbf{X}_1 \in \mathcal{B}_n) && (26)\\
&\geq (\gamma-\delta)^{(L+1)} - \frac{L(L+1)\delta}{2} && (27)
\end{aligned}$$

where (25) follows from the analysis leading to (24), (26) follows from the union bound and the fact that $\mathbf{X}_0, \mathbf{X}_1, \ldots, \mathbf{X}_L$ are independent and (27) follows from the lower bound on the probability of $\mathcal{B}_n$ and (20). For sufficiently small δ, (27) is bounded away from zero, so by (10), $P_e^{(n)}$ is also bounded away from zero for sufficiently large n if $R > 0$.

5. Achievability Proof

Before proceeding to the proof, let define the typical set for Gaussian random variables $X_1, \ldots, X_k$ as:

$$\mathcal{T}_\epsilon^{(n)}(X_1, \ldots, X_k) = \left\{(\mathbf{x}_1, \ldots, \mathbf{x}_k) : \mathbb{E}(X_i X_j) - \epsilon \leq \frac{1}{n}\langle \mathbf{x}_i, \mathbf{x}_j \rangle \leq \mathbb{E}(X_i X_j) + \epsilon \text{ for all } i, j \in [1:k]\right\}. \quad (28)$$

Without loss of generality, assume $P = 1$, $r = 0$ and

$$\Lambda < L, \quad (29)$$

$$R < C\left(\frac{1}{\Lambda + \sigma^2}\right). \quad (30)$$

Note that assuming $r = 0$ makes the code deterministic. Thus, it suffices to achieve the list-code capacity by a $(n, L2^{nR}, L)$ deterministic code construction as follows:

Codebook generation: Fix $\epsilon > \epsilon' > \gamma > 0$. We generate $L2^{nR}$ i.i.d zero mean Gaussian sequences $\mathbf{X}(m)$ with variance $(1-\gamma)$ for each $m \in [L2^{nR}]$.

Encoding: The transmitter sends $\mathbf{X} = \mathbf{X}(m)$ if its power is less than 1, otherwise it sends zero.

Decoding: First, given $\mathbf{y}$, for $1 \leq \ell \leq L$ let set $\mathscr{S}_\ell$ be the set of ℓ-tuple messages $(m_1, \ldots, m_\ell)$ that $(\mathbf{x}(m_1), \ldots, \mathbf{x}(m_\ell), \mathbf{y}) \in \mathcal{T}_\epsilon^{(n)}(X_1, \ldots, X_\ell, Y)$ for any set of zero-mean Gaussian variables $(X_1, \ldots, X_\ell, Y)$ such that

$$\mathrm{Cov}(X_1, \ldots, X_\ell, Y) = \begin{bmatrix} \mathbf{I}_\ell & \mathbf{1}_\ell^T \\ \mathbf{1}_\ell & A \end{bmatrix}_{(\ell+1)\times(\ell+1)} \quad (31)$$

and $1 \leq A \leq 1 + \sigma^2 + \Lambda$.

Now we define the decoding function as

$$\phi(\mathbf{y}) = \underset{\mathcal{L} \in \bigcup_{\ell=1}^{L} \mathscr{S}_\ell}{\arg\min} \left\| \mathbf{y} - \sum_{m \in \mathcal{L}} \mathbf{x}(m) \right\|. \quad (32)$$

where we choose between multiple minimizing $\mathcal{L}$ arbitrarily.

Analysis of the probability of error: To prove that the constructed code is achievable meaning that the probability of error tends to zero as $n \to \infty$, we utilize several lemmas including the following. We provide some necessary codebook properties that hold with high probability in Lemma 1, the proof for which is in the Appendix A.

Lemma 1. *Let $\mathbf{X}(m)$ for $m \in [N]$, $N = L2^{nR}$, be the Gaussian codebook described above. With probability approaching 1 as $n \to \infty$, the codebook satisfies the following, for any $\mathbf{x}, \mathbf{s}$ where $\|\mathbf{s}\|^2 \leq n\Lambda$ and any zero-mean jointly Gaussian random vector $(X, X_1, \ldots, X_\ell, S)$ with positive definite covariance matrices with diagonals at most $(1, 1, \ldots, 1, \Lambda)$ for all $1 \leq \ell \leq L$:*

$$\frac{1}{N}\left|\left\{m : (\mathbf{x}(m), \mathbf{s}) \notin \bigcup_{\substack{(X,S) \text{ independent:} \\ EX^2=1, ES^2 \leq \Lambda}} \mathcal{T}_{\epsilon'}^{(n)}(X, S)\right\}\right| \leq \exp(-n\delta(\epsilon)), \quad (33)$$

$$\left|\left\{m_1 : (\mathbf{x}, \mathbf{x}(m_1), \mathbf{s}) \in \mathcal{T}_\epsilon^{(n)}(X, X_1, S)\right\}\right| \leq \exp\left\{n\left[|R - I(X_1; XS)|^+ + \delta(\epsilon)\right]\right\}, \quad (34)$$

$$\frac{1}{N}\left|\left\{m : (\mathbf{x}(m), \mathbf{x}(m_1), \mathbf{s}) \in \mathcal{T}_\epsilon^{(n)}(X, X_1, S) \text{ for some } m_1 \neq m\right\}\right| \leq 2\exp\{-n\delta(\epsilon)/2\},$$
$$\text{if } I(X; X_1 S) \geq |R - I(X_1; S)|^+ + \delta(\epsilon), \quad (35)$$

$$\left|\left\{(m_1, \ldots, m_\ell) : (\mathbf{x}, \mathbf{x}(m_1), \ldots, \mathbf{x}(m_\ell), \mathbf{s}) \in \mathcal{T}_\epsilon^{(n)}(X, X_1, \ldots, X_\ell, S)\right\}\right| \leq \exp\left\{n\delta(\epsilon)\right\}$$

$$\text{if } R < \min_{k\in\{1,\ldots,\ell\}} I(X_k;S), \tag{36}$$

$$\frac{1}{N}\left|\left\{m : (\mathbf{x}(m),\mathbf{x}(m_1),\ldots,\mathbf{x}(m_\ell),\mathbf{s}) \in \mathcal{T}_\epsilon^{(n)}(X,X_1,\ldots,X_\ell,S) \text{ for some } m_1,\ldots,m_\ell \neq m\right\}\right| \leq \exp\{-n\delta(\epsilon)/2\}, \quad \text{if } I(X;X_1\ldots X_\ell S)\geq\delta(\epsilon) \text{ and } R<\min_{k\in\{1,\ldots,\ell\}} I(X_k;S). \tag{37}$$

The following lemmas can be easily generalized for Gaussian random variables by following the corresponding lemmas in Reference [30] for discrete memoryless random variables.

Conditional Typicality Lemma: Let $(X,Y)\sim f(x,y)$ be jointly Gaussian random variables. Suppose that $\mathbf{x}\in\mathcal{T}_\epsilon^{(n)}(X)$ and $\mathbf{Y}\sim f(\mathbf{y}|\mathbf{x})=\prod_{i=1}^n f_{Y|X}(y_i|x_i)$. Then, for every $\epsilon>\epsilon'$,

$$\lim_{n\to\infty}\mathbb{P}\{(\mathbf{x},\mathbf{Y})\in\mathcal{T}_\epsilon^{(n)}(X,Y)\}=1. \tag{38}$$

Joint Typicality Lemma: Let $(X,Y,Z)\sim f(x,y,z)$ be jointly Gaussian random variables. If $(\tilde{\mathbf{x}},\tilde{\mathbf{y}})$ is a pair of arbitrary sequences and $\tilde{\mathbf{Z}}\sim\prod_{i=1}^n f_{Z|X}(\tilde{z}_i|\tilde{x}_i)$ then there exists $\delta(\epsilon)>0$ that tends to zero as $\epsilon\to 0$ such that

$$\mathbb{P}\{(\tilde{\mathbf{x}},\tilde{\mathbf{y}},\tilde{\mathbf{Z}})\in\mathcal{T}_\epsilon^{(n)}(X,Y,Z)\}\leq\exp(-n(I(Y;Z|X)-\delta(\epsilon))). \tag{39}$$

Assume that the legitimate user transmits message M. Then, the probability of error is upper bounded by the sum of the following L error events probabilities:

$$P_\ell=\mathbb{P}\Bigg\{\|\mathbf{Y}-\mathbf{x}(m_1)-\ldots-\mathbf{x}(m_\ell)\|^2\leq\min_{\substack{\hat{m}_1,\ldots,\hat{m}_\ell:\\(M,\hat{m}_1,\ldots,\hat{m}_\ell)\in\mathscr{S}_{\ell+1}}}\|\mathbf{Y}-\mathbf{x}(M)-\mathbf{x}(\hat{m}_1)-\ldots-\mathbf{x}(\hat{m}_\ell)\|^2 \text{ for some } (m_1,\ldots,m_\ell)\in\mathscr{S}_\ell, m_i\neq M \text{ for all } i\in[\ell]\Bigg\} \text{ for } 1\leq\ell<L, \tag{40}$$

$$P_L=\mathbb{P}\left\{\exists(m_1,\ldots,m_L)\in\mathscr{S}_L : m_\ell\neq M, \text{ for all } \ell\in[L]\right\}. \tag{41}$$

By Lemma 1, we may assume we have a deterministic codebook that satisfies (33)–(37). Consider any state sequence $\mathbf{s}$. By (33), with high probability $(\mathbf{x}(M),\mathbf{s})\in\mathcal{T}_{\epsilon'}^{(n)}(X,S)$ where (X,S) are independent and $\mathbb{E}X^2=1,\mathbb{E}S^2\leq\Lambda$ (33). Thus, by the conditional typicality lemma, for every $\epsilon>\epsilon'$ with high probability $(\mathbf{x}_i,\mathbf{s},\mathbf{V})\in\mathcal{T}_\epsilon^{(n)}(X,S,V)$ where (X,S,V) are mutually independent and $\mathbb{E}V^2=\sigma^2$.

We first bound probability event P_ℓ for $1\leq\ell<L$. Define the shorthand $\vec{X}_\ell=(XX_1\ldots X_\ell SV)$. Let $\mathcal{V}_\ell$ denote a finite set of Gaussian distributions of $\vec{X}_\ell$ that is ϵ-dense in the set of all Gaussian distributions of $\vec{X}_\ell$ with variances at most $(1,1,\ldots,1,\Lambda,\sigma^2)$. Note that the cardinality of $\mathcal{V}_\ell$ does not depend on n. We may upper bound P_ℓ by

$$\sum_{\vec{X}_\ell\in\mathcal{V}_\ell}\frac{1}{N}\sum_{m=1}^N e_{\vec{X}_\ell}(m,\mathbf{s}) \tag{42}$$

where

$$e_{\vec{X}_\ell}(m,\mathbf{s})=\mathbb{P}\Bigg\{(\mathbf{x}(m),\mathbf{x}(m_1),\ldots,\mathbf{x}(m_\ell),\mathbf{s},\mathbf{V})\in\mathcal{T}_\epsilon^{(n)}(\vec{X}_\ell), \|\mathbf{x}(m)+\mathbf{s}+\mathbf{V}-\mathbf{x}(m_1)-\ldots-\mathbf{x}(m_\ell)\|^2\leq\min_{\substack{\hat{m}_1,\ldots,\hat{m}_\ell:\\(m,\hat{m}_1,\ldots,\hat{m}_\ell)\in\mathscr{S}_{\ell+1}}}\|\mathbf{s}+\mathbf{V}-\mathbf{x}(\hat{m}_1)-\ldots-\mathbf{x}(\hat{m}_\ell)\|^2$$

$$\text{for some } (m_1,\ldots,m_\ell)\in\mathscr{S}_\ell \text{ and } m_i\neq m \text{ for all } i\in[\ell]\Big\}. \quad (43)$$

We will show that $\frac{1}{N}\sum_{m=1}^{N} e_{\vec{X}_\ell}(m,\mathbf{s})\to 0$ for all Gaussian vectors $\vec{X}_\ell$ (whether or not they are in $\mathcal{V}_\ell$). We may restrict ourselves to $\vec{X}_\ell$ where

$$(X,S,V) \text{ are mutually independent,} \quad (44)$$
$$\mathbb{E}X^2=\mathbb{E}X_1^2=\ldots=\mathbb{E}X_\ell^2=1,\quad \mathbb{E}V^2=\sigma^2,\quad \mathbb{E}S^2\leq\Lambda \quad (45)$$
$$(X_i, X+S+V-X_i) \text{ are independent for all } i\in[\ell], \quad (46)$$
$$\mathbb{E}(X+S+V-X_i)^2\leq\Lambda+\sigma^2 \text{ for all } i\in[\ell], \quad (47)$$

where (44) holds since the legitimate transmitter, the jammer and the noise are independently generated, (45) follows from the assumptions for $\vec{X}_\ell$, (46) corresponds to $\mathbb{E}X_i(Y-X_i)=0$ following from (31) and the assumption that $(m_1,\ldots,m_\ell)\in\mathscr{S}_\ell$ and (47) is obtained by (31) as follows:

$$\begin{aligned}\mathbb{E}(X+S+V-X_i)^2&=\mathbb{E}(Y-X_i)^2 &(48)\\ &=\mathbb{E}Y^2+\mathbb{E}X_i^2-2\mathbb{E}YX_i &(49)\\ &\leq 1+\sigma^2+\Lambda+1-2 &(50)\\ &=\Lambda+\sigma^2. &(51)\end{aligned}$$

Now, suppose

$$I(XV;SX_1\ldots X_\ell)=0. \quad (52)$$

Then we would have $\mathbb{E}XX_i=0$ for all $i\in[\ell]$. Thus, $(X,X_1,\ldots,X_\ell,S+V-X_1-\ldots-X_\ell)$ are mutually independent since

$$\mathbb{E}X(S+V-X_1-\ldots-X_\ell)=\mathbb{E}X(S+V)=0, \quad (53)$$

and

$$\begin{aligned}\mathbb{E}X_i(S+V-X_1-\ldots-X_\ell)&=\mathbb{E}X_i(Y-X-X_1-\ldots-X_\ell) &(54)\\ &=\mathbb{E}X_i(Y-X_i) &(55)\\ &=0, \text{ for all } i\in[\ell], &(56)\end{aligned}$$

where (53) follows from (44), (55) and (56) follow from (31) and $\mathbb{E}XX_i=0$. Hence, if the message $(m_1,\ldots,m_\ell)$ satisfies the conditions in the probability in (43), then $(m,m_1,\ldots,m_\ell)\in\mathscr{S}_{\ell+1}$. This implies that $(\hat{m}_1,\ldots,\hat{m}_\ell)$ takes on the value $(m_1,\ldots,m_\ell)$ in the minimum, so $\|\mathbf{Y}-\mathbf{x}(m_1)-\ldots-\mathbf{x}(m_\ell)\|^2\leq\|\mathbf{Y}-\mathbf{x}(m_1)-\ldots-\mathbf{x}(m_\ell)-\mathbf{x}(M)\|^2$ and so we must have

$$\mathbb{E}(X+S+V-X_1-\ldots-X_\ell)^2\leq\mathbb{E}(S+V-X_1-\ldots-X_\ell)^2. \quad (57)$$

However, this contradicts the assumptions that X is independent from $S,X_1,\ldots,X_\ell,V$, since

$$\begin{aligned}\mathbb{E}(X+S+V-X_1-\ldots,X_\ell)^2&=\mathbb{E}X^2+\mathbb{E}(S+V-X_1-\ldots,X_\ell)^2 &(58)\\ &=1+\mathbb{E}(S+V-X_1-\ldots,X_\ell)^2. &(59)\end{aligned}$$

Therefore, the assumption in (52) is false and there exists $\eta>0$ such that

$$\eta\leq I(XV;SX_1\ldots X_\ell)=I(XV;X_1\ldots X_\ell|S). \quad (60)$$

Now, consider the following two cases.

Case (a): $R < \min\{I(X_1;S),\ldots,I(X_\ell;S)\}$. By (37), we only need to consider distributions where

$$I(X;X_1\ldots X_\ell S) < \delta(\epsilon). \tag{61}$$

Then for any $m, \mathbf{s}$

$$\begin{aligned} e_{\vec{X}_\ell}(m,\mathbf{s}) &\le \sum_{m_1,\ldots,m_\ell} \mathbb{P}\left\{(\mathbf{x}(m),\mathbf{x}(m_1),\ldots,\mathbf{x}(m_\ell),\mathbf{s},\mathbf{V}) \in \mathcal{T}_\epsilon^{(n)}(X,X_1,\ldots,X_\ell,S,V)\right\} && (62)\\ &\le \exp\left\{-n\left(I(V;X_1\ldots X_\ell|XS)-\delta(\epsilon)\right)\right\} && (63)\\ &= \exp\{-n(I(XV;X_1\ldots X_\ell|S)-I(X;X_1\ldots X_\ell|S)-\delta(\epsilon))\} && (64)\\ &\le \exp\{-n(\eta-2\delta(\epsilon))\}. && (65)\end{aligned}$$

where in (62) the sum is over all $m_1,\ldots,m_\ell : (\mathbf{x}(m),\mathbf{x}(m_1),\ldots,\mathbf{x}(m_\ell),\mathbf{s}) \in \mathcal{T}_\epsilon^{(n)}(X,X_1,\ldots,X_\ell,S)$, in (63) we use (36), the joint typicality lemma and $I(V;XS)=0$ and (65) follows from (60) and (61). Thus, (65) tends to zero exponentially fast for sufficiently small $\delta(\epsilon)$.

Case (b): $R \ge \min\{I(X_1;S),\ldots,I(X_\ell;S)\}$. We may assume without loss of generality that $R \ge I(X_1;S)$. Now, we upper bound (43) by

$$e_{\vec{X}_\ell}(m,\mathbf{s}) \le \sum_{\hat{m}:(\mathbf{x}(m),\mathbf{x}(\hat{m}),\mathbf{s})\in\mathcal{T}_\epsilon^{(n)}(X,X_1,S)} \mathbb{P}\left\{(\mathbf{x}(m),\mathbf{x}(\hat{m}),\mathbf{s},\mathbf{V}) \in \mathcal{T}_\epsilon^{(n)}(X,X_1,S,V)\right\}. \tag{66}$$

Note that by (35), we may narrow the distributions to those with

$$I(X;X_1S) < R - I(X_1;S) + \delta(\epsilon). \tag{67}$$

Therefore,

$$\begin{aligned} e_{\vec{X}_\ell}(m,\mathbf{s}) &\le \exp\left\{n\left[|R-I(X_1;XS)|^+ - I(V;X_1|XS)+2\delta(\epsilon)\right]\right. && (68)\\ &\le \exp\left(n\left[R-I(X_1;XS)-I(V;X_1|XS)+2\delta(\epsilon)\right]\right) && (69)\\ &= \exp(n[R-I(X_1;XSV)+2\delta(\epsilon)]) && (70)\end{aligned}$$

where (68) follows from (34) and the joint typicality lemma, (69) follows since by $R \ge I(X_1;S)$ and (67) we have

$$\begin{aligned} R &> I(X;X_1S)+I(X_1;S)-\delta(\epsilon) && (71)\\ &= I(X;X_1|S)+I(X_1;S)-\delta(\epsilon) && (72)\\ &= I(X_1;XS)-\delta(\epsilon). && (73)\end{aligned}$$

Let $Z = X+S+V-X_1$. From (46)–(47), we get

$$\begin{aligned} I(X_1;XSV) &\ge I(X_1;X_1+Z) && (74)\\ &= C\left(\frac{1}{\mathbb{E}Z^2}\right) && (75)\\ &\ge C\left(\frac{1}{\Lambda+\sigma^2}\right) && (76)\end{aligned}$$

Using this result in (70), we obtain

$$e_{\vec{X}_\ell}(m,\mathbf{s}) \le \exp\left\{n\left[R - C\left(\frac{1}{\Lambda+\sigma^2}\right)+2\delta(\epsilon)\right]\right\} \tag{77}$$

meaning that $e_{\vec{X}_\ell}(m, \mathbf{s})$ is exponentially vanishing if $\delta(\epsilon)$ is sufficiently small and the rate condition in (30) holds.

Now, we consider error probability P_L. Define $\vec{X}_L = (XX_1X_2 \dots X_L SV)$. Let $\mathcal{V}_L$ denote a finite ϵ-dense subset of Gaussian vectors $\vec{X}_L$ with variances at most $1, 1, 1, \dots, 1, \Lambda, \sigma^2$. Thus, P_L can be upper bounded by

$$\sum_{\vec{X}_L \in \mathcal{V}_L} \frac{1}{N} \sum_{m=1}^{N} e_{\vec{X}_L}(m, \mathbf{s}) \tag{78}$$

where

$$e_{\vec{X}_L}(m, \mathbf{s}) = \mathbb{P}\Big\{ (\mathbf{x}(m), \mathbf{x}(m_1), \dots, \mathbf{x}(m_L), \mathbf{s}, \mathbf{V}) \in \mathcal{T}_\epsilon^{(n)}(\vec{X}_L), \\ \text{for some } (m_1, \dots, m_L) \in \mathscr{S}_L \text{ and } m_\ell \neq m \text{ for all } \ell \in [L] \Big\}. \tag{79}$$

Thus, we need to show that $\frac{1}{N}\sum_{m=1}^{N} e_{\vec{X}_L}(m, \mathbf{s})$ vanishes as $n \to \infty$ for all Gaussian vectors $\vec{X}_L$ that satisfy (44)–(47) for all $\ell \in [L]$ and

$$\mathbb{E}X_L^2 = 1, \quad (X_1, \dots, X_L) \text{ are independent,} \tag{80}$$

where (80) follows from (31) in which $\mathrm{Cov}(X_i X_j) = 0$ for all $i \neq j \in [L]$.

Observe that if $I(XV; SX_1 \dots X_L) = 0$, then we would have for all $\ell \in [L]$

$$0 = \mathbb{E}X_\ell(X + S + V - X_\ell) \tag{81}$$
$$= \mathbb{E}X_\ell(S - X_\ell) \tag{82}$$
$$= \mathbb{E}X_\ell S - 1, \tag{83}$$

where (83) follows from (31).

Since $\mathbb{E}X_i X_j = 0$ for all $i, j \in [L]$ and $i \neq j$, the covariance matrix of $(S, X_1, \dots, X_L)$ is equal to

$$\begin{bmatrix} \mathbb{E}S^2 & \mathbf{1}_L^T \\ \mathbf{1}_L & \mathbf{I}_L \end{bmatrix} \tag{84}$$

which has the determinant of $\mathbb{E}S^2 - L$. This determinant should be positive since the covariance matrix $\mathrm{Cov}(S, X_1, \dots, X_L)$ is positive definite. However, since $\mathbb{E}S^2 \leq \Lambda$, this assumption contradicts the assumption that $\Lambda < L$ in (29). Thus, there exists $\eta > 0$ such that

$$\eta \leq I(XV; SX_1 \dots X_L) = I(XV; X_1 \dots X_L | S) \tag{85}$$

where we have used the fact that $I(XV; S) = 0$.

Now, we may consider two cases $R < \min\{I(X_1; S), \dots, I(X_L; S)\}$ and $R \geq \min\{I(X_1; S), \dots, I(X_L; S)\}$. Therefore, using an identical argument as in the cases (a) and (b) for P_ℓ, it follows that $e_{\vec{X}_L}$ is also exponentially vanishing.

Author Contributions: Conceptualization, F.H. and O.K.; methodology, F.H. and O.K.; formal analysis, F.H. and O.K.; investigation, F.H.; resources, O.K.; writing—original draft preparation, F.H.; writing—review and editing, O.K.; supervision, O.K.; project administration, O.K.; funding acquisition, O.K.

Funding: This research was funded by National Science Foundation under grant number CCF-1453718.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Appendix A. Proof of Lemma 1

In order to prove (33), we use our proof in Reference [31] (Lemma 6) for one codebook. Moreover, to obtain (34)–(37), we apply the corresponding proof of these four equations in Reference [19] (Lemma 1) for Gaussian distributions. Note that Reference [19] focuses on discrete alphabets but the same proofs can be extended to Gaussian distributions by quantization of the set of continuous random variables in the following way.

Let $\mathbf{X}_i$ be Gaussian i.i.d. n-length random vector (codebook) independent from each other with $\mathrm{Var}(X) = 1$. Fix $\mathbf{x} \in \mathcal{T}_\epsilon^{(n)}(X), \mathbf{s} \in \mathscr{S}^n$ and a covariance matrix $\mathrm{Cov}(X, X_1, \ldots, X_\ell, S) \in \mathcal{V}^{(\ell+2)\times(\ell+2)}$ such that $\mathscr{S}^n$ is a ν-dense subset of $\mathbb{R}^n$ for $\mathbf{s}$ such that $||\mathbf{s}||^2 \leq n\Lambda$ and $\mathcal{V}^{(\ell+2)\times(\ell+2)}$ is a ν-dense subset of $\mathbb{R}^{(\ell+2)\times(\ell+2)}$ for positive definite covariance matrices with diagonals at most $(1, 1, \ldots, 1, \Lambda)$.

Using the similar proof in Reference [19] (Lemma 1), we obtain for given $\mathbf{x}, \mathbf{s}$ and covariance matrix $\mathrm{Cov}(X, X_1, \ldots, X_\ell, S)$ that the complement of each event in (34)–(37) happens with decreasingly doubly exponential probability for sufficiently large n meaning that

$$\begin{aligned}
&\mathbb{P}\big\{\big|\{m_1 : (\mathbf{x}, \mathbf{x}(m_1), \mathbf{s}) \in \mathcal{T}_\epsilon^{(n)}(X, X_1, S)\}\big| \leq \exp\big\{n\big[|R - I(X_1; XS)|^+ + \delta(\epsilon)\big]\big\}\big\} \\
&\quad < \exp(-\exp(n\sigma(\epsilon))), \qquad \text{(A1)} \\
&\mathbb{P}\big\{\frac{1}{N}\big|\{m : (\mathbf{x}(m), \mathbf{x}(m_1), \mathbf{s}) \in \mathcal{T}_\epsilon^{(n)}(X, X_1, S) \text{ for some } m_1 \neq m\}\big| \leq 2\exp\{-n\delta(\epsilon)/2\}\big\} \\
&\quad < \exp(-\exp(n\sigma(\epsilon))), \text{ if } I(X; X_1 S) \geq |R - I(X_1; S)|^+ + \delta(\epsilon), \qquad \text{(A2)} \\
&\mathbb{P}\big\{\Big|\Big\{(m_1, \ldots, m_\ell) : (\mathbf{x}, \mathbf{x}(m_1), \ldots, \mathbf{x}(m_\ell), \mathbf{s}) \in \mathcal{T}_\epsilon^{(n)}(X, X_1, \ldots, X_\ell, S)\Big\}\Big| \leq \exp\{n\delta(\epsilon)\}\big\} \\
&\quad < \exp(-\exp(n\sigma(\epsilon))) \text{ if } R < \min_{k\in\{1,\ldots,\ell\}} I(X_k; S), \qquad \text{(A3)} \\
&\mathbb{P}\big\{\frac{1}{N}\big|\{m : (\mathbf{x}(m), \mathbf{x}(m_1), \ldots, \mathbf{x}(m_\ell), \mathbf{s}) \in \mathcal{T}_\epsilon^{(n)}(X, X_1, \ldots, X_\ell, S) \text{ for some } m_1, \ldots, m_\ell \neq m\}\big| \\
&\leq \exp\{-n\delta(\epsilon)/2\}\big\} < \exp(-\exp(n\sigma(\epsilon))) \text{if } I(X; X_1 \ldots X_\ell S) \geq \delta(\epsilon) \text{ and } R < \min_{k\in\{1,\ldots,\ell\}} I(X_k; S). \qquad \text{(A4)}
\end{aligned}$$

Then, in order to complete the proof, since for any fixed ν the cardinality of finite set $\mathscr{S}^n$ is only increasingly exponentially in n and the set $\mathcal{V}^{(\ell+2)\times(\ell+2)}$ is finite along with the doubly decreasing exponential probabilities in (A1)–(A4), we derive that with probability approaching to 1, all inequalities in (34)–(37) hold simultaneously for sufficiently large n. Since these inequalities hold for every element in the finite sets $\mathscr{S}^n$ and $\mathcal{V}^{(\ell+2)\times(\ell+2)}$, then for any vector $\mathbf{s}, \mathbf{x}$ and any given covariance matrix $\mathrm{Cov}(X, X_1, \ldots, X_\ell, S)$ (with $\|\mathbf{x}\|^2 = n, \|\mathbf{s}\|^2 \leq n\Lambda$) which is not in its corresponding ν-dense subset, there exists a point in the corresponding ν-dense subset that is close enough to it (in its ν distance neighborhood). Now, by using the continuity properties of all sets, we may conclude that (34)–(37) hold also for any point which is not in the ν-dense subset.

References

1. Blackwell, D.; Breiman, L.; Thomasian, A.J. The Capacities of Certain Channel Classes under Random Coding. *Ann. Math. Statist.* **1960**, *31*, 558–567. [CrossRef]
2. Stiglitz, I. Coding for a class of unknown channels. *IEEE Trans. Inf. Theory* **1966**, *12*, 189–195. [CrossRef]
3. Ahlswede, R.; Wolfowitz, J. Correlated decoding for channels with arbitrarily varying channel probability functions. *Inf. Control* **1969**, *14*, 457–473. [CrossRef]
4. Ahlswede, R. Elimination of correlation in random codes for arbitrarily varying channels. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* **1978**, *44*, 159–175. [CrossRef]
5. Ericson, T. Exponential error bounds for random codes in the arbitrarily varying channel. *IEEE Trans. Inf. Theory* **1985**, *31*, 42–48. [CrossRef]
6. Csiszar, I.; Narayan, P. The capacity of the arbitrarily varying channel revisited: Positivity, constraints. *IEEE Trans. Inf. Theory* **1988**, *34*, 181–193. [CrossRef]

7. Csiszar, I.; Narayan, P. Arbitrarily varying channels with constrained inputs and states. *IEEE Trans. Inf. Theory* **1988**, *34*, 27–34. [CrossRef]
8. Csiszar, I. Arbitrarily varying channels with general alphabets and states. *IEEE Trans. Inf. Theory* **1992**, *38*, 1725–1742. [CrossRef]
9. Sarwate, A.D. Coding against myopic adversaries. In Proceedings of the 2010 IEEE Information Theory Workshop, Cairo, Egypt, 6–8 January 2010; pp. 1–5. [CrossRef]
10. Dey, B.K.; Jaggi, S.; Langberg, M.; Sarwate, A.D. Coding against delayed adversaries. In Proceedings of the 2010 IEEE International Symposium on Information Theory, Austin, TX, USA, 12–18 June 2010; pp. 285–289. [CrossRef]
11. Dey, B.K.; Jaggi, S.; Langberg, M.; Sarwate, A.D. Upper Bounds on the Capacity of Binary Channels with Causal Adversaries. *IEEE Trans. Inf. Theory* **2013**, *59*, 3753–3763. [CrossRef]
12. Jahn, J. Coding of arbitrarily varying multiuser channels. *IEEE Trans. Inf. Theory* **1981**, *27*, 212–226. [CrossRef]
13. Gubner, J.A. On the deterministic-code capacity of the multiple-access arbitrarily varying channel. *IEEE Trans. Inf. Theory* **1990**, *36*, 262–275. [CrossRef]
14. Ahlswede, R.; Cai, N. Arbitrarily varying multiple-access channels. I. Ericson's symmetrizability is adequate, Gubner's conjecture is true. *IEEE Trans. Inf. Theory* **1999**, *45*, 742–749. [CrossRef]
15. MolavianJazi, E. Secure Communications over Arbitrarily Varying Wiretap Channels. Ph.D. Thesis, University of Notre Dame, Notre Dame, Indiana, 2009.
16. Bjelaković, I.; Boche, H.; Sommerfeld, J. Strong secrecy in arbitrarily varying wiretap channels. In Proceedings of the 2012 IEEE Information Theory Workshop, Lausanne, Switzerland, 3–7 September 2012; pp. 617–621. [CrossRef]
17. Nötzel, J.; Wiese, M.; Boche, H. The Arbitrarily Varying Wiretap Channel—Secret Randomness, Stability, and Super-Activation. *IEEE Trans. Inf. Theory* **2016**, *62*, 3504–3531. [CrossRef]
18. Goldfeld, Z.; Cuff, P.; Permuter, H.H. Arbitrarily Varying Wiretap Channels with Type Constrained States. *IEEE Trans. Inf. Theory* **2016**, *62*, 7216–7244. [CrossRef]
19. Hughes, B.L. The smallest list for the arbitrarily varying channel. *IEEE Trans. Inf. Theory* **1997**, *43*, 803–815. [CrossRef]
20. Sarwate, A.D.; Gastpar, M. List-Decoding for the Arbitrarily Varying Channel under State Constraints. *IEEE Trans. Inf. Theory* **2012**, *58*, 1372–1384. [CrossRef]
21. Hughes, B.; Narayan, P. Gaussian arbitrarily varying channels. *IEEE Trans. Inf. Theory* **1987**, *33*, 267–284. [CrossRef]
22. Csiszar, I.; Narayan, P. Capacity of the Gaussian arbitrarily varying channel. *IEEE Trans. Inf. Theory* **1991**, *37*, 18–26. [CrossRef]
23. Ahlswede, R. The capacity of a channel with arbitrarily varying additive Gaussian channel probability functions. In Proceedings of the Transactions of the Sixth Prague Conference on Information Theory, Statistical Decision Functions, Random Processes, Prague, Czech Republic, 19–25 September 1971; pp. 13–21.
24. Hughes, B.; Narayan, P. The capacity of a vector Gaussian arbitrarily varying channel. *IEEE Trans. Inf. Theory* **1988**, *34*, 995–1003. [CrossRef]
25. Nitinawarat, S. On the Deterministic Code Capacity Region of an Arbitrarily Varying Multiple-Access Channel under List Decoding. *IEEE Trans. Inf. Theory* **2013**, *59*, 2683–2693. [CrossRef]
26. Boche, H.; Schaefer, R.F. List decoding for arbitrarily varying multiple access channels with conferencing encoders. In Proceedings of the 2014 IEEE International Conference on Communications (ICC), Sydney, Australia, 10–14 June 2014; pp. 1934–1940. [CrossRef]
27. Schaefer, R.F.; Boche, H. List Decoding for Arbitrarily Varying Broadcast Channels with Receiver Side Information. *IEEE Trans. Inf. Theory* **2014**, *60*, 4472–4487. [CrossRef]
28. Blachman, N.M.; Few, L. Multiple packing of spherical caps. *Mathematika* **1963**, *10*, 84–88. [CrossRef]
29. Hosseinigoki, F.; Kosut, O. Capacity of the Gaussian Arbitrarily-Varying Channel with List Decoding. In Proceedings of the 2018 IEEE International Symposium on Information Theory (ISIT), Vail, CO, USA, 17–22 June 2018; pp. 471–475. [CrossRef]

30. El Gamal, A.; Kim, Y.H. *Network Information Theory*; Cambridge University Press: Cambridge, MA, USA, 2011.
31. Hosseinigoki, F.; Kosut, O. The Gaussian Interference Channel in the Presence of Malicious Jammers. *arXiv* **2017**, arXiv:1712.04133.