A Structured Coding Framework for Communication and Computation Over Continuous Networks

Farhad Shirani, *Member, IEEE*, and S. Sandeep Pradhan, *Fellow, IEEE*

***Abstract*— This work considers an information-theoretic characterization of the set of achievable rates, costs, and distortions in a broad class of distributed communication and function computation scenarios with general continuous-valued sources and channels. A framework is presented which involves fine discretization of the source and channel variables followed by communication over the resulting discretized network. In order to evaluate the resulting achievable regions, convergence results for information measures are provided under the proposed discretization process. Prior works have considered such convergence results for mutual information quantities written in terms of univariate functions of random variables, and sums of independent random variables. These convergence results have been used to derive achievable regions in point-to-point communication scenarios and specific multiterminal scenarios with continuous alphabets. However, the best-known achievability results for distributed communication and function computation scenarios, which are based on structured coding strategies, involve mutual information quantities written in terms of bivariate functions of random variables, e.g., sum of two (not necessarily independent) random variables. A main contribution of this work is to show the convergence of mutual information quantities written in terms of sums of quantized random variables. This is an essential step in evaluating the achievable regions in continuous distributed computation scenarios by generalizing the structured coding strategies which have been previously used to derive the best-known achievable regions in discrete networks. The framework is used to provide achievability results for the problems of function computation over multiple-access channels, distributed source coding, function reconstruction (two-help-one), and multiple-descriptions source coding. In each scenario, discrete structured coding strategies along with the aforementioned convergence results are used to derive inner bounds to set of achievable rates, costs, and distortions. Furthermore, structured coding strategies are considered for distributed function computation scenarios involving computation of non-additive functions. The techniques are used to study an example where the objective is to compute the product of channel inputs over a multiple access channel, and an inner bound to the achievable rate region is evaluated. It is shown that, in contrast to many well-studied scenarios in multiterminal information theory, Gaussian input distribution is outperformed by the uniform input distribution.**

***Index Terms*— Data communication, information theory, rate distortion theory, channel capacity, channel coding, source coding.**

I. Introduction

OVER the past several decades, information theory has provided a framework for the study of the fundamental limits of communication — such as achievable rates, costs, and distortions — and the design of source and channel coding strategies in a wide range of communication scenarios, particularly over discrete source and channel networks. Many of the achievability results rely on the concept of strong typicality, which is based on the frequency of occurrence of symbols in sequences of discrete random variables [1]. The notion of strong typicality does not extend naturally to sequences of continuous variables, and hence cannot be used in studying continuous source and channel networks. To address this, Wyner [2] proposed a method for the study of the point-to-point (PtP) source coding with side-information. Wyner's method involves fine quantization of the source, the side-information, and the auxiliary variables to create a finite-alphabet problem, and then using the achievability results for the finite-alphabet problem to derive performance limits for the original problem using convergence properties of mutual information. In particular, this approach relies on two important techniques to guarantee convergence of the associated mutual information terms: the data processing inequality, and lower semi-continuity of mutual information as a function of probability distributions [3]. This approach has been proven to achieve the performance limits in the context of random *unstructured code ensembles* in point-to-point source coding and channel coding problems.

Our primary objective in this work is to characterize inner bounds to the set of achievable rates, costs, and distortions in multiterminal communication scenarios involving distributed communication and function computation in continuous networks. In contrast to point-to-point communications, in multi-terminal communication settings such as the broadcast channel [4], interference channel [5], variations of the MAC channel [6], [7], distributed source coding problem [8], and

Manuscript received 23 December 2022; revised 14 August 2023; accepted 4 December 2023. Date of publication 15 December 2023; date of current version 16 February 2024. This work was supported by NSF under Grant CCF 2241057 and Grant CCF 2132815. An earlier version of this paper was presented in part at the 2015 IEEE International Symposium on Information Theory (ISIT) [DOI: 10.1109/ISIT.2015.7282722] and in part at 2018 ISIT [DOI: 10.1109/ISIT.2018.8437668]. *(Corresponding author: Farhad Shirani.)*

Farhad Shirani is with the Knight Foundation School of Computing and Information Sciences, Florida International University, Miami, FL 33199 USA (e-mail: fshirani@fiu.edu).

S. Sandeep Pradhan is with the Electrical Engineering and Computer Science Department, University of Michigan, Ann Arbor, MI 48109 USA (e-mail: pradhanv@umich.edu).

Communicated by M. Wigger, Associate Editor for Shannon Theory.

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TIT.2023.3343668.

Digital Object Identifier 10.1109/TIT.2023.3343668

multiple descriptions source coding problem [9], it has been observed that the application of random *structured coding ensembles* — such as random linear coding ensembles — yields improved achievability results over random unstructured code ensembles in the discrete alphabet settings. In fact, it has been shown that in specific multiterminal communication scenarios, any optimality achieving sequence of codes muse be *asymptotically* algebraically closed (e.g., [10, Lemma 1]). Loosely speaking, the reason is that in these multiuser communication scenarios, in order to derive the best-known achievability results, in the discrete case, one needs to incorporate bivariate function computation at certain terminals of the system, which is facilitated by using structured codes which are algebraically closed under the bivariate operation, as opposed to random unstructured codes which do not possess such closure properties [6], [11], [12], [13], [14], [15], [16]. In these characterizations of performance limits, the mutual information quantities involve bivariate functions of random variables. For example, $I(X_1 \oplus X_2; Y)$ appears in the rate expressions in the problem of computation over MAC, where X_1 and X_2 are the channel inputs and Y is the channel output, and $\oplus$ refers to the addition operation of a finite field [13]. Convergence of such mutual information quantities is not guaranteed under the fine quantization of the variables using techniques of [2]. Specifically, these techniques rely on lower semi-continuity of mutual information and the data processing inequality to ensure convergence of mutual information quantities as the quantization step becomes asymptotically small, however, the data processing inequality does not guarantee that $I(\widehat{X}_1 \oplus \widehat{X}_2; \widehat{Y}) \leq I(X_1 \oplus X_2; Y)$, where $\widehat{X}_1, \widehat{X}_2, \widehat{Y}$ are the discretized versions of X_1, X_2, Y, respectively. In this work, we develop a new discretization framework and provide a performance analysis to derive such convergence guarantees. We use the framework to derive achievability results for distributed communication and function computation problems with continuous alphabets via discretization followed by discrete structured coding schemes. It should be pointed out that we do not introduce new structured coding strategies, rather, we apply previously developed structured coding techniques to the discretized network, and use the aforementioned convergence results to derive inner bounds to the achievable region. The preliminary results of this framework appeared in [16] and [17]. In a related work, this idea of discretizing the continuous random variables, and then applying discrete coding strategies has been recently used in the study of the compute-and-forward communication scenario [18].

The proposed discretization method in this work builds upon Wyner's fine quantization technique [2] and our preliminary work in [16] along with structured coding techniques for discrete alphabets. To elaborate, our approach involves taking a collection of jointly continuous random variables with some Markov chain constraints — enforced by the physical separation of distributed terminals — and performing three operations: (a) clipping to produce bounded random variables, (b) smoothing to produce random variables with continuous probability density functions (PDFs), and (c) quantization to produce finite alphabet random variables. Then, we use the structured coding frameworks, specifically linear codes, for discrete alphabets developed in the literature to derive achievable regions written in terms of mutual-information of discretized random variables and their sums, where the sum is with respect to some finite field. We show convergence of mutual information quantities involving random variables and their sums, as the quantization becomes fine and the clipping interval becomes large, to their continuous counterparts, where the sum is with real addition. We further extend this result to general bivariate functions such as products of pairs of variables using a novel embedding concept. In order to be able to apply the discrete coding strategies, one needs to ensure that the discretized random variables satisfy the Markov chains which capture the constraints on collaboration among the various terminals of the given communication problems. To address this, we preform the three discretization operations such that the resulting discretized random variables satisfy the Markov chains which are satisfied by the original continuous variables (Section III).

There are two main challenges in proving the above-mentioned convergence results for information quantities. The first challenge is the presence of Markov chains which need to be preserved throughout the discretization process. To explain further, let us focus on the Berger distributed source coding strategy [19]. Consider a distributed memoryless source (X, Y) with underlying joint distribution $P_{X,Y}$. Let U and V be the single-letter random variables corresponding to the quantization of X and Y, respectively, in the Berger-Tung coding strategy. Recall that the random variables must satisfy the Markov chain $U-X-Y-V$ among the source variables X, Y and the auxiliary variables U, V. If the Markov chain is not satisfied, one cannot use the Markov lemma to ensure joint typicality of the compressed sequences U^n and V^n in the distributed terminals, which is an essential step in deriving the Berger achievable region in the discrete alphabet setting [19]. However, one can see that if the random variables are quantized individually, then the quantized random variables do not satisfy the long Markov chain in general (see Example 1). To address this, we use a novel randomized clipping and discretization operations for the random variables in the middle of the chain, i.e., X and Y. The procedure clips and discretizes the random variables using locally generated independent noise variables with a carefully constructed distribution. This is detailed in Section III. The second challenge is to prove convergence of information quantities involving linear combinations of random variables under the three Markov-chain-preserving operations. We develop an analytical technique to bound mutual information under the clipping operation by iterative decomposition of the mutual information and entropy-power inequality (see Appendix C). Additionally, we develop analytical techniques to prove convergence under discretization by uncovering new connections between information quantities and variational distance (see Lemma 6). Furthermore, we build upon a technique developed in [20] for bounded and independent random variables, where convergence of entropies of sums of independent quantized random variables to that of quantized sum was addressed. In the present case, the random variables are not independent, and must maintain the given Markov structure throughout the discretization process. This requires new methods for proving convergence of information measures under discretization.

The contributions of this work are summarized as follows:

- We prove convergence of mutual information terms involving linear combinations of (not necessarily independent) random variables, under a discretization scheme involving clipping, smoothing, and fine quantization that *preserves* the underlying Markov chain constraints among the random variables. The previous known convergence results for mutual information terms involve univariate functions [2] and sums of independent random variables [20]. These convergence results are crucial in generalizing the previously known structured coding techniques developed for discrete networks and applying them to continuous networks to characterize the fundamental performance limits in multiterminal communication scenarios. (Theorem 1 and Theorem 2).
- We apply this discretization framework to derive inner bounds on the performance limits for general continuous-valued sources and channels in multi-terminal communication scenarios including computation over MAC, distributed source coding, lossy two-help-one problem, and multiple descriptions (MD) source coding, expressed in terms of single-letter information quantities. The achievability results are derived by applying the proposed discretization technique along with previously known *discrete structured coding strategies*. That is, structured coding strategies are applied on the discretized channel and source variables (Theorems 3-7).
- We provide an embedding technique for evaluating the achievable regions in multiterminal communication scenarios involving general bivariate functions such as product of pairs of random variables. We provide an example of computing such a function over an additive MAC with Gaussian noise, and derive an inner bound to the achievable rate region. We observe that in this scenario the uniform input distribution outperforms the Gaussian input distribution. (Section IV)

Other Related Work: Other than the aforementioned fine quantization techniques, a second class of techniques which has been considered for linear quadratic Gaussian (LQG) sources and channels is to use subtractive dithered lattice codes [21], [22]. The drawback of these lattice codes is that (a) they are very specific to the LQG nature of the problem, and hence are not amenable to extensions to non-Gaussian and nonlinear problems, and (b) they are based on the PtP communication perspective, and hence their applications to the multiterminal problems such as distributed source coding and multiple description source coding require a reduction of the latter problems to a sequence of PtP problems [22], [23], [24], [25]. However, it might be challenging to find such reductions for many problems, for example, the multiple description source coding with Gaussian sources and with linear distortion criterion studied in this work (see Section VI). A third approach which has been taken to study PtP source coding, PtP channel coding, and communication over MAC, derives performance limits using weak typicality [26] instead of strong typicality. Weak typicality is based on the empirical entropy of sequences of random variables, and is defined for both discrete and continuous variables. However, the weak typicality is not applicable in many multiterminal communication problems such as distributed source coding, and communication over broadcast channels, since for instance, the Markov lemma [27] (a crucial step in the derivation of achievable regions) is not valid for weakly typical sequences. In [28], an alternative approach was proposed by defining weak-* typicality which extends the notion of weak typicality using bounded functions. The Markov lemma has been shown to hold for weak-* typical sequences. The results were applied to source compression in the presence of side-information. The derivations in [2] and [28] are based on unstructured random code ensembles. The applications of this approach to multiterminal communication problems involving function computation, which require structured coding ensembles to achieve optimality, have not been studied.

The rest of the paper is organized as follows: Section II provides the preliminaries needed for the rest of the paper. Section III develops the first set of main results which form the framework for the analysis of the achievability results in the sequel. Section IV studies the problem of computation over MAC. In Section V, we consider the distributed compression of continuous sources and derive bounds on the fundamental limits of communication in the lossy two-help-one problem. Finally, Section VI considers the multiple descriptions source coding problem. Section VII concludes the paper.

Notation: We represent random variables by capital letters such as X, U and their realizations by small letters such as x, u. Sets are denoted by sans-serif letters such as X, U. The set of natural numbers, and the real numbers are represented by $\mathbb{N}$, and $\mathbb{R}$ respectively. Collections of sets are denoted by calligraphic letters such as $\mathcal{X}, \mathcal{U}$. The Borel sigma-field is denoted by $\mathcal{B}$. For $n \in \mathbb{N}$, we denote the sigma-field generated by $\mathcal{B}^n$ as $\sigma(\mathcal{B}^n)$. The random variable $\mathbb{1}_{\mathsf{E}}$ is the indicator function of the event E. The set of numbers $\{1, 2, \cdots, m\}$, $m \in \mathbb{N}$ is denoted by $[m]$ for brevity. For a given $n \in \mathbb{N}$, the n-length vector $(x_1, x_2, \ldots, x_n)$ is written as x^n. The function $h(\cdot)$ denotes the differential entropy. For the set $\mathsf{A} \subset \mathbb{R}^n$, we write $cl(\mathsf{A})$ to denote the convex closure. For a pair of distributions P_X and Q_X defined on alphabet $\mathcal{X}$, the variational distance is denoted by $TV(P_X, Q_X) := \frac{1}{2}\sum_{x\in\mathcal{X}} |P_X(x) - Q_X(x)|$.

II. Preliminaries

A. Source and Channel Models

We consider continuous memoryless source and channel networks with real-valued inputs and outputs, and without feedback. Such channel networks (source networks) are completely characterized by their associated channel transition probability (source distribution) and input cost functions (output distortion functions). In this paper, we consider the following formulation of the transition probability function.

Definition 1 (Transition Probability): A transition probability is a function $P : \mathbb{R} \times \mathcal{B} \to \mathbb{R}$ such that:

- For each $x \in \mathbb{R}$, $P(\cdot|x) : \mathsf{A} \mapsto P(\mathsf{A}|x)$ is a probability measure on $(\mathbb{R}, \mathcal{B})$.
- For each $\mathsf{A} \in \mathcal{B}$, $P(\mathsf{A}|\cdot) : x \mapsto P(\mathsf{A}|x)$ is a measurable function.

Definition 2 (Memoryless Channel Without Feedback): A channel is characterized by i) a transition probability $P_{Y|X} : \mathbb{R} \times \mathcal{B} \to \mathbb{R}$, and ii) a continuous cost function $\kappa : \mathbb{R} \to \mathbb{R}^+$, where X and Y are the channel input and output, respectively.

Remark 1: As noted in Definition 2, we assume that the cost function is continuous. This smoothness condition implies that $\kappa(x)$ is bounded for all $x \in \mathbb{R}$.

We assume that the channel is memoryless and used without feedback, i.e. the joint probability measure on $(\mathbb{R}^n, \sigma(\mathcal{B}^n))$ is given by the unique extension of the product measure

$$P(Y_i \in \mathsf{A}_i, i \in [n] | X^n = x^n) = \prod_{i=1}^{n} P_{Y|X}(\mathsf{A}_i | x_i), .$$

for all $\mathsf{A}_1, \mathsf{A}_2, \cdots, \mathsf{A}_n \in \mathcal{B}$ given x^n is transmitted over the channel by using the channel n times.

Definition 3 (Joint Channel Probability Measure): For a channel $(P_{Y|X}, \kappa)$, and given probability measure P_X on $(\mathbb{R}, \mathcal{B})$, the joint probability measure P_{XY} on $(\mathbb{R}^2, \sigma(\mathcal{B}^2))$ is the unique extension of the measure on product sets

$$\begin{aligned} P_{XY}(\mathsf{A} \times \mathsf{B}) &= \int_{\mathsf{A}} P_X(dx) P_{Y|X}(\mathsf{B}|x) \\ &= \int_{\mathsf{A}} P_X(dx) \int_{\mathsf{B}} P_{Y|X}(dy|x), \qquad \mathsf{A}, \mathsf{B} \in \mathcal{B}. \end{aligned}$$

We also consider PtP and multiuser source coding scenarios. An information source is characterized by its associated probability measure and distortion function as described below.

Definition 4 (Memoryless Source): A source is characterized by i) a probability measure $P_X : \mathcal{B} \to \mathbb{R}$, and ii) a jointly continuous distortion[1] function $d : \mathbb{R} \times \mathbb{R} \to \mathbb{R}^+$.

Remark 2: As noted in Definition 4, we assume that the distortion function is a jointly continuous. This smoothness condition implies that $d(x, \hat{x})$ is finite for all $x, \hat{x} \in \mathbb{R}$.

B. Structured Coding Ensembles

In various multi-terminal communication settings with discrete alphabets, it has been observed that the application of random structured coding ensembles — such as random linear coding ensembles — yields improved achievability results over random unstructured code ensembles in the discrete alphabet settings [4], [5], [6], [7], [8], [9]. In the next sections, we provide a discretization framework which allows us to discretize continuous networks and use structured coding techniques developed for these discrete networks, along with convergence results for the mutual information terms, to characterize inner-bounds to the achievable regions. In this section, we briefly define structured coding ensembles. A more complete description can be found at [13].

Definition 5 (Structured Code): For a given alphabet $\mathcal{X}$, a code of length n is a subset $\mathcal{C}$ of $\mathcal{X}^n$. A code is said to be structured if $\mathcal{C}$ is closed under a bivariate function $g : \mathcal{X} \times \mathcal{X} \to \mathcal{X}$ applied symbol-wise, i.e., for all $x^n, y^n \in \mathcal{C}$, we have $g^n(x^n, y^n) \in \mathcal{C}$. A code which is not structured is called unstructured.

[1] $\mathbb{R}^+$ denotes the set of non-negative real numbers.

For example, let $\mathcal{X} = \{0, 1\}$, and $g(x_1, x_2) = x_1 \oplus_2 x_2$, for all $x_1, x_2 \in \mathcal{X}$, where $\oplus_2$ denotes addition modulo-2. Then a code that is closed under g is called a binary linear code.

Definition 6 (Code Ensemble): A code ensemble is a collection of K codes $\mathsf{C} = \{\mathcal{C}_1, \ldots, \mathcal{C}_K\}$ defined on the same alphabet along with a probability distribution P on the set $\{1, 2, \ldots, K\}$. A structured code ensemble is one where all codes in the ensemble are structured with respect to the shared bivariate function.

III. Framework for Continuous to Discrete Source and Channel Transformation

This section introduces the components of a discretization framework which is considered in subsequent sections to study communication over continuous sources and channel networks. We prove convergence of mutual information of sums of discretized random variables to that of their continuous counterparts. Theorems 1 and 2 are the main results of this section.

As mentioned in the introduction, many multiterminal communication problems require the use of structured code ensembles to achieve the optimal rate region. The use of generic quantization and clipping functions does not guarantee convergence of the resulting mutual information terms involving sums of random variables which characterize the achievable region of the discretized problem. We consider jointly continuous random variables X, Y, U and V, with a joint PDF $f_{XY} f_{U|X} f_{V|Y}$, so that the variables satisfy the Markov chain $U - X - Y - V$, where $f_{U|X}$ denote the conditional PDF of U given X. Furthermore, we consider a jointly continuous distortion function $d : \mathbb{R}^2 \to \mathbb{R}^+$ and continuous reconstruction function $g : \mathbb{R}^2 \to \mathbb{R}$. This is a very generic scenario, and the results derived in this section are applied in multiterminal communication scenarios such as computation over MAC (Section IV), distributed source coding (Section V), and multiple descriptions source coding (Section VI) to derive achievable regions. We denote the joint probability measure as P_{XYUV}. Considering applications in distributed source coding, we refer to (X, Y) as the source variables, and (U, V) the auxiliary variables.

In order to preserve the Markov chain $U - X - Y - V$, we discretize the variables in two steps. In Section III-A, we quantize the auxiliary variables U and V to $\widehat{U}$ and $\widehat{V}$, respectively. We note that the Markov chain $\widehat{U} - X - Y - \widehat{V}$ holds since $\widehat{U}$ and $\widehat{V}$ are individual transformations of U and V, respectively. In the next step, in Section III-B, we discretize the source variables X and Y to $\widehat{X}$ and $\widehat{Y}$, respectively. We note that the Markov chain $\widehat{U} - \widehat{X} - \widehat{Y} - \widehat{V}$ may not necessarily hold if the source variables are simply discretized without modifying the auxiliary variables. To illustrate this issue more clearly, we provide the following example:

Example 1: Let X be uniformly distributed over the interval $[-1/2, 1/2]$ and $Y = X$. Let $U = 2X$ and $V = 2Y$. Note that the Markov chain $U - X - Y - V$ holds. Consider a naive quantization of all variables with grid-size one, centered at zero. Then, $\widehat{X}$ and $\widehat{Y}$, the quantized versions of X and Y, are equal to 0, whereas $\widehat{U}$ and $\widehat{V}$ are non-trivial ternary variables which are equal to each other with probability one. So, $\widehat{U} - \widehat{X} - \widehat{Y} - \widehat{V}$ does not hold.

To address this, we produce new variables $\overline{U}$ and $\overline{V}$, such that the marginal distribution of $(X,\overline{U})$ and $(Y,\overline{V})$ is preserved, and the Markov chain $\overline{U}-\widehat{X}-\widehat{Y}-\overline{V}$ is satisfied. In subsequent sections, we use discrete structured coding techniques of [8] for source variables $(\widehat{X},\widehat{Y})$ and auxiliary variables $(\overline{U},\overline{V})$ to derive new inner-bound to the achievable regions of several multi-terminal communication problems. As the discretrization becomes more fine, the latter distribution approaches (i.e., convergence in variational distance as shown in the proof of Theorem 2) the original joint distribution of the continuous variables.

A. Discretization of Auxiliary Random Variables

Our objective is to provide characterizations of achievable rate-distortion functions for source coding and achievable rate-cost functions for channel coding by first clipping, smoothing, and then quantizing the associated random variables. We refer to this entire process as discretization. Generic quantization and clipping operations have been used in prior literature (e.g., [29]). We modify these operations in the sequel to develop Markov-chain-preserving discretization procedures. Some of the prior convergence results available in the literature that are used in this work are included in Appendix A for ease of reference. The (generic) quantization operation is defined below.

Definition 7 (Quantization Function): Let $n \in \mathbb{N}$. The quantization function $Q_n:\mathbb{R}\to 2^{-n}\mathbb{Z}$ is defined as

$$Q_n(s)=\arg\min_{a\in 2^{-n}\mathbb{Z}}|s-a|, \qquad s\in\mathbb{R}.$$

Fix $\ell,\ell',\epsilon>0$, and $n\in\mathbb{N}$. We define the clipped variables $U_\ell, V_{\ell'}$ as follows:

$$\widetilde{U}_\ell=\begin{cases}U & \text{if } U\in[-\ell,\ell]\\ U' & \text{Otherwise}\end{cases}, \quad \widetilde{V}'_{\ell'}=\begin{cases}V & \text{if } V\in[-\ell',\ell']\\ V' & \text{Otherwise}\end{cases},$$

where U',V' are independent of each other and U,V, and are generated according to $f_{U'}(\cdot):=f_{U|U\in[-\ell,\ell]}(\cdot)$, and $f_{V'}(\cdot):=f_{V|V\in[-\ell',\ell']}(\cdot)$, respectively. Note that the clipping operation described above yields a continuous random variable as opposed to the generic clipping operation. We take ℓ,ℓ' sufficiently large. In general, a continuous random variable U may not have a continuous PDF. In such scenarios, a useful technique is to 'smoothen' the variable using additive noise.

Remark 3: In this work, we assume that for any continuous random variable X, the PDF approaches infinity in a finite number of points, and that the set of points of discontinuity has (Lebesgue) measure zero. Furthermore, we assume that the continuous random variables have finite variances.

We define the smoothed random variables $\widetilde{U}_{\ell,\epsilon},\widetilde{V}_{\ell',\epsilon}$, where

$$\widetilde{U}_{\ell,\epsilon}:=\widetilde{U}_\ell+\widetilde{N}_{\ell,\epsilon}, \qquad \widetilde{V}_{\ell',\epsilon}:=\widetilde{V}_\ell+\widetilde{N}'_{\ell',\epsilon},$$

$$f_{\widetilde{N}_{\ell,\epsilon}}(\tilde{n})=\frac{1}{2\epsilon}, \quad \tilde{n}\in(-\epsilon,\epsilon),$$

$$f_{\widetilde{N}'_{\ell',\epsilon}}(\tilde{n}')=\frac{1}{2\epsilon}, \quad \tilde{n}'\in(-\epsilon,\epsilon),$$

and the variables $\widetilde{N}_{\ell,\epsilon}$ and $\widetilde{N}'_{\ell',\epsilon}$ are mutually independent of each other, and of X,Y,V,U,U_ℓ and $V_{\ell'}$. This smoothing operation ensures that the variables have a continuous PDF which is required in our analysis. Next, we consider quantizing $\widetilde{U}_{\ell,\epsilon}$ and $\widetilde{V}_{\ell',\epsilon}$ to $\widehat{U}_{\ell,\epsilon,n}=Q_n(\widetilde{U}_{\ell,\epsilon})$ and $\widehat{V}_{\ell,\epsilon,n}=Q_n(\widetilde{V}_{\ell',\epsilon})$, respectively. Note that by construction the Markov chain $\widehat{U}_{\ell,\epsilon,n}-X-Y-\widehat{V}_{\ell,\epsilon,n}$ holds. We have the following theorem which shows the convergence of information measures under the above discretization operation. In proving the theorem, we first develop three important lemmas regarding the convergence of cost and distortion functions and smoothing of random variables which are provided in Appendix B.

Theorem 1: For any $\xi>0$ and all sufficiently large $n,\ell,\ell'>0$, and sufficiently small $\epsilon>0$, the following hold:

$$|I(\widehat{U}_{n,\ell,\epsilon}+\widehat{V}_{n,\ell',\epsilon};\widehat{U}_{n,\ell,\epsilon})-I(U+V;U)|\le\xi, \quad (1)$$

$$|I(\widehat{U}_{n,\ell,\epsilon}+\widehat{V}_{n,\ell',\epsilon};\widehat{V}_{n,\ell',\epsilon})-I(U+V;V)|\le\xi. \quad (2)$$

Proof: For the complete proof please see Appendix C. We provide an outline of the proof steps and techniques in the following. In the first step, we consider randomized clipping and show that for any $\zeta>0$, and all sufficiently large ℓ,ℓ', we have:

$$|I(\widetilde{U}_\ell+\widetilde{V}_{\ell'};\widetilde{U}_\ell)-I(U+V;U)|\le\zeta, \quad (3)$$

$$|I(\widetilde{U}_\ell+\widetilde{V}_{\ell'};\widetilde{V}_{\ell'})-I(U+V;V)|\le\zeta. \quad (4)$$

Note that one cannot directly apply the data processing inequality to show Equations (3) and (4) since $U_\ell+V_{\ell'}$ is not a processed version of $U+V$. As an intermediate step, we show that

$$I(\widetilde{U}_\ell;\widetilde{U}_\ell+\widetilde{V}_{\ell'})\le I(U;U+\widetilde{V}_{\ell'})+2\eta_1+\gamma_1,$$

where $\eta_1\to 0$ as $\ell,\ell'\to\infty$,

$$\begin{aligned}\gamma_1:=\;&P(A_{U,\ell}=0,B_{V,\ell'}=0)I(U';U'+V'|A_{U,\ell}=0,B_{V,\ell'}=0)\\&+P(A_{U,\ell}=0,B_{V,\ell'}=1)I(U';U'+V|A_{U,\ell}=0,B_{V,\ell'}=1),\end{aligned}$$

and we have defined $A_{U,\ell}$ as the indicator of $U\in[-\ell,\ell]$ and $B_{V,\ell'}$ as the indicator of $V\in[-\ell',\ell']$. Next, we bound from above the term γ_1 by applying the law of total variance and using the fact that $U+V$ has finite variance,[2] and show that $\gamma_1\to 0$ as $\ell,\ell'\to\infty$. Finally, we show the desired result by applying the entropy-power inequality and the law of total variance. In the second step, we consider smoothing of the random variables and show that for all $\gamma>0$, and all sufficiently small $\epsilon>0$ we have:

$$|I(\widetilde{U}_{\ell,\epsilon}+\widetilde{V}_{\ell',\epsilon};\widetilde{U}_{\ell,\epsilon})-I(\widetilde{U}_\ell+\widetilde{V}_{\ell'};\widetilde{U}_\ell)|\le\gamma \quad (5)$$

$$|I(\widetilde{U}_{\ell,\epsilon}+\widetilde{V}_{\ell',\epsilon};\widetilde{V}_{\ell',\epsilon})-I(\widetilde{U}_\ell+\widetilde{V}_{\ell'};\widetilde{V}_{\ell'})|\le\gamma. \quad (6)$$

The proof uses the lower semi-continuity [30] of information and a new lemma about convergence of mutual information for additive channels with peak power constraint approaching zero (see Lemma 5 given in Appendix B). In the next step, we consider quantization and show that for any $\gamma>0$, and all sufficiently large $n\in\mathbb{N}$, the following hold:

$$|I(\widehat{U}_{\ell,\epsilon,n}+\widehat{V}_{\ell',\epsilon,n};\widehat{U}_{\ell,\epsilon,n})-I(\widetilde{U}_{\ell,\epsilon}+\widetilde{V}_{\ell',\epsilon};\widetilde{U}_{\ell,\epsilon})|\le\gamma \quad (7)$$

$$|I(\widehat{U}_{\ell,\epsilon,n}+\widehat{V}_{\ell',\epsilon,n};\widehat{V}_{\ell',\epsilon,n})-I(\widetilde{U}_{\ell,\epsilon}+\widetilde{V}_{\ell',\epsilon};\widetilde{V}_{\ell',\epsilon})|\le\gamma. \quad (8)$$

[2]Note that this follows from the fact that $Var(U+V)\le 4\max(Var(U),Var(V))$ and the assumptions made in Remark 3.

The proof of this result is involved and relies on the fact that $f_{\widetilde{U}_{\ell,\epsilon},\widetilde{V}_{\ell',\epsilon}}$ is continuous over a compact support, and hence is uniformly continuous, which is ensured by the aforementioned clipping and smoothing operations. We refer the reader to Step 3 in Appendix C. □

B. Discretization of the Source Variables

In the following, we describe the procedure for discretizing the source variables while ensuring that the long Markov chain holds. Let $\ell, \ell' > 0$, and Z and W be two independent random variables that are independent of the source (X,Y) such that $Z \in [-\ell, \ell]$ with probability one, $W \in [-\ell', \ell']$ with probability one, and the distribution $P_Z P_W$ is given by

$$P_Z(\mathsf{A}) = \frac{P_X(\mathsf{A} \cap [-\ell,\ell])}{P_X([-\ell,\ell])}, \qquad P_W(\mathsf{B}) = \frac{P_Y(\mathsf{B} \cap [-\ell',\ell'])}{P_Y([-\ell',\ell'])},$$

for all events A and B in Borel sigma algebra. The PDFs of the variables Z and W are the truncated versions of that of X and Y, respectively. Define the clipped source variables as:

$$\widetilde{X}_\ell = \begin{cases} X & \text{if } X \in [-\ell,\ell] \\ Z & \text{otherwise} \end{cases} \tag{9}$$

and

$$\widetilde{Y}_{\ell'} = \begin{cases} Y & \text{if } Y \in [-\ell',\ell'] \\ W & \text{otherwise.} \end{cases} \tag{10}$$

Furthermore, let $n \in \mathbb{N}$, and define the quantized and clipped source variables $\widehat{X}_{n,\ell} := Q_n(\widetilde{X}_\ell)$, and $\widehat{Y}_{n,\ell'} := Q_n(\widetilde{Y}_{\ell'})$.

Theorem 2: Consider a quadruple of random variables (X,Y,U,V), where i) (X,Y) are jointly continuous with joint PDF $f_{X,Y}$, and ii) U,V are discrete random variables defined on finite sets U and V, respectively, and iii) the long Markov chain $U - X - Y - V$ holds. Then, for any $\xi > 0$ for all sufficiently large $n, \ell, \ell' > 0$, there exist random variables $\overline{U}_{n,\ell}$ and $\overline{V}_{n,\ell'}$ defined on $\mathsf{U} \times \mathsf{V}$ such that the long Markov chain $\overline{U}_{n,\ell} - \widehat{X}_{n,\ell} - \widehat{Y}_{n,\ell'} - \overline{V}_{n,\ell'}$ holds, and the following conditions are satisfied

$$|I(\overline{U}_{n,\ell} + \overline{V}_{n,\ell'}; \overline{U}_{n,\ell}) - I(U+V;U)| \le \xi, \tag{11}$$

$$|I(\overline{U}_{n,\ell} + \overline{V}_{n,\ell'}; \overline{V}_{n,\ell'}) - I(U+V;V)| \le \xi. \tag{12}$$

Proof: The complete proof is given in Appendix D. We give a brief outline of the proof here. Given the quadruple (X,Y,U,V), we use the discretization process described above to generate $(\widehat{X}_{n,\ell}, \widehat{Y}_{n,\ell})$ from (X,Y) in a distributed way. Then, we generate the random variables $(\widehat{X}_{n,\ell}, \widehat{Y}_{n,\ell}, \overline{U}_{n,\ell}, \overline{V}_{n,\ell})$ satisfying the long Markov Chain $\overline{U}_{n,\ell} - \widehat{X}_{n,\ell} - \widehat{Y}_{n,\ell'} - \overline{V}_{n,\ell'}$ such that

$$\lim_{n,\ell,\ell' \to \infty} TV\left(P_{\widehat{X}_{n,\ell}\widehat{Y}_{n,\ell'}\overline{U}_{n,\ell}\overline{V}_{n,\ell'}}, P_{XYUV}\right) = 0.$$

This relies on a new Pinsker-type inequality relating the mutual information with the variational distance under Markov chain constraints shown in Appendix B (see Lemma 6). Since the random variables U,V have finite alphabets, one can use the continuity of mutual information for finite-alphabet random variables to complete the proof of the theorem. □

Corollary 1: Given a triple of random variables (Y,U,V), for any $\xi > 0$, for all sufficiently large $n, n', \ell, \ell', \ell'' > 0$, and all sufficiently small ϵ, the following condition is satisfied

$$|I(\widehat{U}_{n,\ell,\epsilon} + \widehat{V}_{n,\ell',\epsilon}, \widehat{Y}_{n',\ell''}; \widehat{U}_{n,\ell,\epsilon}) - I(U+V,Y;U)| \le \xi, \tag{13}$$

where $\widehat{Y}_{n',\ell''} = Q_{n'}(\widetilde{Y}_{\ell''})$, and $\widetilde{Y}_{\ell''}$ is defined as in (10).

Proof: First, we apply clipping and quantization of the random variable Y. Then by the data processing inequality and the lower semi-continuity of mutual information, we have

$$|I(U+V, \widehat{Y}_{n',\ell''}; U) - I(U+V,Y;U)| \le \xi,$$

for all sufficiently large n' and ℓ''. Next we note that

$$I(U+V, \widehat{Y}_{n',\ell''}; U) = I(U+V; U|\widehat{Y}_{n',\ell''}) + I(U; \widehat{Y}_{n',\ell''}).$$

Regarding the first term, since $\widehat{Y}_{n',\ell''}$ is has a finite alphabet, we can apply Theorem 2 on each $I(U+V;U|\widehat{Y}_{n',\ell''} = \widehat{y})$ for each value of $\widehat{y}$ and show convergence. Regarding the second term, $I(\widehat{U}_{n,\ell,\epsilon}; \widehat{Y}_{n',\ell''})$ converges to $I(U; \widehat{Y}_{n',\ell''})$ due to lower semi-continuity of mutual information and data processing inequality. □

Remark 4: In this section, we have considered the addition operation as the bivariate function operating on the pair of variables U and V. The results obtained in this case can be extended to a large class of bivariate functions by embedding them into the addition operation, where an embedding is defined below, and using the fact that mutual information is invariant under any one-to-one univariate transformation of its arguments. This is demonstrated through an example in Section IV. This can potentially be further extended by using the Kolmogorov representation theorem [31], [32].

Definition 8 (Embedding into Addition Operation): A bivariate function $g : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ is said to be embedable in the addition operation if there exists a triple of mappings $(h(\cdot), \phi_1(\cdot), \phi_2(\cdot))$ such that $g(x,y) = h(\phi_1(x) + \phi_2(y)), \forall x, y \in \mathbb{R}$.

Example 2 (Embedding the Absolute Value of the Product): Consider the bivariate function $g(x,y) = |xy|, x, y \in \mathbb{R}$. Take $\phi_1(x) = \log|x|, \phi_2(y) = \log|y|, x, y \in \mathbb{R}$ and $h(x) = 2^x, x \in \mathbb{R}$. Then, $(h(\cdot), \phi_1(\cdot), \phi_2(\cdot))$ provides an embedding of $g(\cdot,\cdot)$ into the addition operation.

IV. Computation Over MAC

In this section, we consider a coding problem about the two-transmitter MAC. We consider a simple formulation which is purely a channel coding problem and that captures the essence of the key concepts. In the standard formulation of the multiple-access channel coding, the receiver wishes to recover both the messages reliably. Now consider a variation of this problem, where the decoder is interested in recovering only a single-letter bivariate function $g(\cdot,\cdot)$ of the channel inputs sent by the transmitters reliably. Originally, this problem was formulated in [33], and has been studied extensively with applications to interference channels and relay channels [34], [35], [36]. We demonstrate that structured codes can better facilitate the interaction between the two transmitters to ensure that the decoder recovers the desired information while transmitting information at a larger rate that can be sustained by unstructured codes. Formally, a memoryless stationary two-transmitter MAC, used without feedback, is given by a tuple $(P_{Y|X_1,X_2}, \kappa_1, \kappa_2)$, consisting of the transition probability $P_{Y|X_1,X_2} : \mathbb{R}\times\mathbb{R}\times\mathcal{B} \to \mathbb{R}$, and two continuous cost functions κ_1 and κ_2.

A. Problem Formulation and Main Result

Definition 9: Given a MAC $(P_{Y|X_1,X_2}, \kappa_1, \kappa_2)$, and a bivariate function $g : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, a transmission system with parameters (n, Θ_1, Θ_2) for reliable computation consists of a pair of encoder mappings and a decoder mapping $e_i : \{1, 2, \ldots, \Theta_i\} \to \mathbb{R}_i^n, i = 1, 2$, and $f : \mathbb{R}^n \to \mathbb{R}^n$. A quadruple of rates and costs $(R_1, R_2, \tau_1, \tau_2)$ is said to be achievable if $\forall\epsilon > 0$, and all sufficiently large n, there exists a transmission system with parameters (n, Θ_1, Θ_2) such that for $i = 1, 2$,

$$\frac{1}{n}\log\Theta_i \geq R_i - \epsilon, \quad \frac{1}{\Theta_i}\sum_{j=1}^{\Theta_i}\kappa_i(e_i(j)) \leq \tau_i + \epsilon,$$

$$\sum_{j=1}^{\Theta_1}\sum_{k=1}^{\Theta_2}\frac{1}{\Theta_1\Theta_2} \times P^n_{Y_1,Y_2|X_1,X_2}\left[f(Y^n) \neq g(e_1(j), e_2(k))|e_1(j), e_2(k)\right] \leq \epsilon,$$

for $i = 1, 2$. Let the optimal capacity cost region $\mathsf{C}(\tau_1, \tau_2)$ denote the set of all rate pairs (R_1, R_2) such that $(R_1, R_2, \tau_1, \tau_2)$ is achievable.

From now on, we focus on the class of bivariate functions $g(\cdot, \cdot)$ that are embeddable in the real addition operation with the associated triple of functions $(h(\cdot), \phi_1(\cdot), \phi_2(\cdot))$. In the following, we provide an achievable rate region that is based on linear codes and discretization.

Definition 10: Let $\mathcal{P}(\tau_1, \tau_2)$ denote the collection of distributions $P_{QU_1U_2X_1X_2}$ defined on $\mathsf{Q} \times \mathbb{R}^4$ such that (i) $(U_1X_1)-Q-(U_2X_2)$ form a Markov chain, with Q being a finite set, and (ii) $\mathbb{E}(\kappa_i(X_i)) \leq \tau_i, i \in \{1, 2\}$. For a $P_{QU_1U_2X_1X_2} \in \mathcal{P}$, let $\alpha_F(P_{QU_1U_2X_1X_2})$ denote the set of rate pairs $(R_1, R_2) \in [0, \infty)^2$ that satisfy

$$\begin{aligned} R_1 &\leq I(U_1; Y|U_2Q) + I(Z; Y|U_1U_2Q) \\ &\quad - I(Z; X_2|U_1U_2Q), \\ R_2 &\leq I(U_2; Y|U_1Q) + I(Z; Y|U_1U_2Q) \\ &\quad - I(Z; X_1|U_1U_2Q), \\ R_1 + R_2 &\leq I(U_1U_2; Y|Q) + 2I(Z; Y|U_1U_2Q) \\ &\quad - I(Z; X_1|U_1U_2Q) - I(Z; X_2|U_1U_2Q), \end{aligned}$$

where the mutual information terms are evaluated with $P_{QU_1U_2X_1X_2}P_{Y|X_1X_2}$, and $Z = \phi_1(X_1) + \phi_2(X_2)$. Let the information rate region be defined as

$$\mathsf{R}_F(\tau_1, \tau_2) = \mathrm{cl}\left(\bigcup_{P_{QU_1U_2X_1X_2}\in\mathcal{P}(\tau_1,\tau_2)}\alpha_F(P_{QU_1U_2X_1X_2})\right).$$

It should be noted that the standard capacity region of a multiple access channel is defined by considering the complete reconstruction of both messages, whereas in computation over MAC, only a bivariate function of the two input sequences needs to be reconstructed at the receiver.

Theorem 3: Given a MAC $(P_{Y|X_1,X_2}, \kappa_1, \kappa_2)$, and a bivariate function $g : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, the optimal capacity cost region $\mathsf{C}(\tau_1, \tau_2)$ contains the information rate region $\mathsf{R}_F(\tau_1, \tau_2)$, i.e., $\mathsf{R}_F(\tau_1, \tau_2) \subseteq \mathsf{C}(\tau_1, \tau_2)$.

Proof Outline: We have two coding layers associated with unstructured codes and structured codes. The two layers are combined using superposition coding. The unstructured codes yields the standard rate region for MAC. This can be shown using superposition coding techniques. For the structured coding layer, we discretize and smoothen the channel input variables X_1 and X_2 first, and then discretize the channel output Y to yield a discrete version of the problem with variables $\widehat{X}_{1,n,\ell,\epsilon}, \widehat{X}_{2,n,\ell',\epsilon}$ and $\widehat{Y}_{n',\ell''}$. Now one can use [13, Theorem 4.2] to show that the following rates are achievable for the structured coding layer using nested linear codes over arbitrarily large finite fields, assuming that the messages corresponding the unstructured coding layer are decoded correctly:

$$\begin{aligned} R_1 &\leq H(\widehat{X}_{1,n,\ell,\epsilon}|U_1U_2Q) \\ &\qquad - H(\widehat{X}_{1,n,\ell,\epsilon} + \widehat{X}_{2,n,\ell',\epsilon}|\widehat{Y}_{n',\ell''}U_1U_2Q) \\ &= I(\widehat{X}_{1,n,\ell,\epsilon} + \widehat{X}_{2,n,\ell',\epsilon}; \widehat{Y}_{n',\ell''}|U_1U_2Q) \\ &\qquad - I(\widehat{X}_{1,n,\ell,\epsilon} + \widehat{X}_{2,n,\ell',\epsilon}; \widehat{X}_{2,n,\ell',\epsilon}|U_1U_2Q), \end{aligned}$$

and

$$\begin{aligned} R_2 &\leq H(\widehat{X}_{2,n,\ell',\epsilon}|U_1U_2Q) \\ &\qquad - H(\widehat{X}_{1,n,\ell,\epsilon} + \widehat{X}_{2,n,\ell',\epsilon}|\widehat{Y}_{n',\ell''}U_1U_2Q) \\ &= I(\widehat{X}_{1,n,\ell,\epsilon} + \widehat{X}_{2,n,\ell',\epsilon}; \widehat{Y}_{n',\ell''}|U_1U_2Q) \\ &\qquad - I(\widehat{X}_{1,n,\ell,\epsilon} + \widehat{X}_{2,n,\ell',\epsilon}; \widehat{X}_{1,n,\ell,\epsilon}|U_1U_2Q). \end{aligned}$$

Now using Corollary 1, we see that the mutual information terms involving discrete variables converge to the corresponding terms with continuous variables using the following identity:

$$\begin{aligned} I(X_1 + X_2, Y; X_2) &= I(X_1 + X_2; X_2) + I(Y; X_1, X_2) \\ &\quad - I(Y; X_1 + X_2). \end{aligned}$$

The desired result follows by noting that the real addition is equal to the field addition for the discrete variables with probability approaching one for asymptotically large ℓ.

B. Computation of Products Over MAC

As mentioned in Remark 4, one can apply the techniques developed in the prequel to the computation of bivariate functions of variables other than the addition operation. In the following, as a proof of concept, we characterize an achievable rate region for the computation of products of pairs of random variables over additive MAC with Gaussian noise and with power constraint $\kappa_i(x_i) = x_i^2$ and $\tau_i = 1, i = 1, 2$. To elaborate, we let $Y = X_1 + X_2 + N$ and $Z = g(X_1, X_2) = X_1X_2$, where $N \sim \mathcal{N}(0, \sigma^2)$, and the input alphabet is $\mathbb{R}\backslash\{0\}$.[3] Recall that Example 2 provides an embedding of $|X_1X_2|$ in the real addition operation. In order to compute $g(X_1, X_2)$ we compute the pair $(|X_1X_2|, \mathrm{sign}^+(X_1X_2))$,

[3]Note that if the symbol zero is allowed, then recovering the product is trivial, and an infinite rate is achievable for any signal-to-noise ratio.

TABLE I
THE ACHIEVABLE SYMMETRIC RATES FOR COMPUTING THE PRODUCT OF RANDOM VARIABLES OVER THE NOISELESS ADDITIVE MAC

	Gaussian	Laplacian	Uniform
Rate	1.2540	0.8530	1.3161

where we define[4] $\text{sign}^+(x) := \mathbb{1}(x > 0)$, and leverage the fact that $\text{sign}^+(X_1X_2) = \text{sign}^+(X_1) \oplus_2 \text{sign}^+(X_2)$. We then transmit $|X_1X_2|$ and $\text{sign}^+(X_1X_2)$ separately by using the coding strategy given in the proof of Theorem 3 twice. For simplicity, we will restrict our analysis to a structured coding scheme with trivial auxiliary random variables, i.e., $\mathcal{U}_1 = \mathcal{U}_2 = \emptyset$. We also restrict our attention to symmetric input distributions for ease of computation. We use a superposition of two layers of structured codes. In the first layer, we encode the sign information yielding the rates as follows:

$$R_{s,i} = I(Z_s; Y) - I(Z_s; \text{sign}^+(X_j)), \quad i, j = 1, 2, \quad j \neq i,$$

where $Z_s = \text{sign}^+(X_1) \oplus_2 \text{sign}^+(X_2)$. In the second layer, we encode the information regarding the absolute value of the random variables yielding rates as follows:

$$R_{p,i} = I(Z_p; Y|Z_s) - I(Z_p; |X_j|\big|Z_s), \quad i, j = 1, 2, \quad j \neq i,$$

where $Z_p = |X_1X_2|$. The resulting rates are $R_i = R_{p,i} + R_{s,i}, i = 1, 2$ which can be simplified as given below:

$$R_i = R_{p,i} + R_{s,i} = I(Z; Y) - I(Z; X_j), \quad i, j = 1, 2, j \neq i.$$

Note that characterizing the set of achievable rates in this inner bound involves optimizing over the set of distributions $P_{X_1}P_{X_2}$ under power constraints. We provide achievable rates for three choices of random variables, namely the uniform, Gaussian, and Laplacian variables from the generalized Gaussian family all having zero mean and unit variance. Table I provides the set of achievable rates for the noiseless MAC scenario, i.e. $\sigma^2 = 0$. It is worth noting that the uniform distribution, achieves the best symmetric rate among the three distributions. Note that the Gaussian lies between the Laplacian and the uniform in the family with the shape parameters being $\beta = 2, 1$ and ∞, respectively. As can be observed, the achievable rates follow a similar order. The fact that uniform distribution achieves a higher symmetric rate compared to Gaussian distribution is noteworthy. An outline of the computational steps is given in Appendix E. For the additive MAC with Gaussian noise, Figure 1 shows the achievable rates for the uniform and the Gaussian input distributions as a function of the channel signal-to-noise ratio.

Remark 5: Note that when the channel is noiseless, the capacity to compute the sum is infinite, however to compute the product, the achievable rates in Table I are finite. This illustrates the effect of mismatch between the structure in the channel and the structure in the computational objective.

C. Computation of Sums Over Gaussian MAC Example

Consider the MAC given by $Y = X_1 + X_2 + N$, where N is zero-mean Gaussian with variance σ_N^2. The decoder wishes

[4]Note that this is different from the conventional definition of $\text{sign}^+(\cdot)$.

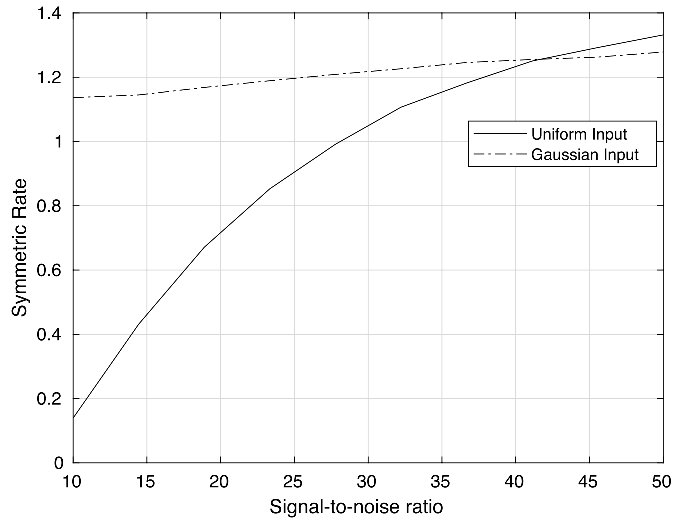


Fig. 1. Achievable symmetric rates for computation of products over an additive MAC with Gaussian noise for Uniform and Gaussian input distributions as a function of SNR.

to compute the sum of the inputs, i.e., $Z = g(X_1, X_2) = X_1 + X_2$. We have power constraints on X_1 and X_2: $\kappa_1(x_1) = x_1^2$ and $\kappa_2(x_2) = x_2^2$, for all $x_1, x_2 \in \mathbb{R}$. Let $\tau_i = P_i$ for $i = 1, 2$. The rates achievable using unstructured code ensembles is given by the standard MAC capacity region given [26, Chapter 15] by

$$\Big\{(R_1, R_2) : R_1 \leq \frac{1}{2}\log\left(1 + \frac{P_1}{\sigma_N^2}\right), R_2 \leq \frac{1}{2}\log\left(1 + \frac{P_2}{\sigma_N^2}\right), \quad R_1 + R_2 \leq \frac{1}{2}\log\left(1 + \frac{P_1 + P_2}{\sigma_N^2}\right)\Big\}.$$

This is achieved using independent Gaussian inputs X_1 and X_2 of variances P_1 and P_2, respectively. Using the same distribution, one can achieve the following rates while employing structured code ensembles.

$$\Big\{(R_1, R_2) : R_1 \leq \frac{1}{2}\log\left(\frac{P_1(P_1 + P_2 + \sigma_N^2)}{(P_1 + P_2)\sigma_N^2}\right), \quad R_2 \leq \frac{1}{2}\log\left(\frac{P_2(P_1 + P_2 + \sigma_N^2)}{(P_1 + P_2)\sigma_N^2}\right)\Big\}.$$

Comparing the sum-rate we see that the structured coding scheme performs better than the unstructured coding scheme when

$$\left(1 + \frac{P_1}{P_2}\right)\left(1 + \frac{P_2}{P_1}\right) \leq 1 + \frac{P_1}{\sigma_N^2} + \frac{P_2}{\sigma_N^2}.$$

For the case when $P_1 = P_2 = P$ boils down to the condition that $\frac{P}{\sigma_N^2} \geq 1.5$.

V. DISTRIBUTED SOURCE CODING

In this section we consider the distributed source coding problem. Our objective is to derive an inner bound to the achievable rate-distortion region for the two-help-one problem [1] with continuous sources using structured codes. As a first step, in Section V-A, we first prove the achievability of the Berger-Tung rate-distortion region for the continuous two-user distributed source coding problem. In Section V-B, we add a structured coding layer and derive an inner bound.

A. Two-User Distributed Source Coding

Next, we consider distributed source coding problem consisting of two correlated and memoryless continuous-valued sources X and Y, characterized by a probability measure P_{XY}, which needs to be compressed distributively into bits to be sent to a joint decoder. The joint decoder wishes to reconstruct the sources with respect to two separable distortion measures $d_x : \mathbb{R} \times \mathbb{R} \to \mathbb{R}^+$ and $d_y : \mathbb{R} \times \mathbb{R} \to \mathbb{R}^+$. The distributed source coding problem is a lossy version of the Slepian-Wolf source coding problem. This is a well-studied problem [19], [37], [38], and we skip the formal definition for conciseness. Well-known inner and outer bounds (called Berger-Tung bounds) on the performance limits of this problem exist in the discrete case [19], [39], [40], [41]. For the continuous source and reconstruction alphabet case, the only available results in the literature are for jointly Gaussian source distributions and quadratic distortion functions [42], [43]. The results presented in the following are valid for general distributions and distortion functions.

Let RD denote the set of all achievable rate and distortion tuples (R_1, R_2, D_1, D_2). We denote $\mathsf{R}(D_1, D_2)$ as the set of all rates (R_1, R_2) such that (R_1, R_2, D_1, D_2) is achievable.

Definition 11: Let $\mathcal{P}(D_1, D_2)$ denote the collection of pairs of transition probabilities $P_{U|X}$ and $P_{V|Y}$, and pairs of continuous reconstruction functions $g_i : \mathbb{R}^2 \to \mathbb{R}$ for $i = 1, 2$, such that $\mathbb{E}d_x(X, \widehat{X}) \leq D_1$, $\mathbb{E}d_y(Y, \widehat{Y}) \leq D_2$, where the expectations are evaluated with the joint measure $P_{XY}P_{U|X}P_{V|Y}$, i.e., with the Markov chain $U - X - Y - V$, and $\widehat{X} = g_1(U, V)$ and $\widehat{Y} = g_2(U, V)$. For a $(P_{U|X}, P_{V|Y}, g_1, g_2) \in \mathcal{P}(D_1, D_2)$, let $\alpha(P_{U|X}, P_{V|Y}, g_1, g_2)$ denote the set of rate pairs $(R_1, R_2) \in [0, \infty)^2$ that satisfy

$$R_1 \geq I(X; U|V),\ R_2 \geq I(Y; V|U), R_1 + R_2 \geq I(XY; UV).$$

Let the information rate region be defined as

$$\mathsf{R}_{QB}(D_1, D_2) = \text{cl}\left(\bigcup_{(P_{U|X}, P_{V|Y}, g_1, g_2) \in \mathcal{P}(D_1, D_2)} \alpha(P_{U|X}, P_{V|Y}, g_1, g_2) \right).$$

Theorem 4: For a given source (P_{XY}, d_x, d_y), we have $\mathsf{R}_{QB}(D_1, D_2) \subseteq \mathsf{R}(D_1, D_2)$.

Proof: Please see Appendix F. □

B. Lossy Two-Help-One Problem

Next, we consider a coding theorem for continuous correlated sources for the two-help-one problem. Consider a triple of memoryless continuous-valued sources (X, Y, Z) characterized by a probability measure P_{XYZ}. Let $d : \mathbb{R}^2 \to \mathbb{R}^+$ be a jointly continuous distortion function. The sources X and Y act as helpers for the third source Z. The sources need to be compressed distributively with rates R_1, R_2 and R_3, respectively, into bits to be sent to a joint decoder. For simplicity we let $R_3 = 0$. The joint decoder wishes to reconstruct the source Z with respect to a distortion function d.

1) Problem Formulation and Main Result:

Definition 12: An (n, Θ_1, Θ_2) transmission system consists of mappings $e_i : \mathbb{R}^n \to \{1, 2, \ldots, \Theta_i\}$, for $i = 1, 2$, and $f : \{1, 2, \ldots, \Theta_1\} \times \{1, 2, \ldots, \Theta_2\} \to \mathbb{R}^n$. A triple (R_1, R_2, D) is said to be achievable if there exists a sequence of $(n, \Theta_{1n}, \Theta_{2n})$ transmission systems such that for $i = 1, 2$,

$$\lim_{n\to\infty} \frac{\log \Theta_i}{n} \leq R_i,\ \lim_{n\to\infty} \mathbb{E}d_n(Z^n, f(e_1(X^n), e_2(Y^n))) \leq D,$$

where $d_n(\cdot, \cdot)$ is the n-letter average distortion, i.e., $d_n(z^n, \hat{z}^n) = \frac{1}{n}\sum_{i=1}^n d(z_i, \hat{z}_i)$ for all $z^n, \hat{z}^n \in \mathbb{R}^n$. Let $\mathsf{R}(D)$ denote the set of rates (R_1, R_2) such that (R_1, R_2, D) is achievable.

We provide a coding theorem for the continuous sources.

Definition 13: Let $\mathcal{P}(D)$ denote the collection of transition probabilities $P_{QU_1V_1UV\widehat{Z}|XY}$ such that (i) $(UU_1) - (XQ) - (YQ) - (VV_1)$ form a Markov chain, (ii) Q is independent of (X, Y), (iii) $\widehat{Z} = g(U_1, V_1, U + V)$ for some function g, and (iv) $\mathbb{E}d(Z, \widehat{Z}) \leq D$, where the expectations are evaluated with distribution $P_{XYZ}P_{QU_1V_1UV\widehat{Z}|XY}$. For a $P_{QU_1V_1UV\widehat{Z}|XY} \in \mathcal{P}(D)$, let $\alpha_F(P_{QU_1V_1UV\widehat{Z}|XY})$ denote the set of rate pairs $(R_1, R_2) \in [0, \infty)^2$ that satisfy

$$\begin{aligned}
R_1 &\geq I(X; UU_1|QV_1) + I(U + V; V|QU_1V_1) \\
&\quad - I(U; V|QU_1V_1), \\
R_2 &\geq I(Y; VV_1|QU_1) + I(U + V; U|QU_1V_1) \\
&\quad - I(U; V|QU_1V_1) \\
R_1 + R_2 &\geq I(XY; UVU_1V_1|Q) + I(U + V; V|QU_1V_1) \\
&\quad + I(U + V; U|QU_1V_1) - I(U; V|QU_1V_1),
\end{aligned}$$

where the mutual information terms are evaluated with $P_{XYZ}P_{QU_1V_1UV\widehat{Z}|XY}$. Let the information rate region be defined as

$$\mathsf{R}_F(D) = \text{cl}\left(\bigcup_{P_{QU_1V_1UV\widehat{Z}|XY} \in \mathcal{P}(D)} \alpha_F(P_{QU_1V_1UV\widehat{Z}|XY}) \right).$$

Theorem 5: For a given source (P_{XYZ}, d) we have $\mathsf{R}_F(D) \subseteq \mathsf{R}(D)$.

Remark 6: The concept of embedding, described in Definition 8, has been taken into account in the above rate-distortion characterization through the function $g(\cdot, \cdot, \cdot)$ and the choice of auxiliary variables in obtaining the reconstruction.

Proof Outline: We propose a coding scheme involving two layers to prove the theorem. The first is the Berger-Tung unstructured coding layer. The second is the structured coding layer that uses nested linear codes. First we discretize the auxilliary variables U and V, and then discretize the source variables X and Y, as described in Section III, to come up with a discrete version of the problem at hand with discrete variables $\overline{U}_{n,\ell}, \overline{V}_{n,\ell'}, \overline{X}_{n,\ell}$, and $\overline{Y}_{n,\ell'}$ as stated in Theorems 1 and 2. The Berger-Tung unstructured coding rates are derived as in Theorem 4. The structured coding is accomplished using nested linear codes. The rates associated with this layer can be understood as follows: for ease of explanation, assume that the unstructured coding auxiliary variables U_1, V_1 and the time-sharing variable Q are trivial. Then, for the discrete communication system, the rates

$$\begin{aligned}
R_1 &\geq H(\overline{U}_{n,\ell} + \overline{V}_{n,\ell'}) - H(\overline{U}_{n,\ell}|\widehat{X}_{n,\ell}) \\
&= I(\widehat{X}_{n,\ell}; \overline{U}_{n,\ell}) - I(\overline{U}_{n,\ell}; \overline{V}_{n,\ell'}) \\
&\quad + I(\overline{V}_{n,\ell'}; \overline{U}_{n,\ell} + \overline{V}_{n,\ell'}),
\end{aligned}$$

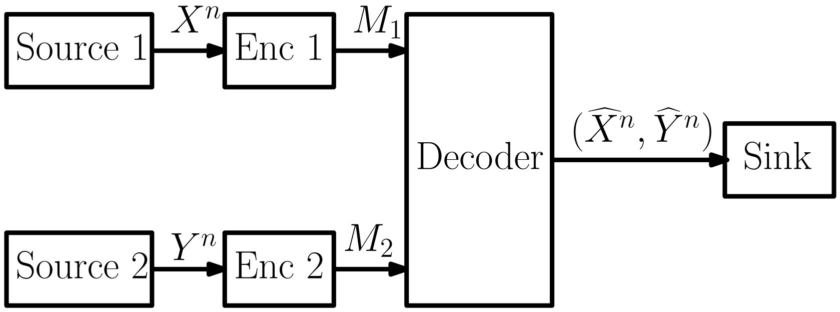


Fig. 2. General distributed source coding problem, where $n \in \mathbb{N}$ is the blocklength.

and similarly

$$\begin{aligned} R_2 &\geq H(\overline{U}_{n,\ell} + \overline{V}_{n,\ell'}) - H(\overline{U}_{n,\ell}|\widehat{Y}_{n,\ell'}) \\ &= I(\widehat{Y}_{n,\ell'}; \overline{V}_{n,\ell'}) - I(\overline{U}_{n,\ell}; \overline{V}_{n,\ell'}) \\ &\quad + I(\overline{U}_{n,\ell}; \overline{U}_{n,\ell} + \overline{V}_{n,\ell'}), \end{aligned}$$

can be achieved using nested linear codes (over arbitrarily large prime fields) along with joint-typical encoding and decoding as given in [13, Theorem 3.7]. Achievability for the original problem with continuous variables follows by Theorems 1, 2, and Theorem 4, and by noting that $P(U+V \in [-\ell, \ell]) \to 1$ as $\ell \to \infty$, so that the field addition for the discrete variables $\overline{U}_{n,\ell}$ and $\overline{V}_{n,\ell}$ is equal to real addition with probability approaching one as $\ell \to \infty$. The two coding layers can be combined using the technique of superposition coding.

2) Gaussian Lossy Two-Help-One Example: Consider a pair of zero-mean, jointly Gaussian, unit-variance correlated sources X and Y with correlation coefficient $\rho > 0$. Let $Z = X - cY$ for some c, and let $d(z, \hat{z}) = (z - \hat{z})^2$. Let us evaluate a subset of the inner bound to the achievable rate-distortion region using a specific test channel. Let us denote $\sigma_Z^2 = 1 + c^2 - 2\rho c$, and let D denote the target distortion. Let us choose $\mathsf{Q} = \phi$, and $U_1 = V_1 = 0$. Moreover consider

$$U = X + Q_1, \quad \text{and} \quad V = cY + Q_2,$$

where Q_1 and Q_2 are independent zero-mean Gaussian random variables that are independent of the pair (X, Y). We take their variances to be q_1 and $\frac{D\sigma_Z^2}{\sigma_Z^2 - D} - q_1$. With this choice we see that $U + V = Z + Q_1 + Q_2$, and we take $\widehat{Z} = \mathbb{E}(Z|U+V) = \frac{\sigma_Z^2 - D}{\sigma_Z^2}(U+V)$, which results in $\mathbb{E}d(Z, \widehat{Z}) = D$. Now let us see the achievable rates.

$$R_1 \geq \frac{1}{2} \log \frac{\sigma_Z^4}{q_1(\sigma_Z^2 - D)}, \quad R_2 \geq \frac{1}{2} \log \frac{\sigma_Z^4}{D\sigma_Z^2 - q_1(\sigma_Z^2 - D)}.$$

Eliminating q_1, we see that the rate distortion tuple (R_1, R_2, D) satisfying the following equations is achievable:

$$2^{-2R_1} + 2^{-2R_2} \leq \left(\frac{\sigma_Z^2}{D}\right)^{-1}.$$

This was obtained in [15] using lattice codes. Here we derived this using nested linear codes and the convergence of random variables.

VI. Multiple Descriptions Source Coding

In this section, we consider the multiple descriptions source coding problem, where given a source X, the encoder wishes to construct a set of $\ell \geq 2$ descriptions of the source, such that given each subset of descriptions, the source can be reconstructed with a desired distortion. The scenario has been studied extensively in the discrete case [9], [44], [45], [46], [47], [48]. We derive an achievable rate-distortion region for general continuous sources, and demonstrate that structured codes achieve a larger rate-distortion region compared to unstructured codes in an example with Gaussian sources and test-channels. It should be noted that the achievable region for the Gaussian multiple descriptions problem with $\ell = 2$ was derived by [49]. The results presented in the following are valid for general distributions and distortion functions.

A. Problem Formulation and Main Result

Definition 14: Let $\ell \geq 2$, $\mathsf{L} = [\ell]$, $\mathcal{L} = 2^{\mathsf{L}} - \phi$, $n \in \mathbb{N}$, and $\Theta_i \in \mathbb{N}, i \in \mathsf{L}$. A coding system with parameters $(n, \Theta_i : i \in \mathsf{L})$ for multiple description coding of a given source $(P_X, d_{\mathsf{N}} : \mathsf{N} \in \mathcal{L})$, consists of ℓ encoder mappings and $2^\ell - 1$ decoder mappings:

$$e_i : \mathsf{X}^n \to [\Theta_i], \qquad f_{\mathsf{N}} : \prod_{i \in \mathsf{N}} [\Theta_i] \to \mathsf{X}^n,$$

where $i \in \mathsf{L}$, and $\mathsf{N} \in \mathcal{L}$. A of rate-distortion tuple $(R_i : i \in \mathsf{L}, D_{\mathsf{N}} : \mathsf{N} \in \mathcal{L}) \in (\mathbb{R}^+)^{\ell + 2^\ell - 1}$ is said to be achievable if for all $\epsilon > 0$, and for all sufficiently large n, there exists a coding system with parameters $(n, \Theta_i : i \in \mathsf{L})$ such that

$$\begin{gathered} \frac{1}{n} \log \Theta_i \leq R_i + \epsilon, \\ \mathbb{E}[d_{\mathsf{N}}(X^n, f_{\mathsf{N}}((e_i(X^n))_{i \in \mathsf{N}}))] \leq D_{\mathsf{N}} + \epsilon, \qquad i \in \mathsf{L}, \mathsf{N} \in \mathcal{L}. \end{gathered}$$

The operational rate-distortion region $\mathsf{R}_{op}(D_{\mathsf{N}} : \mathsf{N} \in \mathcal{L})$ is given by the set of all achievable rate-distortion tuples $(R_i : i \in \mathsf{L}, D_{\mathsf{N}} : \mathsf{N} \in \mathcal{L})$.

We use the discretization techniques developed in prior sections, along with the Sperner[5] Set Coding (SSC) strategy in [13, Theorem 6.2] to derive an achievable rate-distortion region for multiple descriptions coding with continuous sources. The SSC strategy is a generalization of the Zhang-Berger strategy [44] to multiple descriptions scenarios

[5] A family of sets is called a Sperner family of sets if none of its elements is a subset of another element. For instance $\{\{1\}, \{2, 3\}\}$ is a Sperner set, but $\{\{1\}, \{1, 3\}\}$ is not.

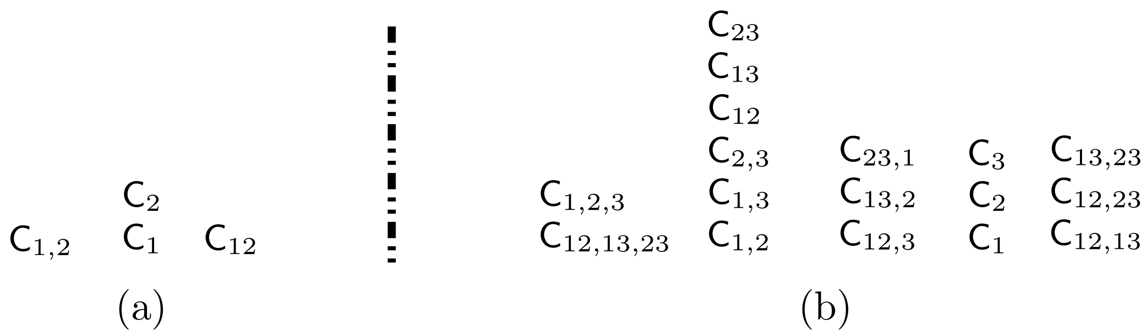


Fig. 3. The SSC codebooks for (a) $\ell = 2$ and (b) $\ell = 3$.

involving more than two descriptions and three decoders. In the Zhang and Berger strategy, a common codebook $\mathsf{C}_{1,2}$ is used to encode the common information decoded at the Decoder 1 and Decoder 2. The scheme uses a total of four codebooks, corresponding to information decoded at all three decoders, each of the individual decoders, and only the joint decoder, respectively. In the SSC strategy, for each subset M of decoders, one codebook C_{M} is generated, which encodes the common information decoded among this subset of decoders. It is shown in [9] that only codebooks for which M is a Sperner set are non-redundant. To elaborate more, we explain the random variables decoded at each decoder for the SSC strategy for the three-descriptions problem, i.e. $\ell = 3$ (See Figure 3). Note that in this case, we have 7 decoders corresponding to the 7 non-empty subsets of $\{1, 2, 3\}$. The complete explanation of the scheme is provided in [9]. Let S_{L} be the Sperner set for $\ell = 3$. It is known that $\mathbf{S}_{\mathsf{L}}$ has 17 elements. Let $U_{\mathcal{M}}, \mathcal{M} \in \mathbf{S}_{\mathsf{L}}$ be a vector of random variables with whose joint distribution with source X is given by $P_{X, U_{\mathcal{M}}, \mathcal{M} \in \mathbf{S}_{\mathsf{L}}}$. The SSC strategy generates 17 independent random codebooks $\mathsf{C}_{\mathcal{M}}, \mathcal{M} \in \mathbf{S}_{\mathsf{L}}$, with blocklength n, where $\mathsf{C}_{\mathcal{M}}$ is generated based on the single letter distribution $P_{U_{\mathcal{M}}}$ for $\mathcal{M} \in \mathbf{S}_{\mathsf{L}}$. Each codebook is binned independently ℓ times, once per description. Given a source sequence X^n, the encoder finds codewords $U^n_{\mathcal{M}} \in \mathsf{C}_{\mathcal{M}}, \mathcal{M} \in \mathbf{S}_{\mathsf{L}}$ which are jointly typical with each other and the source sequence and sends the corresponding bin numbers on each description. Each decoder decodes a subset of the codewords. To elaborate, decoder $\mathsf{N} \in 2^{\mathsf{L}} - \phi$ recovers $U_{\mathcal{M}}$ based on the received bin numbers if $\mathsf{N} \subseteq \mathcal{M}$. Let $\mathbf{M}_{\mathsf{N}}$ be the set of indices of random variables whose corresponding codewords are decoded at decoder N, and let $\widetilde{\mathbf{M}}_{\mathsf{N}}$ be the indices of those which are decodable if we have access to strict subsets of the descriptions received by N. Furthermore, for the collections of families $\mathbf{M}_1, \mathbf{M}_2$ and $\mathbf{M}_3$, we write $[U, V, W]_{(\mathbf{M}_1, \mathbf{M}_2, \mathbf{M}_3)}$ to denote the unordered collection of random variables $\{U_{\mathbf{M}_1}, V_{\mathbf{M}_2}, W_{\mathbf{M}_3}\}$. The following theorem provides an achievable region for the multiple descriptions problem using the discretization process developed in the previous sections along with the SSC strategy with unstructured random codes developed in [13, Theorem 6.2] for discrete sources and test-channels.

Definition 15: Given a source $(P_X, d_{\mathsf{N}} : \mathsf{N} \in \mathcal{L})$, let $\mathcal{P}(D_{\mathsf{N}} : \mathsf{N} \in \mathcal{L})$ denote the collection of pairs $(P, g_{\mathcal{L}})$ of (a) joint distribution P on random variables X and $U_{\mathcal{M}}, \mathcal{M} \in \mathbf{S}_{\mathsf{L}}$ with X-marginal distribution P_X and (b) a set of reconstruction functions $g_{\mathcal{L}} := (g_{\mathsf{N}} : \mathsf{U}_{\{\mathsf{N}\}} \rightarrow \mathsf{X}, \mathsf{N} \in \mathcal{L})$ such that $\mathbb{E} d_{\mathsf{N}}(X, g_{\mathsf{N}}(U_{\{\mathsf{N}\}})) \leq D_{\mathsf{N}}, \forall \mathsf{N} \in \mathcal{L}$, where the expectations are evaluated with the distribution P. For a $(P, g_{\mathcal{L}}) \in \mathcal{P}(D_{\mathsf{N}} : \mathsf{N} \in \mathcal{L})$, define $\alpha_{\mathrm{SS}}(P, g_{\mathcal{L}})$ as the set of rate tuples $(R_i : i \in \mathsf{L})$ satisfying the following constraints for some non-negative real numbers $(\rho_{\mathcal{M},i}, r_{\mathcal{M}})_{i \in \widetilde{\mathcal{M}}, \mathcal{M} \in \mathbf{S}_{\mathsf{L}}}$:

$$\overline{I}(U_{\mathbf{M}}) + I(U_{\mathbf{M}}; X) \leq \sum_{\mathcal{M} \in \mathbf{M}} r_{\mathcal{M}}, \forall \ \mathbf{M} \subset \mathbf{S}_{\mathsf{L}}, \tag{14}$$

$$\sum_{\mathcal{M} \in \mathbf{M}_{\mathsf{N}} \setminus (\mathbf{L} \cup \widetilde{\mathbf{M}}_{\mathsf{N}})} (r_{\mathcal{M}} - \sum_{i \in \widetilde{\mathcal{M}}} \rho_{\mathcal{M},i}) \leq \overline{I}(U_{\mathbf{M}_{\mathsf{N}} \setminus (\mathbf{L} \cup \widetilde{\mathbf{M}})}) + I(U_{\mathbf{M}_{\mathsf{N}} \setminus (\mathbf{L} \cup \widetilde{\mathbf{M}})}; U_{\mathbf{L} \cup \widetilde{\mathbf{M}}}), \forall \mathbf{L} \subset \mathbf{M}_{\mathsf{N}}, \forall \mathsf{N} \in \mathcal{L}, \tag{15}$$

$$R_i = \sum_{\mathcal{M}} \rho_{\mathcal{M},i}, \tag{16}$$

where we have defined $\overline{I}(Z^k) := \sum_{j=1}^{k} I(Z_k; Z^{k-1})$ for a random vector Z^k, $\mathbf{M}_{\mathsf{N}}$ is the set of all codebooks decoded at decoder N, that is $\mathbf{M}_{\mathsf{N}} := \{\mathcal{M} \in \mathbf{S}_{\mathsf{L}} | \exists \mathsf{N}' \subset \mathsf{N}, \mathsf{N}' \in \mathcal{M}\}$, and $\widetilde{\mathbf{M}}_{\mathsf{N}}$ denotes the set of all codebooks decoded at decoders $\mathsf{N}_p \subsetneq \mathsf{N}$ which receive subsets of descriptions received by N, that is $\widetilde{\mathbf{M}}_{\mathsf{N}} := \bigcup_{\mathsf{N}_p \subsetneq \mathsf{N}} \mathbf{M}_{\mathsf{N}_p}$. The mutual information terms are evaluated with the distribution P. Define the Sperner Set Coding rate-distortion region as

$$\mathsf{R}_{\mathrm{SS}}(D_{\mathsf{N}} : \mathsf{N} \in \mathcal{L}) := cl \left(\bigcup_{(P, g_{\mathcal{L}}) \in \mathcal{P}(D_{\mathsf{N}} : \mathsf{N} \in \mathcal{L})} \alpha_{\mathrm{SS}}(P, g_{\mathcal{L}}) \right).$$

Theorem 6: Given a source $(P_X, d_{\mathsf{N}} : \mathsf{N} \in \mathcal{L})$, the operational rate-distortion region contains the information rate-distortion region, i.e., $\mathsf{R}_{\mathrm{SS}}(D_{\mathsf{N}} : \mathsf{N} \in \mathcal{L}) \subseteq \mathsf{R}_{op}(D_{\mathsf{N}} : \mathsf{N} \in \mathcal{L})$.

Proof Outline. Given the random variables X and $U_{\mathcal{M}}, \mathcal{M} \in \mathbf{S}_{\mathsf{L}}$ described in the theorem statement, the transmission system first discretizes the source using techniques developed in the prior sections and then uses the discrete SSC strategy introduced in [13, Theorem 6.2] to achieve the rate-distortion vector in (14), (15), and (16). The mutual-information terms in (14), (15) for the discretized variables converge to that of the continuous variables as the clipping limits are increased asymptotically and the quantization step approaches zero by similar arguments as in the prior sections.

The following theorem provides an achievable region for the multiple descriptions problem using the discretization process developed in the previous sections along with the SSC strategy with both unstructured and structured random codes developed in [13, Theorem 6.8] for discrete sources and test-channels. It should be noted that in [13, Theorem 6.8], the information inequalities are stated in terms of the entropy terms. Whereas in order to be able to apply the results of Theorems 1 and 2, one needs to express the achievable region in terms of mutual information quantities. The following theorem provides such a representation.

Definition 16: Given a source $(P_X, d_{\mathsf{N}} : \mathsf{N} \in \mathcal{L})$, let $\mathcal{P}(D_{\mathsf{N}} : \mathsf{N} \in \mathcal{L})$ denote the collection of pairs $(P, g_{\mathcal{L}})$ of (a) joint distribution P on random variables X, $U_{\mathcal{M}}, \mathcal{M} \in \mathbf{S}_{\mathsf{L}}$, $V_{\mathcal{A}_{\text{in}}}, V_{\mathcal{A}_{\text{out}}}, V_{\mathcal{A}_{\text{sum}}}$, with X-marginal distribution P_X, where $\mathcal{A}_j \in \mathbf{S}_{\mathsf{L}}, j \in \{\text{in}, \text{out}, \text{sum}\}$, are three distinct families, and all the auxiliary random variables take values in $\mathbb{R}$, and (b) a set of reconstruction functions $g_{\mathcal{L}} = \{g_{\mathsf{N}} : \mathsf{U}'_{\{\mathsf{N}\}} \to \mathsf{X}, \mathsf{N} \in \mathcal{L}\}$, such that $V_{\mathcal{A}_{\text{sum}}} = V_{\mathcal{A}_{\text{in}}} + V_{\mathcal{A}_{\text{out}}}$, and $\mathbb{E} d_{\mathsf{N}}(X, g_{\mathsf{N}}(U'_{\{\mathsf{N}\}})) \leq D_{\mathsf{N}}\ \forall \mathsf{N} \in \mathcal{L}$, where $U'_{\{\mathsf{N}\}} := (U_{\{\mathsf{N}\}}, V_{\mathbf{N}})$ and $\mathbf{N} := \{\mathcal{A}_i | \mathcal{A}_i \in \mathbf{M}_{\mathsf{N}}, i \in \{\text{in}, \text{out}, \text{sum}\}\}$, and the expectations are evaluated with P. For a $(P, g_{\mathcal{L}}) \in \mathcal{P}(D_{\mathsf{N}} : \mathsf{N} \in \mathcal{L})$, define $\alpha_F(P, g_{\mathcal{L}})$ as the set of rate tuple $(R_i, i \in \mathsf{L})$ satisfying the following constraints for some non-negative real numbers $(\rho_{\mathcal{M},i}, r_{o,\mathcal{M}})_{i \in \widetilde{\mathcal{M}}, \mathcal{M} \in \mathbf{S}_{\mathsf{L}}}$, $r'_{\mathcal{A}_{\text{in}}}, \rho_{\mathcal{A}_{\text{in}},i}, i \in \widetilde{\mathcal{A}}_{\text{in}}, r'_{\mathcal{A}_{\text{out}}}, \rho_{\mathcal{A}_{\text{out}},i}, i \in \widetilde{\mathcal{A}}_{\text{out}}$:

i) Covering Constraints: for all $\mathbf{M} \subset \mathbf{S}_{\mathsf{L}}, \mathbf{E} \subset \mathbf{A}$ and $\alpha, \beta \in \mathbb{F}_p^+$,

$$\sum_{\mathcal{M} \in \mathbf{M}} r_{\mathcal{M}} + \sum_{\mathcal{E} \in \mathbf{E}} r'_{\mathcal{E}} \geq \bar{I}(U_{\mathbf{M}}, V_{\mathbf{E}}) + I(U_{\mathbf{M}} V_{\mathbf{E}}; X), \tag{17}$$

$$\sum_{\mathcal{M} \in \mathbf{M}} r_{\mathcal{M}} + r'_{\mathcal{A}_{\text{out}}} \geq \bar{I}(U_{\mathbf{M}}, V_{\mathcal{A}_{\text{out}}}) + I(U_{\mathbf{M}} W_{\mathcal{A}_{\text{out}},\alpha,\beta}; X) - I(W_{\mathcal{A}_{\text{sum}},\alpha,\beta}; V_{\mathcal{A}_{\text{in}}} | U_{\mathbf{M}}) + I(V_{\mathcal{A}_{\text{in}}}; V_{\mathcal{A}_{\text{out}}} | U_{\mathbf{M}}). \tag{18}$$

ii) Packing constraints: for all $\overline{\mathbf{L}} \subset \overline{\mathbf{M}}_{\mathsf{N}}, \mathcal{A}_{\text{sum}} \notin \mathbf{M}_{\mathsf{N}}$,

$$\sum_{\mathcal{M} \in \mathbf{M}_{\mathsf{N}} \backslash \widetilde{\mathbf{M}}_{\mathsf{N}} \cup \mathbf{L}} \Big(r_{\mathcal{M}} - \sum_{j \in \widetilde{\mathcal{M}}} \rho_{\mathcal{M},j} \Big) + \sum_{\substack{\mathcal{E} \in \mathbf{M}_{\mathsf{N}} \backslash \widetilde{\mathbf{M}}_{\mathsf{N}} \cup \overline{\mathbf{L}} \\ \bigcap \{\mathcal{A}_i | i \in \{\text{in},\text{out}\}\}}} \Big(r'_{\mathcal{E}} - \sum_{j \in \widetilde{\mathcal{E}}} \rho_{o,\mathcal{E},j} \Big) \leq \bar{I}([UVW]_{\overline{\mathbf{M}}_{\mathsf{N}} \backslash \widetilde{\mathbf{M}}_{\mathsf{N}} \cup \mathbf{L}}) + I([UVW]_{\overline{\mathbf{M}}_{\mathsf{N}} \backslash \widetilde{\mathbf{M}}_{\mathsf{N}} \cup \mathbf{L}}; [UVW]_{\widehat{\mathbf{M}}_{\mathsf{N}} \cup \overline{\mathbf{L}}}), \tag{19}$$

and for all $\overline{\mathbf{L}} \subset \overline{\mathbf{M}}_{\mathsf{N}}, \mathcal{A}_{\text{sum}} \in \mathbf{M}_{\mathsf{N}}, \mathcal{A}_{\text{in}} \notin \mathbf{M}_{\mathsf{N}}, \mathcal{A}_{\text{out}} \notin \mathbf{M}_{\mathsf{N}}$

$$\sum_{\mathcal{M} \in \mathbf{M}_{\mathsf{N}} \backslash \widetilde{\mathbf{M}}_{\mathsf{N}} \cup \mathbf{L}} \Big(r_{\mathcal{M}} - \sum_{j \in \widetilde{\mathcal{M}}} \rho_{\mathcal{M},j} \Big) + r'_{\mathcal{A}_{\text{out}}} - \sum_{j \in \widetilde{\mathcal{A}}_{\text{sum}}} \rho_{o,\mathcal{A}_{\text{sum}},j} \leq \bar{I}(U_{\mathbf{M}_{\mathsf{N}}}, V_{\mathcal{A}_{\text{out}}}) + I(U_{\mathbf{M}_{\mathsf{N}}} W_{\mathcal{A}_{\text{sum}},1,1}; [UVW]_{\widehat{\mathbf{M}}_{\mathsf{N}} \cup \overline{\mathbf{L}}}) - I(W_{\mathcal{A}_{\text{sum}},1,1}; V_{\text{in}} | U_{\mathbf{M}_{\mathsf{N}}}) + I(V_{\text{in}}; V_{\text{out}} | U_{\mathbf{M}_{\mathsf{N}}}), \tag{20}$$

where (a) $R_i = \sum_{\mathcal{M}} \rho_{\mathcal{M},i}$, (b) $\mathbf{A} := \{\mathcal{A}_{\text{in}}, \mathcal{A}_{\text{out}}\}$, (c) $\overline{\mathbf{M}}_{\mathsf{N}} := (\mathbf{M}_{\mathsf{N}}, \{\mathcal{A}_j, j \in \{\text{in}, \text{out}\} | \mathcal{A}_j \in \mathbf{M}_{\mathsf{N}}\}, \{(\mathcal{A}_{\text{sum}}, 1, 1) | \mathcal{A}_{\text{sum}} \in \mathbf{M}_{\mathsf{N}}\})$, (d) $\widehat{\mathbf{M}}_{\mathsf{N}} := \bigcup_{\mathsf{N}' \subsetneq \mathsf{N}} \overline{\mathbf{M}}_{\mathsf{N}'}$, (e) $r'_{\mathcal{A}_{\text{in}}} \leq r'_{\mathcal{A}_{\text{out}}}$, and (f) $W_{\mathcal{A}_3,\alpha,\beta} := \alpha V_{\mathcal{A}_{\text{in}}} + \beta V_{\mathcal{A}_{\text{out}}}$.[6] The mutual information terms are evaluated with the distribution P. Define information rate-distortion region as

$$\mathsf{R}_F(D_{\mathsf{N}} : \mathsf{N} \in \mathcal{L}) := cl \left(\bigcup_{(P, g_{\mathcal{L}}) \in \mathcal{P}(D_{\mathsf{N}} : \mathsf{N} \in \mathcal{L})} \alpha_F(P, g_{\mathcal{L}}) \right).$$

Theorem 7: For a given source $(P_X, d_{\mathsf{N}} : \mathsf{N} \in \mathcal{L})$, the operational rate-distortion region contains the information rate-distortion region, i.e., $\mathsf{R}_F(D_{\mathsf{N}} : \mathsf{N} \in \mathcal{L}) \subseteq \mathsf{R}(D_{\mathsf{N}} : \mathsf{N} \in \mathcal{L})$.

Proof Outline. Given the random variables X and $U_{\mathcal{M}}, \mathcal{M} \in \mathbf{S}_{\mathsf{L}}, V_{\mathcal{A}_{\text{in}}}, V_{\mathcal{A}_{\text{out}}}, V_{\mathcal{A}_{\text{sum}}}$ described in the theorem statement, the transmission system first discretizes the source using techniques developed in the prior sections, and then uses a coding scheme based on nested linear codes considered in [13, Theorem 6.8]. Next, we rewrite the entropy terms in [13, Theorem 6.8] in terms of mutual information terms using the fact that for any triple U, V, X the following holds

$$\begin{aligned} H(\alpha U + \beta V|X) - H(U|X) &= H(\alpha U + \beta V|X) - H(U, V|X) + H(V|X, U) \\ &= H(\alpha U + \beta V|X) - H(\alpha U + \beta V, V|X) + H(V|X, U) \\ &= -H(V|X, \alpha U + \beta V) + H(V|X, U) \\ &= I(\alpha U + \beta V; V|X) - I(U; V|X). \end{aligned}$$

The resulting mutual-information terms in the covering bounds given in (17),(18) and packing bounds given in (19),(20) for the discretized variables converge to that of the continuous variables as the clipping limits are increased asymptotically and the quantization step size approaches zero by similar arguments as in the prior sections.

B. Linear Quadratic Gaussian Examples

1) Vector Gaussian Example: We proceed to show through an example that using structured codes gives gains in terms of the achievable rate-distortion. In this example, we evaluate inner bounds to the achievable rate-distortion region of a vector Gaussian example. The vector Gaussian multiple descriptions problem has been studied extensively in prior works [50], [51], [52]. We consider a specifically tailored vector Gaussian problem which allows for an analytical proof of strict sub-optimality of the unstructured coding scheme as compared to the structured coding scheme (Propositions 1 and 2). The set-up is shown in Figure 4. Here X and Z are independent zero-mean, unit-variance, Gaussian sources. The distortion function for the individual decoders is the mean squared error. Decoder 1 and 2 want to reconstruct X and Z, respectively, with mean squared error less than or equal to P, and Decoder 2 wants to reconstruct $Y = X + Z$ with distortion less than or equal to $2P$. Each of the joint decoders wish to reconstruct X and Z with distortion, given by the mean square error, less than or equal to P.[7]

Proposition 1: The rate triple $(R_1, R_2, R_3) = (\frac{1}{2}\log(\frac{1}{P}), \frac{1}{2}\log(\frac{1}{P}), \frac{1}{2}\log(\frac{2}{P}))$ is achievable using the SSC strategy with structured codes, i.e. $(R_1, R_2, R_3) \in \mathsf{R}_F$.

Proof: The proof follows by taking $\mathcal{A}_{\text{in}} = \{1\}$, $\mathcal{A}_{\text{out}} = \{2\}$, and $\mathcal{A}_{\text{sum}} = \{3\}$ with $V_{\{1\}} = X + N_P$, $V_{\{2\}} = Z + N'_P$ and $V_{\{3\}} = V_{\{1\}} + V_{\{2\}}$ in Theorem 7, where N_P and N'_P are independent Gaussian variables with zero mean and variance P, and all other variables are taken to be trivial. □

Proposition 2: For $P < 1/2$, the rate triple $(R_1, R_2, R_3) = (\frac{1}{2}\log(\frac{1}{P}), \frac{1}{2}\log(\frac{1}{P}), \frac{1}{2}\log(\frac{2}{P}))$ is not achievable with SSC

[6] The collection $\{\mathcal{A}_3, \alpha, \beta\}$ is used as the subscript for W since the random variable is defined using α and β.

[7] Note that here we have considered multiple distortion constraints at the joint decoders. The arguments in Theorems 6 and 7 can be extended to the scenario with multiple distortion constraints in a straightforward manner.

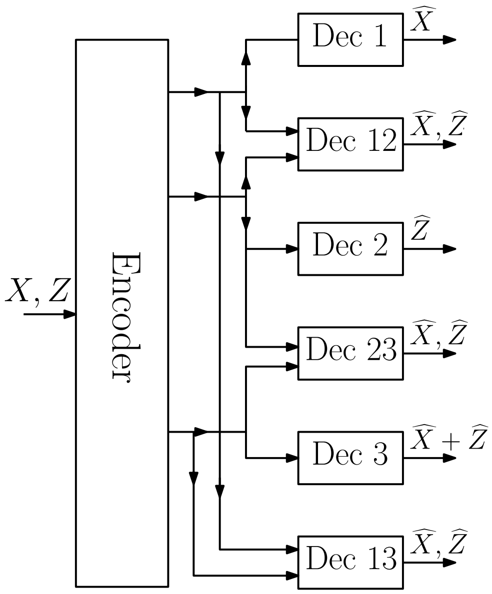
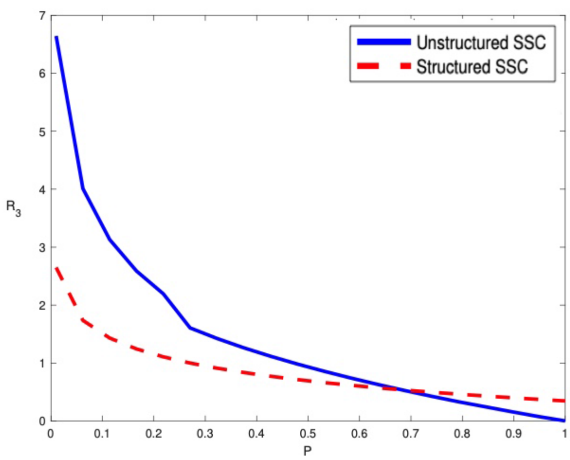


Fig. 4. Example with the Vector Gaussian Source.



P	SSC Unstructured	SSC Structured
0.01	6.643	2.649
0.114	3.130	1.431
0.218	1.816	1.107
0.322	1.425	0.912
0.426	1.11	0.772
0.531	0.858	0.663
0.635	0.631	0.573
0.739	0.427	0.497
0.843	0.245	0.431
0.947	0.0772	0.373
1	0.00	0.346

Fig. 5. Achievable rates R_3 in Example 1 using the SSC strategy with unstructured codes (red curve) and structured codes (blue curve).

strategy using unstructured codes with jointly Gaussian test channels, i.e. (R_1, R_2, R_3) is not in R_{SS} restricted to jointly Gaussian test channels.

Proof: The achievable rates are shown in Figure 5. The details of the proof are provided in Appendix G. □

Note that in the above example, as a result of the independence between the two source components X and Z, the extra covering bound in (18) is redundant. However, this is not always the case, to illustrate this point we investigate the following example.

2) Scalar Gaussian Example: Consider a three-descriptions problem where the source X is a scalar zero-mean unit-variance Gaussian random variable and the distortion is measured with respect to mean squared error at every decoder. We derive a set of achievable rate-distortions using the expressions in Theorem 7 and jointly Gaussian test channels.

Consider the random variables U and V which are jointly Gaussian with X and have the following covariance matrix ($\frac{1}{2} < P < \frac{2}{3}$):

$$\text{Cov}([X, U, V]) = \begin{bmatrix} 1 & 1-P & 1-P \\ 1-P & 1-P & 0 \\ 1-P & 0 & 1-P \end{bmatrix}.$$

We intend to transmit U on the first description (i.e. $V_{\{1\}} = U$), V on the second description (i.e. $V_{\{2\}} = V$) and $U+V$ on the third description (i.e. $V_{\{3\}} = U+V$). In this case the covering bound (18) is not redundant. To see this, note that the covering bound is non-redundant if $I(U+V; V|X) - I(U; V|X) < 0$. Simplifying the inequality, the bound is non-redundant if $\text{Var}(V|X,U) < \text{Var}(V|X,U+V)$. Also,

$$\text{Var}(V|X,U) = \frac{P(1-P)}{2}, \quad \text{Var}(V|X,U+V) = \frac{1-P}{2}$$
$$\Rightarrow \text{Var}(V|X,U) < \text{Var}(V|X,U+V) \iff P < \frac{2}{3},$$

which shows that the bound is non-redundant in this setting. We calculate the achievable rates using Fourier–Motzkin elimination yielding:

$$R_1 = R_2 = \max\left\{\frac{I(UV;X)}{2}, \right.$$
$$\left. I(U;X) - I(\alpha U + \beta V; V|X) + I(U;V|X)\right\},$$
$$R_3 = R_1 - H(U) + H(U+V).$$

We have

$$I(UV;X) = \frac{1}{2}\log(\frac{1}{2P-1}), \quad I(U;X) = \frac{1}{2}\log\frac{1}{P},$$
$$I(U;V|X) - I(\alpha U + \beta V; V|X)$$
$$= \frac{1}{2}\log\left(\frac{\alpha^2 P}{\alpha^2 + \beta^2 - (\alpha+\beta)^2(1-P)}\right).$$

Hence the rates

$$R_1 = R_2 = \max\left\{\frac{1}{2}\log(\frac{P^2}{P+(1-P)^2}), \frac{1}{4}\log(\frac{1}{2P-1})\right\},$$

$R_3 = R_1 + \frac{1}{2}$, and the distortions $D_1 = D_2 = P$, $D_3 = 2P, D_{12} = D_{13} = D_{23} = (2P-1)$ are achievable.

VII. Conclusion

A new framework for deriving the fundamental performance bounds of continuous source and channel networks was introduced. The framework involves fine discretization of the source and channel variables followed by communication over the resulting discretized network. Convergence results for information measures under the proposed discretization process were provided, and these results were used to derive the fundamental limits of computation over MAC, distributed source coding with distortion constraints, the function reconstruction problems (two-help-one), and the multiple-descriptions source coding problem.

Appendix A
Preliminaries on Convergence of Distributions and Cost/Distortion Function

A. Convergence of Distributions and Information Measures

In this appendix, we present several known results on convergence of information measures which are used in the subsequent sections.

Definition 17 (Convergence of Probability Measures): Consider a sequence of probability measures $P_n, n \in \mathbb{N}$, defined on the probability space $(\Omega, \mathcal{F})$,

- **Strong Convergence:** $P_n, n \in \mathbb{N}$ is said to converge strongly to P if

$$\lim_{n\to\infty} P_n(\mathsf{A}) = P(\mathsf{A}), \text{ for all } \mathsf{A} \in \mathcal{F}.$$

- **Convergence in Total Variation:** $P_n, n \in \mathbb{N}$ is said to converge in total variation to P if

$$\lim_{n\to\infty} TV(P_n, P) = 0,$$

where $TV(P,Q) := \sup_{\mathsf{A}} |P(\mathsf{A}) - Q(\mathsf{A})|$ is the total variation between P and Q.

Remark 7: It can be noted convergence in total variation guarantees strong convergence, which in turn guarantees convergence in distribution.

Lemma 1 (Lower Semi-Continuity of Mutual Information [3]): Consider a sequence of pairs of random variables $(S_n, T_n), n \in \mathbb{N}$, defined on $(\mathbb{R}^2, \sigma(\mathcal{B} \times \mathcal{B}))$. If P_{S_n,T_n} converges strongly to $P_{S,T}$, then

$$I(S;T) \le \liminf_{n\to\infty} I(S_n;T_n).$$

Intuitively, if we take $n \to \infty$, then the discrete random variable $Q_n(S)$ converges to the continuous random variable S in distribution, and hence by Lemma 1, for variables S and T, the mutual information $I(Q_n(S); Q_n(T))$ converges to $I(S;T)$. This is stated formally in the following lemma.

Lemma 2 (Convergence of Discretized Variables [30]): For any two random variables (S,T), the sequence $(Q_{n_1}(S), Q_{n_2}(T))$ converges in distribution to (S,T) as $n_1, n_2 \to \infty$. Consequently, if $I(S;T) < \infty$, we have

$$\lim_{n_1,n_2\to\infty} I(Q_{n_1}(S); Q_{n_2}(T)) = I(S;T).$$

Lemma 3 (Convergence of Clipped Variables): Let $\ell_1, \ell_2, u_1, u_2 > 0$, then for any two random variables (S,T) with $I(S;T) < \infty$, we have

$$\lim_{\ell_1,\ell_2,u_1,u_2\to\infty} TV(P_{\widetilde{S}_{\ell_1,u_1},\widetilde{T}_{\ell_2,u_2}}, P_{ST}) = 0$$

and hence

$$\lim_{\ell_1,\ell_2,u_1,u_2\to\infty} I(\widetilde{S}_{\ell_1,u_1}; \widetilde{T}_{\ell_2,u_2}) = I(S;T).$$

Proof: The lemma follows by noting that the total variation between P_{ST} and $P_{S_{\ell_1,u_1}T_{\ell_2,u_2}}$ is given by

$$TV(P_{\widetilde{S}_{\ell_1,u_1},\widetilde{T}_{\ell_2,u_2}}, P_{ST}) = [1 - P\left[(-\ell_1 \le S \le u_1) \cap (-\ell_2 \le T \le u_2)\right]].$$

□

Appendix B
New Auxiliary Results on Convergence of Distributions and Cost/Distortion Functions

A. Convergence of Cost/Distortion Functions and Smoothing of Random Variables

The following lemma is used in evaluating the distortion and cost of communication strategies in continuous networks in the subsequent sections.

Lemma 4 (Convergence of Cost Functions and Distortion Functions): Let S and T be two random variables. For any continuous function $\kappa : \mathbb{R} \to \mathbb{R}^+$ such that $\mathbb{E}(\kappa(S)) < \infty$, there exist two increasing (and approaching ∞) sequences of lengths l_m, u_m such that

$$\lim_{m\to\infty} \mathbb{E}\kappa(\widetilde{S}_{l_m,u_m}) = \mathbb{E}\kappa(S).$$

For any jointly continuous function $d : \mathbb{R}^2 \to \mathbb{R}^+$ such that $\mathbb{E}(d(S,T)) < \infty$, there exist four increasing (and approaching ∞) sequences of lengths l_n, u_n, and l_m, u_m such that

$$\lim_{n\to\infty} \lim_{m\to\infty} \mathbb{E}d(\widetilde{S}_{l_n,u_n}, \widetilde{T}_{l_m,u_m}) = \mathbb{E}d(S,T).$$

Proof: Let us fix $\epsilon > 0$ and let $\alpha_1 = \liminf_{s\to\infty} \kappa(s)$ and $\alpha_2 = \liminf_{s\to\infty} \kappa(-s)$. Since $\mathbb{E}(\kappa(S)) < \infty$, we have

$$\int_{[-\ell,u]^c} \kappa(s) dP_S(s) \le \epsilon, \quad \text{and} \quad \int_{[-\ell,u]^c} dP_S(s) \le \epsilon \tag{21}$$

for all $u > U(\epsilon)$ and $\ell > L(\epsilon)$ for some $U(\epsilon)$ and $L(\epsilon)$.

First consider the case when $\alpha_1 < \infty$ and $\alpha_2 < \infty$. In this case, there exists $U_1(\epsilon)$ such that for all $u > U_1(\epsilon)$, we have

$$\alpha_1 + 2\epsilon > \alpha_1 + \epsilon \ge \left(\inf_{\tilde{u}>u} \kappa(\tilde{u})\right) \ge \alpha_1 - \epsilon.$$

This implies that for all $u > U_1(\epsilon)$ there exists $u^* > u$ such that $\alpha_1 + 2\epsilon > \kappa(u^*)$. Hence, by choosing $u > \max\{U_1(\epsilon), U(\epsilon)\}$, we obtain a $u^* > u$ such that

$$\kappa(u^*)P(S \ge u^*) \le (\alpha_1 + 2\epsilon)P(S \ge u^*)$$
$$= (\alpha_1 - \epsilon + 3\epsilon) \int_{u^*}^{\infty} dP_S(s) \le \int_{u^*}^{\infty} \kappa(s) dP_S(s) + 3\epsilon^2.$$

Similarly, there exists ℓ^* such that

$$\kappa(-\ell^*)P(S \le -\ell^*) \le \int_{-\infty}^{-\ell^*} \kappa(s) dP_S(s) + 3\epsilon^2.$$

Now using the above results, consider

$$\mathbb{E}\kappa(\widetilde{S}_{\ell^*,u^*}) = \int_{-\ell^*+}^{u^*-} \kappa(s) dP_S(s) + \kappa(u^*)P(S \ge u^*) + \kappa(-\ell^*)P(S \le \ell^*) \le 6\epsilon^2 + \mathbb{E}\kappa(S). \tag{22}$$

Moreover,

$$\mathbb{E}\kappa(\widetilde{S}_{\ell^*,u^*}) \geq \int_{-\ell^*+}^{u^*-} \kappa(s)dP_S(s)$$
$$\geq \int \kappa(s)dP_S(s) - 2\epsilon = -2\epsilon + \mathbb{E}\kappa(S).$$

Next consider the case when $\alpha_1 = \infty$. Define

$$u^* := \arg\min_{\{u \geq U(\epsilon)\}} \kappa(u).$$

We have

$$\kappa(u^*)P(S \geq u^*) \leq \int_{u^*}^{\infty} \kappa(s)dP_S(s).$$

Using a similar argument for α_2 we get the desired result (second statement) for the cost function. The second statement follows by similar arguments. Note that there exists a T-measurable function d_1 such that$\mathbb{E}d(S,T) = \mathbb{E}d_1(T)$.

This follows because $\int\int_{\mathsf{A}} d(s,t)dP_{ST}(s,t)$ is a finite measure on A for every $\mathsf{A} \in \mathcal{B}$, and is absolutely continuous with respect to P_T, and d_1 is the corresponding Radon-Nikodym derivative. It follows by similar arguments as in the proof of the statement for the cost function that there exists a sequence of lengths ℓ_m, u_m such that

$$\lim_{m\to\infty} \mathbb{E}d(S, \widetilde{T}_{\ell_m,u_m}) = \mathbb{E}d(S,T).$$

Similarly, it follows that there exists a sequence of lengths ℓ_n, u_n such that

$$\lim_{n\to\infty} \mathbb{E}d(\widetilde{S}_{\ell_n,u_n}, \widetilde{T}_{\ell_m,u_m}) = \mathbb{E}d(S, \widetilde{T}_{\ell_m,u_m}),$$

for every ℓ_m, u_m. This completes the proof. □

Lemma 5 (Smoothing of Random Variables): Consider a bounded continuous random variable U defined on the probability space $([-M,M], \mathcal{B}[-M,M], P_U)$, such that $h(U) < \infty$ and $M > 0$, and let N_ϵ be uniformly distributed over $[-\epsilon,\epsilon], \epsilon > 0$. Assume that U and N_ϵ are independent. Then,

$$\lim_{\epsilon\to 0} I(N_\epsilon; U + N_\epsilon) = 0.$$

Proof: Let $U_\epsilon := U + N_\epsilon$.

$$I(N_\epsilon; U_\epsilon) = h(U_\epsilon) - h(U)$$
$$= \int f_U(u)\log f_U(u)du - \int f_{U_\epsilon}(u)\log f_{U_\epsilon}(u)du.$$

The integral $\int f_{U_\epsilon}(u)\log f_{U_\epsilon}(u)du$ converges to $\int f_U(u))\log f_U(u)du$ as $\epsilon \to 0$ by Fatou's Lemma [53]. To see this, take $g_\epsilon(u) := f_{U_\epsilon}(u)\log f_{U_\epsilon}(u) + 1, u \in [-M-\epsilon, M+\epsilon]$ and $g(x) := f_U(u)\log f_U(u)+1, u \in [-M,M]$, and note that $g_\epsilon(u)$ is non-negative for $u \in [-M-\epsilon, M+\epsilon]$ since $f_{U_\epsilon}(u)\log f_{U_\epsilon}(u) \geq -\frac{1}{e}\log e > -1, u \in \mathbb{R}$. So, by Fatou's lemma $\int_{u\in\mathbb{R}} g(u)du \leq \liminf_{\epsilon\to 0}\int_{u\in\mathbb{R}} g_\epsilon(u)du$ since by construction $g_\epsilon(u)$ converges to $g(u)$ in a pointwise manner almost everywhere as $\epsilon \to 0$. The last statement follows by noting that $f_{U_\epsilon}(\cdot)$ converges to $f_U(\cdot)$ in a pointwise manner almost everywhere as $\epsilon \to 0$ using the assumptions made in Remark 3. This implies that $\int f_U(u)\log f_U(u)du \leq \liminf_{\epsilon\to 0}\int f_{U_\epsilon}(u)\log f_{U_\epsilon}(u)du$. Also, note that $\int f_U(u)\log f_U(u)du \geq \int f_{U_\epsilon}(u) \log f_{U_\epsilon}(u)du$, $\epsilon > 0$ since $I(N_\epsilon; U_\epsilon) \geq 0$. So, $\int f_U(u)\log f_U(u)du \geq \limsup_{\epsilon\to 0}\int f_{U_\epsilon}(u)\log f_{U_\epsilon}(u)du$. Consequently, $I(N_\epsilon; U_\epsilon) \to 0$ as $\epsilon \to 0$. □

Lemma 6: For any quintuple of random variables A, B, C, D and E with a joint distribution that satisfies the Markov chain $(A,B) - C - (DE)$, consider a pair of random variables $\widehat{A}, \widehat{E}$ that are correlated with (B,D) such that $P_{BA} = P_{B\widehat{A}}$, $P_{DE} = P_{D\widehat{E}}$, and $\widehat{A} - B - C - D - \widehat{E}$, then

$$I(A;C|B) + I(E;C|D) \geq \frac{1}{2\ln 2} TV^2(P_{CAE}, P_{C\widehat{A}\widehat{E}}).$$

Proof: We have the following inequalities:

$$I(A;C|B) + I(E;C|D)$$
$$= \sum_{a,b,c,d,e} P_C(c)P_{B|C}(b|c)P_{D|C}(d|c)P_{A|BC}(a|b,c)$$
$$\times P_{E|DC}(e|d,c)\log\frac{P_{A|BC}(a|b,c)P_{E|DC}(e|d,c)}{P_{A|B}(a|b)P_{E|D}(e|d)}$$
$$= \sum_{b,c,d} P_C(c)P_{B|C}(b|c)$$
$$\times P_{D|C}(d|c)D(P_{A|BC}(\cdot|b,c)P_{E|DC}(\cdot|d,c)||P_{A|B}(\cdot|b)P_{E|D}(\cdot|d))$$
$$\overset{(a)}{\geq} \sum_c P_C(c)D(P_{AE|C}(\cdot|c)||P_{\widehat{A},\widehat{E}|C}(\cdot|c))$$
$$\overset{(b)}{\geq} \frac{1}{\ln 2}\sum_c P_C(c)TV^2(P_{A,E|C}(\cdot|c), P_{\widehat{A},\widehat{E}|C}(\cdot|c))$$
$$\overset{(c)}{\geq} \frac{1}{2\ln 2}(\sum_c P_C(c)TV(P_{A,E|C}(\cdot|c), P_{\widehat{A},\widehat{E}|C}(\cdot|c)))^2$$
$$= \frac{1}{2\ln 2}TV^2(P_{A,E,C}, P_{\widehat{A},\widehat{E},C}),$$

where (a) follows from the convexity of relative entropy, (b) follows from Pinsker's inequality and (c) follows from Jensen's inequality. □

Appendix C
Proof of Theorem 1

Proof: Step 1 (Clipping): In this step, we show that for any $\zeta > 0$, and all sufficiently large ℓ, ℓ', we have:

$$|I(\widetilde{U}_\ell + \widetilde{V}_{\ell'}; \widetilde{U}_\ell) - I(U+V;U)| \leq \zeta, \quad (23)$$
$$|I(\widetilde{U}_\ell + \widetilde{V}_{\ell'}; \widetilde{V}_{\ell'}) - I(U+V;V)| \leq \zeta. \quad (24)$$

First, note that $I(\widetilde{U}_\ell + \widetilde{V}_{\ell'}; \widetilde{U}_\ell) - I(U+V;U)$ is bounded from below in the limit as $\ell, \ell' \to \infty$ by 0 due to lower semi-continuity of mutual information. Next, we show that the term is bounded from above by ζ. Define $A_{U,\ell}$ as the indicator of $U \in [-\ell, \ell]$ and $B_{V,\ell'}$ as the indicator of $V \in [-\ell', \ell']$. Fix an arbitrary small positive number η_1. Consider the following arguments:

$$I(\widetilde{U}_\ell; \widetilde{U}_\ell + \widetilde{V}_{\ell'}) \leq I(A_{U,\ell}, \widetilde{U}_\ell; \widetilde{U}_\ell + \widetilde{V}_{\ell'})$$
$$\leq H(A_{U,\ell}) + P(A_{U,\ell} = 0)I(\widetilde{U}_\ell; \widetilde{U}_\ell + \widetilde{V}_{\ell'}|A_{U,\ell} = 0)$$
$$+ P(A_{U,\ell} = 1)I(\widetilde{U}_\ell; \widetilde{U}_\ell + \widetilde{V}_{\ell'}|A_{U,\ell} = 1)$$
$$\leq H(A_{U,\ell}) + P(A_{U,\ell} = 0)I(\widetilde{U}_\ell; \widetilde{U}_\ell + \widetilde{V}_{\ell'}, B_{V,\ell'}|A_{U,\ell} = 0)$$
$$+ P(A_{U,\ell} = 1)I(\widetilde{U}_\ell; \widetilde{U}_\ell + \widetilde{V}_{\ell'}|A_{U,\ell} = 1)$$
$$\leq H(A_{U,\ell}) + H(B_{V,\ell'})$$

$$
\begin{aligned}
&+ P(B_{V,\ell'} = 0|A_{U,\ell} = 0) \\
&\times P(A_{U,\ell} = 0) I(\widetilde{U}_\ell; \widetilde{U}_\ell + \widetilde{V}_{\ell'}|A_{U,\ell} = 0, B_{V,\ell'} = 0) \\
&\quad + P(B_{V,\ell'} = 1|A_{U,\ell} = 0) P(A_{U,\ell} = 0) \\
&\times I(\widetilde{U}_\ell; \widetilde{U}_\ell + \widetilde{V}_{\ell'}|A_{U,\ell} = 0, B_{V,\ell'} = 1) \\
&\quad + P(A_{U,\ell} = 1) I(\widetilde{U}_\ell; \widetilde{U}_\ell + \widetilde{V}_{\ell'}|A_{U,\ell} = 1) \\
&= H(A_{U,\ell}) + H(B_{V,\ell'}) \\
&+ P(B_{V,\ell'} = 0|A_{U,\ell} = 0) P(A_{U,\ell} = 0) \\
&\times I(U'; U' + V'|A_{U,\ell} = 0, B_{V,\ell'} = 0) \\
&\quad + P(B_{V,\ell'} = 1|A_{U,\ell} = 0) P(A_{U,\ell} = 0) \\
&\times I(U'; U' + V|A_{U,\ell} = 0, B_{V,\ell'} = 1) \\
&\quad + P(A_{U,\ell} = 1) I(U; U + \widetilde{V}_{\ell'}|A_{U,\ell} = 1) \\
&\overset{(a)}{\leq} P(B_{V,\ell'} = 0|A_{U,\ell} = 0) P(A_{U,\ell} = 0) \\
&\times I(U'; U' + V'|A_{U,\ell} = 0, B_{V,\ell'} = 0) \\
&\quad + P(B_{V,\ell'} = 1|A_{U,\ell} = 0) P(A_{U,\ell} = 0) \\
&\times I(U'; U' + V|A_{U,\ell} = 0, B_{V,\ell'} = 1) \\
&\quad + P(A_{U,\ell} = 1) \\
&\times I(U; U + \widetilde{V}_{\ell'}|A_{U,\ell} = 1) + 2\eta_1 \\
&\overset{(b)}{\leq} I(U; U + \widetilde{V}_{\ell'}|A_{U,\ell} = 1) + 2\eta_1 + \gamma_1 \\
&\overset{(c)}{\leq} I(A_{U,\ell}; U + \widetilde{V}_{\ell'}) + P(A_{U,\ell} = 1) I(U; U + \widetilde{V}_{\ell'}|A_{U,\ell} = 1) \\
&+ P(A_{U,\ell} = 0) I(U; U + \widetilde{V}_{\ell'}|A_{U,\ell} = 0) + 3\eta_1 + \gamma_1 \\
&= I(U; U + \widetilde{V}_{\ell'}) + 2\eta_1 + \gamma_1,
\end{aligned}
$$

where in (a) we have taken ℓ and ℓ' large enough such that $H(A_{U,\ell}) < \eta_1$ and $H(B_{V,\ell'}) < \eta_1$, in (b), we have defined

$$
\begin{aligned}
\gamma_1 :&= P(A_{U,\ell} = 0, B_{V,\ell'} = 0) \\
&\times I(U'; U' + V'|A_{U,\ell} = 0, B_{V,\ell'} = 0) \\
&\quad + P(A_{U,\ell} = 0, B_{V,\ell'} = 1) \\
&\times I(U'; U' + V|A_{U,\ell} = 0, B_{V,\ell'} = 1).
\end{aligned}
$$

and in (c), we have taken ℓ, ℓ' large enough such that $P(A_{U,\ell} = 0) I(U; U + \widetilde{V}_{\ell'}|A_{U,\ell} = 1) < \eta_1$. We show in the following that such ℓ, ℓ' always exists. It suffices to show that $I(U; U + \widetilde{V}_{\ell'}|A_{U,\ell} = 1) < \infty$:

$$
\begin{aligned}
&I(U; U + \widetilde{V}_{\ell'}|A_{U,\ell} = 1) \leq I(U; B_{V,\ell'}, U + \widetilde{V}_{\ell'}|A_{U,\ell} = 1) \\
&\leq H(B_{V,\ell'}) + P(B_{V,\ell'} = 1|A_{U,\ell} = 1) \\
&\times I(U; U + \widetilde{V}_{\ell'}|A_{U,\ell} = 1, B_{V,\ell'} = 1) \\
&\quad + P(B_{V,\ell'} = 0|A_{U,\ell} = 1) \\
&\times I(U; U + \widetilde{V}_{\ell'}|A_{U,\ell} = 1, B_{V,\ell'} = 0) \\
&= H(B_{V,\ell'}) + P(B_{V,\ell'} = 1|A_{U,\ell} = 1) \\
&\times I(U; U + V|A_{U,\ell} = 1, B_{V,\ell'} = 1) \\
&\quad + P(B_{V,\ell'} = 0|A_{U,\ell} = 1) \\
&\times I(U; U + V'|A_{U,\ell} = 1, B_{V,\ell'} = 0).
\end{aligned}
$$

Note that $I(U; U + V|A_{U,\ell} = 1, B_{V,\ell'} = 1) < \infty$, since

$$
\begin{aligned}
&\infty > I(U; U + V) \geq I(U; U + V|A_{U,\ell}, B_{U,\ell}) \\
&- H(A_{U,\ell}) - H(B_{V,\ell'}) \\
&\geq I(U; U + V|A_{U,\ell}, B_{U,\ell}) - 2\eta_1 \\
&\geq P(A_{U,\ell} = 1, B_{U,\ell} = 1) I(U; U + V|A_{U,\ell} = 1, B_{U,\ell} = 1) \\
&\qquad - 2\eta_1 \\
&\geq \frac{1}{2} I(U; U + V|A_{U,\ell} = 1, B_{U,\ell} = 1) - 2\eta_1,
\end{aligned}
$$

for large enough ℓ and ℓ'. Furthermore, we have using the entropy power inequality [26],

$$
\begin{aligned}
&I(U; U + V'|A_{U,\ell} = 1, B_{V,\ell'} = 0) \\
&= h(U + V'|A_{U,\ell} = 1, B_{V,\ell'} = 0) \\
&\qquad - h(V'|A_{U,\ell} = 1, B_{V,\ell'} = 0) \\
&= h(U + V'|A_{U,\ell} = 1, B_{V,\ell'} = 0) - h(V') \\
&\leq \frac{1}{2} \log 2\pi e \mathrm{Var}(U + V'|A_{U,\ell} = 1, B_{V,\ell'} = 0) \\
&\qquad + \frac{1}{2} |\log 2\pi e \mathrm{Var}(V')| \\
&\overset{(a)}{\leq} \frac{1}{2} \log 2\pi e \frac{1}{P(A_{U,\ell} = 1, B_{V,\ell'} = 0)} \mathrm{Var}(U + V') \\
&\qquad + \frac{1}{2} |\log 2\pi e \mathrm{Var}(V')| \\
&\leq \frac{1}{2} \log 2\pi e \frac{1}{P(A_{U,\ell} = 1, B_{V,\ell'} = 0)} (\mathrm{Var}(U) + \mathrm{Var}(V')) \\
&\qquad + \frac{1}{2} |\log 2\pi e \mathrm{Var}(V')| \\
&\overset{(b)}{\leq} \frac{1}{2} \log 2\pi e \frac{1}{P(A_{U,\ell} = 1, B_{V,\ell'} = 0)} \times \\
&(\mathrm{Var}(U) + \frac{\mathrm{Var}(V)}{P(B_{V,\ell'} = 1)}) + \frac{1}{2} |\log 2\pi e \frac{\mathrm{Var}(V)}{P(B_{V,\ell'} = 1)}|, \\
&\leq \frac{1}{2} \log 2\pi e \frac{1}{P(A_{U,\ell} = 1, B_{V,\ell'} = 0)} (\mathrm{Var}(U) + \frac{\mathrm{Var}(V)}{1 - \eta_1}) \\
&\quad + \frac{1}{2} |\log 2\pi e \frac{\mathrm{Var}(V)}{1 - \eta_1}|,
\end{aligned}
$$

where in (a) and (b) we have used the law of total variance. As a result,

$$
P(B_{V,\ell'} = 0|A_{U,\ell} = 1) I(U; U + V'|A_{U,\ell} = 1, B_{V,\ell'} = 0) < \infty,
$$

for sufficiently large ℓ, ℓ', since $\mathrm{Var}(U), \mathrm{Var}(V) < \infty$ by assumption as explained in Remark 3.

Next, we show that γ_1 can be made arbitrarily small for large enough ℓ, ℓ'. We consider the first mutual information term in γ_1 as follows.

$$
\begin{aligned}
&I(U'; U' + V'|A_{U,\ell} = 0, B_{U,\ell'} = 0) \\
&= I(U'; U' + V') = h(U' + V') - h(V')
\end{aligned}
$$

We have:

$$
\begin{aligned}
&h(V') = h(V|B_{V,\ell'} = 1) \\
&= \frac{1}{P(B_{V,\ell'} = 1)} (h(V) - P(B_{V,\ell'} = 0) h(V|B_{V,\ell'} = 0) \\
&\qquad - H(B_{V,\ell'})),
\end{aligned}
$$

and $h(V|B_{V,\ell'} = 0)$ can be bounded from above as follows:

$$
\begin{aligned}
&h(V|B_{V,\ell'} = 0) \leq \frac{1}{2} \log 2\pi e \mathrm{Var}(V|B_{V,\ell'} = 0) \\
&\overset{(a)}{\leq} \frac{1}{2} \log 2\pi e \frac{1}{P(B_{V,\ell'} = 0)} \mathrm{Var}(V) \\
&\leq \frac{1}{2} \log 2\pi e \frac{1}{1 - \eta_1} \mathrm{Var}(V),
\end{aligned}
$$

where in (a) we have used the law of total variance. This implies that $h(V') \geq h(V) - \eta_1$ as $\ell \to \infty$. So, $h(V')$ is bounded from below. Next, we bound $h(U'+V')$ from above:

$$h(U' + V') \leq \frac{1}{2} \log 2\pi e \mathrm{Var}(U' + V'),$$

and using law of total variance, we have:

$$\begin{aligned}
&\mathrm{Var}(U' + V') = \mathrm{Var}(U') + \mathrm{Var}(V') \\
&= \mathrm{Var}(U|A_{U,\ell} = 1) + \mathrm{Var}(V|B_{V,\ell'} = 1) \\
&\leq \frac{1}{P(A_{U,\ell} = 1)}\mathrm{Var}(U) + \frac{1}{P(B_{V,\ell'} = 1)}\mathrm{Var}(V) \\
&\leq \frac{1}{1-\eta_1}(\mathrm{Var}(U) + \mathrm{Var}(V)).
\end{aligned}$$

As a result, we have

$$\begin{aligned}
&P(A_{U,\ell} = 0, B_{V,\ell'} = 0)I(U'; U' + V'|A_{U,\ell} = 0, B_{V,\ell'} = 0) \\
&\leq P(A_{U,\ell} = 0, B_{V,\ell'} = 0)\times \\
&\left(\frac{1}{2}\log 2\pi e \frac{1}{1-\eta_1}\left(\mathrm{Var}(U) + \mathrm{Var}(V)\right) - h(V)\right) \to 0
\end{aligned}$$

as $\ell \to \infty$.

Next, we consider the second mutual information term $I(U'; U' + V|A_{U,\ell} = 0, B_{V,\ell'} = 1)$ in γ_1. Note that:

$$\begin{aligned}
&h(U' + V|A_{U,\ell} = 0, B_{V,\ell'} = 1) \\
&\leq \frac{1}{2}\log 2\pi e \mathrm{Var}(U' + V|A_{U,\ell} = 0, B_{V,\ell'} = 1) \\
&= \frac{1}{2}\log 2\pi e (\mathrm{Var}(U') + \mathrm{Var}(V|A_{U,\ell} = 0, B_{V,\ell'} = 1)) \\
&= \frac{1}{2}\log 2\pi e (\mathrm{Var}(U|A_{U,\ell} = 1) + \mathrm{Var}(V|A_{U,\ell} = 0, B_{V,\ell'} = 1)) \\
&\leq \frac{1}{2}\log 2\pi e \left(\frac{\mathrm{Var}(U)}{P(A_{U,\ell} = 1)} + \frac{\mathrm{Var}(V)}{P(A_{U,\ell} = 0, B_{V,\ell'} = 1)}\right).
\end{aligned}$$

So, $P(A_{U,\ell} = 0, B_{V,\ell'} = 1)h(U'+V|A_{U,\ell} = 0, B_{V,\ell'} = 1) \to 0$ as $\ell, \ell' \to \infty$. Also, $P(A_{U,\ell} = 0, B_{V,\ell'} = 1)h(V|A_{U,\ell} = 0, B_{V,\ell'} = 1) \to 0$ as $\ell, \ell' \to \infty$ following similar arguments as above. Hence, $P(A_{U,\ell} = 0, B_{V,\ell'} = 1)I(U'; U'+V|A_{U,\ell} = 0, B_{V,\ell'} = 1) \to 0$ as $\ell, \ell' \to \infty$ and consequently $\gamma_1 \to 0$ as $\ell, \ell' \to \infty$.

Next we focus on $I(U; U + \widetilde{V}_{\ell'})$:

$$\begin{aligned}
&I(U; U + \widetilde{V}_{\ell'}) \leq I(U; U + \widetilde{V}_{\ell'}, B_{V,\ell'}) \\
&\leq H(B_{V,\ell'}) + P(B_{V,\ell'} = 0)I(U; U + V'|B_{V,\ell'} = 0) \\
&+ P(B_{V,\ell'} = 1)I(U; U + \widetilde{V}_{\ell'}|B_{V,\ell'} = 1) \\
&\leq P(B_{V,\ell'} = 1)I(U; U + \widetilde{V}_{\ell'}|B_{V,\ell'} = 1) + \gamma_2
\end{aligned}$$

where

$$\gamma_2 := H(B_{V,\ell'}) + P(B_{V,\ell'} = 0)I(U; U + V'|B_{V,\ell'} = 0).$$

Note that:

$$\begin{aligned}
&h(U + V'|B_{V,\ell'} = 0) \leq \frac{1}{2}\log 2\pi e \mathrm{Var}(U + V'|B_{V,\ell'} = 0) \\
&\leq \frac{1}{2}\log 2\pi e (\mathrm{Var}(U|B_{V,\ell'} = 0) + \mathrm{Var}(V'|B_{V,\ell'} = 0)) \\
&\stackrel{(a)}{\leq} \frac{1}{2}\log 2\pi e \left(\frac{\mathrm{Var}(U)}{P(B_{V,\ell'} = 0)} + \frac{\mathrm{Var(V)}}{P(B_{V,\ell'} = 1)}\right) \\
&\leq \frac{1}{2}\log 2\pi e \left(\frac{\mathrm{Var}(U)}{P(B_{V,\ell'} = 0)} + \frac{\mathrm{Var(V)}}{1-\eta_1}\right),
\end{aligned}$$

where in (a) we have used the fact that by construction V' has PDF $f_{V|B_{V,\ell'}=1}$ and used the law of total variance to conclude that

$$\begin{aligned}
&\mathrm{Var}(V'|B_{V,\ell'} = 0) = \mathrm{Var}(V') = \mathrm{Var}(V|B_{V,\ell'} = 1) \\
&\leq \frac{\mathrm{Var}(V)}{P(B_{V,\ell'} = 1)} \leq \frac{\mathrm{Var}(V)}{1-\eta_1},
\end{aligned}$$

and that

$$\begin{aligned}
&\mathrm{Var}(U) \geq \mathbb{E}(\mathrm{Var}(U|B_{V,\ell'})) \geq P(B_{V,\ell'} = 0)\mathrm{Var}(U|B_{V,\ell'} = 0) \\
&\Rightarrow \frac{\mathrm{Var}(U)}{P(B_{V,\ell'} = 0)} \geq \mathrm{Var}(U|B_{V,\ell'} = 0).
\end{aligned}$$

So, $P(B_{V,\ell'} = 0)h(U + V'|B_{V,\ell'} = 0) \to 0$ as $\ell, \ell' \to \infty$, and in turn using the arguments used above regarding $h(V')$ we infer that $\gamma_2 \to 0$ as $\ell, \ell' \to \infty$.

Next, consider $P(B_{V,\ell'} = 1)I(U; U + \widetilde{V}_{\ell'}|B_{V,\ell'} = 1)$ as follows:

$$\begin{aligned}
&P(B_{V,\ell'} = 1)I(U; U + \widetilde{V}_{\ell'}|B_{V,\ell'} = 1) \\
&= P(B_{V,\ell'} = 1)I(U; U + V|B_{V,\ell'} = 1) \\
&\leq I(U; U + V, B_{V,\ell'}) \\
&\leq I(U; U + V) + H(B_{V,\ell'}) \\
&\leq I(U; U + V) + \eta_1.
\end{aligned}$$

Step 2 (Smoothing): In this step, we show that for all $\gamma > 0$, and all sufficiently small $\epsilon > 0$, we have:

$$|I(\widetilde{U}_{\ell,\epsilon} + \widetilde{V}_{\ell,\epsilon}; \widetilde{U}_{\ell,\epsilon}) - I(\widetilde{U}_\ell + \widetilde{V}_\ell; \widetilde{U}_\ell)| \leq \gamma, \quad (25)$$

$$|I(\widetilde{U}_{\ell,\epsilon} + \widetilde{V}_{\ell,\epsilon}; \widetilde{V}_{\ell,\epsilon}) - I(\widetilde{U}_\ell + \widetilde{V}_\ell; \widetilde{V}_\ell)| \leq \gamma. \quad (26)$$

We argue that the CDF $F_{X,Y,\widetilde{U}_{\ell,\epsilon},\widetilde{V}_{\ell,\epsilon}} \to F_{X,Y,\widetilde{U}_\ell,\widetilde{V}_\ell}$ as $\epsilon \to 0$. We show convergence for $F_{\widetilde{U}_{\ell,\epsilon}}$. The convergence for the joint distribution follows by similar arguments. To show this, let $u \in [-\ell, \ell]$ be a point of continuity of $F_{\widetilde{U}_\ell}(\cdot)$. Then, by construction, we have $F_{\widetilde{U}_{\ell,\epsilon}}(u) \to F_{\widetilde{U}_\ell}(u)$ as $\epsilon \to 0$ as shown below:

$$\begin{aligned}
&P(\widetilde{U}_\ell \leq u - \epsilon) \leq P(\widetilde{U}_{\ell,\epsilon} \leq u) \leq P(\widetilde{U}_\ell \leq u + \epsilon) \\
&\Rightarrow P(\widetilde{U}_\ell \leq u - \epsilon) - F_{\widetilde{U}_\ell}(u) \\
&\leq P(\widetilde{U}_{\ell,\epsilon} \leq u) - F_{\widetilde{U}_\ell}(u) \\
&\leq P(\widetilde{U}_\ell \leq u + \epsilon) - F_{\widetilde{U}_\ell}(u) \\
&\Rightarrow P(\widetilde{U}_\ell \leq u - \epsilon) - F_{\widetilde{U}_\ell}(u + \epsilon) \leq P(\widetilde{U}_{\ell,\epsilon} \leq u) - F_{\widetilde{U}_\ell}(u) \\
&\leq P(\widetilde{U}_\ell \leq u + \epsilon) - F_{\widetilde{U}_\ell}(u - \epsilon) \\
&\Rightarrow |F_{\widetilde{U}_{\ell,\epsilon}}(u) - F_{\widetilde{U}_\ell}(u)| \leq P(u - \epsilon < \widetilde{U}_\ell \leq u + \epsilon).
\end{aligned}$$

Towards showing (25), note that

$$\begin{aligned}
&I(\widetilde{U}_\ell; \widetilde{U}_\ell + \widetilde{V}_\ell) = I(\widetilde{U}_\ell; \widetilde{U}_\ell + \widetilde{V}_\ell|\widetilde{N}_{\ell,\epsilon}, \widetilde{N}'_{\ell,\epsilon}) \\
&= I(\widetilde{U}_{\ell,\epsilon}; \widetilde{U}_{\ell,\epsilon} + \widetilde{V}_{\ell,\epsilon}|\widetilde{N}_{\ell,\epsilon}, \widetilde{N}'_{\ell,\epsilon}) \\
&= I(\widetilde{U}_{\ell,\epsilon}, \widetilde{N}_{\ell,\epsilon}, \widetilde{N}'_{\ell,\epsilon}; \widetilde{U}_{\ell,\epsilon} + \widetilde{V}_{\ell,\epsilon}) - I(\widetilde{N}_{\ell,\epsilon}, \widetilde{N}'_{\ell,\epsilon}; \widetilde{U}_{\ell,\epsilon} + \widetilde{V}_{\ell,\epsilon}) \\
&\geq I(\widetilde{U}_{\ell,\epsilon}, \widetilde{N}_{\ell,\epsilon}, \widetilde{N}'_{\ell,\epsilon}; \widetilde{U}_{\ell,\epsilon} + \widetilde{V}_{\ell,\epsilon}) - I(\widetilde{N}_{\ell,\epsilon}, \widetilde{N}'_{\ell,\epsilon}; \widetilde{U}_{\ell,\epsilon}, \widetilde{V}_{\ell,\epsilon}) \\
&= I(\widetilde{U}_{\ell,\epsilon}, \widetilde{N}_{\ell,\epsilon}, \widetilde{N}'_{\ell,\epsilon}; \widetilde{U}_{\ell,\epsilon} + \widetilde{V}_{\ell,\epsilon}) \\
&- I(\widetilde{N}_{\ell,\epsilon}, \widetilde{N}'_{\ell,\epsilon}; \widetilde{U}_{\ell,\epsilon}) - I(\widetilde{N}_{\ell,\epsilon}, \widetilde{N}'_{\ell,\epsilon}; \widetilde{V}_{\ell,\epsilon}|\widetilde{U}_{\ell,\epsilon}) \\
&= I(\widetilde{U}_{\ell,\epsilon}, \widetilde{N}_{\ell,\epsilon}, \widetilde{N}'_{\ell,\epsilon}; \widetilde{U}_{\ell,\epsilon} + \widetilde{V}_{\ell,\epsilon})
\end{aligned}$$

$$
\begin{aligned}
&- I(\widetilde{N}_{\ell,\epsilon};\widetilde{U}_{\ell,\epsilon}) - I(\widetilde{N}_{\ell,\epsilon},\widetilde{N}'_{\ell,\epsilon};\widetilde{V}_{\ell,\epsilon}|\widetilde{U}_{\ell,\epsilon})\\
&= I(\widetilde{U}_{\ell,\epsilon};\widetilde{U}_{\ell,\epsilon}+\widetilde{V}_{\ell,\epsilon}) + I(\widetilde{N}_{\ell,\epsilon},\widetilde{N}'_{\ell,\epsilon};\widetilde{U}_{\ell,\epsilon}+\widetilde{V}_{\ell,\epsilon}|\widetilde{U}_{\ell,\epsilon})\\
&- I(\widetilde{N}_{\ell,\epsilon};\widetilde{U}_{\ell,\epsilon}) - I(\widetilde{N}_{\ell,\epsilon},\widetilde{N}'_{\ell,\epsilon};\widetilde{V}_{\ell,\epsilon}|\widetilde{U}_{\ell,\epsilon})\\
&= I(\widetilde{U}_{\ell,\epsilon};\widetilde{U}_{\ell,\epsilon}+\widetilde{V}_{\ell,\epsilon}) - I(\widetilde{N}_{\ell,\epsilon};\widetilde{U}_{\ell,\epsilon}),
\end{aligned}
$$

and $I(\widetilde{N}_{\ell,\epsilon};\widetilde{U}_{\ell,\epsilon}) \to 0$ as $\epsilon \to 0$ by Lemma 5. The proof of (25) follows by lower semi-continuity of mutual information. Equation (26) can be proved using a similar argument. For later convenience, note that the joint PDF of $(\widetilde{U}_{\ell,\epsilon},\widetilde{V}_{\ell,\epsilon})$ is jointly continuous on a compact support, and hence uniformly continuous.

Step 3 (Quantization): In this step, we discretize $\widetilde{U}_{\ell,\epsilon},\widetilde{V}_{\ell,\epsilon}$ to $\widehat{U}_{\ell,\epsilon,n},\widehat{V}_{\ell,\epsilon,n}$ by applying $Q_n(\cdot)$. We show that for any $\gamma>0$, and all sufficiently large $n\in\mathbb{N}$, the following hold:

$$|I(\widehat{U}_{\ell,\epsilon,n}+\widehat{V}_{\ell,\epsilon,n};\widehat{U}_{\ell,\epsilon,n}) - I(\widetilde{U}_{\ell,\epsilon}+\widetilde{V}_{\ell,\epsilon};\widetilde{U}_{\ell,\epsilon})| \le \gamma, \quad (27)$$

$$|I(\widehat{U}_{\ell,\epsilon,n}+\widehat{V}_{\ell,\epsilon,n};\widehat{V}_{\ell,\epsilon,n}) - I(\widetilde{U}_{\ell,\epsilon}+\widetilde{V}_{\ell,\epsilon};\widetilde{V}_{\ell,\epsilon})| \le \gamma. \quad (28)$$

We will show Equation (27) using the approach taken in [20] to study entropy of linear combinations of independent continuous variables. The proof of (28) follows by a similar argument. We drop the subscript on $Q_n(\cdot)$ when there is no ambiguity.

Define $\mod_Q(\widetilde{U}_{\ell,\epsilon}) := \widetilde{U}_{\ell,\epsilon} - \widehat{U}_{\ell,\epsilon,n}$ and $\mod_Q(\widetilde{V}_{\ell,\epsilon}) := \widetilde{V}_{\ell,\epsilon} - \widehat{V}_{\ell,\epsilon,n}$, and the variables $C := Q(\widetilde{U}_{\ell,\epsilon}+\widetilde{V}_{\ell,\epsilon})$, $D := \widehat{U}_{\ell,\epsilon,n}+\widehat{V}_{\ell,\epsilon,n}$, $E = Q(\mod_Q(\widetilde{U}_{\ell,\epsilon}) + \mod_Q(\widetilde{V}_{\ell,\epsilon}))$. Note that $E\in\{\frac{-1}{N},0,\frac{1}{N}\}$ by construction, where $N := 2^n$. We will show that i) $H(C)-H(D)\to 0$ as $N\to\infty$, and ii) $H(\widehat{U}_{\ell,\epsilon,n},C) - H(\widehat{U}_{\ell,\epsilon,n},D) \to 0$ as $N\to\infty$. This implies that $I(\widehat{U}_{\ell,\epsilon,n};\widehat{U}_{\ell,\epsilon,n}+\widehat{V}_{\ell,\epsilon,n}) - I(\widehat{U}_{\ell,\epsilon,n};Q(\widetilde{U}_{\ell,\epsilon}+\widetilde{V}_{\ell,\epsilon})) \to 0$ as $n\to\infty$. Consequently, by data processing inequality and lower semi-continuity of mutual information, we have $I(\widehat{U}_{\ell,\epsilon,n};\widehat{U}_{\ell,\epsilon,n}+\widehat{V}_{\ell,\epsilon,n}) - I(\widetilde{U}_{\ell,\epsilon};\widetilde{U}_{\ell,\epsilon}+\widetilde{V}_{\ell,\epsilon}) \to 0$.

First, we will show that $H(C)-H(D)\to 0$ as $N\to\infty$. Note that $C = D+E$ using the distributive property of lattices [54]. As a result,

$$H(C)-H(D) = I(C;E) - I(D;E).$$

First, let us consider $I(D;E)$ as follows:

$$
\begin{aligned}
I(D;E) &= I(\widehat{U}_{\ell,\epsilon,n}+\widehat{V}_{\ell,\epsilon,n};Q(\mod_Q(\widetilde{U}_{\ell,\epsilon})+\mod_Q(\widetilde{V}_{\ell,\epsilon})))\\
&\le I(\widehat{U}_{\ell,\epsilon,n}+\widehat{V}_{\ell,\epsilon,n};\ \mod_Q(\widetilde{U}_{\ell,\epsilon})+\mod_Q(\widetilde{V}_{\ell,\epsilon}))\\
&\le I(\widehat{U}_{\ell,\epsilon,n},\widehat{V}_{\ell,\epsilon,n};\ \mod_Q(\widetilde{U}_{\ell,\epsilon}),\ \mod_Q(\widetilde{V}_{\ell,\epsilon})),
\end{aligned}
$$

which goes to 0 as $N\to\infty$ using Lemma 5 in [20]. Next, let us consider $I(C;E)$. Using Proposition 12 in [55], we have:

$$I(C;E) \le (\log 3 - 1)T(E;C) + h_b(T(E;C)),$$

where h_b is the binary entropy function, and $T(E;C) := TV(P_{E,C},P_EP_C)$, where $P_{E,C}$ is the probability mass function of the pair (E,C). So, it suffices to show that $T(E;C)\to 0$ as $N\to\infty$. Note that

$$
\begin{aligned}
&TV(P_{E,C},P_EP_C)\\
&= \sum_{e\in\{\frac{-1}{N},0,\frac{1}{N}\}} P(E=e)TV(P_C,P_{C|E}(\cdot|e))\\
&\le \sum_{e\in\{\frac{-1}{N},0,\frac{1}{N}\}} TV(P_C,P_{C|E}(\cdot|e))\\
&\le \sum_{e,e'\in\{\frac{-1}{N},0,\frac{1}{N}\}} TV(P_{C|E}(\cdot|e'),P_{C|E}(\cdot|e))\\
&= \sum_{e,e'\in\{\frac{-1}{N},0,\frac{1}{N}\}} \sum_d |P_{D|E}(d-e'|e') - P_{D|E}(d-e|e)|\\
&\le \sum_{e,e'\in\{\frac{-1}{N},0,\frac{1}{N}\}} \sum_d |P_{D|E}(d-e'|e') - P_{D|E}(d-e'|e)|\\
&+ |P_{D|E}(d-e'|e) - P_{D|E}(d-e|e)|\\
&= \sum_{e,e'\in\{\frac{-1}{N},0,\frac{1}{N}\}} \sum_d |P_{D|E}(d|e') - P_{D|E}(d|e)|\\
&+ |P_{D|E}(d-e'|e) - P_{D|E}(d-e|e)|. \qquad (29)
\end{aligned}
$$

We investigate the first term in the summation in Equation (29).

$$
\begin{aligned}
&\sum_{e,e'\in\{\frac{-1}{N},0,\frac{1}{N}\}} \sum_d |P_{D|E}(d|e') - P_{D|E}(d|e)|\\
&\le \sum_{e,e'\in\{\frac{-1}{N},0,\frac{1}{N}\}} \sum_d |P_{D|E}(d|e') - P_D(d)|\\
&+ |P_D(d) - P_{D|E}(d|e)|,
\end{aligned}
$$

which goes to 0 as $N\to\infty$ due to Pinsker's inequality. To see this, fix $\eta>0$, and let N be large enough so that $I(D;E)<\eta$. Note that such N exists since $\lim_{N\to\infty} I(D;E)=0$ as shown above. Due to Pinsker's inequality, we have:

$$
\begin{aligned}
\eta \ge I(D;E) &= \sum_e P(E=e)D(P_{D|E}(\cdot|e)||P_D)\\
&\ge 2(\ln 2)\sum_e P(E=e)TV^2(P_{D|E}(\cdot|e),P_D)\\
&\ge 2(\ln 2)P(E=e')TV^2(P_{D|E}(\cdot|e'),P_D),
\end{aligned}
$$

for all $e'\in\{\frac{-1}{N},0,\frac{1}{N}\}$. Furthermore, we show that $|P(E=0)-\frac{3}{4}|\to 0$ and $|P(E=\frac{1}{N})-\frac{1}{8}|\to 0$, and $|P(E=\frac{-1}{N})-\frac{1}{8}|\to 0$ as $N\to\infty$:

$$
\begin{aligned}
P(E=\frac{1}{N}) &= \int_{u,v:Q(\mod_Q(u)+\mod_Q(v)))=\frac{1}{N}} f_{\widetilde{U}_\ell,\widetilde{V}_\ell}(u,v)dudv\\
&= \sum_{i,j=1}^{N}\int_{u,v\in E_i\times E_j:Q(\mod_Q(u)+\mod_Q(v)))=\frac{1}{N}} f_{\widetilde{U}_{\ell,\epsilon},\widetilde{V}_{\ell,\epsilon}}(u,v)dudv\\
&= \sum_{i=1}^{N}\sum_{j=1}^{N}\int_{u,v\in E_i\times E_j:u+v>e_i+e_j+\frac{1}{2N}} f_{\widetilde{U}_{\ell,\epsilon},\widetilde{V}_{\ell,\epsilon}}(u,v)dudv\\
&\le \sum_{i=1}^{N}\sum_{j=1}^{N}\int_{u,v\in E_i\times E_j:u+v<e_i+e_j-\frac{1}{2N}} f_{\widetilde{U}_{\ell,\epsilon},\widetilde{V}_{\ell,\epsilon}}(u,v)dudv+\delta_N\\
&= P(E=-\frac{1}{N})+\delta_N,
\end{aligned}
$$

where in the last inequality we have used the fact that $f_{\widetilde{U}_{\ell,\epsilon},\widetilde{V}_{\ell,\epsilon}}$ is continuous over a compact support, and hence uniformly continuous, to argue the existence of δ_N such that $\delta_N\to 0$ as $N\to 0$. Similarly $P(E=\frac{-1}{N}) \le P(E=\frac{1}{N})+\delta_N$. Furthermore,

$$P(E=0) = \int_{u,v:Q(\mod_Q(u)+\mod_Q(v)))=0} f_{\widetilde{U}_{\ell,\epsilon},\widetilde{V}_{\ell,\epsilon}}(u,v)dudv$$

$$= \sum_{i,j=1}^{N} \int_{u,v \in E_i \times E_j : Q(\text{mod}_Q(u) + \text{mod}_Q(v))) = 0} f_{\widetilde{U}_{\ell,\epsilon},\widetilde{V}_{\ell,\epsilon}}(u,v) du dv$$
$$= \sum_{i,j=1}^{N} \int_{u,v \in E_i \times E_j : e_i + e_j + \frac{-1}{2N} < u+v < e_i+e_j+\frac{1}{2N}} f_{\widetilde{U}_{\ell,\epsilon},\widetilde{V}_{\ell,\epsilon}}(u,v) du dv$$
$$\leq 6 \sum_{i,j=1}^{N} \int_{u,v \in E_i \times E_j : u+v < e_i+e_j - \frac{1}{2N}} f_{\widetilde{U}_{\ell,\epsilon},\widetilde{V}_{\ell,\epsilon}}(u,v) du dv + \delta_N$$
$$= 6P(E = \frac{-1}{N}) + \delta_N.$$

Similarly, $6P(E = \frac{-1}{N}) \leq P(E = 0) + \delta_N$. So, $|P(E = 0) - \frac{3}{4}| \leq \delta_N$ and $|P(E = \frac{1}{N}) - \frac{1}{8}| \leq \delta_N$, and $|P(E = \frac{-1}{N}) - \frac{1}{8}| \leq \delta_N$. Consequently, $TV(P_{D|E}(\cdot|e'), P_D) \to 0$ as $N \to \infty$ for all $e' \in \{\frac{-1}{N}, 0, \frac{1}{N}\}$.

Next, we consider the second term in Equation (29). Fix $e, e' \in \{\frac{-1}{N}, 0, \frac{1}{N}\}$. Note that

$$\begin{aligned} &D + e - e' \\ &= Q(\widetilde{U}_{\ell,\epsilon}) + Q(\widetilde{V}_{\ell,\epsilon}) + e - e' \\ &= Q(\widetilde{U}_{\ell,\epsilon} + e) + Q(\widetilde{V}_{\ell,\epsilon} - e'), \end{aligned}$$

So,

$$\begin{aligned} &\sum_d |P_{D|E}(d - e'|e) - P_{D|E}(d - e|e)| \\ &= \sum_d |P_{D|E}(d + e - e'|e) - P_{D|E}(d|e)| \\ &= \sum_d |P(Q(\widetilde{U}_{\ell,\epsilon} + e) + Q(\widetilde{V}_{\ell,\epsilon} - e') = d | E = e) \\ &- P(Q(\widetilde{U}_{\ell,\epsilon}) + Q(\widetilde{V}_{\ell,\epsilon}) = d) | E = e)| \end{aligned} \tag{30}$$

$$\begin{aligned} &\overset{(a)}{\leq} 2 \sup_{\mathsf{A}} |P_{\widetilde{U}_{\ell,\epsilon},\widetilde{V}_{\ell,\epsilon}|E}(\mathsf{A} + \{(e, -e')\}|e), - P_{\widetilde{U}_{\ell,\epsilon},\widetilde{V}_{\ell,\epsilon}|E}(\mathsf{A}|e)| \\ &= 2 \sup_{\mathsf{A}} \frac{1}{P(E = e)} |P_{\widetilde{U}_{\ell,\epsilon},\widetilde{V}_{\ell,\epsilon}}((\mathsf{A} + \{(e, -e')\}) \cap \{E = e\}) \\ &- P_{\widetilde{U}_{\ell,\epsilon},\widetilde{V}_{\ell,\epsilon}}(\mathsf{A} \cap \{E = e\})| \\ &\overset{(b)}{\leq} 2 \sup_{\mathsf{A}} \frac{1}{P(E = e)} |P_{\widetilde{U}_{\ell,\epsilon},\widetilde{V}_{\ell,\epsilon}}(\mathsf{A} + \{(e, -e')\}) - P_{\widetilde{U}_{\ell,\epsilon},\widetilde{V}_{\ell,\epsilon}}(\mathsf{A})| \\ &= \frac{1}{P(E = e)} \\ &\times \int_{u,v} |f_{\widetilde{U}_{\ell,\epsilon},\widetilde{V}_{\ell,\epsilon}}(u+e, v-e') - f_{\widetilde{U}_{\ell,\epsilon},\widetilde{V}_{\ell,\epsilon}}(u,v)| du dv, \end{aligned} \tag{31}$$

where in (a) we have used the data processing inequality for variational distance, and in (b) we have used the fact that the supremum over A is larger than that over $\mathsf{A} \cap \{E = e\}$. Note that the last term goes to 0 as $N \to \infty$ due to uniform continuity of $f_{\widetilde{U}_{\ell,\epsilon},\widetilde{V}_{\ell,\epsilon}}$.

We have thus shown that $T(C; E) \to 0$ as $N \to \infty$, and hence $I(C; E) \to 0$, and consequently, $H(C) \to H(D)$ as $N \to \infty$. Next, we will show that $H(\widehat{U}_{\ell,\epsilon,n}, C) - H(\widehat{U}_{\ell,\epsilon,n}, D) \to 0$. Similar to the previous part, we have:

$$\begin{aligned} &H(\widehat{U}_{\ell,\epsilon,n}, C) - H(\widehat{U}_{\ell,\epsilon,n}, D) \\ &= I(\widehat{U}_{\ell,\epsilon,n}, C; E) - I(\widehat{U}_{\ell,\epsilon,n}, D; E). \end{aligned}$$

The second term $I(\widehat{U}_{\ell,\epsilon,n}, D; E)$ goes to 0 as $N \to \infty$ by a similar argument as in the previous case. For the first term, similarly it suffices to show that $TV(P_{E,C,\widehat{U}_{\ell,\epsilon,n}}, P_E P_{C,\widehat{U}_{\ell,\epsilon,n}}) \to 0$ as $N \to \infty$. We have:

$$\begin{aligned} &TV(P_{E,C,\widehat{U}_{\ell,\epsilon,n}}, P_E P_{C,\widehat{U}_{\ell,\epsilon,n}}) \\ &= \sum_{e \in \{\frac{-1}{N}, 0, \frac{1}{N}\}} P(E = e) TV(P_{C,\widehat{U}_{\ell,\epsilon,n}}, P_{C,\widehat{U}_{\ell,\epsilon,n}|E}(\cdot|e)) \\ &\leq \sum_{e,e' \in \{\frac{-1}{N}, 0, \frac{1}{N}\}} TV(P_{C,\widehat{U}_{\ell,\epsilon,n}|E}(\cdot|e'), P_{C,\widehat{U}_{\ell,\epsilon,n}|E}(\cdot|e)) \\ &= \sum_{e,e' \in \{\frac{-1}{N}, 0, \frac{1}{N}\}} \sum_d \sum_u |P_{D,\widehat{U}_{\ell,\epsilon,n}|E}(d - e', u|e') \\ &- P_{D,\widehat{U}_{\ell,\epsilon,n}|E}(d - e, u|e)| \\ &\leq \sum_{e,e' \in \{\frac{-1}{N}, 0, \frac{1}{N}\}} \sum_d \sum_u |P_{\widehat{U}_{\ell,\epsilon,n},\widehat{V}_{\ell,\epsilon,n}|E}(u, d - e' - u|e') \\ &- P_{\widehat{U}_{\ell,\epsilon,n},\widehat{V}_{\ell,\epsilon,n}|E}(u, d - e' - u|e)| \\ &\quad + |P_{\widehat{U}_{\ell,\epsilon,n},\widehat{V}_{\ell,\epsilon,n},|E}(u, d - e' - u|e) \\ &- P_{\widehat{U}_{\ell,\epsilon,n},\widehat{V}_{\ell,\epsilon,n}|E}(u, d - e - u|e)| \\ &= \sum_{e,e' \in \{\frac{-1}{N}, 0, \frac{1}{N}\}} \sum_v \sum_u |P_{\widehat{U}_{\ell,\epsilon,n},\widehat{V}_{\ell,\epsilon,n}|E}(u, v|e') \\ &- P_{\widehat{U}_{\ell,\epsilon,n},\widehat{V}_{\ell,\epsilon,n}|E}(u, v|e)| \\ &\quad + |P_{\widehat{U}_{\ell,\epsilon,n},\widehat{V}_{\ell,\epsilon,n}|E}(u, v - e'|e) \\ &- P_{\widehat{U}_{\ell,\epsilon,n},\widehat{V}_{\ell,\epsilon,n}|E}(u, v - e|e)|. \end{aligned} \tag{32}$$

We will focus on the first term in equation (32):

$$\begin{aligned} &\sum_{e,e' \in \{\frac{-1}{N}, 0, \frac{1}{N}\}} \sum_{u,v} |P_{\widehat{U}_{\ell,\epsilon,n},\widehat{V}_{\ell,\epsilon,n}|E}(u, v|e') \\ &- P_{\widehat{U}_{\ell,\epsilon,n},\widehat{V}_{\ell,\epsilon,n}|E}(u, v|e)| \\ &\leq \sum_{e,e' \in \{\frac{-1}{N}, 0, \frac{1}{N}\}} |P_{\widehat{U}_{\ell,\epsilon,n},\widehat{V}_{\ell,\epsilon,n}|E}(u, v|e') - P_{\widehat{U}_{\ell,\epsilon,n},\widehat{V}_{\ell,\epsilon,n}}(u, v)| \\ &\quad + |P_{\widehat{U}_{\ell,\epsilon,n},\widehat{V}_{\ell,\epsilon,n}}(u, v) - P_{\widehat{U}_{\ell,\epsilon,n},\widehat{V}_{\ell,\epsilon,n}|E}(u, v|e)|, \end{aligned}$$

where the last two terms go to 0 as $N \to \infty$ due to Pinsker's inequality and the fact that $I(\widehat{U}_{\ell,\epsilon,n}, D; E) \to 0$ as $N \to \infty$. Next we note the second term in (32) goes to 0 by a similar argument as in Equation (31) and uniform continuity of $f_{\widetilde{U}_{\ell,\epsilon},\widetilde{V}_{\ell,\epsilon}}$. As a result, $H(\widehat{U}_{\ell,\epsilon,n}, C) - H(\widehat{U}_{\ell,\epsilon,n}, D) \to 0$. This completes the proof. □

Appendix D
Proof of Theorem 2

We provide a proof of Equation (11). The proof of Equation (12) follows by symmetry.

Step 1 (Clipping X and Y, and Generating $\overline{U}_\ell$ and $\overline{V}_\ell$): Let Z, W, $\widetilde{X}_\ell$, and $\widetilde{Y}_{\ell'}$ be as defined in Section III-B. Let $(\overline{U}_\ell, \overline{V}_{\ell'})$ be random variables that are correlated with (X, Y, Z, W) such that the distribution of $\overline{U}_\ell$ given $\widetilde{X}_\ell$ is given by

$$P_{\overline{U}_\ell|\widetilde{X}_\ell}(\cdot|x) = P_{U|X}(\cdot|x), \quad x \in [-\ell, \ell]$$

and the distribution of $\overline{V}_{\ell'}$ given $\widetilde{Y}_{\ell'}$ is given by

$$P_{\overline{V}_{\ell'}|\widetilde{Y}_{\ell'}}(\cdot|y) = P_{V|Y}(\cdot|y), \quad y \in [-\ell', \ell'].$$

One can check that the following Markov chain holds: $\overline{U}_\ell - \widetilde{X}_\ell - (X,Y,Z,W) - \widetilde{Y}_{\ell'} - \overline{V}_{\ell'}$. Furthermore, for any quadruple of events A, B, C, and D, the random variables $(\widetilde{X}_\ell, \widetilde{Y}_{\ell'}, \overline{U}_\ell, \overline{V}_{\ell'})$ have the following distribution:

$$\begin{aligned}
&P_{\widetilde{X}_\ell,\widetilde{Y}_{\ell'},\overline{U}_\ell,\overline{V}_{\ell'}}(\mathsf{A},\mathsf{B},\mathsf{C},\mathsf{D})\\
&= P_{\widetilde{X}_\ell,\widetilde{Y}_{\ell'},\overline{U}_\ell,\overline{V}_{\ell'},X,Y}(\mathsf{A},\mathsf{B},\mathsf{C},\mathsf{D},\mathbb{R},\mathbb{R})\\
&= P_{\widetilde{X}_\ell,\widetilde{Y}_{\ell'},\overline{U}_\ell,\overline{V}_{\ell'},X,Y}(\mathsf{A},\mathsf{B},\mathsf{C},\mathsf{D},\mathbb{R}\cap[-\ell,\ell],\mathbb{R}\cap[-\ell',\ell'])\\
&+ P_{\widetilde{X}_\ell,\widetilde{Y}_{\ell'},\overline{U}_\ell,\overline{V}_{\ell'},X,Y}(\mathsf{A},\mathsf{B},\mathsf{C},\mathsf{D},\mathbb{R}\cap[-\ell,\ell]^c,\mathbb{R}\cap[-\ell',\ell'])\\
&+ P_{\widetilde{X}_\ell,\widetilde{Y}_{\ell'},\overline{U}_\ell,\overline{V}_{\ell'},X,Y}(\mathsf{A},\mathsf{B},\mathsf{C},\mathsf{D},\mathbb{R}\cap[-\ell,\ell],\mathbb{R}\cap[-\ell',\ell']^c)\\
&+ P_{\widetilde{X}_\ell,\widetilde{Y}_{\ell'},\overline{U}_\ell,\overline{V}_{\ell'},X,Y}(\mathsf{A},\mathsf{B},\mathsf{C},\mathsf{D},\mathbb{R}\cap[-\ell,\ell]^c,\mathbb{R}\cap[-\ell',\ell']^c)\\
&= P_{X,Y,U,V}(\mathsf{A}\cap[-\ell,\ell],\mathsf{B}\cap[-\ell',\ell'],\mathsf{C},\mathsf{D})\\
&+ P_{\widetilde{X}_\ell,\widetilde{Y}_{\ell'},\overline{U}_\ell,\overline{V}_{\ell'},X,Y}(\mathsf{A},\mathsf{B},\mathsf{C},\mathsf{D},\mathbb{R}\cap[-\ell,\ell]^c,\mathbb{R}\cap[-\ell',\ell'])\\
&+ P_{\widetilde{X}_\ell,\widetilde{Y}_{\ell'},\overline{U}_\ell,\overline{V}_{\ell'},X,Y}(\mathsf{A},\mathsf{B},\mathsf{C},\mathsf{D},\mathbb{R}\cap[-\ell,\ell],\mathbb{R}\cap[-\ell',\ell']^c)\\
&+ P_{\widetilde{X}_\ell,\widetilde{Y}_{\ell'},\overline{U}_\ell,\overline{V}_{\ell'},X,Y}(\mathsf{A},\mathsf{B},\mathsf{C},\mathsf{D},\mathbb{R}\cap[-\ell,\ell]^c,\mathbb{R}\cap[-\ell',\ell']^c)
\end{aligned}$$

Note that

$$\begin{aligned}
&0 \le P_{\widetilde{X}_\ell,\widetilde{Y}_{\ell'},\overline{U}_\ell,\overline{V}_{\ell'},X,Y}(\mathsf{A},\mathsf{B},\mathsf{C},\mathsf{D},\mathbb{R}\cap[-\ell,\ell]^c,\mathbb{R}\cap[-\ell',\ell'])\\
&+ P_{\widetilde{X}_\ell,\widetilde{Y}_{\ell'},\overline{U}_\ell,\overline{V}_{\ell'},X,Y}(\mathsf{A},\mathsf{B},\mathsf{C},\mathsf{D},\mathbb{R}\cap[-\ell,\ell],\mathbb{R}\cap[-\ell',\ell']^c)\\
&+ P_{\widetilde{X}_\ell,\widetilde{Y}_{\ell'},\overline{U}_\ell,\overline{V}_{\ell'},X,Y}(\mathsf{A},\mathsf{B},\mathsf{C},\mathsf{D},\mathbb{R}\cap[-\ell,\ell]^c,\mathbb{R}\cap[-\ell',\ell']^c)\\
&\le 1 - P_{X,Y}(\mathbb{R}\cap[-\ell,\ell],\mathbb{R}\cap[-\ell',\ell']),
\end{aligned}$$

which approaches 0 as $\ell, \ell' \to \infty$. As a result,

$$\lim_{\ell,\ell'\to\infty} P_{\widetilde{X}_\ell,\widetilde{Y}_{\ell'},\overline{U}_\ell,\overline{V}_{\ell'}}(\mathsf{A},\mathsf{B},\mathsf{C},\mathsf{D}) = P_{X,Y,U,V}(\mathsf{A},\mathsf{B},\mathsf{C},\mathsf{D}), \tag{33}$$

for all A, B, C, D. Hence as $\ell, \ell' \to \infty$, $P_{\widetilde{X}_\ell,\widetilde{Y}_{\ell'},\overline{U}_\ell,\overline{V}_{\ell'}}$ converges strongly to $P_{X,Y,U,V}$. Recalling that $\overline{U}_\ell$ and $\overline{V}_\ell$ have finite alphabets and using the continuity of mutual information for finite alphabet variables we have

$$\lim_{\ell,\ell'\to\infty} I(\overline{U}_\ell;\overline{U}_\ell+\overline{V}_{\ell'}) = I(U;U+V),$$
$$\lim_{\ell,\ell'\to\infty} I(\overline{V}_{\ell'};\overline{U}_\ell+\overline{V}_{\ell'}) = I(V;U+V).$$

Step 2 (Discretizing X and Y): Next we quantize $\widetilde{X}_\ell$ and $\widetilde{Y}_{\ell'}$ into $\widehat{X}_{n,\ell}$ and $\widehat{Y}_{n,\ell'}$ and enforce the Markov chain. Now using

$$I(\widetilde{X}_\ell\widetilde{Y}_{\ell'}\overline{V}_{\ell'};\overline{U}_\ell|\widehat{X}_{n,\ell}) = I(\widetilde{X}_\ell,\widetilde{Y}_{\ell'}\overline{V}_{\ell'};\overline{U}_\ell) - I(\widehat{X}_{n,\ell};\overline{U}_\ell),$$

and Theorem 2 we have

$$\lim_{n\to\infty} I(\widetilde{X}_\ell\widetilde{Y}_{\ell'}\overline{V}_{\ell'};\overline{U}_\ell|\widehat{X}_{n,\ell}) = I(\widetilde{Y}_{\ell'}\overline{V}_{\ell'};\overline{U}_\ell|\widetilde{X}_\ell) = 0, \tag{34}$$

and similarly,

$$\lim_{n\to\infty} I(\widetilde{X}_\ell\widetilde{Y}_{\ell'}\overline{U}_\ell;\overline{V}_{\ell'}|\widehat{Y}_{n,\ell'}) = I(\widetilde{X}_\ell,\overline{U}_\ell;\overline{V}_{\ell'}|\widetilde{Y}_{\ell'}) = 0. \tag{35}$$

Define $\overline{U}_{n,\ell}$ and $\overline{V}_{n,\ell'}$ as random variables having the same alphabet as $\overline{U}_\ell$ and $\overline{V}_{\ell'}$, and that are jointly correlated with $(\widehat{X}_{n,\ell},\widetilde{X}_\ell,\widetilde{Y}_{\ell'},\widehat{Y}_{n,\ell'})$ according to the probability distribution that satisfies (i) the Markov chain $\overline{V}_{n,\ell'} - \widehat{Y}_{n,\ell'} - \widetilde{Y}_\ell - \widetilde{X}_\ell - \widehat{X}_{n,\ell} - \overline{U}_{n,\ell}$, (ii) the pair $(\widehat{X}_{n,\ell},\overline{U}_{n,\ell})$ has the same distribution as the pair $(\widehat{X}_{n,\ell},\overline{U}_\ell)$, and (iii) the pair $(\widehat{Y}_{n,\ell'},\overline{V}_{n,\ell'})$ has the same distribution as the pair $(\widehat{Y}_{n,\ell'},\overline{V}_{\ell'})$. We use Lemma 6 as follows. From Equations (34) and (35), by taking the quintuple $A=\overline{U}_\ell$, $B=\widehat{X}_{n,\ell}$, $C=(\widetilde{X}_\ell,\widetilde{Y}_\ell)$, $D=\widehat{Y}_{n,\ell'}$, and $E=\overline{V}_{\ell'}$, we have

$$\lim_{n\to\infty} V\left(P_{\widehat{X}_{n,\ell}\widehat{Y}_{n,\ell'}\overline{U}_{n,\ell}\overline{V}_{n,\ell'}}, P_{\widetilde{X}_\ell\widetilde{Y}_{\ell'}\overline{U}_\ell,\overline{V}_{\ell'}}\right) = 0,$$

and using the continuity of mutual information for finite alphabets, we have

$$\lim_{n\to\infty} I(\overline{U}_{n,\ell};\overline{U}_{n,\ell}+\overline{V}_{n,\ell'}) = I(\overline{U}_\ell;\overline{U}_\ell+\overline{V}_{\ell'}),$$
$$\lim_{n\to\infty} I(\overline{V}_{n,\ell'};\overline{U}_{n,\ell}+\overline{V}_{n,\ell'}) = I(\overline{V}_{\ell'};\overline{U}_\ell+\overline{V}_{\ell'}).$$

This completes the proof.

Appendix E
Computational Steps for Computing Products Over Additive MAC

In this appendix we will provide some of the details regarding the computational steps for computing the product of pairs of uniform, Gaussian, and Laplacian random variables over additive MAC with Gaussian noise. We wish to compute two mutual information quantities: $I(Z;Y)$ and $I(Z;X_1)$. We describe the steps in computing $I(Z;Y)$. The steps to compute $I(Z;X_1)$ follow by similar techniques. We estimate the mutual information empirically estimating $\mathbb{E}(\log\frac{f_{Z,Y}(Z,Y)}{f_Z(Z)f_Y(Y)})$ by averaging over randomly generated samples Z and Y. That is, we generate X_1 and X_2 randomly and independently, based on their underlying distribution, e.g. zero-mean and unit-variance uniform, Gaussian, or Laplacian distribution. Then, we compute $Z = X_1X_2$ and $Y = X_1 + X_2 + N$, where N is a zero-mean Gaussian noise with variance $10^{-0.1\mathrm{SNR}}$. Next, we compute the joint PDF $f_{Z,Y}(Z,Y)$ and the marginals $f_Z(Z)$ and $f_Y(Y)$. For Gaussian and Laplacian pairs of variables, analytical expressions for $f_Z(\cdot)$ are given in [56] and [57]. In order to find $f_{Z,Y}$ and f_Y, we use the analytical expressions for $f_{Z,Y|N}(z,y|n) = f_{Z,X_1+X_2}(z,y-n)$ and $f_{Y|N}(y|n) = f_{X_1+X_2}(y-n)$, given in [56] and [57], and empirically estimate $f_{Z,Y} = \mathbb{E}(f_{Z,Y|N})$ and $f_Y = \mathbb{E}(f_{Y|N})$, respectively. We estimate the information quantities based on 5×10^5 Monte-Carlo trials, and the PDFs based on 10^5 Monte-Carlo trials.

Appendix F
Proof of Theorem 4

Proof: We first prove the following lemma which shows the convergence of expected distortion under the discretization procedure.

Lemma 7: Under the discretization procedure of the auxiliary variables U and V considered in Theorem 1, for any $\xi > 0$, and all sufficiently large $n, \ell, \ell' > 0$, and sufficiently small $\epsilon > 0$, the following hold:

$$|\mathbb{E}(d_1(X,g_1(\widehat{U}_{n,\ell,\epsilon},\widehat{V}_{n,\ell',\epsilon}))) - \mathbb{E}(d_1(X,g_1(U,V)))| \le \xi, \tag{36}$$

$$|\mathbb{E}(d_2(Y,g_2(\widehat{U}_{n,\ell,\epsilon},\widehat{V}_{n,\ell',\epsilon}))) - \mathbb{E}(d_2(Y,g_2(U,V)))| \le \xi. \tag{37}$$

Proof: The lemma follows by convergence in distribution of $(X,\widehat{U}_{n,\ell,\epsilon},\widehat{V}_{n,\ell',\epsilon})$ to (X,U,V) and $(Y,\widehat{U}_{n,\ell,\epsilon},\widehat{V}_{n,\ell',\epsilon})$ to (Y,U,V), which was shown in Appendix C, along with the

Portmanteau theorem [58], [59] and the arguments in the proof of Lemma 4. □

Lemma 8: Under the discretization procedure of the sources X and Y considered in Theorem 2, and for a fixed finite alphabet auxiliary variables $\widehat{U}$ and $\widehat{V}$ with the Markov chain, for any $\xi > 0$, and all sufficiently large $n, \ell, \ell' > 0$, and sufficiently small $\epsilon > 0$, the following hold:

$$|\mathbb{E}(d_1(\widehat{X}_{n,\ell}, g_1(\overline{U}_{n,\ell}, \overline{V}_{n,\ell'}))) - \mathbb{E}(d_1(X, g_1(\widehat{U}, \widehat{V})))| \leq \xi. \quad (38)$$

$$|\mathbb{E}(d_2(\widehat{Y}_{n,\ell}, g_2(\overline{U}_{n,\ell}, \overline{V}_{n,\ell'}))) - \mathbb{E}(d_2(Y, g_2(\widehat{U}, \widehat{V})))| \leq \xi. \quad (39)$$

Proof: Note that using Equation (33) in Appendix D, along with the Portmanteau Theorem, for the clipped variables, we have:

$$\lim_{\ell,\ell' \to \infty} |\mathbb{E}d(\widetilde{X}_\ell, g_1(\overline{U}_\ell, \overline{V}_{\ell'})) - \mathbb{E}d(X, g_1(U, V))| = 0. \quad (40)$$

For the quantized variables, since convergence in variational distance, shown in Appendix D, implies convergence in distribution, the lemma follows using Portmanteau theorem [58], [59] and Lemma 4. □

Fix $\xi > 0$, and define $d_1'(x, u, v) := d_1(x, g(u, v))$ and $d_2'(y, u, v) := d_2(y, g(u, v))$ for all $x, y, u, v, \in \mathbb{R}^4$. Using the procedure described in Section III, we perform clipping, smoothing, and discretization of random variables (U, V) into $\widehat{U}_{n',\ell',\epsilon'}$ and $\widehat{V}_{n',\ell',\epsilon}$ by choosing the parameters (n', ℓ', ϵ) appropriately. Furthermore, using the procedure described in Section III, we perform clipping, and discretization of the source variables (X, Y) to produce the quadruple $(\widehat{X}_{n,\ell}, \widehat{Y}_{n,\ell}, \overline{U}'_{n',\ell',\epsilon'}, \overline{V}'_{n',\ell',\epsilon'})$ by choosing the parameters (n, ℓ) appropriately. For ease of notation, we will drop the subscripts from the random variables when there is no ambiguity. Using the data-processing inequality, the lower semi-continuity of mutual information and Theorems 1 and 2 and Lemmas 7 and 8, we have:

$$\begin{aligned}
&|I(\widehat{X}; \overline{U}) - I(X; U)| \leq 2\xi, \quad |I(\widehat{Y}; \overline{V}) - I(Y; V)| \leq 2\xi, \\
&|I(\overline{U}; \overline{V}) - I(U; V)| \leq 2\xi, \quad (41) \\
&\mathbb{E}d_1'(X, U, V) + \xi \geq \mathbb{E}d_1'(\widehat{X}, \overline{U}, \overline{V}), \\
&\mathbb{E}d_2'(Y, U, V) + \xi \geq \mathbb{E}d_2'(\widehat{Y}, \overline{U}, \overline{V}). \quad (42)
\end{aligned}$$

We can use the coding theorem in [19] to show that the rate-distortion tuple given by

$$\begin{aligned}
&R_1 \geq I(\widehat{X}; \overline{U}) - I(\overline{U}; \overline{V}), \\
&R_2 \geq I(\widehat{Y}; \overline{V}) - I(\overline{U}; \overline{V}), \\
&R_1 + R_2 \geq I(\widehat{X}; \overline{U}) + I(\widehat{Y}; \overline{V}) - I(\overline{U}; \overline{V}), \\
&D_1 \geq \mathbb{E}d_1'(\widehat{X}, \overline{U}, \overline{V}), \qquad D_2 \geq \mathbb{E}d_2'(\widehat{Y}, \overline{U}, \overline{V}),
\end{aligned}$$

is achievable for the finite-alphabet source $(\widehat{X}, \widehat{Y}, d_1, d_2)$. Based on Equations (41) the rates are within 2ξ of the ones in the theorem statement. To show that the claimed distortions for the reconstruction of continuous source are achievable, consider a transmission system with parameter (m, Θ_1, Θ_2) for compressing the finite-alphabet source such that

$$\frac{1}{m} \sum_{i=1}^{m} \mathbb{E}d_1'(\widehat{X}_i, \overline{U}_i', \overline{V}_i') \leq \mathbb{E}d_1'(\widehat{X}, \overline{U}, \overline{V}) + \xi, \quad (43)$$

where $(\overline{U}'^m, \overline{V}'^m) := f(e_1(\widehat{X}^m), e_2(\widehat{Y}^m))$, and e_1 and e_2 denote the encoders and f the decoder.

For the source (X, Y, d_1, d_2) we obtain an $(m, \Theta_1', \Theta_2')$ transmission system TS_c as follows. We assume that the encoder and decoder share common randomness. From (X, Y) we create $(\widehat{X}, \widehat{Y})$ and use TS_d. Let $T_1 = \mathbb{1}_{\{X \in [-\ell, \ell]\}}$. The encoder of X in TS_c sends information to the decoder in two parts. The first part is $e(\widehat{X}^m)$ and the second part is a compressed (almost lossless) version of T_1^m. If the sequence T_1^m is typical, the encoder sends the index of the sequence in the typical set, otherwise, it sends the index 0. This extra piece of information requires $h_b(P_X([-\ell, \ell]) + \xi$ bits per sample. Similar encoding strategy is used at the other encoder. The decoder is constructed as follows. Let $(\check{U}^m, \check{V}^m)$ denote the reconstruction vector. If the decoder receives index 0 in the second part from either of the encoders, then it uses an arbitrary constant c as a reconstruction, i.e., $\check{U}_i = c, \check{V}_i = c'$ for all i. Otherwise, it can reconstruct (T_1^m, T_2^m) reliably. If $T_{1i} = T_{2i} = 1$, then the reconstruction is $\check{U}_i = \overline{U}_i'$ and $\check{V}_i = \overline{V}_i'$, otherwise it is arbitrary constants (c, c'), i.e., $\check{U}_i = c, \check{V}_i = c'$.

Assume that the parameters of the transmission system that of the discrete source are such that

- $h_b(P_X([-\ell, \ell])) \leq \xi$, and $h_b(P_Y([-\ell, \ell])) \leq \xi$
- $P(\mathsf{A}^c)\mathbb{E}d_1'(X_i, c, c'|\mathsf{A}^c) \leq \xi$, and $P(\mathsf{A}^c)\mathbb{E}d_2'(Y_i, c, c'|\mathsf{A}^c) \leq \xi$ for all $1 \leq i \leq m$, where A denotes the event that (T_1^m, T_2^m) is jointly typical.
- $P(\mathsf{B}_i^c)\mathbb{E}d_1'(X_i, c, c'|\mathsf{B}_i^c) \leq \xi$, and $P(\mathsf{B}_i^c)\mathbb{E}d_2'(Y_i, c, c'|\mathsf{B}_i^c) \leq \xi$ for all $1 \leq i \leq m$, where B_i denote the event $(X_i, Y_i) \in [-\ell, \ell] \times [-\ell, \ell]$.
- $|d_1'(x_1, b_1, b_2) - d_1'(x_2, b_1, b_2)| \leq \xi$, for all (a) $x_1, x_2 \in [-\ell, \ell]$, (b) $b_1 \in [-\ell', \ell']$, (c) $b_2 \in [-\ell', \ell']$, and (d) $|x_1 - x_2| \leq \frac{1}{2^n}$.
- $|d_2'(y_1, b_1, b_2) - d_2'(y_2, b_1, b_2)| \leq \xi$, for all (a) $x_1, x_2 \in [-\ell, \ell]$, (b) $b_1 \in [-\ell', \ell']$, (c) $b_2 \in [-\ell', \ell']$, and (d) $|x_1 - x_2| \leq \frac{1}{2^n}$.

Note that the constants c, c' satisfying the conditions in the second and third bullets exist by the assumption that the expected distortion is finite so that there exist c, c' such that $\mathbb{E}(d_1'(X, c, c')) < \infty$ and $\mathbb{E}(d_2'(Y, c, c')) < \infty$. Let $P_{X_i, Y_i|\mathsf{A}, \mathsf{B}_i}$ denote the probability distribution of (X_i, Y_i) given the event A and B_i. Consider for any $i \in \{1, 2, \ldots, m\}$,

$$\begin{aligned}
&\mathbb{E}d_1'(X_i, \check{U}_i, \check{V}_i) \\
&\leq P(\mathsf{A}^c)\mathbb{E}d_1'(X_i, c, c'|\mathsf{A}^c) + P(B_i^c)\mathbb{E}d_1'(X_i, c, c'|\mathsf{B}_i^c) \\
&\qquad + P(\mathsf{A} \cap \mathsf{B}_i)\mathbb{E}d_1'(X_i, \check{U}_i, \check{V}_i|\mathsf{A}, \mathsf{B}_i) \\
&\overset{a}{\leq} 2\xi + P(\mathsf{A} \cap \mathsf{B}_i)\mathbb{E}d_1'(X_i, \check{U}_i, \check{V}_i|\mathsf{A} \cap \mathsf{B}_i) \\
&= 2\xi + P(\mathsf{A} \cap \mathsf{B}_i)\mathbb{E}d_1'(X_i, \overline{U}_i', \overline{V}_i'|\mathsf{A} \cap \mathsf{B}_i) \\
&\leq 2\xi + P(\mathsf{B}_i)\mathbb{E}d_1'(X_i, \overline{U}_i', \overline{V}_i'|\mathsf{B}_i) \\
&\overset{(b)}{=} 2\xi + P(\mathsf{B}_i) \sum_{i,b,b'} P_i(b, b'|\zeta(i)) \int_{A(i)} d_1'(x, b, b') \frac{dP_{X_i}(x)}{P(\mathsf{B}_i)} \\
&\overset{(c)}{\leq} 2\xi + P(\mathsf{B}_i) \left[\mathbb{E}d_1'(\widehat{X}_i, \overline{U}_i, \overline{V}_i) + \xi\right],
\end{aligned}$$

where we have following arguments: (a) follows from third and fourth bullets from the previous page. In (b) we have denoted the conditional probability of $\overline{U}_i, \overline{V}_i$ given $\widehat{X}_i$ as P_i. (c) follows from the fifth bullet from the previous page. Finally, we have

$$\frac{1}{m}\sum_{i=1}^{m}\mathbb{E}d_1'(X_i,\check{U}_i,\check{V}_i) \leq 2\xi + \left[\mathbb{E}d_1'(\widehat{X},\overline{U},\overline{V})) + 2\xi\right] \leq 5\xi + \mathbb{E}d_1'(X,U,V).$$

The proof for the distortion for reconstructing Y follows by similar arguments. This completes the desired proof. □

APPENDIX G
PROOF OF PROPOSITION 2

The proof follows similar steps as the one given in [9, Example 3]. We provide an outline in the following. Note that Decoders $\{1\}$,$\{2\}$ and $\{1,2\}$ operate at optimal PtP rate-distortion. So by the same arguments as in steps one through five in [9, Example 3], all codebooks except $C_{\{1\},\{3\}}, C_{\{2\},\{3\}}, C_{\{1\}}, C_{\{2\}}, C_{\{3\}}$ can be eliminated. For ease of notation, we denote the corresponding random variables for these codebooks as $U_{1,3}, U_{2,3}, U_1, U_2, U_3$, respectively. Note that due to optimality at Decoders $\{1\}$, $\{2\}$ and $\{1,2\}$, we must have: $(U_{1,3}, U_1) - X - Z - (U_{2,3}, U_2)$.

In order to evaluate the achievable rates using Gaussian test channels, let $U_{1,3} = X + Q_{1,3}$ and $U_{2,3} = Z + Q_{2,3}$, where $Q_{1,3}, Q_{2,3}$ are two Gaussian variables which are independent of each other and of X, Z with variances $\theta_1, \theta_2 > 0$, respectively. Note that $E(X|U_{1,3}) = \frac{1}{1+\theta_1}U_{1,3}$ and $E(Z|U_{2,3}) = \frac{1}{1+\theta_2}U_{2,3}$. In order to achieve the desired distortion at Decoders $\{1\}$ and $\{2\}$, we must have $U_1 = X - \frac{1}{1+\theta_1}U_{1,3} + Q_1$ and $U_2 = Z - \frac{1}{1+\theta_1}U_{2,3} + Q_2$, where Q_1 and Q_2 are Gaussian variables with zero mean and variance P, independent of each other and all other variables. Then, the reconstructions $\hat{X} = U_{1,3} + U_1$ and $\hat{Z} = U_{2,3} + U_2$ satisfy the distortion constraints at Decoder $\{1\}$ and Decoder $\{2\}$, respectively. The Gaussian variable U_3 can be decomposed in terms of $X, Z, Q_{1,3}, Q_{2,3}, Q_1, Q_2, Q_3$, where Q_3 is an independent Gaussian variable with zero mean and unit variance, so that $U_3 = \alpha_1 X + \alpha_2 Z + \alpha_3 Q_{1,3} + \alpha_4 Q_{2,3} + \alpha_5 Q_1 + \alpha_6 Q_2 + \alpha_7 Q_3$ for some $\alpha_i \in \mathbb{R}, i \in [7]$. Then, the reconstruction at Decoder $\{3\}$ of $X+Z$ which minimizes the distortion is given by:

$$\begin{aligned}(\widehat{X+Z})_3 &:= \mathbb{E}(X+Z|U_{1,3},U_{2,3},U_3)\\ &= \Sigma_{X+Z,U_{1,3}U_{2,3}U_3}\Sigma^{-1}_{U_{1,3},U_{2,3},U_3}\begin{bmatrix}U_{1,3} & U_{2,3} & U_3\end{bmatrix}^T\\ &= \frac{1}{1+\theta_1}U_{1,3} + \frac{1}{1+\theta_2}W_{2,3} + \frac{\alpha_1+\alpha_2}{Var(U_3)}U_3,\end{aligned}$$

where $\Sigma_{X+Z,U_{1,3}U_{2,3}U_3} := \mathbb{E}((X+Z)[U_{1,3},U_{2,3},U_3])$ and $\Sigma_{U_{1,3},U_{2,3},U_3}$ is the covariance matrix of $U_{1,3}, U_{2,3}, U_3$. Similarly, the reconstructions at Decoder $\{1,3\}$ are:

$$\begin{aligned}\widehat{X}_{1,3} &:= \mathbb{E}(X|U_{1,3},U_{2,3},U_1,U_3)\\ &= \Sigma_{X,U_{1,3}U_{2,3}U_1U_3}\Sigma^{-1}_{U_{1,3},U_{2,3},U_1,U_3}\begin{bmatrix}U_{1,3} & U_{2,3} & U_1 & U_3\end{bmatrix}^T,\\ \widehat{Z}_{1,3} &:= \mathbb{E}(Z|U_{1,3},U_{2,3},U_1,U_3)\\ &= \Sigma_{Z,U_{1,3}U_{2,3}U_1U_3}\Sigma^{-1}_{U_{1,3},U_{2,3},U_1,U_3}\begin{bmatrix}U_{1,3} & U_{2,3} & U_1 & U_3\end{bmatrix}^T.\end{aligned}$$

The reconstructions $\widehat{X}_{2,3}, \widehat{Z}_{2,3}$ at Decoder $\{2,3\}$ can be written in a similar fashion. Furthermore, using the covering and packing bounds in Theorem 6, we have:

$$\begin{aligned}R_3 &\geq I(X,Z;U_3,U_{1,3},U_{2,3}) + I(U_1,U_2;U_3|U_{1,3},U_{2,3},X,Z)\\ &= \frac{1}{2}\log\frac{|\Sigma_{X,Z}||\Sigma_{U_{1,3}U_{2,3}U_3}|}{|\Sigma_{X,Z,U_{1,3},U_{2,3}U_3}|}\\ &\quad + I(Q_1,Q_2;\alpha_5Q_1+\alpha_6Q_2+\alpha_7Q_3)\\ &= \frac{1}{2}\log\frac{|\Sigma_{U_{1,3}U_{2,3}U_3}|}{|\Sigma_{X,Z,U_{1,3},U_{2,3}U_3}|} + \frac{1}{2}\log\left(\alpha_5^2P + \alpha_6^2P + \alpha_7^2\right).\end{aligned}$$

Computer-assisted optimization over $\alpha_i, i \in [7], \theta_1, \theta_2$ for $P = 0.5$ yields the value 0.9317 which is strictly less than $\frac{1}{2}\log\frac{2}{P} = 1$. The achievable rates for other values of $P \in [0,1]$ are plotted in Figure 5. □

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers and an associate editor for their constructive comments which led to a much improved manuscript. They would like to thank Mohsen Heidari for his helpful comments and suggestions in preparing this manuscript.

REFERENCES

[1] I. Csiszár and J. Korner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. New York, NY, USA: Academic, 1981.
[2] A. D. Wyner, "The rate-distortion function for source coding with side information at the decoder—II: General sources," *Inf. Control*, vol. 38, no. 1, pp. 60–80, Jul. 1978.
[3] M. S. Pinsker, *Information and Information Stability of Random Variables and Processes*. San Francisco, CA, USA: Holden-Day, 1964.
[4] A. Padakandla and S. S. Pradhan, "Achievable rate region for three user discrete broadcast channel based on coset codes," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2013, pp. 1277–1281.
[5] A. Padakandla, A. G. Sahebi, and S. S. Pradhan, "A new achievable rate region for the 3-user discrete memoryless interference channel," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2012, pp. 2256–2260.
[6] B. Nazer and M. Gastpar, "Computation over Gaussian multiple-access channels," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2007, pp. 2391–2395.
[7] T. Philosof and R. Zamir, "On the loss of single-letter characterization: The dirty multiple access channel," *IEEE Trans. Inf. Theory*, vol. 55, no. 6, pp. 2442–2454, Jun. 2009.
[8] D. Krithivasan and S. S. Pradhan, "Distributed source coding using Abelian group codes: A new achievable rate-distortion region," *IEEE Trans. Inf. Theory*, vol. 57, no. 3, pp. 1495–1519, Mar. 2011.
[9] F. Shirani and S. S. Pradhan, "An achievable rate-distortion region for multiple descriptions source coding based on coset codes," *IEEE Trans. Inf. Theory*, vol. 64, no. 5, pp. 3781–3809, May 2018.
[10] F. Shirani and S. S. Pradhan, "Trade-off between communication and cooperation in the interference channel," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2016, pp. 2214–2218.
[11] J. Korner and K. Marton, "How to encode the modulo-two sum of binary sources (Corresp.)," *IEEE Trans. Inf. Theory*, vol. IT-25, no. 2, pp. 219–221, Mar. 1979.
[12] T. Philosof, R. Zamir, U. Erez, and A. J. Khisti, "Lattice strategies for the dirty multiple access channel," *IEEE Trans. Inf. Theory*, vol. 57, no. 8, pp. 5006–5035, Aug. 2011.
[13] S. S. Pradhan, A. Padakandla, and F. Shirani, "An algebraic and probabilistic framework for network information theory," *Found. Trends Commun. Inf. Theory*, vol. 18, no. 2, pp. 173–379, 2021.
[14] S. Sridharan, A. Jafarian, S. Vishwanath, S. A. Jafar, and S. Shamai (Shitz), "A layered lattice coding scheme for a class of three user Gaussian interference channels," in *Proc. 46th Annu. Allerton Conf. Commun., Control, Comput.*, Sep. 2008, pp. 531–538.
[15] D. Krithivasan and S. S. Pradhan, "Lattices for distributed source coding: Jointly Gaussian sources and reconstruction of a linear function," *IEEE Trans. Inf. Theory*, vol. 55, no. 12, pp. 5628–5651, Dec. 2009.

[16] F. Shirani and S. S. Pradhan, "Lattices from linear codes and fine quantization: General continuous sources and channels," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2018, pp. 2356–2360.

[17] F. Shirani, M. Heidari, and S. S. Pradhan, "New lattice codes for multiple-descriptions," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2015, pp. 1580–1584.

[18] A. Pastore, S. H. Lim, C. Feng, B. Nazer, and M. Gastpar, "A discretization approach to compute–forward," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2021, pp. 2208–2213.

[19] T. Berger, *Multiterminal Source Coding*. New York, NY, USA: Spring-Verlag, 1977.

[20] A. V. Makkuva and Y. Wu, "Equivalence of additive-combinatorial linear inequal. For Shannon entropy and differential entropy," *IEEE Trans. Inf. Theory*, vol. 64, no. 5, pp. 3579–3589, May 2018.

[21] U. Erez and R. Zamir, "Achieving $\frac{1}{2}$ log (1 + SNR) on the AWGN channel with lattice encoding and decoding," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2293–2314, Sep. 2004.

[22] R. Zamir, S. Shamai (Shitz), and U. Erez, "Nested linear/lattice codes for structured multiterminal binning," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1250–1276, Jun. 2002.

[23] J. Østergaard and R. Zamir, "Multiple-description coding by dithered delta–sigma quantization," *IEEE Trans. Inf. Theory*, vol. 55, no. 10, pp. 4661–4675, Oct. 2009.

[24] S. N. Diggavi, N. J. A. Sloane, and V. A. Vaishampayan, "Asymmetric multiple description lattice vector quantizers," *IEEE Trans. Inf. Theory*, vol. 48, no. 1, pp. 174–191, Feb. 2002.

[25] J. Chen, C. Tian, T. Berger, and S. S. Hemami, "Multiple description quantization via Gram–Schmidt orthogonalization," *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5197–5217, Dec. 2006.

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

[27] S.-Y. Tung, "Multiterminal source coding," Ph.D. dissertation, School Elect. Eng., Cornell Univ., 1978.

[28] P. Mitran, "On a Markov lemma and typical sequences for Polish alphabets," *IEEE Trans. Inf. Theory*, vol. 61, no. 10, pp. 5342–5356, Oct. 2015.

[29] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*. vol. 159. Cham, Switzerland: Springer, 2012.

[30] R. M. Gray, *Entropy and Information Theory*. Cham, Switzerland: Springer, 2011.

[31] A. N. Kolmogorov, "On the representation of continuous functions of several variables by superpositions of continuous functions of a smaller number of variables," *Proc. USSR Acad. Sci.*, pp. 179–182, 1956.

[32] V. I. Arnold, "On functions of three variables," in *Collected Works: Representations of Functions, Celestial Mechanics and KAM Theory, 1957–1965*. Springer, 2009, pp. 5–8.

[33] B. Nazer and M. Gastpar, "Computation over multiple-access channels," *IEEE Trans. Inf. Theory*, vol. 53, no. 10, pp. 3498–3516, Oct. 2007.

[34] B. Nazer and M. Gastpar, "Compute-and-forward: Harnessing interference through structured codes," *IEEE Trans. Inf. Theory*, vol. 57, no. 10, pp. 6463–6486, Oct. 2011.

[35] S. A. Jafar et al, "Interference alignment—A new look at signal dimensions in a communication network," *Found. Trends® Commun. Inf. Theory*, vol. 7, no. 1, pp. 1–134, 2011.

[36] A. Padakandla and S. S. Pradhan, "Computing sum of sources over an arbitrary multiple access channel," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2013, pp. 2144–2148.

[37] D. Slepian and J. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inf. Theory*, vol. IT-19, no. 4, pp. 471–480, Jul. 1973.

[38] A. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inf. Theory*, vol. IT-22, no. 1, pp. 1–10, Jan. 1976.

[39] A. B. Wagner, B. G. Kelly, and Y. Altug, "Distributed rate-distortion with common components," *IEEE Trans. Inf. Theory*, vol. 57, no. 7, pp. 4035–4057, Jul. 2011.

[40] F. Shirani and S. S. Pradhan, "A new achievable rate-distortion region for distributed source coding," *IEEE Trans. Inf. Theory*, vol. 67, no. 7, pp. 4485–4503, Jul. 2021.

[41] A. B. Wagner and V. Anantharam, "An improved outer bound for multiterminal source coding," *IEEE Trans. Inf. Theory*, vol. 54, no. 5, pp. 1919–1937, May 2008.

[42] S. Tavildar, P. Viswanath, and A. B. Wagner, "The Gaussian many-help-one distributed source coding problem," *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 564–581, Jan. 2010.

[43] J. Wang, J. Chen, and X. Wu, "On the sum rate of Gaussian multiterminal source coding: New proofs and results," *IEEE Trans. Inf. Theory*, vol. 56, no. 8, pp. 3946–3960, Aug. 2010.

[44] Z. Zhang and T. Berger, "New results in binary multiple descriptions," *IEEE Trans. Inf. Theory*, vol. IT-33, no. 4, pp. 502–521, Jul. 1987.

[45] A. El Gamal and T. M. Cover, "Multiple user information theory," *Proc. IEEE*, vol. 68, no. 12, pp. 1466–1483, Apr. 1980.

[46] E. Akyol, K. Viswanatha, and K. Rose, "Combinatorial message sharing and random binning for multiple description coding," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2012, pp. 1371–1375.

[47] S. S. Pradhan, R. Puri, and K. Ramchandran, "n-channel symmetric multiple descriptions—Part I:(n,k) source-channel erasure codes," *IEEE Trans. Inf. Theory*, vol. 50, no. 1, pp. 47–61, Jan. 2004.

[48] C. Tian and J. Chen, "New coding schemes for the symmetric K-description problem," *IEEE Trans. Inf. Theory*, vol. 56, no. 10, pp. 5344–5365, Oct. 2010.

[49] L. Ozarow, "On a source-coding problem with two channels and three receivers," *Bell Syst. Tech. J.*, vol. 59, no. 10, pp. 1909–1921, Dec. 1980.

[50] H. Wang and P. Viswanath, "Vector Gaussian multiple description with individual and central receivers," *IEEE Trans. Inf. Theory*, vol. 53, no. 6, pp. 2133–2153, Jun. 2007.

[51] A. B. Wagner, S. Tavildar, and P. Viswanath, "Rate region of the quadratic Gaussian two-encoder source-coding problem," *IEEE Trans. Inf. Theory*, vol. 54, no. 5, pp. 1938–1961, May 2008.

[52] Y. Xu, J. Chen, and Q. Wang, "The sum rate of vector Gaussian multiple description coding with tree-structured covariance distortion constraints," *IEEE Trans. Inf. Theory*, vol. 63, no. 10, pp. 6547–6560, Oct. 2017.

[53] H. L. Royden and P. Fitzpatrick, *Real Analysis*, vol. 32. New York, NY, USA: Macmillan, 1988.

[54] R. Zamir, *Lattice Coding for Signals and Networks: A Structured Coding Approach to Quantization, Modulation, and Multiuser Information Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2014.

[55] Y. Polyanskiy and Y. Wu, "Dissipation of information in channels with input constraints," *IEEE Trans. Inf. Theory*, vol. 62, no. 1, pp. 35–55, Jan. 2016.

[56] S. Nadarajah, "The linear combination, product and ratio of Laplace random variables," *Statistics*, vol. 41, no. 6, pp. 535–545, Dec. 2007.

[57] M. K. Simon, *Probability Distributions Involving Gaussian Random Variables: A Handbook for Engineers and Scientists*. Cham, Switzerland: Springer, 2002.

[58] K. Achim, *Probability Theory: A Comprehensive Course*. NY, USA: Wiley, 2014.

[59] P. Billingsley, *Convergence of Probability Measures*. Hoboken, NJ, USA: Wiley, 2013.

Farhad Shirani (Member, IEEE) received the B.S. degree in electrical engineering from the Sharif University of Technology and the M.Sc. degree in applied mathematics and the Ph.D. degree from the University of Michigan, Ann Arbor, MI, USA. He was a Lecturer and a Post-Doctoral Research Fellow with the University of Michigan in 2017. He was a Research Assistant Professor with New York University from 2017 to 2020. He is currently an Assistant Professor with the Knight Foundation School of Computing and Information Sciences, Florida International University. His recent work includes developing information-theoretic methods for the analysis of fundamental limits of web privacy, the design of receiver architectures for energy-efficient communication over MIMO systems, and the design of algorithms for opportunistic multi-user scheduling under various fairness constraints. His research interests include privacy and security, wireless communications, and machine learning.

S. Sandeep Pradhan (Fellow, IEEE) received the M.E. degree from the Indian Institute of Science in 1996 and the Ph.D. degree from the University of California at Berkeley in 2001. From 2002 to 2008, he was an Assistant Professor. From 2008 to 2015, he was an Associate Professor with the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI, USA, where he is currently a Professor. His research interests include network information theory, coding theory, and quantum information theory. He was an Associate Editor of IEEE TRANSACTIONS ON INFORMATION THEORY in the area of Shannon theory from 2014 to 2016.