Distributed Simulation of Continuous Random Variables

Cheuk Ting Li, *Student Member, IEEE*, and Abbas El Gamal, *Fellow, IEEE*

***Abstract*— We establish the first known upper bound on the exact and Wyner's common information of n continuous random variables in terms of the dual total correlation between them (which is a generalization of mutual information). In particular, we show that when the pdf of the random variables is log-concave, there is a constant gap of $n^2 \log e + 9n \log n$ between this upper bound and the dual total correlation lower bound that does not depend on the distribution. The upper bound is obtained using a computationally efficient dyadic decomposition scheme for constructing a discrete common randomness variable W from which the n random variables can be simulated in a distributed manner. We then bound the entropy of W using a new measure, which we refer to as the erosion entropy.**

***Index Terms*— Exact common information, Wyner's common information, log-concave distribution, dual total correlation, channel synthesis.**

I. Introduction

THIS paper is motivated by the following question. Alice would like to simulate a random variable X_1 and Bob would like to simulate another random variable X_2 such that (X_1, X_2) are jointly Gaussian with a prescribed mean and covariance matrix. Can these two random variables be simulated in a distributed manner with only a finite amount of common randomness between Alice and Bob?

We answer this question in the affirmative for n continuous random variables under certain conditions on their joint pdf, including when it is log-concave such as Gaussian.

The general distributed randomness generation setup we consider is depicted in Figure 1. There are n agents (e.g., processors in a computer cluster or nodes in a communication network) that observe an i.i.d. Bern(1/2) sequence $B_1, B_2, \ldots$. Agent $i \in [1:n]$ wishes to simulate the random variable X_i using this sequence and its local randomness, which is independent of $B_1, B_2, \ldots$ and at other agents' local randomness, such that $X^n = (X_1, \ldots, X_n)$ follows a prescribed distribution *exactly*.

To perform this distributed randomness generation task, the agents adopt an agreed upon *discrete distribution generating tree* [1]. At time $t \in \{1, 2, \ldots\}$, they observe B_t and traverse the tree accordingly. Let T be the time at which they reach a leaf node of the tree. The agents then simultaneously generate their random variables using the common randomness sequence B^T and their local randomness. The distributed randomness simulation problem is to find the minimum expected number of common random bits $\mathsf{E}[T]$ to perform the randomness generation task and the scheme that achieves this minimum.

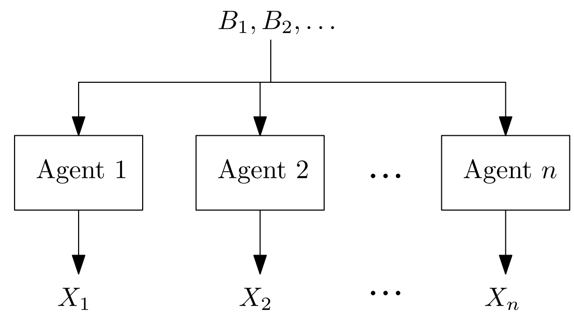

Fig. 1. Distributed randomness generation setting. An i.i.d. sequence $B_1, B_2, \ldots$ is broadcast to n agents and agent $i \in [1:n]$ generates X_i using B^T and its local randomness, where B^T is the sequence corresponding to a path from the root to a leaf of the agreed upon discrete distribution generating tree.

Using the well-known Knuth-Yao result on the optimal discrete distribution generating tree [1], we obtain

$$G(X_1; \ldots; X_n) \le \min \mathsf{E}[T] \le G(X_1; \ldots; X_n) + 2,$$

where

$$G(X_1; \ldots; X_n) = \min_{W:\, X_1 \perp\!\!\!\perp X_2 \perp\!\!\!\perp \ldots \perp\!\!\!\perp X_n \mid W} H(W). \tag{1}$$

is the *common entropy* [2]. Hence in this paper we will focus on studying $G(X_1; \ldots; X_n)$ instead of $\min \mathsf{E}[T]$. Note that the "common randomness variable" W here is generated using the sequence B^T, and agent i can generate X_i according to the conditional distribution given W.

The common entropy between two random variables was investigated in [2]. Computing $G(X_1; X_2)$, even for moderate size random variable alphabets, can be computationally difficult since it involves minimizing a concave function over a non-convex set; see [2] for some cases where G can be computed and for some properties that can be exploited to compute it. Hence the main difficulty in constructing a scheme that achieves G (within 2 bits) for a given (X_1, X_2) distribution is finding the optimal common randomness W that achieves it.

Manuscript received May 4, 2016; revised March 24, 2017; accepted May 22, 2017. Date of publication August 3, 2017; date of current version September 13, 2017. C. T. Li was supported by the Hong Kong Alumni Stanford Graduate Fellowship. This paper was presented at the 2016 IEEE International Symposium on Information Theory.

The authors are with the Department of Electrical Engineering, Stanford University, Stanford, CA 94305 USA (e-mail: ctli@stanford.edu; abbas@ee.stanford.edu).

Communicated by T. Liu, Associate Editor for Shannon Theory.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2017.2735438

For discrete random variables (X_1, X_2), it can be readily shown that

$$I(X_1; X_2) \le J(X_1;X_2) \le G(X_1; X_2) \le \min\{H(X_1), H(X_2)\}, \tag{2}$$

where

$$J(X_1; X_2) = \min_{W:\, X_1 \perp\!\!\!\perp X_2 | W} I(W; X_1, X_2) \tag{3}$$

is Wyner's common information [3], which is the minimum amount of common randomness rate needed to generate the discrete memoryless source (DMS) (X_1, X_2) with asymptotically vanishing normalized relative entropy. The notion of exact common information rate $\bar{G}(X_1; X_2) = \lim_n G(X_1^n; X_2^n)/n$, which is the asymptotic minimum amount of common randomness rate needed to generate the DMS (X_1, X_2) exactly, was also introduced in [2]. It was shown that: (i) in general $J \le \bar{G} \le G$, (ii) G can be strictly larger than $\bar{G}$, and (iii) in some cases $\bar{G}(X_1; X_2) = J(X_1; X_2)$. It is not known, however, if $\bar{G}(X_1; X_2) = J(X_1; X_2)$ in general. As such, we do not consider $\bar{G}$ further in this paper.

In [4] and [5] Wyner's common information was extended to n discrete random variables

$$J(X_1; \ldots; X_n) = \min_{W:\, X_1 \perp\!\!\!\perp X_2 \perp\!\!\!\perp \ldots \perp\!\!\!\perp X_n | W} I(W; X_1, \ldots, X_n).$$

We can generalize the bounds in (2) to n random variables to obtain

$$\begin{aligned} I_{\mathrm{D}}(X_1; X_2; \ldots; X_n) &\le J(X_1; X_2; \ldots; X_n) \\ &\le G(X_1; X_2; \ldots; X_n) \\ &\le \min_i H(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n), \end{aligned} \tag{4}$$

where I_{D} is the *dual total correlation* [6]—a generalization of mutual information defined as

$$\begin{aligned} &I_{\mathrm{D}}(X_1; X_2; \ldots; X_n) \\ &= H(X_1, \ldots, X_n) - \sum_{i=1}^{n} H(X_i | X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n). \end{aligned}$$

Details of the derivation of the lower bound in (4) can be found in Appendix A.

The definitions of common entropy G and Wyner common information J continue to hold for continuous random variables. The operational implications of Wyner's common information for two continuous random variables are studied in [7]. Wyner's common information between scalar jointly Gaussian random variables is computed in [7], and the result is extended to Gaussian vectors in [8], and to outputs of additive Gaussian channels in [9]. Moreover, the lower bound on J in (4) continues to hold for continuous random variables after replacing the entropy H in the definition of I_{D} with the differential entropy h. However, since the entropy of continuous random variables is infinite, there is no known upper bound on G (or J) for continuous random variables, and it is not even clear under what conditions G is finite.

In this paper, which is a more complete version of [10], we devise a computationally efficient scheme for constructing a common randomness variable W for distributed simulation of n continuous random variables and establish upper bounds on its entropy, which in turn provide upper bounds on G. In particular we establish the following upper bound on G when the pdf of X^n is *log-concave*

$$I_{\mathrm{D}} \le J \le G \le I_{\mathrm{D}} + n^2 \log_2 e + 9n \log_2 n. \tag{5}$$

For $n = 2$, this bound reduces to

$$I(X_1; X_2) \le J(X_1; X_2) \le G(X_1; X_2) \le I(X_1; X_2) + 24.$$

Applying this result to two jointly Gaussian random variables shows that only a finite amount of common randomness is needed for their distributed simulation.

Note that the above upper bound also provides an upper bound on Wyner's common information between n continuous random variables with log-concave pdf. This result is interesting in its own right since computing Wyner's common information for n continuous random variables is very difficult in general and there is no previously known upper bound on it.

Our distributed randomness simulation scheme uses a dyadic decomposition procedure to construct the common randomness variable W. For X^n uniformly distributed over a set A, our decomposition method partitions A into hypercubes. The common randomness W is defined as the position and size of the hypercube that contains $X_1, \ldots, X_n$. Conditioned on W, the random variables X^n are independent and uniformly distributed over line segments, which when combined with local randomness facilitates distributed exact simulation. Since bounding $H(W)$ directly is quite difficult, we bound it using the *erosion entropy* of the set, which is a new measure that is shift-invariant. This scheme is extended to non-uniformly distributed X^n by performing the same dyadic decomposition on the positive part of the hypograph of the pdf of X^n. The entropy of the resulting W is then bounded using a quantity called *truncated differential entropy* to obtain the upper bound in (5).

The cardinality of the random variable W needed for exact distributed simulation of continuous random variables is in general infinite. By terminating the dyadic decomposition at a finite iteration, however, we show that the random variables can be approximately simulated using a fixed length scheme such that for log-concave pdfs, the total variation distance between the simulated and prescribed pdfs can be bounded as a function of the dual total correlation and the number of common random bits. This result provides an upper bound on the one-shot version of Wyner's common information with total variation constraint.

The rest of the paper is organized as follows. In Section II, we introduce the aforementioned dyadic decomposition scheme and establish an upper bound on G when the random variables are uniformly distributed over an orthogonally convex set. In Section III, we extend this bound to non-uniform distributions with orthogonally concave pdf and establish our main result on log-concave pdfs. In Section IV, we establish an upper bound on the one-shot version of Wyner's common information with total variation constraint. In Appendix B, we provide details on the implementation of the coding scheme for constructing the common randomness variable.

A. Notation

Throughout this paper, we assume that log is base 2 and the entropy H is in bits. We use the notation: $[a:b] = [a,b] \cap \mathbb{Z}$ and $X_{[1:n]\backslash i} = (X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)$.

A set $A \subseteq \mathbb{R}^n$ is said to be *orthogonally convex* if for any line L parallel to one of the n axes, $L \cap A$ is a connected set (empty, a point, or an interval). A function f is said to be orthogonally concave if its *hypograph* $\{(x, \alpha) : x \in \mathbb{R}^n, \alpha \le f(x)\}$ is orthogonally convex.

We denote the i-th standard basis vector of $\mathbb{R}^n$ by e_i. We denote the volume of a Lebesgue measurable set $A \subseteq \mathbb{R}^n$ by $\mathrm{V}_n(A) = \int_{\mathbb{R}^n} \mathbf{1}_A(x)dx$. If $A \subseteq B \subseteq \mathbb{R}^n$, where B is an m-dimensional affine subspace, we denote the m-dimensional volume of A by $\mathrm{V}_m(A) = \int_B \mathbf{1}_A(x)dx$.

We define the projection of a point $x \in \mathbb{R}^n$ as

$$\mathrm{P}_{i_1,\ldots,i_k}(x) = (x_{i_1}, \ldots, x_{i_k}) \in \mathbb{R}^k,$$

and the projection of a set $A \subseteq \mathbb{R}^n$ onto the dimensions $i_1, \ldots, i_k$ as

$$\mathrm{P}_{i_1,\ldots,i_k}(A) = \left\{(x_{i_1}, \ldots, x_{i_k}) : x \in A\right\} \subseteq \mathbb{R}^k.$$

We use the shorthand notation

$$\begin{aligned} \mathrm{P}_{\backslash i}(A) &= \mathrm{P}_{1,2,\ldots,i-1,i+1,\ldots,n}(A), \\ \mathrm{VP}_{i_1,\ldots,i_k}(A) &= \mathrm{V}_k(\mathrm{P}_{i_1,\ldots,i_k}(A)), \\ \mathrm{VP}_{\backslash i}(A) &= \mathrm{V}_{n-1}(\mathrm{P}_{\backslash i}(A)). \end{aligned}$$

For $A, B \subseteq \mathbb{R}^n$, $A + B$ denotes the Minkowski sum $\{a + b : a \in A, b \in B\}$, and for $x \in \mathbb{R}^n$, $A + x = \{a + x : a \in A\}$. For $\gamma \in \mathbb{R}$, $\gamma A = \{\gamma a : a \in A\}$. For $M \in \mathbb{R}^{n \times n}$, $MA = \{Ma : a \in A\}$. The *erosion* of the set A by the set B is defined as $A \ominus B = \{x \in \mathbb{R}^n : B + x \subseteq A\}$.

For a set $A \subseteq \mathbb{R}^n$ where $0 \in A$, the radial function $\rho_A : \mathbb{R}^n \to \mathbb{R}$ is defined as $\rho_A(x) = \sup\{\lambda \ge 0 : \lambda x \in A\}$.

II. Uniform Distribution Over a Set

In this section, we study the case where the desired distribution is the uniform distribution over a set $A \subseteq \mathbb{R}^n$.

We first define the dyadic decomposition of a set, which is the building block of our distributed randomness simulation scheme.

Definition 1 (Dyadic Decomposition): For $v \in \mathbb{Z}^n$ and $k \in \mathbb{Z}$, we define the hypercube $C_{k,v} = 2^{-k}([0,1]^n + v) \subset \mathbb{R}^n$ (see Figure 2). For a set $A \subseteq \mathbb{R}^n$ with a boundary of measure zero and $k \in \mathbb{Z}$, define the set

$$D_k(A) = \left\{v \in \mathbb{Z}^n : C_{k,v} \subseteq A \text{ and } C_{k-1,\lfloor v/2 \rfloor} \nsubseteq A\right\},$$

where $\lfloor v/2 \rfloor$ is the vector formed by the entries $\lfloor v_i/2 \rfloor$.

The *dyadic decomposition* of A is the partitioning of A into hypercubes $\left\{C_{k,v}\right\}$ such that $v \in D_k(A)$ and $k \in \mathbb{Z}$. Since every point x in the interior of A is contained in some hypercube in A, the interior is contained in $\cup_{k \in \mathbb{Z},\, v \in D_k(A)} C_{k,v}$, and the set of points in A not covered by the hypercubes has measure zero.

For $X^n \sim \mathrm{Unif}(A)$, denote by $C_{K,V}$, $V \in D_K(A)$, the hypercube that contains X^n and let the *dyadic decomposition random variable* $W_{\mathrm{D}} = (K, V)$. Then conditioned on $W_{\mathrm{D}} = (k, v)$, $X^n \sim \mathrm{Unif}(C_{k,v})$, that is, $X_1, \ldots, X_n$ are conditionally independent given W_{D}. Hence, we can use the dyadic decomposition to perform distributed randomness simulation as follows.

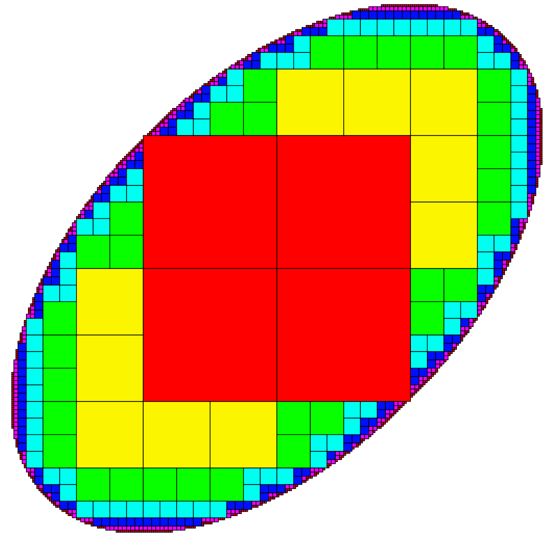

Fig. 2. Dyadic decomposition of the uniform pdf over the ellipse in Example 1.

1) The agents agree on the dyadic decomposition of the set A and on an optimal discrete distribution generating tree [1] that assigns outcomes of the random bits $b_1, b_2, \ldots$ into values of $w_A = (k, v)$, that is, labels for the hypercubes.
2) Upon observing $b_1, b_2, \ldots$, agent i generates $w_A = (k, v)$ using the discrete distribution generating tree.
3) Agent $i \in [1:n]$ generates $X_i \sim \mathrm{Unif}[2^{-k}v_i, 2^{-k}(v_i + 1)]$, that is, uniformly distributed over its corresponding side of the hypercube.

The implementation details of this scheme are provided in Appendix B.

Since $X_1, \ldots, X_n$ are conditionally independent given W_{D}, we have

$$G(X_1; \ldots; X_n) \le H(W_{\mathrm{D}}).$$

To illustrate the above dyadic decomposition scheme, consider the following.

Example 1: Let $X^n \sim \mathrm{Unif}(A)$, where A is an ellipse, i.e., $A = \left\{x \in \mathbb{R}^2 : x^T K x < 1\right\}$ and K is a positive definite matrix. Figure 2 illustrates the dyadic decomposition for $K = \begin{bmatrix} 4/3 & -2/3 \\ -2/3 & 4/3 \end{bmatrix}$. Figure 3 plots the pmf of the constructed W_{D} on a log-log scale (w_i is the i-th most probable w). As can be seen, the tail of the pmf of W_{D} roughly follows a straight line, that is, the pmf of W_{D} follows a power law tail $p(w_i) \propto i^{-\alpha}$ with $\alpha \approx 2$. Hence $H(W_{\mathrm{D}})$ is finite.

The exact value of the entropy of the dyadic decomposition random variable W_{D} is very difficult to compute in general. If the set A is orthogonally convex, however, we can bound $H(W_{\mathrm{D}})$ in terms of the volumes of the set A and its projections as follows.

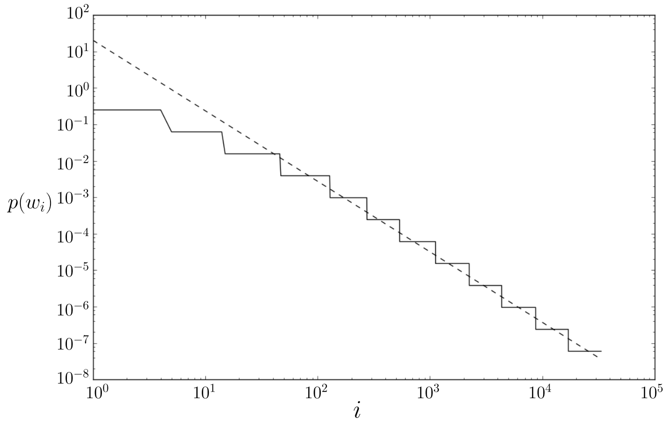

Fig. 3. The pmf of W for the dyadic decomposition of the uniform pdf over the ellipse in Example 1.

Theorem 1: Let $A \subseteq \mathbb{R}^n$ be an orthogonally convex set with $0 < \mathrm{V}_n(A) < \infty$ and $X^n \sim \mathrm{Unif}(A)$, then

$$H(W_\mathrm{D}) \le n \log\left(\sum_{i=1}^{n} \mathrm{VP}_{\setminus i}(A)\right) - (n-1)\log \mathrm{V}_n(A) + (2+\log e)n.$$

Moreover,

$$G(X_1; \cdots; X_n) \le n \log\left(\sum_{i=1}^{n} \mathrm{VP}_{\setminus i}(A)\right) - (n-1)\log \mathrm{V}_n(A) + n\log e. \quad (6)$$

The second bound is obtained using a randomization technique as in Proposition 2. It is slightly sharper than the first bound, but at the expense of being an existence result that does not directly correspond to the entropy of the dyadic decomposition of A.

Before we prove this theorem, consider the following.

Example 1 (Continued): Applying (6) to the uniform pdf over the ellipse $A = \{x \in \mathbb{R}^2 : x^T K x \le 1\}$, we obtain

$$\begin{aligned} H(W_\mathrm{D}) &\le 2\log\left(\sum_{i=1}^{2} \mathrm{VP}_{\setminus i}(A)\right) - \log \mathrm{V}_2(A) + 4 + 2\log e \\ &= 2\log\left(2\sqrt{\frac{K_{11}}{\det K}} + 2\sqrt{\frac{K_{22}}{\det K}}\right) \\ &\quad - \log\left(\pi\sqrt{\frac{1}{\det K}}\right) + 4 + 2\log e \\ &= \log\left(\pi^{-1}\frac{\left(\sqrt{K_{11}}+\sqrt{K_{22}}\right)^2}{\sqrt{\det K}}\right) + 6 + 2\log e. \end{aligned}$$

Comparing this to the mutual information for the uniform pdf over the ellipse, we have

$$I(X_1; X_2) = \log\left(\pi e^{-1}\sqrt{\frac{K_{11}K_{22}}{\det K}}\right).$$

Note that the gap between $H(W_\mathrm{D})$ and $I(X_1; X_2)$ depends on the ratio between $\left(\sqrt{K_{11}}+\sqrt{K_{22}}\right)^2$ and $\sqrt{K_{11}K_{22}}$, which becomes very large when $K_{11} \gg K_{22}$. For example, if $K = \mathrm{diag}(10000, 1)$, then $\sqrt{K_{11}K_{22}} = 100$ and $I(X_1; X_2) \approx 0.21$. On the other hand, $\left(\sqrt{K_{11}}+\sqrt{K_{22}}\right)^2 = 10201$ and the bound on $H(W_\mathrm{D})$ is 13.02. In Section II-B we show that this gap can be reduced and bounded by a constant by appropriately scaling A.

A. Proof of Theorem 1

We begin with an outline of the proof:

1) The difficulty in evaluating $H(W_\mathrm{D})$ arises from the complexity of the pmf of W_D. To see this, let $X^n \sim \mathrm{Unif}(A)$ and Ψ be the side length of the dyadic hypercube containing the point X^n. Then the entropy of W can be expressed as
$$H(W_\mathrm{D}) = \mathsf{E}[-\log\ (\Psi^n/\mathrm{V}_n(A))].$$
The expectation in this expression is taken over the pmf of Ψ, which is quite complicated because the dyadic hypercubes are positioned at grid points. This makes $H(W_\mathrm{D})$ dependent on shifting and scaling, which do not affect the common entropy G.
To simplify the analysis, we introduce the erosion entropy defined as
$$\mathsf{E}[-\log\ (\Phi)],$$
where $\Phi = \sup\{\phi : [0,\phi]^n + X \subseteq A\}$, that is, the side length of the largest hypercube positioned at X^n in A. Note that this quantity is easier to analyze since it is shift-invariant because the hypercube can be positioned at any point in A. A more general definition of erosion entropy is given below and some of its properties are listed in Proposition 1. Proposition 2 shows that $H(W_\mathrm{D})$ can be bounded by erosion entropy.
2) The next step in the proof is to bound the erosion entropy for orthogonally convex A. Note that Φ is small only when X^n is close to the boundary of A, hence $\mathsf{P}\{\Phi \le \gamma\}$ can be bounded by the volume of the boundary or the volumes of the projections of A, under the orthogonal convexity assumption. Lemma 1 bounds the volume of the erosion of A by a hypercube $\mathrm{V}_n(A \ominus [0,\gamma]^n) = \mathrm{V}_n(A)(1 - \mathsf{P}\{\Phi \le \gamma\})$ using volumes of the projections of A. This result is then applied to bound the erosion entropy $\mathsf{E}[-\log\ (\Phi)]$ to complete the proof of the theorem.

We are now ready to give the details of the proof.

Definition 2 (Erosion Entropy): The *erosion entropy* of a set $A \subseteq \mathbb{R}^n$ with $0 < \mathrm{V}_n(A) < \infty$ by a convex set $\Delta \subseteq \mathbb{R}^n$ is defined as

$$h_{\ominus\Delta}(A) = \int_{-\infty}^{\infty}\left(\mathbf{1}\{t \ge 0\} - \frac{\mathrm{V}_n\left(A \ominus 2^{-t}\Delta\right)}{\mathrm{V}_n(A)}\right)dt,$$

where $A \ominus \Delta = \{x \in \mathbb{R}^n : \Delta + x \subseteq A\}$ is the erosion of A by Δ.

The erosion entropy roughly measures the ratio of the surface area to the volume of the set A. To see this, assume that

$0 \in \Delta$ and let $X^n \sim \mathrm{Unif}(A)$ and $\Phi = \sup\{\phi : \phi\Delta + X \subseteq A\}$, then we can rewrite the erosion entropy as

$$\begin{aligned} h_{\ominus\Delta}(A) &= \int_{-\infty}^{\infty} \left(\mathbf{1}\{t \ge 0\} - \mathsf{P}\left\{X^n \in A \ominus 2^{-t}\Delta\right\}\right) dt \\ &= \int_{-\infty}^{\infty} \left(\mathbf{1}\{t \ge 0\} - \mathsf{P}\left\{2^{-t}\Delta + X^n \subseteq A\right\}\right) dt \\ &= \int_{-\infty}^{\infty} \left(\mathbf{1}\{t \ge 0\} - \mathsf{P}\{t \ge -\log \Phi\}\right) dt \\ &= \mathsf{E}[-\log(\Phi)]. \end{aligned}$$

If we further assume that $\Delta = [0, 1]^n$, then a large erosion entropy means that the largest hypercube positioned at a random point in A is small, which suggests that A has a large surface area to volume ratio.

We now state some basic properties of the erosion entropy.

Proposition 1: For a set $A \subseteq \mathbb{R}^n$ with $0 < \mathrm{V}_n(A) < \infty$, convex sets $\Delta, \Delta_1, \Delta_2 \subseteq \mathbb{R}^n$, and non-singular $M \in \mathbb{R}^{n\times n}$, the erosion entropy satisfies the following.

1) *Monotonicity. If $\Delta_1 \subseteq \Delta_2$, then $h_{\ominus\Delta_1}(A) \le h_{\ominus\Delta_2}(A)$.*
2) *Scaling. $h_{\ominus\beta\Delta}(\alpha A) = h_{\ominus\Delta}(A) + \log(\beta/\alpha)$.*
3) *Linear transformation. $h_{\ominus MB}(MA) = h_{\ominus\Delta}(A)$.*
4) *Union. If $A_1, \ldots, A_k \subseteq \mathbb{R}^n$ are disjoint, then*
$$h_{\ominus\Delta}\left(\bigcup_{i=1}^{k} A_i\right) \le \sum_{i=1}^{k} \frac{\mathrm{V}_n(A_i)}{\mathrm{V}_n(\bigcup_j A_j)} \cdot h_{\ominus\Delta}(A_i).$$
Equality holds when the closures of $A_1, \ldots, A_k \subseteq \mathbb{R}^n$ are disjoint.
5) *Reduction to differential entropy. If $X^n \sim \mathrm{Unif}(A)$, and $A \cap L$ is connected for all lines L parallel to the n-th axis, then*
$$h_{\ominus\{0\}^{n-1}\times[0,1]}(A) = h(X^{n-1}) - \log \mathrm{V}_n(A) + \log e.$$
As a result, for general continuous random variables X^n with pdf f,
$$h_{\ominus\{0\}^{n}\times[0,1]}(\mathrm{hyp}_+ f) = h(X^n) + \log e,$$
where $\mathrm{hyp}_+ f = \{(x, \alpha) : x \in \mathbb{R}^n,\ 0 < \alpha < f(x)\} \subseteq \mathbb{R}^{n+1}$.

The proofs of these properties are given in Appendix C.

In the following proposition we show that $H(W_{\mathrm{D}})$ can be upper bounded using the erosion entropy. Moreover we show that the erosion entropy is the average of the dyadic decomposition entropy under random shifting and scaling.

Proposition 2: For a set $A \subseteq \mathbb{R}^n$ with a boundary of measure zero, we have

$$H(W_{\mathrm{D}}) \le \log \mathrm{V}_n(A) + n h_{\ominus[0,1]^n}(A) + 2n.$$

Moreover, for any $T \in \mathbb{Z}$, $T > (1/n)\log \mathrm{V}_n(A) + 1$, when $U^n \sim \mathrm{Unif}[0, 2^T]$ i.i.d., $\Theta \sim \mathrm{Unif}[0, 1]$ independent of U^n and $\Lambda = 2^{\Theta}$, let $\tilde{W}_{\mathrm{D}}$ be the dyadic decomposition random variable corresponding to the random set $\Lambda A + U$, then we have

$$E_{\Lambda,U}\left[H(\tilde{W}_{\mathrm{D}})\right] = \log \mathrm{V}_n(A) + n h_{\ominus[0,1]^n}(A).$$

The proof of this proposition is given in Appendix D. If the set A is not orthogonally convex but can be partitioned into orthogonally convex sets, then the property of erosion entropy of union of sets in Proposition 1 can be used to bound $H(W_{\mathrm{D}})$.

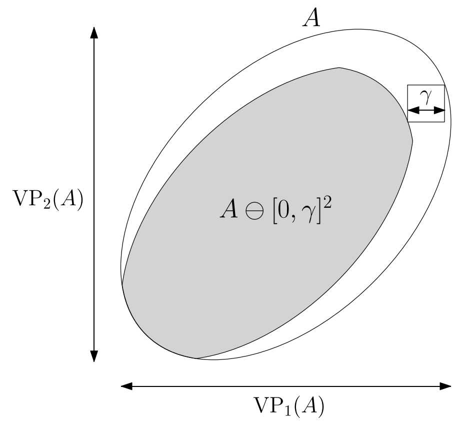

Fig. 4. The erosion $A \ominus [0, \gamma]^2$ when A is an ellipse.

To prove Theorem 1, we need the following lemma, which bounds the volume of the erosion of A by a hypercube.

Lemma 1: For any orthogonally convex set $A \subseteq \mathbb{R}^n$ with $0 < \mathrm{V}_n(A) < \infty$ and $\gamma \ge 0$,

$$\begin{aligned} \mathrm{V}_n\left(A \ominus [0, \gamma]^n\right) \ge \mathrm{V}_n(A) - \sum_{i=1}^{n} \int_{\mathrm{P}_{\setminus i}(A)} \min\{\gamma,\\ \mathrm{V}_1\left(A \cap (\mathrm{span}(\mathrm{e}_i) + x)\right)\} dx_{[1:n]\setminus i}, \end{aligned}$$

where $\mathrm{span}(\mathrm{e}_1) + x = \{(\alpha, x_2, x_3, \ldots, x_n) : \alpha \in \mathbb{R}\} \subseteq \mathbb{R}^n$. *As a result,*

$$\mathrm{V}_n\left(A \ominus [0, \gamma]^n\right) \ge \mathrm{V}_n(A) - \gamma \sum_{i=1}^{n} \mathrm{VP}_{\setminus i}(A).$$

The proof of this lemma is given in Appendix E. For example, when $n = 2$, we have $\mathrm{V}_2\left(A \ominus [0, \gamma]^2\right) \ge \mathrm{V}_2(A) - \gamma\,\mathrm{VP}_1(A) - \gamma\,\mathrm{VP}_2(A)$, giving a bound on the volume of erosion by the "width" $\mathrm{VP}_1(A)$ and "height" $\mathrm{VP}_2(A)$; see Figure 4.

To complete the proof of Theorem 1, note that by Proposition 2, the theorem can be proved by bounding $h_{\ominus[0,1]^n}(A)$. By Lemma 1, we have

$$\begin{aligned} &h_{\ominus[0,1]^n}(A) \\ &= \int_{-\infty}^{\infty} \left(\mathbf{1}\{t \ge 0\} - \frac{\mathrm{V}_n\left(A \ominus [0, 2^{-t}]^n\right)}{\mathrm{V}_n(A)}\right) dt \\ &\le \int_{-\infty}^{\infty} \left(\mathbf{1}\{t \ge 0\} - \max\left(0,\ 1 - 2^{-t}\frac{\sum_{i=1}^{n} \mathrm{VP}_{\setminus i}(A)}{\mathrm{V}_n(A)}\right)\right) dt \\ &= \int_{-\infty}^{\infty} \left(\mathbf{1}\{t \ge 0\} - \max\left(0, 1 - 2^{-\left(t - \log\left(\frac{\sum_i \mathrm{VP}_{\setminus i}(A)}{\mathrm{V}_n(A)}\right)\right)}\right)\right) dt \\ &= \int_{-\infty}^{\infty} \left(\mathbf{1}\left\{t \ge -\log\left(\frac{\sum_i \mathrm{VP}_{\setminus i}(A)}{\mathrm{V}_n(A)}\right)\right\}\right. \\ &\qquad\qquad \left. - \max\left(0, 1 - 2^{-t}\right)\right) dt \\ &= \log\left(\frac{\sum_i \mathrm{VP}_{\setminus i}(A)}{\mathrm{V}_n(A)}\right) \\ &\quad + \int_{-\infty}^{\infty} \left(\mathbf{1}\{t \ge 0\} - \max\left(0, 1 - 2^{-t}\right)\right) dt \\ &= \log\left(\frac{\sum_i \mathrm{VP}_{\setminus i}(A)}{\mathrm{V}_n(A)}\right) + \int_{0}^{\infty} 2^{-t} dt \\ &= \log\left(\frac{\sum_i \mathrm{VP}_{\setminus i}(A)}{\mathrm{V}_n(A)}\right) + \log e. \end{aligned}$$

For the second result, note that the randomization in Proposition 2 (i.e., considering $\Lambda A + U$, a randomly shifted and scaled version of A) does not affect the right hand side of Theorem 1, which completes the proof of the theorem.

B. Scaling

A shortcoming of dyadic decomposition is that it uses only hypercubes, not hyper-rectangles. Consider Example 1 with $K = \text{diag}(10000, 1)$, which corresponds to a very thin ellipse. Since only small squares can fit in this thin ellipse, $H(W_\text{D})$ would be quite large. However, if we scale X_1 by a factor of 100 before applying dyadic decomposition, then the ellipse will be scaled to a circle, and $H(W_\text{D})$ would be considerably smaller. As such, in this section, we present a tighter bound on the common entropy between continuous random variables by first scaling A along each dimension, that is, by performing a linear transformation DA where D is a diagonal matrix. This corresponds to scaling the random variable X_i by D_{ii}, $i \in [1:n]$, before applying the scheme. Since this scaling can be part of the agreed upon protocol (together with the dyadic decomposition of DA and the discrete distribution tree), each agent i can simply scale back its generated random variable $D_{ii}X_i$ to generate X_i.

Let W_DS be the dyadic decomposition random variable corresponding to uniform distribution over the scaled set DA. While Theorem 1 uses the volume of projections to bound W_D, we introduce a more refined quantity called truncated differential entropy to be used in a sharper bound, which will be useful in proving our main result in Section III.

Definition 3 (Truncated Differential Entropy): Let $X^n \sim f(x^n)$ and define its truncated differential entropy $\tilde{h}_\zeta(X^n)$ for $\zeta \in (0, 1]$, as

$$\tilde{h}_\zeta(X^n) = \int_{\mathbb{R}^n} -\zeta^{-1} \min\{\xi,\ f(x)\} \log\Big(\zeta^{-1} \min\{\xi,\ f(x)\}\Big)\, dx,$$

where $\xi > 0$ such that

$$\int_{\mathbb{R}^n} \min\{\xi,\ f(x)\}\, dx = \zeta,$$

i.e., $\tilde{h}_\zeta(X^n)$ is the differential entropy of the pdf $\zeta^{-1}\min\{\xi,\ f(x)\}$. Also define

$$\tilde{h}_0(X^n) = \lim_{\zeta \to 0} \tilde{h}_\zeta(X^n) = \log \mathrm{V}_n\{x : f(x) > 0\}.$$

Note that $\tilde{h}_\zeta(X^n)$ is monotonically decreasing in ζ from $\tilde{h}_0(X^n)$ (the entropy of the uniform pdf on the support of X^n) to $\tilde{h}_1(X^n) = h(X^n)$. The proof of this fact is given in Appendix F.

We now state the main result of this section, which shows that the gap between $H(W_\text{DS})$ and I_D depends on how close $\tilde{h}_{1/n}(X_{[1:n]\setminus i})$ and $h(X_{[1:n]\setminus i})$, $i \in [1:n]$, are to each other.

Theorem 2: For any orthogonally convex set $A \subseteq \mathbb{R}^n$ with $0 < \mathrm{V}_n(A) < \infty$, $X^n \sim \text{Unif}(A)$, there exists a diagonal matrix $D \in \mathbb{R}^{n\times n}$ with positive diagonal entries such that the entropy of the dyadic decomposition of $DA = \{Dx : x \in A\}$ is bounded by

$$H(W_\text{DS}) \le \sum_{i=1}^{n} \tilde{h}_{1/n}(X_{[1:n]\setminus i}) - (n-1)\log \mathrm{V}_n(A) + n\log n + (2+\log e)n.$$

Equivalently,

$$H(W_\text{DS}) \le I_\text{D}(X_1; \cdots; X_n) + \sum_{i=1}^{n}\Big(\tilde{h}_{1/n}(X_{[1:n]\setminus i}) - h(X_{[1:n]\setminus i})\Big) + n\log n + (2+\log e)n.$$

Moreover,

$$I_\text{D} \le J \le G \le I_\text{D} + \sum_{i=1}^{n}\Big(\tilde{h}_{1/n}(X_{[1:n]\setminus i}) - h(X_{[1:n]\setminus i})\Big) + n\log n + n\log e.$$

The proof of this theorem and a method for finding D are given in Appendix G.

The improvement of Theorem 2 over Theorem 1 is twofold: First, by applying a suitable scaling, we bring the summation term in the bound outside the log, giving a better bound. Second, in Theorem 1, the bound is in terms of $\mathrm{VP}_{\setminus i}(A) = 2^{\tilde{h}_0(X_{[1:n]\setminus i})}$. Theorem 2 improves upon it by using $\tilde{h}_{1/n}(X_{[1:n]\setminus i})$ instead (which is smaller than $\tilde{h}_0(X_{[1:n]\setminus i})$ by the monotonicity property). We will see in Section III that this improvement is essential to the nonuniform pdf case.

We illustrate the bound in Theorem 2 in the following.

Example 1 (Continued): Applying Theorem 2 to the uniform pdf over the ellipse $A = \{x \in \mathbb{R}^2 : x^T K x \le 1\}$, we have

$$\begin{aligned}
H(W_\text{DS}) &\le \sum_{i=1}^{2} \tilde{h}_{1/2}(X_{[1:2]\setminus i}) - \log \mathrm{V}_2(A) + 6 + 2\log e \\
&\le \sum_{i=1}^{2} \log\big(\mathrm{VP}_{\setminus i}(A)\big) - \log \mathrm{V}_2(A) + 6 + 2\log e \\
&= \log\left(2\sqrt{\frac{K_{11}}{\det K}}\right) + \log\left(2\sqrt{\frac{K_{22}}{\det K}}\right) \\
&\quad - \log\left(\pi\sqrt{\frac{1}{\det K}}\right) + 6 + 2\log e \\
&= \log\left(\pi^{-1}\sqrt{\frac{K_{11}K_{22}}{\det K}}\right) + 8 + 2\log e.
\end{aligned}$$

In comparison, the mutual information is

$$I(X_1; X_2) = \log\left(\pi e^{-1}\sqrt{\frac{K_{11}K_{22}}{\det K}}\right),$$

and the gap between $H(W_\text{DS})$ and $I(X_1; X_2)$ is bounded by a constant. Figure 5 plots the values of $H(W_\text{DS})$ (calculated by finding all squares in the dyadic decomposition with side length at least 2^{-11}, which yields a precise estimate), the upper bound in Theorem 2, and $I(X_1; X_2)$ for $K = \frac{1}{1-t^2}\begin{bmatrix} 1 & -t \\ -t & 1 \end{bmatrix}$, $t \in [0, 1]$.

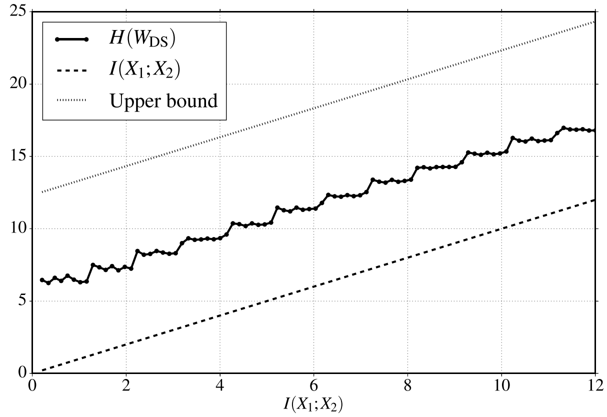

Fig. 5. Plot of the entropy of the dyadic decomposition $H(W_{\mathrm{DS}})$, mutual information $I(X_1; X_2)$, and the upper bound in Theorem 2 against $I(X_1; X_2)$ for Example 1.

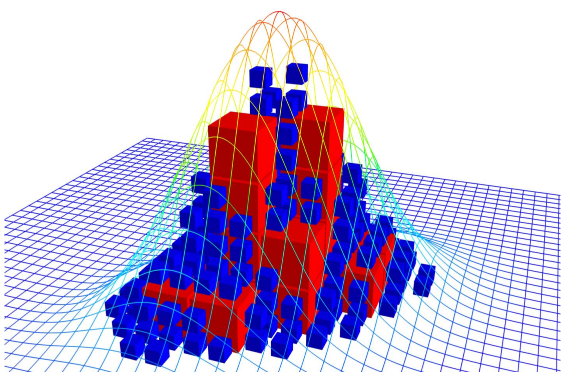

Fig. 6. Dyadic decomposition of $\mathrm{hyp}_+ f$ for the Gaussian pdf f in Example 2.

III. Nonuniform Distributions

In this section, we extend our results to the case in which the pdf of X^n is not necessarily uniform. Let $X^n \sim f(x^n)$ and let the support of f be A. We add a random variable Z such that $(X_1, \ldots, X_n, Z) \sim \mathrm{Unif}(\mathrm{hyp}_+ f)$, where $\mathrm{hyp}_+ f$ is the *positive strict hypograph* defined as

$$\mathrm{hyp}_+ f = \left\{(x, \alpha) : x \in \mathbb{R}^n,\ 0 < \alpha < f(x)\right\} \subseteq \mathbb{R}^{n+1}.$$

Note that the marginal pdf of X^n is f. Assuming that $\mathrm{hyp}_+ f$ is orthogonally convex, i.e., f is orthogonally concave, we can apply the results for the uniform pdf case in Section II. To illustrate this extension, consider the following.

Example 2: Let (X_1, X_2) be zero mean Gaussian with covariance matrix $K = \begin{bmatrix} 1/8 & 1/16 \\ 1/16 & 1/8 \end{bmatrix}$. Figure 6 plots the cubes with side length $\geq 2^{-3}$ of the dyadic decomposition of the positive strict hypograph of this pdf. Note that the cubes are scaled down so as to show the ones behind them. Figure 7 plots the pmf of $W_{\mathrm{hyp}_+ f}$ in log-log scale. As in Example 1, the tail of the pmf follows a power law and $H(W_{\mathrm{hyp}_+ f})$ is finite.

We now show that if f is log-concave (i.e., $x \mapsto \log f(x)$ is concave), then the difference between the entropy of the dyadic decomposition and the dual total correlation is bounded by a constant that depends only on n.

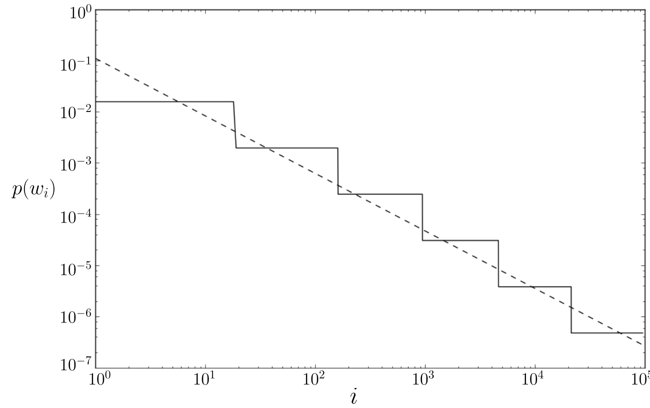

Fig. 7. The pmf of W for the dyadic decomposition of the hypograph of the Gaussian pdf in Example 2.

Theorem 3: If the pdf of X^n is log-concave, then there exists a diagonal matrix $D \in \mathbb{R}^{(n+1)\times(n+1)}$ with positive diagonal entries such that the dyadic decomposition random variable W_{DH} corresponding to the set $D\mathrm{hyp}_+ f$ satisfies

$$\begin{aligned} H(W_{\mathrm{DH}}) &\leq I_{\mathrm{D}}(X_1; \cdots; X_n) + n^2 \log e \\ &\quad + n\left(\log n + \log(n+1) + e + 2\log e + 2\right) + 2 + \log e \\ &\leq I_{\mathrm{D}}(X_1; \cdots; X_n) + n^2 \log e + 12 n \log n. \end{aligned}$$

Moreover,

$$I_{\mathrm{D}} \leq J \leq G \leq I_{\mathrm{D}} + n^2 \log e + 9 n \log n.$$

Theorem 3 will be proved by invoking Theorem 2. Theorem 1 is not sufficient here because $\mathrm{VP}_{\setminus (n+1)}(\mathrm{hyp}_+ f)$ can be infinite for distributions with unbounded support. To apply Theorem 2 on $\mathrm{hyp}_+ f$, we have to bound $\tilde{h}_{1/(n+1)}(X^n)$ and $\tilde{h}_{1/(n+1)}(X_{[1:n]\setminus i}, Z)$ (we will see later that it suffices to bound $\tilde{h}_{1/(n+1)}(X_{[1:n]\setminus i}, Z)$ by $\mathrm{VP}_{\setminus i}(\mathrm{hyp}_+ f)$ for $i \leq n$). We first state a result in [11], which bounds the differential entropy of a log-concave pdf by its maximum density.

Lemma 2: For any log-concave pdf f of X^n,

$$-\log\Big(\sup_{x\in\mathbb{R}^n} f(x)\Big) \leq h(X^n) \leq -\log\Big(\sup_{x\in\mathbb{R}^n} f(x)\Big) + n \log e.$$

We can generalize Lemma 2 to bound the conditional differential entropy as follows.

Lemma 3: For any log-concave pdf f of X^n, and $1 \leq m \leq n$,

$$\begin{aligned} 0 &\leq h(X_{m+1}^n \mid X^m) + \log\Big(\int_{\mathbb{R}^m} \sup_{\tilde{x}_{m+1}^n \in \mathbb{R}^{n-m}} f(x^m, \tilde{x}_{m+1}^n) dx^m\Big) \\ &\leq n \log e + \log \binom{n}{m}. \end{aligned}$$

The proof of this lemma is given in Appendix H. By setting $m = n-1$, the integral term becomes $\mathrm{VP}_{\setminus n}(\mathrm{hyp}_+ f)$. Hence we can use this lemma to bound $\mathrm{VP}_{\setminus i}(\mathrm{hyp}_+ f)$ for $i = 1, \ldots, n$, which in turn bounds $\tilde{h}_{1/(n+1)}(X_{[1:n]\setminus i}, Z)$.

Next, we establish a bound on the difference between the differential entropy and the truncated differential entropy.

Lemma 4: For any log-concave pdf f,

$$\tilde{h}_\zeta(X^n) - h(X^n) \le \log\zeta + \nu + n\log e,$$

where $\nu \ge 0$ satisfies

$$\Gamma(n+1,\, \nu) = \zeta \cdot \Gamma(n+1),$$

and $\Gamma(n,z) = \int_z^\infty t^{n-1}e^{-t}dt$ is the incomplete gamma function, and $\Gamma(n) = \Gamma(n,0)$ is the gamma function. Moreover, if $\zeta \ge e^{-(e-2)n}$, then

$$\tilde{h}_\zeta(X^n) - h(X^n) \le \log\zeta + (e + \log e)n.$$

The proof of this lemma is given in Appendix I.

We now proceed to the proof of Theorem 3.

Proof of Theorem 3: Let $(X^n, Z) \sim \mathrm{Unif}(\mathrm{hyp}_+ f)$. Applying Theorem 2 on $\mathrm{hyp}_+ f$, we have

$$\begin{aligned} H(W_{\mathrm{DH}}) &\le \tilde{h}_{1/(n+1)}(X^n) + \sum_{i=1}^n \tilde{h}_{1/(n+1)}(X_{[1:n]\setminus i},\, Z) \\ &\quad + (n+1)\log(n+1) + (2+\log e)(n+1) \\ &\le \tilde{h}_{1/(n+1)}(X^n) + \sum_{i=1}^n \log\left(\mathrm{VP}_{\setminus i}(\mathrm{hyp}_+ f)\right) \\ &\quad + (n+1)\log(n+1) + (2+\log e)(n+1). \end{aligned}$$

Consider the term $\tilde{h}_{1/(n+1)}(X^n)$. Since $1/(n+1) \ge e^{-(e-2)n}$, by Lemma 4 we have

$$\tilde{h}_{1/(n+1)}(X^n) - h(X^n) \le -\log(n+1) + (e+\log e)n.$$

Consider the term $\log\left(\mathrm{VP}_{\setminus n}(\mathrm{hyp}_+ f)\right)$. By Lemma 3, we have

$$\begin{aligned} &\log\left(\mathrm{VP}_{\setminus n}(\mathrm{hyp}_+ f)\right) \\ &\quad = \log \int_{\mathbb{R}^{n-1}} \sup_{\tilde{x}_n} f(x_1^{n-1}, \tilde{x}_n) dx_1^{n-1} \\ &\quad \le -h(X_n | X_1^{n-1}) + n\log e + \log\binom{n}{n-1} \\ &\quad = -h(X_n | X_1^{n-1}) + n\log e + \log n. \end{aligned}$$

Hence

$$\begin{aligned} &H(W_{\mathrm{DH}}) \\ &\quad \le h(X^n) - \log(n+1) + (e+\log e)n \\ &\qquad + \sum_{i=1}^n \left(-h(X_i | X_{[1:n]\setminus i}) + n\log e + \log n\right) \\ &\qquad + (n+1)\log(n+1) + (2+\log e)(n+1) \\ &\quad = I_{\mathrm{D}}(X_1; \cdots; X_n) + n^2\log e \\ &\qquad + n\left(\log n + \log(n+1) + e + 2\log e + 2\right) + 2 + \log e. \end{aligned}$$

The second result can be proved using the randomization in Proposition 2, i.e., considering $\Lambda A + U$, a randomly shifted and scaled version of A. ■

Note that the result in Theorem 3 can be readily extended to mixtures of log-concave pdfs using the union property in Proposition 1. It is not possible to obtain a constant bound on the gap between I_{D} and G for arbitrary pdfs, however. To see this, let $(\tilde{X}_1, \tilde{X}_2) \in \{1, \ldots, 2^m\}^2$ be two discrete random variables with pmf $\begin{bmatrix} 1/3 & 1/3 \\ 1/3 & 0 \end{bmatrix}^{\otimes m}$, i.e., $(\tilde{X}_1, \tilde{X}_2)$ consists of m i.i.d. copies of two random variables with pmf $\begin{bmatrix} 1/3 & 1/3 \\ 1/3 & 0 \end{bmatrix}$. Now let $X_1 = \tilde{X}_1 + Z_1$ and $X_2 = \tilde{X}_2 + Z_2$, where $Z_1, Z_2 \sim \mathrm{Unif}[0,1]$. Then we have $I(X_1; X_2) = I(\tilde{X}_1; \tilde{X}_2) = (\log 3 - 4/3)m$, $J(X_1; X_2) = J(\tilde{X}_1; \tilde{X}_2) = (2/3)m$, and the gap between I and J grows linearly in m. Since $G \ge J$, the gap between I and G grows at least linearly in m.

IV. Approximate Distributed Simulation

The cardinality of W for exact simulation of n continuous random variables is in general infinite (i.e., the number of common random bits $B_1, B_2, \ldots$ needed is unbounded). We show that if the exact simulation requirement is relaxed by only requiring that the total variation distance between the distributions of the simulated and the prescribed random variables to be small, then distributed simulation is possible with a fixed number of common random bits.

We define the approximate distributed simulation problem as follows. There are n agents that have access to common randomness B^N i.i.d. Bern(1/2). Agent $i \in [1:n]$ wishes to simulate the random variable $\tilde{X}_i$ using B^N and its local randomness, which is independent of B^N and local randomness at other agents, such that the total variation between the distributions of $\tilde{X}^n$ and X^n is bounded as

$$d_{\mathrm{TV}}\left((\tilde{X}_1, \ldots, \tilde{X}_n),\, (X_1, \ldots, X_n)\right) \le \epsilon,$$

for some $\epsilon > 0$. The problem is to find the conditions under which the length-distance pair (N, ϵ) is achievable.

We can find sufficient conditions under which (N, ϵ) is achievable by terminating the dyadic decomposition scheme described in the previous sections after a finite number of iterations, that is, by discarding all hypercubes smaller than a prescribed size. The following proposition gives the length-distance pairs achievable by this truncated dyadic decomposition scheme in terms of $H(W_{\mathrm{D}})$ for uniform pdf over A (or $H(W_{\mathrm{DH}})$ for non-uniform pdfs), which in turn can be bounded using Theorem 1 or 2 for uniform distributions, or Theorem 3 for log-concave distributions.

Theorem 4: Let $X^n \sim \mathrm{Unif}(A)$, where $A \subseteq \mathbb{R}^n$ with a boundary of measure zero. The truncated dyadic decomposition scheme can achieve the length-distance pair (N, ϵ) if

$$\epsilon \ge 2^{-N} + N^{-1}H(W_{\mathrm{D}}).$$

Proof: Fix an arbitrary w_0. Define $\tilde{W}_{\mathrm{D}}$ with pmf

$$p_{\tilde{W}_{\mathrm{D}}}(w) = \begin{cases} 2^{-N}\left\lfloor 2^N p_{W_{\mathrm{D}}}(w)\right\rfloor & \text{if } w \ne w_0, \\ 1 - \sum_{w' \ne w_0} 2^{-N}\left\lfloor 2^N p_{W_{\mathrm{D}}}(w')\right\rfloor & \text{if } w = w_0. \end{cases}$$

Since the values of $p_{\tilde{W}_{\mathrm{D}}}(w)$ are multiples of 2^{-N}, $\tilde{W}_{\mathrm{D}}$ can be generated using B^N. The truncated scheme uses $\tilde{W}_{\mathrm{D}}$ instead of W_{D} and the operations performed by the agents are otherwise unchanged. We have

$$\begin{aligned} &d_{\mathrm{TV}}\left((\tilde{X}_1, \ldots, \tilde{X}_n),\, (X_1, \ldots, X_n)\right) \\ &\quad \le d_{\mathrm{TV}}\left(W_{\mathrm{D}},\, \tilde{W}_{\mathrm{D}}\right) \\ &\quad = \sum_w \max\left\{p_{W_{\mathrm{D}}}(w) - p_{\tilde{W}_{\mathrm{D}}}(w),\, 0\right\} \\ &\quad \le \sum_w \left(p_{W_{\mathrm{D}}}(w) - 2^{-N}\left\lfloor 2^N p_{W_{\mathrm{D}}}(w)\right\rfloor\right) \end{aligned}$$

$$\leq \sum_{w} p_{W_{\mathrm{D}}}(w) \cdot \min\left\{1,\, 2^{-N}/p_{W_{\mathrm{D}}}(w)\right\}$$
$$\leq \sum_{w} p_{W_{\mathrm{D}}}(w) \cdot \left(2^{-N} - N^{-1}\log p_{W_{\mathrm{D}}}(w)\right)$$
$$= 2^{-N} + N^{-1}H(W_{\mathrm{D}}).$$

■

V. Conclusion

We proposed a scheme for distributed simulation of continuous random variables based on dyadic decomposition. We established a bound on the expected number of common randomness bits used in terms of the dual total correlation for the class of log concave pdfs. As a result, the gap between exact and Wyner's common information and dual total correlation can be bounded for this set of distributions.

Our results readily translate to the exact, one-shot version of the channel synthesis problem in [4], [12], and [15] without common randomness in which we wish to simulate a channel $f_{Y|X}(y|x)$ with input distribution $f_X(x)$. Given the input $X \sim f_X$, the encoder produces the codeword W using a prefix-free code. Upon receiving W, the decoder produces the output $\hat{Y}$ such that $(X, \hat{Y}) \sim f_X f_{Y|X}$. The problem again is to find the minimum entropy of W. A consequence of our results is that an additive Gaussian noise channel with Gaussian input, can be exactly simulated using only a finite expected codeword length.

We have seen in Section II-B that performing different scalings on each X_i can reduce $H(W_{\mathrm{D}})$. More generally, applying a bijective transformation $g_i(x_i)$ to each random variable before using the dyadic decomposition scheme may help reduce $H(W_{\mathrm{D}})$ further. For example, applying the copula transform [14] $g_i(x) = F_{X_i}(x)$ such that $g_i(X_i) \sim \mathrm{Unif}[0,1]$ has the benefit that when the X_i's are close to independent, the pdf is close to a constant function over the unit hypercube, which is likely to result in a smaller $H(W_{\mathrm{D}})$.

Appendix

A. Bounding Common Information by Dual Total Correlation

We show that $I_{\mathrm{D}}(X_1; \cdots ; X_n) \leq J(X_1; \cdots ; X_n)$. For general random variables, the dual total correlation is defined as

$$I_{\mathrm{D}}(X_1; \cdots ; X_n) = \sum_{i=1}^{n-1} I(X_i; X_{i+1}^n \mid X_1^{i-1}).$$

To prove the inequality, let W be a random variable such that $X_1, \ldots, X_n$ are conditionally independent given W, then

$$\begin{aligned}
&I(W; X^n)\\
&\quad = I(X_1; X_2^n) + I(W; X_2^n|X_1) + I(W; X_1|X_2^n)\\
&\quad \geq I(X_1; X_2^n) + I(W; X_2^n|X_1)\\
&\quad = I(X_1; X_2^n) + I(X_2; X_3^n|X_1)\\
&\qquad + I(W; X_3^n|X_1^2) + I(W; X_2|X_1, X_3^n)\\
&\quad \geq I(X_1; X_2^n) + I(X_2; X_3^n|X_1) + I(W; X_3^n|X_1^2)\\
&\qquad \vdots\\
&\quad \geq I(X_1; X_2^n) + I(X_2; X_3^n|X_1) + \cdots + I(X_{n-1}; X_n|X_1^{n-2}).
\end{aligned}$$

B. Dyadic Decomposition Algorithm Details

We present the simulation algorithm for $X^n \sim \mathrm{Unif}(A)$ using a similar idea as the interval algorithm for random number generation [15].

Input: Agent i, common randomness $b_1, b_2, \ldots$, $A \subseteq [0,1]^n$
Output: random variate x_i, number of bits used t

1) $\upsilon \leftarrow (0, \ldots, 0)$, $k \leftarrow 0$, $\alpha \leftarrow 0$, $(\mu, \nu) \leftarrow (0, 1)$, $t \leftarrow 0$
2) While $C_{k,\upsilon} = 2^{-k}([0,1]^n + \upsilon) \nsubseteq A$:
3) For each $\tilde{\upsilon} \in \{0,1\}^n + 2\upsilon = \{2\upsilon_1, 2\upsilon_1 + 1\} \times \cdots \times \{2\upsilon_n, 2\upsilon_n + 1\}$:
4) $p_{\tilde{\upsilon}} \leftarrow \mathrm{V}_n(A \cap C_{k+1,\tilde{\upsilon}})/\mathrm{V}_n(A)$
5) While $[\mu, \nu] \nsubseteq [\alpha + \sum_{\hat{\upsilon} \prec \tilde{\upsilon}} p_{\hat{\upsilon}},\, \alpha + \sum_{\hat{\upsilon} \preceq \tilde{\upsilon}} p_{\hat{\upsilon}}]$ for all $\tilde{\upsilon} \in \{0,1\}^n + 2\upsilon$ ($\prec$ is lexicographical order)
6) $\tilde{w} \leftarrow b_{t+1}$, $t \leftarrow t + 1$
7) $(\mu, \nu) \leftarrow (\mu + (\nu - \mu)\tilde{w}/2,\ \mu + (\nu - \mu)(\tilde{w} + 1)/2)$
8) $\upsilon \leftarrow \tilde{\upsilon}$ where $[\mu, \nu] \subseteq [\alpha + \sum_{\hat{\upsilon} \prec \tilde{\upsilon}} p_{\hat{\upsilon}},\ \alpha + \sum_{\hat{\upsilon} \preceq \tilde{\upsilon}} p_{\hat{\upsilon}}]$
9) $\alpha \leftarrow \alpha + \sum_{\hat{\upsilon} \prec \tilde{\upsilon}} p_{\hat{\upsilon}}$, $k \leftarrow k + 1$
10) Output randomly generated x_i according to $\mathrm{Unif}[2^{-k}\upsilon_i,\ 2^{-k}(\upsilon_i + 1)]$ and t

The above algorithms assume $A \subseteq [0,1]^n$. The case where A is unbounded can be handled by first generating the integer parts $\lfloor X^n \rfloor = (\lfloor X_1 \rfloor, \ldots, \lfloor X_n \rfloor)$ of $X^n \sim \mathrm{Unif}(A)$, then run the algorithms on $A \cap ([0,1]^n + \lfloor X^n \rfloor)$. The algorithms can be applied to non-uniform pdfs by letting A to be $\mathrm{hyp}_+ f$ scaled according to Theorem 2.

C. Proof of Proposition 1

The monotonicity property and the linear transformation property follow directly from the definition of erosion entropy.

For the scaling property, consider

$$\begin{aligned}
&h_{\ominus\beta\Delta}(\alpha A)\\
&\quad = \int_{-\infty}^{\infty}\left(\mathbf{1}\{t \geq 0\} - \frac{\mathrm{V}_n\left(\alpha A \ominus 2^{-t}\beta\Delta\right)}{\mathrm{V}_n(\alpha A)}\right)dt\\
&\quad = \int_{-\infty}^{\infty}\left(\mathbf{1}\{t \geq 0\} - \frac{\mathrm{V}_n\left(A \ominus 2^{-t+\log(\beta/\alpha)}\Delta\right)}{\mathrm{V}_n(A)}\right)dt\\
&\quad = \int_{-\infty}^{\infty}\left(\mathbf{1}\{t \geq -\log(\beta/\alpha)\} - \frac{\mathrm{V}_n\left(A \ominus 2^{-t}\Delta\right)}{\mathrm{V}_n(A)}\right)dt\\
&\quad \geq h_{\ominus\Delta}(A) + \log(\beta/\alpha).
\end{aligned}$$

For the union property, consider

$$\begin{aligned}
&h_{\ominus\Delta}\left(\bigcup_{i=1}^{k} A_i\right)\\
&= \int_{-\infty}^{\infty}\left(\mathbf{1}\{t \geq 0\} - \frac{\mathrm{V}_n\left(\left(\bigcup_i A_i\right) \ominus 2^{-t}\Delta\right)}{\sum_i \mathrm{V}_n(A_i)}\right)dt\\
&\leq \int_{-\infty}^{\infty}\left(\mathbf{1}\{t \geq 0\} - \frac{\mathrm{V}_n\left(\bigcup_i \left(A_i \ominus 2^{-t}\Delta\right)\right)}{\sum_i \mathrm{V}_n(A_i)}\right)dt\\
&= \int_{-\infty}^{\infty}\left(\mathbf{1}\{t \geq 0\} - \frac{\sum_i \mathrm{V}_n\left(A_i \ominus 2^{-t}\Delta\right)}{\sum_i \mathrm{V}_n(A_i)}\right)dt\\
&= \int_{-\infty}^{\infty}\left(\sum_i \frac{\mathrm{V}_n(A_i)}{\sum_j \mathrm{V}_n(A_j)}\left(\mathbf{1}\{t \geq 0\} - \frac{\mathrm{V}_n\left(A_i \ominus 2^{-t}\Delta\right)}{\mathrm{V}_n(A_i)}\right)\right)dt\\
&= \sum_i \frac{\mathrm{V}_n(A_i)}{\sum_j \mathrm{V}_n(A_j)} \cdot h_{\ominus\Delta}(A_i).
\end{aligned}$$

Equality holds if $\left(\bigcup_i A_i\right) \ominus 2^{-t}\Delta = \bigcup_i \left(A_i \ominus 2^{-t}\Delta\right)$, which is true when the closures of $A_1, \ldots, A_k$ are disjoint.

For the reduction to differential entropy property, let $X^n \sim \mathrm{Unif}(A)$, and $A \cap L$ is connected for all lines L parallel to the n-th axis, then

$$\begin{aligned}
&h_{\ominus\{0\}^{n-1}\times[0,1]}(A) \\
&= \int_{-\infty}^{\infty}\left(\mathbf{1}\{t \geq 0\} - \frac{\mathrm{V}_n\left(A \ominus \left(\{0\}^{n-1}\times[0,2^{-t}]\right)\right)}{\mathrm{V}_n(A)}\right)dt \\
&= \int_{-\infty}^{\infty}\Bigg(\mathbf{1}\{t \geq 0\} \\
&\quad - \int_{\mathbb{R}^{n-1}} \frac{\mathrm{V}_1\left(\left(A \ominus \{0\}^{n-1}\times[0,2^{-t}]\right)\cap\left(\{x_1^{n-1}\}\times\mathbb{R}\right)\right)}{\mathrm{V}_n(A)} \\
&\qquad dx_1^{n-1}\Bigg)dt \\
&= \int_{-\infty}^{\infty}\Bigg(\mathbf{1}\{t \geq 0\} \\
&\quad - \int_{\mathbb{R}^{n-1}} \max\left\{f_{X_1^{n-1}}(x_1^{n-1}) - 2^{-t}/\mathrm{V}_n(A),\, 0\right\}dx_1^{n-1}\Bigg)dt \\
&= \int_{\mathbb{R}^{n-1}} f_{X_1^{n-1}}(x_1^{n-1})\int_{-\infty}^{\infty}\Big(\mathbf{1}\{t \geq 0\} \\
&\quad - \max\left\{1 - 2^{-t}/(\mathrm{V}_n(A) f_{X_1^{n-1}}(x_1^{n-1})),\, 0\right\}\Big)dt\,dx_1^{n-1} \\
&= \int_{\mathbb{R}^{n-1}} f_{X_1^{n-1}}(x_1^{n-1})\int_{-\infty}^{\infty}\Bigg(\mathbf{1}\{t \geq 0\} \\
&\quad - \max\left\{1 - 2^{-(t+\log \mathrm{V}_n(A) + \log f_{X_1^{n-1}}(x_1^{n-1}))}, 0\right\}\Bigg)dt\,dx_1^{n-1} \\
&= \int_{\mathbb{R}^{n-1}} f(x_1^{n-1})\int_{-\infty}^{\infty}\Big(\mathbf{1}\left\{t \geq \log \mathrm{V}_n(A) + \log f(x_1^{n-1})\right\} \\
&\quad - \max\left\{1 - 2^{-t},\, 0\right\}\Big)dt\,dx_1^{n-1} \\
&= \int_{\mathbb{R}^{n-1}} f(x_1^{n-1})\Big(-\log \mathrm{V}_n(A) - \log f(x_1^{n-1}) \\
&\quad + \int_{-\infty}^{\infty}\left(\mathbf{1}\{t \geq 0\} - \max\left\{1-2^{-t},\, 0\right\}\right)dt\Big)dx_1^{n-1} \\
&= \int_{\mathbb{R}^{n-1}} f(x_1^{n-1})\Big(-\log \mathrm{V}_n(A) - \log f(x_1^{n-1}) \\
&\quad + \int_{0}^{\infty} 2^{-t}dt\Big)dx_1^{n-1} \\
&= h(X_1^{n-1}) - \log \mathrm{V}_n(A) + \log e.
\end{aligned}$$

D. Proof of Proposition 2

The entropy of W_{D} can be expressed as

$$\begin{aligned}
&H(W_{\mathrm{D}}) \\
&\quad = -\sum_{k\in\mathbb{Z}}\sum_{v\in D_k(A)} \mathsf{P}\{W_{\mathrm{D}} = (k,v)\}\log \mathsf{P}\{W_{\mathrm{D}} = (k,v)\} \\
&\quad = -\sum_{k\in\mathbb{Z}}\sum_{v\in D_k(A)} \frac{2^{-nk}}{\mathrm{V}_n(A)}\log\frac{2^{-nk}}{\mathrm{V}_n(A)} \\
&\quad = \sum_{k\in\mathbb{Z}} \frac{2^{-nk}\,|D_k(A)|}{\mathrm{V}_n(A)}\left(nk + \log \mathrm{V}_n(A)\right) \\
&\quad = \log \mathrm{V}_n(A) + \frac{1}{\mathrm{V}_n(A)}\sum_{k\in\mathbb{Z}} nk2^{-nk}\,|D_k(A)|.
\end{aligned}$$

Note that

$$\begin{aligned}
\sum_{l=-\infty}^{k} 2^{-nl}\,|D_l(A)| &= 2^{-nk}\left|\left\{v \in \mathbb{Z}^n : C_{k,v} \subseteq A\right\}\right| \\
&\geq 2^{-nk}\left|\left\{v \in \mathbb{Z}^n : C_{k-1,(v-w)/2} \subseteq A\right\}\right|.
\end{aligned}$$

for any $w \in [0,1]^n$, since $C_{k,v} \subseteq C_{k-1,(v-w)/2}$. Note that the $(v-w)/2$ in the subscript may not have integer entries, but still the same definition $C_{k,v} = 2^{-k}([0,1]^n + v)$ can be applied. Also

$$\begin{aligned}
&\int_{[0,1]^n}\left|\left\{v \in \mathbb{Z}^n : C_{k-1,(v-w)/2} \subseteq A\right\}\right| dw \\
&\quad = \sum_{v\in\mathbb{Z}^n}\int_{[0,1]^n}\mathbf{1}\left\{C_{k-1,(v-w)/2} \subseteq A\right\}dw \\
&\quad = 2^n\int_{\mathbb{R}^n}\mathbf{1}\left\{C_{k-1,w} \subseteq A\right\}dw \\
&\quad = 2^n 2^{n(k-1)}\mathrm{V}_n\left(A \ominus [0,\, 2^{-(k-1)}]^n\right) \\
&\quad = 2^{nk}\mathrm{V}_n\left(A \ominus [0,\, 2^{-(k-1)}]^n\right).
\end{aligned}$$

Hence

$$\begin{aligned}
\sum_{l=-\infty}^{k} 2^{-nl}\,|D_l(A)| &\geq \mathrm{V}_n\left(A \ominus [0,\, 2^{-(k-1)}]^n\right), \\
\sum_{l=k+1}^{\infty} 2^{-nl}\,|D_l(A)| &\leq \mathrm{V}_n(A) - \mathrm{V}_n\left(A \ominus [0,\, 2^{-(k-1)}]^n\right).
\end{aligned}$$

Note that $H(W_{\mathrm{D}}) = H(W_{(1/2)A})$, and also the right-hand-side of the proposition remains the same when A is replaced by $(1/2)A$. Without loss of generality assume A is small enough such that $\mathrm{V}_n(A) \leq 1$, so $D_k(A) = \emptyset$ for $k < 0$.

$$\begin{aligned}
&H(W_{\mathrm{D}}) \\
&= \log \mathrm{V}_n(A) + \frac{1}{\mathrm{V}_n(A)}\sum_{k=0}^{\infty} nk2^{-nk}\,|D_k(A)| \\
&= \log \mathrm{V}_n(A) + \frac{n}{\mathrm{V}_n(A)}\sum_{k=0}^{\infty}\sum_{l=k+1}^{\infty} 2^{-nl}\,|D_l(A)| \\
&\leq \log \mathrm{V}_n(A) + \frac{n}{\mathrm{V}_n(A)}\sum_{k=0}^{\infty}\left(\mathrm{V}_n(A) - \mathrm{V}_n\left(A \ominus [0,\, 2^{-(k-1)}]^n\right)\right) \\
&\leq \log \mathrm{V}_n(A) + \frac{n}{\mathrm{V}_n(A)}\int_{-2}^{\infty}\left(\mathrm{V}_n(A) - \mathrm{V}_n\left(A \ominus [0,\, 2^{-t}]^n\right)\right)dt \\
&= \log \mathrm{V}_n(A) + n\cdot h_{\ominus[0,1]^n}(A) + 2n.
\end{aligned}$$

To prove the second result, consider

$$\begin{aligned}
&\sum_{l=-\infty}^{k} 2^{-nl}\,|D_l\left(\Lambda(A+U)\right)| \\
&\quad = 2^{-nk}\left|\left\{v \in \mathbb{Z}^n : C_{k,v} \subseteq \Lambda(A+U)\right\}\right| \\
&\quad = 2^{-nk}\left|\left\{v \in \mathbb{Z}^n : \Lambda^{-1}C_{k,v} - U \subseteq A\right\}\right|.
\end{aligned}$$

Assuming $k \geq -T$ and taking expectation over U, we obtain

$$\begin{aligned}
&\mathsf{E}\left[\sum_{l=-\infty}^{k} 2^{-nl}\left|D_l\left(\Lambda A+U\right)\right| \,\middle|\, \Lambda\right] \\
&= 2^{-nT}\int_{[0,2^T]^n} 2^{-nk}\left|\left\{v\in\mathbb{Z}^n : C_{k,v}-u\subseteq \Lambda A\right\}\right| du \\
&= 2^{-nT}\int_{[0,2^T]^n} 2^{-nk}\left|\left\{v : 2^{-k}\left([0,1]^n+v-2^k u\right)\subseteq \Lambda A\right\}\right| du \\
&= \int_{[0,1]^n} 2^{-nk}\left|\left\{v : 2^{-k}\left([0,1]^n+v-2^{T+k} u\right)\subseteq \Lambda A\right\}\right| du \\
&\overset{(a)}{=} \int_{\mathbb{R}^n} 2^{-nk}\mathbf{1}\left\{2^{-k}\left([0,1]^n+u\right)\subseteq \Lambda A\right\} du \\
&= \int_{\mathbb{R}^n} \mathbf{1}\left\{2^{-k}[0,1]^n+u\subseteq \Lambda A\right\} du \\
&= \mathrm{V}_n\left(\Lambda A\ominus 2^{-k}[0,1]^n\right) \\
&= \Lambda^n \mathrm{V}_n\left(A\ominus \Lambda^{-1}2^{-k}[0,1]^n\right),
\end{aligned}$$

where (a) follows since 2^{T+k} is a non-negative integer. Since $2^{nT} > \mathrm{V}_n(2A)$, $D_k(\Lambda A+U)=\emptyset$ for $k<-T$. Hence, we have

$$\begin{aligned}
&H(\tilde{W}_{\mathrm{D}}) \\
&= \log \mathrm{V}_n(\Lambda A) + \frac{1}{\mathrm{V}_n(\Lambda A)}\sum_{k=-T}^{\infty} nk2^{-nk}\left|D_k(\Lambda A+U)\right| \\
&= \log \mathrm{V}_n(A) + n\log\Lambda + \frac{n}{\Lambda^n \mathrm{V}_n(A)} \\
&\quad \cdot \sum_{k=-T}^{\infty}\left(\mathbf{1}\{k\geq 0\}\,\Lambda^n\mathrm{V}_n(A) - \sum_{l=-\infty}^{k} 2^{-nl}\left|D_l(\Lambda A+U)\right|\right).
\end{aligned}$$

Taking expectation over U, we obtain

$$\begin{aligned}
\mathsf{E}\left[H(\tilde{W}_{\mathrm{D}})\,\middle|\,\Lambda\right] &= \log \mathrm{V}_n(A) + n\log\Lambda + \frac{n}{\Lambda^n \mathrm{V}_n(A)} \\
&\quad \cdot \sum_{k=-T}^{\infty}\Big(\mathbf{1}\{k\geq 0\}\,\Lambda^n\mathrm{V}_n(A) \\
&\quad -\Lambda^n\mathrm{V}_n\left(A\ominus\Lambda^{-1}2^{-k}[0,1]^n\right)\Big) \\
&= \log \mathrm{V}_n(A) + n\log\Lambda + n\sum_{k=-T}^{\infty}\Big(\mathbf{1}\{k\geq 0\} \\
&\quad - \frac{\mathrm{V}_n\left(A\ominus\Lambda^{-1}2^{-k}[0,1]^n\right)}{\mathrm{V}_n(A)}\Big).
\end{aligned}$$

Taking expectation over Λ, we have

$$\begin{aligned}
&\mathsf{E}\left[H(\tilde{W}_{\mathrm{D}})\right] \\
&= \log \mathrm{V}_n(A) + \mathsf{E}\left[n\log\Lambda\right] \\
&\quad + n\,\mathsf{E}\left[\sum_{k=-T}^{\infty}\left(\mathbf{1}\{k\geq 0\} - \frac{\mathrm{V}_n\left(A\ominus 2^{-(k+\Theta)}[0,1]^n\right)}{\mathrm{V}_n(A)}\right)\right] \\
&= \log \mathrm{V}_n(A) + \frac{n}{2} \\
&\quad + n\left(\int_{-T}^{\infty}\left(\mathbf{1}\{\theta\geq 0\} - \frac{\mathrm{V}_n\left(A\ominus 2^{-\theta}[0,1]^n\right)}{\mathrm{V}_n(A)}\right)d\theta - \frac{1}{2}\right) \\
&= \log \mathrm{V}_n(A) + n\int_{-T}^{\infty}\left(\mathbf{1}\{\theta\geq 0\} - \frac{\mathrm{V}_n\left(A\ominus 2^{-\theta}[0,1]^n\right)}{\mathrm{V}_n(A)}\right)d\theta \\
&\overset{(a)}{=} \log \mathrm{V}_n(A) + n\int_{-\infty}^{\infty}\left(\mathbf{1}\{\theta\geq 0\} - \frac{\mathrm{V}_n\left(A\ominus 2^{-\theta}[0,1]^n\right)}{\mathrm{V}_n(A)}\right)d\theta \\
&= \log \mathrm{V}_n(A) + n\cdot h_{\ominus[0,1]^n}(A).
\end{aligned}$$

where (a) follows since $2^{nT} > \mathrm{V}_n(2A)$, hence $A\ominus 2^{-\theta}[0,1]^n=\emptyset$ for $\theta\leq -T$.

E. Proof of Lemma 1

We first prove the following claim: For any orthogonally convex set $A\subseteq\mathbb{R}^n$ with $0<\mathrm{V}_n(A)<\infty$ and $\gamma\geq 0$, the set $A\ominus\left([0,\gamma]\times\{0\}^{n-1}\right)\subseteq A$ is orthogonally convex, and

$$\begin{aligned}
&\mathrm{V}_n\left(A\ominus\left([0,\gamma]\times\{0\}^{n-1}\right)\right) \\
&\quad = \mathrm{V}_n(A) - \int_{\mathrm{P}_{[2:n]}(A)} \min\left\{\gamma, \mathrm{V}_1\left(A\cap(\mathrm{span}(\mathrm{e}_1)+x)\right)\right\} dx_2^n,
\end{aligned} \tag{7}$$

where $\mathrm{span}(\mathrm{e}_1)+x=\{(\alpha,x_2,x_3,\ldots,x_n) : \alpha\in\mathbb{R}\}\subseteq\mathbb{R}^n$.

By the orthogonal convexity of A,

$$\begin{aligned}
A\ominus\left([0,\gamma]\times\{0\}^{n-1}\right) &= \{x : x+\alpha\mathrm{e}_1\in A \quad \text{for all } \alpha\in[0,\gamma]\} \\
&= \{x : x\in A,\ x+\gamma\,\mathrm{e}_1\in A\} \\
&= A\cap(A-\gamma\,\mathrm{e}_1)
\end{aligned}$$

is the intersection of two orthogonally convex sets, and is therefore orthogonally convex. Also

$$\begin{aligned}
&\mathrm{V}_n(A) - \mathrm{V}_n\left(A\ominus\left([0,\gamma]\times\{0\}^{n-1}\right)\right) \\
&\quad = \mathrm{V}_n\left\{x\in A : x+\gamma\,\mathrm{e}_1\notin A\right\} \\
&\quad = \int_{\mathrm{P}_{[2:n]}(A)} \mathrm{V}_1\Big\{x_1\in\mathbb{R} : (x_1,\tilde{x}_2,\ldots,\tilde{x}_n)\in A, \\
&\qquad (x_1+\gamma,\ \tilde{x}_2,\ldots,\tilde{x}_n)\notin A\Big\} d\tilde{x}_2^n \\
&\quad = \int_{\mathrm{P}_{[2:n]}(A)} \mathrm{V}_1\Big\{x\in A\cap(\mathrm{span}(\mathrm{e}_1)+\tilde{x}) : \\
&\qquad x+\gamma\,\mathrm{e}_1\notin A\cap(\mathrm{span}(\mathrm{e}_1)+\tilde{x})\Big\} d\tilde{x}_2^n \\
&\quad = \int_{\mathrm{P}_{[2:n]}(A)} \min\left\{\gamma\ ,\ \mathrm{V}_1\left(A\cap(\mathrm{span}(\mathrm{e}_1)+\tilde{x})\right)\right\} d\tilde{x}_2^n,
\end{aligned}$$

where the last equality follows since $A\cap(\mathrm{span}(\mathrm{e}_1)+\tilde{x})$ is connected. Hence the claim is proved.

Now we proceed to the second coordinate. By invoking (7) twice, we obtain

$$\begin{aligned}
&\mathrm{V}_n(A) - \mathrm{V}_n\left(A\ominus\left([0,\gamma]^2\times\{0\}^{n-2}\right)\right) \\
&\quad = \mathrm{V}_n(A) - \mathrm{V}_n\left(A\ominus\left([0,\gamma]\times\{0\}^{n-1}\right)\right) \\
&\qquad + \mathrm{V}_n\left(A\ominus\left([0,\gamma]\times\{0\}^{n-1}\right)\right) \\
&\qquad - \mathrm{V}_n\left(A\ominus\left([0,\gamma]^2\times\{0\}^{n-2}\right)\right)
\end{aligned}$$

$$= \int_{\mathrm{P}_{\setminus 1}(A)} \min\{\gamma\ ,\ \mathrm{V}_1(A \cap (\mathrm{span}(\mathrm{e}_1)+x))\}\,dx$$
$$+ \int_{\mathrm{P}_{\setminus 2}(A \ominus ([0,\gamma]\times\{0\}^{n-1}))} \min\Big\{\gamma,$$
$$\mathrm{V}_1\left(A \ominus \left([0,\gamma]\times\{0\}^{n-1}\right) \cap (\mathrm{span}(\mathrm{e}_2)+x)\right)\Big\}\,dx$$
$$\le \int_{\mathrm{P}_{\setminus 1}(A)} \min\{\gamma\ ,\ \mathrm{V}_1(A \cap (\mathrm{span}(\mathrm{e}_1)+x))\}\,dx$$
$$+ \int_{\mathrm{P}_{\setminus 2}(A)} \min\{\gamma\ ,\ \mathrm{V}_1(A \cap (\mathrm{span}(\mathrm{e}_2)+x))\}\,dx,$$

where the last inequality is due to $A \ominus \left([0,\gamma]\times\{0\}^{n-1}\right) \subseteq A$. By repeating this argument for each axis, we obtain the second bound.

F. Proof of Monotonicity of $\tilde{h}_\zeta(X^n)$

Let $\zeta_1 \le \zeta_2$ and ξ_1, ξ_2 correspond to ζ_1, ζ_2, then $\zeta_1^{-1}\xi_1 \le \zeta_2^{-1}\xi_2$. We have $\zeta_1^{-1}\min\{\xi_1,\ f(x)\} \le \zeta_2^{-1}\min\{\xi_2,\ f(x)\}$ if and only if $f(x) \ge \xi_1\zeta_1^{-1}\zeta_2$. Hence

$$\int_{\mathbb{R}^n} -\zeta_2^{-1}\min\{\xi_2,\ f(x)\}\log\left(\zeta_2^{-1}\min\{\xi_2,\ f(x)\}\right)dx$$
$$= \int_{\mathbb{R}^n} -\zeta_2^{-1}\min\{\xi_2,\ f(x)\}$$
$$\log\left(\zeta_1\xi_1^{-1}e^{-1}\zeta_2^{-1}\min\{\xi_2,\ f(x)\}\right)dx + \log\left(\zeta_1\xi_1^{-1}e^{-1}\right)$$
$$\le \int_{\mathbb{R}^n} -\zeta_1^{-1}\min\{\xi_1,\ f(x)\}$$
$$\log\left(\zeta_1\xi_1^{-1}e^{-1}\zeta_1^{-1}\min\{\xi_1,\ f(x)\}\right)dx + \log\left(\zeta_1\xi_1^{-1}e^{-1}\right)$$
$$= \int_{\mathbb{R}^n} -\zeta_1^{-1}\min\{\xi_1,\ f(x)\}\log\left(\zeta_1^{-1}\min\{\xi_1,\ f(x)\}\right)dx,$$

where the inequality is because $t \mapsto -t\log(\zeta_1\xi_1^{-1}e^{-1}t)$ is increasing when $t \le \zeta_1^{-1}\xi_1$, and decreasing when $t \ge \zeta_1^{-1}\xi_1$.

G. Proof of Theorem 2

We first prove the following claim on $H(W_{\mathrm{D}})$ involving truncated differential entropy

$$H(W_{\mathrm{D}}) \le n\left(H(\zeta_1,\ldots,\zeta_n) + \sum_{i=1}^n \zeta_i \tilde{h}_{\zeta_i}(X_{[1:n]\setminus i})\right)$$
$$- (n-1)\log \mathrm{V}_n(A) + (2+\log e)n,$$

where

$$\zeta_i = \int_{\mathbb{R}^{n-1}} \min\left\{f_{X_{[1:n]\setminus i}}(x_{[1:n]\setminus i})\ ,\ \xi\right\}dx_{[1:n]\setminus i}$$

for a suitable $\xi > 0$ such that $\sum \zeta_i = 1$. By Proposition 2, the claim can be proved by bounding $h_{\ominus[0,1]^n}(A)$. Note that by Lemma 1,

$$h_{\ominus[0,1]^n}(A)$$
$$= \int_{-\infty}^{\infty}\left(\mathbf{1}\{t\ge 0\} - \frac{\mathrm{V}_n\left(A\ominus[0,\ 2^{-t}]^n\right)}{\mathrm{V}_n(A)}\right)dt$$
$$\le \int_{-\infty}^{\infty}\Bigg(\mathbf{1}\{t\ge 0\} - \frac{1}{\mathrm{V}_n(A)}\max\Bigg(0, \mathrm{V}_n(A) - \sum_{i=1}^n \int_{\mathrm{P}_{\setminus i}(A)}$$
$$\min\left\{2^{-t},\ \mathrm{V}_1(A\cap(\mathrm{span}(\mathrm{e}_i)+x))\right\}dx_{[1:n]\setminus i}\Bigg)\Bigg)dt$$
$$= \int_{-\infty}^{\infty}\Bigg(\mathbf{1}\{t\ge 0\} - \max\Bigg(0,\ 1 - \sum_{i=1}^n \int_{\mathrm{P}_{\setminus i}(A)}$$
$$\min\left\{\frac{2^{-t}}{\mathrm{V}_n(A)}\ ,\ f_{X_{[1:n]\setminus i}}(x_{[1:n]\setminus i})\right\}dx_{[1:n]\setminus i}\Bigg)\Bigg)dt$$
$$\overset{(a)}{=} \int_{-\infty}^{\infty}\Bigg(-\mathbf{1}\{t<0\} + \sum_{i=1}^n \int_{\mathrm{P}_{\setminus i}(A)} \min\Bigg\{\frac{2^{-t}}{\mathrm{V}_n(A)},\ \xi,$$
$$f_{X_{[1:n]\setminus i}}(x_{[1:n]\setminus i})\Bigg\}dx_{[1:n]\setminus i}\Bigg)dt$$
$$= \sum_{i=1}^n \int_{-\infty}^{\infty}\Bigg(-\mathbf{1}\{t<0\}\cdot\zeta_i + \int_{\mathrm{P}_{\setminus i}(A)} \min\Bigg\{\frac{2^{-t}}{\mathrm{V}_n(A)},\ \xi,$$
$$f_{X_{[1:n]\setminus i}}(x_{[1:n]\setminus i})\Bigg\}dx_{[1:n]\setminus i}\Bigg)dt$$
$$= \sum_{i=1}^n \int_{\mathrm{P}_{\setminus i}(A)}\int_{-\infty}^{\infty}\Bigg(-\mathbf{1}\{t<0\}\cdot\min\left\{\xi,\ f_{X_{[1:n]\setminus i}}(x_{[1:n]\setminus i})\right\}$$
$$+ \min\left\{\frac{2^{-t}}{\mathrm{V}_n(A)},\ \xi,\ f_{X_{[1:n]\setminus i}}(x_{[1:n]\setminus i})\right\}\Bigg)dt\,dx_{[1:n]\setminus i}$$
$$= \sum_{i=1}^n \int_{\mathrm{P}_{\setminus i}(A)} -\min\left\{\xi,\ f_{X_{[1:n]\setminus i}}(x_{[1:n]\setminus i})\right\}$$
$$\cdot\left(\log\left(\mathrm{V}_n(A)\min\left\{\xi,\ f_{X_{[1:n]\setminus i}}(x_{[1:n]\setminus i})\right\}\right) - \log e\right)dx_{[1:n]\setminus i}$$
$$= -\log \mathrm{V}_n(A) + H(\zeta_1,\ldots,\zeta_n)$$
$$+ \sum_{i=1}^n \zeta_i \int_{\mathrm{P}_{\setminus i}(A)} -\zeta_i^{-1}\min\left\{\xi,\ f_{X_{[1:n]\setminus i}}(x_{[1:n]\setminus i})\right\}$$
$$\cdot\log\left(\zeta_i^{-1}\min\left\{\xi,\ f_{X_{[1:n]\setminus i}}(x_{[1:n]\setminus i})\right\}\right)dx_{[1:n]\setminus i} + \log e$$
$$= -\log \mathrm{V}_n(A) + H(\zeta_1,\ldots,\zeta_n) + \sum_{i=1}^n \zeta_i \tilde{h}_{\zeta_i}(X_{[1:n]\setminus i}) + \log e,$$

where (a) follows by the definition of ξ. The claim follows.

We proceed to prove Theorem 2. Let $D = \mathrm{diag}(d_1,\ldots,d_n)$. Assuming $\prod_i d_i = 1$, then by the claim,

$$H(W_{\mathrm{DS}})$$
$$\le n\left(H(\zeta_1,\ldots,\zeta_n) + \sum_{i=1}^n \zeta_i\left(\tilde{h}_{\zeta_i}(X_{[1:n]\setminus i}) + \sum_{j\ne i}\log d_j\right)\right)$$
$$-, (n-1)\log \mathrm{V}_n(A) + (2+\log e)n$$
$$\le n\sum_{i=1}^n \zeta_i\left(\tilde{h}_{\zeta_i}(X_{[1:n]\setminus i}) - \log d_i\right)$$
$$- (n-1)\log \mathrm{V}_n(A) + n\log n + (2+\log e)n,$$

where

$$\zeta_i = \Big(\prod_{j\ne i} d_j\Big)$$
$$\cdot\int_{\mathbb{R}^{n-1}} \min\left\{\Big(\prod_{j\ne i} d_j\Big)^{-1} f_{X_{[1:n]\setminus i}}(x_{[1:n]\setminus i}), \xi\right\}dx_{[1:n]\setminus i}$$
$$= \int_{\mathbb{R}^{n-1}} \min\left\{f_{X_{[1:n]\setminus i}}(x_{[1:n]\setminus i})\ ,\ \xi d_i^{-1}\right\}dx_{[1:n]\setminus i}$$

for a suitable $\xi > 0$ such that $\sum \zeta_i = 1$. Let $\alpha_1, \ldots, \alpha_n > 0$ such that

$$\int_{\mathbb{R}^{n-1}} \min\left\{f_{X_{[1:n]\setminus i}}(x_{[1:n]\setminus i}),\ \alpha_i\right\} dx_{[1:n]\setminus i} = \frac{1}{n}.$$

Set $\xi = \left(\prod_j a_j\right)^{1/n}$, $d_i = \alpha_i^{-1}\xi$, then we have $\zeta_i = 1/n$,

$$H(W_{\mathrm{DS}}) \le \sum_{i=1}^{n} \tilde{h}_{1/n}(X_{[1:n]\setminus i}) - (n-1)\log \mathrm{V}_n(A) + n\log n + (2+\log e)n.$$

H. Proof of Lemma 3

Before proving this lemma, we first prove the following claim on the volume of a convex set. For any convex set $A \subseteq \mathbb{R}^n$ where $0 \in A$ and $1 \le m \le n$, let $\tilde{A} = \left\{x_{m+1}^n : (0^m,\ x_{m+1}^n) \in A\right\}$, then

$$\mathrm{V}_n(A) \ge \binom{n}{m}^{-1} \cdot \mathrm{V}_{n-m}(\tilde{A}) \cdot \mathrm{VP}_{[1:m]}(A).$$

Now we prove the claim. Denote the section of A as

$$S_A(\tilde{x}_{m+1}^n) = \left\{x^m : (x^m,\ \tilde{x}_{m+1}^n) \in A\right\} \subseteq \mathbb{R}^m.$$

Note that

$$\begin{aligned}
&\mathrm{V}_n(A)\\
&= \int_{S^{m-1}} \int_0^\infty \left(\int_{\{\tilde{x}_{m+1}^n :\, (rx^m,\ \tilde{x}_{m+1}^n)\in A\}} d\tilde{x}_{m+1}^n\right) r^{m-1} dr dx^m\\
&= \int_{S^{m-1}} \int_{\{(r,\ \tilde{x}_{m+1}^n):\, r\ge 0,\ (rx^m,\ \tilde{x}_{m+1}^n)\in A\}} r^{m-1} d(r,\ \tilde{x}_{m+1}^n) dx^m.
\end{aligned}$$

Consider the set

$$S_{\mathrm{rad},A}(x^m) = \left\{(r,\ \tilde{x}_{m+1}^n) : r \ge 0,\ (rx^m,\ \tilde{x}_{m+1}^n) \in A\right\}.$$

It is the intersection of A and a half-space, and hence it is convex. By definition of radial function and projection, there exists $\hat{x}_{m+1}^n(x^n)$ such that

$$(\rho_{\mathrm{P}_{[1:m]}(A)}(x^m),\ \hat{x}_{m+1}^n(x^n)) \in S_{\mathrm{rad},A}(x^m).$$

Also by definition of $\tilde{A}$,

$$\{0\} \times \tilde{A} \subseteq S_{\mathrm{rad},A}(x^m).$$

Hence the convex hull of $\left\{(\rho_{\mathrm{P}_{[1:m]}(A)}(x^m),\ \hat{x}_{m+1}^n(x^n))\right\} \cup \left(\{0\}\times\tilde{A}\right)$ is a subset of $S_{\mathrm{rad},A}(x^m)$. The convex hull can be expressed as

$$\begin{aligned}
\Big\{(r,\ \tilde{x}_{m+1}^n) :\ & 0 \le r \le \rho_{\mathrm{P}_{[1:m]}(A)}(x^m),\\
& \tilde{x}_{m+1}^n \in \left(1 - r\rho_{\mathrm{P}_{[1:m]}(A)}^{-1}(x^m)\right)\tilde{A}\\
& + r\rho_{\mathrm{P}_{[1:m]}(A)}^{-1}(x^m)\cdot \hat{x}_{m+1}^n(x^n)\Big\}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\mathrm{V}_n(A)\\
&= \int_{S^{m-1}} \int_{S_{\mathrm{rad},A}(x^m)} r^{m-1} d(r,\ \tilde{x}_{m+1}^n) dx^m\\
&\ge \int_{S^{m-1}} \int_0^{\rho_{\mathrm{P}_{[1:m]}(A)}(x^m)} \int_{\left(1-r\rho_{\mathrm{P}_{[1:m]}(A)}^{-1}(x^m)\right)\tilde{A} + r\rho_{\mathrm{P}_{[1:m]}(A)}^{-1}(x^m)\cdot\hat{x}_{m+1}^n(x^n)}\\
&\qquad d\tilde{x}_{m+1}^n \cdot r^{m-1} dr dx^m\\
&= \int_{S^{m-1}} \int_0^{\rho_{\mathrm{P}_{[1:m]}(A)}(x^m)} \mathrm{V}_{n-m}(\tilde{A})\\
&\qquad \cdot \left(1 - r\rho_{\mathrm{P}_{[1:m]}(A)}^{-1}(x^m)\right)^{n-m} r^{m-1} dr dx^m\\
&= \int_{S^{m-1}} \rho_{\mathrm{P}_{[1:m]}(A)}^m(x^m) \mathrm{V}_{n-m}(\tilde{A}) \cdot \mathrm{B}(m,\ n-m+1) dx^m\\
&= m \cdot \mathrm{B}(m,\ n-m+1) \cdot \mathrm{V}_{n-m}(\tilde{A}) \cdot \mathrm{VP}_{[1:m]}(A),
\end{aligned}$$

where

$$\mathrm{B}(a,b) = \int_0^1 t^{a-1}(1-t)^{b-1} dt = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

is the beta function. The claim follows.

We proceed to prove Lemma 3. To prove the lower bound, consider

$$\begin{aligned}
&h(X_{m+1}^n \mid X^m)\\
&\quad= \int_{\mathbb{R}^m} f_{X^m}(x^m) h(X_{m+1}^n \mid X^m = x^m) dx^m\\
&\quad\ge \int_{\mathbb{R}^m} f_{X^m}(x^m) \cdot -\log \sup_{\tilde{x}_{m+1}^n \in \mathbb{R}^{n-m}} f(\tilde{x}_{m+1}^n \mid x^m) dx^m\\
&\quad\ge -\log \int_{\mathbb{R}^m} f_{X^m}(x^m) \sup_{\tilde{x}_{m+1}^n \in \mathbb{R}^{n-m}} f(\tilde{x}_{m+1}^n \mid x^m) dx^m\\
&\quad= -\log \int_{\mathbb{R}^m} \sup_{\tilde{x}_{m+1}^n \in \mathbb{R}^{n-m}} f(x^m,\ \tilde{x}_{m+1}^n) dx^m.
\end{aligned}$$

Now we prove the upper bound. By Lemma 2,

$$h(X^n) \le -\log\Big(\sup_{x^n \in \mathbb{R}^n} f(x^n)\Big) + n\log e, \tag{8}$$

$$h(X^m) \ge -\log\Big(\sup_{x^m \in \mathbb{R}^m} f_{X^m}(x^m)\Big). \tag{9}$$

Without loss of generality, assume that $\sup_{x^m\in\mathbb{R}^m} f_{X^m}(x^m)$ is attained at $x^m = 0$ and $\sup_{x_{m+1}^n} f(0^m, x_{m+1}^n)$ is attained at $x_{m+1}^n = 0$. Denote the super level set

$$L_z^+(f) = \left\{x^n : f(x^n) \ge z\right\}.$$

Since f is log-concave, $L_z^+(f)$ is convex. Define

$$\tilde{L}_z^+(f) = \left\{x_{m+1}^n : (0^m,\ x_{m+1}^n) \in L_z^+(f)\right\}.$$

By the claim we proved earlier,

$$\begin{aligned}
&\int_{\mathbb{R}^n} f(x^n)\\
&= \int_0^\infty \mathrm{V}_n\left(L_z^+(f)\right) dz\\
&\ge \int_{\{z:\, 0\in L_z^+(f)\}} \mathrm{V}_n\left(L_z^+(f)\right) dz\\
&\ge \int_{\{z:\, 0\in L_z^+(f)\}} \binom{n}{m}^{-1} \cdot \mathrm{V}_{n-m}(\tilde{L}_z^+(f)) \cdot \mathrm{VP}_{[1:m]}(L_z^+(f)) dz
\end{aligned}$$

$$
\begin{aligned}
&= \binom{n}{m}^{-1} \cdot \int_0^{f(0)} \mathrm{V}_{n-m}(\tilde{L}_z^+(f)) \cdot \mathrm{VP}_{[1:m]}(L_z^+(f)) dz \\
&\overset{(a)}{\geq} \binom{n}{m}^{-1} \cdot \left(\int_0^{f(0)} \mathrm{V}_{n-m}(\tilde{L}_z^+(f)) dz \right) \\
&\quad \cdot \left(\frac{1}{f(0)} \int_0^{f(0)} \mathrm{VP}_{[1:m]}(L_z^+(f)) dz \right) \\
&\overset{(b)}{\geq} \binom{n}{m}^{-1} f_{X^m}(0) \left(\frac{1}{\sup_x f(x)} \int_0^{\sup_x f(x)} \mathrm{VP}_{[1:m]}(L_z^+(f)) dz \right) \\
&= \binom{n}{m}^{-1} f_{X^m}(0) \frac{1}{\sup_x f(x)} \int_{\mathbb{R}^m} \sup_{\tilde{x}_{m+1}^n \in \mathbb{R}^{n-m}} f(x^m, \tilde{x}_{m+1}^n) dx^m \\
&\overset{(c)}{\geq} \binom{n}{m}^{-1} 2^{-h(X^m)} 2^{h(X^n)} e^{-n} \int_{\mathbb{R}^m} \sup_{\tilde{x}_{m+1}^n \in \mathbb{R}^{n-m}} f(x^m, \tilde{x}_{m+1}^n) dx^m.
\end{aligned}
$$

where (a) is due to Chebyshev's sum inequality, since both $\mathrm{V}_{n-m}(\tilde{L}_z^+(f))$ and $\mathrm{VP}_{[1:m]}(L_z^+(f))$ are non-increasing in z, (b) is due to $\int_0^{f(0)} \mathrm{V}_{n-m}(\tilde{L}_z^+(f)) dz = \int_{\mathbb{R}^{n-m}} f(0^m, \tilde{x}_{m+1}^n) d\tilde{x}_{m+1}^n = f_{X^m}(0)$ since $\sup_{\tilde{x}_{m+1}^n} f(0^m, \tilde{x}_{m+1}^n) = f(0)$, and $\mathrm{VP}_{[1:m]}(L_z^+(f))\, dz$ is non-increasing in z, and (c) is due to (8) and (9). The result follows from $\int_{\mathbb{R}^n} f(x^n) = 1$.

I. Proof of Lemma 4

As in the definition of $\tilde{h}_\zeta$, let $\xi > 0$ such that

$$\int_{\mathbb{R}^n} \min\{\xi,\ f(x)\}\, dx = \zeta.$$

Without loss of generality, assume $\sup_{x \in \mathbb{R}^n} f(x) = f(0)$ and let $\alpha = f(0)$. Let

$$A = \left\{ x \in \mathbb{R}^n :\ f(x) \geq \xi \right\}.$$

By log-concavity of f, we know A is convex, and we have

$$\mathrm{V}_n(A) = \frac{1}{n} \cdot \int_{S^{n-1}} \rho_A^n(x) dx,$$

where $S^{n-1} = \{x \in \mathbb{R}^n : \|x\|_2 = 1\}$ is the unit sphere. For $x \in A$, by definition of ρ_A, we know $x/(\rho_A^{-1}(x) + \epsilon) \in A$ for any $\epsilon > 0$,

$$
\begin{aligned}
f(x) &= f\left(\left(1 - \rho_A^{-1}(x) - \epsilon\right) \cdot 0 + \left(\rho_A^{-1}(x) + \epsilon\right) \cdot \frac{x}{\rho_A^{-1}(x) + \epsilon} \right) \\
&\geq (f(0))^{1-\rho_A^{-1}(x)-\epsilon} \cdot \left(f\left(\frac{x}{\rho_A^{-1}(x) + \epsilon} \right) \right)^{\rho_A^{-1}(x)+\epsilon} \\
&\geq \alpha^{1-\rho_A^{-1}(x)-\epsilon} \xi^{\rho_A^{-1}(x)+\epsilon}.
\end{aligned}
$$

Therefore

$$f(x) \geq \alpha^{1-\rho_A^{-1}(x)} \xi^{\rho_A^{-1}(x)}.$$

Hence

$$
\begin{aligned}
&\int_A f(x) dx \\
&= \int_{S^{n-1}} \int_0^{\rho_A(x)} f(rx) \cdot r^{n-1} dr dx \\
&\geq \int_{S^{n-1}} \int_0^{\rho_A(x)} \alpha^{1-\rho_A^{-1}(rx)} \xi^{\rho_A^{-1}(rx)} r^{n-1} dr dx \\
&= \int_{S^{n-1}} \int_0^{\rho_A(x)} \alpha^{1-r\rho_A^{-1}(x)} \xi^{r\rho_A^{-1}(x)} r^{n-1} dr dx \\
&= \int_{S^{n-1}} \rho_A^n(x) \int_0^1 \alpha^{1-r} \xi^r r^{n-1} dr dx \\
&= \int_{S^{n-1}} \rho_A^n(x) \alpha (-\log(\xi/\alpha))^{-n} \left(\Gamma(n) - \Gamma(n, -\log(\xi/\alpha))\right) dx \\
&= n\alpha \left(-\log(\xi/\alpha)\right)^{-n} \left(\Gamma(n) - \Gamma(n,\ -\log(\xi/\alpha))\right) \mathrm{V}_n(A),
\end{aligned}
$$

where $\Gamma(n, z) = \int_z^\infty t^{n-1} e^{-t} dt$ is the incomplete gamma function, and $\Gamma(n) = \Gamma(n, 0)$ is the gamma function.

On the other hand, for $x \notin A$, then for any $\epsilon > 0$, we have $x/(\rho_A^{-1}(x) - \epsilon) \notin A$,

$$
\begin{aligned}
\xi &\geq f\left(\frac{x}{\rho_A^{-1}(x) - \epsilon} \right) \\
&= f\left(\left(1 - \frac{1}{\rho_A^{-1}(x) - \epsilon}\right) \cdot 0 + \frac{1}{\rho_A^{-1}(x) - \epsilon} \cdot x \right) \\
&\geq \alpha^{1-1/(\rho_A^{-1}(x)-\epsilon)} \cdot (f(x))^{1/(\rho_A^{-1}(x)-\epsilon)},
\end{aligned}
$$

Therefore

$$f(x) \leq \alpha^{1-\rho_A^{-1}(x)} \xi^{\rho_A^{-1}(x)}.$$

Hence

$$
\begin{aligned}
&\int_{\mathbb{R}^n \setminus A} f(x) dx \\
&= \int_{S^{n-1}} \int_{\rho_A(x)}^\infty f(rx) \cdot r^{n-1} dr dx \\
&\leq \int_{S^{n-1}} \int_{\rho_A(x)}^\infty \alpha^{1-\rho_A^{-1}(rx)} \xi^{\rho_A^{-1}(rx)} r^{n-1} dr dx \\
&= \int_{S^{n-1}} \int_{\rho_A(x)}^\infty \alpha^{1-r\rho_A^{-1}(x)} \xi^{r\rho_A^{-1}(x)} r^{n-1} dr dx \\
&= \int_{S^{n-1}} \rho_A^n(x) \int_1^\infty \alpha^{1-r} \xi^r r^{n-1} dr dx \\
&= \int_{S^{n-1}} \rho_A^n(x) \left(\alpha \left(-\log(\xi/\alpha)\right)^{-n} \Gamma(n,\ -\log(\xi/\alpha)) \right) dx \\
&= n\alpha \left(-\log(\xi/\alpha)\right)^{-n} \Gamma(n,\ -\log(\xi/\alpha)) \mathrm{V}_n(A).
\end{aligned}
$$

Recall that $\int_{\mathbb{R}^n} \min\{\xi,\ f(x)\}\, dx = \zeta$,

$$
\begin{aligned}
\zeta &= \int_{\mathbb{R}^n} \min\{\xi,\ f(x)\}\, dx \\
&= \int_{\mathbb{R}^n \setminus A} f(x) dx + \xi \mathrm{V}_n(A) \\
&\leq \left(n\alpha \left(-\log(\xi/\alpha)\right)^{-n} \Gamma(n,\ -\log(\xi/\alpha)) + \xi \right) \mathrm{V}_n(A).
\end{aligned}
$$

Also

$$
\begin{aligned}
\zeta &= \int_{\mathbb{R}^n} \min\{\xi,\ f(x)\}\, dx \\
&= 1 - \int_A f(x) dx + \xi \mathrm{V}_n(A) \\
&\leq 1 - \left(n\alpha \nu^{-n} \left(\Gamma(n) - \Gamma(n, \nu)\right) - \xi \right) \mathrm{V}_n(A),
\end{aligned}
$$

where $\nu = -\log(\xi/\alpha)$. Since $\zeta \le ac$ and $\zeta \le 1 - bc$ implies $\zeta \le a/(a+b)$, we have

$$\begin{aligned}\zeta &\le \frac{n\alpha\nu^{-n}\Gamma(n,\nu)+\xi}{\left(n\alpha\nu^{-n}\Gamma(n,\nu)+\xi\right)+\left(n\alpha\nu^{-n}\left(\Gamma(n)-\Gamma(n,\nu)\right)-\xi\right)}\\ &= \frac{n\Gamma(n,\nu)+e^{-\nu}\nu^{n}}{n\Gamma(n)}\\ &= \frac{\Gamma(n+1,\ \nu)}{\Gamma(n+1)}.\end{aligned}$$

By Lemma 2, $h(X^n) \ge -\log\alpha$. Recall that $\tilde{h}_\zeta(X^n)$ is the entropy of the pdf $\tilde{f}(x) = \zeta^{-1}\min\{\xi,\ f(x)\}$, which is also log-concave. Hence by Lemma 2, $\tilde{h}_\zeta(X^n) \le -\log(\zeta^{-1}\xi) + n\log e$. As a result,

$$\begin{aligned}\tilde{h}_\zeta(X^n) - h(X^n) &\le -\log\left(\zeta^{-1}\xi/\alpha\right) + n\log e\\ &\le \log\zeta + \nu + n\log e.\end{aligned}$$

To prove the second bound, assume that $\zeta \ge e^{-(e-2)n}$ and $\nu > en$. We use the bound

$$\Gamma(a,z) < Bz^{a-1}e^{-z}$$

for $a > 1,\ B > 1,\ z > B(a-1)/(B-1)$ due to [16].

Substituting $a = n+1,\ z = en$, and $B = e$, we have $\Gamma(n+1,\nu) < \Gamma(n+1,en) < e(en)^n e^{-en}$. We also know that $\Gamma(n+1) \ge n^n e^{-(n-1)}$, hence

$$\begin{aligned}\frac{\Gamma(n+1,\ \nu)}{\Gamma(n+1)} &< \frac{e(en)^n e^{-en}}{n^n e^{-(n-1)}}\\ &= e^{-(e-2)n},\end{aligned}$$

which leads to a contradiction and $\nu \le en$. The result follows.

ACKNOWLEDGEMENT

The authors wish to thank the anonymous reviewers for several comments that have greatly helped improve this paper.

REFERENCES

[1] D. E. Knuth and A. C. Yao, "The complexity of nonuniform random number generation," in *Algorithms and Complexity*. New York, NY, USA: Academic, 1976, pp. 357–428.
[2] G. R. Kumar, C. T. Li, and A. El Gamal, "Exact common information," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2014, pp. 161–165.
[3] A. D. Wyner, "The common information of two dependent random variables," *IEEE Trans. Inf. Theory*, vol. IT-21, no. 2, pp. 163–179, Mar. 1975.
[4] P. W. Cuff, H. H. Permuter, and T. M. Cover, "Coordination capacity," *IEEE Trans. Inf. Theory*, vol. 56, no. 9, pp. 4181–4206, Sep. 2010.
[5] W. Liu, G. Xu, and B. Chen, "The common information of N dependent random variables," in *Proc. 48th Annu. Allerton Conf. Commun., Control, Comput.*, Sep./Oct. 2010, pp. 836–843.
[6] T. S. Han, "Nonnegative entropy measures of multivariate symmetric correlations," *Inf. Control*, vol. 36, no. 2, pp. 133–156, 1978.
[7] G. Xu, W. Liu, and B. Chen, "Wyner's common information for continuous random variables—A lossy source coding interpretation," in *Proc. 45th Annu. Conf. Inf. Sci. Syst.*, Mar. 2011, pp. 1–6.
[8] S. Satpathy and P. Cuff, "Gaussian secure source coding and Wyner's common information," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2015, pp. 116–120.
[9] P. Yang and B. Chen, "Wyner's common information in Gaussian channels," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2014, pp. 3112–3116.
[10] C. T. Li and A. El Gamal, "Distributed simulation of continuous random variables," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2016, pp. 565–569.
[11] S. Bobkov and M. Madiman, "The entropy per coordinate of a random vector is highly constrained under convexity conditions," *IEEE Trans. Inf. Theory*, vol. 57, no. 8, pp. 4940–4954, Aug. 2011.
[12] P. Cuff, "Distributed channel synthesis," *IEEE Trans. Inf. Theory*, vol. 59, no. 11, pp. 7071–7096, Nov. 2013.
[13] P. Harsha, R. Jain, D. McAllester, and J. Radhakrishnan, "The communication complexity of correlation," *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 438–449, Jan. 2010.
[14] A. Sklar, *Fonctions de Répartition à n Dimensions et Leurs Marges*, vol. 8. Paris, France: Univ. Paris, 1959.
[15] T. S. Han and M. Hoshi, "Interval algorithm for random number generation," *IEEE Trans. Inf. Theory*, vol. 43, no. 2, pp. 599–611, Mar. 1997.
[16] P. Natalini and B. Palumbo, "Inequalities for the incomplete gamma function," *Math. Inequalities Appl.*, vol. 3, no. 1, pp. 69–77, 2000.

Cheuk Ting Li (S'12) received the B.Sc. degree in mathematics and B.Eng. degree in information engineering from The Chinese University of Hong Kong in 2012, and the M.S. degree in electrical engineering from Stanford University in 2014. He is currently working toward the Ph.D. degree at Stanford University. His research interests include generation of random variables, one-shot schemes in information theory and wireless communications.

Abbas El Gamal (S'71–M'73–SM'83–F'00) is the Fortinet Founders Chair of the Department of Electrical Engineering and the Hitachi America Professor in the School of Engineering at Stanford University. He received his B.Sc. Honors degree from Cairo University in 1972, and his M.S. in Statistics and Ph.D. in Electrical Engineering both from Stanford University in 1977 and 1978, respectively. From 1978 to 1980, he was an Assistant Professor of Electrical Engineering at USC. From 2003 to 2012, he was the Director of the Information Systems Laboratory at Stanford University. His research contributions have been in network information theory, FPGAs, and digital imaging devices and systems. He has authored or coauthored over 230 papers and holds 35 patents in these areas. He is coauthor of the book Network Information Theory (Cambridge Press 2011). He is a member of the US National Academy of Engineering and a Fellow of the IEEE. He received several honors and awards for his research contributions, including the 2016 IEEE Richard Hamming Medal, the 2014 Viterbi Lecture, the 2013 Shannon Memorial Lecture, the 2012 Claude E. Shannon Award, the inaugural Padovani Lecture, and the 2004 INFOCOM Paper Award. He served on the Board of Governors of the Information Theory Society from 2009 to 2016 and was President in 2014.