A Universal Coding Scheme for Remote Generation of Continuous Random Variables

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

***Abstract*— Alice selects an arbitrary pdf f and uses a stochastic encoder to generate a prefix-free codeword M, which is sent to Bob so that he can generate a single instance of the random variable $X \sim f$. We describe a universal coding scheme for this setup which works for any f, and establish an upper bound on its expected codeword length when the pdf f is bounded, orthogonally concave (which includes quasiconcave pdf), and has a finite first absolute moment. A dyadic decomposition scheme is used to express the pdf as a mixture of uniform pdfs over hypercubes. Alice randomly selects a hypercube according to its weight, encodes its position and size into M, and sends it to Bob who generates X uniformly over the hypercube. Compared to previous results on channel simulation, our coding scheme applies to any continuous distribution and does not require two-way communication or shared randomness. Applying our coding scheme to classical simulation of quantum entanglement, we obtain a tighter bound on the average codeword length than previously known.**

***Index Terms*— Universal code, channel simulation, communication complexity, simulation of quantum entanglement.**

I. Introduction

CONSIDER the one-shot remote random variable generation setting depicted in Figure 1. Alice and Bob both agree on a set of distributions $\mathscr{P}$ (over a discrete or continuous set). Alice selects a distribution $p \in \mathscr{P}$ and wishes to allow Bob to generate a random variable X according to this distribution. To accomplish this goal, Alice and Bob use an agreed upon *universal* coding scheme in which Alice uses a stochastic encoder to assign to each $p \in \mathscr{P}$ a codeword $M \in \{0,1\}^*$ from an agreed upon prefix-free code and Bob uses a stochastic decoder to generate a single instance of $X \sim p$ from the received codeword M. Let $L(M)$ be the length of M in bits. The question we aim to answer in this paper is under what conditions does a coding scheme exists such that for every distribution $p \in \mathscr{P}$, Bob can generate $X \sim p$ with finite expected codeword length $\mathsf{E}_p(L(M))$?

The answer to this question clearly depends on the set of distributions $\mathscr{P}$. Consider the following two simple special cases:

1. Let $\mathscr{P}$ be the set of probability mass functions over the positive integers, then we can use the following "generate–compress" strategy. Alice generates $X \sim p$ and then uses a universal code over the positive integers, e.g., [1], [2], to encode X into M. Upon receiving M, Bob recovers X. Using these codes, the expected codeword length $\mathsf{E}_p(L(M))$ is finite as long as $\mathsf{E}_p(\log X)$ is finite. Note that this scheme uses a stochastic encoder but a deterministic decoder.
2. Let X be continuous and the class of pdfs $\mathscr{P}$ has a finite (or countable) cardinality, then we can use the following "compress–generate" strategy. Alice encodes the index of p into M. Upon receiving M, Bob first recovers p then uses it to generates X. Note that this scheme uses a deterministic encoder but a stochastic decoder.

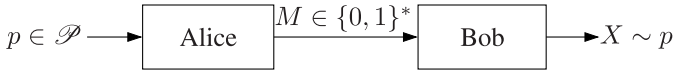

Fig. 1. Universal remote generation of random variables.

If we index the set $\mathscr{P}$ by $\theta \in \Theta$, then our setting can be viewed as a one-shot synthesis (or simulation) of a channel from θ to X with only one-way communication and without common randomness. Several channel simulation scenarios have been previously studied in classical and quantum information theory. In [3], Bennett *et al.* considered the asymptotic setting and established the reverse Shannon theorem, which states that k uses of a channel with capacity C can be simulated using $kC + o(k)$ bits of communication with unlimited amount of common randomness. In [4], Winter studied the asymptotic case with limited common randomness and θ_i distributed according to a given distribution. He showed that $kI(\theta; X) + o(k)$ bits of communication and $kH(X|\theta) + o(k)$ bits of common randomness suffice. Subsequently, Cuff [5] and Bennet *et al.* [6] independently characterized the entire tradeoff region between communication and common randomness for the same setting.

For the one-shot channel simulation setting, schemes based on rejection sampling were developed by Steiner [7], who assumed that Alice and Bob share unlimited common randomness; and by Massar *et al.* [8], who assumed two-way communication between Alice and Bob. Harsha *et al.* [9] established a one-shot version of the reverse Shannon theorem using rejection sampling. These rejection sampling schemes, however, are sensitive to the size of $\mathscr{P}$—a large size $\mathscr{P}$ leads to a high rejection rate, which in turn leads to a high computation time.

Note that the aforementioned asymptotic and one-shot channel simulation schemes are not universal since a scheme designed for a channel from θ to X is guaranteed to work only when the simulated distribution lies in the convex hull of the set of output distributions $\{p(x|\theta)\}$.

Manuscript received October 28, 2016; accepted January 2, 2018. Date of publication February 8, 2018; date of current version March 15, 2018. This paper was presented at the 2016 IEEE Information Theory Workshop.

C. T. Li was with the Department of Electrical Engineering, Stanford University, Stanford, CA 94305 USA. He is now with the Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, Berkeley, CA 94720 USA (e-mail: ctli@berkeley.edu).

A. El Gamal is with the Department of Electrical Engineering, Stanford University, Stanford, CA 94305 USA (e-mail: abbas@ee.stanford.edu).

Communicated by D. L. Neuhoff, Associate Editor for Source Coding.

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.2018.2803752

In this paper, which is an extended and more complete version of [10], we present a universal coding scheme for remote generation of continuous random variables (over scalars or vectors), which we will refer to as *universal dyadic coding scheme*. When $\mathscr{P}$ is restricted to the set of orthogonally concave pdfs $\{f_X(x)\}$, (which includes quasiconcave), we are able to establish an upper bound on the expected codeword length of M in terms of $\sup_x f_X(x)$ and $\mathsf{E}(\|X\|_\infty)$. We use a dyadic decomposition scheme to express the selected pdf as a mixture of uniform distributions over hypercubes. Alice first selects a hypercube from this mixture at random according to its weight, then encodes its position and size into a codeword M using an agreed upon universal code over the integers. Upon receiving M, Bob finds the hypercube and generates X uniformly over it.

In [11], a similar dyadic decomposition scheme was introduced for distributed simulation of continuous random variables according to an agreed upon pdf in a non-universal setting. In Section II we provide a more detailed comparison between the dyadic scheme in [11] and the one used in this paper.

To further motivate our setup and the universal dyadic coding scheme, consider the following two applications.

Application 1 (Classical Simulation of Quantum Entanglement): The simulation of correlations induced by quantum entanglement using classical communication has been widely studied, for example, in [7], [12], and [13]. Consider the Bell state $|\Phi^+\rangle = (|0\rangle_A|0\rangle_B + |1\rangle_A|1\rangle_B)/\sqrt{2}$ of a pair of qubits [14], one held by Alice and the other held by Bob. If Alice measures her qubit in the direction θ_A (unknown to Bob) to obtain $Y_A \in \{\pm 1\}$ and Bob measures his qubit in the direction θ_B (unknown to Alice) to obtain $Y_B \in \{\pm 1\}$, then $\mathsf{P}\{Y_A = 1\} = \mathsf{P}\{Y_B = 1\} = 1/2$ and $\mathsf{E}[Y_A Y_B] = -\cos(\theta_A - \theta_B)$. By Bell's theorem, it is impossible to simulate the joint distribution of (Y_A, Y_B) for all θ_A and θ_B using a classical common randomness source (local hidden variables) between Alice and Bob in place of the qubits. However, such simulation is possible if we instead allow Alice to send a codeword M to Bob. By a modification of the expression in [15] and letting $X \in [0, 2\pi]$ be a random variable with conditional pdf

$$f(x|y_A;\theta_A) = \frac{1}{2}\max\{\cos(y_A(x-\theta_A)),\, 0\}, \tag{1}$$

and $Y_B = -\operatorname{sgn}(\cos(X - \theta_B))$, then (Y_A, Y_B) follows the desired distribution. Hence Alice can generate Y_A and use our universal remote generation coding scheme to encode $f(x|Y_A;\theta_A)$ into M to allow Bob to generate X and Y_B. Using Theorem 3 in Section IV, we show that the expected number of bits is bounded as $\mathsf{E}(L(M)) \le 12.31$, and using numerical computation we show that $\mathsf{E}(L(M)) \le 8.96$ is achievable. In comparison, the scheme in [8], which also does not require common randomness between Alice and Bob, but requires two-way communication (we only allow one-way), provides a looser upper bound of 20 bits on the average number of bits needed.

Application 2 (Minimax Mixed Strategy With a Helper): In decision theory, it is sometimes desirable to adopt a mixed strategy in which the decision is chosen at random. Suppose the payoff $g(X, \theta)$ depends on the agent's decision X and an unknown parameter θ selected from a set Θ (which may be chosen by an adversary). The optimal minimax mixed strategy to choose X is

$$F_X^*(x) = \arg\max_{F_X(x)} \inf_{\theta\in\Theta} \mathsf{E}\left[g(X,\theta)\right].$$

Now suppose the parameter $\theta = (\theta_1, \theta_2)$, where θ_1 is known to a helper (Alice) but θ_2 is not known to Alice or the decision agent (Bob). Alice wishes to help Bob generate the decision X with the optimal pdf given θ_1,

$$F_X^*(x;\theta_1) = \arg\max_{F_X(x)} \inf_{\theta_2} \mathsf{E}\left[g(X,\theta_1,\theta_2)\right].$$

To help Bob generate X using this optimal pdf, Alice can use our universal remote generation coding scheme. For example, consider the payoff function

$$g(x,\theta) = \begin{cases} e^{2\theta - x} & \text{if } x \ge \theta \\ 0 & \text{if } x < \theta, \end{cases}$$

where $\theta \ge 0$. If nothing else is known about θ, then the optimal minimax mixed strategy would be to choose X according to the pdf $f_X^*(x) = e^{-x}$ for $x \ge 0$, which guarantees a payoff of $1/2$. Now assume that Alice knows that $\theta \ge a \ge 0$ (we can let $a = \theta_1$, $\theta_2 \ge 0$, $\theta = \theta_1 + \theta_2$), then the optimal mixed strategy is $f_X^*(x;a) = e^{-(x-a)}$ for $x \ge a$, which results in a payoff of $(1/2)e^a$. As shown in Theorem 2 in Section III, Alice can use our universal remote generation coding scheme to enable Bob to generate X according to this pdf using no more than $\log(a+1)+2\log(\log(a+1)+12)+23$ bits on average. Our universal coding scheme can also be used to perform mixed strategies in other scenarios, e.g., Nash equilibrium [16] in non-cooperative games.

The rest of the paper is organized as follows. For clarity of presentation, we first present the construction of our universal dyadic coding scheme for uniform distributions over subsets of $\mathbb{R}^n$ and upper bound its expected codeword length. In Section III, we extend our scheme to non-uniform distributions, establishing an upper bound on the expected codeword length for orthogonally concave pdfs. In Section IV, we present a variant of our scheme for distributions with a uniform bounded support and apply it to the simulation of the Bell state. Finally, in Section V, we present a lower bound on the expected codeword length in terms of the relative entropy between the actual and the implicit distribution of our scheme.

A. Notation

Throughout this paper, we assume that log is base 2 and entropy H is in bits. Log in base e is written as ln. We use the notation: $[a : b] = [a, b] \cap \mathbb{Z}$,

A set $A \subseteq \mathbb{R}^n$ is 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 orthogonally concave if the hypograph $\{(x,\alpha) : x \in \mathbb{R}^n,\ \alpha \le f(x)\}$ is orthogonally convex.

We denote the volume of a Lebesgue measurable set $A \subseteq \mathbb{R}^n$ by $\mathrm{V}_n(A) = \int_{\mathbb{R}^n} \mathbf{1}_A(x)dx$. For $A, B \subseteq \mathbb{R}^n$, $A + B$

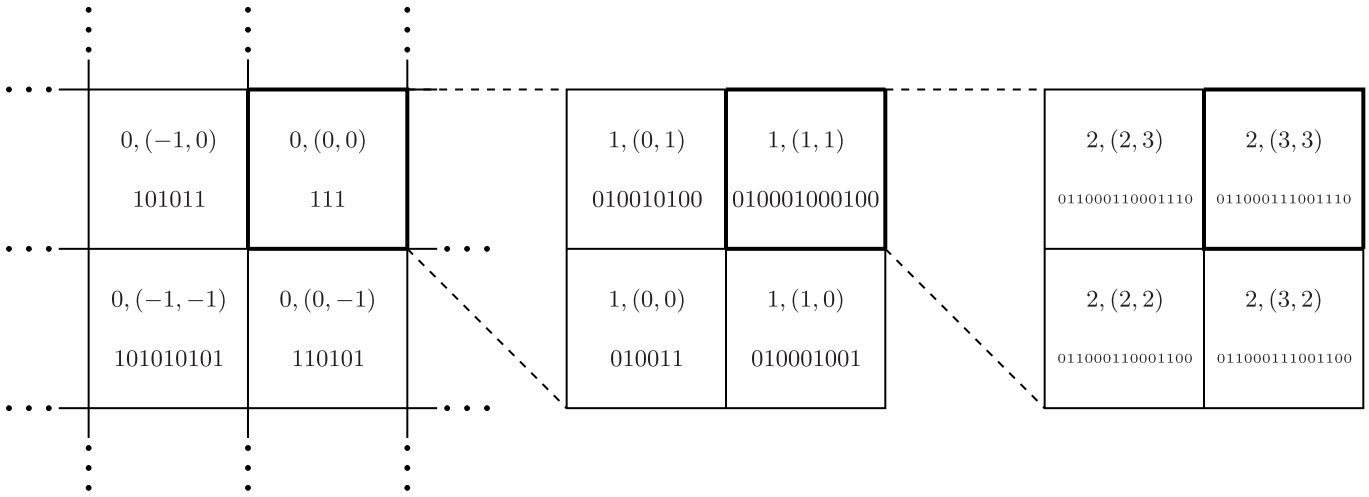

Fig. 2. Dyadic squares $C_{k,v}$ used in the dyadic decomposition for $n = 2$, showing their associated k, v and codeword assignments.

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* $A \ominus B$ is defined as $\{x \in \mathbb{R}^n : B + x \subseteq A\}$.

II. Uniform Distributions

In this section, we develop our universal dyadic coding scheme for the set of uniform pdfs over finite volume sets $A \subseteq \mathbb{R}^n$. We first introduce the dyadic decomposition of a set [11], which is the building block of our coding scheme.

Definition 1 (Dyadic Decomposition): For $v \in \mathbb{Z}^n$ and $k \in \mathbb{Z}$, define the hypercube $C_{k,v} = 2^{-k}([0, 1]^n + v) \subset \mathbb{R}^n$. 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 of A 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.

Our scheme uses a universal code over the integers to encode the position and size of the hypercubes. In particular, we will use the signed Elias delta code defined as follows [2]. Let

$$g_{\gamma+}(k) = 0^N \parallel 1 \parallel a_{N-1}a_{N-2}\ldots a_0,$$
$$g_{\delta+}(k) = g_{\gamma+}(N + 1) \parallel a_{N-1}a_{N-2}\ldots a_0.$$

Then the signed Elias code is

$$g_\delta(k) = \begin{cases} g_{\delta+}(1 - 2k) & \text{if } k \le 0, \\ g_{\delta+}(2k) & \text{if } k > 0, \end{cases}$$

where $a_N a_{N-1}\ldots a_0$ is the binary representation of k. The length of the codeword $g_\delta(k)$ is

$$L(g_\delta(k)) = \lfloor \log(2|k| + 1)\rfloor + 2\lfloor \log(\lfloor \log(2|k| + 1)\rfloor + 1)\rfloor + 1. \quad (2)$$

We are now ready to define the universal dyadic coding scheme for the set of uniform pdfs.

Universal dyadic coding scheme for uniform pdfs. The universal dyadic coding scheme for the set of uniform pdfs over positive, finite volume subsets $A \subset \mathbb{R}^n$ with a boundary of measure zero consists of:

1) A stochastic encoder that generates $\tilde{x}$ according to the observed uniform pdf over A. It then finds (k, v) such that $v \in D_k(A)$ and $\tilde{x} \in C_{k,v}$. The encoder then maps (k, v) into a codeword m which consists of the concatenation of signed Elias delta codewords for $k, v_1, \ldots, v_n$, i.e., $m = g_C(k, v) = g_\delta(k) \parallel g_\delta(v_1) \parallel \cdots \parallel g_\delta(v_n)$.
2) A stochastic decoder that upon receiving m recovers (v, k) and generates x according to a uniform pdf over $C_{k,v}$.

The dyadic decomposition for $\mathbb{R}^2$ and the assignments of codeword to the squares are illustrated in Figure 2.

The following illustrates how our scheme is used for a given pdf.

Example 1: Consider a uniform pdf over the ellipse $A = \left\{x \in \mathbb{R}^2 : x^T K x < 1\right\}$, $K = \begin{bmatrix} 4/3 & -2/3 \\ -2/3 & 4/3 \end{bmatrix}$. Figure 3 depicts the universal dyadic coding scheme for this pdf. The encoder first generates a point in the ellipse uniformly at random, and then sends the codeword representing the square containing the point. The expected codeword length (computed by listing all squares in the dyadic decomposition with side length at least 2^{-16}) is 15.6. Note that the entropy of M, $H(M) = 6.35$ is significantly smaller since the code is universal.

The length of the codeword of the universal dyadic coding scheme depends on the magnitude of k and $v_1, \ldots, v_n$, (which depends on k and $\|x\|$), hence the length can be bounded using k and $\|x\|$. In [11], it is shown that the expected value of k can be bounded using the following quantity.

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

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

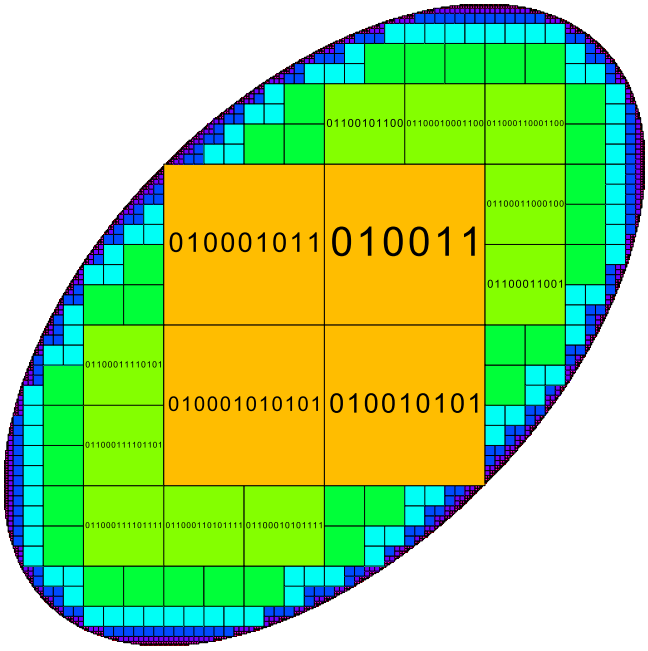

Fig. 3. Universal dyadic coding scheme on the uniform distribution over an ellipse.

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

If A is orthogonally convex, the erosion entropy of A by the hypercube $[0,1]^n$ can be bounded by the expected norm of the uniform distribution on A, as shown in the following.

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

$$h_{\ominus[0,1]^n}(A) \le (n-1)\log \mathsf{E}\left[\|X\|_\infty\right] - \log \mathrm{V}_n(A) + 4n.$$

The proof of this lemma is given in Appendix V-A. We now use the erosion entropy to bound the expected codeword length of the universal dyadic coding scheme.

Theorem 1: The expected codeword length of the universal dyadic coding scheme for uniform pdfs for $X \sim \mathrm{Unif}(A)$ is upper bounded as

$$\begin{aligned}
&\mathsf{E}[L(M)] \\
&\quad \le n\ell_\delta\left(h + \mathsf{E}\left[\log\left(\|X\|_\infty + \mathrm{V}_n^{1/n}(A)\right)\right] + 4\right) \\
&\qquad + \ell_\delta\left(\log\left(h + 2\max\left\{\log \mathrm{V}_n^{1/n}(A),\, 0\right\} + \frac{5}{2}\right) + 2\right) \\
&\quad \le n\ell_\delta\left(h + \log \mathsf{E}[\|X\|_\infty] + 8\right) \\
&\qquad + \ell_\delta\left(\log\left(h + 2\max\left\{\log \mathsf{E}[\|X\|_\infty],\, 0\right\} + 9\right) + 2\right),
\end{aligned}$$

where $\ell_\delta(t) = t + 2\log t$ and $h = h_{\ominus[0,1]^n}(A)$.

Theorem 1 shows that the expected codeword length depends on the erosion entropy, the expected magnitude of X, and the volume of the set. Intuitively, the erosion entropy measures the complexity of the set (or loosely speaking its surface area to volume ratio). However, the erosion entropy is invariant under shifting. Since our universal scheme is sensitive to the position of A as well its shape, the bound in Theorem 1 depends also on the expected magnitude of X. The function $\ell_\delta(t)$ in Theorem 1 comes from the length of the Elias delta code in (2). Other universal codes for integers may be used in place of Elias delta code, and result in a different bound.

We now prove Theorem 1.

Proof of Theorem 1: Let $x \in A$. Consider the length of the codeword for (v,k) with $v \in D_k(A)$ and $x \in C_{k,v}$. We have $x \in 2^{-k}([0,1]^n + v)$, hence $\|x\|_\infty \ge 2^{-k}\max_i(|v_i + 1/2| - 1/2)$. Since $2^{-nk} \le \mathrm{V}_n(A)$, $k \ge -(1/n)\log \mathrm{V}_n(A)$. Let

$$\tau = 2\max\left\{(1/n)\log \mathrm{V}_n(A),\, 0\right\},$$

then $|k| \le k + \tau$. From (2), the length of the codeword for (v,k) is

$$\begin{aligned}
&L(g_C(v,k)) \\
&\le \lfloor \log(2|k|+1)\rfloor + 2\lfloor \log(\lfloor \log(2|k|+1)\rfloor + 1)\rfloor \\
&\quad + \sum_{i=1}^n \left(\lfloor \log(2|v_i|+1)\rfloor + 2\lfloor \log(\lfloor \log(2|v_i|+1)\rfloor + 1)\rfloor\right) \\
&\quad + n + 1 \\
&\overset{(a)}{\le} \log(|k| + 1/2) + 2\log(\log(|k|+1/2)+2) \\
&\quad + \sum_{i=1}^n \left(\log\left(\left|v_i + \frac{1}{2}\right| + \frac{1}{2}\right)\right. \\
&\quad \left. + 2\log\left(\log\left(\left|v_i + \frac{1}{2}\right| + \frac{1}{2}\right) + 2\right)\right) + 2n + 2 \\
&\le \log(k + \tau + 1/2) + 2\log(\log(k+\tau+1/2)+2) \\
&\quad + \sum_{i=1}^n \left(\log\left(2^k\left(\|x\|_\infty + 2^{-k}\right)\right)\right. \\
&\quad \left. + 2\log\left(\log\left(2^k\left(\|x\|_\infty + 2^{-k}\right)\right) + 2\right)\right) + 2n + 2 \\
&\le \log(k + \tau + 1/2) + 2\log(\log(k+\tau+1/2)+2) \\
&\quad + n\left(\log\left(2^k\left(\|x\|_\infty + \mathrm{V}_n^{1/n}(A)\right)\right)\right. \\
&\quad \left. + 2\log\left(\log\left(2^k\left(\|x\|_\infty + \mathrm{V}_n^{1/n}(A)\right)\right) + 2\right)\right) + 2n + 2 \\
&= n\left(\log\left(\|x\|_\infty + \mathrm{V}_n^{1/n}(A)\right) + k + 2\right. \\
&\quad \left. + 2\log\left(\log\left(\|x\|_\infty + \mathrm{V}_n^{1/n}(A)\right) + k + 2\right)\right) \\
&\quad + \log(k+\tau+1/2) + 2 + 2\log(\log(k+\tau+1/2)+2) \\
&= n\ell_\delta\left(\log\left(\|x\|_\infty + \mathrm{V}_n^{1/n}(A)\right) + k + 2\right) \\
&\quad + \ell_\delta(\log(k+\tau+1/2)+2) \qquad (3)
\end{aligned}$$

where (a) follows by the fact that $\lfloor \log(2|i|+1)\rfloor \le \log(|2i+1|+1)$.

Let $X \sim \mathrm{Unif}(A)$, and V, K be such that $V \in D_K(A)$ and $X \in C_{K,V}$. Then we have

$$\begin{aligned}
&\mathsf{E}[L(Enc(\mathrm{Unif}(A))] \\
&\quad = \mathsf{E}[L(g_C(V,K)] \\
&\quad \le n\,\mathsf{E}\left[\ell_\delta\left(\log\left(\|X\|_\infty + \mathrm{V}_n^{1/n}(A)\right) + K + 2\right)\right] \\
&\qquad + \mathsf{E}\left[\ell_\delta(\log(K+\tau+1/2)+2)\right] \\
&\quad \le n\ell_\delta\left(\mathsf{E}\left[\log\left(\|X\|_\infty + \mathrm{V}_n^{1/n}(A)\right)\right] + \mathsf{E}[K] + 2\right) \\
&\qquad + \ell_\delta(\log(\mathsf{E}[K] + \tau + 1/2) + 2)
\end{aligned}$$

by Jensen's inequality and the concavity of ℓ_δ. We now proceed to bound

$$\mathsf{E}[K] = \tfrac{1}{\mathrm{V}_n(A)}\textstyle\sum_{k=-\infty}^{\infty} k \cdot 2^{-nk}\,|D_k(A)|.$$

Consider

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

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. The same definition of $C_{k,v} = 2^{-k}([0,1]^n + v)$ can still be applied, however. 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

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

and

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

As a result,

$$\begin{aligned}
&\mathsf{E}[K] \\
&= \frac{1}{\mathrm{V}_n(A)} \sum_{k=-\infty}^{\infty} k \cdot 2^{-nk} |D_k(A)| \\
&= \frac{1}{\mathrm{V}_n(A)} \sum_{k=-\infty}^{\infty} \left(\mathbf{1}\{k \geq 0\} \mathrm{V}_n(A) - \sum_{l=-\infty}^{k} 2^{-nl} |D_l(A)|\right) \\
&\leq \frac{1}{\mathrm{V}_n(A)} \sum_{k=-\infty}^{\infty} \left(\mathbf{1}\{k \geq 0\} \mathrm{V}_n(A) - \mathrm{V}_n\left(A \ominus [0,\, 2^{-(k-1)}]^n\right)\right) \\
&\leq \frac{1}{\mathrm{V}_n(A)} \int \left(\mathbf{1}\{t \geq 0\} \mathrm{V}_n(A) - \mathrm{V}_n\left(A \ominus [0,\, 2^{-t}]^n\right)\right) dt + 2 \\
&= h_{\ominus [0,1]^n}(A) + 2.
\end{aligned}$$

We have

$$\mathsf{E}[L(M)] \leq n\ell_\delta\left(h + \mathsf{E}\left[\log\left(\|X\|_\infty + \mathrm{V}_n^{1/n}(A)\right)\right] + 4\right) + \ell_\delta\left(\log\left(h + 2\max\left\{\log \mathrm{V}_n^{1/n}(A),\, 0\right\} + \frac{5}{2}\right) + 2\right).$$

It remains to bound $\mathrm{V}_n^{1/n}(A)$ by $\mathsf{E}[\|X\|_\infty]$. By the Markov inequality,

$$\begin{aligned}
&\mathsf{E}[\|X\|_\infty] \\
&\quad \geq \frac{1}{4}\mathrm{V}_n^{1/n}(A) \cdot \mathrm{P}\left\{\|X\|_\infty \geq \frac{1}{4}\mathrm{V}_n^{1/n}(A)\right\} \\
&\quad = \frac{1}{4}\mathrm{V}_n^{1/n}(A) \cdot \frac{\mathrm{V}_n\left\{x \in A : \|x\|_\infty \geq (1/4)\mathrm{V}_n^{1/n}(A)\right\}}{\mathrm{V}_n(A)} \\
&\quad \geq \frac{1}{4}\mathrm{V}_n^{1/n}(A) \cdot \frac{\mathrm{V}_n(A) - \left((1/2)\mathrm{V}_n^{1/n}(A)\right)^n}{\mathrm{V}_n(A)} \\
&\quad \geq \frac{1}{8}\mathrm{V}_n^{1/n}(A). \qquad (4)
\end{aligned}$$

Hence,

$$\begin{aligned}
&\mathsf{E}[L(M)] \\
&\leq n\ell_\delta(h + \log \mathsf{E}[\|X\|_\infty] + 4 + \log 9) \\
&\quad + \ell_\delta\left(\log\left(h + 2\max\{\log \mathsf{E}[\|X\|_\infty] + 3,\, 0\} + \frac{5}{2}\right) + 2\right). \qquad (5)
\end{aligned}$$

This completes the proof of the theorem. ■

Combining Lemma 1 and Theorem 1, we can bound the expected length of the universal dyadic coding scheme for orthogonally convex sets.

Corollary 1: The expected codeword length of the universal dyadic coding scheme for uniform pdfs applied to an orthogonally convex $A \subseteq \mathbb{R}^n$ is upper bounded as

$$\begin{aligned}
&E[L(M)] \\
&\leq n\ell_\delta\big((n-1)\log r + \log(\|\hat{x}\|_\infty + r) - \log \mathrm{V}_n(A) + 4n + 8\big) \\
&\quad + \ell_\delta\big(\log\big((n-1)\log r + 2\max\{r,\, 0\} \\
&\quad - \log \mathrm{V}_n(A) + 4n + 9\big) + 2\big).
\end{aligned}$$

for any $\hat{x} \in \mathbb{R}^n$, where $\ell_\delta(t) = t + 2\log t$ and $r = \mathsf{E}\left[\|X - \hat{x}\|_\infty\right]$.

Proof of Corollary 1: By Lemma 1,

$$h_{\ominus [0,1]^n}(A) \leq (n-1)\log \mathsf{E}\left[\|X - \hat{x}\|_\infty\right] - \log \mathrm{V}_n(A) + 4n.$$

By Theorem 1,

$$\begin{aligned}
&\mathsf{E}[L(M)] \\
&\leq n\ell_\delta\left(h + \log\left(\mathsf{E}[\|X\|_\infty] + \mathrm{V}_n^{1/n}(A)\right) + 4\right) \\
&\quad + \ell_\delta\left(\log\left(h + 2\max\left\{\log \mathrm{V}_n^{1/n}(A),\, 0\right\} + \frac{5}{2}\right) + 2\right) \\
&\leq n\ell_\delta(h + \log \mathsf{E}[\|X\|_\infty] + 8) \\
&\quad + \ell_\delta\left(\log\left(h + 2\max\left\{\log \mathsf{E}[\|X - \hat{x}\|_\infty],\, 0\right\} + 9\right) + 2\right) \\
&\leq n\ell_\delta\big((n-1)\log \mathsf{E}\left[\|X - \hat{x}\|_\infty\right] + \log \mathsf{E}[\|X\|_\infty] \\
&\quad - \log \mathrm{V}_n(A) + 4n + 8\big) + \ell_\delta\big(\log\big((n-1)\log \mathsf{E}\left[\|X - \hat{x}\|_\infty\right] \\
&\quad + 2\max\left\{\log \mathsf{E}[\|X - \hat{x}\|_\infty],\, 0\right\} - \log \mathrm{V}_n(A) + 4n + 9\big) + 2\big).
\end{aligned}$$

■

An added benefit of our universal dyadic coding scheme is that X can be generated in a distributed manner. Suppose X is an n-dimensional vector $X_1, \ldots, X_n$. Instead of having one decoder wishing to generate X, we have n decoders that

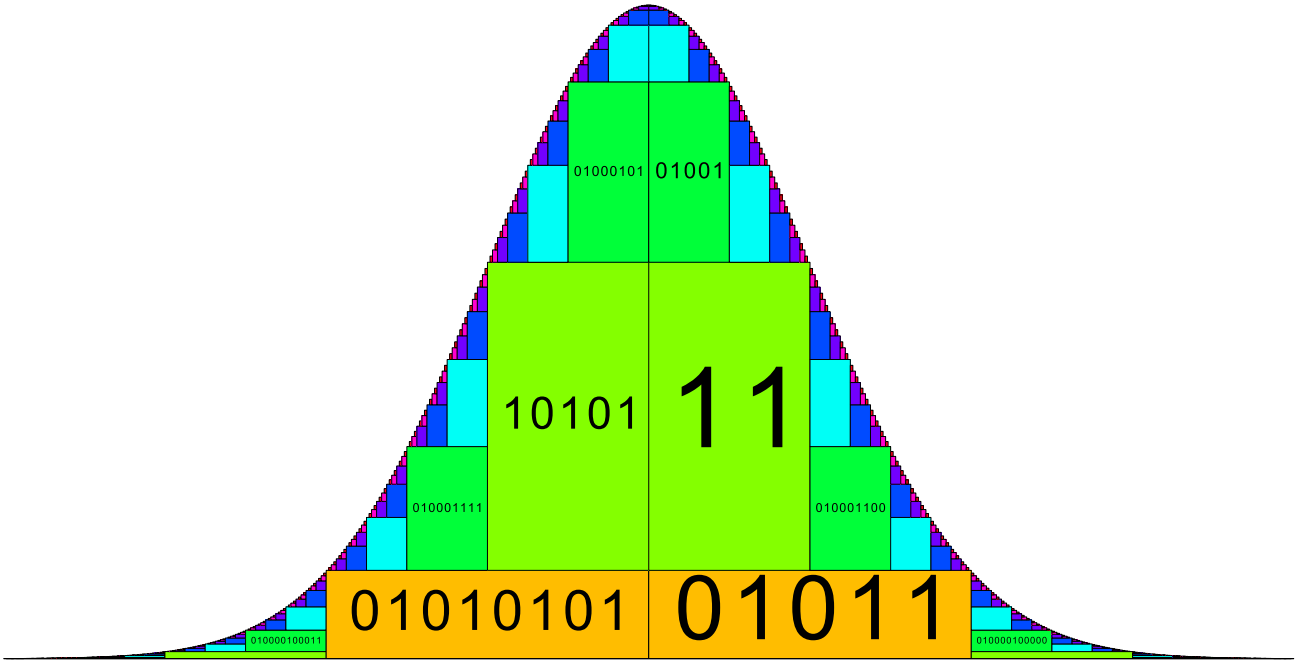

Fig. 4. Universal dyadic coding scheme on $\mathcal{N}(0,1)$.

all receive M and decoder i wishes to generate only X_i, $i \in [1:n]$. Such distributed generation is possible using our universal dyadic coding scheme since decoder i can generate X_i uniformly over the interval $[2^{-k}v_i,\ 2^{-k}(v_i+1)]$ without any need to cooperate with other decoders. In [11]. we described a dyadic decomposition coding scheme for distributed generation of a given pdf. The scheme in this paper differs from that in [11] in several aspects.

- The scheme in this paper is universal, while the scheme in [11] is constructed for a given pdf known to both the encoder and the decoders.
- In [11] we used an optimal prefix free code, such as a Huffman code, to encode the hypercubes, while in this paper we use a universal code over the integers since the distribution on the hypercubes is not known a priori.
- In [11], we can perform scaling (and bijective transformations) on each variable X_i before applying the dyadic decomposition scheme. It is not possible to perform such preprocessing here since the decoder would not know the scaling factor or the bijective transformation used.
- In the analysis of the expected codeword length in [11], it suffices to consider only the distribution of the sizes of the hypercubes. In our universal scheme, both the size and the position of the hypercube affect the length of the codeword assigned to it.

III. Non-Uniform Distributions

In this section, we extend the results of the previous section to the case where the pdf of $X = (X_1, \ldots, X_n)$ is selected from a set of arbitrary (not necessarily uniform) pdfs. The key idea in extending our scheme is the following. Note that in general, any pdf can be written as a mixture of uniform pdfs. Let $Z \sim f_Z$, where $f_Z(z) = \mathrm{V}_n(L_z^+(f))$ for $z \geq 0$ and $L_z^+(f) = \{x \in \mathbb{R}^n : f(x) \geq z\}$ is the superlevel set of f. Let $X|\{Z = z\} \sim \mathrm{Unif}(L_z^+(f))$, then we have $X \sim f(x)$. Hence $f(x)$ can be expressed as a mixture of uniform distributions over $L_z^+(f)$ for different values of z. Alice can first generate $Z \sim f_Z$, then apply the universal dyadic coding scheme for uniform distributions on $L_Z^+(f)$. The scheme is formally defined as follows.

Universal dyadic coding scheme for general pdfs. The universal dyadic coding scheme for the set of almost everywhere continuous pdfs $\mathscr{P}$ consists of:

1) A stochastic encoder that generates $\tilde{x}$ according to the observed f and generates z uniformly in $[0, f(\tilde{x})]$, and finds (k, v) such that $v \in D_k(L_z^+(f))$ and $\tilde{x} \in C_{k,v}$. The encoder maps (k, v) into a codeword m that consists of the concatenation of the signed Elias delta codewords for $k, v_1, \ldots, v_n$, i.e., $m = g_C(k, v) = g_\delta(k) \,\|\, g_\delta(v_1) \,\|\, \cdots \,\|\, g_\delta(v_n)$.
2) A stochastic decoder that upon receiving m recovers (v, k) and generates x uniformly over $C_{k,v}$.

We illustrate this scheme in the following.

Example 2: Assume that the selected pdf is Gaussian with zero mean and unit variance. Figure 4 depicts the universal dyadic coding scheme for this pdf. The horizontal and vertical axes represent x and z, respectively. The encoder sends the codeword for the rectangle containing (x, z). The expected codeword length (computed by listing all intervals in the dyadic decomposition with length at least 2^{-20}) is 7.06.

As a consequence of Theorem 1, we have the following bound on the expected codeword length.

Theorem 2: The expected codeword length of the universal dyadic coding scheme for $X \sim f(x)$ is upper bounded as

$$\begin{aligned} &\mathsf{E}[L(M)] \\ &\leq n\ell_\delta\left(\mathsf{E}_Z[h] + \log \mathsf{E}[\|X\|_\infty] + 8\right) \\ &\quad + \ell_\delta\left(\log\left(\mathsf{E}_Z[h] + 2\max\{\log \mathsf{E}[\|X\|_\infty],\, 0\} + 10\right) + 2\right), \end{aligned}$$

where $\ell_\delta(t) = t + 2\log t$ and $h = h_{\ominus[0,1]^n}(L_Z^+(f))$ is a random variable, where $Z \sim f_Z$, $f_Z(z) = \mathrm{V}_n(L_z^+(f))$ for $z \geq 0$.

Proof: Using (5),

$$\begin{aligned}
&\mathsf{E}[L(M)] \\
&\le n\,\mathsf{E}_Z\left[\ell_\delta\left(h + \log \mathsf{E}[\|X\|_\infty \mid Z] + 8\right)\right] \\
&\quad + \mathsf{E}_Z\left[\ell_\delta\left(\log\left(h + 2\max\left\{\log \mathsf{E}[\|X\|_\infty],\, 0\right\} + 8.5\right) + 2\right)\right] \\
&\le n\ell_\delta\left(\mathsf{E}_Z[h] + \log \mathsf{E}[\|X\|_\infty] + 8\right) \\
&\quad + \ell_\delta\Big(\log\Big(\mathsf{E}_Z[h] + 2\,\mathsf{E}_Z\left[\max\left\{\log \mathsf{E}[\|X\|_\infty \mid Z],\, 0\right\}\right] \\
&\quad + 8.5\Big) + 2\Big).
\end{aligned}$$

To bound $\mathsf{E}_Z\left[\max\left\{\log \mathsf{E}[\|X\|_\infty \mid Z],\, 0\right\}\right]$, define a concave function $q : [0, \infty) \to [0, \infty)$,

$$q(t) = \begin{cases} te^{-1}\log e & \text{if } t \le e \\ \log t & \text{if } t > e. \end{cases}$$

Note that

$$\max\{\log t,\, 0\} \le q(t) \le \max\{\log t,\, 0\} + e^{-1}\log e.$$

Hence,

$$\begin{aligned}
&\mathsf{E}_Z\left[\max\left\{\log \mathsf{E}[\|X\|_\infty \mid Z],\, 0\right\}\right] \\
&\qquad \le \mathsf{E}_Z\left[q\left(\mathsf{E}[\|X\|_\infty \mid Z]\right)\right] \\
&\qquad \le q\left(\mathsf{E}[\|X\|_\infty]\right) \\
&\qquad \le \max\left\{\log \mathsf{E}[\|X\|_\infty],\, 0\right\} + e^{-1}\log e.
\end{aligned}$$

Note that we also need $L_z^+(f)$ to have a boundary of measure zero for almost all z, in order for the coding scheme to succeed almost surely. This is implied by the almost everywhere continuity of $f(x)$. The proof of this claim is given in Appendix V-B. ∎

We can also generalize Corollary 1 to orthogonally concave pdfs (which includes quasiconcave pdfs) as follows.

Corollary 2: The expected codeword length of the universal dyadic coding scheme for $X \sim f(x)$, where f is orthogonally concave, is upper bounded as

$$\begin{aligned}
&E[L(M)] \\
&\le n\ell_\delta\left((n-1)\log r + \log(\left\|\hat{x}\right\|_\infty + r) + h(Z) + 4n + 8\right) \\
&\quad + \ell_\delta\Big(\log\left((n-1)\log r + 2\max\left\{\log r,\, 0\right\}\right. \\
&\quad \left. + h(Z) + 4n + 10\right) + 2\Big).
\end{aligned}$$

for any $\hat{x} \in \mathbb{R}^n$, where $\ell_\delta(t) = t + 2\log t$, $r = \mathsf{E}\left[\left\|X - \hat{x}\right\|_\infty\right]$, $Z \sim f_Z$, $f_Z(z) = \mathrm{V}_n(L_z^+(f))$ for $z \ge 0$. As a result, if $\sup_x f(x) < \infty$,

$$\begin{aligned}
&E[L(M)] \\
&\le n\ell_\delta\left((n-1)\log r + \log(\left\|\hat{x}\right\|_\infty + r) + \log\sup_x f(x)\right. \\
&\quad \left. + 4n + 8\right) + \ell_\delta\Big(\log\left((n-1)\log r + 2\max\left\{\log r,\, 0\right\}\right. \\
&\quad \left. + \log\sup_x f(x) + 4n + 10\right) + 2\Big).
\end{aligned}$$

IV. Bounded Support Distributions

In this section, we present a variant of the universal dyadic coding scheme for a set of distributions with a uniform bound on their supports. Without loss of generality, assume $\mathscr{P}$ consists of the set of pdfs over $[0, 1]^n$. Since the υ in the definition of our universal dyadic coding scheme (corresponding to the position of the hypercube) is bounded, we can use a fixed length code to encode $(\upsilon_1, \ldots, \upsilon_n)$. This allows us to reduce the expected codeword length.

Universal dyadic coding scheme for pdfs over the unit hypercube. The universal dyadic coding scheme for pdfs over $[0, 1]^n$ consists of:

1) A stochastic encoder that generates $\tilde{x}$ according to the observed f and generates z uniformly in $[0, f(\tilde{x})]$, and finds (k, υ) such that $\upsilon \in D_k(L_z^+(f))$ and $\tilde{x} \in C_{k,\upsilon}$. The encoder then maps (k, υ) into a codeword which consists of the concatenation of the unsigned Elias gamma codeword for $k + 1$, and the k-bit binary representations of $\upsilon_1, \ldots, \upsilon_n$, i.e., $m = g_C(k, \upsilon) = g_\gamma + (k + 1) \,\|\, g_{\mathrm{b},k}(\upsilon_1) \,\|\, \cdots \,\|\, g_{\mathrm{b},k}(\upsilon_1)$, where $g_{\mathrm{b},k}(i)$ is the binary representation of i with k bits, possibly with leading zeros.
2) A stochastic decoder that upon observing m recovers (υ, k) and generates x uniformly over $C_{k,\upsilon}$.

Since the length of the unsigned Elias gamma codeword $g_\gamma(k + 1)$ is $2\lfloor\log(k + 1)\rfloor + 1$, the length of m is

$$L(m) = nk + 2\lfloor\log(k + 1)\rfloor + 1.$$

The expected codeword length is upper bounded as follows.

Theorem 3: The expected codeword length of the universal dyadic coding scheme for pdfs over the unit hypercube for $X \sim f(x)$, where f is orthogonally concave, is upper bounded as

$$\begin{aligned}
E[L(M)] &\le n\left(h(Z) + \log n + \log e + 2\right) \\
&\quad + 2\log\left(h(Z) + \log n + \log e + 3\right) + 1,
\end{aligned}$$

where $Z \sim f_Z$, $f_Z(z) = \mathrm{V}_n(L_z^+(f))$ for $z \ge 0$. As a result, if $\sup_x f(x) < \infty$,

$$\begin{aligned}
E[L(M)] &\le n\left(\log\sup_x f(x) + \log n + \log e + 2\right) \\
&\quad + 2\log\left(\log\sup_x f(x) + \log n + \log e + 3\right) + 1.
\end{aligned}$$

Proof: In [11. Th. 1], it was shown that the erosion entropy for orthogonally convex A is bounded as

$$h_{\ominus[0,1]^n}(A) \le \log\left(\frac{\sum_{i=1}^n \mathrm{VP}_{\setminus i}(A)}{\mathrm{V}_n(A)}\right) + \log e,$$

where $\mathrm{VP}_{\setminus i}(A) = \{(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n) : x \in A\}$. Let $Z \sim f_Z$, $f_Z(z) = \mathrm{V}_n(L_z^+(f))$, then

$$h_{\ominus[0,1]^n}(L_z^+(f)) \le -\log \mathrm{V}_n(L_z^+(f)) + \log n + \log e.$$

Hence

$$\begin{aligned}
&\mathsf{E}[L(M)] \\
&\le n\,\mathsf{E}[K] + 2\log(\mathsf{E}[K] + 1) + 1 \\
&\le n\left(\mathsf{E}[h_{\ominus[0,1]^n}(L_Z^+(f))] + 2\right) \\
&\quad + 2\log\left(\mathsf{E}[h_{\ominus[0,1]^n}(L_Z^+(f))] + 3\right) + 1 \\
&\le n\left(\mathsf{E}[-\log \mathrm{V}_n(L_Z^+(f))] + \log n + \log e + 2\right) \\
&\quad + 2\log\left(\mathsf{E}[-\log \mathrm{V}_n(L_Z^+(f))] + \log n + \log e + 3\right) + 1 \\
&= n\left(h(Z) + \log n + \log e + 2\right) \\
&\quad + 2\log\left(h(Z) + \log n + \log e + 3\right) + 1.
\end{aligned}$$

∎

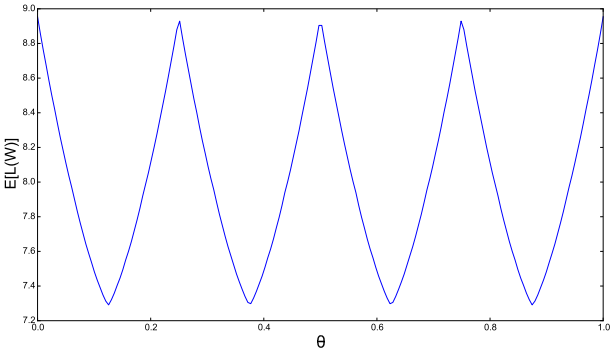

Fig. 5. Universal dyadic coding scheme for pdfs over the unit hypercube applied to the simulation of the Bell state.

As an example, we apply this result to simulating the Bell state in Application 1

$$f(x\,|\,\theta) = \pi \max\{\cos(2\pi(x-\theta)),\, 0\},$$

fitted to the interval $[0, 1]$. Although this pdf is not orthogonally concave, it can be decomposed into at most two orthogonally concave parts with disjoint support, hence the expected codeword length is the weighted average of the expected codeword lengths for those two pieces, which incurs a penalty of at most 1 bit (for communicating which piece to generate from). By Theorem 3,

$$\mathsf{E}\,[L(M)] \le \log\pi + \log e + 2 + 2\log(\log\pi + \log e + 3) + 2 \approx 12.31.$$

Figure 5 plots the numerical values of $\mathsf{E}\,[L(M)]$ versus θ computed by listing all intervals in the dyadic decomposition with length at least 2^{-17}. As can be seen, $\mathsf{E}\,[L(M)] \le 8.96$ for all θ.

V. Lower Bound on Expected Codeword Length

In the previous sections we focused on schemes for universal remote generation of continuous random variables and upper bounds on their expected codeword length. In this section, we give a lower bound on the expected codeword length of a universal remote generation scheme in terms of its implicit distribution, which is an analog to the implicit distribution of a prefix-free code. For a prefix-free code $E : \{1, 2, \ldots\} \to \{0, 1\}^*$ over the positive integers, its implicit distribution (see [17]) is given by

$$p_{\mathrm{Im}}(x) = \Big(\sum_{x'=1}^{\infty} 2^{-L(E(x'))}\Big)^{-1} 2^{-L(E(x))}.$$

When the code is applied on a random integer with distribution $p(x)$, its expected length is lower bounded by

$$\mathsf{E}_p(L(E(X))) \ge H(p) + D(p\|p_{\mathrm{Im}}). \tag{6}$$

Now we define the implicit distribution of a universal remote generation coding scheme. Consider a universal remote generation coding scheme with a prefix-free codeword set $C \subseteq \{0, 1\}^*$. Upon receiving $m \in C$, Bob generates $X|\{M = m\} \sim \tilde{p}_m$ according to a distribution $\tilde{p}_m(dx)$. Define the implicit distribution of this scheme as

$$p_{\mathrm{Im}}(dx) = \Big(\sum_{m'\in C} 2^{-L(m')}\Big)^{-1} \sum_{m\in C} 2^{-L(m)}\,\tilde{p}_m(dx).$$

We now show that the expected codeword length $\mathsf{E}_p(L(M))$ is lower bounded by the relative entropy between p and p_{Im}.

Proposition 1: For a universal remote generation scheme with an implicit distribution p_{Im}, the average codeword length for $p \in \mathscr{P}$ is lower bounded as

$$\mathsf{E}_p(L(M)) \ge D(p\|p_{\mathrm{Im}}).$$

Proof: Consider the input distribution p. Assume the encoder outputs m with probability $a(m)$. Then $p = \sum_{m\in C} a(m)\tilde{p}_m$, $\mathsf{E}_p(L(M)) = \sum_{m\in C} a(m)L(m)$. By convexity of relative entropy,

$$\begin{aligned} D(p\|p_{\mathrm{Im}}) &= D\Big(\sum_{m\in C} a(m)\tilde{p}_m \,\|\, p_{\mathrm{Im}}\Big) \\ &\le \sum_{m\in C} a(m) D\left(\tilde{p}_m\|p_{\mathrm{Im}}\right) \\ &\le \sum_{m\in C} a(m)L(m) \\ &= \mathsf{E}_p(L(M)). \end{aligned}$$

■

Proposition 1 means that a universal remote generation scheme can only be designed to work best on one distribution (the implicit distribution) and the distributions that are close to it. If the implicit distribution is concentrated around its mode and the actual distribution is centered at a mode far from the mode of the implicit distribution, then the expected length will be large. Nevertheless, if the implicit distribution has a slow, power law decay (similar to a universal code over the integers [1], [2]), then it will impose a smaller penalty on distributions not centered at the mode of the implicit distribution.

Consider the (unbounded support) dyadic universal code. Using (3), the implicit distribution in this case is approximately given by the pdf

$$\begin{aligned} &f_{\mathrm{Im}}(x) \\ &\quad\propto \sum_{(\upsilon,k):\, x\in C_{k,\upsilon}} 2^{-L(g_C(\upsilon,k))}\cdot 2^{nk} \\ &\quad\approx \sum_{(\upsilon,k):\, x\in C_{k,\upsilon}} 2^{-\left(n\ell_\delta\left(\log\left(2^k\|x\|_\infty+1\right)\right)+\ell_\delta(\log(|k|+1))\right)}\cdot 2^{nk} \\ &\quad= \sum_{k=-\infty}^{\infty} 2^{-n\log\left(\|x\|_\infty+2^{-k}\right)-2n\log\left(k+\log\left(\|x\|_\infty+2^{-k}\right)\right)-\ell_\delta(\log(|k|+1))} \\ &\quad\propto 2^{-(n\log(\|x\|_\infty)+2n\log\log(\|x\|_\infty))} \\ &\quad= \|x\|_\infty^{-n}\left(\log\|x\|_\infty\right)^{-2n}. \end{aligned}$$

for large $\|x\|_\infty$ (and $|x_i|$ are not too far from $\|x\|_\infty$ for all i), where $\ell_\delta(t) = t + 2\log t$. Hence $f_{\mathrm{Im}}(x)$ has a power law decay. Proposition 1 gives the lower bound on the expected codeword length for generating $X \sim f$,

$$D\left(f_X \,\|\, f_{\mathrm{Im}}\right) \approx -h(X) + n\,\mathsf{E}\left[\ell_\delta(\log\|X\|_\infty)\right],$$

Comparing this to Theorem 1, we see that the upper bound is close to the lower bound when $\|X\|_\infty$ is the dominant term.

Note that Proposition 1 continues to hold even when Alice and Bob are allowed to share unlimited common randomness (denoted by the random variable Q). Suppose the prefix-free

codeword set when $Q = q$ is $C_q \subseteq \{0,1\}^*$. Upon receiving $m \in C$, Bob generates $X|\{Q = q, M = m\} \sim \tilde{p}_{q,m}$. The implicit distribution is

$$p_{\mathrm{Im}}(dx) = \int_{\mathcal{Q}} \Big(\sum_{m' \in C_q} 2^{-L(m')} \Big)^{-1} \sum_{m \in C_q} 2^{-L(m)} \tilde{p}_{q,m}(dx) p_Q(dq).$$

Proposition 1 still holds due to the convexity of relative entropy. Comparing this lower bound to the average length of the rejection sampling scheme in [9] (which requires common randomness), which achieves

$$\mathsf{E}_p(L(M)) \le D(p\|p^*) + 2\log(D(p\|p^*)+1) + O(1)$$

for some p^*. Hence, the lower bound is quite tight when unlimited common randomness is allowed.

APPENDIX

A. Proof of Lemma 1

The lemma is trivial when $n = 1$ since A can only be an interval. Hence we assume $n \ge 2$.

From the definition of erosion entropy,

$$\begin{aligned}
&h_{\ominus[0,1]^n}(A) \\
&= \int_{-\infty}^{\infty} \left(\mathbf{1}\{t \ge 0\} - \frac{\mathrm{V}_n\left(A \ominus 2^{-t}[0,1]^n\right)}{\mathrm{V}_n(A)} \right) dt \\
&= \int_{-\infty}^{\infty} t \cdot d\left(\frac{\mathrm{V}_n\left(A \ominus 2^{-t}[0,1]^n\right)}{\mathrm{V}_n(A)} \right) \\
&\overset{\text{(i)}}{=} \frac{1}{\mathrm{V}_n(A)} \int_0^{\infty} (\log s) \cdot d\mathrm{V}_n\left(A \ominus [0,s]^n\right) \\
&= \frac{1}{\mathrm{V}_n(A)} \int_0^{\infty} (\log s) \cdot d\Bigg(\sum_{i=1}^{n} \Big(\mathrm{V}_n\left(A \ominus \{0\}^{n-i} \times [0,s]^i\right) \\
&\quad - \mathrm{V}_n\left(A \ominus \{0\}^{n-i+1} \times [0,s]^{i-1}\right) \Big) \Bigg) \\
&= \frac{1}{\mathrm{V}_n(A)} \int_0^{\infty} (-\log s) \left(\sum_{i=1}^{n} \mathrm{V}_{n-1}\mathrm{P}_{\setminus i}\left(A \ominus [0,s]^n\right) \right) ds
\end{aligned}$$

where (i) is by substituting $s = 2^{-t}$, and $\mathrm{P}_{\setminus i}(A) = \{(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n) : x \in A\} \subseteq \mathbb{R}^{n-1}$. Let

$$q(s) = \frac{\sum_{i=1}^{n} \mathrm{V}_{n-1}\mathrm{P}_{\setminus i}\left(A \ominus [0,s]^n\right)}{\mathrm{V}_n(A)}.$$

Then we have $\int_0^{\infty} q(s)ds = 1$, $\int_0^{\infty} (-\log s) q(s) ds = h_{\ominus[0,1]^n}(A)$. Also

$$\begin{aligned}
&\mathsf{E}\left[\|X\|_{\infty}\right] \\
&= \frac{1}{\mathrm{V}_n(A)} \int_A \|x\|_{\infty} dx \\
&= \frac{-1}{\mathrm{V}_n(A)} \int_0^{\infty} \frac{d}{ds} \left(\int_{A \ominus [0,s]^n} \|x\|_{\infty} dx \right) ds \\
&= \frac{-1}{\mathrm{V}_n(A)} \int_0^{\infty} \sum_{i=1}^{n} \\
&\quad \times \frac{\partial}{\partial s_i} \left(\int_{A \ominus [0,s_1] \times \cdots [0,s_n]} \|x\|_{\infty} dx \right) \Big|_{(s_1,\ldots,s_n)=(s,\ldots,s)} ds \\
&\ge \frac{-1}{\mathrm{V}_n(A)} \int_0^{\infty} \sum_{i=1}^{n} \\
&\quad \times \frac{\partial}{\partial s_i} \left(\int_{A \ominus [0,s_1] \times \cdots [0,s_n]} \left\| x_{[1:n]\setminus i} \right\|_{\infty} dx \right) \Big|_{(s_1,\ldots,s_n)=(s,\ldots,s)} ds \\
&= \frac{1}{\mathrm{V}_n(A)} \int_0^{\infty} \left(\sum_{i=1}^{n} \int_{\mathrm{P}_{\setminus i}(A \ominus [0,s]^n)} \left\| x_{[1:n]\setminus i} \right\|_{\infty} dx_{[1:n]\setminus i} \right) ds \\
&\overset{\text{(i)}}{\ge} \frac{1}{\mathrm{V}_n(A)} \int_0^{\infty} \left(\sum_{i=1}^{n} \frac{1}{8} \mathrm{V}_{n-1} \mathrm{P}_{\setminus i}^{1+1/(n-1)}\left(A \ominus [0,s]^n\right) \right) ds \\
&\ge \frac{1}{8\mathrm{V}_n(A)} \int_0^{\infty} n^{\frac{-1}{n-1}} \left(\sum_{i=1}^{n} \mathrm{V}_{n-1}\mathrm{P}_{\setminus i}\left(A \ominus [0,s]^n\right) \right)^{\frac{n}{n-1}} ds \\
&= \frac{1}{8} n^{-1/(n-1)} \mathrm{V}_n^{1/(n-1)}(A) \int_0^{\infty} q^{n/(n-1)}(s) ds,
\end{aligned}$$

where (i) is by (4) in the proof of Theorem 1. Let

$$\tilde{q}(s) = \begin{cases} \left(\beta(\log e)^{n-1}\Gamma(n)\right)^{-1} (-\log(s/\beta))^{n-1} & \text{if } s \le \beta \\ 0 & \text{if } s > \beta, \end{cases}$$

where $\beta = e^n 2^{-h}$, $h = h_{\ominus[0,1]^n}(A)$. Then $\int_0^{\infty} \tilde{q}(s)ds = 1$, $\int_0^{\infty} (-\log s)\tilde{q}(s)ds = h$. Hence,

$$\begin{aligned}
&\int_0^{\infty} q^{\frac{n}{n-1}}(s)ds \\
&\ge \int_0^{\beta} q^{\frac{n}{n-1}}(s)ds \\
&= \int_0^{\beta} \tilde{q}(s)^{\frac{n}{n-1}} \left(q(s)/\tilde{q}(s)\right)^{\frac{n}{n-1}} ds \\
&\overset{\text{(i)}}{\ge} \left(\int_0^{\beta} \tilde{q}(s)^{\frac{n}{n-1}} ds \right) \left(\left(\int_0^{\beta} \tilde{q}(s)^{\frac{n}{n-1}} ds \right)^{-1} \right. \\
&\quad \left. \times \int_0^{\beta} \tilde{q}(s)^{\frac{n}{n-1}} \left(q(s)/\tilde{q}(s)\right) ds \right)^{\frac{n}{n-1}} \\
&= \left(\int_0^{\beta} \tilde{q}(s)^{\frac{n}{n-1}} ds \right)^{\frac{-1}{n-1}} \left(\int_0^{\beta} \tilde{q}(s)^{1/(n-1)} q(s) ds \right)^{\frac{n}{n-1}} \\
&= \left(n\left(\beta\Gamma(n)\right)^{\frac{-1}{n-1}} \right)^{\frac{-1}{n-1}} \\
&\quad \cdot \left((\log e)^{-1} \left(\beta\Gamma(n)\right)^{\frac{-1}{n-1}} \int_0^{\beta} (-\log s + \log\beta) q(s) ds \right)^{\frac{n}{n-1}} \\
&\ge \left(n\left(\beta\Gamma(n)\right)^{\frac{-1}{n-1}} \right)^{\frac{-1}{n-1}} \\
&\quad \cdot \left((\log e)^{-1} \left(\beta\Gamma(n)\right)^{\frac{-1}{n-1}} \int_0^{\infty} (-\log s + \log\beta) q(s) ds \right)^{\frac{n}{n-1}} \\
&= \left(n\left(\beta\Gamma(n)\right)^{\frac{-1}{n-1}} \right)^{\frac{-1}{n-1}} \\
&\quad \cdot \left((\log e)^{-1} \left(\beta\Gamma(n)\right)^{\frac{-1}{n-1}} (h + \log\beta) \right)^{\frac{n}{n-1}} \\
&= n\left(\beta\Gamma(n)\right)^{\frac{-1}{n-1}},
\end{aligned}$$

where (i) is by weighted power mean inequality. As a result,

$$\begin{aligned}
&\mathsf{E}\left[\|X\|_\infty\right] \\
&\geq \frac{1}{8}n^{-1/(n-1)}\mathrm{V}_n^{1/(n-1)}(A)\int_0^\infty q^{n/(n-1)}(s)ds \\
&\geq \frac{1}{8}n^{(n-2)/(n-1)}\mathrm{V}_n^{1/(n-1)}(A)\left(e^n 2^{-h}\Gamma(n)\right)^{-1/(n-1)}, \\
h &\leq (n-1)\left(\log \mathsf{E}\left[\|X\|_\infty\right]-\log\left(\frac{1}{8}n^{(n-2)/(n-1)}\mathrm{V}_n^{1/(n-1)}(A)\right)\right) \\
&\quad + \log\Gamma(n) + n\log e \\
&= (n-1)\log \mathsf{E}\left[\|X\|_\infty\right] - \log \mathrm{V}_n(A) + 3(n-1) \\
&\quad - (n-2)\log n + \log\Gamma(n) + n\log e \\
&\leq (n-1)\log \mathsf{E}\left[\|X\|_\infty\right] - \log \mathrm{V}_n(A) + 3(n-1) \\
&\quad - (n-2)\log n + (n\log n - (n-1)\log e) + n\log e \\
&= (n-1)\log \mathsf{E}\left[\|X\|_\infty\right] - \log \mathrm{V}_n(A) + 3(n-1) \\
&\quad + 2\log n + \log e \\
&\leq (n-1)\log \mathsf{E}\left[\|X\|_\infty\right] - \log \mathrm{V}_n(A) + 4n.
\end{aligned}$$

B. Proof of the Claim on Measure Zero Boundary in Theorem 2

We will prove that if f is a pdf which is continuous almost everywhere, then $L_z^+(f)$ has a boundary of measure zero for almost all z. Assume the contrary that there exist an uncountable $G \subseteq [0,\infty)$ such that $\mathrm{V}_n(\partial L_z^+(f)) > 0$ for all $z \in G$ (note that $\partial L_z^+(f)$ is a Borel set and thus measurable). Then we show that there exists $z_1 \neq z_2 \in G$ such that $\mathrm{V}_n(\partial L_{z_1}^+(f) \cap \partial L_{z_2}^+(f)) > 0$ (which follows from σ-finiteness, though we include a proof here for completeness). To show the claim, note that for any $z \in G$, there exists a hypercube $[0,1]^n + v$, $v \in \mathbb{Z}^n$ such that $\partial L_z^+(f) \cap ([0,1]^n + v)$ has nonzero measure. Hence there exists a hypercube $[0,1]^n + v$ such that $\partial L_z^+(f) \cap ([0,1]^n + v)$ has nonzero measure for an uncountable set of z's. Since an uncountable collection of positive numbers must contain a finite subcollection with sum greater than 1, we can select $z_1, \ldots, z_m$ such that $\sum_i \mathrm{V}_n(\partial L_{z_i}^+(f) \cap ([0,1]^n + v)) > 1$, and hence there exists two of these sets with an intersection of nonzero measure.

Now we have $z_1 < z_2 \in G$ such that $\mathrm{V}_n(\partial L_{z_1}^+(f) \cap \partial L_{z_2}^+(f)) > 0$. Assume there exists x in the intersection at which f is continuous, since $x \in \partial L_{z_2}^+(f)$, there exist sequence $y_i \to x$ with $f(y_i) \geq z_2$, and hence $f(x) \geq z_2$. Also since $x \in \partial L_{z_1}^+(f) \subseteq \mathrm{cl}\{y : f(y) < z_1\}$, there exist sequence $\tilde{y}_i \to x$ with $f(\tilde{y}_i) < z_1$, and hence $f(x) \leq z_1$, leading to a contradiction. Therefore f is discontinuous in $\partial L_{z_1}^+(f) \cap \partial L_{z_2}^+(f)$, contradicting the assumption that f is continuous almost everywhere. Therefore $L_z^+(f)$ has a boundary of measure zero for almost all z.

ACKNOWLEDGEMENT

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

REFERENCES

[1] V. I. Levenshtein, "On the redundancy and delay of decodable coding of natural numbers," *Problems Cybern.*, vol. 20, pp. 173–179, 1968.

[2] P. Elias, "Universal codeword sets and representations of the integers," *IEEE Trans. Inf. Theory*, vol. 21, no. 2, pp. 194–203, Mar. 1975.

[3] C. H. Bennett, P. W. Shor, J. A. Smolin, and A. V. Thapliyal, "Entanglement-assisted capacity of a quantum channel and the reverse Shannon theorem," *IEEE Trans. Inf. Theory*, vol. 48, no. 10, pp. 2637–2655, Oct. 2002.

[4] A. Winter. (2002). "Compression of sources of probability distributions and density operators." [Online]. Available: https://arxiv.org/abs/quant-ph/0208131

[5] P. Cuff, "Distributed channel synthesis," *IEEE Trans. Inf. Theory*, vol. 59, no. 11, pp. 7071–7096, Nov. 2013.

[6] C. H. Bennett, I. Devetak, A. W. Harrow, P. W. Shor, and A. Winter, "The quantum reverse Shannon theorem and resource tradeoffs for simulating quantum channels," *IEEE Trans. Inf. Theory*, vol. 60, no. 3, pp. 2926–2959, May 2014.

[7] M. Steiner, "Towards quantifying non-local information transfer: Finite-bit non-locality," *Phys. Lett. A*, vol. 270, no. 5, pp. 239–244, Jun. 2000.

[8] S. Massar, D. Bacon, N. J. Cerf, and R. Cleve, "Classical simulation of quantum entanglement without local hidden variables," *Phys. Rev. A*, vol. 63, no. 5, p. 052305, Apr. 2001.

[9] 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.

[10] C. T. Li and A. El Gamal, "A universal coding scheme for remote generation of continuous random variables," in *Proc. IEEE Inf. Theory Workshop*, Sep. 2016, pp. 384–388.

[11] C. T. Li and A. El Gamal. (2016). "Distributed simulation of continuous random variables." [Online]. Available: http://arxiv.org/abs/1601.05875

[12] T. Maudlin, "Bell's inequality, information transmission, and prism models," in *Proc. Biennial Meeting Philosophy Sci. Assoc. (PSA)*, 1992, pp. 404–417.

[13] G. Brassard, R. Cleve, and A. Tapp, "Cost of exactly simulating quantum entanglement with classical communication," *Phys. Rev. Lett.*, vol. 83, no. 9, p. 1874, 1999.

[14] J. S. Bell, "On the Einstein-Podolsky-Rosen paradox," *Physics*, vol. 1, no. 3, pp. 195–200, 1964.

[15] M. Feldmann, "New loophole for the Einstein-Podolsky-Rosen paradox," *Found. Phys. Lett.*, vol. 8, no. 1, pp. 41–53, Feb. 1995.

[16] J. Nash, "Non-cooperative games," *Ann. Math.*, vol. 54, no. 2, pp. 286–295, Sep. 1951.

[17] D. J. C. MacKay, *Information Theory, Inference and Learning Algorithms*. Cambridge, U.K.: Cambridge Univ. Press, 2003.

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. and Ph.D. degree in electrical engineering from Stanford University in 2014 and 2018 respectively. He is currently a postdoctoral scholar at the Department of Electrical Engineering and Computer Sciences, University of California, Berkeley. His research interests include generation of random variables, one-shot schemes in information theory, wireless communications and information-theoretic secrecy.

Abbas El Gamal (S'71–M'73–SM'83–F'00) is 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. From 2012-2017, he was the Fortinet Founders Chair of the Department of Electrical Engineering. 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.