Exact Channel Simulation under Exponential Cost

Spencer Hill, Fady Alajaji, and Tamás Linder

Abstract—**In this work, we generalize results in exact channel simulation to an exponential communication cost, specifically to Campbell's average codeword length $L(t)$ of order t [1], and to Rényi's entropy. In exact channel simulation, given a source of shared randomness, a sender wishes to communicate a message to the receiver so they can generate a sample from the conditional distribution $\mathrm{P}_{Y|X}$. We lower bound the Campbell cost of channel simulation using any sampling algorithm, showing that it grows approximately as $I_{t+1}(X;Y)$, where I_α is an appropriately defined α-mutual information of order α. Using the Poisson functional representation of Li and El Gamal [2], we prove an upper bound on $L(t)$ whose leading α-mutual information term has order within ϵ of the lower bound. Our results reduce to the bounds of Harsha et al. [3] as $t \to 0$. We also provide numerical examples for the additive white Gaussian noise channel, demonstrating that the upper and lower bounds are typically within 5-10 bits of each other.**

Index Terms—**Channel simulation, Rényi entropy, Rényi divergence, α-mutual information, Poisson functional representation, common randomness, variable-length codes, exponential cost.**

I. Introduction

We consider the problem of exact channel simulation under an exponential communication cost. As visualized in Fig. 1, given the communication channel $\mathrm{P}_{Y|X}$ and a source of randomness U with distribution P_U shared by both encoder and decoder, upon input $x \sim \mathrm{P}_X$ the sender communicates a binary message word M such that the receiver can generate a sample $Y \sim \mathrm{P}_{Y|X}(\,\cdot \mid x)$. This problem is called *channel simulation* because it uses a digital (noiseless) communication channel to simulate the behaviour of a noisy communication channel. We use the qualifier *exact* to emphasize that the sample produced must have precise distribution $\mathrm{P}_{Y|X}(\,\cdot \mid x)$, as opposed to *approximate* channel simulation, where Y is allowed to follow $\mathrm{P}_{Y|X}(\,\cdot \mid x)$ only approximately [4].

As studied in [2], [3], [5], among others, the key question is how to design a communication protocol such that the expected length of M, $\mathbb{E}[|M|]$, is minimized. Naively communicating a sample from $\mathrm{P}_{Y|X}(\,\cdot \mid x)$ using lossless coding is often impossible, as in principle, the sample can have a continuous distribution. Instead, protocols typically choose the shared randomness to be a sequence $\{U_i\}_{i\geq 1}$ independent and identically distributed (i.i.d.) according to the marginal P_Y and transmit an index $K \in \mathbb{N}$ such that the Kth sample in the shared randomness has exact distribution $\mathrm{P}_{Y|X}(\,\cdot \mid x)$, i.e., $U_K \sim \mathrm{P}_{Y|X}(\,\cdot \mid x)$. Then, with $\mathcal{C} : \mathbb{N} \to \{0,1\}^*$ denoting a uniquely decodable binary variable-length code and $M = \mathcal{C}(K)$ a binary message with length $|M|$, the goal is to select K and the code $\mathcal{C}$ such that $\mathbb{E}[|M|]$ is minimized. Channel simulation, also called reverse channel coding [6] or relative entropy coding [7], has recently gained significant theoretical and practical attention due to its promise as an alternative for quantization in deep learning-based compression systems [7]. There has been a corresponding research effort to understand the theoretical limits of channel simulation and its related problems [8], [9].

Generalizing, one may also be concerned about costs other than the expected message length. Motivated by Campbell's lossless source coding theorem [1], we consider one-shot exact channel simulation under an exponential storage cost. Such a cost is appropriate for applications where buffer overflow can occur [10]–[13], and therefore, the cost of long codewords is especially significant [14]–[16]. Here for a given $t > 0$, we aim to minimize the average codeword length of order t, also called the normalized cumulant generating function of codeword lengths [14]–[16], given by

$$L(t) = \frac{1}{t}\log\Big(\mathbb{E}[2^{t|M|}]\Big). \tag{1}$$

As $t \to 0$ in (1), by L'Hôpital's rule, we recover the average codeword length, $\mathbb{E}[|M|]$. In this work, we provide Rényi generalizations of prior results to the Campbell cost $L(t)$. In Section II, we formally define the problem of exact channel simulation, review some of the relevant literature, and outline our contributions. In Section III, we lower bound $L(t)$ for exact channel simulation using any sampling algorithm. In Section IV, we use the Poisson functional representation to prove an upper bound on $L(t)$ whose leading Rényi divergence term has order within ϵ of the lower bound. We also describe an operational procedure for encoding the index using a universal code for integers and prove an upper bound on $L(t)$ that reduces to the bound of Harsha et al. [3] as $t \to 0$. In Section V, we give numerical examples in the case of the additive white Gaussian noise (AWGN) channel. These examples show that qualitatively, the lower and upper bounds match in shape and are within 5-10 bits of each other for most $0 < \alpha < 1$, even when $L(t)$ is very large.

Notation: For P, Q probability distributions, we write $P \ll Q$ to indicate that P is absolutely continuous with respect to Q, in which case $\mathrm{d}P/\mathrm{d}Q$ denotes the Radon-Nikodym derivative. For random variables X and Y, we write $X \perp Y$ to indicate that they are independent. We let $\{0,1\}^*$ denote the collection of all finite-length binary words, i.e., $\{0,1\}^* = \{0, 1, 00, 01, 10, 11, 000, \ldots\}$, and for a message $M \in \{0,1\}^*$ we write $|M|$ to denote its length. We write $\mathcal{N}(\mu, \sigma^2)$ for the normal distribution with mean μ and variance σ^2. Throughout, $\log$ is in base 2, all information measures are

The authors are with the Mathematics and Statistics Department at Queen's University, Kingston, Ontario, Canada. This work was supported by the Natural Sciences and Engineering Research Council of Canada. Email: {spencer.hill, fa, tamas.linder}@queensu.ca.

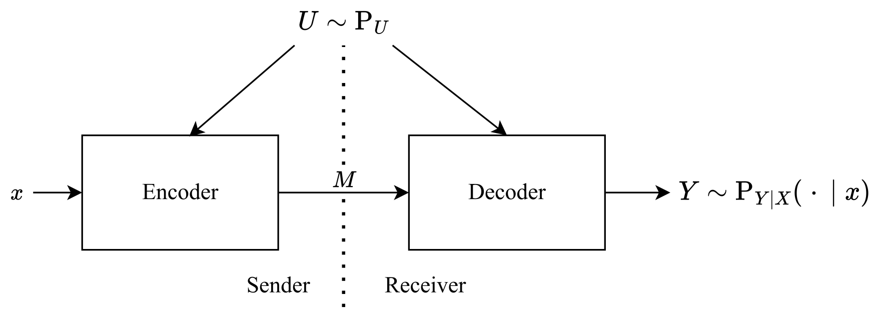


Fig. 1: Block diagram of the problem of exact channel simulation.

in bits, and $\ln$ denotes the natural logarithm. $\mathbb{N}$ denotes the natural numbers $\mathbb{N} = \{1, 2, \ldots\}$ and Γ is the gamma function.

II. Preliminaries

A. Rényi Information Measures

Let X be a discrete random variable with alphabet $\mathcal{X}$ and distribution P_X. For $\alpha \in (0,1) \cup (1,\infty)$, we define the Rényi entropy of order α of X as [17]

$$H_\alpha(X) = \frac{1}{1-\alpha} \log\left(\sum_{x\in\mathcal{X}} \mathrm{P}_X(x)^\alpha\right). \tag{2}$$

$H_\alpha(X)$ is nonincreasing in α, and if $H_\alpha(X) < \infty$ for some $0 < \alpha < 1$, then $\lim_{\alpha\nearrow 1} H_\alpha(X) = H(X)$, recovering the Shannon entropy. In [1], Campbell provided an operational meaning to the Rényi entropy by connecting it with the average code length of order t, $L(t)$.

Proposition 1 ([1]). *Suppose the discrete random variable X is encoded using a uniquely decodable binary code $\mathcal{C} : \mathcal{X} \to \{0,1\}^*$ such that each codeword has length $n_x := |\mathcal{C}(x)|$, for $x \in \mathcal{X}$. Then the average code length of order $t > 0$ given by*

$$L(t) = \frac{1}{t} \log\left(\sum_{x\in\mathcal{X}} \mathrm{P}_X(x) 2^{tn_x}\right), \tag{3}$$

satisfies

$$H_\alpha(X) \le L(t), \tag{4}$$

where $\alpha = \frac{1}{t+1}$. Moreover, there exists a uniquely decodable binary code $\mathcal{C}$ such that

$$L(t) < H_\alpha(X) + 1. \tag{5}$$

We note that while Campbell's original proof was for X with finite alphabet, it can easily be extended to the case of a countably infinite alphabet. For simplicity, we will often refer to $L(t)$ as just the Campbell cost. $L(t)$ is a nondecreasing function of t, and is strictly increasing in t unless all the lengths n_x are equal. As a result, if X takes countably infinite values, then $L(t) \to \infty$ as $t \to \infty$. If X is finite-valued, then $L(t) \to \max_{x\in\mathcal{X}} n_x$ as $t \to \infty$.

Rényi also defined the divergence of order α between probability distributions P and Q, $D_\alpha(P||Q)$, as

$$D_\alpha(P||Q) = \frac{1}{\alpha - 1} \log\left(\mathbb{E}_{X\sim Q}\left[\left(\frac{\mathrm{d}P}{\mathrm{d}Q}(X)\right)^\alpha\right]\right),$$

where $\alpha \in (0,1)\cup(1,\infty)$ and it is assumed that $P \ll Q$. Note that $\lim_{\alpha\nearrow 1} D_\alpha(P||Q) = D(P||Q)$, i.e., the KL divergence is recovered when $\alpha \nearrow 1$ [18].

B. Problem Definition

Let X and Y be random variables whose alphabets are the respective Polish spaces $\mathcal{X}$ and $\mathcal{Y}$. Suppose that X and Y have joint distribution P_{XY}, marginal distributions P_X and P_Y, and conditional distribution $\mathrm{P}_{Y|X}(\,\cdot\,|x)$ such that $\mathrm{P}_{Y|X}(\,\cdot\,|x) \ll \mathrm{P}_Y$ for P_X-almost all x. The sender is provided with $x \in \mathcal{X}$ generated according to P_X and transmits a uniquely decodable binary message M to the receiver so that they can generate a single sample $y \in \mathcal{Y}$ with exact distribution $\mathrm{P}_{Y|X}(\,\cdot\,|x)$. We assume that both sender and receiver have access to an unlimited source of common randomness, $U \sim \mathrm{P}_U$. We wish to minimize the expected Campbell cost,

$$\mathbb{E}_X[L(t \mid X)], \tag{6}$$

where $L(t \mid X = x)$ is the expected Campbell cost of M given input $x \in \mathcal{X}$. The minimization is over all common randomness U and encoder-decoder pairs such that the message M is generated by a prefix-free code and the simulated channel is exact. Observe that as $t \to 0$ (equivalently $\alpha \to 1$) in (6) we recover $\mathbb{E}_X[\mathbb{E}[|M| \mid X]] = \mathbb{E}[|M|]$. We henceforth refer to this problem setup as *exact channel simulation*.

An important subclass of channel simulation algorithms is sampling algorithms. In sampling algorithms, the shared randomness is the i.i.d. sequence $\{U_i\}_{i\ge 1}$ distributed according to the marginal P_Y and the message is an index $K \in \mathbb{N}$ such that the Kth sample has the required distribution, i.e., $U_K \sim \mathrm{P}_{Y|X}(\,\cdot \mid x)$. Most state-of-the-art exact channel simulation algorithms capable of simulating general channels are sampling algorithms, including (among others) the Poisson functional representation [2], rejection sampling [3], [5], and greedy Poisson rejection sampling [19]. Notable examples of channel simulation algorithms which are not sampling algorithms include subtractive dithered quantization [20] and polar codes for channel simulation [21], both of which simulate a limited family of channels (additive uniform noise and binary output channels, respectively).

C. Prior Work

The problem of channel simulation was first studied by Wyner [22], who considered the asymptotic version (communicating n samples at a time) without common randomness but allowing error between the simulated and target distributions. Winter [23] showed that, by allowing a shared source of unlimited randomness, it is possible to simulate the channel with small error at an asymptotic cost equal to the mutual information. Bennett et al. [24] showed that the same result holds even for exact channel simulation. More recently, Cuff [25], Bennett et al. [26], and Yu and Tan [27] studied the tradeoff between communication rate and amount of common randomness in the asymptotic regime. Sriramu and Wagner [9] and Flamich et al. [28] examined the optimal rate of exact channel synthesis in the asymptotic case. There has also been recent work, motivated by applications in deep learning-based compression, on developing algorithms which are computationally efficient, e.g. [6], [19], [29]–[32]. We refer the reader to [33] for a survey of channel simulation methods and results.

In our problem of one-shot exact channel simulation, a simple application of the data processing inequality shows the lower bound $\mathbb{E}[|M|] \geq I(X;Y)$ [3]. The first achievability bound is due to Harsha et al. [3], who described a communication protocol for discrete X and Y for which

$$\mathbb{E}[|M|] \leq I(X;Y) + (1+\epsilon)\log(I(X;Y)+1) + c_\epsilon,$$

with c_ϵ a constant depending on ϵ. Their proof effectively amounts to showing that for each $x \in \mathcal{X}$,

$$\mathbb{E}[\log K \mid X = x] \leq D(\mathrm{P}_{Y|X}(\,\cdot\mid x) \,||\, \mathrm{P}_Y) + 2\log e,$$

then encoding K using a universal code and taking expectation over $X \sim \mathrm{P}_X$. As noted in [2], by using instead a power law code one can bound $\mathbb{E}[|M|] \leq I(X;Y) + \log(I(X;Y)+1) + 7.78$ for discrete X and Y. The Poisson functional representation, first proposed by Li and El Gamal in [2], has since been used to prove a tighter upper bound on $\mathbb{E}[|M|]$ [34].

Definition 1 (Poisson functional representation [35])**.** *Let $\mathcal{U}$ be a Polish space and P, Q two probability distributions on $\mathcal{U}$ with $P \ll Q$. Let $\{T_i\}_{i\geq 1}$ be a rate-one Poisson process and let $\{U_i\}_{i\geq 1}$ be an i.i.d. sequence distributed according to Q and independent of $\{T_i\}_{i\geq 1}$. Define*

$$K \coloneqq \operatorname*{arg\,min}_{i\geq 1} \frac{T_i}{\frac{\mathrm{d}P}{\mathrm{d}Q}(U_i)}. \tag{7}$$

It is shown in [35, Appendix A] that under this definition, $U_K \sim P$. For $P = \mathrm{P}_{Y|X}(\,\cdot\mid x)$, $Q = \mathrm{P}_Y$, and K generated according to Definition 1 and encoded using the optimal prefix-free code for the Zipf distribution, [34] showed that

$$\mathbb{E}[|M|] \leq I(X;Y) + \log(I(X;Y)+2) + 3.$$

Moreover, [2] showed that there exist distributions for which $\mathbb{E}[|M|] \geq I(X;Y) + \log(I(X;Y)+1) - 1$, i.e., the $\log$ term is necessary in general.

D. Contributions

In this work, we study the Campbell cost, $L(t)$, of one-shot exact channel simulation algorithms. Although channel simulation and its associated minimum communication cost are well-studied problems, to the best of our knowledge, this natural extension (from linear cost to an exponential cost in the message lengths) has not been investigated in the literature. Our main contributions can be summarized as follows:

- We lower bound $\mathbb{E}_X[L(t \mid X)]$ for channel simulation using any sampling algorithm, showing that it grows approximately as $\mathbb{E}_X[D_{\frac{1}{\alpha}}(\mathrm{P}_{Y|X}(\,\cdot\mid X) \,||\, \mathrm{P}_Y)]$, where $\alpha = \frac{1}{1+t}$.
- The lower bound motivates the definition of an α-mutual information, $I_\alpha(X;Y) \coloneqq \mathbb{E}_X[D_\alpha(\mathrm{P}_{Y|X}(\,\cdot\mid x) \,||\, \mathrm{P}_Y)]$, for which we prove several properties one expects of any α-mutual information.
- We use the Poisson functional representation to prove the upper bound $\mathbb{E}_X[L(t \mid X)] \leq (1+\epsilon) I_{\frac{1+\epsilon(1-\alpha)}{\alpha}}(X;Y) + c_1(\alpha,\epsilon)$, for any $0 < \alpha < 1$ and $\epsilon > 0$, where $c_1(\alpha,\epsilon)$ is a constant depending on α and ϵ. Using numerical examples, we demonstrate that this upper bound is within 5-10 bits of the lower bound, even for distributions where $L(t)$ is large.
- We describe an operational procedure for encoding the communicated index using a universal code that gives $L(t) \leq I_{\frac{2-\alpha}{\alpha}}(X;Y) + (1+\epsilon)\log(I(X;Y)+1) + c_2(\epsilon)$, for any $2/3 < \alpha < 1$ and $0 < \epsilon \leq \frac{3\alpha-2}{2-2\alpha}$, where $c_2(\epsilon)$ is a constant. We show that this upper bound reduces to the bound of Harsha et al. [3] as $\alpha \to 1$.

III. Lower Bound on the Campbell Cost

We present our main results as bounds on the Campbell cost $L(t)$, but using Proposition 1 can write them as bounds on $H_\alpha(K)$ within one bit (this will be formalized in Corollary 1). In this section, we will prove a lower bound on the Campbell cost of exact channel simulation of any sampling algorithm. Note that for all $0 < \alpha < 1$ and with $t = \frac{1-\alpha}{\alpha}$, one has the trivial lower bound

$$L(t) \geq H_\alpha(K) \geq H(K) \geq I(X;Y),$$

where the final inequality follows as in [3]. However, such a lower bound loses the dependency on α, which is key to understanding $H_\alpha(K)$ and $L(t)$.

Theorem 1. *Let X and Y be random variables with conditional distribution $P_{Y|X}$ satisfying the assumptions of the exact channel simulation problem (see Section II-B). Suppose that we are using any sampling algorithm, i.e., the shared source of randomness is the i.i.d. sequence $\{U_i\}_{i\geq 1}$ distributed according to P_Y and the message is an index $K \in \mathbb{N}$ such that $U_K \sim P_{Y|X}(\,\cdot\mid x)$. Then, for any $0 < \alpha < 1$ with $t = \frac{1-\alpha}{\alpha}$,* *any uniquely decodable binary encoding of K has expected Campbell cost*

$$\begin{aligned}&\mathbb{E}_X[L(t \mid X)]\\&\geq \mathbb{E}_X[D_{\frac{1}{\alpha}}(P_{Y|X}(Y \mid X) \,||\, P_Y)] + \frac{\alpha}{1-\alpha}\log(\alpha) - 1. \quad (8)\end{aligned}$$

Proof. We fix $x \in \mathcal{X}$ and show that $L(t \mid X = x) \geq D_{1/\alpha}(\mathrm{P}_{Y|X}(\,\cdot \mid x) \,||\, \mathrm{P}_Y) + \frac{\alpha}{1-\alpha}\log(\alpha) - 1$; taking expectation with respect to $X \sim \mathrm{P}_X$ will give the result. We use the following lemma, adapted from [36], whose proof can be found in [37].

Lemma 1. *Let $P \ll Q$ be probability distributions on a Polish space $\mathcal{U}$ and suppose that both sender and receiver have access to an i.i.d. sequence $\{U_i\}_{i\geq 1}$ drawn from Q. Let K be the output of any sampling algorithm, i.e., $U_K \sim P$. Then, for any $0 < \alpha < 1$ and bijection $g : \mathbb{N} \to \mathbb{N}$,*

$$\mathbb{E}[(g(K))^\alpha] \geq \frac{1}{1+\alpha} 2^{\alpha D_{\alpha+1}(P||Q)}. \quad (9)$$

Let $g : \mathbb{N} \to \mathbb{N}$ be the bijection which orders $\tilde{K} := g(K)$ so that its probability distribution is nonincreasing. The optimal one-to-one encoding of $\tilde{K}$ (under the Campbell cost) will have codeword lengths $\lfloor \log(k+1) \rfloor$. The cost of this encoding lower bounds the Campbell cost of any uniquely decodable encoding of $\tilde{K}$ (equivalently, lower bounds the Campbell cost of encoding K) as

$$\begin{aligned}L(t) &\geq \frac{1}{t}\log\left(\sum_{k=1}^{\infty} \tilde{p}(k) 2^{t\lfloor \log(k+1) \rfloor}\right)\\&\geq \frac{1}{t}\log\left(\mathbb{E}[\tilde{K}^t]\right) - 1\\&\geq \frac{1}{t}\log\left(\frac{1}{1+t} 2^{tD_{t+1}(\mathrm{P}_{Y|X}(\,\cdot\,|x)||\mathrm{P}_Y)}\right) - 1. \quad (10)\end{aligned}$$

where (10) follows from Lemma 1 with $P = \mathrm{P}_{Y|X}(\,\cdot \mid x)$ and $Q = \mathrm{P}_Y$. Substituting $\alpha = \frac{1}{1+t}$ gives the desired inequality. □

As $\alpha \to 1$ in (8), we recover the bound $\mathbb{E}[|M|] \geq \mathbb{E}_X[D(\mathrm{P}_{Y|X}(Y \mid X) \,||\, \mathrm{P}_Y)] - \frac{1}{\ln(2)} - 1 = I(X;Y) - \frac{1}{\ln(2)} - 1$, the same lower bound as in [3] with a small penalty term of $\frac{-1}{\ln(2)} - 1$. To the best of our knowledge, the leading term in (8), $\mathbb{E}_X[D_{\frac{1}{\alpha}}(\mathrm{P}_{Y|X}(Y \mid X) \,||\, \mathrm{P}_Y)]$, is not a previously defined α-mutual information [38], [39]. It is similar to the Augustin-Csiszár α-mutual information, which is defined as the minimum $\min_Q \mathbb{E}_X[D_\alpha(\mathrm{P}_{Y|X}(\,\cdot \mid X) \,||\, Q)]$ [39], but can be more accurately thought of as a Rényi generalization of $I(X;Y) = \mathbb{E}_X[D(\mathrm{P}_{Y|X}(\,\cdot \mid X) \,||\, \mathrm{P}_Y)]$. We define

$$I_\alpha(X;Y) := \mathbb{E}_X[D_\alpha(\mathrm{P}_{Y|X}(Y \mid X) \,||\, \mathrm{P}_Y)]$$

and will refer to the lower bound in (8) as $\mathrm{LB}^\alpha := I_{\frac{1}{\alpha}}(X;Y) + \frac{\alpha}{1-\alpha}\log(\alpha) - 1$. The following proposition collects some important properties of $I_\alpha(X;Y)$, justifying its definition as an α-mutual information.

Proposition 2. *Let X and Y be general random variables with $P_{Y|X}(\,\cdot \mid x) \ll P_Y$ for P_X-almost every x, and let $\alpha \in (0,1) \cup (1,\infty)$. Then, $I_\alpha(X;Y)$ satisfies the following properties.*

- *(I)* $\lim_{\alpha\to 1} I_\alpha(X;Y) = I(X;Y)$. ***(Recovers MI)***
- *(II)* $I_\alpha(X;Y) \geq 0$. ***(Nonnegativity)***
- *(III)* $I_\alpha(X;Y) = 0 \Leftrightarrow X \perp Y$. ***(Positivity)***
- *(IV) For* $\alpha_1 < \alpha_2$, $I_{\alpha_1}(X;Y) \leq I_{\alpha_2}(X;Y)$. ***(Monotonicity)***
- *(V) For X discrete,* $I_\alpha(X;X) = H(X)$. ***(Self-information)***
- *(VI) For a random variable Z forming the Markov chain $X \to Y \to Z$ with $P_{Z|X}(\,\cdot \mid x) \ll P_Z$ for P_X-almost every x,* $I_\alpha(X;Z) \leq I_\alpha(X;Y)$. ***(Data processing inequality)***
- *(VII) If the pairs of random variables $\{(X_i, Y_i)\}_{1\leq i\leq n}$ are independent, then* $I_\alpha(X_1,\ldots,X_n;Y_1,\ldots Y_n) = \sum_{i=1}^n I_\alpha(X_i;Y_i)$. ***(Additivity)***

Proof. (I), (II), (IV), (VI), (VII): Follow as immediate consequences of properties of the Rényi divergence (recovery of KL divergence, nonnegativity, monotonicity, data processing inequality, and additivity [18]), which apply pointwise for each $x \in \mathcal{X}$ and therefore also in expectation.

(III): If $X \perp Y$, $\mathrm{P}_{Y|X}(\,\cdot \mid x) = \mathrm{P}_Y$ for every $x \in \mathcal{X}$, hence $D_\alpha(\mathrm{P}_{Y|X}(\,\cdot \mid x) \,||\, \mathrm{P}_Y) = D_\alpha(\mathrm{P}_Y \,||\, \mathrm{P}_Y) = 0$ and $I_\alpha(X,Y) = 0$. Conversely, if $I_\alpha(X;Y) = 0$ then, by the equality condition of the Rényi divergence being 0, $\mathrm{P}_{Y|X}(\,\cdot \mid x) = \mathrm{P}_Y$ for P_X-almost $x \in \mathcal{X}$ and $X \perp Y$.

(V): Let X be discrete, then $\mathrm{P}_{X|X}(\,\cdot \mid x) = \mathbb{1}_x$ (where $\mathbb{1}_{\{\cdot\}}$ denotes the indicator function) and $D_\alpha(\mathbb{1}_x||\mathrm{P}_X) = \frac{1}{\alpha-1}\log\left(\sum_{x'\in\mathcal{X}} \mathbb{1}_x(x')^\alpha\, \mathrm{P}_X(x')^{1-\alpha}\right) = -\log \mathrm{P}_X(x)$. Thus, $I_\alpha(X;X) = \mathbb{E}_X[-\log \mathrm{P}_X] = H(X)$. □

Like most α-mutual informations, $I_\alpha(X;Y)$ is not symmetric and does not admit a clean chain rule. Theorems 1 and 2 give $I_\alpha(X;Y)$ an operational meaning for the channel simulation problem.

IV. Upper Bounds using the Poisson Functional Representation

Theorem 1 tells us that the minimum Campbell cost of channel simulation using any sampling algorithm grows approximately as $I_{\frac{1}{\alpha}}(X;Y)$. The natural question is whether any channel simulation algorithms exist that achieve this lower bound. Suppose that we could bound $\mathbb{E}_X[L(t \mid X)] \leq I_{\frac{1}{\alpha}}(X;Y) + C$ for some constant C and all random variables X, Y. Then, taking $\alpha \to 1$ (equivalently $t \to 0$) would yield $\mathbb{E}[|M|] \leq I(X;Y) + C$, a bound shown to be invalid by an explicit counter-example [2]. Therefore, it is reasonable to conjecture that a tight and general upper bound, valid for all X and Y, does not exist for the Campbell cost, either. Instead, we have the following upper bound on $L(t)$ using the Poisson functional representation.

Theorem 2. *Let X and Y be random variables satisfying the assumptions of the exact channel simulation problem and let K be the output of the Poisson functional representation given that the shared randomness between sender and receiver is the i.i.d. sequence $\{U_i\}_{i\geq 1}$ distributed according to P_Y. Then, for* *any $0<\alpha<1$ and $\epsilon>0$, there exists a uniquely decodable encoding of K such that*

$$\mathbb{E}_X[L(t \mid X)] \leq (1+\epsilon) I_{\frac{1+\epsilon(1-\alpha)}{\alpha}}(X;Y) + c_1(\alpha,\epsilon), \quad (11)$$

with $t=\frac{1-\alpha}{\alpha}$ and $c_1(\alpha,\epsilon)$ a constant term defined as

$$c_1(\alpha,\epsilon) \coloneqq \begin{cases} (1+\epsilon)\log e + 1 + \log\left(1+\frac{1}{\epsilon}\right), & \frac{1}{2}<\alpha<1 \text{ and } 0<\epsilon<\frac{2\alpha-1}{1-\alpha} \\ \frac{\alpha}{1-\alpha}\log\left(\Gamma\left(\frac{1+\epsilon(1-\alpha)}{\alpha}\right)\right) +4+3\epsilon-\frac{2\alpha}{1-\alpha}+\log\left(1+\frac{1}{\epsilon}\right), & 0<\alpha<\frac{1}{2} \text{ or } \epsilon \geq \frac{2\alpha-1}{1-\alpha} \end{cases} \quad (12)$$

We will refer to the upper bound in (11) as $\mathrm{UB}_1^\alpha \coloneqq (1+\epsilon) I_{\frac{1+\epsilon(1-\alpha)}{\alpha}}(X;Y) + c_1(\alpha,\epsilon)$. Before proving Theorem 2, we state two lemmas which will be used to upper bound the moments of K.

Lemma 2. *Let K be the output of the Poisson functional representation given two probability distributions P,Q on the Polish space $\mathcal{U}$ with $P \ll Q$. Then,*

$$\mathbb{E}[\log K] \leq D(P||Q)+1, \quad (13)$$

and for $0<\alpha<1$,

$$\mathbb{E}[K^\alpha] \leq 2^{\alpha D_{\alpha+1}(P||Q)} + \alpha. \quad (14)$$

Equation (13) was proved in [34], and the upper bound in (14) follows from [35, Prop. 4] after substituting $j=1$. We note that for $0<\alpha<1$ and $g(k)=k$, the Poisson functional representation in (14) almost achieves the lower bound on the α-moment of any sampling algorithm in Lemma 1.

Lemma 3. *Let $X \sim$ Geo(p) be a geometric random variable with parameter $0<p<1$. Then, for any $r \geq 1$,*

$$\mathbb{E}[X^r] \leq 2^{r-1}\left(\frac{\Gamma(r+1)}{p^r}+1\right).$$

The proof of Lemma 3 can be found in [37].

Proof of Theorem 2. As in the proof of Theorem 1, we condition on $X=x$ and will show that $L(t \mid X=x) \leq (1+\epsilon) D_{\frac{1+\epsilon(1-\alpha)}{\alpha}}(\mathrm{P}_{Y|X}(\,\cdot \mid x) \,||\, \mathrm{P}_Y) + c_1(\alpha,\epsilon)$; taking expectation with respect to $X \sim \mathrm{P}_X$ will show the theorem. We will prove the desired inequality in two cases. In both cases, we will relate the Campbell cost $L(t)$ to the expected moments of K^r, with $r=\frac{1-\alpha}{\alpha}(1+\epsilon)$. For $r<1$ (equivalently $\epsilon<\frac{2\alpha-1}{1-\alpha}$), we will bound the moments of K using (14) in Lemma 2, while in the case $r \geq 1$ (equivalently $\epsilon \geq \frac{2\alpha-1}{1-\alpha}$) we will use the bound on the moments of a geometric random variable in Lemma 3.

Fix $1/2<\alpha<1$ and $0<\epsilon<\frac{2\alpha-1}{1-\alpha}$. Then, by the Kraft inequality, there exists a uniquely decodable code $\mathcal{C}:\mathbb{N}\to\{0,1\}^*$ with codeword lengths $|\mathcal{C}(k)| \leq (1+\epsilon)\log k + 1 + \log\left(1+\frac{1}{\epsilon}\right)$. Let $t=\frac{1-\alpha}{\alpha}$. Then, $0<t<1$ and $0<\epsilon<\frac{2\alpha-1}{1-\alpha}=\frac{1-t}{t}$ imply that $t(1+\epsilon)<1$, so we can apply the upper bound on $\mathbb{E}[K^\alpha]$ in (14) to get

$$\begin{aligned} &L(t \mid X=x) && (15)\\ &\leq \frac{1}{t}\log\left(\sum_{k=1}^{\infty} p(k) 2^{t(1+\epsilon)\log k + t + t\log\left(1+\frac{1}{\epsilon}\right)}\right) \\ &\leq \frac{1}{t}\log\left(2^{(1+\epsilon)tD_{(1+\epsilon)t+1}(\mathrm{P}_{Y|X}(\,\cdot\,|x)||\mathrm{P}_Y)} + (1+\epsilon)t\right) \\ &\quad + 1 + \log\left(1+\frac{1}{\epsilon}\right) && (16)\\ &\leq (1+\epsilon) D_{(1+\epsilon)t+1}(\mathrm{P}_{Y|X}(\,\cdot \mid x) \,||\, \mathrm{P}_Y) + (1+\epsilon)\log e \\ &\quad + 1 + \log\left(1+\frac{1}{\epsilon}\right). && (17) \end{aligned}$$

Here, (16) follows from (14) and (17) from $\log(x+a) \leq \log(x) + a\log e$ for all $x \geq 1$ and $a>0$. Noting that $(1+\epsilon)t+1=\frac{1+\epsilon(1-\alpha)}{\alpha}$ gives the desired bound.

Suppose now that $0<\alpha<1$ and $\epsilon \geq \frac{2\alpha-1}{1-\alpha}$, so that with $t=\frac{1-\alpha}{\alpha}$, $(1+\epsilon)t \geq 1$. From [35, Eq. 29] (after substituting $j=1$), for any $u \in \mathcal{U}$,

$$K \mid \{U_K=u\} \sim \mathrm{Geo}(\beta(u)), \quad (18)$$

for $\beta(u)=\mathbb{E}_{U\sim \mathrm{P}_Y}\left[\max\left\{\frac{\mathrm{dP}_{Y|X}(\,\cdot\,|x)}{\mathrm{dP}_Y}(u), \frac{\mathrm{dP}_{Y|X}(\,\cdot\,|x)}{\mathrm{dP}_Y}(U)\right\}\right]^{-1}$. By Lemma 3, for $r=(1+\epsilon)t \geq 1$,

$$\begin{aligned} &\mathbb{E}[K^r \mid U_K=u] \\ &\leq 2^{r-1}\left(\frac{\Gamma(r+1)}{\beta(u)^r}+1\right) \\ &\leq \Gamma(r+1)2^{2r-2}\left(\left(\frac{\mathrm{dP}_{Y|X}(\,\cdot \mid x)}{\mathrm{dP}_Y}(u)\right)^r+1\right)+2^{r-1}. \end{aligned}$$

Taking expectation with respect to $U_K \sim \mathrm{P}_{Y|X}(\,\cdot \mid x)$, we get that

$$\mathbb{E}[K^r] \leq \Gamma(r+1)2^{2r-2}\left(2^{rD_{r+1}(\mathrm{P}_{Y|X}(\,\cdot\,|x)||\mathrm{P}_Y)}+1\right)+2^{r-1}. \quad (19)$$

Again encoding K using a uniquely decodable code $\mathcal{C}$ with codeword lengths $|\mathcal{C}(k)| \leq (1+\epsilon)\log k + 1 + \log\left(1+\frac{1}{\epsilon}\right)$ and writing $r=(1+\epsilon)t$ and $c(\epsilon)=1+\log\left(1+\frac{1}{\epsilon}\right)$, we have that

$$\begin{aligned} &L(t \mid X=x) \\ &\leq \frac{1}{t}\log(\mathbb{E}[K^r]) + c(\epsilon) \\ &\leq \frac{1}{t}\log\Big(\Gamma(r+1)2^{2r-2}\left(2^{rD_{r+1}(\mathrm{P}_{Y|X}(\,\cdot\,|x)||\mathrm{P}_Y)}+1\right) \\ &\quad + 2^{r-1}\Big) + c(\epsilon) && (20)\\ &\leq \frac{1}{t}\log\left(\Gamma(r+1)2^{2r-2}\left(2^{rD_{r+1}(\mathrm{P}_{Y|X}(\,\cdot\,|x)||\mathrm{P}_Y)}+1\right)\right) \\ &\quad + \frac{1}{t}\log\left(2^{r-1}\right) + c(\epsilon) && (21)\\ &\leq \frac{1}{t}\log\left(2^{rD_{r+1}(\mathrm{P}_{Y|X}(\,\cdot\,|x)||\mathrm{P}_Y)}\right) + \frac{1}{t}\log(\Gamma(r+1)) \\ &\quad + \frac{2r-2}{t} + \frac{1}{t} + \frac{r-1}{t} + c(\epsilon). && (22) \end{aligned}$$

Here, (20) follows by (19), while (21) and (22) both follow from the inequality $\log(x+1) \leq \log(x) + 1$ for all $x \geq 1$. Substituting $r = (1+\epsilon)t$ and $t = \frac{1-\alpha}{\alpha}$ gives the statement of the theorem. $\square$

The α-mutual information term in UB_1^α has order $\frac{1+\epsilon(1-\alpha)}{\alpha}$, which as $\epsilon \to 0$ goes to $\frac{1}{\alpha}$, the order of the α-mutual term in LB^α. We see that the Poisson functional representation almost achieves the lower bound on the Campbell cost among all sampling-based channel simulation schemes. However, $c_1(\alpha,\epsilon) \to \infty$ as $\epsilon \to 0$, thus UB_1^α is not tight. Numerical examples in Section V demonstrate that, after minimizing over $\epsilon > 0$, UB_1^α is typically within 5-10 bits of LB^α. As $\alpha \to 1$, we obtain the upper bound on the expected message length $\mathbb{E}[|M|] \leq (1+\epsilon)I(X;Y) + (1+\epsilon)\log e + 1 + \log\big(1+\frac{1}{\epsilon}\big)$, for any $\epsilon > 0$. This bound is generally looser than the bound from the strong functional representation lemma [2], but better than the upper bound $\mathbb{E}[|M|] \leq I(X;Y) + (1+\epsilon)\log(I(X;Y)+1) + c_\epsilon$ by Harsha et al. [3]. However, UB_1^α could be operationally difficult to achieve, as it requires constructing a uniquely decodable encoding of the positive integers with codeword lengths $|\mathcal{C}(k)| \approx (1+\epsilon)\log k$ for all k. The bounds of [2] and [3] have the operational advantage of encoding K using a power law code and universal code. In our problem, encoding K using a universal code gives the following upper bound, which is strictly worse than UB_1^α.

Theorem 3. *Let X and Y be random variables satisfying the assumptions of the exact channel simulation problem and let K be the output of the Poisson functional representation given that the shared randomness between sender and receiver is the i.i.d. sequence $\{U_i\}_{i\geq 1}$ distributed according to P_Y. Then, for any $2/3 < \alpha < 1$ and $0 < \epsilon \leq \frac{3\alpha-2}{2-2\alpha}$, encoding K using a universal code $\mathcal{C}$ with codeword lengths $|\mathcal{C}(k)| = \log k + (1+\epsilon)\log\log(k+1) + O(1)$ gives*

$$\begin{aligned}\mathbb{E}_X[L(t \mid X)] \leq\ & I_{\frac{2-\alpha}{\alpha}}(X;Y) \\ & + (1+\epsilon)\log(I(X;Y)+1) + c_2(\epsilon),\end{aligned}$$

with $t = \frac{1-\alpha}{\alpha}$ and $c_2(\epsilon) = 3 + \epsilon + \log\Big(\frac{\ln(2)}{\epsilon} + \frac{3}{2}\Big)$.

Proof. Fix $X = x$ and let $2/3 < \alpha < 1$ and $0 < \epsilon \leq \frac{3\alpha-2}{2-2\alpha}$. We will bound $L(t \mid X = x)$ using a universal coding of the natural numbers and then take expectation with respect to $X \sim \mathrm{P}_X$ to get the desired result. In particular, using the prefix-free encoding $\mathcal{C}$ described in [40, Ex. 1.11.16] (see also [3]) we can set $|\mathcal{C}(k)| \leq \log k + (1+\epsilon)\log\log(k+1) + 1 + \log\Big(\frac{\ln(2)}{\epsilon} + \frac{3}{2}\Big)$ for any $\epsilon > 0$. Since $t = \frac{1-\alpha}{\alpha}$, $0 < t < 1/2$. Then,

$$\begin{aligned}
& L(t \mid X = x) \\
& \leq \frac{1}{t}\log\Big(\mathbb{E}\left[K^t \log(K+1)^{(1+\epsilon)t}\right]\Big) \\
& \qquad + 1 + \log\left(\frac{\ln(2)}{\epsilon} + \frac{3}{2}\right) \\
& \leq \frac{1}{t}\log\left(\sqrt{\mathbb{E}[K^{2t}]\mathbb{E}[\log(K+1)^{2(1+\epsilon)t}]}\right) \\
& \qquad + 1 + \log\left(\frac{\ln(2)}{\epsilon} + \frac{3}{2}\right) \qquad (23) \\
& = \frac{1}{2t}\left(\log\big(\mathbb{E}[K^{2t}]\big) + \log\Big(\mathbb{E}[\log(K+1)^{2(1+\epsilon)t}]\Big)\right) \\
& \qquad + 1 + \log\left(\frac{\ln(2)}{\epsilon} + \frac{3}{2}\right). \qquad (24)
\end{aligned}$$

Here, (23) follows from the Cauchy-Schwartz inequality. Using Jensen's inequality (which applies as $2(1+\epsilon)t \leq 1$) we can bound

$$\begin{aligned}
& \frac{1}{2t}\log\left(\mathbb{E}[\log(K+1)^{2(1+\epsilon)t}]\right) \\
& \leq \frac{1}{2t}\log\left(\mathbb{E}[\log(K+1)]^{2(1+\epsilon)t}\right) \\
& = (1+\epsilon)\log\left(\mathbb{E}[\log(K+1)]\right) \\
& \leq (1+\epsilon)\log\left(\mathbb{E}[\log K] + 1\right) \qquad (25) \\
& \leq (1+\epsilon)\log\big(D(\mathrm{P}_{Y|X}(\,\cdot \mid x) \,\|\, \mathrm{P}_Y) + 1 + 1\big) \qquad (26) \\
& \leq (1+\epsilon)\log\big(D(\mathrm{P}_{Y|X}(\,\cdot \mid x) \,\|\, \mathrm{P}_Y) + 1\big) + 1 + \epsilon. \qquad (27)
\end{aligned}$$

Here, (25) and (27) both follow from the inequality $\log(x+1) \leq \log(x) + 1$ for all $x \geq 1$ and (26) follows from the upper bound on $\mathbb{E}\left[\log K\right]$ in Lemma 2, (13). In the first term of (24), because $0 < t < 1/2$ by assumption, we can apply the upper bound on $\mathbb{E}\left[K^\alpha\right]$ in Lemma 2 to get that

$$\begin{aligned}
\frac{1}{2t}\log\big(\mathbb{E}[K^{2t}]\big) & \leq \frac{1}{2t}\log\Big(2^{2tD_{2t+1}(P\|Q)} + 2t\Big) \\
& \leq D_{2t+1}(\mathrm{P}_{Y|X}(\,\cdot \mid x) \,\|\, \mathrm{P}_Y) + 1, \qquad (28)
\end{aligned}$$

where (28) holds again by $\log(x+a) \leq \log(x) + a$. Combining (27) and (28), and noting $2t+1 = \frac{2-\alpha}{\alpha}$, we obtain the bound

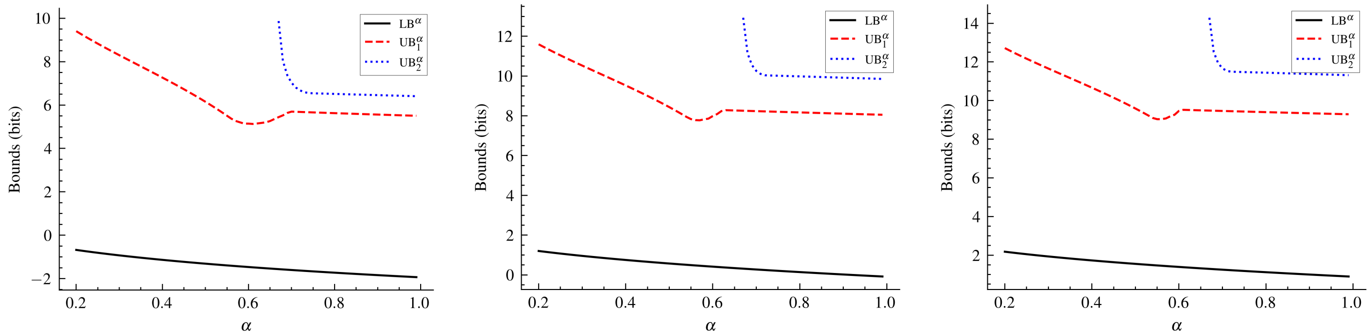
$$\begin{aligned}
& L(t \mid X = x) \\
& \leq D_{\frac{2-\alpha}{\alpha}}(\mathrm{P}_{Y|X}(\,\cdot \mid x) \,\|\, \mathrm{P}_Y) \\
& \quad + (1+\epsilon)\log\big(D(\mathrm{P}_{Y|X}(\,\cdot \mid x) \,\|\, \mathrm{P}_Y) + 1\big) + c_2(\epsilon).
\end{aligned}$$

Taking expectation with respect to $X \sim \mathrm{P}_X$ gives

$$\begin{aligned}
& \mathbb{E}_X[L(t \mid X)] \\
& \leq I_{\frac{2-\alpha}{\alpha}}(X;Y) \\
& \quad + (1+\epsilon)\mathbb{E}_X[\log\big(D(\mathrm{P}_{Y|X}(\,\cdot \mid x) \,\|\, \mathrm{P}_Y) + 1\big)] + c_2(\epsilon) \\
& \leq I_{\frac{2-\alpha}{\alpha}}(X;Y) + (1+\epsilon)\log(I(X;Y)+1) + c_2(\epsilon), \qquad (29)
\end{aligned}$$

where (29) follows from applying Jensen's inequality to $\log$. $\square$

We will refer to the upper bound in Theorem 3 as $\mathrm{UB}_2^\alpha := I_{\frac{2-\alpha}{\alpha}}(X;Y) + (1+\epsilon)\log(I(X;Y)+1) + c_2(\epsilon)$. As $\alpha \to 1$, UB_1^α reduces to $\mathbb{E}[|M|] \leq I(X;Y) + (1+\epsilon)\log(I(X;Y)+1) + c_2(\epsilon)$ for any $\epsilon > 0$, the same upper bound as in [3]. Note that UB_2^α is valid only for $2/3 < \alpha < 1$, whereas UB_1^α is valid for all $0 < \alpha < 1$. Moreover, for all $2/3 < \alpha < 1$, $\mathrm{UB}_1^\alpha < \mathrm{UB}_2^\alpha$. To see why, observe that for $\epsilon < 1$, $\frac{1+\epsilon(1-\alpha)}{\alpha} < \frac{2-\alpha}{\alpha}$, i.e., the order of the leading α-mutual information is strictly less in UB_1^α. Since $I_\alpha(X;Y)$ is nondecreasing in α (see Proposition 2), one has $I_{\frac{1+\epsilon(1-\alpha)}{\alpha}}(X;Y) \leq I_{\frac{2-\alpha}{\alpha}}(X;Y)$. As $c_1(\alpha,\epsilon) < c_2(\epsilon)$ for all $2/3 < \alpha < 1$, there exists $\epsilon > 0$ such that UB_1^α is better even



(a) $\sigma^2 = 1$ and $\rho^2 = 1$ (b) $\sigma^2 = 25$ and $\rho^2 = 1$ (c) $\sigma^2 = 100$ and $\rho^2 = 1$

Fig. 2: Comparison of the bounds on the Campbell cost for the AWGN channel, for $0.2 < \alpha < 1$.

for small $I_\alpha(X;Y)$. Numerical results, shown in Section V, illustrate this conclusion.

While Theorems 1 to 3 are bounds on the Campbell cost, we can use Proposition 1 to obtain bounds on $H_\alpha(K)$.

Corollary 1. *Let X and Y be random variables satisfying the assumptions of the exact channel simulation problem. Suppose that a sampling algorithm is used, i.e., the shared randomness is the i.i.d. sequence $\{U_i\}_{i\geq 1}$ distributed according to P_Y and upon input $x \in \mathcal{X}$ an index K is outputted such that $U_K \sim P_{Y|X}(\,\cdot \mid x)$. Then, for any $0 < \alpha < 1$, $\mathbb{E}_X[H_\alpha(K \mid X)] > LB^\alpha - 1$, with LB^α given in Theorem 1. Moreover for K the output of the Poisson functional representation, for any $0 < \alpha < 1$ and $\epsilon > 0$, we have that $\mathbb{E}_X[H_\alpha(K \mid X)] \leq UB_1^\alpha$, with UB_1^α given in Theorem 2.*

Proof. The upper bound is immediate after substitution in (4) in Proposition 1. The lower bound follows from (5) in Proposition 1, specifically that for any sampling algorithm outputting K, there exists an encoding of K such that $L(t) < H_\alpha(K)+1$. Then, the lower bound $L(t) \geq \mathrm{LB}^\alpha$ of Theorem 1 applies to this sampling algorithm and encoding, meaning $H_\alpha(K) > \mathrm{LB}^\alpha - 1$. □

V. Numerical Examples

In this section, we give numerical examples comparing LB^α, UB_1^α, and UB_2^α from Theorems 1 to 3. We consider the AWGN channel with input $X \sim \mathcal{N}(0,\sigma^2)$ and output $Y = X + Z$, $Z \sim \mathcal{N}(0,\rho^2)$. Here, $\sigma^2, \rho^2 \in \mathbb{R}_{\geq 0}$ are the respective variances of the input and additive noise. One can compute that $\mathrm{P}_{Y|X}(\,\cdot \mid x) = \mathcal{N}(x,\rho^2)$ and the marginal distribution is $\mathrm{P}_Y = \mathcal{N}(0,\sigma^2+\rho^2)$. Trying to losslessly communicate a sample from a normal distribution is not possible; however, by allowing a access to a shared source of randomness, we can communicate the sample with finite cost. By the formula for the Rényi divergence between two Gaussians [41, p. 45] (see also [42]),

$$\begin{aligned} &D_\alpha(\mathrm{P}_{Y|X}(\,\cdot \mid x) \,||\, \mathrm{P}_Y) \\ &= \frac{1}{2}\ln\left(1+\frac{\sigma^2}{\rho^2}\right) + \frac{\ln\left(\frac{\sigma^2+\rho^2}{\rho^2+\alpha\sigma^2}\right)}{2(\alpha-1)} + \frac{1}{2}\frac{\alpha x^2}{\rho^2+\alpha\sigma^2}, \end{aligned}$$

and we can take expectation with respect to $X \sim \mathrm{P}_X$ to get that

$$I_\alpha(X;Y) = \frac{1}{2}\ln\left(1+\frac{\sigma^2}{\rho^2}\right) + \frac{\ln\left(\frac{\sigma^2+\rho^2}{\rho^2+\alpha\sigma^2}\right)}{2(\alpha-1)} + \frac{1}{2}\frac{\alpha\sigma^2}{\rho^2+\alpha\sigma^2}.$$

One can confirm that as $\alpha \to 1$, we recover the standard expression $I(X;Y) = \frac{1}{2}\log\left(1+\frac{\sigma^2}{\rho^2}\right)$. Fig. 2 compares the bounds for fixed $\rho^2 = 1$ and three increasing choices of σ^2; note that the true minimum $\mathbb{E}_X[L(t \mid X)]$ of any sampling algorithm is guaranteed to lie between UB_1^α and LB^α. We have only shown the bounds for $0.2 < \alpha < 1$ for display reasons, as UB_1^α goes to ∞ as $\alpha \to 0$. In both upper bounds, for each value of α, the bound is optimized over ϵ to find the minimum value. As discussed in Section IV, in all cases UB_1^α is tighter than UB_2^α for all $2/3 < \alpha < 1$.

VI. Conclusion

In this paper, we have generalized bounds on the expected communication cost of one-shot exact channel simulation to the Campbell cost and Rényi's entropy. Such bounds are useful in situations where one wishes to disproportionally penalize long codewords, such as applications with buffer overflow. There are several interesting directions of future work stemming from these results. Most obviously, it is an open question if one can prove the lower bound $\mathbb{E}_X[L(t \mid X)] \geq I_{1/\alpha}(X;Y)$ for any channel simulation algorithm, not just sampling schemes. We conjecture that such a bound is true, but it is not an immediate application of the data processing inequality like in the Shannon case. It is also interesting to consider whether one can tighten UB_1^α to be in the spirit of [2], specifically so that it reduces to $\mathbb{E}[|M|] \leq I(X;Y)+\log(I(X;Y)+1)+O(1)$ as $\alpha \to 1$. More generally, the strong functional representation lemma has been applied to several problems outside of one-shot exact channel simulation, most notably one-shot variable-length lossy source coding and multiple description coding [2]. It would be interesting to see if the strong functional representation lemma can play a role in information-theoretic coding problems using the Campbell cost or Rényi entropy.

References

[1] L. L. Campbell, "A coding theorem and Rényi's entropy," *Information and Control*, vol. 8, no. 4, pp. 423–429, 1965.

[2] C. T. Li and A. E. Gamal, "Strong functional representation lemma and applications to coding theorems," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, 2017, pp. 589–593.

[3] P. Harsha, R. Jain, D. McAllester, and J. Radhakrishnan, "The communication complexity of correlation," *IEEE Transactions on Information Theory*, vol. 56, no. 1, pp. 438–449, 2010.

[4] G. Flamich and L. Wells, "Some notes on the sample complexity of approximate channel simulation," in *Proc. IEEE International Symposium on Information Theory Workshops (ISIT-W)*. IEEE, 2024, pp. 1–6.

[5] M. Braverman and A. Garg, "Public vs private coin in bounded-round information," in *Proc. International Colloquium on Automata, Languages, and Programming*, 2014, pp. 502–513.

[6] L. Theis and N. Y. Ahmed, "Algorithms for the communication of samples," in *Proc. International Conference on Machine Learning*, 2022, pp. 21 308–21 328.

[7] G. Flamich, M. Havasi, and J. M. Hernández-Lobato, "Compressing images by encoding their latent representations with relative entropy coding," *Advances in Neural Information Processing Systems*, vol. 33, pp. 16 131–16 141, 2020.

[8] D. Goc and G. Flamich, "On channel simulation with causal rejection samplers," in *Proc. IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2024, pp. 1682–1687.

[9] S. M. Sriramu and A. B. Wagner, "Optimal redundancy in exact channel synthesis," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, 2024, pp. 1913–1918.

[10] F. Jelinek, "Buffer overflow in variable length coding of fixed rate sources," *IEEE Transactions on Information Theory*, vol. 14, no. 3, pp. 490–501, 1968.

[11] P. Humblet, "Generalization of Huffman coding to minimize the probability of buffer overflow (corresp.)," *IEEE Transactions on Information Theory*, vol. 27, no. 2, pp. 230–232, 1981.

[12] A. C. Blumer and R. J. McEliece, "The Rényi redundancy of generalized Huffman codes," *IEEE Transactions on Information Theory*, vol. 34, no. 5, pp. 1242–1249, 1988.

[13] A. Somazzi, P. Ferragina, and D. Garlaschelli, "On nonlinear compression costs: when Shannon meets Rényi," *IEEE Access*, 2024.

[14] T. A. Courtade and S. Verdú, "Cumulant generating function of codeword lengths in optimal lossless compression," in *Proc. 2014 IEEE International Symposium on Information Theory*, 2014, pp. 2494–2498.

[15] ——, "Variable-length lossy compression and channel coding: Non-asymptotic converses via cumulant generating functions," in *Proc. 2014 IEEE International Symposium on Information Theory*, 2014, pp. 2499–2503.

[16] S. Saito and T. Matsushima, "Non-asymptotic bounds of cumulant generating function of codeword lengths in variable-length lossy compression," *IEEE Transactions on Information Theory*, vol. 69, no. 4, pp. 2113–2119, 2022.

[17] A. Rényi, "On measures of entropy and information," in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, volume 1: Contributions to the Theory of Statistics*, vol. 4. University of California Press, 1961, pp. 547–562.

[18] T. Van Erven and P. Harremos, "Rényi divergence and Kullback-Leibler divergence," *IEEE Transactions on Information Theory*, vol. 60, no. 7, pp. 3797–3820, 2014.

[19] G. Flamich, "Greedy Poisson rejection sampling," *Advances in Neural Information Processing Systems*, vol. 36, pp. 37 089–37 127, 2023.

[20] R. Zamir and M. Feder, "On universal quantization by randomized uniform/lattice quantizers," *IEEE Transactions on Information Theory*, vol. 38, no. 2, pp. 428–436, 2002.

[21] S. Sriramu, R. Barsz, E. Polito, and A. Wagner, "Fast channel simulation via error-correcting codes," *Advances in Neural Information Processing Systems*, vol. 37, pp. 107 932–107 959, 2024.

[22] A. Wyner, "The common information of two dependent random variables," *IEEE Transactions on Information Theory*, vol. 21, no. 2, pp. 163–179, 1975.

[23] A. Winter, "Compression of sources of probability distributions and density operators," *arXiv preprint quant-ph/0208131*, 2002.

[24] 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 transactions on Information Theory*, vol. 48, no. 10, pp. 2637–2655, 2002.

[25] P. Cuff, "Distributed channel synthesis," *IEEE Transactions on Information Theory*, vol. 59, no. 11, pp. 7071–7096, 2013.

[26] 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 Transactions on Information Theory*, vol. 60, no. 5, pp. 2926–2959, 2014.

[27] L. Yu and V. Y. Tan, "Exact channel synthesis," *IEEE Transactions on Information Theory*, vol. 66, no. 5, pp. 2799–2818, 2019.

[28] G. Flamich, S. M. Sriramu, and A. B. Wagner, "The redundancy of non-singular channel simulation," *arXiv preprint arXiv:2501.14053*, 2025.

[29] G. Flamich, S. Markou, and J. M. Hernández-Lobato, "Fast relative entropy coding with a* coding," in *Proc. International Conference on Machine Learning*, 2022, pp. 6548–6577.

[30] ——, "Faster relative entropy coding with greedy rejection coding," *Advances in Neural Information Processing Systems*, vol. 36, pp. 50 558–50 569, 2023.

[31] S. Sriramu, R. Barsz, E. Polito, and A. Wagner, "Fast channel simulation via error-correcting codes," *Advances in Neural Information Processing Systems*, vol. 37, pp. 107 932–107 959, 2024.

[32] G. Flamich and L. Theis, "Adaptive greedy rejection sampling," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, 2023, pp. 454–459.

[33] C. T. Li, *Channel Simulation: Theory and Applications to Lossy Compression and Differential Privacy*. Now Publishers, Inc., 2024, vol. 21, no. 6.

[34] ——, "Pointwise redundancy in one-shot lossy compression via Poisson functional representation," *arXiv preprint arXiv:2401.14805*, 2024.

[35] C. T. Li and V. Anantharam, "A unified framework for one-shot achievability via the Poisson matching lemma," *IEEE Transactions on Information Theory*, vol. 67, no. 5, pp. 2624–2651, 2021.

[36] J. Liu and S. Verdú, "Rejection sampling and noncausal sampling under moment constraints," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, 2018, pp. 1565–1569.

[37] S. Hill, F. Alajaji, and T. Linder, "Communication complexity of exact sampling under Rényi information," *arXiv preprint arXiv:2506.12219*, 2025.

[38] S. Verdú, "α-mutual information," in *Proc. Information Theory and Applications Workshop (ITA)*, 2015, pp. 1–6.

[39] I. Csiszár, "Generalized cutoff rates and Rényi's information measures," *IEEE Transactions on information theory*, vol. 41, no. 1, pp. 26–34, 2002.

[40] M. Li and P. Vitányi, *An Introduction to Kolmogorov Complexity and its Applications*. Springer, 2008, vol. 3.

[41] F. Liese and I. Vajda, *Convex Statistical Distances*. Teubner, 1987.

[42] M. Gil, F. Alajaji, and T. Linder, "Rényi divergence measures for commonly used univariate continuous distributions," *Information Sciences*, vol. 249, pp. 124–131, 2013.