Some Useful Integral Representations for Information-Theoretic Analyses

Neri Merhav Igal Sason

The Andrew and Erna Viterbi Faculty of Electrical Engineering
Technion – Israel Institute of Technology
Technion City, Haifa 3200003, Israel
E-mail: {merhav,sason}@ee.technion.ac.il

Abstract

This work is an extension of our earlier article, where a well–known integral representation of the logarithmic function was explored, and was accompanied with demonstrations of its usefulness in obtaining compact, easily–calculable, exact formulas for quantities that involve expectations of the logarithm of a positive random variable. Here, in the same spirit, we derive an exact integral representation (in one or two dimensions) of the moment of a non–negative random variable, or the sum of such independent random variables, where the moment order is a general positive real, not necessarily an integer. The proposed formula is applied to a variety of examples with an information–theoretic motivation, and it is shown how it facilitates their numerical evaluations. In particular, when applied to the calculation of a moment of the sum of a large number, n, of non–negative random variables, it is clear that integration over one or two dimensions, as suggested by our proposed integral representation, is significantly easier than the alternative of integrating over n dimensions, as needed in the direct calculation of the desired moment.

Index Terms: Logarithmic expectation, moment–generating function, differential Rényi entropy, jamming, estimation errors, concentration inequalities, multivariate Cauchy distributions, randomized guessing.

I. INTRODUCTION

In mathematical analyses associated with many problems in information theory and related fields, one is often faced with the need to compute expectations of logarithmic functions of composite random variables (see, e.g., [11], [13], [16], [17], [18], [22], [25], [33]), or moments of such random variables, whose order may be a general positive real, not even necessarily an integer (see, e.g., [1], [2], [3], [5], [6], [7], [8], [9], [10], [20], [28], [29], [31], [32]).

In the case of the logarithmic function, the common practice is either to resort approximate evaluations, provided by upper and lower bounds on the desired expression (for example, by using Jensen's inequality), or to approximate the calculations by using the Taylor series expansion of the function $\ln x$. More recently, it has become popular to use the replica trick (see, e.g., [19, Chapter 8]), which is a non-rigorous, but useful technique, borrowed from statistical physics.

In our earlier work [22], we have demonstrated how the following well–known integral representation of the logarithmic function,

$$\ln x = \int_0^\infty \left(e^{-u} - e^{-ux}\right) \frac{\mathrm{d}u}{u}, \quad x > 0, \tag{1}$$

can be useful in a variety of application areas in the field of information theory, including both source and channel coding, as well as other aspects of this field. To calculate the expectation, $\mathbb{E}\{\ln X\}$, where X is a positive random variable, the idea is simply to invoke the integral representation (1) and to commute the expectation and integration operators, i.e.,

$$\mathbb{E}\{\ln X\} = \int_0^\infty \left(e^{-u} - \mathbb{E}\{e^{-uX}\}\right) \frac{\mathrm{d}u}{u}, \tag{2}$$

thereby replacing the calculation of $\mathbb{E}\{\ln X\}$ by the calculation of the moment–generating function (MGF), $\mathbb{E}\{e^{-uX}\}$ for all $u \geq 0$, which is often a lot easier to express in closed form. Moreover, in frequently encountered situations where X is given by the sum of n independently identically distributed (i.i.d.) random variables, the MGF of X is given by the n–th power of the MGF of a single random variable in the sum that forms X. This reduces the dimension of the integration from n (in the original expression) to a single dimension of the integration over u. Interestingly, this integral representation has also been used in the statistical physics literature (see, e.g., [12], [19, p. 140], [30]), but not as much as the replica trick.

In this paper, we proceed in the same spirit as in [22], and we extend the scope to propose an integral representation of a general moment of a non–negative random variable, X, namely, the expectation, $\mathbb{E}\{X^\rho\}$ for a given $\rho > 0$. Obviously, when ρ is integer, this moment is simply given by the ρ–th order derivative of the MGF of X, calculated at the origin, as is very well known. However, the integral representation we propose, in this work, applies to any non–integer, positive ρ, and here too, it replaces the direct calculation of $\mathbb{E}\{X^\rho\}$ by integration of an expression that involves the MGF of X. We refer to this representation as an *extension* of (2), as the latter can be obtained as a special case of the formula for $\mathbb{E}\{X^\rho\}$, by invoking the identity

$$\mathbb{E}\{\ln X\} = \lim_{\rho \to 0} \frac{\mathbb{E}\{X^\rho\} - 1}{\rho}, \tag{3}$$

or alternatively, the identity,

$$\mathbb{E}\{\ln X\} = \lim_{\rho \to 0} \frac{\ln[\mathbb{E}\{X^\rho\}]}{\rho}. \tag{4}$$

While the proposed integral representation of $\mathbb{E}\{X^\rho\}$ can be readily obtained from [14, p. 363, Identity (3.434.1)] in the range $\rho \in (0, 1)$, the non–trivial extension we propose for $\rho > 1$ is new to the best of our knowledge.

As in [22], the proposed integral representation is applied to a variety of examples with an information–theoretic motivation, and it is shown how it facilitates the numerical evaluations. In particular, similarly as in the case of the logarithmic function, when applied to the calculation of a moment of the sum of a large number, n, of non–negative random variables, it is clear that integration over one or two dimensions, as suggested by our proposed integral representation, is significantly easier than the alternative of integrating over n dimensions, as needed in the direct calculation of the desired moment. Furthermore, single or double-dimensional integrals can be instantly and accurately calculated using built-in numerical integration procedures.

The outline of the remaining part of this paper is as follows. In Section II, we provide the mathematical background associated with the integral representation in general. In Section III, we demonstrate this integral representation in applications, including: moments of guesswork, moments of estimation errors, differential Rényi entropies of generalized multivariate Cauchy

distributions, and mutual information calculations of a certain model of a jammed channel. Each one of these examples occupies one subsection of Section III. The integral representations in this paper are not limited to the examples in Section III, and such representations can be proved useful in other information–theoretic problems (see, e.g., [22] and references therein).

II. Statistical Moments of Arbitrary Positive Orders

It is well known that any integer–order moment of a random variable X can be calculated from its MGF

$$M_X(u) := \mathbb{E}\big\{e^{uX}\big\}, \quad u \in \mathbb{R}, \tag{5}$$

by using its ρ–th order derivative, calculated at $u = 0$, i.e.,

$$\mathbb{E}\{X^\rho\} = M_X^{(\rho)}(0), \quad \rho \in \mathbb{N}. \tag{6}$$

Quite often, however, there is a theoretical and practical interest to calculate positive non-integral moments of non-negative random variables. We next obtain a closed–form integral expression of the ρ–th moment of a non-negative random variable X, as a functional of its MGF, for any positive real ρ. Before we proceed, it should be noted that for $\rho \in (0, 1)$, such an expression is available in handbooks of standard tables of integrals, for example, in [14, p. 363, Identity (3.434.1)]. The first innovation here, however, is in a non–trivial extension of this formula for all $\rho > 0$ as an expression that involves a one–dimensional integral. It should be noted that although the definition of a non–integer moment of a RV is also given by a one–dimensional integral (or a sum, depending on whether the RV is discrete or continuous), the utility of our formula is, e.g., in expressing the ρ-th moment of a sum of non–negative and independent random variables as a one–dimensional integral, instead of an n–dimensional integral which is obtained by the direct definition. This new formula serves as the basic building block in all our information-theoretic applications throughout this paper.

We first define the Beta and Gamma functions (see, e.g., [14, Section 8.3] and [23, Chapter 5]):

$$\Gamma(u) := \int_0^\infty t^{u-1} e^{-t} \, \mathrm{d}t, \quad u > 0, \tag{7}$$

$$B(u,v) := \int_0^1 t^{u-1}(1-t)^{v-1}\,\mathrm{d}t = \frac{\Gamma(u)\,\Gamma(v)}{\Gamma(u+v)}, \quad u,v>0. \tag{8}$$

Theorem 1: Let X be a non-negative random variable with an MGF $M_X(\cdot)$, and let $\rho > 0$ be a non-integer real. Then,

$$\mathbb{E}\{X^\rho\} = \frac{1}{1+\rho}\sum_{\ell=0}^{\lfloor\rho\rfloor}\frac{\alpha_\ell}{B(\ell+1,\rho+1-\ell)} + \frac{\rho\,\sin(\pi\rho)\,\Gamma(\rho)}{\pi}\int_0^\infty \frac{1}{u^{\rho+1}}\left(\sum_{j=0}^{\lfloor\rho\rfloor}\left\{\frac{(-1)^j\,\alpha_j}{j!}\,u^j\right\}e^{-u} - M_X(-u)\right)\mathrm{d}u, \tag{9}$$

where for all $j \in \{0,1,\ldots,\}$

$$\alpha_j := \mathbb{E}\big\{(X-1)^j\big\} \tag{10}$$

$$= \frac{1}{j+1}\sum_{\ell=0}^{j}\frac{(-1)^{j-\ell}\,M_X^{(\ell)}(0)}{B(\ell+1,j-\ell+1)}. \tag{11}$$

Proof: See Appendix A. ∎

Remark 1: The proof of (9) in Appendix A does not apply to $\rho \in \mathbb{N}$ (see (A.7), (A.8) etc., where the denominators vanish for $\rho \in \mathbb{N}$). In the latter case, by referring to the second term on the right–hand side of (9), we get $\sin(\pi\rho) = 0$ and also the integral diverges (specifically, for $\rho \in \mathbb{N}$, the integrand scales like $\frac{1}{u}$ for u that is sufficiently close to zero), yielding an expression of the type $0 \cdot \infty$. However, taking a limit in (9) where we let ρ tend to an integer, and applying L'Hôpital's rule can reproduce the well–known result in (6).

Corollary 1: For any $\rho \in (0,1)$,

$$\mathbb{E}\{X^\rho\} = 1 + \frac{\rho}{\Gamma(1-\rho)}\int_0^\infty \frac{e^{-u} - M_X(-u)}{u^{1+\rho}}\,\mathrm{d}u. \tag{12}$$

Proof: Eq. (12) is due to Theorem 1, and by using (A.20), (A.22) (see Appendix A) and $\alpha_0 := 1$, which give

$$\Gamma(\rho)\,\Gamma(1-\rho) = \frac{\pi}{\sin(\pi\rho)}, \tag{13}$$

$$\frac{1}{1+\rho}\,\frac{\alpha_0}{B(1,\rho+1)} = \frac{1}{1+\rho}\,\frac{\Gamma(\rho+2)}{\Gamma(\rho+1)} = 1. \tag{14}$$

∎

Remark 2: Corollary 1 also follows from [14, p. 363, Identity (3.434.1)] (see [22, Section 4]).

Corollary 2: [22] Let X be a positive random variable. Then,

$$\mathbb{E}\{\ln X\} = \int_0^\infty \frac{e^{-u} - M_X(-u)}{u}\,\mathrm{d}u. \tag{15}$$

A proof of (15) is presented in [22, Section 2], based on the integral representation of the logarithmic function in (1), and by interchanging the integration and the expectation. It can be alternatively proved by using Corollary 1, and the identity $\ln x = \lim\limits_{\rho\to 0}\frac{x^\rho - 1}{\rho}$ for $x > 0$. Identity (15) has many useful information–theoretic applications on its own right, as demonstrated in [22], and here we add even some more. The current work is an extension and further development of [22], whose main theme is in exploiting Theorem 1 and studying its information–theoretic applications, as well as some more applications of the logarithmic expectation.

III. Applications

In this section, we exemplify the usefulness of the integral representation of the ρ-th moment in Theorem 1 and the logarithmic expectation in several problem areas in information theory and statistics. These include analyses of randomized guessing, estimation errors, Rényi entropy of n-dimensional generalized Cauchy distributions, and finally, calculations of the mutual information for channels with a certain jammer model. To demonstrate the direct computability of the relevant quantities, we also present graphs of their numerical calculations.

A. Moments of Guesswork

Consider the problem of guessing the realization of a random variable which takes on values in a finite alphabet, using a sequence of yes/no questions of the form "Is $X = x_1$?", "Is $X = x_2$?", etc., until a positive response is provided by a party that observes the actual realization of X. Given a distribution of X, a commonly used performance metric for this problem is the expected number of guesses or, more generally, the ρ-th moment of the number of guesses until X is guessed successfully. When it comes to guessing random vectors, say, of length n, minimizing the moments of the number of guesses by different (deterministic or randomized) guessing strategies has several applications and motivations in information theory, such as sequential decoding, guessing passwords, etc., and it is also strongly related to lossless source coding

(see, e.g., [1], [2], [3], [5], [6], [15], [21], [27], [28], [29], [31], [32]). In this vector case, the moments of the number of guesses behave as exponential functions of the vector dimension, n, at least asymptotically, as n grows without bound. For random vectors with i.i.d. components, the best achievable asymptotic exponent of the ρ-th guessing moment is expressed in [1] by using the Rényi entropy of X of order $\widetilde{\rho} := \frac{1}{1+\rho}$. Arikan assumed in [1] that the distribution of X is known, and analyzed the optimal deterministic guessing strategy, which orders the guesses according to non–increasing probabilities. Refinements of the exponential bounds in [1] with tight upper and lower bounds on the guessing moments for optimal deterministic guessing were recently derived in [28]. In the sequel, we refer to randomized guessing strategies, rather than deterministic strategies, and we aim to derive exact, calculable expressions for their associated guessing moments (as it is later explained in this subsection).

Let the random variable X take on values in a finite alphabet $\mathcal{X}$. Consider a random guessing strategy where the guesser sequentially submits a sequence of independently drawn random guesses according to a certain probability distribution, $\widetilde{P}(\cdot)$, defined on $\mathcal{X}$. Randomized guessing strategies have the advantage that they can be used by multiple asynchronous agents which submit their guesses concurrently (see [21] and [27]).

In this subsection, we consider the setting of randomized guessing, and obtain an exact representation of the guessing moment in the form of a one–dimensional integral. Let $x \in \mathcal{X}$ be any realization of X and let the guessing distribution, $\widetilde{P}$, be given. The random number, G, of independent guesses until success has a geometric distribution:

$$\Pr\{G = k|x\} = \big[1 - \widetilde{P}(x)\big]^{k-1} \widetilde{P}(x), \quad k \in \mathbb{N}, \tag{16}$$

and so, the corresponding MGF is equal to

$$\begin{aligned} M_G(u|x) &= \sum_{k=1}^{\infty} e^{ku} \Pr\{G = k|x\} \\ &= \frac{\widetilde{P}(x)}{e^{-u} - \big(1 - \widetilde{P}(x)\big)}, \quad u < \ln \frac{1}{1 - \widetilde{P}(x)}. \end{aligned} \tag{17}$$

In view of (9)–(11) and (17), for $x \in \mathcal{X}$ and non–integer $\rho > 0$,

$$\mathbb{E}\{G^{\rho}|x\} = \frac{1}{1+\rho} \sum_{\ell=0}^{\lfloor \rho \rfloor} \frac{\alpha_{\ell}}{B(\ell+1, \rho+1-\ell)} \tag{18}$$

$$+\frac{\rho\sin(\pi\rho)\,\Gamma(\rho)}{\pi}\int_0^\infty \frac{1}{u^{\rho+1}}\left(\sum_{j=0}^{\lfloor\rho\rfloor}\left\{\frac{(-1)^j\,\alpha_j}{j!}\;u^j\right\}e^{-u}-\frac{\widetilde{P}(x)}{e^u-\big(1-\widetilde{P}(x)\big)}\right)\mathrm{d}u,$$

with $\alpha_0 := 1$, and for $j \in \mathbb{N}$

$$\begin{aligned}\alpha_j &:= \mathbb{E}\big\{\big(G-1\big)^j|X=x\big\}\\ &=\sum_{k=1}^{\infty}(k-1)^j\,\big(1-\widetilde{P}(x)\big)^{k-1}\,\widetilde{P}(x)\\ &=\widetilde{P}(x)\,\mathrm{Li}_{-j}\big(1-\widetilde{P}(x)\big).\end{aligned}\tag{19}$$

In (19) $\mathrm{Li}_{-j}(\cdot)$ is a polylogarithm (see, e.g., [23, Section 25.12]), which is given by

$$\mathrm{Li}_{-j}(x)=\Big(x\,\frac{\mathrm{d}}{\mathrm{d}x}\Big)^j\,\frac{x}{1-x},\quad \forall\, j\in\mathbb{N}\cup\{0\},\tag{20}$$

with $\Big(x\,\frac{\mathrm{d}}{\mathrm{d}x}\Big)^j$ denoting differentiation with respect to x and multiplication of the derivative by x, repeatedly j times. In particular, we have

$$\mathrm{Li}_0(x)=\frac{x}{1-x},\quad \mathrm{Li}_{-1}(x)=\frac{x}{(1-x)^2},\quad \mathrm{Li}_{-2}(x)=\frac{x(1+x)}{(1-x)^3},\tag{21}$$

and so on. The function $\mathrm{Li}_{-j}(x)$ is a built–in function in the Matlab and Mathematica softwares, which is expressed as $\mathrm{polylog}(-j,x)$. By Corollary 1, if $\rho\in(0,1)$, then (18) is simplified to

$$\mathbb{E}\{G^\rho|x\}=1+\frac{\rho}{\Gamma(1-\rho)}\int_0^\infty\frac{e^{-u}-e^{-2u}}{u^{\rho+1}\big[\big(1-\widetilde{P}(x)\big)^{-1}-e^{-u}\big]}\,\mathrm{d}u.\tag{22}$$

Let P denote the distribution of X. Averaging over X to get the unconditional ρ–th moment using (22), one obtains for all $\rho\in(0,1)$,

$$\mathbb{E}\{G^\rho\}=1+\frac{\rho}{\Gamma(1-\rho)}\int_0^1\frac{1-z}{(-\ln z)^{\rho+1}}\sum_{x\in\mathcal{X}}\frac{P(x)\big(1-\widetilde{P}(x)\big)}{1-z\big(1-\widetilde{P}(x)\big)}\,\mathrm{d}z,\tag{23}$$

where (23) is obtained by using the substitution $z := e^{-u}$. A suitable expression of such an integral is similarly obtained, for all $\rho>0$, by averaging (18) over X. In comparison, a direct calculation of the ρ–th moment gives

$$\mathbb{E}\{G^\rho\}=\sum_{x\in\mathcal{X}}P(x)\,\mathbb{E}\{G^\rho|x\}=\sum_{k=1}^{\infty}\sum_{x\in\mathcal{X}}k^\rho\big(1-\widetilde{P}(x)\big)^{k-1}\,\widetilde{P}(x)\,P(x).\tag{24}$$

The double sum in (24) involves a numerical computation of an infinite series, where the number of terms required to obtain a good approximation increases with ρ, and needs to be determined. The right–hand side of (23), on the other hand, involves integration over $[0,1]$. For

every practical purpose, however, definite integrals in one or two dimensions can be calculated instantly using built-in numerical integration procedures in MATLAB, Maple, Mathematica, or any other mathematical software tools, and the computational complexity of the integral in (23) is not affected by ρ.

As a complement to (18) (which applies to a non-integral and positive ρ), we obtain that the ρ-th moment of the number of randomized guesses, with $\rho \in \mathbb{N}$, is equal to

$$\begin{aligned}\mathbb{E}\{G^\rho|x\} &= \mathbb{E}\big\{\big[(G-1)+1\big]^\rho \,|\, x\big\} \\ &= \sum_{j=0}^{\rho} \binom{\rho}{j} \mathbb{E}\big\{(G-1)^j|x\big\} \\ &= \sum_{j=0}^{\rho} \binom{\rho}{j} \alpha_j \\ &= 1 + \widetilde{P}(x) \sum_{j=1}^{\rho} \left\{ \binom{\rho}{j} \mathrm{Li}_{-j}\big(1-\widetilde{P}(x)\big) \right\}, \end{aligned} \tag{25}$$

where (25) follows from (19) and since $\alpha_0 = 1$. By averaging over X,

$$\mathbb{E}\{G^\rho\} = 1 + \sum_{x \in cX} \left\{ P(x)\widetilde{P}(x) \sum_{j=1}^{\rho} \left\{ \binom{\rho}{j} \mathrm{Li}_{-j}\big(1-\widetilde{P}(x)\big) \right\} \right\}. \tag{26}$$

To conclude, (18) and its simplification in (22) for $\rho \in (0,1)$ give calculable one–dimensional integral expressions for the ρ–th guessing moment with any $\rho > 0$. This refers to a randomized guessing strategy whose practical advantages were further explained in [21] and [27]. This avoids the need of numerical calculations of infinite sums. A Further simplification for $\rho \in \mathbb{N}$ is provided in (25) and (26), expressed in closed form as a function of polylogarithms.

B. Moments of Estimation Errors

Let $X_1, \ldots, X_n$ be i.i.d. random variables with an unknown expectation θ to be estimated, and consider the simple estimator,

$$\widehat{\theta}_n = \frac{1}{n} \sum_{i=1}^{n} X_i. \tag{27}$$

For given $\rho > 0$, we next derive an easily-calculable expression of the ρ-th moment of the estimation error.

Let $D_n := \big(\widehat{\theta}_n - \theta\big)^2$ and $\rho' := \frac{\rho}{2}$. By Theorem 1, if $\rho > 0$ is a non-integral multiple of 2, then

$$\begin{aligned}
&\mathbb{E}\big\{\big|\widehat{\theta}_n - \theta\big|^{\rho}\big\} \\
&= \mathbb{E}\big\{D_n^{\rho'}\big\} \qquad (28) \\
&= \frac{2}{2+\rho} \sum_{\ell=0}^{\lfloor \rho/2 \rfloor} \frac{\alpha_\ell}{B\big(\ell+1, \rho/2+1-\ell\big)} \\
&\quad + \frac{\rho}{2\pi}\, \sin\Big(\frac{\pi\rho}{2}\Big)\, \Gamma\Big(\frac{\rho}{2}\Big) \int_0^\infty \frac{1}{u^{\rho/2+1}} \left(\sum_{j=0}^{\lfloor \rho/2 \rfloor} \left\{ \frac{(-1)^j\, \alpha_j}{j!}\, u^j \right\} e^{-u} - M_{D_n}(-u) \right) \mathrm{d}u, \qquad (29)
\end{aligned}$$

where

$$M_{D_n}(-u) = \mathbb{E}\big\{\exp\big(-u\big(\widehat{\theta}_n - \theta\big)^2\big)\big\}, \quad \forall\, u \geq 0, \tag{30}$$

$\alpha_0 := 1$, and for all $j \in \mathbb{N}$ (see (11))

$$\alpha_j = \frac{1}{j+1} \sum_{\ell=0}^{j} \frac{(-1)^{j-\ell}\, M_{D_n}^{(\ell)}(0)}{B(\ell+1, j-\ell+1)}. \tag{31}$$

By Corollary 1 and (28), if in particular $\rho \in (0,2)$, then the right–hand side of (29) is simplified to

$$\mathbb{E}\big\{\big|\widehat{\theta}_n - \theta\big|^{\rho}\big\} = 1 + \frac{\rho}{2\,\Gamma(1-\frac{1}{2}\,\rho)} \int_0^\infty u^{-(1+\frac{1}{2}\rho)} \left[e^{-u} - M_{D_n}(-u)\right] \mathrm{d}u, \tag{32}$$

and, for all $k \in \mathbb{N}$,

$$\mathbb{E}\big\{|\widehat{\theta}_n - \theta|^{2k}\big\} = M_{D_n}^{(k)}(0). \tag{33}$$

In view of (28)–(33), obtaining a closed–form expression for the ρ-th moment of the estimation error, for an arbitrary $\rho > 0$, hinges on the calculation of the right side of (30) for all $u \geq 0$. To this end, we invoke the identity

$$e^{-uz^2} = \frac{1}{2\sqrt{\pi u}} \int_{-\infty}^{\infty} e^{-j\omega z - \omega^2/(4u)}\, \mathrm{d}\omega, \quad \forall\, u > 0,\, z \in \mathbb{R}, \tag{34}$$

which is the MGF of a zero–mean Gaussian random variable with variance $\frac{1}{2u}$. Together with (30), it gives (see (see Appendix B.1)

$$M_{D_n}(-u) = \frac{1}{2\sqrt{\pi u}} \int_{-\infty}^{\infty} e^{-j\omega\theta}\, \phi_X^n\Big(\frac{\omega}{n}\Big)\, e^{-\omega^2/(4u)}\, \mathrm{d}\omega, \quad \forall\, u > 0, \tag{35}$$

where X is a generic random variable with the same distribution as of X_i for all i.

The combination of (29)–(33) enables to calculate exactly the ρ-th moment $\mathbb{E}\big\{|\widehat{\theta}_n - \theta|^\rho\big\}$, for any given $\rho > 0$, in terms of a two-dimensional integral. Combining (32) and (35) yields, for all $\rho \in (0,2)$,

$$\begin{aligned}&\mathbb{E}\big\{\big|\widehat{\theta}_n - \theta\big|^\rho\big\}\\ &= 1 + \frac{\rho}{2\,\Gamma(1-\frac{1}{2}\,\rho)} \int_0^\infty \int_{-\infty}^\infty u^{-(\rho/2+1)} \Big[\tfrac{1}{2}\, e^{-u-|\omega|} - \frac{1}{2\sqrt{\pi u}}\, \phi_X^n\Big(\frac{\omega}{n}\Big)\, e^{-j\omega\theta - \omega^2/(4u)}\Big]\, \mathrm{d}\omega\, \mathrm{d}u,\end{aligned} \tag{36}$$

where we have used the identity $\int_{-\infty}^\infty \frac{1}{2}\, e^{-|\omega|}\, \mathrm{d}\omega = 1$ in the derivation of the first term of the integral on the right–hand side of (36).

As an example, consider the case where $\{X_i\}_{i=1}^n$ are i.i.d. Bernoulli random variables with

$$\mathbb{P}\{X_1 = 1\} = \theta, \quad \mathbb{P}\{X_1 = 0\} = 1 - \theta \tag{37}$$

where the characteristic function is given by

$$\phi_X(u) := \mathbb{E}\big\{e^{juX}\big\} = 1 + \theta\big(e^{ju} - 1\big), \quad u \in \mathbb{R}. \tag{38}$$

Thanks to the availability of the exact expression, we can next compare the exact ρ-th moment of the estimation error $|\widehat{\theta}_n - \theta|$, with the following closed–form upper bound (see Appendix B.2) and thereby assess its tightness:

$$\mathbb{E}\big\{\big|\widehat{\theta}_n - \theta\big|^\rho\big\} \le K(\rho,\theta) \cdot n^{-\rho/2}, \tag{39}$$

which holds for all $n \in \mathbb{N}$, $\rho > 0$ and $\theta \in [0,1]$, with

$$K(\rho,\theta) := \rho\, \Gamma\Big(\frac{\rho}{2}\Big)\, \big(2\theta\,(1-\theta)\big)^{\rho/2}. \tag{40}$$

Figures 1 and 2 display plots of $\mathbb{E}\,\big|\widehat{\theta}_n - \theta\big|$ as a function of θ and n, in comparison to the upper bound (39). The difference in the plot of Figure 1 is significant except for the boundaries of the interval $[0,1]$, where both the exact value and the bound vanish. Figure 2 indicates that the exact value of $\mathbb{E}\,\big|\widehat{\theta}_n - \theta\big|$, for large n, scales like $\sqrt{n}$; this is reflected from the apparent parallelism of the curves in both graphs, and by the upper bound (39).

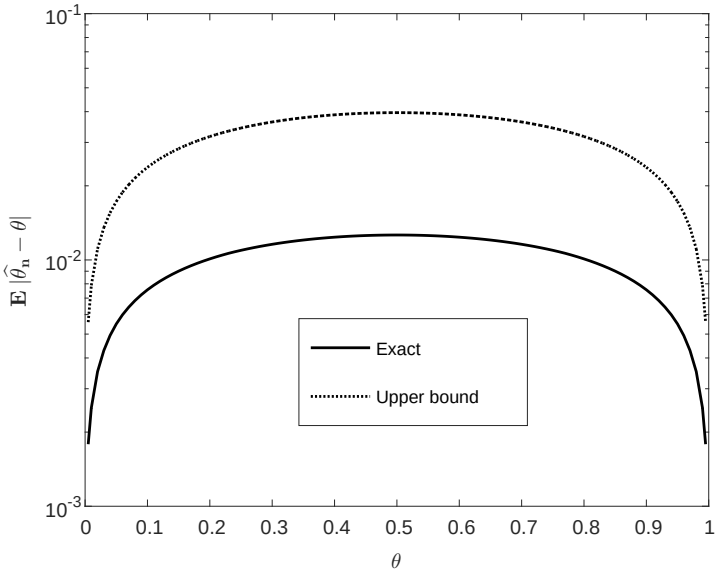

Fig. 1. The exact value of $\mathbb{E}\left|\widehat{\theta}_n - \theta\right|$ (see (36) and (38)) in comparison to the upper bound (39) as functions of $\theta \in [0,1]$ with $n = 1000$.

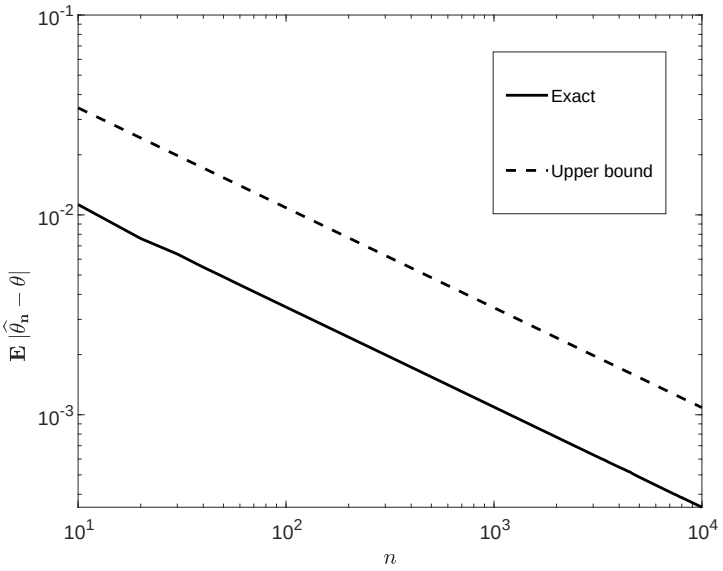

Fig. 2. A plot of $\mathbb{E}\left|\widehat{\theta}_n - \theta\right|$ (see (36) and (38)) versus the upper bound (39) as functions of n with $\theta = \frac{1}{4}$.

To conclude, this subsection provides an exact, double–integral expression for the ρ-th moment of the estimation error of the expectation of n i.i.d. random variables. In other words, the dimension of the integral does not increase with n, and it is a calculable expression. We further compare our expression with an upper bound that stems from concentration inequalities. Although the scaling of the bound as a polynomial of n is correct, the difference between the

exact expression and the bound is significant (see Fig. 1 and 2).

C. Rényi Entropy of Extended Multivariate Cauchy Distributions

Let $X^n = (X_1, \ldots, X_n)$ be a random vector whose probability density function is of the form

$$f(x^n) = \frac{C_n}{[1 + \sum_{i=1}^n g(x_i)]^q}, \quad x^n = (x_1, \ldots, x_n) \in \mathbb{R}^n, \tag{41}$$

for a certain function $g\colon \mathbb{R} \to [0, \infty)$, and a positive constant q such that

$$\int_{\mathbb{R}^n} \frac{1}{[1 + \sum_{i=1}^n g(x_i)]^q} \, \mathrm{d}x^n < \infty. \tag{42}$$

We refer to this kind of density (see also [22, Section 3.1]) as a *generalized multivariate Cauchy density* because the multivariate Cauchy density function is the special case pertaining to the choices $g(x) = x^2$ and $q = \frac{1}{2}(n+1)$. The differential Shannon entropy of the generalized multivariate Cauchy density was derived in [22, Section 3.1] using the integral representation of the logarithm (1), where it was presented as a two–dimensional integral.

We next extend the analysis of [22] to differential Rényi entropies of an arbitrary positive order α (recall that the differential Rényi entropy is specialized to the differential Shannon entropy at $\alpha = 1$ [26]). We show that, for the generalized multivariate Cauchy density, the differential Rényi entropy can be presented as a two–dimensional integral, rather than an n–dimensional integral. Defining

$$Z(t) := \int_{-\infty}^{\infty} e^{-t g(x)} \, \mathrm{d}x, \quad t > 0, \tag{43}$$

we get from (41) (see [22, Section 3.1]) that

$$C_n = \frac{\Gamma(q)}{\displaystyle\int_0^{\infty} t^{q-1} e^{-t} Z^n(t) \, \mathrm{d}t}. \tag{44}$$

For $g(x) = |x|^\theta$, with a fixed $\theta > 0$, (43) implies that

$$Z(t) = \frac{2\,\Gamma(1/\theta)}{\theta\, t^{1/\theta}}. \tag{45}$$

In particular, for $\theta = 2$ and $q = \frac{1}{2}(n+1)$, we get the multivariate Cauchy density from (41). In this case, it follows from (45) that $Z(t) = \sqrt{\frac{\pi}{t}}$ for $t > 0$, and from (44)

$$C_n = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\pi^{(n+1)/2}}. \tag{46}$$

For $\alpha \in (0,1) \cup (1,\infty)$, the (differential) Rényi entropy of order α is given by

$$\begin{aligned} h_\alpha(X^n) &:= \frac{1}{1-\alpha} \log \int_{\mathbb{R}^n} f^\alpha(x^n)\,\mathrm{d}x^n \\ &= \frac{1}{1-\alpha} \log \mathbb{E}\big[f^{\alpha-1}(X^n)\big]. \end{aligned} \tag{47}$$

Using the Laplace transform relation,

$$\frac{1}{s^q} = \frac{1}{\Gamma(q)} \int_0^\infty t^{q-1} e^{-st}\,\mathrm{d}t, \quad \forall\, q>0,\ \mathrm{Re}(s)>0, \tag{48}$$

we obtain that, for $\alpha > 1$ (see Appendix C),

$$\begin{aligned} h_\alpha(X^n) = \frac{\alpha}{\alpha-1} \log \int_0^\infty t^{q-1} e^{-t} Z^n(t)\,\mathrm{d}t + \frac{\log\Gamma\big(q(\alpha-1)\big)}{\alpha-1} - \log\Gamma(q) \\ - \frac{1}{\alpha-1} \log \int_0^\infty \int_0^\infty t^{q(\alpha-1)-1} u^{q-1} e^{-(t+u)}\, Z^n(t+u)\,\mathrm{d}u\,\mathrm{d}t. \end{aligned} \tag{49}$$

Otherwise, if $\alpha \in (0,1)$, we distinguish between the following two cases:

1) If $\alpha = 1 - \frac{m}{q}$ for some $m \in \{1,\ldots,q-1\}$, then

$$\begin{aligned} h_\alpha(X^n) = \frac{\alpha}{1-\alpha} \log C_n - \frac{1}{1-\alpha} \log\Gamma(q) \\ + \frac{1}{1-\alpha} \log\left(\sum_{\ell=0}^{m} \left\{ (-1)^{m-\ell} \int_0^\infty t^{q-1} e^{-t} \varphi_n^{(\ell)}(t)\,\mathrm{d}t \right\} \right), \end{aligned} \tag{50}$$

with

$$\varphi_n(t) := Z^n(t), \quad \forall\, t \geq 0. \tag{51}$$

2) Otherwise (i.e., if $\rho := q(1-\alpha) \notin \mathbb{N}$), then

$$\begin{aligned} &h_\alpha(X^n) \\ &= -\log C_n + \frac{1}{1-\alpha} \log\Bigg(\frac{1}{1+\rho} \sum_{\ell=0}^{\lfloor\rho\rfloor} \frac{\beta_\ell(n)}{B(\ell+1,\rho+1-\ell)} \\ &\qquad + \frac{\rho\,\sin(\pi\rho)\,\Gamma(\rho)}{\pi} \int_0^\infty \frac{e^{-u}}{u^{\rho+1}} \left(\sum_{j=0}^{\lfloor\rho\rfloor} \left\{ \frac{(-1)^j\,\beta_j(n)}{j!}\, u^j \right\} \right. \\ &\qquad\qquad \left. - \frac{C_n}{\Gamma(q)} \int_0^\infty t^{q-1} e^{-t} Z^n(t+u)\,\mathrm{d}t \right) \Bigg), \end{aligned} \tag{52}$$

where $\beta_0 := 1$, and for all $j \in \mathbb{N}$

$$\beta_j(n) := \frac{C_n}{\Gamma(q)} \sum_{\ell=0}^{j} \left\{ \frac{(-1)^{j-\ell}}{B(\ell+1,j-\ell+1)} \sum_{k=0}^{\ell} \left\{ (-1)^{\ell-k} \binom{\ell}{k} \int_0^\infty t^{q-1} e^{-t} \varphi_n^{(k)}(t)\,\mathrm{d}t \right\} \right\}. \tag{53}$$

The proof of the integral expressions of the Rényi entropy of order $\alpha \in (0,1)$, as given in (49)–(53), is provided in Appendix C.

Once again, the advantage of these expressions, which do not seem to be very easy (at least on the face of it), is that they only involve one– or two–dimensional integrals, rather than an expression of an n–dimensional integral (as it could have been in the case of an n–dimensional density).

D. Mutual Information Calculations for Communication Channels with Jamming

Consider a channel that is fed by an input vector $X^n = (X_1, \ldots, X_n) \in \mathcal{X}^n$ and generates an output vector $Y^n = (Y_1, \ldots, Y_n) \in \mathcal{Y}^n$, where $\mathcal{X}$ and $\mathcal{Y}$ are either finite, countably infinite or continuous alphabets, and $\mathcal{X}^n$ and $\mathcal{Y}^n$ are their n-th order Cartesian powers. Let the conditional probability distribution of the channel be given by

$$p_{Y^n|X^n}(y^n|x^n) = \frac{1}{n}\sum_{i=1}^{n}\left\{\prod_{j\neq i} q_{Y|X}(y_j|x_j)\; r_{Y|X}(y_i|x_i)\right\}, \tag{54}$$

where $r_{Y|X}(\cdot|\cdot)$ and $q_{Y|X}(\cdot|\cdot)$ are given conditional probability distributions of Y given X, $x^n = (x_1, \ldots, x_n) \in \mathcal{X}^n$ and $y^n = (y_1, \ldots, y_n) \in \mathcal{Y}^n$. This channel model refers to a discrete memoryless channel (DMC), which is nominally given by

$$q_{Y^n|X^n}(y^n|x^n) = \prod_{i=1}^{n} q_{Y|X}(y_i|x_i), \tag{55}$$

where one of the transmitted symbols is jammed at a uniformly distributed random time, i, and the transition distribution of the jammed symbol is given by $r_{Y|X}(y_i|x_i)$ instead of $q_{Y|X}(y_i|x_i)$. The restriction to a single jammed symbol is made merely for the sake of simplicity, but it can easily be extended.

We wish to evaluate how the jamming affects the mutual information $I(X^n;Y^n)$. Clearly, when one talks about jamming, it should be worse, but this is not part of the mathematical model, where the relation between r and q has not been specified. Let the input distribution be given by the product form

$$p_{X^n}(x^n) = \prod_{i=1}^{n} p_X(x_i), \quad x^n \in \mathcal{X}^n. \tag{56}$$

The mutual information (in nats) is given by

$$
\begin{aligned}
&I(X^n;Y^n)\\
&= h(Y^n) - h(Y^n|X^n) \qquad (57)\\
&= \int_{\mathcal{X}^n\times\mathcal{Y}^n} p_{X^n,Y^n}(x^n,y^n)\,\ln p_{Y^n|X^n}(y^n|x^n)\,\mathrm{d}x^n\,\mathrm{d}y^n - \int_{\mathcal{Y}^n} p_{Y^n}(y^n)\,\ln p_{Y^n}(y^n)\,\mathrm{d}y^n. \qquad (58)
\end{aligned}
$$

For simplicity of notation, we henceforth omit the domains of integration whenever they are clear from the context. We have,

$$
\begin{aligned}
&\int p_{X^n,Y^n}(x^n,y^n)\,\ln p_{Y^n|X^n}(y^n|x^n)\,\mathrm{d}x^n\,\mathrm{d}y^n\\
&= \int p_{X^n,Y^n}(x^n,y^n)\,\ln\!\left(\frac{p_{Y^n|X^n}(y^n|x^n)}{q_{Y^n|X^n}(y^n|x^n)}\right)\mathrm{d}x^n\,\mathrm{d}y^n\\
&\quad + \int p_{X^n,Y^n}(x^n,y^n)\,\ln q_{Y^n|X^n}(y^n|x^n)\,\mathrm{d}x^n\,\mathrm{d}y^n. \qquad (59)
\end{aligned}
$$

By using the logarithmic expectation in (15), and the following equality (see (54) and (55)):

$$
\frac{p_{Y^n|X^n}(y^n|x^n)}{q_{Y^n|X^n}(y^n|x^n)} = \frac{1}{n}\sum_{i=1}^{n}\frac{r_{Y|X}(y_i|x_i)}{q_{Y|X}(y_i|x_i)}, \qquad (60)
$$

we obtain (see Appendix D.1)

$$
\begin{aligned}
&\int p_{X^n,Y^n}(x^n,y^n)\,\ln\!\left(\frac{p_{Y^n|X^n}(y^n|x^n)}{q_{Y^n|X^n}(y^n|x^n)}\right)\mathrm{d}x^n\,\mathrm{d}y^n\\
&= \int_0^\infty \frac{1}{u}\left[e^{-u} - f^{n-1}\!\left(\frac{u}{n}\right) g\!\left(\frac{u}{n}\right)\right]\mathrm{d}u, \qquad (61)
\end{aligned}
$$

where, for $u \geq 0$,

$$
f(u) := \int p_X(x)\,q_{Y|X}(y|x)\,\exp\!\left(-\frac{u\,r_{Y|X}(y|x)}{q_{Y|X}(y|x)}\right)\mathrm{d}x\,\mathrm{d}y, \qquad (62)
$$

$$
g(u) := \int p_X(x)\,r_{Y|X}(y|x)\,\exp\!\left(-\frac{u\,r_{Y|X}(y|x)}{q_{Y|X}(y|x)}\right)\mathrm{d}x\,\mathrm{d}y. \qquad (63)
$$

Moreover, owing to the product form of q_n, it is shown in Appendix D.2 that

$$
\begin{aligned}
&\int p_{X^n,Y^n}(x^n,y^n)\,\ln q_{Y^n|X^n}(y^n|x^n)\,\mathrm{d}x^n\,\mathrm{d}y^n\\
&= \int p_X(x)\,r_{Y|X}(y|x)\,\ln q_{Y|X}(y|x)\,\mathrm{d}x\,\mathrm{d}y\\
&\quad + (n-1)\int p_X(x)\,q_{Y|X}(y|x)\,\ln q_{Y|X}(y|x)\,\mathrm{d}x\,\mathrm{d}y. \qquad (64)
\end{aligned}
$$

Combining (59), (61) and (64), we express $h(Y^n|X^n)$ as a double integral over $\mathcal{X} \times \mathcal{Y}$, independently of n (rather than an integration over $\mathcal{X}^n \times \mathcal{Y}^n$):

$$\begin{aligned} h(Y^n|X^n) = &\int_0^\infty \frac{1}{u} \Big[f^{n-1}\Big(\frac{u}{n}\Big)\, g\Big(\frac{u}{n}\Big) - e^{-u} \Big]\, \mathrm{d}u \\ &- \int p_X(x)\, r_{Y|X}(y|x)\, \ln q_{Y|X}(y|x)\, \mathrm{d}x\, \mathrm{d}y \\ &- (n-1) \int p_X(x)\, q_{Y|X}(y|x)\, \ln q_{Y|X}(y|x)\, \mathrm{d}x\, \mathrm{d}y. \end{aligned} \tag{65}$$

We next calculate the differential channel output entropy, $h(Y^n)$, induced by $p_{Y^n|X^n}(\cdot|\cdot)$. From Appendix D.3,

$$p_{Y^n}(y^n) = \prod_{j=1}^n v(y_j) \cdot \frac{1}{n} \sum_{i=1}^n \frac{w(y_i)}{v(y_i)}, \tag{66}$$

where, for all $y \in \mathcal{Y}$,

$$v(y) := \int q_{Y|X}(y|x)\, p_X(x)\, \mathrm{d}x, \tag{67}$$

$$w(y) := \int r_{Y|X}(y|x)\, p_X(x)\, \mathrm{d}x. \tag{68}$$

By (1), the following identity holds for every positive random variable Z (see Appendix D.3):

$$\mathbb{E}\{Z \ln Z\} = \int_0^\infty \frac{1}{u} \Big[M_Z'(0)\, e^{-u} - M_Z'(-u) \Big]\, \mathrm{d}u \tag{69}$$

where $M_Z(u) := \mathbb{E}\{e^{uZ}\}$. By setting $Z := \frac{1}{n} \sum_{i=1}^n \frac{w(V_i)}{v(V_i)}$ where $\{V_i\}_{i=1}^n$ are i.i.d. random variables with the density function v, some algebraic manipulations give (see Appendix D.3)

$$\begin{aligned} h(Y^n) = &\int_0^\infty \frac{1}{u} \Big[t^{n-1}\Big(\frac{u}{n}\Big)\, s\Big(\frac{u}{n}\Big) - e^{-u} \Big]\, \mathrm{d}u \\ &- \int w(y)\, \ln v(y)\, \mathrm{d}y - (n-1) \int v(y) \ln v(y)\, \mathrm{d}y, \end{aligned} \tag{70}$$

where

$$s(u) := \int w(y)\, \exp\Big(-\frac{u\, w(y)}{v(y)} \Big)\, \mathrm{d}y, \quad u \geq 0, \tag{71}$$

$$t(u) := \int v(y)\, \exp\Big(-\frac{u\, w(y)}{v(y)} \Big)\, \mathrm{d}y, \quad u \geq 0. \tag{72}$$

Combining (57), (65) and (70), we obtain the mutual information for the channel with jamming, which is given by

$$I_p(X^n; Y^n) = \int_0^\infty \frac{1}{u} \Big[t^{n-1}\Big(\frac{u}{n}\Big)\, s\Big(\frac{u}{n}\Big) - f^{n-1}\Big(\frac{u}{n}\Big)\, g\Big(\frac{u}{n}\Big) \Big]\, \mathrm{d}u$$

$$+ \int p_X(x)\, r_{Y|X}(y|x)\, \ln q_{Y|X}(y|x)\, \mathrm{d}x\, \mathrm{d}y - \int w(y)\, \ln v(y)\, \mathrm{d}y$$
$$+ (n-1)\Big[\int p_X(x)\, q_{Y|X}(y|x)\, \ln q_{Y|X}(y|x)\, \mathrm{d}x\, \mathrm{d}y - \int v(y) \ln v(y)\, \mathrm{d}y\Big]. \quad (73)$$

We next exemplify our results in the case where q is a binary symmetric channel (BSC) with crossover probability $\delta \in (0, \frac{1}{2})$, and p is a BSC with a larger crossover probability, $\varepsilon \in (\delta, \frac{1}{2}]$. We assume that the input bits are i.i.d. and equiprobable. The specialization of our analysis to this setup is provided in Appendix D.4, showing that the mutual information of the channel p_{X^n,Y^n}, fed by the binary symmetric source, is given by

$$I_p(X^n;Y^n) = n\ln 2 - d(\varepsilon\|\delta) - h_{\mathrm{b}}(\varepsilon) - (n-1)h_{\mathrm{b}}(\delta) \quad (74)$$
$$+ \int_0^\infty \bigg\{ e^{-u} - \bigg[(1-\delta)\, \exp\Big(-\frac{(1-\varepsilon)u}{(1-\delta)n}\Big) + \delta\, \exp\Big(-\frac{\varepsilon u}{\delta n}\Big)\bigg]^{n-1}$$
$$\cdot \bigg[(1-\varepsilon)\, \exp\Big(-\frac{(1-\varepsilon)u}{(1-\delta)n}\Big) + \varepsilon\, \exp\Big(-\frac{\varepsilon u}{\delta n}\Big)\bigg]\bigg\}\, \frac{\mathrm{d}u}{u},$$

where $h_{\mathrm{b}}\colon [0,1] \to [0, \ln 2]$ is the binary entropy function

$$h_{\mathrm{b}}(x) := -x\ln(x) - (1-x)\ln(1-x), \quad x \in [0,1] \quad (75)$$

with the convention that $0 \ln 0 = 0$, and

$$d(\varepsilon\|\delta) := \varepsilon\, \ln\Big(\frac{\varepsilon}{\delta}\Big) + (1-\varepsilon)\, \ln\Big(\frac{1-\varepsilon}{1-\delta}\Big), \quad (\delta,\varepsilon) \in [0,1]^2 \quad (76)$$

denotes the binary relative entropy. By the data processing inequality, the mutual information in (74) is smaller than that of the BSC with crossover probability δ:

$$I_q(X^n;Y^n) = n\big(\ln 2 - h_{\mathrm{b}}(\delta)\big). \quad (77)$$

Fig. 3 refers to the case where $\delta = 10^{-3}$ and $n = 128$. Here $I_q(X^n;Y^n) = 87.71$ nats, and $I_p(X^n;Y^n)$ is decreased by 2.88 nats due to the jammer (see Fig. 3).

To conclude, this subsection studies the change in the mutual information $I(X^n;Y^n)$ due to jamming, relative to the mutual information associated with the nominal channel without jamming. Due to the integral representations provided in our analysis, the calculation of the mutual information finally depends on one–dimensional integrals, as opposed to the original n-dimensional integrals, pertaining to the expressions that define the associated differential entropies.

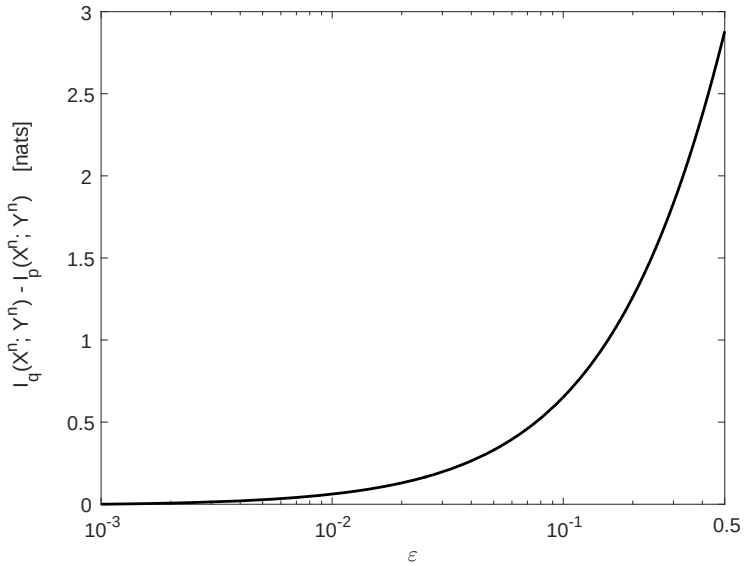

Fig. 3. The degradation in mutual information for $n = 128$. The jammer–free channel q is a BSC with crossover probability $\delta = 10^{-3}$, and r is a BSC with crossover probability $\varepsilon \in \left(\delta, \frac{1}{2}\right]$. The input bits are i.i.d. and equiprobable. The degradation in $I(X^n; Y^n)$ (nats) is displayed as a function of ε.

APPENDIX A
PROOF OF THEOREM 1

Let $\rho > 0$ be a non-integer real, and define the function $F_\rho \colon (0, \infty) \to \mathbb{R}$ as follows:

$$F_\rho(\mu) := \int_0^\infty \frac{1}{u^{\rho+1}} \left(e^{-\mu u} - \sum_{j=0}^{\lfloor \rho \rfloor} \left\{ \frac{(-1)^j}{j!} (\mu - 1)^j u^j \right\} e^{-u} \right) \mathrm{d}u, \quad \mu > 0, \tag{A.1}$$

with the convention that $0^0 := \lim\limits_{x \to 0^+} x^x = 1$. By the Taylor series expansion of $e^{-\mu u}$ as a function of μ around $\mu = 1$, we find that for small positive u, the integrand of (A.1) scales like $u^{-(\rho - \lfloor \rho \rfloor)}$ with $\rho - \lfloor \rho \rfloor \in (0, 1)$. Furthermore, for large u, the same integrand scales like $u^{-(\rho+1)} e^{-\min\{\mu, 1\} u}$. This guarantees the convergence of the integral, and so $F_\rho(\cdot)$ is well-defined and finite in the interval $(0, \infty)$.

From (A.1), $F_\rho(1) = 0$ (for $\mu = 1$, the integrand of (A.1) is identically zero on $(0, \infty)$). Differentiation ℓ times with respect to μ, under the integration sign with $\ell \in \left\{0, \ldots, \lfloor \rho \rfloor\right\}$, gives

$$F_\rho^{(\ell)}(\mu) = \int_0^\infty \frac{1}{u^{\rho+1}} \left[(-1)^\ell u^\ell e^{-\mu u} - \sum_{j=\ell}^{\lfloor \rho \rfloor} \left\{ \frac{(-1)^j}{(j-\ell)!} \cdot (\mu - 1)^{j-\ell} u^j \right\} e^{-u} \right] \mathrm{d}u, \tag{A.2}$$

which implies that

$$F_\rho^{(\ell)}(1) = 0, \qquad \ell = 0, \ldots, \lfloor \rho \rfloor. \tag{A.3}$$

We next calculate $F_\rho^{(k)}(\mu)$ for $k := \lfloor \rho \rfloor + 1$ and $\mu > 0$:

$$\begin{aligned} F_\rho^{(k)}(\mu) &= \int_0^\infty \frac{1}{u^{\rho+1}} \frac{\partial^k}{\partial \mu^k} \left\{ e^{-\mu u} - \sum_{j=0}^{\lfloor \rho \rfloor} \left\{ \frac{(-1)^j}{j!} \cdot (\mu - 1)^j u^j \right\} e^{-u} \right\} \mathrm{d}u \\ &= \int_0^\infty \frac{(-u)^k e^{-\mu u}}{u^{\rho+1}} \, \mathrm{d}u \\ &= (-1)^k \int_0^\infty u^{k-\rho-1} e^{-\mu u} \, \mathrm{d}u \\ &= (-1)^k \int_0^\infty \Big(\frac{t}{\mu} \Big)^{k-\rho-1} e^{-t} \, \mu^{-1} \, \mathrm{d}t \\ &= (-1)^k \mu^{\rho-k} \, \Gamma(k-\rho). \end{aligned} \tag{A.4}$$

Hence, from (A.3) and (A.4),

$$F_\rho(1) = \ldots = F_\rho^{(\lfloor \rho \rfloor)}(1) = 0, \tag{A.5}$$

$$F_\rho^{(k)}(\mu) = (-1)^k \mu^{\rho-k} \, \Gamma(k-\rho), \qquad k := \lfloor \rho \rfloor + 1, \ \mu > 0. \tag{A.6}$$

By integrating both sides of (A.6) with respect to μ, successively k times, (A.5) implies that

$$F_\rho(\mu) = \frac{(-1)^k \, \Gamma(k-\rho) \, \mu^\rho}{\prod_{i=0}^{k-1} (\rho - i)} + \sum_{i=0}^{k-1} c_i(\rho) \, (\mu - 1)^i, \qquad k := \lfloor \rho \rfloor + 1, \ \mu > 0, \tag{A.7}$$

with some integration constants $\big\{ c_i(\rho) \big\}_{i=0}^{k-1}$. Since $F_\rho(1) = 0$ (see (A.5)), (A.7) implies that

$$c_0(\rho) = \frac{(-1)^{k+1} \, \Gamma(k-\rho)}{\prod_{i=0}^{k-1} (\rho - i)}, \tag{A.8}$$

and since (by assumption) ρ is a non-integer, the denominator on the right–hand side of (A.8) is non-zero. Moreover, since $F_\rho^{(\ell)}(1) = 0$ for all $\ell \in \{1, \ldots k-1\}$ (see (A.5)), differentiation of both sides of (A.7) ℓ times at $\mu = 1$ yields

$$c_\ell(\rho) := \frac{(-1)^{k+1} \, \Gamma(k-\rho) \, \prod_{i=0}^{\ell-1} (\rho - i)}{\ell! \, \prod_{i=0}^{k-1} (\rho - i)}, \qquad \ell = 1, \ldots, k-1. \tag{A.9}$$

Substituting (A.8) and (A.9) into (A.7) gives

$$F_\rho(\mu) = \frac{(-1)^k \, \Gamma(k-\rho)}{\prod_{i=0}^{k-1} (\rho - i)} \left[\mu^\rho - 1 - \sum_{\ell=1}^{k-1} \left\{ \frac{1}{\ell!} \prod_{i=0}^{\ell-1} (\rho - i) \, (\mu - 1)^\ell \right\} \right], \quad \mu > 0. \tag{A.10}$$

Combining (A.1) with (A.10) and rearranging terms, we obtain

$$\mu^{\rho} = 1 + \sum_{\ell=1}^{k-1} \left\{ \frac{1}{\ell!} \prod_{i=0}^{\ell-1} (\rho - i)\, (\mu - 1)^{\ell} \right\} + \frac{(-1)^{k-1} \prod_{i=0}^{k-1} (\rho - i)}{\Gamma(k - \rho)} \int_0^{\infty} \frac{1}{u^{\rho+1}} \left(\sum_{j=0}^{\lfloor \rho \rfloor} \left\{ \frac{(-1)^j}{j!} (\mu - 1)^j u^j \right\} e^{-u} - e^{-\mu u} \right) \mathrm{d}u. \quad \text{(A.11)}$$

Setting $\mu := X \geq 0$, and taking expectations of both sides of (A.11) yield (see (5) and (10))

$$\mathbb{E}\{X^{\rho}\} = 1 + \sum_{\ell=1}^{k-1} \left\{ \frac{1}{\ell!} \prod_{i=0}^{\ell-1} (\rho - i)\, \alpha_{\ell} \right\} + \frac{(-1)^{k-1} \prod_{i=0}^{k-1} (\rho - i)}{\Gamma(k - \rho)} \int_0^{\infty} \frac{1}{u^{\rho+1}} \left(\sum_{j=0}^{\lfloor \rho \rfloor} \left\{ \frac{(-1)^j\, \alpha_j}{j!}\, u^j \right\} e^{-u} - M_X(-u) \right) \mathrm{d}u. \quad \text{(A.12)}$$

We next rewrite and simplify both terms in the right side of (A.12) as follows:

$$1 + \sum_{\ell=1}^{k-1} \left\{ \frac{1}{\ell!} \prod_{i=0}^{\ell-1} (\rho - i)\, \alpha_{\ell} \right\}$$

$$= 1 + \sum_{\ell=1}^{k-1} \left\{ \frac{1}{\Gamma(\ell+1)}\, \frac{\Gamma(\rho+1)}{\Gamma(\rho-\ell+1)} \cdot \alpha_{\ell} \right\} \quad \text{(A.13)}$$

$$= 1 + \frac{1}{1+\rho} \sum_{\ell=1}^{k-1} \left\{ \frac{1}{\Gamma(\ell+1)}\, \frac{\Gamma(\rho+2)}{\Gamma(\rho-\ell+1)} \cdot \alpha_{\ell} \right\} \quad \text{(A.14)}$$

$$= 1 + \frac{1}{1+\rho} \sum_{\ell=1}^{k-1} \frac{\alpha_{\ell}}{B(\ell+1, \rho-\ell+1)} \quad \text{(A.15)}$$

$$= \frac{1}{1+\rho} \sum_{\ell=0}^{k-1} \frac{\alpha_{\ell}}{B(\ell+1, \rho-\ell+1)}, \quad \text{(A.16)}$$

and

$$\frac{(-1)^{k-1} \prod_{i=0}^{k-1} (\rho - i)}{\Gamma(k - \rho)}$$

$$= \frac{(-1)^{k-1}\, \Gamma(\rho+1)}{\Gamma(k-\rho)\, \Gamma(\rho-k+1)} \quad \text{(A.17)}$$

$$= (-1)^{k-1}\, \Gamma(\rho+1) \cdot \frac{\sin\big(\pi(k-\rho)\big)}{\pi} \quad \text{(A.18)}$$

$$= \frac{\rho\, \sin(\pi\rho)\, \Gamma(\rho)}{\pi}. \quad \text{(A.19)}$$

Eqs. (A.13), (A.14), (A.17) and (A.19) are based on the recursion (see, e.g., [14, page 904, Identity (8.331)])

$$\Gamma(x+1) = x\,\Gamma(x), \qquad x > 0, \tag{A.20}$$

(A.15) relies on the relation between the Beta and Gamma functions in (8); (A.16) is based on the following equality (see (8), (A.20), and recall that $\Gamma(1) = 1$):

$$B(1, \rho+1) = \frac{\Gamma(1)\,\Gamma(\rho+1)}{\Gamma(\rho+2)} = \frac{1}{\rho+1}, \tag{A.21}$$

and, finally, (A.19) holds by using the identity (see, e.g., [14, page 905, Identity (8.334)])

$$\Gamma(x)\,\Gamma(1-x) = \frac{\pi}{\sin(\pi x)}, \quad \forall\, x \in (0,1), \tag{A.22}$$

with $x := k - \rho = \lfloor \rho \rfloor + 1 - \rho \in (0,1)$ (since, by assumption, ρ is a non-integer). Combining (A.12)–(A.19) gives (9) (recall that $\alpha_0 := 1$, and $k - 1 := \lfloor \rho \rfloor$ holds by (A.6)).

We finally prove (11). By (10), for all $j \in \mathbb{N}$,

$$\begin{aligned}
\alpha_j &= \mathbb{E}\big\{(X-1)^j\big\} \\
&= \sum_{\ell=0}^{j} (-1)^{j-\ell} \binom{j}{\ell}\, \mathbb{E}\big\{X^\ell\big\} \\
&= \sum_{\ell=0}^{j} \frac{(-1)^{j-\ell}\,\Gamma(j+1)\; M_X^{(\ell)}(0)}{\Gamma(\ell+1)\,\Gamma(j-\ell+1)} \\
&= \frac{1}{j+1} \sum_{\ell=0}^{j} \frac{(-1)^{j-\ell}\,\Gamma(j+2)\; M_X^{(\ell)}(0)}{\Gamma(\ell+1)\,\Gamma(j-\ell+1)} \\
&= \frac{1}{j+1} \sum_{\ell=0}^{j} \frac{(-1)^{j-\ell}\, M_X^{(\ell)}(0)}{B(\ell+1, j-\ell+1)}.
\end{aligned} \tag{A.23}$$

APPENDIX B
COMPLEMENTARY DETAILS OF THE ANALYSIS IN SECTION III-B

B.1 *Proof of Eq.* (35)

For all $u > 0$,

$$M_{D_n}(-u) = \mathbb{E}\Big\{\exp\Big(-u\big(\widehat{\theta}_n - \theta\big)^2\Big)\Big\} \tag{B.1}$$

$$= \mathbb{E}\bigg\{\frac{1}{2\sqrt{\pi u}} \int_{-\infty}^{\infty} e^{j\omega(\widehat{\theta}_n - \theta)}\, e^{-\omega^2/(4u)}\, \mathrm{d}\omega\bigg\} \tag{B.2}$$

$$= \frac{1}{2\sqrt{\pi u}} \int_{-\infty}^{\infty} e^{-j\omega\theta}\, \mathbb{E}\big\{ e^{j\omega\widehat{\theta}_n} \big\}\, e^{-\omega^2/(4u)}\, \mathrm{d}\omega \tag{B.3}$$

$$= \frac{1}{2\sqrt{\pi u}} \int_{-\infty}^{\infty} e^{-j\omega\theta}\, \mathbb{E}\bigg\{ \exp\Big(\frac{j\omega}{n} \sum_{i=1}^{n} X_i \Big) \bigg\}\, e^{-\omega^2/(4u)}\, \mathrm{d}\omega \tag{B.4}$$

$$= \frac{1}{2\sqrt{\pi u}} \int_{-\infty}^{\infty} e^{-j\omega\theta}\, \phi_X^n\bigg(\frac{\omega}{n} \bigg)\, e^{-\omega^2/(4u)}\, \mathrm{d}\omega, \tag{B.5}$$

where (B.1) is (30); (B.2) relies on (34); (B.3) holds by interchanging expectation and integration; (B.4) is due to (27), and (B.5) holds by the assumption that $X_1, \ldots, X_n$ are i.i.d.

B.2 *Derivation of the Upper bound in* (39)

For all $\rho > 0$,

$$\begin{aligned} \mathbb{E}\big\{ \big|\widehat{\theta}_n - \theta\big|^\rho \big\} &= \int_0^\infty \mathbb{P}\big(\big|\widehat{\theta}_n - \theta\big|^\rho \geq t \big)\, \mathrm{d}t \\ &= \int_0^\infty \mathbb{P}\big(\big|\widehat{\theta}_n - \theta\big|^\rho \geq \varepsilon^\rho \big)\, \rho\, \varepsilon^{\rho-1}\, \mathrm{d}\varepsilon \\ &= \int_0^\infty \mathbb{P}\big(\big|\widehat{\theta}_n - \theta\big| \geq \varepsilon \big)\, \rho\, \varepsilon^{\rho-1}\, \mathrm{d}\varepsilon. \end{aligned} \tag{B.6}$$

We next use the Chernoff bound for upper bounding $\mathbb{P}\big(\big|\widehat{\theta}_n - \theta\big| \geq \varepsilon \big)$ for all $\varepsilon > 0$,

$$\begin{aligned} \mathbb{P}\big(\widehat{\theta}_n - \theta \geq \varepsilon \big) &= \mathbb{P}\bigg(\sum_{i=1}^{n} (X_i - \theta) \geq n\varepsilon \bigg) \\ &\leq \inf_{s \geq 0} \bigg\{ e^{-sn\varepsilon}\, \mathbb{E}\bigg\{ \exp\Big(s \sum_{i=1}^{n} (X_i - \theta) \Big) \bigg\} \bigg\} \\ &= \inf_{s \geq 0} \bigg\{ e^{-sn\varepsilon} \prod_{i=1}^{n} \mathbb{E}\Big\{ e^{s(X_i - \theta)} \Big\} \bigg\} \\ &= \inf_{s \geq 0} \Big\{ e^{-sn\varepsilon} \Big(\theta\, e^{s(1-\theta)} + (1-\theta)\, e^{-s\theta} \Big)^n \Big\} \\ &= \inf_{s \geq 0} \Big\{ e^{-ns\varepsilon + nH_\theta(s)} \Big\} \end{aligned} \tag{B.7}$$

with $\theta \in [0, 1]$, and

$$H_\theta(s) := \ln\Big(\theta\, e^{s(1-\theta)} + (1-\theta)\, e^{-s\theta} \Big), \quad s \geq 0. \tag{B.8}$$

We now use an upper bound on $H_\theta(s)$ for every $s \geq 0$. By Theorem 3.2 and Lemma 3.3 in [4] (see also [24, Lemma 2.4.6]), we have

$$H_\theta(s) \leq C(\theta)\, s^2 \tag{B.9}$$

with

$$C(\theta) := \begin{cases} 0, & \text{if } \theta = 0, \\ \dfrac{1-2\theta}{4\ln\Big(\frac{1-\theta}{\theta}\Big)}, & \text{if } \theta \in \big(0, \frac{1}{2}\big), \\ \frac{1}{2}\,\theta(1-\theta), & \text{if } \theta \in \big[\frac{1}{2}, 1\big]. \end{cases} \tag{B.10}$$

Combining (B.7) and (B.9) yields

$$\begin{aligned} \mathbb{P}\big(\widehat{\theta}_n - \theta \geq \varepsilon\big) &\leq \inf_{s \geq 0} \Big\{ e^{-n\varepsilon s + nC(\theta)s^2} \Big\} \\ &= \exp\bigg(-\frac{n\varepsilon^2}{4C(\theta)}\bigg). \end{aligned} \tag{B.11}$$

Similarly, it is easy to show that the same Chernoff bound applies also to $\mathbb{P}\big(\widehat{\theta}_n - \theta \leq -\varepsilon\big)$, which overall gives

$$\mathbb{P}\big(\big|\widehat{\theta}_n - \theta\big| \geq \varepsilon\big) \leq 2\exp\bigg(-\frac{n\varepsilon^2}{4C(\theta)}\bigg). \tag{B.12}$$

Inequality (B.12) is a refined version of Hoeffding's inequality (see [24, Section 2.4.4]), which is derived for the Bernoulli distribution (see (B.7)) and by invoking the Chernoff bound; moreover, (B.12) coincides with Hoeffding's inequality in the special case $\theta = \frac{1}{2}$ (which, from (B.10), yields $C(\theta) = \frac{1}{8}$). In view of the fact that (B.12) forms a specialization of [24, Theorem 2.4.7], it follows that the Bernoulli case is the worst one (in the sense of leading to the looser upper bound) among all probability distributions whose support is the interval $[0, 1]$ and whose expected value is $\theta \in [0, 1]$. However, in the Bernoulli case, a simple symmetry argument applies for improving the bound (B.12) as follows. Since $\{X_i\}$ are i.i.d., Bernoulli with mean θ, then obviously, $\{1 - X_i\}$ are Bernoulli, i.i.d. with mean $1 - \theta$ and (from (27))

$$\widehat{\theta}_n(1 - X_1, \ldots, 1 - X_n) = 1 - \widehat{\theta}_n(X_1, \ldots, X_n), \tag{B.13}$$

which implies that the error estimation is identical in both cases. Hence, $\mathbb{P}\big(\big|\widehat{\theta}_n - \theta\big| \geq \varepsilon\big)$ is symmetric around $\theta = \frac{1}{2}$. It can be verified that

$$\min\big\{C(\theta),\, C(1-\theta)\big\} = \tfrac{1}{2}\,\theta(1-\theta), \quad \forall\, \theta \in [0, 1], \tag{B.14}$$

which follows from (B.10) and since $C(\theta) > C(1-\theta)$ for all $\theta \in (0, \frac{1}{2})$ (see [24, Fig. 2.1]). In view of (B.14) and the above symmetry consideration, the upper bound in (B.12) is improved

for values of $\theta \in (0, \frac{1}{2})$, which therefore gives

$$\mathbb{P}\big(\big|\widehat{\theta}_n - \theta\big| \geq \varepsilon\big) \leq 2\exp\bigg(-\frac{n\varepsilon^2}{2\theta(1-\theta)}\bigg), \quad \forall\, \theta \in [0,1],\ \varepsilon > 0. \tag{B.15}$$

From (27), the probability in (B.15) vanishes if $\theta = 0$ or $\theta = 1$. Consequently, for $\rho > 0$,

$$\begin{aligned}
\mathbb{E}\big\{\big|\widehat{\theta}_n - \theta\big|^{\rho}\big\} &= \int_0^{\infty} \mathbb{P}\big(\big|\widehat{\theta}_n - \theta\big| \geq \varepsilon\big)\, \rho\, \varepsilon^{\rho-1}\, \mathrm{d}\varepsilon && \text{(B.16)}\\
&\leq \int_0^{\infty} 2\exp\bigg(-\frac{n\varepsilon^2}{2\theta(1-\theta)}\bigg)\, \rho\, \varepsilon^{\rho-1}\, \mathrm{d}\varepsilon && \text{(B.17)}\\
&= \rho\, \big(2\theta(1-\theta)\big)^{\rho/2} \int_0^{\infty} u^{\rho/2-1}\, e^{-u}\, \mathrm{d}u \cdot n^{-\rho/2} && \text{(B.18)}\\
&= \rho\, \Gamma\Big(\frac{\rho}{2}\Big)\big(2\theta(1-\theta)\big)^{\rho/2} \cdot n^{-\rho/2} && \text{(B.19)}\\
&= K(\rho,\theta) \cdot n^{-\rho/2}, && \text{(B.20)}
\end{aligned}$$

where (B.16)–(B.20) hold, respectively, due to (B.6), (B.15), the substitution $u := \frac{n\varepsilon^2}{2\theta(1-\theta)}$, (7) and (40).

APPENDIX C
COMPLEMENTARY DETAILS OF THE ANALYSIS IN SECTION III-C

We start by proving (49). In view of (47), for $\alpha \in (0,1) \cup (1,\infty)$

$$h_\alpha(X^n) = \frac{1}{1-\alpha}\, \log \mathbb{E}\big[f^{\alpha-1}(X^n)\big], \tag{C.1}$$

where $X^n := (X_1, \ldots, X_n)$. For $\alpha > 1$, we get

$$\begin{aligned}
&\mathbb{E}\big[f^{\alpha-1}(X^n)\big]\\
&= C_n^{\alpha-1}\, \mathbb{E}\bigg\{\bigg[1 + \sum_{i=1}^{n} g(X_i)\bigg]^{q(1-\alpha)}\bigg\} && \text{(C.2)}\\
&= C_n^{\alpha-1} \int_{\mathbb{R}^n} f(x^n) \cdot \frac{1}{\Gamma\big(q(\alpha-1)\big)} \int_0^{\infty} t^{q(\alpha-1)-1} \exp\bigg\{-\bigg(1 + \sum_{i=1}^{n} g(x_i)\bigg)t\bigg\}\, \mathrm{d}t && \text{(C.3)}\\
&= \frac{C_n^{\alpha-1}}{\Gamma\big(q(\alpha-1)\big)} \int_0^{\infty} t^{q(\alpha-1)-1}\, e^{-t}\, \mathbb{E}\bigg[\exp\bigg(-t\sum_{i=1}^{n} g(x_i)\bigg)\bigg]\, \mathrm{d}t. && \text{(C.4)}
\end{aligned}$$

where (C.2) holds due to (41); (C.3) follows from (48), and (C.4) holds by swapping order of integrations. Furthermore, from (41) and (48),

$$f(x^n) = \frac{C_n}{\Big(1 + \sum_{i=1}^{n} g(x_i)\Big)^q}$$

$$= \frac{C_n}{\Gamma(q)} \int_0^\infty u^{q-1} e^{-u} \exp\Big(-u \sum_{i=1}^n g(x_i)\Big) \mathrm{d}u, \quad \forall\, x^n \in \mathbb{R}^n, \tag{C.5}$$

and it follows from (C.5) and by swapping order of integrations,

$$\begin{aligned}
&\mathbb{E}\left[\exp\Big(-t \sum_{i=1}^n g(X_i)\Big)\right] \\
&= \frac{C_n}{\Gamma(q)} \int_0^\infty u^{q-1} e^{-u} \int_{\mathbb{R}^n} \exp\Big(-(t+u) \sum_{i=1}^n g(x_i)\Big) \mathrm{d}x^n \, \mathrm{d}u \\
&= \frac{C_n}{\Gamma(q)} \int_0^\infty u^{q-1} e^{-u} \left\{\prod_{i=1}^n \int_{-\infty}^\infty \exp\big(-(t+u) g(x_i)\big) \, \mathrm{d}x_i\right\} \mathrm{d}u \\
&= \frac{C_n}{\Gamma(q)} \int_0^\infty u^{q-1} e^{-u} \left(\int_{-\infty}^\infty \exp\big(-(t+u)\, g(x)\big) \, \mathrm{d}x\right)^n \mathrm{d}u \\
&= \frac{C_n}{\Gamma(q)} \int_0^\infty u^{q-1} e^{-u} Z^n(t+u) \, \mathrm{d}u
\end{aligned} \tag{C.6}$$

where (C.6) holds by the definition of $Z(\cdot)$ in (43). Finally, combining (44), (C.1), (C.4) and (C.6) gives (49).

The proof of (50)–(53) is a straightforward calculation which follows by combining (C.1), (C.2), (C.6) and Theorem 1 (we replace $\{\alpha_j\}$ in Theorem 1 with $\{\beta_j(n)\}$ in order not to confuse with the order α of the Rényi entropy of X^n).

APPENDIX D
CALCULATIONS OF THE n-DIMENSIONAL INTEGRALS IN SECTION III-D

D.1 *Proof of Eqs.* (61)–(63)

$$\begin{aligned}
&\int p_{X^n,Y^n}(x^n, y^n) \, \ln\left(\frac{p_{Y^n|X^n}(y^n|x^n)}{q_{Y^n|X^n}(y^n|x^n)}\right) \mathrm{d}x^n \, \mathrm{d}y^n \\
&= \int p_{X^n,Y^n}(x^n, y^n) \, \ln\left(\frac{1}{n} \sum_{i=1}^n \frac{r_{Y|X}(y_i|x_i)}{q_{Y|X}(y_i|x_i)}\right) \mathrm{d}x^n \, \mathrm{d}y^n \\
&= \int_0^\infty \frac{1}{u} \left[e^{-u} - \int p_{X^n,Y^n}(x^n, y^n) \exp\left(-\frac{u}{n} \sum_{i=1}^n \frac{r_{Y|X}(y_i|x_i)}{q_{Y|X}(y_i|x_i)}\right) \mathrm{d}x^n \, \mathrm{d}y^n\right] \mathrm{d}u
\end{aligned} \tag{D.1}$$

$$\begin{aligned}
&= \int_0^\infty \frac{1}{u} \left[e^{-u} - \int \frac{1}{n} \sum_{i=1}^n \left\{\prod_{j \neq i} q_{Y|X}(y_j|x_j)\, p_X(x_j) \cdot r_{Y|X}(y_i|x_i)\, p_X(x_i)\right\}\right. \\
&\qquad \left. \cdot \exp\left(-\frac{u}{n} \sum_{i=1}^n \frac{r_{Y|X}(y_i|x_i)}{q_{Y|X}(y_i|x_i)}\right) \mathrm{d}x^n \, \mathrm{d}y^n\right] \mathrm{d}u
\end{aligned} \tag{D.2}$$

$$= \int_0^\infty \frac{1}{u} \left[e^{-u} - \int \frac{1}{n} \sum_{i=1}^{n} \left\{ \prod_{j \neq i} q_{Y|X}(y_j|x_j)\, p_X(x_j)\, \exp\!\left(-\frac{u}{n} \frac{r_{Y|X}(y_j|x_j)}{q_{Y|X}(y_j|x_j)} \right) \right.\right.$$
$$\left.\left. \cdot\, r_{Y|X}(y_i|x_i)\, p_X(x_i) \exp\!\left(-\frac{u}{n} \frac{r_{Y|X}(y_i|x_i)}{q_{Y|X}(y_i|x_i)} \right) \right\} \mathrm{d}x^n\, \mathrm{d}y^n \right] \mathrm{d}u \tag{D.3}$$

$$= \int_0^\infty \frac{1}{u} \left[e^{-u} - \frac{1}{n} \sum_{i=1}^{n} \left\{ \prod_{j \neq i} \int q_{Y|X}(y_j|x_j)\, p_X(x_j)\, \exp\!\left(-\frac{u}{n} \frac{r_{Y|X}(y_j|x_j)}{q_{Y|X}(y_j|x_j)} \right) \mathrm{d}x_j\, \mathrm{d}y_j \right.\right.$$
$$\left.\left. \cdot \int r_{Y|X}(y_i|x_i)\, p_X(x_i) \exp\!\left(-\frac{u}{n} \frac{r_{Y|X}(y_i|x_i)}{q_{Y|X}(y_i|x_i)} \right) \mathrm{d}x_i\, \mathrm{d}y_i \right\} \right] \mathrm{d}u \tag{D.4}$$

$$= \int_0^\infty \frac{1}{u} \left[e^{-u} - \frac{1}{n} \sum_{i=1}^{n} \left\{ \left(\int q_{Y|X}(y|x)\, p_X(x)\, \exp\!\left(-\frac{u}{n} \frac{r_{Y|X}(y|x)}{q_{Y|X}(y|x)} \right) \mathrm{d}x\, \mathrm{d}y \right)^{n-1} \right.\right.$$
$$\left.\left. \cdot \int r_{Y|X}(y|x)\, p_X(x) \exp\!\left(-\frac{u}{n} \frac{r_{Y|X}(y|x)}{q_{Y|X}(y|x)} \right) \mathrm{d}x\, \mathrm{d}y \right\} \right] \mathrm{d}u \tag{D.5}$$

$$= \int_0^\infty \frac{1}{u} \left[e^{-u} - \left(\int q_{Y|X}(y|x)\, p_X(x)\, \exp\!\left(-\frac{u}{n} \frac{r_{Y|X}(y|x)}{q_{Y|X}(y|x)} \right) \mathrm{d}x\, \mathrm{d}y \right)^{n-1} \right.$$
$$\left. \cdot \int r_{Y|X}(y|x)\, p_X(x) \exp\!\left(-\frac{u}{n} \frac{r_{Y|X}(y|x)}{q_{Y|X}(y|x)} \right) \mathrm{d}x\, \mathrm{d}y \right] \mathrm{d}u \tag{D.6}$$

$$= \int_0^\infty \frac{1}{u} \left[e^{-u} - f^{n-1}\!\left(\frac{u}{n}\right) g\!\left(\frac{u}{n}\right) \right] \mathrm{d}u, \tag{D.7}$$

where $f(\cdot)$ and $g(\cdot)$ are defined in (62) and (63), respectively. Consequently, $f(0) = g(0) = 1$, and $0 \leq f(u), g(u) \leq 1$ for all $u > 0$.

D.2 *Proof of Eq.* (64)

$$\int p_{X^n,Y^n}(x^n, y^n)\, \ln q_{Y^n|X^n}(y^n|x^n)\, \mathrm{d}x^n\, \mathrm{d}y^n$$
$$= \int p_{X^n,Y^n}(x^n, y^n) \sum_{j=1}^{n} \ln q_{Y|X}(y_j|x_j)\, \mathrm{d}x^n\, \mathrm{d}y^n \tag{D.8}$$
$$= \int \prod_{\ell=1}^{n} p_X(x_\ell) \cdot \frac{1}{n} \sum_{i=1}^{n} \left\{ \prod_{\ell \neq i} q_{Y|X}(y_\ell|x_\ell)\, r_{Y|X}(y_i|x_i) \right\} \sum_{j=1}^{n} \ln q_{Y|X}(y_j|x_j)\, \mathrm{d}x^n\, \mathrm{d}y^n \tag{D.9}$$
$$= \int \prod_{\ell=1}^{n} p_X(x_\ell) \cdot \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \left\{ \prod_{\ell \neq i} q_{Y|X}(y_\ell|x_\ell) \cdot r_{Y|X}(y_i|x_i)\, \ln q_{Y|X}(y_j|x_j) \right\} \mathrm{d}x^n\, \mathrm{d}y^n \tag{D.10}$$

$$= \frac{1}{n} \int_{\mathcal{X}^n} \prod_{\ell=1}^{n} p_X(x_\ell) \left(\sum_{i=1}^{n} \sum_{j=1}^{n} \int_{\mathcal{Y}^n} \prod_{\ell \neq i} q_{Y|X}(y_\ell|x_\ell) \cdot r_{Y|X}(y_i|x_i) \, \ln q_{Y|X}(y_j|x_j) \, \mathrm{d}y^n \right) \mathrm{d}x^n \tag{D.11}$$

We next calculate the inner integral on the right–hand side of (D.11). For $i = j$,

$$\begin{aligned} &\int_{\mathcal{Y}^n} \prod_{\ell \neq i} q_{Y|X}(y_\ell|x_\ell) \cdot r_{Y|X}(y_i|x_i) \, \ln q_{Y|X}(y_j|x_j) \, \mathrm{d}y^n \\ &= \prod_{\ell \neq i} \int_{\mathcal{Y}} q_{Y|X}(y_\ell|x_\ell) \, \mathrm{d}y_\ell \cdot \int_{\mathcal{Y}} r_{Y|X}(y_i|x_i) \, \ln q_{Y|X}(y_i|x_i) \, \mathrm{d}y_i \\ &= \int_{\mathcal{Y}} r_{Y|X}(y|x_i) \, \ln q_{Y|X}(y|x_i) \, \mathrm{d}y, \end{aligned} \tag{D.12}$$

else,

$$\begin{aligned} &\int_{\mathcal{Y}^n} \prod_{\ell \neq i} q_{Y|X}(y_\ell|x_\ell) \cdot r_{Y|X}(y_i|x_i) \, \ln q_{Y|X}(y_j|x_j) \, \mathrm{d}y^n \\ &= \prod_{\ell \notin \{i,j\}} \int_{\mathcal{Y}} q_{Y|X}(y_\ell|x_\ell) \, \mathrm{d}y_\ell \cdot \int_{\mathcal{Y}} q_{Y|X}(y_j|x_j) \, \ln q_{Y|X}(y_j|x_j) \, \mathrm{d}y_j \cdot \int_{\mathcal{Y}} r_{Y|X}(y_i|x_i) \, \mathrm{d}y_i \\ &= \int_{\mathcal{Y}} q_{Y|X}(y|x_j) \, \ln q_{Y|X}(y|x_j) \, \mathrm{d}y. \end{aligned} \tag{D.13}$$

Hence, from (D.11)–(D.13),

$$\begin{aligned} &\int p_{X^n,Y^n}(x^n, y^n) \, \ln q_{Y^n|X^n}(y^n|x^n) \, \mathrm{d}x^n \, \mathrm{d}y^n \\ &= \frac{1}{n} \int_{\mathcal{X}^n} \prod_{\ell=1}^{n} p_X(x_\ell) \left(\sum_{i=1}^{n} \int_{\mathcal{Y}} r_{Y|X}(y|x_i) \, \ln q_{Y|X}(y|x_i) \, \mathrm{d}y \right. \\ &\qquad \left. + \sum_{i=1}^{n} \sum_{j \neq i} \int_{\mathcal{Y}} q_{Y|X}(y|x_j) \, \ln q_{Y|X}(y|x_j) \, \mathrm{d}y \right) \mathrm{d}x^n \\ &= \frac{1}{n} \left[\sum_{i=1}^{n} \left\{ \prod_{\ell \neq i} \int_{\mathcal{X}} p_X(x_\ell) \, \mathrm{d}x_\ell \cdot \int_{\mathcal{X} \times \mathcal{Y}} r_{Y|X}(y|x_i) \, \ln q_{Y|X}(y|x_i) \, p_X(x_i) \, \mathrm{d}x_i \, \mathrm{d}y \right\} \right. \\ &\qquad \left. + \sum_{i=1}^{n} \sum_{j \neq i} \left\{ \prod_{\ell \neq j} \int_{\mathcal{X}} p_X(x_\ell) \, \mathrm{d}x_\ell \cdot \int_{\mathcal{X} \times \mathcal{Y}} p_X(x_j) \, q_{Y|X}(y|x_j) \, \ln q_{Y|X}(y|x_j) \, \mathrm{d}x_j \, \mathrm{d}y \right\} \right] \\ &= \frac{1}{n} \left[\sum_{i=1}^{n} \int_{\mathcal{X} \times \mathcal{Y}} r_{Y|X}(y|x) \, \ln q_{Y|X}(y|x) \, p_X(x) \, \mathrm{d}x \, \mathrm{d}y \right. \\ &\qquad \left. + \sum_{i=1}^{n} \sum_{j \neq i} \int_{\mathcal{X} \times \mathcal{Y}} p_X(x) \, q_{Y|X}(y|x) \, \ln q_{Y|X}(y|x) \, \mathrm{d}x \, \mathrm{d}y \right] \end{aligned}$$

$$= \int_{\mathcal{X}\times\mathcal{Y}} p_X(x)\, r_{Y|X}(y|x)\, \ln q_{Y|X}(y|x)\, \mathrm{d}x\, \mathrm{d}y$$
$$+ (n-1) \int_{\mathcal{X}\times\mathcal{Y}} p_X(x)\, q_{Y|X}(y|x)\, \ln q_{Y|X}(y|x)\, \mathrm{d}x\, \mathrm{d}y. \tag{D.14}$$

D.3 *Proof of Eqs.* (66)–(72)

$$\begin{aligned} p_{Y^n}(y^n) &= \int p_{Y^n|X^n}(y^n|x^n)\, p_{X^n}(x^n)\, \mathrm{d}x^n \\ &= \frac{1}{n}\sum_{i=1}^{n}\left\{\prod_{j\neq i}\int q_{Y|X}(y_j|x_j)\, p_X(x_j)\, \mathrm{d}x_j \cdot \int r_{Y|X}(y_i|x_i) P_X(x_i)\, \mathrm{d}x_i\right\} \\ &= \frac{1}{n}\sum_{i=1}^{n}\left\{\prod_{j\neq i} v(y_j)\cdot w(y_i)\right\} \\ &= \prod_{j=1}^{n} v(y_j)\cdot \frac{1}{n}\sum_{i=1}^{n}\frac{w(y_i)}{v(y_i)}, \quad \forall\, y^n \in \mathcal{Y}^n, \end{aligned} \tag{D.15}$$

where $v(\cdot)$ and $w(\cdot)$ are probability densities on $\mathcal{Y}$, as defined in (67) and (68), respectively. This proves (66).

We next prove (69), which is used to calculate the entropy of Y^n with the density $p_{Y^n}(\cdot)$ in (D.15). In view of the integral representation of the logarithmic function in (1), and by interchanging the order of the integrations, we get that for a positive random variable Z

$$\begin{aligned} \mathbb{E}\big\{Z\ln Z\big\} &= \int_0^\infty \frac{1}{u}\cdot \mathbb{E}\big\{Z\left(e^{-u}-e^{-uZ}\right)\big\}\, \mathrm{d}u \\ &= \int_0^\infty \frac{\mathbb{E}\big\{Z\big\}\, e^{-u} - \mathbb{E}\big\{Ze^{-uZ}\big\}}{u}\, \mathrm{d}u \\ &= \int_0^\infty \frac{M_Z'(0)\, e^{-u} - M_Z'(-u)}{u}\, \mathrm{d}u, \end{aligned} \tag{D.16}$$

which proves (69). Finally, we prove (70). In view of (D.15),

$$\begin{aligned} h(Y^n) &= -\int p_{Y^n}(y^n)\, \ln p_{Y^n}(y^n)\, \mathrm{d}y^n \\ &= -\int \prod_{j=1}^{n} v(y_j)\cdot \frac{1}{n}\sum_{i=1}^{n}\frac{w(y_i)}{v(y_i)}\cdot\left[\ln\left(\prod_{j=1}^{n} v(y_j)\right) + \ln\left(\frac{1}{n}\sum_{i=1}^{n}\frac{w(y_i)}{v(y_i)}\right)\right]\mathrm{d}y^n \\ &= -\int \prod_{j=1}^{n} v(y_j)\cdot \frac{1}{n}\sum_{i=1}^{n}\frac{w(y_i)}{v(y_i)}\cdot\left[\sum_{j=1}^{n}\ln v(y_j) + \ln\left(\frac{1}{n}\sum_{i=1}^{n}\frac{w(y_i)}{v(y_i)}\right)\right]\mathrm{d}y^n \\ &= -\int \prod_{\ell=1}^{n} v(y_\ell)\cdot \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{w(y_i)\, \ln v(y_j)}{v(y_i)}\, \mathrm{d}y^n \end{aligned}$$

$$-\int \prod_{j=1}^{n} v(y_j) \cdot \frac{1}{n} \sum_{i=1}^{n} \frac{w(y_i)}{v(y_i)} \cdot \ln\left(\frac{1}{n} \sum_{i=1}^{n} \frac{w(y_i)}{v(y_i)} \right) \mathrm{d}y^n. \tag{D.17}$$

A calculation of the first integral on the right–hand side of (D.17) gives

$$\begin{aligned}
&\int \prod_{\ell=1}^{n} v(y_\ell) \cdot \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{w(y_i)\ \ln v(y_j)}{v(y_i)}\ \mathrm{d}y^n \\
&= \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \int \prod_{\ell=1}^{n} v(y_\ell) \cdot \frac{w(y_i)\ \ln v(y_j)}{v(y_i)}\ \mathrm{d}y^n \\
&= \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \int \prod_{\ell \neq i} v(y_\ell) \cdot w(y_i)\ \ln v(y_j)\, \mathrm{d}y^n.
\end{aligned} \tag{D.18}$$

For $i = j$, the inner integral on the right–hand side of (D.18) satisfies

$$\begin{aligned}
&\int \prod_{\ell \neq i} v(y_\ell) \cdot w(y_i)\ \ln v(y_j)\, \mathrm{d}y^n \\
&= \prod_{\ell \neq i} \int v(y_\ell)\, \mathrm{d}y_\ell \cdot \int w(y_i)\ \ln v(y_i)\, \mathrm{d}y_i \\
&= \int w(y)\ \ln v(y)\, \mathrm{d}y,
\end{aligned} \tag{D.19}$$

and for $i \neq j$,

$$\begin{aligned}
&\int \prod_{\ell \neq i} v(y_\ell) \cdot w(y_i)\ \ln v(y_j)\, \mathrm{d}y^n \\
&= \prod_{\ell \neq i,j} \int v(y_\ell)\, \mathrm{d}y_\ell \cdot \int w(y_i)\, \mathrm{d}y_i \cdot \int v(y_j)\ \ln v(y_j)\, \mathrm{d}y_j \\
&= \int v(y)\ \ln v(y)\, \mathrm{d}y.
\end{aligned} \tag{D.20}$$

Therefore combining (D.18)–(D.20) gives

$$\begin{aligned}
&\int \prod_{\ell=1}^{n} v(y_\ell) \cdot \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{w(y_i)\ \ln v(y_j)}{v(y_i)}\ \mathrm{d}y^n \\
&= \int w(y)\ \ln v(y)\, \mathrm{d}y + (n-1) \int v(y)\ \ln v(y)\, \mathrm{d}y.
\end{aligned} \tag{D.21}$$

Finally, we calculate the second integral on the right–hand side of (D.17). Let μ_n be the probability density function defined as

$$\mu_n(y^n) := \prod_{j=1}^{n} v(y_j), \quad y^n \in \mathcal{Y}^n, \tag{D.22}$$

and let

$$Z := \frac{1}{n}\sum_{i=1}^{n}\frac{w(V_i)}{v(V_i)} \tag{D.23}$$

where $\{V_i\}_{i=1}^n$ are i.i.d. $\mathcal{Y}$-valued random variables with a probability density function v. Then, in view of (69), the second integral on the right–hand side of (D.17) satisfies

$$\begin{aligned}
&\int\prod_{j=1}^{n}v(y_j)\cdot\frac{1}{n}\sum_{i=1}^{n}\frac{w(y_i)}{v(y_i)}\cdot\ln\left(\frac{1}{n}\sum_{i=1}^{n}\frac{w(y_i)}{v(y_i)}\right)\,\mathrm{d}y^n\\
&=\mathbb{E}\big\{Z\ln Z\big\}\\
&=\int_0^\infty\frac{M_Z'(0)\,e^{-u}-M_Z'(-u)}{u}\,\mathrm{d}u.
\end{aligned}\tag{D.24}$$

The MGF of Z is equal to

$$\begin{aligned}
M_Z(u)&=\int_{\mathcal{Y}^n}\prod_{i=1}^{n}v(r_i)\,\exp\left(\frac{u}{n}\sum_{i=1}^{n}\frac{w(r_i)}{v(r_i)}\right)\mathrm{d}\underline{r}\\
&=\prod_{i=1}^{n}\int_{\mathcal{Y}}v(r_i)\,\exp\left(\frac{u}{n}\,\frac{w(r_i)}{v(r_i)}\right)\mathrm{d}r_i\\
&=K^n\Big(\frac{u}{n}\Big),
\end{aligned}\tag{D.25}$$

where

$$K(u):=\int_{\mathcal{Y}}v(y)\,\exp\left(\frac{u\,w(y)}{v(y)}\right)\mathrm{d}y,\quad\forall\,u\in\mathbb{R},\tag{D.26}$$

and consequently, (D.26) yields

$$\begin{aligned}
M_Z'(u)&=K^{n-1}\Big(\frac{u}{n}\Big)\,K'\Big(\frac{u}{n}\Big)\\
&=\left(\int v(y)\,\exp\left(\frac{u\,w(y)}{v(y)}\right)\mathrm{d}y\right)^{n-1}\int w(y)\,\exp\left(\frac{u\,w(y)}{v(y)}\right)\mathrm{d}y,
\end{aligned}\tag{D.27}$$

and

$$M_Z'(0)=1.\tag{D.28}$$

Therefore, combining (D.24)–(D.28) gives the following single-letter expression for the second multi-dimensional integral on the right–hand side of (D.17):

$$\int\prod_{j=1}^{n}v(y_j)\cdot\frac{1}{n}\sum_{i=1}^{n}\frac{w(y_i)}{v(y_i)}\cdot\ln\left(\frac{1}{n}\sum_{i=1}^{n}\frac{w(y_i)}{v(y_i)}\right)\mathrm{d}y^n=\int_0^\infty\frac{1}{u}\left[e^{-u}-t^{n-1}\Big(\frac{u}{n}\Big)\,s\Big(\frac{u}{n}\Big)\right]\mathrm{d}u,\tag{D.29}$$

where the functions $s(\cdot)$ and $t(\cdot)$ are defined in (71) and (72), respectively. Combining (D.17), (D.21) and (D.29) gives (70).

D.4 *Specialization to a BSC with Jamming*

In the BSC example considered, we have $\mathcal{X} = \mathcal{Y} = \{0, 1\}$, and

$$r_{Y|X}(y|x) = \varepsilon\, 1\{x \neq y\} + (1 - \varepsilon)\, 1\{x = y\}, \tag{D.30}$$

$$q_{Y|X}(y|x) = \delta\, 1\{x \neq y\} + (1 - \delta)\, 1\{x = y\}, \tag{D.31}$$

where $1\{\text{relation}\}$ is the indicator function that is equal to 1 if the relation holds, and to zero otherwise. Recall that we assume $0 < \delta < \varepsilon \leq \frac{1}{2}$. Let

$$p_X(0) = p_X(1) = \tfrac{1}{2}, \tag{D.32}$$

be the binary symmetric source (BSS). From (62) and (63), for $u \geq 0$,

$$\begin{aligned} f(u) &= \sum_{x,y} p_X(x)\, q_{Y|X}(y|x)\, \exp\biggl(-\frac{u\, r_{Y|X}(y|x)}{q_{Y|X}(y|x)}\biggr) \\ &= (1 - \delta)\, \exp\biggl(-\frac{(1 - \varepsilon)u}{1 - \delta}\biggr) + \delta\, \exp\biggl(-\frac{\varepsilon u}{\delta}\biggr), \end{aligned} \tag{D.33}$$

$$\begin{aligned} g(u) &= \sum_{x,y} p_X(x)\, r_{Y|X}(y|x)\, \exp\biggl(-\frac{u\, r_{Y|X}(y|x)}{q_{Y|X}(y|x)}\biggr) \\ &= (1 - \varepsilon)\, \exp\biggl(-\frac{(1 - \varepsilon)u}{1 - \delta}\biggr) + \varepsilon\, \exp\biggl(-\frac{\varepsilon u}{\delta}\biggr). \end{aligned} \tag{D.34}$$

Furthermore, we get from (D.30), (D.31) and (D.32) that

$$\begin{aligned} -\sum_{x,y} p_X(x)\, r_{Y|X}(y|x)\, \ln q_{Y|X}(y|x) &= -\varepsilon \ln \delta - (1 - \varepsilon) \ln(1 - \delta) \\ &= d(\varepsilon\|\delta) + h_{\mathrm{b}}(\varepsilon), \end{aligned} \tag{D.35}$$

and

$$-\sum_{x,y} p_X(x)\, r_{Y|X}(y|x)\, \ln q_{Y|X}(y|x) = h_{\mathrm{b}}(\delta). \tag{D.36}$$

Substituting (D.33)–(D.36) into (65) (where integrals in (65) are replaced by sums) gives

$$H(Y^n|X^n) = d(\varepsilon\|\delta) + h_{\mathrm{b}}(\varepsilon) + (n - 1)\, h_{\mathrm{b}}(\delta) + \int_0^\infty \Bigl[f^{n-1}\Bigl(\frac{u}{n}\Bigr)\, g\Bigl(\frac{u}{n}\Bigr) - e^{-u}\Bigr]\, \frac{\mathrm{d}u}{u}. \tag{D.37}$$

Since the input is a BSS, due to the symmetry of the channel (54), the output is also a BSS. This implies that (in units of nats)

$$H(Y^n) = n \ln 2. \tag{D.38}$$

As a sanity check, we verify it by using (70). From (67) and (68), for $y \in \{0, 1\}$,

$$v(y) = p_X(0)\, q_{Y|X}(y|0) + p_X(1)\, q_{Y|X}(y|1) = \tfrac{1}{2}, \tag{D.39}$$

$$w(y) = p_X(0)\, r_{Y|X}(y|0) + p_X(1)\, r_{Y|X}(y|1) = \tfrac{1}{2}, \tag{D.40}$$

and, from (71) and (72), it consequently follows that

$$s(u) = w(0)\, \exp\left(-\frac{u\, w(0)}{v(0)}\right) + w(1)\, \exp\left(-\frac{u\, w(1)}{v(1)}\right) = e^{-u}, \quad \forall\, u \geq 0, \tag{D.41}$$

and also

$$t(u) = e^{-u}, \quad \forall\, u \geq 0. \tag{D.42}$$

It can be verified that substituting (D.39)–(D.42) into (70) reproduces (D.38). Finally, subtracting (D.37) from (D.38) gives (74).

REFERENCES

[1] E. Arikan, "An inequality on guessing and its application to sequential decoding," *IEEE Transactions on Information Theory*, vol. 42, no. 1, pp. 99–105, January 1996.

[2] E. Arikan and N. Merhav, "Guessing subject to distortion," *IEEE Transactions on Information Theory*, vol. 44, no. 3, pp. 1041–1056, May 1998.

[3] E. Arikan and N. Merhav, "Joint source-channel coding and guessing with application to sequential decoding," *IEEE Transactions on Information Theory*, vol. 44, no. 5, pp. 1756–1769, September 1998.

[4] D. Berend and A. Kontorovich, "On the concentration of the missing mass," *Electronic Communications in Probability*, vol. 18, paper 3, pp. 1–7, 2013.

[5] S. Boztaş, "Comments on "An inequality on guessing and its application to sequential decoding"," *IEEE Transactions on Information Theory*, vol. 43, no. 6, pp. 2062–2063, November 1997.

[6] A. Bracher, E. Hof and A. Lapidoth, "Guessing attacks on distributed–storage systems," *IEEE Transactions on Information Theory*, vol. 65, no. 11, pp. 6975–6998, November 2019.

[7] C. Bunte and A. Lapidoth, "Encoding tasks and Rényi entropy," *IEEE Trans. on Information Theory*, vol. 60, no. 9, pp. 5065–5076, September 2014.

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

[9] T. Courtade and S. Verdú, "Cumulant generating function of codeword lengths in optimal lossless compression," *Proceedings of the 2014 IEEE International Symposium on Information Theory*, pp. 2494–2498, Honolulu, Hawaii, USA, July 2014.

[10] T. Courtade and S. Verdú, "Variable-length lossy compression and channel coding: Non-asymptotic converses via cumulant generating functions," *Proceedings of the 2014 IEEE International Symposium on Information Theory*, pp. 2499–2503, Honolulu, Hawaii, USA, July 2014.

[11] A. Dong, H. Zhang, D. Wu, and D. Yuan, "Logarithmic expectation of the sum of exponential random variables for wireless communication performance evaluation," *Proc. 2015 IEEE 82nd Vehicular Technology Conference*, Boston, MA, USA, September 2015.

[12] S. E. Esipov and T. J. Newman, "Interface growth and Burgers turbulence: the problem of random initial conditions," *Phys. Rev. E*, vol. 48, no. 2, pp. 1046–1050, August 1993.

[13] R. J. Evans, J. Boersma, N. M. Blachman and A. A. Jagers, "The entropy of a Poisson distribution," *SIAM Review*, vol. 30, no. 2, pp. 314–317, June 1988.

[14] I. S. Gradshteyn and I. M. Ryzhik, *Tables of Integrals, Series, and Products*, Eighth Edition, Elsevier, 2014.

[15] M. K. Hanawal and R. Sundaresan, "Guessing revisited: a large deviations approach," *IEEE Transactions on Information Theory*, vol. 57, no. 1, pp. 70–78, January 2011.

[16] C. Knessl, "Integral representations and asymptotic expansions for Shannon and Rényi entropies," *Applied Mathematical Letters*, vol. 11, no. 2, pp. 69–74, 1998.

[17] A. Lapidoth and S. Moser, "Capacity bounds via duality with applications to multiple-antenna systems on flat fading channels," *IEEE Transactions on Information Theory*, vol. 49, no. 10, pp. 24262467, October 2003.

[18] A. Martinez, "Spectral efficiency of optical direct detection," *Journal of the Optical Society of America B*, vol. 24, no. 4, pp. 739–749, April 2007.

[19] M. Mézard and A. Montanari, *Information, Physics, and Computation*, Oxford University Press, New-York, USA, 2009.

[20] N. Merhav, "Lower bounds on exponential moments of the quadratic error in parameter estimation," *IEEE Transactions on Information Theory*, vol. 64, no. 12, pp. 7636–7648, December 2018.

[21] N. Merhav and A. Cohen, "Universal randomized guessing with application to asynchronous decentralized brute-force attacks," *IEEE Transactions on Information Theory*, vol. 66, no. 1, pp. 114–129, January 2020.

[22] N. Merhav and I. Sason, "An integral representation of the logarithmic function with applications in information theory," *Entropy*, vol. 22, no. 1, paper 51, pp. 1–22, January 2020.

[23] F. W. J. Olver, D. W. Lozier, R. F. Boisvert and C. W. Clark, *NIST Handbook of Mathematical Functions*, NIST (National Institute of Standards and Technology) and Cambridge University Press, New York, USA, 2010.

[24] M. Raginsky and I. Sason, *Concentration of Measure Inequalities in Information Theory, Communications and Coding: Third Edition*, pp. 1–261, Foundations and Trends in Communications and Information Theory, NOW Publishers, Delft, 2019.

[25] A. Rajan and C. Tepedelenlioğlu, "Stochastic ordering of fading channels through the Shannon transform," *IEEE Transactions on Information Theory*, vol. 61, no. 4, pp. 1619–1628, April 2015.

[26] A. Rényi, "On measures of entropy and information," *Proceedings of the Fourth Berkeley Symposium on Probability Theory and Mathematical Statistics*, pp. 547–561, Berkeley, California, USA, 1961.

[27] S. Salamatian, W. Huleihel, A. Beirami, A. Cohen and M. Médard, "Why botnets work: distributed brute-force attacks need no synchronization," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 9, pp. 2288–2299, September 2019.

[28] I. Sason and S. Verdú, "Improved bounds on lossless source coding and guessing moments via Rényi measures," *IEEE Transactions on Information Theory*, vol. 64, no. 6, pp. 4323–4346, June 2018.

[29] I. Sason, "Tight bounds on the Rényi entropy via majorization with applications to guessing and compression," *Entropy*, vol. 20, no. 12, paper 896, pp. 1–25, November 2018.

[30] J. Song, S. Still, R. D. H. Rojas, I. P. Castillo, and M. Marsili, "Optimal work extraction and mutual information in a generalized Szilárd engine," arXiv:1910.0419v1, October 9, 2019.

[31] R. Sundaresan, "Guessing under source uncertainty," *IEEE Transactions on Information Theory*, vol. 53, no. 1, pp. 269–287, January 2007.

[32] R. Sundaresan, "Guessing based on length functions," *Proceedings of the 2007 IEEE International Symposium on Information Theory*, pp. 716–719, Nice, France, June 2007.

[33] P. H. Zadeh and R. Hosseini, "Expected logarithm of central quadratic form and its use in KL-divergence of some distributions," *Entropy*, vol. 18, no. 8, paper 288, pp. 1–25, August 2016.