Thinning, photonic beamsplitting, and a general discrete Entropy power Inequality

Thinning, photonic beamsplitting, and a general discrete entropy power inequality

Saikat Guha
Raytheon BBN Technologies
Email: sguha@bbn.com

Jeffrey H. Shapiro
Massachusetts Institute of Technology
Email: jhs@mit.edu

Raúl García-Patrón Sánchez
Université Libre de Bruxelles
Email: rgarciap@ulb.ac.be

Abstract—**Many partially-successful attempts have been made to find the most natural discrete-variable version of Shannon's entropy power inequality (EPI). We develop an axiomatic framework from which we deduce the natural form of a discrete-variable EPI and an associated entropic monotonicity in a discrete-variable central limit theorem. In this discrete EPI, the geometric distribution, which has the maximum entropy among all discrete distributions with a given mean, assumes a role analogous to the Gaussian distribution in Shannon's EPI. The entropy power of X is defined as the mean of a geometric random variable with entropy $H(X)$. The crux of our construction is a discrete-variable version of Lieb's scaled addition $X \boxplus_\eta Y$ of two discrete random variables X and Y with $\eta \in (0,1)$. We discuss the relationship of our discrete EPI with recent work of Yu and Johnson who developed an EPI for a restricted class of random variables that have ultra-log-concave (ULC) distributions. Even though we leave open the proof of the aforesaid natural form of the discrete EPI, we show that this discrete EPI holds true for variables with arbitrary discrete distributions when the entropy power is redefined as $e^{H(X)}$ in analogy with the continuous version. Finally, we show that our conjectured discrete EPI is a special case of the yet-unproven Entropy Photon-number Inequality (EPnI), which assumes a role analogous to Shannon's EPI in capacity proofs for Gaussian bosonic (quantum) channels.**

I. Introduction

The Entropy Power Inequality (EPI) for statistically independent real *continuous*-valued random variables X and Y,

$$v(X+Y) \geq v(X) + v(Y) \tag{1}$$

was first stated by Shannon [1], where $v(X) = e^{2h(X)}/(2\pi e)$ is the *entropy power* of X, with $h(X) = -\int_{-\infty}^{\infty} p_X(x) \log p_X(x) dx$ being the differential entropy of X. Equality in (1) holds if and only if X and Y are Gaussian. Using $\mathcal{E}_G(t) \equiv \frac{1}{2}\log(2\pi e t)$ the entropy of a Gaussian random variable with variance t, we get $v(X) = \mathcal{E}_G^{-1}(h(X))$. In other words, the entropy power of X is the variance of a Gaussian random variable that has the same differential entropy as that of X. The EPI was subsequently proved by Stam [2] and Blachman [3]. The EPI (1) has proven to be a powerful tool in information theory. It has found use in converse proofs of various Gaussian channel coding theorems involving continuous-valued sources and channels, especially in the respective converse proofs. Some prominent examples are: the capacity region of the scalar Gaussian broadcast channel [4], its generalization to the MIMO case [5], [6], the private capacity of the Gaussian wiretap channel [7], capacity region of the Gaussian interference channel [8], Gaussian source multiple-description problem [9], [10], and rate distortion region for multi-terminal Gaussian source coding [11].

Due to this success of Shannon's EPI, it is natural to speculate if there is an EPI for statistically-independent discrete-valued random variables X and Y defined on the set of natural numbers $\mathbb{N}_0 \equiv \{0, 1, \ldots\}$, which has analogous implications as above in discrete-valued information and coding theorems.

There have been several attempts to provide an EPI for discrete random variables. It is simple to see that (1) does not hold as is when X and Y are arbitrary discrete random variables and the differential entropy $h(X)$ is replaced by the discrete Shannon entropy $H(X)$; a simple counterexample occurring for when X and Y are deterministic. EPIs with Bernoulli sources were proven where the "$+$" operation in (1) was taken modulo 2 [12], [13], [14], [15]. Harremoës and Vignat retained the use of "$+$" as integer addition, and proved that (1) holds when X and Y are independent binomial $\text{Bin}(n_X, 1/2)$ and $\text{Bin}(n_Y, 1/2)$ variables, upon redefining $v(X)$ in (1) as $e^{2H(X)}/(2\pi e)$, with $H(X)$ being the discrete entropy [16]. A discrete EPI was proven for arbitrary i.i.d. integer-valued random variables, albeit of a non-standard asymmetric form [17]. Yu and Johnson developed a natural form of the discrete EPI but one that holds only for ultra-log concave (ULC) distributed random variables [18], [19], [20], [21]. Recently, Woo and Madiman showed that an inequality very similar to (1) holds for uniform distributions over finite subsets of the integers [22].

In this paper, we present an axiomatic framework to develop a natural generalization of the EPI for arbitrary discrete random variables and an associated central limit theorem (CLT). We begin with an overview of Shannon's EPI in Section II. Section III provides an intuitive development of our discrete EPI. In Section IV, we provide proofs leveraging the algebra of quantum optics, and a close connection of our discrete EPI to a quantum EPI that naturally emerges in capacity proofs for transmitting classical information over bosonic channels [23].

II. Review of the Entropy Power Inequality

Using the following scaling identity for entropy power:

$$v(\sqrt{\beta}X) = \beta v(X), \tag{2}$$

it is straightforward to see that Ineq. (1) can be restated as:

$$v(X \boxplus_\eta Y) \geq \eta v(X) + (1-\eta) v(Y), \text{ where} \tag{3}$$

$$X \boxplus_\eta Y \equiv \sqrt{\eta} X + \sqrt{1-\eta}\, Y \tag{4}$$

is a scaled addition operation. The 'linear' form of the EPI,

$$h(X \boxplus_\eta Y) \geq \eta h(X) + (1-\eta) h(Y), \tag{5}$$

i.e., the fact that differential entropy is concave with respect to normalized linear combinations, was first stated by Lieb [24], a rigorous proof of which, and the proof of the fact that (5) is equivalent to (1), was later provided by Dembo, Cover and Thomas [25]. That (1), hence (3) implies (5) follows simply from the application of $\mathcal{E}_G(\cdot)$ to both sides of (3), and using the concavity of the logarithm function. The converse implication requires the entropy of a scaled random variable, $h(aX) = h(X) + \log|a|$ [26]. Consider random variables $X^* = X/\sqrt{\eta}$ and $Y^* = Y/\sqrt{1-\eta}$, with $\eta = v(X)/\left(v(X) + v(Y)\right)$. Using the expression for $h(aX)$ stated above, it readily follows that $h(X^*) = h(Y^*)$. We then have, using (5),

$$\begin{aligned} h(X+Y) &= h\left(\sqrt{\eta} X^* + \sqrt{1-\eta}\, Y^*\right) && (6) \\ &\geq \eta h(X^*) + (1-\eta) h(Y^*) = h(X^*). && (7) \end{aligned}$$

Applying $\mathcal{E}_G^{-1}$ to both sides, and using (2), yields the EPI (1).

Scaling plays an important role in many proofs of the EPI. The fact that the EPI can be stated in terms of scaled random variables as in inequality (7) was implicit in Verdú and Guo's proof of the EPI [26], but Johnson and Yu later realized the significance of this form to construct an EPI for ULC discrete random variables, as we will describe in Section III [21].

Artstein, Ball, Barthe and Naor proved a stronger form of the EPI (1) for sums of independent continuous random variables [27], a special case of which was the first rigorous proof of *monotonicity* of the convergence of differential entropy in the CLT, i.e., for i.i.d. X_i, the entropy of the normalized sum $h\left(Y_n\right)$, with $Y_n = \sum_{i=1}^n X_i/\sqrt{n}$, is monotone increasing in n, and converges to the entropy of the Gaussian (which has the maximum entropy for a given variance) as $n \to \infty$. Note that $H\left(X_1 + X_2\right)/\sqrt{2} \geq H(X_1)$ for i.i.d. X_1 and X_2 follows from (5), repeated application of which shows that $h\left(Y_{2^k}\right)$ is nondecreasing in k. This cruder version of monotonicity is already sufficient to prove the CLT [28]. Alternative proofs of monotonicity of $h(Y_n), \forall n \in \mathbb{N}_0$ were later given by Tulino and Verdú [29], and by Madiman and Barron [30].

III. EPI FOR DISCRETE RANDOM VARIABLES

A. An axiomatic development of the discrete EPI

The main ingredient needed for a natural discrete-variable generalization of Shannon's EPI, i.e., (3) and (5) is a:

- **scaled addition**: An appropriate definition of $X \boxplus_\eta Y$ for X and Y both defined on $\mathbb{N}_0$, and $\eta \in (0,1)$.

Further, the definition of $\boxplus_\eta$ should be extendable to a random vector $\boldsymbol{X} = (X_1, \ldots, X_n)$, i.e., $\boxplus_{\boldsymbol{\eta}} \boldsymbol{X}$, where $\boldsymbol{\eta} \equiv (\eta_1, \ldots, \eta_n)$ with $\eta_i \geq 0$ and $\sum_{i=1}^n \eta_i = 1$, such that: (1) it reduces to the bivariate case for $n = 2$, viz., $\boxplus_{(\eta, 1-\eta)}(X_1, X_2) = X_1 \boxplus_\eta X_2$; (2) it is commutative in the sense that $\boxplus_{\boldsymbol{\eta}} \boldsymbol{X}$ is invariant under an arbitrary (yet identical) permutation of the entries of $\boldsymbol{\eta}$ and $\boldsymbol{X}$; and (3) it is well behaved under a CLT:

- ***limiting* distribution**: A distribution $p_{L,\lambda}[k]$, $k \in \mathbb{N}_0$ should exist that can be defined solely as a function of its mean λ, and is the limiting distribution in a CLT under $\boxplus_{\boldsymbol{\eta}}$ addition. In other words, $p_{L,\lambda}[k]$ is the distribution of $Y_n \equiv \boxplus_{(\frac{1}{n}, \ldots, \frac{1}{n})} \boldsymbol{X}$, as $n \to \infty$, for i.i.d. arbitrarily-distributed λ-mean random variables $\boldsymbol{X} \equiv \{X_i\}$.

For a complete analogy with Shannon's EPI, one would want $H(Y_n)$, the entropy of Y_n, to be monotonically increasing in $n \in \{1, 2, \ldots\}$. Thus, $p_{L,\lambda}$ should be the distribution with the maximum entropy for a given mean λ. We already know that the geometric distribution $p_{L,\lambda}[k] = (1+\lambda)^{-1}\left(\lambda/(1+\lambda)\right)^k$, $k \in \mathbb{N}_0$, has this property. So, we would like the $\boxplus_{\boldsymbol{\eta}}$ operation for which the above CLT holds with the geometric distribution being the limiting distribution. One would then define:

- **entropy power**: $V(X)$ of X as the mean of a random variable with distribution $p_{L,\lambda}$ that has entropy $H(X)$.

Note that in order for the above to make sense, the entropy of the limiting distribution $p_{L,\lambda}$ should be monotonic (increasing) in its mean λ. With the above, the following should hold:

- **entropy power inequality**: The EPI for discrete random variables, i.e., (3) and (5) should hold with the Shannon entropy power $v(X)$ replaced by the discrete entropy power $V(X)$ and the differential entropy $h(X)$ replaced by the discrete Shannon entropy $H(X)$. Further, equality in both aforesaid forms of the discrete EPI should hold when X and Y both are distributed according to the limiting distribution $p_{L,\lambda}$ (possibly with different means).

The above discussion suggests that once we have the 'correct' definition of the $\boxplus_{\boldsymbol{\eta}}$ operation for discrete random variables (i.e., one that is well behaved under the CLT with the limiting distribution being geometric), one would immediately obtain the natural discrete generalization of Shannon's EPI.

In the above framework for discrete-valued random variables, we chose to peg the definitions to the mean as opposed to the variance (as in the case of continuous random variables). In the discrete case, the law of small numbers (see footnote 1) and the corresponding maximum entropy property both require 'thinning' the mean, whereas in the continuous case, the central limit theorem requires the thinning of the variance, which is achieved by multiplication by $\sqrt{\eta}$. Many have realized that the Rényi thinning operation T_η is the natural discrete-variable equivalent of multiplication by $\sqrt{\eta}$, and has the desirable effect of thinning the mean by a factor η [18], [19], [20], [21], [31].

Definition 1: Given $\eta \in (0,1)$, and $X \in \mathbb{N}_0$, the random variable $Y \in \mathbb{N}_0$, obtained by η-thinning of X (denoted, $Y = T_\eta X$), has the distribution of the sum $\sum_{n=1}^{X} Z_n$, where Z_i are binary $\{0,1\}$ valued i.i.d. Bernoulli$(\eta, 1-\eta)$ random variables that are independent of X. The p.m.f. of Y is: $p_Y[n] = \sum_{k=n}^{\infty} p_X[k] \binom{k}{n} \eta^n (1-\eta)^{k-n}$.

B. Yu and Johnson's prior work on the discrete EPI

Yu and Johnson developed a promising line of approach to the discrete EPI for ultra-log concave (ULC) random variables (i.e., those with p.m.f.s $p_X[n], n \in \mathbb{N}_0$, for which $np_X[n]/p_X[n-1]$ is decreasing as n increases). They defined the scaled addition

$$X \boxplus_\eta Y = T_\eta X + T_{1-\eta} Y, \tag{8}$$

which extends naturally to multiple variables and is well behaved under the CLT with the associated limiting distribution being Poisson. For i.i.d. mean-λ $\{X_i\}$ with p.m.f. $p_X[n]$, $Y_n \equiv \boxplus_{(\frac{1}{n},\ldots,\frac{1}{n})} \boldsymbol{X} = \sum_{i=1}^n T_{1/n} X_i = T_{1/n}\left(\sum_{i=1}^n X_i\right)$ converges to the Poisson distribution of mean λ as $n \to \infty$ [1]. Since the limiting distribution is not geometric, the $\boxplus_\eta$ operation in (8) does not satisfy the criteria in Section III-A. However, within the class of all ULC random variables of mean λ, the Poisson(λ) random variable maximizes the entropy [18]. Furthermore, $H(Y_n)$ increases monotonically in n if $p_X[n]$ is ULC [19]. Motivated by this, Yu and Johnson defined entropy power $V_p(X) = \mathcal{E}_p^{-1}(H(X))$ in terms of the entropy $\mathcal{E}_p(\lambda)$ of a Poisson(λ) random variable, with the hope that the straightforward equivalents (1), (3) and (5) would hold, with X and Y restricted to ULC random variables. They proved that the linear form (concavity of entropy) (5) holds [19], i.e.,

$$H(X \boxplus_\eta Y) \geq \eta H(X) + (1-\eta) H(Y), \tag{9}$$

for all ULC independent X and Y. This was the first major step towards a discrete EPI. The equivalents of (1) and (3),

$$\begin{aligned} V_p(X+Y) &\geq V_p(X) + V_p(Y), \text{ and} && (10)\\ V_p(X \boxplus_\eta Y) &\geq \eta V_p(X) + (1-\eta) V_p(Y) && (11) \end{aligned}$$

were naturally conjectured [19], but were shown later *not* to hold in general, even for ULC X and Y [21].

The key step in going from the EPI (1) to the linear form (5) was the scaling identity for the entropy power (2). If such an identity were to hold for any ULC variable X, i.e., $V_p(T_\eta X) = \eta V_p(X)$, then (9) would imply the conjectured discrete EPI (11). But since the conjecture was found to be false, the scaling identity above cannot be true. However, the following one-sided version of it was proved for ULC X:

$$V_p(T_\eta X) \geq \eta V_p(X), \ \eta \in (0,1). \tag{12}$$

which was termed the restricted discrete EPI [21]. Even though (11) does not hold in general, the restated form (7) does hold for discrete ULC variables [21]: Given ULC independent X and Y, if $\exists\ X^*$ and Y^* s.t. $X = T_\eta X^*$ and $Y = T_{1-\eta} Y^*$ for some $\eta \in (0,1)$, s.t. $H(X^*) = H(Y^*)$, then

$$H(X+Y) \geq H(X^*), \tag{13}$$

with equality iff X and Y are Poisson. The reason why (13) holds but not (11), is that finding η, X^* and Y^* that satisfy the aforesaid conditions is not always possible. This is unlike the continuous version (7), for which η, X^* and Y^* can always be constructed that satisfy the conditions, as shown in Section II.

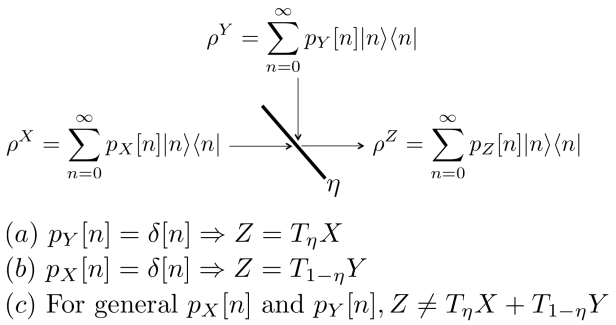
$\rho^Y = \sum_{n=0}^{\infty} p_Y[n]|n\rangle\langle n|$
$\rho^X = \sum_{n=0}^{\infty} p_X[n]|n\rangle\langle n|$
$\rho^Z = \sum_{n=0}^{\infty} p_Z[n]|n\rangle\langle n|$
η
$(a)\ p_Y[n] = \delta[n] \Rightarrow Z = T_\eta X$
$(b)\ p_X[n] = \delta[n] \Rightarrow Z = T_{1-\eta} Y$
(c) For general $p_X[n]$ and $p_Y[n], Z \neq T_\eta X + T_{1-\eta} Y$

Fig. 1. Two independent photon-number-diagonal states ρ^X and ρ^Y combined on a beamsplitter with transmissivity η results in a number-diagonal state ρ^Z at the output. The distribution of the photon number at the output is a particular scaled addition of the two input distributions, explained in Section III-C, which is at the core of our discrete EPI.

C. The natural discrete generalization of Shannon's EPI

Our goal is to develop a $\boxplus_\eta$ operation that: (i) is well behaved under a CLT with the geometric distribution as the limiting distribution; and (ii) satisfies Ineqs. (9) and (11) for X and Y with *arbitrary* discrete distributions, and with the entropy power $V_g(X) = \mathcal{E}_g^{-1}(H(X))$ defined in terms of $\mathcal{E}_g(\lambda) = (1+\lambda)\log(1+\lambda) - \lambda\log(\lambda)$, the entropy of the geometric distribution with mean λ.

Let us consider the following 1-to-1 transform between a discrete distribution $p_X[n]$, $n \in \mathbb{N}_0$ and a circularly-symmetric continuous distribution $p_{\boldsymbol{X}_c}(\boldsymbol{r})$, $\boldsymbol{r} \in \mathbb{C}$ (proof in Section IV):

$$\begin{aligned} p_{\boldsymbol{X}_c}(\boldsymbol{r}) &= \frac{1}{\pi}\sum_{n=0}^{\infty} p_X[n] \frac{e^{-|\boldsymbol{r}|^2}|\boldsymbol{r}|^{2n}}{n!}, \text{and} && (14)\\ p_X[n] &= \frac{1}{\pi}\int p_{\boldsymbol{X}_c}(\boldsymbol{r}) \mathcal{L}_n\left(|\boldsymbol{s}|^2\right) e^{\boldsymbol{r}\boldsymbol{s}^* - \boldsymbol{r}^*\boldsymbol{s}} d^2\boldsymbol{r}\, d^2\boldsymbol{s} && (15) \end{aligned}$$

where $\mathcal{L}_n(x)$ is the n-th Laguerre polynomial. Let us denote $\boldsymbol{X}_c = \mathcal{T}(X)$ and $X = \mathcal{T}^{-1}(\boldsymbol{X}_c)$ [2]. We define scaled addition $\boxplus_\eta$ for discrete random variables X and Y as follows[3]:

$$X \boxplus_\eta Y \equiv \mathcal{T}^{-1}\left(\mathcal{T}(X) \boxplus_\eta \mathcal{T}(Y)\right), \tag{16}$$

where the $\boxplus_\eta$ on the RHS of (16) is the usual continuous-variable definition as in (4). The following is simple to verify:

Proposition 2: If $p_Y[n] = \delta[n]$ is a delta function at $n = 0$, $X \boxplus_\eta Y = T_\eta X$. Similarly, if $p_X[n] = \delta[n]$, $X \boxplus_\eta Y = T_{1-\eta} Y$.

Our definition of $\boxplus_\eta$ coincides with the Yu-Johnson definition when one of the two variables is 0 with probability 1, but instead of directly adding the integer-valued scaled random variables $T_\eta X$ and $T_{1-\eta} Y$, we add the real-valued scaled

[1] The distribution of $\sum_{i=1}^n X_i$ is the n-fold convolution of $p_X[n]$. When $X_i \sim$ Bernoulli(p), $\sum_{i=1}^n X_i \sim$ Binomial(n, p) and $T_{1/n}\left(\sum X_i\right) \sim$ Binomial($n, p/n$) $\to$ Poisson(p) as $n \to \infty$. This special case is the classical Binomial-to-Poisson convergence, known as the "law of small numbers" [31].

[2] We are abusing notation a bit: $\mathcal{T}$ is not a function that maps the random variable X to the random variable $\boldsymbol{X}_c$. It is a transform that takes a *function*, a discrete p.m.f. $p_X[n]$, to another function, a continuous p.d.f. $p_{\boldsymbol{X}_c}(\boldsymbol{r})$.

[3] We will continue to use the same notation for the scaled addition, $\boxplus_\eta$, but in Section III-C onwards, the definition in Eq. (16) will be implied for $\boxplus_\eta$.

random variables $\sqrt{\eta}\boldsymbol{X}_c$ and $\sqrt{1-\eta}\boldsymbol{Y}_c$ (and transform back to the integer domain using (15)), where $\boldsymbol{X}_c = \mathcal{T}(X)$ and $\boldsymbol{Y}_c = \mathcal{T}(Y)$. Generalizing our $\boxplus_\eta$ operation for multiple variables is straightforward. Following the steps outlined in Section III-A, we consider a CLT under this scaled addition.

Theorem 3: For i.i.d. arbitrarily-distributed λ-mean random variables $\boldsymbol{X} \equiv \{X_i\}$, the p.m.f. of $Y_n \equiv \boxplus_{(\frac{1}{n},\ldots,\frac{1}{n})}\boldsymbol{X}$ converges to the geometric distribution of mean λ as $n \to \infty$.

We have the following conjecture on monotonicity of entropy:

Conjecture 4: $H(Y_n)$ increases monotonically with $n \in \{1, 2, \ldots\}$ and converges to $\mathcal{E}_g(\lambda)$, as $n \to \infty$.

This is analogous to Harremoës *et al.*'s law of small numbers [31] and Johnson-Yu's discrete entropic-monotonicity result [21], but no longer restricted to ULC random variables.

Theorem 5: The linear form of the discrete EPI holds for arbitrary independent discrete-valued random variables X, Y:

$$H(X \boxplus_\eta Y) \geq \eta H(X) + (1-\eta)H(Y). \tag{17}$$

We will show in Section IV that (17) follows as a special case of a recent result on a quantum version of the EPI [32]. Inequality (17) with $\eta = 1/2$ implies $H\left(Y_{2^{k+1}}\right) \geq H\left(Y_{2^k}\right)$, which is sufficient to prove Theorem 3 but only proves Conjecture 4 for n increasing in power-of-2 steps.

Conjecture 6: The following is the natural discrete generalization of Shannon's EPI, which holds true for arbitrary independent discrete-valued random variables X and Y:

$$V_g(X \boxplus_\eta Y) \geq \eta V_g(X) + (1-\eta)V_g(Y). \tag{18}$$

Even though we do not provide a proof of (18), we will show in Section IV that (18) is a simple special case of the yet-unproven entropy photon number inequality (EPnI) [23], [33].

De Palma *et al.* recently proved the following restricted one-sided version of (18), analogous to Yu and Johnson's Ineq. (12):

$$V_g\left(T_\eta X\right) \geq \eta V_g(X),\ \eta \in (0,1). \tag{19}$$

See Theorem 23 of Ref. [34] for a proof. There has been a suite of recent progress in quantum versions of the EPI [23], [33], [35], [36], [37], [32], [38], [39], [40], [34], and an eventual proof of the EPnI will imply the validity of the discrete EPI (18) in Conjecture 6. Finally, we have:

Theorem 7: Using a (somewhat unnatural) definition of entropy power, $V_e(X) = e^{H(X)}$—in analogy with the continuous entropy power $v(X)$—the EPI statement in (18) holds true for arbitrary independent discrete random variables X, Y:

$$V_e(X \boxplus_\eta Y) \geq \eta\, V_e(X) + (1-\eta)\, V_e(Y). \tag{20}$$

We show in Section IV that this result follows as a special case of a quantum EPI result recently proved in [38], [39].

IV. Proofs of discrete EPI results

In this Section, we will provide proofs of the various statements in Section III-C. Even though it is possible to prove them directly, it is much easier to leverage the mathematics of quantum optics. Let us consider a beamsplitter of transmissivity $\eta \in (0,1)$ that mixes two modes whose annihilation operators are $\hat{x}$ and $\hat{y}$, to produce an output mode $\hat{z} = \sqrt{\eta}\,\hat{x} + \sqrt{1-\eta}\,\hat{y}$. Assume that the quantum states of the two input modes, $\rho^{XY} = \rho^X \otimes \rho^Y$ are statistically independent. The density operators ρ^X and ρ^Y are infinite-dimensional, unit-trace, positive, Hermitian matrices, whose matrix elements we will express in the complete orthonormal *Fock* (or photon-number) basis $|n\rangle$, $n = 0, 1, \ldots, \infty$. The von Neumann entropy of ρ^X, $S(\rho^X) = -\mathrm{Tr}\left(\rho^X \log \rho^X\right) = H(\{\lambda_i^X\}) = -\sum_i \lambda_i^X \log \lambda_i^X$, where $\left\{\lambda_i^X\right\}$ are the eigenvalues of ρ^X. For a number diagonal state $\rho^X = \sum_{n=0}^{\infty} p_X[n]|n\rangle\langle n|$, $S(\rho^X) = H(X)$, Shannon entropy of the discrete random variable X.

Theorem 8: Consider independent X and Y with p.m.f.s $p_X[n]$ and $p_Y[n]$, $n \in \mathbb{N}_0$. Consider mixing independent number-diagonal states $\rho^X = \sum_{n=0}^{\infty} p_X[n]|n\rangle\langle n|$ and $\rho^Y = \sum_{n=0}^{\infty} p_Y[n]|n\rangle\langle n|$ on a beamsplitter of transmissivity η. Then the output state is also number-diagonal, i.e., $\rho^Z = \sum_{n=0}^{\infty} p_Z[n]|n\rangle\langle n|$, and the output number distribution $p_Z[n]$ is that of the random variable $Z = X \boxplus_\eta Y$, with our definition of $\boxplus_\eta$ given in (16). In other words, the scaled addition in Eq. (16) has a physical interpretation—it is how a beamsplitter of transmissivity η 'adds' photon number distributions of two independent number-diagonal input states.

Proof: The Husimi function of a quantum state ρ^X, $p_{\boldsymbol{X}_c}(\boldsymbol{r}) = \frac{1}{\pi}\langle \boldsymbol{r}|\rho^X|\boldsymbol{r}\rangle$, $\boldsymbol{r} \in \mathbb{C}$, can be interpreted as the p.d.f. of a continuous-valued random variable $\boldsymbol{X}_c$. Here, $|\boldsymbol{r}\rangle$ is the *coherent state* of complex amplitude $\boldsymbol{r}$, an eigenstate of the annihilation operator $\hat{x}$. When $\rho^X = \sum_{n=0}^{\infty} p_X[n]|n\rangle\langle n|$ is number-diagonal, its Husimi function $p_{\boldsymbol{X}_c}(\boldsymbol{r})$ is given by Eq. (14), which is a circularly-symmetric function in the phase space. The anti-normally-ordered characteristic function of ρ^X is then $\chi_A^X(\boldsymbol{s}) = \int p_{\boldsymbol{X}_c}(\boldsymbol{r}) e^{\boldsymbol{r}\boldsymbol{s}^* - \boldsymbol{r}^*\boldsymbol{s}} d^2\boldsymbol{r}$. Using the operator Fourier inverse to express the state $\rho^X = \int \chi_A^X(\boldsymbol{s}) e^{-\boldsymbol{s}\hat{x}^\dagger} e^{\boldsymbol{s}^*\hat{x}} d^2\boldsymbol{s}/\pi$, writing down the number-basis diagonal elements $\langle n|\rho^X|n\rangle = p_X[n]$, and using the identity $\langle n|e^{-\boldsymbol{s}\hat{x}^\dagger} e^{\boldsymbol{s}^*\hat{x}}|n\rangle = \mathcal{L}_n\left(|\boldsymbol{s}|^2\right)$, we obtain Eq. (15). So, we have the 1-to-1 transform, $\boldsymbol{X}_c = \mathcal{T}(X)$ and $X = \mathcal{T}^{-1}\left(\boldsymbol{X}_c\right)$. Physically, X and $\boldsymbol{X}_c$ are random variables corresponding to the outcomes of (ideal) photon number measurement and (ideal) optical heterodyne detection measurement respectively, on the state ρ^X. Because the Husimi function of a quantum state is unique, the above transform relation implies that the Husimi function of a state is circularly symmetric if and only if it is diagonal in the number basis.

Next, we observe that the Husimi function of the output state ρ^Z is given by the scaled convolution $p_{\boldsymbol{Z}_c}(\boldsymbol{r}) = \frac{1}{\eta} p_{\boldsymbol{X}_c}\left(\frac{\boldsymbol{r}}{\sqrt{\eta}}\right) * \frac{1}{(1-\eta)} p_{\boldsymbol{Y}_c}\left(\frac{\boldsymbol{r}}{\sqrt{1-\eta}}\right)$, which implies that the respective transformed variables are related by the scaled addition, $\boldsymbol{Z}_c = \sqrt{\eta}\,\boldsymbol{X}_c + \sqrt{1-\eta}\,\boldsymbol{Y}_c \equiv \boldsymbol{X}_c \boxplus_\eta \boldsymbol{Y}_c$ [35]. Given ρ^X and ρ^Y are both number diagonal (and hence have circularly-symmetric Husimi functions), ρ^Z must also have a circularly-symmetric Husimi function, and hence be number diagonal. Therefore, the number distribution of ρ^Z is that of a random variable $Z = \mathcal{T}^{-1}(\mathcal{T}(X) \boxplus_\eta \mathcal{T}(Y))$. ■

Remark 9: It is simple to verify that, for number-diagonal ρ^X, if $p_Y[n] = \delta[0]$, the output state's number distribution $p_Z[n]$ is the η-thinned version of $p_X[n]$, $Z = T_\eta X$. Similarly, with $p_X[n] = \delta[0]$ and number-diagonal ρ^Y, $Z = T_{1-\eta} Y$. Interestingly however, if neither ρ^X nor ρ^Y is the vacuum state, the output number distribution is *not* a simple addition of the thinned input distributions, i.e., $Z \neq T_\eta X + T_{1-\eta} Y$ (The RHS is Yu and Johnson's $\boxplus_\eta$); rather it is given by (16).

Proof of the discrete CLT in Theorem 3 follows from Theorem 5.10 of [35] when restricted to number-diagonal inputs and using the definition of $\boxplus_\eta$ in Eq. (16).

Let us consider the following two definitions of entropy power for a quantum state, and associated entropy power inequalities for each definition:

(a) *entropy photon number*, $\mathcal{V}_g(\rho^X) = \mathcal{E}_g^{-1}\left(S(\rho^X)\right)$, where $\mathcal{E}_g(\lambda) = (1+\lambda)\log(1+\lambda) - \lambda\log\lambda$ is the von Neumann entropy of a thermal state $\rho_{t,\lambda} = \sum_{n=0}^{\infty} p_{t,\lambda}[n]|n\rangle\langle n|$ of mean photon number λ, with $p_{t,\lambda}[n] = (1+\lambda)^{-1}(\lambda/(1+\lambda))^n$. Since $\rho_{t,\lambda}$ is diagonal in the number basis, $\mathcal{E}_g(\lambda)$ is also the Shannon entropy of the geometric distribution of mean λ, as used in Section III-C.

(b) *quantum entropy power*, $\mathcal{V}_e(\rho^X) = \mathcal{E}_e^{-1}\left(S(\rho^X)\right) = e^{S(\rho^X)}$, where $\mathcal{E}_e(\lambda) = \log(\lambda)$, defined in analogy with the Gaussian entropy function $\mathcal{E}_G(t)$ used to define the classical entropy power $v(X)$ that appears in the original EPI (1).

Entropy photon number inequality (EPnI) [23]—For a pair of independent states ρ^X and ρ^Y input to a beamsplitter of transmissivity η, producing the state ρ^Z at the output, Guha, Shapiro and Erkmen conjectured the following [23]:

$$\mathcal{V}_g\left(\rho^Z\right) \geq \eta\,\mathcal{V}_g\left(\rho^X\right) + (1-\eta)\,\mathcal{V}_g\left(\rho^Y\right), \tag{21}$$

proving that equality occurs if ρ^X and ρ^Y are thermal states. The EPnI plays a role analogous to the EPI in converse proofs of capacities of Gaussian bosonic channels. A general proof of EPnI remains open, but it was proved for Gaussian-state inputs [23]. It was shown that a special case of the EPnI,

$$\mathcal{V}_g\left(T_\eta \rho^X\right) \geq \eta \mathcal{V}_g\left(\rho^X\right), \tag{22}$$

a statement analogous to (12), would complete the converse proofs of the capacity region of the multi-user bosonic broadcast channel [33], [35], and the triple tradeoff region of the pure-loss bosonic channel [41]. Inequality (22) was proved recently by De Palma *et al.* [34], which when applied to number-diagonal inputs, implies (19) as a special case.

Quantum entropy power inequality (q-EPI) [32]—Koenig and Smith conjectured (and provided a partial proof for) the following, by replacing the $\mathcal{V}_g$ in the EPnI by $\mathcal{V}_e$ [32]:

$$\mathcal{V}_e\left(\rho^Z\right) \geq \eta\,\mathcal{V}_e\left(\rho^X\right) + (1-\eta)\,\mathcal{V}_e\left(\rho^Y\right). \tag{23}$$

A complete proof of the q-EPI (23), and a multi-input generalization thereof, were provided by De Palma *et al.* [38], [39]. However, unlike the EPnI, which emerges naturally in the bosonic Gaussian channel capacity converses, the q-EPI only implies upper bounds (or outer bounds, in case of broadcast channels) to the respective capacities [42] (or capacity regions, for broadcast channels [39]). The q-EPI implies the following linear form, analogous to (5), by applying $\mathcal{E}_e(\cdot)$ on both sides of (23) and using the concavity of the log function:

$$S\left(\rho^Z\right) \geq \eta\, S\left(\rho^X\right) + (1-\eta)\, S\left(\rho^Y\right). \tag{24}$$

If the EPnI were proven true, the concavity of $\mathcal{E}_g(\cdot)$ would also imply (24). The linear form (24) when applied to number-diagonal inputs implies (17) in Theorem 5. Similarly, the q-EPI (23) implies the discrete EPI (20). By employing (24) for $\eta = 1/2$ recursively on pairs of identical input states, we get: $S\left(\rho^{Y_{2^{k+1}}}\right) \geq S\left(\rho^{Y_{2^k}}\right)$, $k = 0, 1, \ldots$, which provides a partial proof of a conjecture by Guha and Shapiro on entropic monotonicity in a quantum CLT (Theorem 5.10 of [35]), with n increasing in power-of-2 increments. A complete proof of the aforesaid conjecture, when applied to identical independent number-diagonal states, will imply Conjecture 4 as a special case. Finally, a proof of the EPnI would imply the natural discrete-variable generalization of Shannon's EPI (18), which would satisfy all the desirable properties stated in Section III-A and would hold for all discrete random variables and not be restricted to ULC distributed random variables as in the Yu-Johnson discrete EPI.

V. Conclusions

Shannon's entropy power inequality (EPI) found many applications in proving coding theorem converses for many Gaussian channel and source coding problems. The Entropy Photon-number Inequality (EPnI) was shown to assume a role analogous to Shannon's EPI in capacity converse proofs for transmitting classical information over Gaussian bosonic (quantum) channels. Even though the general form of the EPnI remains unproven, several special cases of it have been proven in the recent years.

Many attempts have been made to find the most natural discrete-variable version of Shannon's entropy power inequality (EPI). In this paper, we developed an axiomatic framework from which we deduced the natural form of a discrete-variable EPI and an associated entropic monotonicity in a discrete-variable central limit theorem. In this discrete EPI, the geometric distribution, which has the maximum entropy among all discrete distributions with a given mean, assumes a role analogous to the Gaussian distribution in Shannon's EPI, and the thermal state in the EPnI. We defined the entropy power of a discrete random variable X as the mean of a geometric random variable with entropy $H(X)$. The crux of our construction is a discrete-variable version of Lieb's scaled addition $X \boxplus_\eta Y$ of two discrete random variables X and Y with $\eta \in (0, 1)$. We discussed the relationship of our discrete EPI with recent work of Yu and Johnson who developed an EPI for a restricted class of random variables that have ultra-log-concave (ULC) distributions, and pegged their definition of the entropy power to the entropy function of the Poisson distribution, which attains the maximum entropy for a given mean, among the class of ULC random variables. Even though we left open the proof of the aforesaid natural form of the discrete EPI that we conjectured in this paper, we showed that this discrete EPI holds true for discrete random variables with arbitrary discrete distributions when the entropy power is redefined as $e^{H(X)}$ in analogy with the continuous version. Finally, we showed that our conjectured discrete EPI is a special case of the EPnI, corresponding to the case when two input quantum states to the EPnI are independent number-diagonal states.

ACKNOWLEDGEMENTS

S.G. and R.G.-P. acknowledge the organizers of the third Beyond i.i.d. Workshop held at BIRS, Banff in July 2015, where this work took seed, and thank Rupert L. Frank and Elliott Lieb for inspiring discussions on the EPI at that workshop. The authors thank Vittorio Giovannetti and Seth Lloyd for useful discussions. SG would like to acknowledge the DARPA Quiness program funded under US Army Research Office award number W31P4Q-12-1-0019. J.H.S. acknowledges support from the Air Force Office of Scientific Research (AFOSR) grant number FA9550-14-1-0052. R.G.-P. is Research Associate of the F.R.S.-FNRS.

This document does not contain technology or technical data controlled under either the U.S. International Traffic in Arms Regulations or the U.S. Export Administration Regulations.

REFERENCES

[1] C. E. Shannon, "A mathematical theory of communication," Bell Syst. Tech. J., **27**, pp. 379–423 and 623–656 (1948).

[2] A. J. Stam, "Some inequalities satisfied by the quantities of information of Fisher and Shannon," Inform. Contr., **2**, 101–112 (1959).

[3] N. M. Blachman, "The convolution inequality for entropy powers," IEEE Trans. Inform. Theory, IT-**11**, 267–271 (1965).

[4] P. Bergmans, "A simple converse for broadcast channels with additive white Gaussian noise," IEEE Transactions on Information Theory, vol. 20, no. 2, 279–280 (1974).

[5] H. Weingarten, Y. Steinberg, and S. Shamai (Shitz), "The capacity region of the Gaussian multiple-input multiple-output broadcast channel," IEEE Transactions on Information Theory, vol. 52, no. 9, 3936–3964 (2006).

[6] M. Mohseni and J. M. Cioffi, "A proof of the converse for the capacity of Gaussian MIMO broadcast channels," in Proceedings of the IEEE International Symposium on Information Theory, Seattle, 881–885 (2006).

[7] S. Leung-Yan-Cheong and M. Hellman, "The Gaussian wire-tap channel," IEEE Transactions on Information Theory, vol. 24, no. 4, 451–456 (1978).

[8] M. H. M. Costa, "On the Gaussian interference channel," IEEE Transactions on Information Theory, vol. 31, no. 5, 607–615 (1985).

[9] R. Zamir, "Gaussian codes and Shannon bounds for multiple descriptions," IEEE Transactions on Information Theory, vol. 45, no. 7, 2629–2636 (1999).

[10] L. Ozarow, "On a source-coding problem with two channels and three receivers," Bell System Technical Journal, vol. 59, no. 10, 1909–1921 (1980).

[11] Y. Oohama, "Gaussian multiterminal source coding," IEEE Transactions on Information Theory, vol. 43, no. 6, 1912–1923 (1997).

[12] A. Wyner and J. Ziv, "A theorem on the entropy of certain binary sequences and applications: Part I, and Part II" IEEE Trans. Inform. Theory, IT-**19**, 769–772; 772–777 (1973).

[13] H.S. Witsenhausen, "Entropy inequalities for discrete channels," IEEE Trans. Inform. Theory, IT-**20**, 610–616 (1974).

[14] R. Ahlswede and J. Köruer, "On the connection between the entropies of input and output distributions of discrete memoryless channels," Proc. of the Fifth Conference on Probability Theory, Brasov (1974).

[15] S. Shamai and A. Wyner, "A binary analog to the entropy-power inequality," IEEE Trans. Inform. Theory, IT-**36**, 1428–1430 (1990).

[16] P. Harremoeës and C. Vignat, "An entropy power inequality for the binomial family," Jour. of Inequalities in Pure and Applied Math., Vol. **4**, Issue 5, Article 93 (2003).

[17] S. Haghighatshoar, E. Abbe, and E. Telatar, "A new entropy power inequality for integer-valued random variables", Proc. ISIT, 589–593 (2013).

[18] O. T. Johnson, "Log-concavity and the maximum entropy property of the Poisson distribution," Stoch. Proc. Appl., **117**, 6 791–802 (2007).

[19] Y. Yu and O. T. Johnson, "Concavity of entropy under thinning," Proc. ISIT, 144–148 (2009).

[20] Y. Yu, "Monotonic convergence in an information-theoretic law of small numbers," IEEE Trans. Inform. Theory, **55**, 12 5412–5422 (2009).

[21] O. Johnson and Y. Yu, "Monotonicity, thinning and discrete versions of the Entropy Power Inequality," IEEE Trans. on Inform. Theory, vol. **56**, no. 11, 5387–5395 (2010).

[22] J. O. Woo and M. Madiman, "A discrete entropy power inequality for uniform distributions", Proc. ISIT, 1625–1629 (2015).

[23] S. Guha, J. H. Shapiro, and B. I. Erkmen, "Capacity of the bosonic wiretap channel and the entropy photon-number inequality," Proc. ISIT, 91–95 (2008).

[24] E. Lieb, "Proof of an entropy conjecture of Wehrl," Comm. Math. Phys., vol. **62**, pp. 35–41 (1978).

[25] A. Dembo, T. M. Cover, and J. A. Thomas, "Information theoretic inequalities," IEEE Trans. Inf. Th., **37**, 6, 1501–1518 (1991).

[26] S. Verdú and D. Guo, "A simple proof of the entropy-power inequality," IEEE Trans. Inform. Theory, vol. **52**, no. 5, pp. 2165–2166 (2006).

[27] S. Artstein, K. M. Ball, F. Barthe, and A. Naor, "Solution of Shannon's problem on the monotonicity of entropy," J. Amer. Math. Soc., vol. **17**, no. 4, 975–982 (2004).

[28] R. Shimizu, "On Fisher's amount of information for location family", Statistical Distributions in Scientific Work, G. P. Patil *et al.*, Ed. Dordrecht, The Netherlands: Reidel, vol. **3**, 305–312 (1975).

[29] A. Tulino and S. Verdú, "Monotonic decrease of the non-Gaussianness of the sum of independent random variables: a simple proof," IEEE Trans. Inform. Theory, vol. **52**, no. 9, pp. 4295–4297 (2006).

[30] M. Madiman and A. Barron, "Generalized entropy power inequalities and monotonicity properties of information," IEEE Trans. Inform. Theory, vol. **53**, no. 7, 2317–2329 (2007).

[31] P. Harremoës, O. Johnson, and I. Kontoyiannis, "Thinning and the law of small numbers," Proc. ISIT, 1491–1495 (2007).

[32] R. Koenig and G. Smith, "The entropy power inequality for quantum systems," IEEE Trans. Inf. Th., **60**, 3, 1536–1548 (2014).

[33] S. Guha, J. H. Shapiro, and B. I. Erkmen, "Classical capacity of bosonic broadcast communication and a new minimum output entropy conjecture," Phys. Rev. A **76**, 032303 (2007).

[34] G. De Palma, D. Trevisan, and V. Giovannetti, "Gaussian states minimize the output entropy of the one-mode quantum attenuator", arXiv:1605.00441 [quant-ph] (2016).

[35] S. Guha, "Multiple-user quantum information theory for optical communication channels", MIT Ph.D. thesis (2008).

[36] V. Giovannetti, S. Guha, S. Lloyd, L. Maccone, and J. H. Shapiro, "Minimum output entropy of bosonic channels: a conjecture," Phys. Rev. A **70**, 032315 (2004).

[37] V. Giovannetti, A. S. Holevo and R. García-Patrón Sánchez, "A solution of Gaussian optimizer conjecture for quantum channels," Commun. Math. Phys., August 2014.

[38] G. De Palma, A. Mari, S. Lloyd, and V. Giovannetti, "The multi-mode quantum Entropy Power Inequality," Phys. Rev. A **91**, 032320 (2015).

[39] G. De Palma, A. Mari, and V. Giovannetti, "A generalization of the entropy power inequality to bosonic quantum systems," Nat. Photonics **8**, 958–964 (2014).

[40] G. De Palma, D. Trevisan, and V. Giovannetti, "Passive states optimize the output of bosonic Gaussian quantum channels," arXiv:1511.00293 [quant-ph] (2015).

[41] M. M. Wilde, P. Hayden and S. Guha, "Information Trade-Offs for Optical Quantum Communication", Phys. Rev. Lett. **108**, 140501 (2012).

[42] R. Koenig and G. Smith, "Limits on classical communication from quantum entropy power inequalities," Nat. Photonics **7**, 142–146 (2013).