The Entropy Power Inequality for Quantum Systems

Robert König and Graeme Smith, *Member, IEEE*

Abstract—When two independent analog signals, X and Y are added together giving $Z = X + Y$, the entropy of Z, $H(Z)$, is not a simple function of the entropies $H(X)$ and $H(Y)$, but rather depends on the details of X and Y's distributions. Nevertheless, the entropy power inequality (EPI), which states that $e^{2H(Z)} \geq e^{2H(X)} + e^{2H(Y)}$, gives a very tight restriction on the entropy of Z. This inequality has found many applications in information theory and statistics. The quantum analogue of adding two random variables is the combination of two independent bosonic modes at a beam splitter. The purpose of this paper is to give a detailed outline of the proof of two separate generalizations of the EPI to the quantum regime. Our proofs are similar in spirit to the standard classical proofs of the EPI, but some new quantities and ideas are needed in the quantum setting. In particular, we find a new quantum de Bruijin identity relating entropy production under diffusion to a divergence-based quantum Fisher information. Furthermore, this Fisher information exhibits certain convexity properties in the context of beam splitters.

Index Terms—Quantum information, entropy-power inequality, differential entropy, Gaussian channels.

I. INTRODUCTION

A. The Classical Entropy Power Inequality: Formulation

In his 1948 paper [1], Shannon identified the (differential) entropy

$$H(X) = -\int f_X(x) \log f_X(x) dx$$

of a random variable X with support on $\mathbb{R}^n$ as the relevant measure for its information content. Using differential entropies, Shannon obtained an expression for the capacity of a noisy channel involving continuous signals. As a corollary, he explicitly computed the capacity of the white thermal noise channel. These are among his most well-known contributions to continous-variable classical information theory. However, Shannon also recognized another fundamental property of differential entropy called the entropy power inequality (EPI).

Consider a situation where an input signal (or random variable) X is combined with an independent signal Y, resulting in an output $X + Y$ with probability density function given by the convolution

$$f_{X+Y}(z) = \int f_X(x) f_Y(z-x) d^n x \qquad \text{for all} \quad z \in \mathbb{R}^n. \quad (1)$$

We may think of the operation

$$(f_X, f_Y) \mapsto f_{X+Y} \quad (2)$$

as an 'addition law' on the space of probability density functions. Shannon's entropy power inequality relates the entropies of the input- and output variables in this process. It is commonly stated as

$$e^{2H(X+Y)/n} \geq e^{2H(X)/n} + e^{2H(Y)/n}. \quad (3)$$

It should be stressed that inequality (3) holds without assumptions on the particular form of the probability density functions f_X, f_Y beyond certain regularity conditions. Specialized to the case where one of the arguments in (2), say, f_Y, is a fixed probability density function (e.g., a Gaussian), inequality (3) immediately gives a lower bound on the output entropy of the so-called additive noise channel $X \mapsto X + Y$ (e.g., the white thermal noise channel) in terms of the entropy $H(X)$ of the input signal. Given this fact and the fundamental nature of the operation (2), it is is hardly suprising that the EPI (3) has figured prominently in an array of information-theoretic works since its introduction in 1948. For example, it has been used to obtain converses for the capacity of the Gaussian broadcast channel, and to show convergence in relative entropy for the central limit theorem [2]. Other uses of the entropy power inequality in multi-terminal information theory may be found e.g., in [3].

The expression entropy power stems from the fact that a centered normal (or Gaussian) distribution $f_X(x) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\|x\|^2/(2\sigma^2)}$ on $\mathbb{R}^n$ with variance σ^2 has entropy $H(X) = \frac{n}{2} \log 2\pi e \sigma^2$. Hence the quantity $\frac{1}{2\pi e} e^{2H(X)/n}$ is the variance of X, commonly referred to as its power or average energy. For a general random variable Z, the quantity $\frac{1}{2\pi e} e^{2H(Z)/n}$ is the power of a Gaussian variable with the same entropy as Z. This fact has been connected [4] to Brunn-Minkowski-type inequalities bounding the volume of the set-sum of two sets in terms of their spherical counterparts, which in turn inspired generalizations to free probability theory [5]. In the following, we will omit the factor $\frac{1}{2\pi e}$ and refer to $e^{2H(Z)/n}$ as the *entropy power of Z*.

The entropy power inequality (3) can be reexpressed as a kind of 'concativity' property of the entropy power under a slightly modified 'addition rule' $(X, Y) \mapsto X \boxplus_{1/2} Y$. To define the latter, let μX denote the random variable with probability density $f_{\mu X}(x) = \frac{1}{\mu} f_X(\frac{x}{\mu})$, $x \in \mathbb{R}^n$ for any scalar $\mu > 0$. We define $X \boxplus_{1/2} Y$ as the result of rescaling

both X and Y by $1/\sqrt{2}$ and subsequently taking their sum (see Eq. (4) below). Because differential entropies satisfy the scaling property $H(\mu X) = H(X) + n \log \mu$ for $\mu > 0$, this leads to the following reformulation of the entropy power inequality (3):

$$e^{2H(X \boxplus_{1/2} Y)/n} \geq \frac{1}{2} e^{2H(X)/n} + \frac{1}{2} e^{2H(Y)/n}. \qquad \text{(cEPI)}$$

Observe that the rhs. is the equal-weight average of the entropy powers of X and Y.

It turns out that the entropy $H(X)$ itself satisfies a concativity inequality analogous to (cEPI) (see (cEPI′) below). More generally, introduce a parameter $0 < \lambda < 1$ controlling the relative weight of X and Y and consider the addition rule

$$(f_X, f_Y) \mapsto f_{X \boxplus_\lambda Y} \quad \text{where} \quad f_{X \boxplus_\lambda Y} = f_{\sqrt{\lambda} X + \sqrt{1-\lambda} Y}. \qquad (4)$$

Note that this map is covariance-preserving when applied to two random-variables X and Y whose first moments and covariances coincide, a property not shared by convolution without rescaling (i.e., (2)).

The concativity inequality for the entropy takes the form

$$H(X \boxplus_\lambda Y) \geq \lambda H(X) + (1-\lambda)H(Y) \quad \text{for} \quad 0 < \lambda < 1. \qquad \text{(cEPI′)}$$

The set of inequalities (cEPI′) can be shown [6], [7] to be equivalent to (cEPI). We will refer to both (cEPI), (cEPI′) as entropy power inequalities; both express a type of concativity under the addition law (4).

B. The Quantum Entropy Power Inequality: Formulation

Here we formulate and subsequently outline a proof of a quantum version of the entropy power inequality. It is phrased in terms of the von Neumann entropy

$$S(X) = -\operatorname{tr}(\rho_X \log \rho_X)$$

of a state ρ_X on a separable infinite-dimensional Hilbert space describing n bosonic degrees of freedom. In quantum optics terminology, it gives a relation between the entropies of states going in and coming out of a beamsplitter. Central to this program is the identification of the correct definitions applying to the quantum case. This is what we consider our main contribution.

Our 'quantum addition law' generalizes (4) and is most conveniently described for the case of two single-mode ($n = 1$) states ρ_X and ρ_Y (although the generalization to $n > 1$ modes is straightforward, see Section II-B). It takes the following form: the Gaussian unitary U_λ corresponding to a beamsplitter with transmissivity $0 < \lambda < 1$ is applied to the two input modes. This is followed by tracing out the second mode. Formally, we have (see Section II-B for more precise definitions)

$$(\rho_X, \rho_Y) \mapsto \rho_{X \boxplus_\lambda Y} = \mathcal{E}_\lambda(\rho_X \otimes \rho_Y)$$
$$\text{where} \quad \mathcal{E}_\lambda(\rho) = \operatorname{tr}_2 U_\lambda \rho U_\lambda^\dagger. \qquad (5)$$

This choice of 'addition law' is motivated by the fact that the resulting position- and momentum operators (Q, P) of the output mode in the Heisenberg picture are given by

$$Q = \sqrt{\lambda} Q_1 + \sqrt{1-\lambda} Q_2$$
$$P = \sqrt{\lambda} P_1 + \sqrt{1-\lambda} P_2 \qquad (6)$$

in terms of the original operators of the first (Q_1, P_1) and second (Q_2, P_2) mode. Hence (6) indeed mimics the classical addition rule (4) in phase space.

Having defined an addition law, we can state the generalizations of (cEPI) and (cEPI′) to the quantum setting. They are

$$e^{S(X \boxplus_{1/2} Y)/n} \geq \frac{1}{2} e^{S(X)/n} + \frac{1}{2} e^{S(Y)/n}, \qquad \text{(qEPI)}$$
$$S(X \boxplus_\lambda Y) \geq \lambda S(X) + (1-\lambda)S(Y) \quad \text{for} \quad \lambda \in [0, 1]. \qquad \text{(qEPI′)}$$

Observe that (qEPI) differs from (cEPI) by the absence of a factor 2 in the exponents. This may be attributed to the fact that the phase space of n bosonic modes is $2n$-dimensional. We emphasize, however, that neither (qEPI) nor (qEPI′) appear to follow in a straightforward manner from their classical counterparts. A further distinction from the classical case is that, to the best of our knowledge, Eq. (qEPI) and (qEPI′) do not appear to be equivalent. Note also that Guha established (qEPI′) for the special case $\lambda = 1/2$ using different techniques [8, Section 5.4.2].

We will not discuss implications of these inequalities here (but see [9] for an upper bound on classical capacities), but expect them to be widely applicable. We point out that the importance of a quantum analog of the entropy power inequality in the context of additive noise channels was recognized in ealier work by Guha et al. [10]. These authors proposed a different generalization of (cEPI), motivated by the fact that the average photon number $\operatorname{tr}(a^\dagger a \rho)$ of a state ρ is the natural analog of the variance of a (centered) probability distribution. In analogy to the definition of (classical) entropy power, they defined the photon number of a state ρ as the average photon number of a Gaussian state σ with identical entropy. Their conjectured photon number inequality states that this quantity is concave under the addition rule $(X, Y) \mapsto X \boxplus_\lambda Y$. The photon number inequality has been shown to hold in the case of Gaussian states [8].

We find that our more literal generalization (qEPI) (for $\lambda = 1/2$) allows us to translate a proof of the classical entropy power inequality to the quantum setting. The formulation (qEPI) also yields tight bounds on classical capacities [9] for thermal noise associated with the transmissivity-1/2-beamsplitter. We conjecture that the quantity $e^{S(X)/n}$ is concave with respect to the addition rule $\boxplus_\lambda$ for all $0 < \lambda < 1$ (a fact which is trivial in the classical case due to the scaling property of entropy). If shown to be true, this establishes tight bounds on the classical capacity of thermal noise channels, for all transmissivities.

C. The Classical Entropy Power Inequality: Proof Sketch

It is instructive to review the basic elements of the proof of the classical entropy power inequality. Roughly, there are two different proof methods: the first one relies on Young's

inequality in the version given by Beckner [11], and Brascamp and Lieb [12]. The second one relies on the relation between entropy and Fisher information, known as de Bruijin's identity. For a detailed account of the general form of these two approaches, we refer to [13]. While [13] gives a good overview, it is worth mentioning that more recently, Rioul has found proofs of the EPI that manage to rely entirely on the use of 'standard' information measures such as mutual information [14]. His work also involves ideas from the second approach.

Here we focus on the second type of proof, originally introduced by Stam [15] and further simplified and made more rigorous by Blachman [16] and others [2], [13]. Its main components can be summarized as follows:

(i) **Gaussian perturbations:** For a random variable X on $\mathbb{R}^n$ define

$$X_t = X + \sqrt{t}Z \quad t \geq 0 \tag{7}$$

where Z be a standard normal with variance $\sigma^2 = 1$. It is well-known that the family of distributions $f_t = f_{X_t}$ satisfies the heat or diffusion equation

$$\frac{\partial}{\partial t} f_t = \Delta f_t, \qquad \Delta = \sum_{j=1}^N \frac{\partial^2}{\partial x_j^2} \tag{8}$$

with initial condition $f_0(x) = f_X(x)$ for all $x \in \mathbb{R}^N$. In particular, diffusion acts as a one-parameter semigroup on the set of probability density functions, i.e.,

$$X_{t_1+t_2} = (X_{t_1})_{t_2} \qquad \text{for all} \quad t_1, t_2 \geq 0. \tag{9}$$

Time evolution under (8) smoothes out initial spatial inhomogenities. In the limit $t \to \infty$, the distribution of X_t becomes 'simpler': it approaches a Gaussian, and the asymptotic scaling of its entropy $H(X_t) \sim g(t)$ is independent of the initial distribution f_X. This is crucial in the proof of the EPI.

A second essential feature of the evolution (8) is that it commutes with convolution: we have (for any $0 < \lambda < 1$ and $t > 0$)

$$X_t \boxplus_\lambda Y_t \equiv (X \boxplus_\lambda Y)_t \tag{10}$$

for the time-evolved versions (X_t, Y_t) of two independent random variables (X, Y).

(ii) **de Bruijin's identity:** A third property of the diffusion process (7) concerns the rate of entropy increase for infinitesimally small times, $t \to 0$. de Bruijin's identity relates this rate to a well-known estimation-theoretic quantity. It states that

$$\frac{d}{dt}\Big|_{t=0} H(X + \sqrt{t}Z) = \frac{1}{2} J(X) \qquad \text{where}$$

$$J(X) = \int (\nabla f(x))^T (\nabla f(x)) \cdot \frac{1}{f(x)} d^N x. \tag{11}$$

The quantity $J(X)$ is the *Fisher information* $J(X) = J(f^{(\theta)}; \theta)|_{\theta=0}$ associated with the family $\{f^{(\theta)}\}_{\theta \in \mathbb{R}}$ of distributions

$$f^{(\theta)}(x) = f(x - \theta \cdot \vec{e}) \quad \text{for all} \quad x \in \mathbb{R}^n. \tag{12}$$

where $\vec{e} = (1, \ldots, 1)$. In other words, the random variables $\{X^{(\theta)}\}_\theta$ are obtained by translating X by an unknown amount θ in each direction, and $J(X)$ relates to the problem of estimating θ (by the Cramér-Rao bound). More importantly, the quantity $J(X)$ inherits various general properties of Fisher information, generally defined as

$$J(f^{(\theta)}; \theta)\big|_{\theta=\theta_0} = \int \left(\frac{\partial}{\partial \theta} \log f^{(\theta)}(x)\right)^2 f^{(\theta)}(x) dx \Big|_{\theta=\theta_0} \tag{13}$$

for a parameterized family $\{f^{(\theta)}\}$ of probability distributions on $\mathbb{R}$ (this is replaced by a matrix in the case $n > 1$).

The link between entropy and Fisher information expressed by de Bruijin's identity is appealing in itself. Its usefulness in the proof of the EPI stems from yet another compatibility property with convolution. Translations as defined by (12) can be equivalently applied before or after applying addition (convolution), that is, we have

$$(X \boxplus_\lambda Y)^{(\theta)} = X^{(\sqrt{\lambda}\theta)} \boxplus_\lambda Y^{(\sqrt{1-\lambda}\theta)} \tag{14}$$

for $0 < \lambda < 1$ and $\theta \in \mathbb{R}$. In other words, adding (convolving) X and Y and then translating by θ is equivalent to adding up translated versions $X^{(\sqrt{\lambda}\theta)}, Y^{(\sqrt{1-\lambda}\theta)}$.

(iii) **Convexity inequality for Fisher information**: This shows that Fisher information behaves in a 'convex' manner under the addition law (4), i.e.,

$$J(X \boxplus_\lambda Y) \leq \lambda J(X) + (1 - \lambda) J(Y). \tag{15}$$

A related inequality for the addition law when $\lambda = 1/2$ is known as Stam's inequality; it takes the form

$$\frac{2}{J(X \boxplus_{1/2} Y)} \geq \frac{1}{J(X)} + \frac{1}{J(Y)}. \tag{16}$$

Proofs of such inequalities have been given in [4], [15], and [16]. An insightful proof was provided by Zamir [17], relying only on basic properties of Fisher information. The most fundamental of those is the information-processing inequality, which states that Fisher information is non-increasing under the application of a channel (stochastic map) $\mathcal{E}$,

$$J(\mathcal{E}(f^{(\theta)}); \theta)\big|_{\theta=\theta_0} \leq J(f^{(\theta)}; \theta)\big|_{\theta=\theta_0}. \tag{17}$$

The Fisher information inequalities (15) and (16) are simple consequences of this data-processing inequality and the compatibility (14) of convolution with translations. For example, the quantities on the two sides of (15) are the Fisher informations of the two families of distributions $\{X^{(\theta)}\}_\theta$ and $\{(X^{(\sqrt{\lambda}\theta)}, Y^{(\sqrt{1-\lambda}\theta)})\}_\theta$, and are directly associated with both sides of the compatiblity identity (14).

The entropy inequalities (cEPI) and (cEPI′) follow immediately from (i)–(iii). Concretely, let us describe the argument of Blachman [16] adapted to the proof of (cEPI′) (The derivation of (cEPI) is identical in spirit, but slightly more involved.

It uses (16) instead of (15).). Note that a similar proof which avoids the use of asymptotically divergent quantities is provided in [13], yet we find that the following version is more amenable to a quantum generalization.

The core idea is to consider perturbed or diffused versions

$$X_t = X + \sqrt{t} Z_1$$
$$Y_t = Y + \sqrt{t} Z_2,$$

of the orignal random variables (X, Y), where $t \geq 0$ and where Z_1, Z_2 are independent random variables with standard normal distribution. This is useful because asymptotically as $t \to \infty$, the entropy power inequality is satisfied trivially for the pair (X_t, Y_t) by (i). We call this the infinite-diffusion limit. de Bruijin's identity (ii) and the Fisher information inequality (iii) provide an understanding of the involved entropies in the opposite limit, where the diffusion only acts for infinitesimally short times. This allows one to extend the result from infinitely large times down to all values of t including 0. In the latter case, there is no diffusion and we are dealing with the original random variables (X, Y).

(*Sketch Proof of* (cEPI′)): In more detail, introduce the quantity

$$\delta(t) := H(X_t \boxplus_\lambda Y_t) - \lambda H(X_t) - (1 - \lambda) H(Y_t) \quad (18)$$

measuring the amount of potential violation of the entropy power inequality for the pair (X_t, Y_t). The entropy power inequality (cEPI′) for (X_t, Y_t) is equivalent to the statement $\delta(t) \geq 0$. Since $(X_0, Y_0) \equiv (X, Y)$, the entropy power inequality (cEPI′) amounts to the inequality

$$\delta(0) \geq 0. \quad (19)$$

To show (19), observe that in the infinite-diffusion limit $t \to \infty$, we have

$$\lim_{t \to \infty} \delta(t) = 0 \quad (20)$$

as all entropies $H(X_t) \sim H(Y_t) \sim H(X_t \boxplus_\lambda Y_t)$ have the same asymptotic scaling. Here we used the compatibility (10) of diffusion and the addition rule (4). de Bruijin's identity (i) and the convexity inequality for Fisher information (ii) imply that the function $\delta(t)$ is non-increasing, i.e,

$$\dot{\delta}(t) \leq 0 \qquad \text{for all} \quad t \geq 0. \quad (21)$$

Indeed, we can use the semigroup property (9) of diffusion to show that it is sufficient to consider infinitesimal perturbations, i.e., the derivative of $\delta(t)$ at $t = 0$. But the condition

$$\dot{\delta}(0) \leq 0$$

is simply the Fisher information inequality (15), thanks to de Bruijin's identity (11) and the compatibility (10) of diffusion with convolution. Hence one obtains (21) which together with Eq. (20) implies the claim (19). □

D. The Quantum Entropy Power Inequality: Proof Elements

Here we argue that a careful choice of corresponding quantum-mechanical concepts guarantees that the remarkable compatibility properties between convolution, diffusion, translation, entropy and Fisher information all have analogous quantum formulations. This yields a proof of the quantum entropy power inequalities (qEPI) and (qEPI′) that follows exactly the structure of the classical proof, yet involves fundamentally quantum-mechanical objects.

In this section, we briefly motivate and discuss the choices required.

(i) **A quantum diffusion process:** For a state ρ of n modes, we define

$$\rho_t = e^{t\mathcal{L}}(\rho) \qquad t \geq 0 \quad (22)$$

as the result of evolving for time t under the one-parameter semigroup $\{e^{t\mathcal{L}}\}_{t \geq 0}$ generated by the Markovian Master equation

$$\frac{d}{dt}\rho_t = \mathcal{L}(\rho_t)$$
$$\mathcal{L}(\cdot) = -\frac{1}{4}\sum_{j=1}^{n} \left([Q_j, [Q_j, \cdot]] + [P_j, [P_j, \cdot]]\right). \quad (23)$$

This choice of Liouvillean $\mathcal{L}$ is motivated (as in [18]) by the fact that it is the natural quantum counterpart of the (classical) Laplacian (cf. (8))

$$\Delta(\cdot) = \frac{1}{4}\sum_{j=1}^{n} \left(\{q_j, \{q_j, \cdot\}\} + \{p_j, \{p_j, \cdot\}\}\right)$$

acting on probability density function on the symplectic phase space $\mathbb{R}^{2n}$ of n classical particles (the factor $\frac{1}{4}$ is introduced here for mathematical convenience). Indeed, $\mathcal{L}$ is obtained from the Dirac correspondence $\{\cdot, \cdot\} \mapsto \frac{1}{i\hbar}[\cdot, \cdot]$ between the Poisson bracket and the commutator, as well as phase space variables and canonical operators, $(q_j, p_j) \mapsto (Q_j, P_j)$.

Alternatively, one can argue that the Wigner functions $\{W_t\}$ of the states $\{\rho_t\}$ defined by the evolution (22) obey the heat equation (8) on $\mathbb{R}^{2n}$.

Hall [18] has previously examined the relation between the Master equation (23) and classical Fisher information. He argued that the position- and momentum density functions $f_t(x) = \langle x|\rho_t|x \rangle$ and $f_t(p) = \langle p|\rho_t|p \rangle$ obey the heat equation (8). Correspondingly, he obtained de Bruijin-type identities for the measurement-entropies when measuring ρ_t in the (overcomplete) position- and momentum eigenbases, respectively. These results are, however, insufficient for our purposes as they refer to purely classical objects (derived from quantum states).

(ii) **Divergence-based quantum Fisher information:** The information processing inequality (17) is well-known to be the defining property of classical Fisher information: A theorem by Čencov [19], [20] singles out the Fisher information metric as the unique monotone metric (under stochastic maps) on the manifold of (parametrized) probability density functions. Unfortunately, such a characterization does not hold in the quantum setting: Petz [21] established a one-to-one correspondence between monotone metrics for quantum states and the set of matrix-monotone functions.

It turns out that the correct choice of quantum Fisher information in this context is given by the second derivative of the relative entropy or divergence $S(\rho\|\sigma) = \mathrm{tr}(\rho\log\rho - \rho\log\sigma)$, that is,

$$J(\rho^{(\theta)};\theta)\Big|_{\theta=\theta_0} = \frac{d^2}{d\theta^2}S(\rho^{(\theta_0)}\|\rho^{(\theta)})\Big|_{\theta=\theta_0}. \quad (24)$$

This choice is guided by the fact that the classical Fisher information (13) can also be written in this form, that is, $J(f^{(\theta)};\theta)\big|_{\theta=\theta_0} = \frac{d^2}{d\theta^2}D(f^{(\theta_0)}\|f^{(\theta)})\big|_{\theta=\theta_0}$, where $D(f\|g) = \int f(x)\log\frac{f(x)}{g(x)}$ is the classical relative entropy. We call the quantity (24) the *divergence-based (quantum) Fisher information*.

Lesniewski and Ruskai [22] have studied metrics obtained by differentiation of a 'contrast functional' $S(\cdot\|\cdot)$ in a similar manner as (24). They found that every monotone Fisher information arises from a quasi-entropy in this way. In particular, the data-processing inequality (17) is a consequence of their general result. We will discuss an independent elementary proof of the data processing inequality for the divergence-based Fisher information in Section IV.

(iii) **Translations of quantum states:** These are simply translations in phase space. We use translated versions of a state ρ displaced in each direction,

$$\rho^{(\theta,Q_j)} = e^{i\theta P_j}\rho e^{-i\theta P_j}$$
$$\rho^{(\theta,P_j)} = e^{-i\theta Q_j}\rho e^{i\theta Q_j} \qquad \text{for} \quad j = 1,\ldots,n.$$

Clearly, this choice is dictated by the Heisenberg action on the mode operators,

$$e^{i\theta P_j}Q_j e^{-i\theta P_j} = Q_j + \theta$$
$$e^{-i\theta Q_j}P_j e^{i\theta Q_j} = P_j + \theta,$$

in analogy to the addition law (6).

In the remainder of this paper, we verify that these definitions indeed satisfy the necessary properties to provide a proof of the quantum inequalities (qEPI) and (qEPI'). A word of mathematical caution is in order here: several times we will need to interchange a derivative with an integral or infinite sum, typically taking a derivative inside a trace. This can be justified as long as the functions involved are sufficiently smooth. In order to focus on the main ideas in our argument, rather than pursue a detailed justification of the interchange of limits we will simply restrict our attention to families of states that are sufficiently smooth. This involves little loss of generality since we are interested in the entropy of states satisfying an energy bound. On this set of states, entropy is continuous, so one can hope to approximate non-smooth functions with smooth ones to obtain the unrestricted result.[1] As a result, smoothness requirements are not usually considered in proofs of the classical EPI [7], [**?**], [15], [16] or studies of related quantum ideas [23], [24]. Developing a more mathematically thorough version of these arguments appears to be a formidable task (even in the classical case, see [2]) which is left to future work.

---

[1]Alternatively, one could introduce a high-photon-number cutoff (well higher than the energy bound of the states under consideration) to make the state space finite, where these problems disappear, and take the limit as the cutoff grows.

II. BASIC DEFINITIONS

A. Continuous-Variable Quantum Information

A quantum state of n modes has $2n$ canonical degrees of freedom. Let Q_k and P_k be the "position" and "momentum" operators of the k-th mode, for $k = 1,\ldots,n$, acting on the Hilbert space $\mathcal{H}_n$ associated with the system. Defining $R = (Q_1, P_1, \ldots, Q_n, P_n)$, these operators obey the commutation relations

$$[R_k, R_\ell] = iJ_{k\ell} \qquad \text{where} \qquad J = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}^{\oplus n},$$

with $^{\oplus n}$ denoting the n-fold direct sum. The *Weyl displacement operators* are defined by

$$D(\xi) = e^{i\xi\cdot JR} \qquad \text{for} \quad \xi \in \mathbb{R}^{2n}.$$

These are unitary operators satisfying the relations

$$D(\xi)D(\eta) = e^{-\frac{i}{2}\xi\cdot J\eta}D(\xi+\eta)$$

and

$$D(\xi)R_k D(\xi)^\dagger = R_k + \xi_k I \qquad \text{or} \qquad (25)$$
$$\big[D(\xi), R_k\big] = \xi_k D(\xi). \qquad (26)$$

This explains the terminology. Any state ρ takes the form

$$\rho = (2\pi)^{-n}\int\chi_\rho(\xi)D(-\xi)d^{2n}\xi, \qquad (27)$$

where

$$\chi_\rho(\xi) = \mathrm{tr}(D(\xi)\rho) \qquad (28)$$

is the *characteristic function* of ρ. The characteristic function χ_ρ is well-defined for any trace-class operator, but the map $\rho \mapsto \chi_\rho$ can be extended to all Hilbert-Schmidt class operators ρ by continuity. The map is an isometry between the Hilbert space of these operators and the space $L^2(\mathbb{R}^{2n})$ of square-integrable functions on the phase space:

$$\mathrm{tr}(\rho^\dagger\sigma) = (2\pi)^{-n}\int\overline{\chi_\rho(\xi)}\chi_\sigma(\xi) \qquad \text{for all} \quad \rho,\sigma. \quad (29)$$

Observe that for Hermitian $\rho = \rho^\dagger$, we have $\chi_\rho(\xi) = \overline{\chi_\rho(-\xi)}$ since $D(\xi)^\dagger = D(-\xi)$

A *Gaussian state* ρ is uniquely characterized by the vector $d = (d_1,\ldots,d_{2n}) \in \mathbb{R}^{2n}$ of first moments

$$d_k = \mathrm{tr}(\rho R_k) \qquad (30)$$

and its covariance matrix γ with entries

$$\gamma_{k\ell} = \mathrm{tr}(\rho\{R_k - d_k, R_\ell - d_\ell\}), \qquad (31)$$

where $\{A, B\} = AB + BA$. Its characteristic function is

$$\chi(\xi) = \exp\left(i\xi\cdot Jd - \xi\cdot J^T\gamma J\xi/4\right). \qquad (32)$$

The covariance matrix γ of a quantum state satisfies the operator inequality

$$\gamma \geq iJ, \qquad (33)$$

which is Heisenberg's uncertainty relation. Conversely, any matrix γ satisfying (33) and vector $d \in \mathbb{R}^{2n}$ define a Gaussian quantum state via (32). Complex conjugating (33) and adding it to itself shows that the covariance matrix $\gamma \geq 0$ is

positive definite. According to Williamson's theorem [25], there is a symplectic matrix $S \in Sp(2n, \mathbb{R})$, i.e., a matrix satisfying $SJS^T = J$, such that

$$S\gamma S^T = \text{diag}(\nu_1, \nu_1, \nu_2, \nu_2, \ldots, \nu_n, \nu_n) \quad \text{with} \quad \nu_j \geq 0 \quad (34)$$

is diagonal. Eq. (34) is called the *Williamson normal form* of γ, and $\{\nu_j\}_j$ are referred to as the *Williamson eigenvalues*.

Symplectic transformations S are in one-to-one correspondence with linear transformations $R \mapsto R' := SR$ preserving the canonical commutation relations. Each $S \in Sp(2n, \mathbb{R})$ defines a unitary U_S on $\mathcal{H}_n$ which realizes this map by conjugation, i.e.,

$$U_S^\dagger R_k U_S = \sum_{\ell=1}^{2n} S_{k\ell} R_\ell =: R_k'. \quad (35)$$

The unitary U_S is unique up to a global phase.

If $S \in Sp(2n, \mathbb{R})$ brings the covariance matrix of a Gaussian state ρ into Williamson normal form as in (34), then $\{R_k'\}_k$ are the *eigenmodes* of ρ. In terms of the creation, annihilation and number operators

$$a_k^\dagger = \frac{1}{\sqrt{2}}(Q_k' + i P_k') \qquad a_k = \frac{1}{\sqrt{2}}(Q_k' - i P_k')$$

$$n_k = a_k^\dagger a_k, \qquad \text{for} \quad k = 1, \ldots, n,$$

the state takes the form

$$U_S \rho U_S^\dagger = \bigotimes_{k=1}^n \frac{e^{-\beta_k n_k}}{\text{tr } e^{-\beta_k n_k}}. \quad (36)$$

In this expression, the inverse temperatures β_k are defined in terms of the symplectic eigenvalues as

$$(e^{\beta_k} - 1)^{-1} = (\nu_k - 1)/2 := N(\nu_k) \quad \text{for} \quad k = 1, \ldots, n. \quad (37)$$

The quantity (37) is the *mean photon number*, i.e., the expectation value $\text{tr}(U_S \rho U_S^\dagger a_k^\dagger a_k)$. Since the number operators have the spectral decomposition $n_k = \sum_{n \geq 0} n|n\rangle\langle n|$ with integer eigenvalues, the entropy $S(\rho) = -\text{tr}(\rho \log \rho)$ of a Gaussian state ρ can be evaluated from (36) and (37) as

$$S(\rho) = \sum_{k=1}^n g(N(\nu_k)) \quad (38)$$

where $g(N) := (N+1) \log(N+1) - N \log N$. Note that this quantity does not depend on the displacement vector d.

A completely positive trace-preserving map (CPTPM) $\mathcal{E}$ on $\mathcal{H}$ is called *Gaussian* if it preserves the set of Gaussian states, that is, the state $\rho' := \mathcal{E}(\rho)$ is Gaussian for all Gaussian input states ρ. By definition, the set of Gaussian operations is closed under composition. A Gaussian operation $\mathcal{E}$ is completely specified by its (Heisenberg) action on mode operators. The action of $\mathcal{E}$ on Gaussian states is determined by a triple (X, Y, ξ), where $\xi \in \mathbb{R}^n$ is an arbitrary vector and X, Y are real $2n \times 2n$ matrices satisfying $Y^T = Y$ and $Y + i(J - XJX^T) \geq 0$. For a Gaussian state ρ with covariance matrix γ and displacement vector d, the Gaussian state $\mathcal{E}(\rho)$ is described by

$$\gamma \mapsto \gamma' := X\gamma X^T + Y \quad (39)$$
$$d \mapsto d' := Xd + \xi. \quad (40)$$

More generally, the action on a general state ρ is determined by the action

$$\chi_\rho(\xi) \mapsto \chi_{\mathcal{E}(\rho)}(\xi) = \chi_\rho(X\xi) e^{-\frac{1}{4}\xi \cdot Y\xi}. \quad (41)$$

on characteristic functions. In the special case where $\mathcal{E}(\rho) = U\rho U^\dagger$ is unitary, we have $Y = 0$ and $X =: S \in Sp(2n, \mathbb{R})$. We call such a unitary U Gaussian and sometimes write $U = U_S$ to indicate it is defined by S (cf. (35)).

A general Gaussian operation $\mathcal{E}$ on $\mathcal{H}_n$ has a Stinespring dilation with a Gaussian unitary U_S, $S \in Sp(2(n+m), \mathbb{R})$ acting on the Hilbert space $\mathcal{H}_n \otimes \mathcal{H}_m$ of $n + m$ modes (for some m) and a Gaussian ancillary state ρ_B on $\mathcal{H}_m$, i.e., it is of the form

$$\mathcal{E}(\rho) = \text{tr}_B(U_S(\rho \otimes \rho_B)U_S^\dagger). \quad (42)$$

Since the operations of taking the tensor product and the partial trace correspond to taking the direct sum or a submatrix on the level of covariance matrices, identities (39) and (42) translate into

$$\left[S(\gamma \otimes \gamma_B)S^T \right]_{2n} = X\gamma X^T + Y,$$

where $[\cdot]_{2n}$ denotes the leading principle $2n \times 2n$ submatrix and γ_B is the covariance matrix of ρ_B. It is again convenient to express this in terms of characteristic functions. First consider the partial trace: if ρ_{n+m} has characteristic function $\chi_{\rho_{n+m}}(\xi, \xi')$, where $\xi \in \mathbb{R}^{2n}$, $\xi' \in \mathbb{R}^{2m}$, then the partial trace $\rho_n = \text{tr}_m \rho_{n+m}$ has characteristic function

$$\chi_{\rho_n}(\xi) = \chi_{\rho_{n+m}}(\xi, 0^{2m}).$$

With (41), we get the transformation rule

$$\chi_\rho \mapsto \chi_{\mathcal{E}(\rho)}(\xi) := (\chi \otimes \chi_B)(S(\xi, 0^{2m}))e^{-\frac{1}{4}(\xi, 0^{2m}) \cdot Y(\xi, 0^{2m})}, \quad (43)$$

where χ_B is the characteristic function of the Gaussian ancillary state ρ_B. Here $(\chi \otimes \chi_B)(\xi, \xi') := \chi(\xi) \cdot \chi_B(\xi')$ is the characteristic function of the product state $\rho \otimes \rho_B$.

B. Quantum Addition Using the Beamsplitter: Definition

In this section, we specify the quantum addition operation (5) in more detail. It takes takes two n-mode states ρ_X, ρ_Y and outputs an n-mode state $\rho_{X \boxplus_\lambda Y} = \mathcal{E}_\lambda(\rho_X \otimes \rho_Y)$, where $0 < \lambda < 1$.

To define the CPTP map $\mathcal{E}_\lambda$, consider the transmissivity λ-beam splitter. This is a Gaussian unitary $U_{\lambda,n}$, whose action on $2n$ modes is defined by the symplectic matrix

$$S_{\lambda,n} = \begin{pmatrix} \sqrt{\lambda}I_n & \sqrt{1-\lambda}I_n \\ \sqrt{1-\lambda}I_n & -\sqrt{\lambda}I_n \end{pmatrix} \otimes I_2, \quad (44)$$

where we use a tensor product $\{e_1^X, \ldots, e_n^X, e_1^Y, \ldots, e_1^Y\} \otimes \{q, p\}$ to represent the two sets of modes $(q_1^X, p_1^X, \ldots, q_n^X, p_n^X)$ and $(q_1^Y, p_1^Y, \ldots, q_n^Y, p_n^Y)$. Note that the symplectic matrix takes the form $I_n \otimes J$, with

$$J = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

With respect to the n pairs $(q_j^X, p_j^X), (q_j^Y, q_j^Y)$ of modes, for $j = 1, \ldots, n$, we have $S_{\lambda,n} = S_{\lambda,1}^{\oplus n}$, hence $U_{\lambda,n} = U_{\lambda,1}^{\otimes n}$ corresponds to independently applying a beam-splitter to each pair of modes. In the following, we will omit the subscript n, and simply write $\xi_X = (\xi_X^{Q_1}, \xi_X^{P_1}, \ldots, \xi_X^{Q_n}, \xi_X^{P_n})$ and $\xi_Y =$

$(\xi_Y^{Q_1}, \xi_Y^{P_1}, \ldots, \xi_Y^{Q_n}, \xi_Y^{P_n})$ for the collection of phase space variables associated with first and second set of n modes.

The map $\mathcal{E}_\lambda$ is defined by conjugation with U_λ and tracing out the second set of modes, i.e., it is

$$\mathcal{E}_\lambda(\rho_{XY}) = \mathrm{tr}_Y \, U_\lambda \rho_{XY} U_\lambda^\dagger. \tag{45}$$

Clearly, this is a Gaussian map. A bipartite Gaussian state ρ_{XY} with covariance matrix and displacement vector

$$\gamma = \begin{pmatrix} \gamma_X & \gamma_{XY} \\ \gamma_{YX} & \gamma_Y \end{pmatrix}, \qquad d = (d_X, d_Y) \tag{46}$$

gets mapped into a Gaussian state $\mathcal{E}_\lambda(\rho_{XY})$ with

$$\begin{aligned} \gamma' &= \lambda \gamma_X + (1-\lambda)\gamma_Y + \sqrt{\lambda(1-\lambda)}(\gamma_{XY} + \gamma_{YX}) \\ d' &= \sqrt{\lambda} d_X + \sqrt{1-\lambda} d_Y \end{aligned} \tag{47}$$

This completely determines the action of $\mathcal{E}_\lambda$ on Gaussian inputs, but we will also be interested in more general inputs of product form. To get an explicit expression, consider two states ρ_X, ρ_Y with characteristic functions χ_X, χ_Y. According to (41) and (44), the state $U_\lambda(\rho_X \otimes \rho_Y)U^\dagger$ has characteristic function

$$\begin{aligned} \chi_{U_\lambda(\rho_X \otimes \rho_Y)U^\dagger}(\xi_X, \xi_Y) &= \chi_X(\sqrt{\lambda}\xi_X + \sqrt{1-\lambda}\xi_Y) \\ &\quad \cdot \chi_Y(\sqrt{1-\lambda}\xi_X - \sqrt{\lambda}\xi_Y). \end{aligned} \tag{48}$$

It follows that $\mathcal{E}_\lambda(\rho \otimes \rho)$ has characteristic function

$$\chi_{\mathcal{E}_\lambda(\rho_X \otimes \rho_Y)}(\xi) = \chi_X(\sqrt{\lambda}\xi) \cdot \chi_Y(\sqrt{1-\lambda}\xi). \tag{49}$$

III. QUANTUM DIFFUSION EQUATION

Consider the diffusion Liouvillean

$$\begin{aligned} \mathcal{L}(\rho) &= -\frac{1}{4} \sum_{j=1}^{n} \left([Q_j, [Q_j, \cdot]] + [P_j, [P_j, \cdot]] \right) \\ &= -\frac{1}{4} \sum_{j=1}^{2n} [R_j, [R_j, \rho]]. \end{aligned} \tag{50}$$

defined on n modes (cf. (23)). We first establish the relevant properties of the one-parameter semigroup $\{e^{t\mathcal{L}}\}_{t \geq 0}$, where the CPTP map $e^{t\mathcal{L}}$ describes evolution under the Markovian master equation

$$\frac{d}{dt}\rho(t) = \mathcal{L}(\rho(t)). \tag{51}$$

for time t. We call this the diffusion semigroup. We will subsequently show that the maps $e^{t\mathcal{L}}$ are compatible with beamsplitters (Section III-A), and analyze the asymptotic scaling $S(\rho(t))$ of solutions $\rho(t) = e^{t\mathcal{L}}(\rho)$ of (51) (Section III-B).

The following is well-known (see [26], [27], where $\mathcal{L}$ is given in terms of creation- and annihilation operators).

Lemma III.1 Let $\mathcal{L}$ be the Liouvillean (50). Then

(i) The Liouvillean $\mathcal{L}$ is Hermitian with respect to the Hilbert-Schmidt inner product, i.e.,

$$\mathrm{tr}(\rho \mathcal{L}(\sigma)) = \mathrm{tr}(\mathcal{L}(\rho)\sigma) \tag{52}$$

for all states ρ, σ.

(ii) Let ρ_0 be a state with characteristic function $\chi_{\rho_0}(\xi)$, and let $\rho(t) = e^{t\mathcal{L}}(\rho_0)$ denote the solution of the Master

equation (51) with initial value $\rho(0) = \rho_0$. Then $e^{t\mathcal{L}}(\rho_0)$ has characteristic function

$$\chi^{(t)}(\xi) = \chi_{\rho_0}(\xi) \exp\left(-\|\xi\|^2 t/4\right), \tag{53}$$

where $\|\xi\|^2 = \sum_{j=1}^{2n} \xi_j^2$.

(iii) For any $t \geq 0$, the CPTPM $e^{t\mathcal{L}}$ is a Gaussian map acting on covariance matrices and displacement vectors by

$$\begin{aligned} \gamma &\mapsto \gamma' = \gamma + t I_{2n} \\ d &\mapsto d' = d. \end{aligned} \tag{54}$$

Proof: If $\chi_\rho(\xi)$ is the characteristic function of a state ρ, we can express the commutator $[R_j, \rho]$ as

$$[R_j, \rho] = \frac{1}{(2\pi)^n} \int \chi_\rho(\xi)\xi_j D(-\xi)d^2\xi.$$

by using (27) and the displacement property (26). Iterating this argument gives

$$[R_j, [R_j, \rho]] = \frac{1}{(2\pi)^n} \int \chi_\rho(\xi)\xi_j^2 D(-\xi)d^2\xi,$$

hence

$$\mathcal{L}(\rho) = -\frac{1}{4(2\pi)^n} \int \chi_\rho(\xi)\|\xi\|^2 D(-\xi)d^2\xi. \tag{55}$$

In other words, the operator $\mathcal{L}(\rho)$ has characteristic function $\chi_{\mathcal{L}(\rho)}(\xi) = -\frac{1}{4}\chi_\rho(\xi)\|\xi\|^2$. Together with (29), this immediately implies (52).

Let $\rho(t)$ be the operator with characteristic function (53). By commuting derivative and integration, we have

$$\begin{aligned} \frac{d}{dt}\rho(t) &= \frac{d}{dt}\frac{1}{(2\pi)^n} \int \chi_{\rho_0}(\xi) \exp(-\|\xi\|^2 t/4)D(-\xi)d^2\xi \\ &= -\frac{1}{4(2\pi)^n} \int \chi_{\rho_0}(\xi)\|\xi\|^2 \exp(-\|\xi\|^2 t/4)D(-\xi)d^2\xi \\ &= -\frac{1}{4(2\pi)^n} \int \chi_{\rho(t)}(\xi)\|\xi\|^2 D(-\xi)d^2\xi \end{aligned}$$

Combined with (55), this shows that $\rho(t)$ is indeed the solution to the Liouville equation, i.e., $\rho(t) = e^{t\mathcal{L}}(\rho)$. This proves (ii).

Finally, property (iii) follows from (ii) since multiplying a Gaussian characteristic function by a Gaussian preserves its Gaussianity. □

A. Compatibility With the Beamsplitter

In the same way as convolution is compatible with the heat equation (cf. (10)), beamsplitters are compatible with diffusion defined by the Liouvillean $\mathcal{L}$. This can be understood as a consequence of the fact that, on the level of Wigner functions (i.e., the Fourier transform of the characteristic function), both the beam-splitter map $\mathcal{E}_\lambda$ and the diffusion $e^{t\mathcal{L}}$ are described by a convolution. Here we give a compact proof based on the fact that these maps are Gaussian.

Lemma III-2. Let $t_X, t_Y \geq 0$, $0 < \lambda < 1$ and let $\mathcal{L}$ be the diffusion Liouvillean acting on n modes. Then

$$\mathcal{E}_\lambda \circ (e^{t_X \mathcal{L}} \otimes e^{t_Y \mathcal{L}}) \equiv e^{t\mathcal{L}} \circ \mathcal{E}_\lambda \tag{56}$$

where $t = \lambda t_X + (1-\lambda)t_Y$.

Proof: According to Sections II-B and III, both $\mathcal{E}_\lambda$ and $e^{t\mathcal{L}}$ are Gaussian maps, hence this is true also for both sides of identity (56). This means it suffices to verify that they agree on all Gaussian input states ρ_{XY} described by (46). The

claim follows immediately from (47) and (54): Indeed, both $\mathcal{E}_\lambda(e^{tX\mathcal{L}} \otimes e^{tY\mathcal{L}})(\rho_{XY})$ and $e^{t\mathcal{L}}(\mathcal{E}_\lambda(\rho_{XY}))$ are described by

$$\gamma' = \lambda\gamma_X + (1-\lambda)\gamma_Y + tI$$
$$d' = \sqrt{\lambda}d_X + \sqrt{1-\lambda}d_Y.$$

□

B. Scaling of the Entropy as $t \to \infty$

The following result will be essential for the proof of the entropy power inequality. It derives from a combination of arguments from [23] and [28], as well as the maximum entropy principle [29].

Theorem III.3 (Entropy scaling under diffusion) Let ρ be an arbitrary (not necessarily Gaussian) state, whose covariance matrix has symplectic eigenvalues $\nu_1, \ldots, \nu_n$. Let $t > 2$. Then

$$ng(N(t-1)) \leq S(e^{t\mathcal{L}}(\rho)) \leq \sum_{k=1}^{n} g(N(t+\nu_k)), \quad (57)$$

where $g(N) = (N+1)\log(N+1) - N\log N$ is the entropy of one-mode state with mean photon number N and $N(\nu) = (\nu - 1)/2$.

Proof: The lower bound is essentially the lower bound on the minimum output entropy from [23, (33)], generalized to several modes. We recall the necessary definitions in the Appendix. Any state can be written in the form

$$\rho = \int Q(\eta)\sigma(\eta)d^{2n}\eta \quad \text{with}$$
$$\sigma(\eta) = \frac{1}{(2\pi)^n} \int e^{\|\xi\|^2/4}e^{i\xi \cdot J\eta}D(-\xi)d^{2n}\xi,$$

where Q is a probability distribution, i.e., $Q(\eta) \geq 0$ and $\int Q(\eta)d^{2n}\eta = 1$. The operators $\sigma(\eta)$ are generally not quantum states. However, by linearity of the superoperator $e^{t\mathcal{L}}$, we can compute

$$e^{t\mathcal{L}}(\sigma(\eta)) = \frac{1}{(2\pi)^n} \int e^{-(t-1)/4 \cdot \|\xi\|^2}e^{i\xi \cdot J\eta}D(-\xi)d^{2n}\xi$$

according to Lemma III.1, which shows that $e^{t\mathcal{L}}(\sigma(\eta))$ is a displaced thermal state with covariance matrix $(t-1)I_{2n}$ if $t > 2$. Because entropy is concave, we therefore get

$$S(e^{t\mathcal{L}}(\rho)) \geq \int S(e^{t\mathcal{L}}(\sigma(\eta)))d^{2n}\eta = ng(N(t-1)),$$

and the claim follows.

For the proof of the upper bound in (57), we can assume without loss of generality that ρ is Gaussian. This is because the maximum entropy principle [29] states that for any state ρ, we have

$$S(\rho) \leq S(\rho^G), \quad (58)$$

where ρ^G is a Gaussian state with covariance matrix identical to ρ. The remainder of the proof closely follows arguments in [28]. It is a well-known fact that the entropy of a state ρ can be computed by taking the limit

$$S(\rho) = \lim_{p \to 1^+} \frac{1}{1-p} \log \operatorname{tr} \rho^p. \quad (59)$$

For a Gaussian state ρ whose covariance matrix has symplectic eigenvalues $\nu = (\nu_1, \ldots, \nu_n)$, the expression on the rhs. is equal to [30]

$$\operatorname{tr}\rho^p = \prod_{k=1}^{n} \frac{2^p}{f_p(\nu_j)} := \frac{2^{pn}}{F_p(\nu)} \quad \text{and where}$$
$$f_p(\nu) = (\nu+1)^p - (\nu-1)^p \quad (60)$$

for any $p \geq 1$. Weak submajorization, written $x \prec^w y$ for two vectors $x, y \in \mathbb{R}^n$, is defined by the condition

$$x \prec^w y \quad \Leftrightarrow \quad \sum_{j=1}^{k} x_j^{\downarrow} \leq \sum_{j=1}^{k} y_j^{\downarrow} \quad \text{for} \quad k = 1, \ldots, n,$$

where $x_1^{\downarrow} \geq x_2^{\downarrow} \geq \cdots \geq x_n^{\downarrow}$ and $y_1^{\downarrow} \geq y_2^{\downarrow} \geq \cdots \geq y_n^{\downarrow}$ are decreasing rearrangements of x and y, respectively. In [28], it is argued that the function $\nu \mapsto F_p(\nu)$ respects the partial order imposed by weak submajorization on $\mathbb{R}^n$, i.e., it has the property

$$F_p(x) \geq F_p(y) \quad \text{if} \quad x \prec^w y. \quad (61)$$

Furthermore, [28, Theorem 1] states that for any $2n \times 2n$ real positive symmetric matrices A and B, the vector of symplectic eigenvalues of their sum $A + B$ is submajorized by the sum of the corresponding vectors of A and B, respectively, i.e.,

$$\nu(A + B) \prec^w (\nu(A) + \nu(B)). \quad (62)$$

Combining (59), (60), (61) and (62) gives the following statement. Let $\rho[\gamma]$ denote the centered Gaussian state with covariance matrix γ. Let $\gamma_A > 0$ and $\gamma_B > 0$ be positive covariance matrices with symplectic eigenvalues ν^A, ν^B. Then

$$S(\rho[\gamma_A + \gamma_B]) \leq S\left(\rho\left[\bigoplus_{j=1}^{n}(\nu_j^A + \nu_j^B)I_2\right]\right). \quad (63)$$

The upper bound (57) follows immediately from (58) and (63) applied to $\gamma_A = \gamma$ and $\gamma_B = tI_{2n}$ (a valid covariance matrix for $t \geq 1$). This is because for an initial state ρ with covariance matrix γ, the time-evolved state $e^{t\mathcal{L}}(\rho)$ has covariance matrix $\gamma + tI_{2n}$ according to Lemma III.1. □

Using Theorem III.3, we can show that the asymptotic scaling of the entropy (or the entropy power) is independent of the initial state ρ, in the following sense.

Corollary III-4. The entropy and the entropy power of $e^{t\mathcal{L}}(\rho)$ grow logarithmically respectively linearly as $t \to \infty$, with an asymptotic time-dependence independent of the initial state ρ. More precisely, we have

$$\left|S(e^{t\mathcal{L}}(\rho))/n - (1 - \log 2 + \log t)\right| \leq O(1/t)$$
$$\left|\frac{1}{t}e^{S(e^{t\mathcal{L}}(\rho))/n} - \frac{e}{2}\right| \leq O(1/t).$$

where the constants in $O(\cdot)$ depend on ρ.

Proof: From the Taylor series expansions

$$\log(1 + \epsilon) = \epsilon + O(\epsilon^2)$$
$$(1 + \epsilon)^{1/\epsilon} = e - \frac{e}{2}\epsilon + O(\epsilon^2)$$

we get

$$g(N) = \log N + 1 + O(1/N)$$
$$e^{g(N)} = \left(\frac{1}{2} + N\right) e + O(\frac{1}{N}) \quad \text{for} \quad N \to \infty.$$

Replacing the rhs of (57) with the upper bound $\sum_{k=1}^{n} g(N(t + v_k)) \leq n \max_k g(N(t + v_k)) =: ng(N(t + v_*))$ and using $N(v) = (v - 1)/2$ therefore gives

$$\log(\frac{t}{2} - 1) + 1 - O(1/t) \leq S(e^{t\mathcal{L}}(\rho))/n$$
$$\leq \log(\frac{t + v_* - 1}{2}) + 1 + O(1/t)$$

and

$$\frac{t}{2}e - \frac{1}{2}e - O(\frac{1}{t}) \leq e^{S(e^{t\mathcal{L}}(\rho))/n} \leq \frac{t}{2}e + \frac{v_*}{2}e + O(\frac{1}{t})$$

for $t \to \infty$, for some constant v_* depending on ρ. The claim follows from this. $\square$

A simple consequence of Corollary III-4. is that $e^{t\mathcal{L}}(\rho)$ converges in relative entropy to a Gaussian state: we have

$$\lim_{t \to \infty} S(e^{t\mathcal{L}}(\rho) \| e^{t\mathcal{L}}(\rho)^G) = 0, \quad (64)$$

where σ^G denotes the 'Gaussification' of σ, i.e., the Gaussian state with identical first and second moments. Indeed, because $e^{t\mathcal{L}}$ is a Gaussian map (Lemma III.1), we have $e^{t\mathcal{L}}(\rho)^G = e^{t\mathcal{L}}(\rho^G)$ for any state ρ. Therefore,

$$\begin{aligned} S(e^{t\mathcal{L}}(\rho) \| e^{t\mathcal{L}}(\rho)^G) &= S(e^{t\mathcal{L}}(\rho) \| e^{t\mathcal{L}}(\rho^G)) \\ &= -S(e^{t\mathcal{L}}(\rho)) - \text{tr}(e^{t\mathcal{L}}(\rho) \log e^{t\mathcal{L}}(\rho^G)) \\ &= -S(e^{t\mathcal{L}}(\rho)) - \text{tr}(e^{t\mathcal{L}}(\rho)^G \log e^{t\mathcal{L}}(\rho^G)) \\ &= -S(e^{t\mathcal{L}}(\rho)) - \text{tr}(e^{t\mathcal{L}}(\rho^G) \log e^{t\mathcal{L}}(\rho^G)) \\ &= -S(e^{t\mathcal{L}}(\rho)) + S(e^{t\mathcal{L}}(\rho^G)). \end{aligned}$$

In the third identity, we used the fact that $\log e^{t\mathcal{L}}(\rho^G)$ is quadratic in the mode operators. Statement (64) now follows from Corollary III-4..

IV. DIVERGENCE-BASED QUANTUM FISHER INFORMATION

In this section, we introduce the divergence-based Fisher information and establish its main properties. We restrict our focus to what is needed for the proof of the entropy power inequality. We refer to the literature for substantially more general results concerning quasi-entropies and quantum Fisher information (see [21], [22], [31], [32]).

Recall that the relative entropy or divergence of two (invertible) density operators ρ, σ is defined as $S(\rho \| \sigma) = \text{tr}\, \rho (\log \rho - \log \sigma)$. The divergence is nonnegative and faithful, i.e.,

$$S(\rho \| \sigma) \geq 0 \quad \text{and} \quad S(\rho \| \sigma) = 0 \quad \text{if and only if} \quad \rho = \sigma. \quad (65)$$

Furthermore, it is monotone, i.e., non-increasing under the application of a CPTPM $\mathcal{E}$,

$$S(\mathcal{E}(\rho) \| \mathcal{E}(\sigma)) \leq S(\rho \| \sigma). \quad (66)$$

Properties (65) and (66) tell us that $S(\cdot \| \cdot)$ may be thought of as a measure of closeness. A third important property of divergence is its additivity under tensor products,

$$S(\rho_1 \otimes \rho_2 \| \sigma_1 \otimes \sigma_2) = S(\rho_1 \| \sigma_1) + S(\rho_2 \| \sigma_2). \quad (67)$$

Consider a smooth one-parameter family $\theta \mapsto \rho^{(\theta)}$ of states. Using divergence to quantify how much these states change as we deviate from a basepoint $\theta_0 \in \mathbb{R}$, it is natural to consider the function $\theta \mapsto S(\rho^{(\theta_0)} \| \rho^{(\theta)})$ in the neighborhood of θ_0. In the following, we assume that it is twice differentiable at θ_0 (but see comment below). According to (65), this function vanishes for $\theta = \theta_0$, and is nonnegative everywhere. This implies that its first derivative vanishes, i.e.,

$$\frac{d}{d\theta} S(\rho^{(\theta_0)} \| \rho^{(\theta)})\Big|_{\theta=\theta_0} = 0. \quad (68)$$

One is therefore led to consider the second derivative, which we will denote by

$$J(\rho^{(\theta)}; \theta)|_{\theta=\theta_0} = \frac{d^2}{d\theta^2} S(\rho^{(\theta_0)} \| \rho^{(\theta)})\Big|_{\theta=\theta_0}. \quad (69)$$

We call the quantity (69) the divergence-based Fisher information of the family $\{\rho^{(\theta)}\}_\theta$. Its properties are as follows:

Lemma IV.1 (Reparametrization formula) For any constant c, we have

$$J(\rho^{(c\theta)}; \theta)|_{\theta=\theta_0} = c^2 J(\rho^{(\theta)}; \theta)|_{\theta=\theta_0}$$
$$J(\rho^{(\theta+c)}; \theta)|_{\theta=\theta_0} = J(\rho^{(\theta)}; \theta)|_{\theta=\theta_0+c}.$$

Proof: This is immediate from Definition (69). $\square$

Lemma IV.2 (Non-negativity) The Fisher information satisfies

$$J(\rho^{(\theta)}; \theta)\Big|_{\theta=\theta_0} \geq 0. \quad (70)$$

Proof: Define $f_{\theta_0}(\theta) = S(\rho^{(\theta_0)} \| \rho^{(\theta)})$. We know that

$$f_{\theta_0}(\theta) \geq 0 \quad \text{for all } \theta \quad \text{and} \quad f_{\theta_0}(\theta_0) = 0.$$

The claim follows from this by writing the second derivative as a limit

$$\begin{aligned} \frac{d^2}{d\theta^2} S(\rho^{(\theta_0)} \| \rho^{(\theta)})\Big|_{\theta=\theta_0} &= \frac{d^2}{d\theta^2} f_{\theta_0}(\theta)\Big|_{\theta=\theta_0} \\ &= \lim_{\epsilon \to 0} \frac{f_{\theta_0}(\theta_0 + \epsilon) - 2 f_{\theta_0}(\theta_0) + f_{\theta_0}(\theta_0 - \epsilon)}{\epsilon^2}. \end{aligned} \quad (71)$$

$\square$

Lemma IV.3 (Additivity) For a family $\{\rho_A^{(\theta)} \otimes \rho_B^{(\theta)}\}_\theta$ of bipartite product states, we have

$$J(\rho_A^{(\theta)} \otimes \rho_B^{(\theta)}; \theta)\Big|_{\theta=\theta_0} = J(\rho_A^{(\theta)}; \theta)\Big|_{\theta=\theta_0} + J(\rho_B^{(\theta)}; \theta)\Big|_{\theta=\theta_0}. \quad (72)$$

Proof: This directly follows from the additivity (67) of divergence. $\square$

Most importantly, the divergence-based Fisher information satisfies the data processing inequality.

Theorem IV.4 (Data processing for Fisher information). Let $\mathcal{E}$ be an arbitrary CPTPM. Then

$$J(\mathcal{E}(\rho^{(\theta)}); \theta)\Big|_{\theta=\theta_0} \leq J(\rho^{(\theta)}; \theta)\Big|_{\theta=\theta_0}.$$

Proof: The proof proceeds in the same way as a well-known proof for the information-processing property of classical Fisher information (see [33]). We show here that our assumption that $f_{\theta_0}(\theta) = S(\rho^{(\theta_0)} \| \rho^{(\theta)})$ is twice differentiable at $\theta = \theta_0$ is sufficient to give the desired inequality. From (68) and the definition of Fisher information, we conclude that

$$S(\rho^{(\theta_0)} \| \rho^{(\theta)}) = \tfrac{1}{2} J(\rho^{(\theta)}; \theta)\Big|_{\theta=\theta_0} \cdot (\theta - \theta_0)^2 + o(|\theta - \theta_0|^2). \tag{73}$$

Let us argue that an analogous identity holds for the family $\{\mathcal{E}(\rho^{(\theta)})\}_\theta$. Define $g_{\theta_0}(\theta) = S(\mathcal{E}(\rho^{\theta_0}) \| \mathcal{E}(\rho^\theta))$. We know from (65), (68) and the monotonicity (66) that

$$0 \leq g_{\theta_0}(\theta) \leq f_{\theta_0}(\theta) \quad \text{for all } \theta \text{ and}$$
$$g_{\theta_0}(\theta_0) = f_{\theta_0}(\theta_0) = \frac{d}{d\theta} f_{\theta_0}(\theta)\Big|_{\theta=\theta_0} = 0. \tag{74}$$

Therefore, the derivative of g_{θ_0} is

$$\frac{d}{d\theta} g_{\theta_0}(\theta)\big|_{\theta=\theta_0} = \lim_{\epsilon \to 0} \frac{g_{\theta_0}(\theta_0 + \epsilon) - g_{\theta_0}(\theta_0)}{\epsilon}$$
$$= \lim_{\epsilon \to 0} \frac{g_{\theta_0}(\theta_0 + \epsilon)}{\epsilon} \geq 0.$$

On the other hand, (74) also gives

$$\frac{d}{d\theta} g_{\theta_0}(\theta)\big|_{\theta=\theta_0} = \lim_{\epsilon \to 0} \frac{g_{\theta_0}(\theta_0 + \epsilon)}{\epsilon} \leq \lim_{\epsilon \to 0} \frac{f_{\theta_0}(\theta_0 + \epsilon)}{\epsilon}$$
$$= \frac{d}{d\theta} f_{\theta_0}(\theta)\big|_{\theta=\theta_0} = 0,$$

hence

$$\frac{d}{d\theta} g_{\theta_0}(\theta)\Big|_{\theta=\theta_0} = 0.$$

Similarly, writing the second derivative of g_{θ_0} as a limit as in (71) and using (74) gives

$$0 \leq \frac{d^2}{d\theta^2} g_{\theta_0}(\theta)\Big|_{\theta=\theta_0} \leq \frac{d^2}{d\theta^2} f_{\theta_0}(\theta)\Big|_{\theta=\theta_0}.$$

In summary, the function g_{θ_0} has vanishing first and bounded second derivative, which shows that

$$S(\mathcal{E}(\rho^{(\theta_0)}) \| \mathcal{E}(\rho^{(\theta)})) = \frac{1}{2} J(\mathcal{E}(\rho^{(\theta)}); \theta)\Big|_{\theta=\theta_0}$$
$$\cdot (\theta - \theta_0)^2 + o(|\theta - \theta_0|^2) \tag{75}$$

The claim of the theorem now follows from (73), (75) and the data processing inequality (66). □

One may worry about the differentiability of the function $\theta \mapsto S(\rho_{\theta_0} \| \rho_\theta)$ for an arbitrary smooth one-parameter family $\{\rho^{(\theta)}\}_\theta$ on an infinite-dimensional Hilbert space. However, we do not need this full generality here. Throughout, we are only concerned with covariant families of states

$$\rho^{(\theta)} = e^{i(\theta - \theta_0)H} \rho^{(\theta_0)} e^{-i(\theta - \theta_0)H} \quad \text{for all} \quad \theta, \theta_0 \in \mathbb{R} \tag{76}$$

generated by a Hamiltonian H. For a family of the form (76), we can easily compute the derivative

$$\frac{d}{d\theta} S(\rho^{(\theta_0)} \| \rho^{(\theta)})\Big|_{\theta=\theta_0} = - \text{tr}\left(\rho^{(\theta_0)}[iH, \log \rho^{(\theta_0)}]\right), \tag{77}$$

In accordance with (68), this can be seen to vanish by inserting the spectral decomposition of $\rho^{(\theta_0)}$. We can also give an explicit expression for the second derivative, the divergence-based Fisher information. It will be convenient to state this as a lemma.

Lemma V.5. Let $\{\rho^{(\theta)}\}_\theta$ be a covariant family of states as in (76). Then

$$J(\rho^{(\theta)}; \theta)|_{\theta=\theta_0} = \text{tr}(\rho^{(\theta_0)}[H, [H, \log \rho^{(\theta_0)}]]). \tag{78}$$

Proof: Define $U(\theta) = e^{i(\theta - \theta_0)H}$. Using $U(\theta)HU(\theta)^\dagger = H$, we calculate

$$\frac{d}{d\theta} \log \rho^{(\theta)} = [iH, U(\theta)(\log \rho^{(\theta_0)})U(\theta)^\dagger]$$
$$= U(\theta)[iH, \log \rho^{(\theta_0)}]U(\theta)^\dagger,$$

hence

$$\frac{d^2}{d\theta^2} \log \rho^{(\theta)} = [iH, U(\theta)[iH, \log \rho^{(\theta_0)}]U(\theta)^\dagger].$$

Therefore

$$\frac{d^2}{d\theta^2} S(\rho^{(\theta_0)} \| \rho^{(\theta)})$$
$$= - \text{tr}(\rho^{(\theta_0)} \frac{d^2}{d\theta^2} \log \rho^{(\theta)})$$
$$= - \text{tr}(\rho^{(\theta_0)} \left[iH, U(\theta)[iH, \log \rho^{(\theta_0)}]U(\theta)^\dagger\right])$$

and evaluating this at $\theta = \theta_0$ gives the claim. □

V. THE QUANTUM DE BRUIJIN IDENTITY

The quantum analog of the translation rule (12) is defined in terms of phase space translations: For $R \in \{Q_j, P_j\}$ define the displacement operator in the direction R as

$$D_R(\theta) = \begin{cases} e^{i\theta P_j} & \text{if } R = Q_j \\ e^{-i\theta Q_j} & \text{if } R = P_j. \end{cases} \tag{79}$$

For a state ρ, consider the family of translated states

$$\rho^{(\theta, R)} = D_R(\theta)\rho D_R(\theta)^\dagger \quad \theta \in \mathbb{R} \tag{80}$$

and its Fisher information $J(\rho^{(\theta, R)}; \theta)\big|_{\theta=0}$. With a slight abuse of terminology, we will call the quantity

$$J(\rho) := \sum_{k=1}^{2n} J(\rho^{(\theta, R_k)}; \theta)\big|_{\theta=0} \tag{81}$$

obtained by summing over all phase space directions the *Fisher information* of ρ. The quantum version of de Bruijin's identity (11) then reads as follows:

Theorem V.1 (Quantum de Bruijin) Let $\mathcal{L}$ be the Liouvillean (50). The rate of entropy increase when evolving under the one-parameter semigroup $\{e^{t\mathcal{L}}\}_{t \geq 0}$ from the initial state ρ is given by

$$\frac{d}{dt} S(e^{\mathcal{L}t}(\rho))\big|_{t=0} = \tfrac{1}{4} J(\rho). \tag{82}$$

Proof: We use the expression (see [34])

$$\frac{d}{dt} S(e^{t\mathcal{L}}(\rho))\big|_{t=0} = - \text{tr}(\mathcal{L}(\rho) \log \rho). \tag{83}$$

for the rate of entropy increase under a one-parameter semigroup $e^{t\mathcal{L}}$. Because $\mathcal{L}$ is Hermitian (Lemma III.1), this is equal to

$$\frac{d}{dt} S(e^{t\mathcal{L}}(\rho))\big|_{t=0} = - \text{tr}(\rho \mathcal{L}(\log \rho)), \tag{84}$$

The claim now follows from the definition (50) of the Liouvillean $\mathcal{L}$ and Lemma V.5. □

VI. FISHER INFORMATION INEQUALITIES FOR THE BEAMSPLITTER

A. Compatibility of the Beam Splitter With Translations

We first establish a slight generalization of the compatibility condition (14) between the addition rule defined by the CPTPM $\mathcal{E}_\lambda$ (cf. (45)) and translations D_R in phase space (cf. (79)).

Lemma VI.1 Let $\theta, w_X, w_Y \in \mathbb{R}$ and consider the CPTP maps

$$\mathcal{F}(\rho) = \mathcal{E}_\lambda \left((D_R(w_X \theta) \otimes D_R(w_Y \theta)) \, \rho \, (D_R(w_X \theta) \otimes D_R(w_Y \theta))^\dagger \right)$$

and

$$\mathcal{G}(\rho) = D_R(w\theta) \mathcal{E}_\lambda(\rho) D_R(w\theta)^\dagger$$

where $w = \sqrt{\lambda} w_X + \sqrt{1-\lambda} w_Y$. Then $\mathcal{F} \equiv \mathcal{G}$. In particular,

$$\mathcal{E}_\lambda(\rho_X^{(w_X \theta, R)} \otimes \rho_Y^{(w_Y \theta, R)}) = (\mathcal{E}_\lambda(\rho_X \otimes \rho_Y))^{(w\theta, R)}$$

for any two n-mode states ρ_X and ρ_Y, where $\rho \mapsto \rho^{(\theta, R)}$ is defined by (80).

Proof: Since both $\mathcal{F}$ and $\mathcal{G}$ are Gaussian (as compositions of Gaussian operations), it suffices to show that they agree on Gaussian inputs. Note that $\rho \mapsto D_R(\theta) \rho D_R(\theta)^\dagger$ leaves the covariance matrix ρ invariant while adding $d' = d + \theta \cdot e_R$ to the displacement vector, where e_R is the standard basis vector in the R-direction. This, together with (47), immediately implies that $\mathcal{F} \equiv \mathcal{G}$. □

B. Stam Inequality and Convexity Inequality for Fisher Information

The following Fisher information inequality for the quantity $J(\rho)$ (cf. (81)) is a straightforward consequence of the compatibility of beam-splitters with translations, and the data processing inequality for Fisher information. The arguments used here are identical to those applied in Zamir's proof [17] (described transparently in [35] for the special case of interest here) of Stam's inequality. The following statement implies a quantum version of the latter and a convexity inequality of the form (15).

Theorem VI.2 (Quantum Fisher information inequality) Let $w_X, w_Y \in \mathbb{R}$ and $0 < \lambda < 1$. Let ρ_X, ρ_Y be two n-mode states. Then

$$w^2 J(\mathcal{E}_\lambda(\rho_X \otimes \rho_Y)) \leq w_X^2 J(\rho_X) + w_Y^2 J(\rho_Y)$$

where $w = \sqrt{\lambda} w_X + \sqrt{1-\lambda} w_Y$.

Proof: Let $R \in \{Q_j, P_j\}_{j=1}^n$ and consider the family of product states $\{\rho_X^{(w_X \theta, R)} \otimes \rho_Y^{(w_Y \theta, R)}\}_\theta$ and the family $\{(\mathcal{E}_\lambda(\rho_X \otimes \rho_Y))^{(w\theta, R)}\}_\theta$. By Lemma VI.1, the latter is obtained by data processing from the former. By the data processing inequality for Fisher information, this implies

$$J\left((\mathcal{E}_\lambda(\rho_X \otimes \rho_Y))^{(w\theta, R)}\right)\Big|_{\theta=0} \leq J(\rho_X^{(w_X \theta, R)} \otimes \rho_Y^{(w_Y \theta, R)}; \theta)\Big|_{\theta=0},$$

and by the properties of Fisher information, this is equivalent to

$$w^2 J\left((\mathcal{E}_\lambda(\rho_X \otimes \rho_Y))^{(w\theta, R)}\right)\Big|_{\theta=0}$$
$$\leq w_X^2 J(\rho_X^{(\theta, R)})|_{\theta=0} + w_Y^2 J(\rho_Y^{(\theta, R)}; \theta)\Big|_{\theta=0}.$$

The claim is obtained by summing over all phase space directions R. □

The following convexity inequality for Fisher information is an immediate consequence of Theorem VI.2, obtained by setting $w_X = \sqrt{\lambda}$ and $w_Y = \sqrt{1-\lambda}$.

Corollary VI-3 (Convexity inequality for Fisher information) Let ρ_X, ρ_Y be n-mode states and $0 < \lambda < 1$. Then

$$J(\mathcal{E}_\lambda(\rho_X \otimes \rho_Y)) \leq \lambda J(\rho_X) + (1-\lambda) J(\rho_Y). \quad (85)$$

We also obtain the following quantum analog of the classical Stam inequality (16).

Corollary VI-4 (Quantum Stam inequality) Let ρ_X, ρ_Y be arbitrary states and consider the $50 : 50$-beamsplitter map $\mathcal{E} = \mathcal{E}_{1/2}$. Then

$$\frac{2}{J(\mathcal{E}(\rho_X \otimes \rho_Y))} \geq \frac{1}{J(\rho_X)} + \frac{1}{J(\rho_Y)}.$$

Proof: This follows from Theorem VI.2 by setting $(\lambda = 1/2)$ and

$$w_X = \frac{J(\rho_X)^{-1}}{J(\rho_X)^{-1} + J(\rho_Y)^{-1}} \quad \text{and}$$

$$w_Y = \frac{J(\rho_Y)^{-1}}{J(\rho_X)^{-1} + J(\rho_Y)^{-1}}. \qquad □$$

VII. QUANTUM ENTROPY POWER INEQUALITIES

Having introduced the right quantum counterparts of the relevant classical concepts, the proof of the quantum entropy power inequalities is straightforward. We first provide a proof of the version (qEPI$'$) for the transmissivity λ-beamsplitter. It closely follows an argument given by Dembo, Cover and Thomas in [13] for the classical statement (cEPI$'$), but relies on a slight adaptation that allows us to generalize it to the quantum setting.

Theorem VII.1 (Quantum entropy power inequality for transmissivity λ) Let $0 < \lambda < 1$ and let $\mathcal{E}_\lambda$ be the map (45) associated with a beamsplitter of transmissivity λ. Let ρ_X and ρ_Y be arbitrary n-mode states. Then

$$S(\mathcal{E}_\lambda(\rho_X \otimes \rho_Y)) \geq \lambda S(\rho_X) + (1-\lambda) S(\rho_Y). \quad (86)$$

Proof: Define

$$\delta(t) = S(e^{t\mathcal{L}}(\mathcal{E}_\lambda(\rho_X \otimes \rho_Y))) - \lambda S(e^{t\mathcal{L}}(\rho_X))$$
$$- (1-\lambda) S(e^{t\mathcal{L}}(\rho_Y)).$$

We want to show that $\delta(0) \geq 0$. Since $\lim_{t \to \infty} \delta(t) = 0$ by the asymptotic scaling of these entropies (Corollary III-4.), it suffices to show that

$$\frac{d}{dt}\delta(t) \leq 0 \qquad \text{for all} \quad t \geq 0. \quad (87)$$

Because $\{e^{t\mathcal{L}}\}_t$ is a semigroup, we get

$$4\frac{d}{dt}\delta(t) = J(e^{t\mathcal{L}}(\mathcal{E}_\lambda(\rho_X \otimes \rho_Y))) - \lambda J(e^{t\mathcal{L}}(\rho_X))$$
$$- (1-\lambda) J(e^{t\mathcal{L}}(\rho_Y)),$$

by the quantum de Bruijin identity (Theorem V.1). Because of the compatibility of $\mathcal{E}_\lambda$ with $e^{t\mathcal{L}}$ (Lemma III-2.), we can rewrite this as

$$4\frac{d}{dt}\delta(t) = J(\mathcal{E}_\lambda(e^{t\mathcal{L}}(\rho_X) \otimes e^{t\mathcal{L}}(\rho_Y))) - \lambda J(e^{t\mathcal{L}}(\rho_X))$$
$$- (1-\lambda)J(e^{t\mathcal{L}}(\rho_Y)).$$

Using the convexity inequality for Fisher information (Corollary VI-3), the inequality (87) follows. $\qquad\square$

Specializing Theorem VII.1 to Gaussian states, we obtain a concativity inequality for the entropy with respect to the covariance matrix. That is, denoting by $\rho[\gamma]$ the centered Gaussian state with covariance matrix γ, we have

$$S(\lambda\rho[\gamma_1] + (1-\lambda)\rho[\gamma_2]) \geq \lambda S(\rho[\gamma_1]) + (1-\lambda)S(\rho[\gamma_2])$$
$$(88)$$

for all $0 < \lambda < 1$.

Finally, we can obtain the quantum analog (qEPI) of the classical entropy power inequality (cEPI) for the 50 : 50-beamsplitter, again mimicking a known classical proof.

Theorem VII.2 (Entropy power inequality for 50:50 beamsplitter) Consider the map $\mathcal{E} = \mathcal{E}_{1/2}$ associated with a 50:50 beamsplitter. Let ρ_X, ρ_Y be two n-mode states. Then

$$e^{S(\mathcal{E}(\rho_X \otimes \rho_Y))/n} \geq \frac{1}{2}e^{S(\rho_X)/n} + \frac{1}{2}e^{S(\rho_Y)/n}. \quad (89)$$

Proof: The proof follows Blachman [16], and is reproduced here for the reader's convenience. Define the functions

$$F \mapsto E_X(F) := \exp\left(S(e^{F\mathcal{L}}(\rho_X))/n\right)$$
$$G \mapsto E_Y(G) := \exp\left(S(e^{G\mathcal{L}}(\rho_Y))/n\right)$$
$$H \mapsto E_Z(H) := \exp\left(S(e^{H\mathcal{L}}(\mathcal{E}(\rho_X \otimes \rho_Y)))/n\right) \quad (90)$$

expressing the entropy powers of the states obtained by letting diffusion act on ρ_X, ρ_Y and $\mathcal{E}(\rho_X \otimes \rho_Y)$ for times F, G and $H \geq 0$, respectively. According to Corollary III-4., these functions have identical scaling

$$E(T) \sim const. + Te/2 \qquad \text{for } T \to \infty. \quad (91)$$

Moreover, since E_X and E_Y are continuous functions on the positive real axis, Peano's existence theorem shows that the initial value problems

$$\begin{array}{ll} \dot{F}(t) = E_X(F(t)), & F(0) = 0 \\ \dot{G}(t) = E_Y(G(t)), & G(0) = 0 \end{array} \quad (92)$$

have (not necessarily unique) solutions $F(\cdot), G(\cdot)$. Fix a pair of solutions and set

$$H(t) = (F(t) + G(t))/2. \quad (93)$$

Since

$$\dot{F}(t) \geq e^{g(F(t)-1)} > 0 \qquad \text{and} \qquad \dot{G}(t) \geq e^{g(G(t)-1)} > 0$$

according to Theorem III.3, these functions diverge, i.e.,

$$\lim_{t\to\infty} F(t) = \lim_{t\to\infty} G(t) = \lim_{t\to\infty} H(t) = \infty. \quad (94)$$

Consider the function

$$\delta(t) = \frac{E_X(F(t)) + E_Y(G(t))}{2E_Z(H(t))}.$$

With the initial conditions (92), it follows that the claim (89) is equivalent to

$$\delta(0) \geq 1. \quad (95)$$

Inequality (95) follows from two claims: first, we have

$$\lim_{t\to\infty} \delta(t) = 1,$$

as follows immediately from the asymptotic scaling (91) of the entropy powers E, as well as (94) and the choice (93) of $H(t)$. Second, we claim that the derivative of δ satisfies

$$\frac{d}{dt}\delta(t) \geq 0 \quad \text{for all} \quad t \geq 0. \quad (96)$$

Define, in analogy to (90), the Fisher informations of the states ρ_X, ρ_Y and $\mathcal{E}(\rho_X \otimes \rho_Y)$ after diffusion for times F, G and $H \geq 0$ as

$$J_X(F) := J\left(e^{F\mathcal{L}}(\rho_X)\right)$$
$$J_Y(G) := J\left(e^{G\mathcal{L}}(\rho_Y)\right)$$
$$J_Z(H) := J\left(e^{H\mathcal{L}}(\mathcal{E}(\rho_X \otimes \rho_Y))\right). \quad (97)$$

Using the fact that $\{e^{T\mathcal{L}}\}_T$ is a semigroup, we can restate the quantum Bruijin identity (Theorem V.1) as

$$\dot{E}_A(T) = \frac{1}{4n}E_A(T)J_A(T) \quad \text{where} \quad A \in \{X, Y, Z\} \quad (98)$$

Note also that $J_X(F), J_Y(G)$ and $J_Z(H)$ are related by Stam's inequality if $H = (F+G)/2$ is the average (93) of F and G: because

$$e^{H\mathcal{L}}(\mathcal{E}(\rho_X \otimes \rho_Y)) = \mathcal{E}(e^{F\mathcal{L}}(\rho_X) \otimes e^{G\mathcal{L}}(\rho_Y)),$$

by the compatibility of the beamsplitter with diffusion (Lemma III-2.), we obtain

$$J_Z(H) \leq \frac{2J_X(F)J_Y(G)}{J_X(F) + J_Y(G)} \quad (99)$$

using the quantum Stam inequality (Corollary VI-4) and the nonnegativity of divergence-based Fisher information. Computing the derivative of δ,

$$\dot{\delta}(t) = \frac{\dot{E}_X(F(t))\dot{F}(t) + \dot{E}_Y(G(t))\dot{G}(t)}{2E_Z(H(t))}$$
$$- \frac{E_X(F(t)) + E_Y(G(t))}{2E_Z(H(t))^2} \cdot \dot{E}_Z(H(t))\dot{H}(t),$$

inserting (98) and suppressing the δ-dependence leads to

$$8n\dot{\delta} = E_X(F)J_X(F)\dot{F} + E_Y(G)J_Y(G)\dot{G}$$
$$- (E_X(F) + E_Y(G)) \cdot E_Z(H)J_Z(H)\dot{H}$$
$$= E_X(F)^2 J_X(F) + E_Y(G)^2 J_Y(G)$$
$$- \frac{1}{2}(E_X(F) + E_Y(G))^2 J_Z(H),$$

where we inserted the definition (92) of F and G. Replacing $J_Z(H)$ by the upper bound (99) and further suppressing dependencies finally gives

$$8n\dot{\delta} \geq E_X^2 J_X + E_Y^2 J_Y - (E_X + E_Y)^2 \frac{J_X J_Y}{J_X + J_Y}$$
$$= \frac{(E_X J_X - E_Y J_Y)^2}{J_X + J_Y},$$

hence (96) follows from the nonnegativity of the Fisher information. $\square$

APPENDIX

Define the Q-function in terms of the characteristic function by the expression

$$Q(\xi) = \frac{1}{(2\pi)^{2n}} \int e^{-iJ\xi\cdot\eta}\chi(\eta)e^{-\|\eta\|^2/4}d^{2n}\eta. \qquad (100)$$

Identity (100) can be inverted to give

$$\chi(\eta) = e^{\|\eta\|^2/4}\int e^{i\eta\cdot J\xi}Q(\xi)d^{2n}\xi. \qquad (101)$$

Inserting the definition (28) of the characteristic function into (100) gives

$$Q(\xi) = \frac{1}{(2\pi)^n}\operatorname{tr}\left(\rho\frac{1}{(2\pi)^n}\int e^{i\eta\cdot J\xi}D(-\eta)e^{-\|\eta\|^2/4}d^{2n}\eta\right)$$

Comparing this with (32), we conclude that

$$Q(\xi) = \frac{1}{(2\pi)^n}\langle\xi|\rho|\xi\rangle,$$

where $|\xi\rangle = D(\xi)|0\rangle$ is a coherent state.

This shows that $Q(\xi) \geq 0$ for all $\xi \in \mathbb{R}^{2n}$. Furthermore, since the coherent states satisfy the completeness relation

$$I = \frac{1}{(2\pi)^n}\int|\xi\rangle\langle\xi|d^{2n}\xi,$$

we have $\int Q(\xi)d^{2n}\xi = 1$, i.e., Q is a probability density.

The state ρ can be expressed in terms of Q using (101) and (27): we have

$$\rho = \int Q(\eta)\sigma(\eta)d^{2n}\eta \qquad \text{where}$$

$$\sigma(\eta) = \frac{1}{(2\pi)^n}\int e^{\|\xi\|^2/4}e^{i\xi\cdot J\eta}D(-\xi)d^{2n}\xi. \qquad (102)$$

REFERENCES

[1] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 379–423, Jul. 1948.

[2] A. R. Barron, "Entropy and the central limit theorem," *Ann. Probab.*, vol. 14, no. 1, pp. 336–342, 1986.

[3] A. E. Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2012.

[4] M. H. M. Costa and T. M. Cover, "On the similarity of the entropy power inequality and the Brunn-Minkowski inequality (Corresp.)," *IEEE Trans. Inf. Theory*, vol. 30, no. 6, pp. 837–839, Nov. 1984.

[5] S. Szarek and D. Voiculescu, "Volumes of restricted Minkowski sums and the free analogue of the entropy power inequality," *Commun. Math. Phys.*, vol. 178, no. 3, pp. 563–570, 1996.

[6] E. H. Lieb, "Proof of an entropy conjecture of Wehrl," *Commun. Math. Phys.*, vol. 62, no. 1, pp. 35–41, 1978.

[7] S. Verdu and D. Guo, "A simple proof of the entropy-power inequality," *IEEE Trans. Inf. Theory*, vol. 52, no. 5, pp. 2165–2166, May 2006.

[8] S. Guha, "Multiple-user quantum information theory for optical communication channels," Ph.D. dissertation, Dept. Electr. Eng. Comput. Sci., MIT, Cambridge, MA, USA, 2008.

[9] R. König and G. Smith, "Limits on classical communication from quantum entropy power inequalities," *Nature Photon.*, vol. 7, pp. 142–146, Jan. 2013.

[10] S. Guha, B. Erkmen, and J. Shapiro, "The entropy photon-number inequality and its consequences," in *Proc. Inf. Theory Appl. Workshop*, 2008, pp. 128–130.

[11] W. Beckner, "Inequalities in Fourier analysis," *Ann. Math.*, vol. 102, no. 1, pp. pp. 159–182, 1975.

[12] H. Brascamp and E. Lieb, "Best constants in Young's inequality, its converse, and its generalization to more than three functions," *Adv. Math.*, vol. 20, no. 2, pp. 151–172, May 1976.

[13] A. Dembo, T. Cover, and J. Thomas, "Information theoretic inequalities," *IEEE Trans. Inf. Theory*, vol. 37, no. 6, pp. 1501–1518, Nov. 1991.

[14] O. Rioul, "Information theoretic proofs of entropy power inequalities," *IEEE Trans. Inf. Theory*, vol. 57, no. 1, pp. 33–55, Jan. 2011.

[15] A. Stam, "Some inequalities satisfied by the quantities of information of Fisher and Shannon," *Inf. Control*, vol. 2, no. 2, pp. 101–112, 1959.

[16] N. Blachman, "The convolution inequality for entropy powers," *IEEE Trans. Inf. Theory*, vol. 11, no. 2, pp. 267–271, Apr. 1965.

[17] R. Zamir, "A proof of the Fisher information inequality via a data processing argument," *IEEE Trans. Inf. Theory*, vol. 44, no. 3, pp. 1246–1250, May 1998.

[18] M. J. W. Hall, "Quantum properties of classical Fisher information," *Phys. Rev. A*, vol. 62, p. 012107, Jun 2000.

[19] N. N. Čencov, *Statistical Decision Rules and Optimal Inference* (Translations of Mathematical Monographs), vol. 53. Providence, RI, USA: AMS, 1981, p. 499.

[20] L. L. Campbell, "An extended Čencov characterization of the information metric," *Proc. Amer. Math. Soc.*, vol. 98, no. 1, pp. 135–141, 1986.

[21] D. Petz, "Monotone metrics on matrix spaces," *Linear Algebra Appl.*, vol. 244, pp. 81–96, 1996.

[22] A. Lesniewski and M. B. Ruskai, "Monotone Riemannian metrics and relative entropy on non-commutative probability spaces," *J. Math. Phys.*, vol. 40, pp. 5702–5724, Aug. 1999.

[23] V. Giovannetti, S. Guha, S. Lloyd, L. Maccone, and J. H. Shapiro, "Minimum output entropy of bosonic channels: A conjecture," *Phys. Rev. A*, vol. 70, p. 032315, Sep. 2004.

[24] V. Giovannetti, A. S. Holevo, S. Lloyd, and L. Maccone, "Generalized minimal output entropy conjecturefor one-mode Gaussian channels: Definitions and some exact results," *J. Phys. A, Math. Theoretical*, vol. 43, no. 41, p. 032315, Oct. 2010.

[25] J. Williamson, "On the algebraic problem concerning the normal forms of linear dynamical systems," *Amer. J. Math.*, vol. 58, no. 1, pp. 141–163, 1936.

[26] P. Vanheuverzwijn, "Generators for quasi-free completely positive semi-groups," *Ann. Inst. Henri Poincaré Sec. A, Phys. Théorique*, vol. 29, no. 1, pp. 123–138, 1978.

[27] S. Lloyd, V. Giovannetti, L. Maccone, N. J. Cerf, S. Guha, R. Garcia-Patron, *et al.* (2009). *The bosonic minimum output entropy conjecture and Lagrangian minimization* [Online]. Available: http://arxiv.org/abs/0906.2458

[28] T. Hiroshima, "Additivity and multiplicativity properties of some Gaussian channels for Gaussian inputs," *Phys. Rev. A*, vol. 73, p. 012330, Jan. 2006.

[29] M. M. Wolf, G. Giedke, and J. I. Cirac, "Extremality of Gaussian quantum states," *Phys. Rev. Lett.*, vol. 96, p. 080502, Mar. 2006.

[30] A. Holevo, M. Sohma, and O. Hirota, "Error exponents for quantum channels with constrained inputs," *Rep. Math. Phys.*, vol. 46, no. 3, pp. 343–358, 2000.

[31] D. Petz and C. Ghinea, "Introduction to quantum Fisher information," *Quantum Probab. Rel. Topics*, vol. 27, pp. 261–281, Nov. 2009.

[32] F. Hiai, M. Mosonyi, D. Petz, and C. Beny, "Quantum f-divergences and error correction," *Rev. Math. Phys.*, vol. 23, no. 7, pp. 691–747, 2011.

[33] M. Raginsky. (2011, Jul.). Divergence in everything: Cramér-Rao from data processing. *Inf. Struct.* [Online]. Available: http://infostructurailst.wordpress.com/2011/07/

[34] H. Spohn, "Entropy production for quantum dynamical semigroups," *J. Math. Phys.*, vol. 19, no. 5, pp. 1227–1230, 1978.

[35] A. Kagan and T. Yu, "Some inequalities related to the Stam inequality," *Appl. Math.*, vol. 53, no. 3, pp. 195–205, 2008.

Robert König received his diploma in theoretical physics from the Swiss Federal Institute of Technology (ETH) Zurich in 2003. He subsequently worked as a research and teaching assistant at the department of theoretical computer science at ETH before moving to Cambridge, U.K. After completing his PhD in 2007, he was a postdoctoral scholar at the Institute for Quantum Information, Caltech until 2011. In 2012, he joined the Institute for Quantum Computing and the Department of Applied Mathematics at the University of Waterloo after spending a year at IBM T.J. Watson research. He is interested in all mathematical, physical and computer-science related aspects of quantum information.

Graeme Smith is a Research Staff Member in the Physics of Information group at IBM TJ Watson Research Center. He was previously a postdoc at both IBM and Bristol and received a PhD in physics from Caltech in 2006. Graeme is interested in computation and communication in noisy settings, quantum Shannon theory, coding theory, cryptography, quantum characterization, validation and verification, quantum estimation and detection, optical communications, and the physics of information.