Continuous-variable entanglement distillation and noncommutative central limit theorems

Earl T. Campbell,[1,*] Marco G. Genoni,[2] and Jens Eisert[1]
[1]*Dahlem Center for Complex Quantum Systems, Freie Universität Berlin, 14195 Berlin, Germany*
[2]*QOLS, Blackett Laboratory, Imperial College London, London SW7 2BW, United Kingdom*
(Received 17 December 2012; published 24 April 2013)

Entanglement distillation transforms weakly entangled noisy states into highly entangled states, a primitive to be used in quantum repeater schemes and other protocols designed for quantum communication and key distribution. In this work, we present a comprehensive framework for continuous-variable entanglement distillation schemes that convert noisy non-Gaussian states into Gaussian ones in many iterations of the protocol. Instances of these protocols include (a) the recursive-Gaussifier protocol, (b) the temporally reordered recursive-Gaussifier protocol, and (c) the pumping-Gaussifier protocol. The flexibility of these protocols gives rise to several beneficial trade-offs related to success probabilities or memory requirements, which can be adjusted to reflect experimental demands. Despite these protocols involving measurements, we relate the convergence in this protocol to new instances of noncommutative central limit theorems, in a formalism that we lay out in great detail. Implications of the findings for quantum repeater schemes are discussed.

DOI: 10.1103/PhysRevA.87.042330 PACS number(s): 03.67.Ac, 03.67.Bg, 42.50.Ex

I. INTRODUCTION

Photons, with information encoded in continuous-variable degrees of freedom, can travel great distance without significant decoherence. We can, using beam splitters, phase shifters, and detectors, coherently manipulate photons and make measurements. Specifically in the continuous-variable regime, brighter sources of light are available than for single photon, discrete, light sources. These features have motivated research into the usefulness of photonic systems for quantum cryptography, communication, and distributed quantum information processing [1,2]. Discrete protocols, for finite-dimensional systems with arbitrary quantum control, do not typically have exact analogs but rather cousins in the linear optical setting. Any two qubit entangled state can be distilled by local operations [3], whereas distillation of entangled Gaussian states using linear optics is impossible [4–6]. Soon after these impossibility proofs were obtained, it was discovered that an initially non-Gaussian state could, using only linear optics, enable entanglement distillation [7–9]. The original distillation protocol, which is conditioned on detectors finding no photons, outputs a state that evolves toward a Gaussian. Over the years, this protocol has inspired several variants that have been found to exhibit the same Gaussification phenomena [10,11]. Similar "no-go" results [12] prohibit the distillation of highly squeezed states using only passive linear optics, although with relaxed constraints some proposals are possible [13].

Leaving the realm of purely Gaussian operations is essential for entanglement distillation, but unfortunately non-Gaussian operations are much more experimentally challenging. Therefore, it is desirable to keep non-Gaussian operations to a minimum. In the aforementioned protocols, and those considered herein, only the initial noisy resource needs to be non-Gaussian. A source of Gaussian entangled states, such as those emitted by a pumped parametric downconverter, can be probabilistically de-Gaussified by adding or subtracting single photons through the use of single photon detectors and/or sources [8,11,14–17]. An additional benefit of de-Gaussification is that it too can increase the entanglement and other figures of merit, such as the teleportation fidelity [18–21]. Some matter systems (e.g., Ref. [22]) also provide a more direct source of non-Gaussian entangled photons. These are the most experimentally feasible means of non-Gaussian state preparation, but the potential advantage of exploiting more exotic forms of non-Gaussianity has also been considered [11,21,23,24]. The need for non-Gaussian operations extends beyond distillation problems, and they are required to violate locality [25–27] and to outperform classical computers [28–32]. These applications have kindled an interest in the idea of Wigner function negativity as a resource [33].

Until now, known protocols that Gaussify and distill entanglement have the feature of being recursive. To execute these protocols to greater depth requires greater memory storage requirements. The quantum states are combined via a treelike process of pairwise distillation, with each branch demanding additional memory. In the finite-dimensional setting, entanglement pumping protocols [34–42] offer the option of compressing the spatial memory requirement, even down to three to four qubits per location, at the cost of reduced efficiency and increased temporal overheads. Recently, a continuous variable analog of entanglement pumping, the compact distillery scheme, has been proposed [43]. This scheme requires storage of only two modes per location at any moment in time. However, this pumping protocol is not a direct analog of the Gaussification protocols. In particular, the compact distillery does not Gaussify and allows only a modest increase in entanglement.

Here, we extend and further develop the techniques of Ref. [44] where the class of Gaussification protocols was vastly broadened and shown to work in virtue of quantum central limit theorems. This work broadens the class of Gaussifier protocols, and in doing so introduces the concept of a pumping Gaussifier that only requires two modes of memory per location. Unlike the compact distillery scheme, our pumping protocol still Gaussifies and is capable of the same large increases of entanglement possible with the recursive Gaussifier. Surprisingly, the pumping Gaussifer outputs the same

*earltcampbell@gmail.com

 ©2013 American Physical Society

final state as the more well-known recursive Gaussifiers. This makes pumping Gaussifers extremely promising protocols that are especially attractive for experiments with only a small number of modes. We also comment on implications of our findings to devising novel schemes for long distance quantum communication via quantum repeater networks. Despite considerable research on continuous-variable (CV) entanglement distillation, surprisingly these techniques have not previously been explicitly applied to design of quantum repeaters. Indeed, here, we provide the first concrete evidence in the CV context that using quantum repeaters can achieve greater distances of communication than direct transmission.

On a technical level, the approach taken here is complementary to, but subtly distinct from, our earlier results [44]. In particular, compared to these earlier results, the relationship between quantum central limit theorems and Gaussification protocols requires a smaller and simpler set of assumptions required of the physical system. Center stage is taken by a class of noncommutative central limit theorems, which are general enough to capture all of the aforementioned situations of state manipulation, including postselecting measurements. The requirements for a quantum central limit theorem to be valid will be highlighted and discussed in great detail. We remark that these techniques are closely related to those used to prove the extremality principle [45], which asserts that for entanglement measures satisfying very specific properties, Gaussian states have the least entanglement of all states with the same second moments.

II. CONTINUOUS-VARIABLE SYSTEMS AND PHASE SPACE

Here, we introduce our notation and briefly introduce some phase space concepts used throughout. For more details see Refs. [1,5,46]. For a single mode of a CV system, two important observables are

$$\hat{X} = (\hat{a} + \hat{a}^\dagger)/\sqrt{2}, \tag{1}$$

$$\hat{P} = i(\hat{a}^\dagger - \hat{a})/\sqrt{2}, \tag{2}$$

which are analogs of position and momentum in simple harmonic oscillators, with $\hat{a}$ and $\hat{a}^\dagger$ being the photonic annihilation and creation operators. For m optical modes, the set of $2m$ quadrature operators is denoted as a vector of operators

$$\hat{\mathbf{Q}} = (\hat{Q}_1, \hat{Q}_2, \ldots, \hat{Q}_{2m-1}, \hat{Q}_{2m}) = (\hat{X}_1, \hat{P}_1, \ldots, \hat{X}_m, \hat{P}_m). \tag{3}$$

For a quantum state ρ, the expectation values of these quadratures are denoted by a set of $2m$ real numbers

$$[\mathbf{d}_\rho]_k = \mathrm{tr}(\hat{Q}_k \rho), \tag{4}$$

which are called the first moments of ρ. Typically, we are interested in states with zero first moments, so $\mathbf{d}_\rho = 0$. The second moments, akin to variances, are captured by the covariance matrix

$$[\Gamma_\rho]_{j,k} = 2\,\mathrm{Re}\{\mathrm{tr}[(\hat{Q}_j - [\mathbf{d}_\rho]_j)(\hat{Q}_k - [\mathbf{d}_\rho]_k)\rho]\}, \tag{5}$$

which for states with zero first moments simplifies to

$$[\Gamma_\rho]_{j,k} = \mathrm{tr}[(\hat{Q}_j \hat{Q}_k + \hat{Q}_k \hat{Q}_j)\rho]. \tag{6}$$

It is easy to verify that, for physical states, the covariance matrix is real and symmetric.

The first and second moments only partially describe the quantum state, but a complete description can be achieved by using one of a plethora of phase space representations. In particular, we make use of characteristic functions $\chi_\rho : \mathbb{R}^{2m} \to \mathbb{C}$ such that

$$\chi_\rho(\mathbf{r}) = \mathrm{tr}[D(\mathbf{r})\rho], \tag{7}$$

where $D_\mathbf{r}$ is the unitary displacement or Weyl operator

$$D_\mathbf{r} = \exp(i \cdot \hat{\mathbf{Q}}) = \exp\left(i \sum_j r_j \hat{Q}_j\right). \tag{8}$$

We say a state is Gaussian if and only if its characteristic function has a Gaussian shape, which entails

$$\chi_\rho(\mathbf{r}) = \exp(i\mathbf{r} \cdot \mathbf{d}_\rho - \mathbf{r}^T \Gamma_\rho \mathbf{r}/4). \tag{9}$$

Any state outside this set is said to be non-Gaussian. Notable Gaussian states include the vacuum and the coherent states. The Wigner function, which is perhaps more widely known, is simply the Fourier transform of the characteristic function. Since the Fourier transform maps the set of Gaussian functions to itself, the definition of Gaussian states is equivalent if stated in terms of Wigner functions. For our purposes the characteristic function is the most useful choice of phase space representation.

Regarding dynamics, we say a unitary is Gaussian if it has the form $U = \exp(iH)$, where H is Hermitian and quadratic in annihilation and creation operators. The canonical example of a Gaussian measurement is a homodyne, or quadrature, measurement of an observable $\hat{Q}_j$. More general Gaussian measurements can be related to quadrature measurements by use of Gaussian unitaries and ancillary Gaussian states. For example, so called eight-port homodyne measurements project onto the coherent states and can be implemented by using two quadrature measurements and an ancillary mode in the vacuum state.

The most general kind of Gaussian operations are Gaussian channels (completely positive maps). This class of physical operations is most naturally defined by using the Choi-Jamiolkowski (CJ) isomorphism [47,48] between quantum states and channels. For a channel $\mathcal{E}$ mapping m-mode quantum states to m-mode quantum states, the CJ state is

$$\Phi_\mathcal{E} = (\mathbb{1} \otimes \mathcal{E})\Phi, \tag{10}$$

where $\Phi = |\phi\rangle\langle\phi|^{\otimes m}$ is a pure unnormalized operator with

$$|\phi\rangle = \sum_{n=0}^{\infty} |n,n\rangle. \tag{11}$$

Conversely, for all $\Phi_\mathcal{E}$ there exists a unique quantum channel $\mathcal{E}$ specified by the isomorphism, such that

$$\mathcal{E}_\Phi(\rho) = \mathrm{tr}^B[\Phi^{T_B}(\mathbb{1} \otimes \rho)], \tag{12}$$

where T_B is a partial transpose with respect to B. When the Gaussian completely positive (CP) map acts on a Gaussian state ρ with covariance matrix Γ_ρ, it has been shown [4–6] that the output state ρ' is also Gaussian with covariance matrix

$$\Gamma_{\rho'} = \gamma_{AA} - \gamma_{AB}(\gamma_{BB} + \Gamma_\rho)^{-1}\gamma_{AB}^T, \tag{13}$$

where

$$\gamma = \begin{pmatrix} \gamma_{AA} & \gamma_{AB} \\ \gamma_{AB}^T & \gamma_{BB} \end{pmatrix} \tag{14}$$

is the covariance matrix of Φ^{T_B} shown as a block matrix with respect to the partition between systems A and B. The expression for $\Gamma_{\rho'}$ takes the form of a Schur complement, which often arises in matrix problems and Gaussian integration [49]. The partial transpose has a simple effect on covariance matrices, and so explicitly calculating the partial transposed state can be circumvented. Partial transposition, in the Heisenberg picture, takes $\hat{P} \mapsto -\hat{P}$ for every momentum operator acting on system B. Assume we know $\Phi_{\mathcal{E}}/\mathrm{tr}(\Phi_{\mathcal{E}})$ and its covariance matrix $\tilde{\gamma}$. It follows that the partial transposed state $\Phi_{\mathcal{E}}^{T_B}/\mathrm{tr}(\Phi_{\mathcal{E}})$ has covariance matrix $\gamma = \Lambda\tilde{\gamma}\Lambda$, where $\Lambda = \mathbb{1}_A \oplus T_B$ and $T_B = \mathrm{diag}(1,-1,\ldots,1,-1)$.

III. BUILDING BLOCKS

This section introduces the basic building blocks of the protocols considered herein. Each building block is specified by the following: an operator Π called the filter; a value R for the beam-splitter reflectivity; and a choice of two m-mode states that may be outputs from previous building blocks. Throughout this article, any building blocks combined into a larger protocol will use the same filter Π, which must be an invertible operator proportional to a separable Gaussian state with zero first moments. Such filters always, as shown in Ref. [50], have a decomposition of the form

$$\Pi = \int P(\mathbf{r})\Pi_{\mathbf{r}}, \tag{15}$$

where $P(\mathbf{r})$ is a classical, and Gaussian, probability distribution and

$$\Pi_{\mathbf{r}} = D_{\mathbf{r}}|\psi\rangle\langle\psi|D_{\mathbf{r}}^{\dagger} \tag{16}$$

for some pure separable Gaussian $|\psi\rangle$. The set of operators $\{\Pi_{\mathbf{r}}\}$ specifies the positive operator-valued measure (POVM) measurement to be used in the building block. Recall that eight-port homodyne measurements implement a similar POVM where $|\psi\rangle$ is the vacuum state, and so the desired POVM is always equivalent, up to a local Gaussian unitary, to eight-port homodyne measurement. The weighting $P(\mathbf{r})$ is a function of the measurement outcome $\Pi_{\mathbf{r}}$ and dictates the postselection strategy used in the building block. Another important special case is the one where Π approximates the vacuum arbitrarily well, which is the situation considered in Refs. [7,8].

Implementation of a building block is outlined in Fig. 1 and is as follows:

(1) Take two m-mode quantum states ρ_A (modes Aj) and ρ_B (modes Bj).

(2) Each of the m parties mixes their two modes on a beam splitter of reflectivity R.

(3) On each of the beam splitters, take the output from the B modes and locally implement the Gaussian measurement with local POVM elements $\{\Pi_{\mathbf{r}}\}$.

(4) Given measurement outcome data $\mathbf{r}$, postselect declaring a success with probability $P(\mathbf{r})$.

(5) Take the unmeasured A modes and output from the building block.

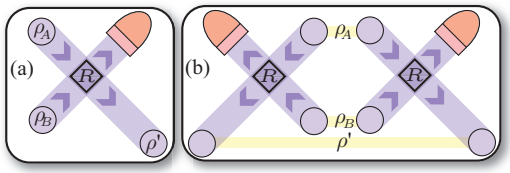

FIG. 1. (Color online) An implementation of an individual building block with beam-splitter reflectivity R for (a) single-mode states and (b) two-mode states. Generalization to m-mode states is straightforward as each additional party performs the same local unitaries.

Here, we have labeled the $2m$ modes as $\{A1,\ldots,Am,B1,\ldots,Bm\}$ and modes sharing the same numerical index share the same physical location. When successful, the building block outputs a state

$$\rho' \propto \int P(\mathbf{r})\mathrm{tr}_B[U(\rho_A \otimes \rho_B)U^{\dagger}(\mathbb{1} \otimes \Pi_{\mathbf{r}})]d\mathbf{r}. \tag{17}$$

The unitary U represents the effect of the beam splitters such that for all j,

$$U^{\dagger}\hat{a}_{Aj}U = \sqrt{T}\hat{a}_{Aj} + \sqrt{R}\hat{a}_{Bj}, \tag{18}$$

where $T = 1 - R$. Taking the integral over measurement outcomes inside the partial trace and using Eq. (15) we have

$$\rho' \propto \mathrm{tr}_B[U(\rho_A \otimes \rho_B)U^{\dagger}(\mathbb{1} \otimes \Pi)]. \tag{19}$$

Unfortunately, the effect of this map can be difficult to analytically evaluate. The root of the technicalities is related to the fact that U and $\mathbb{1} \otimes \Pi$ do not commute. However, following the insights of Ref. [44], we know that by moving to phase space and working with a different object from ρ' the effect of the map can be simplified. This is the key insight that renders the analysis feasible. In Ref. [44] the characteristic function of the non-Hermitian object $\rho'\Pi$ was considered. This work follows parallel reasoning but instead we consider the Hermitian object $\Pi^{1/2}\rho'\Pi^{1/2}$ and its characteristic function. We make use of

$$\tau' = \frac{P\rho' P}{\mathrm{tr}(P\rho' P)} \tag{20}$$

for the normalized and Hermitian filtered object, with

$$P = \Pi^{1/2}. \tag{21}$$

This object is then

$$\tau' \propto \mathrm{tr}_B[(P \otimes \mathbb{1})U(\rho_A \otimes \rho_B)U^{\dagger}(P \otimes \Pi)]. \tag{22}$$

Splitting $\Pi = PP$ and using the cyclicity of the trace we have the more symmetric formula

$$\tau' \propto \mathrm{tr}_B[(P \otimes P)U(\rho_A \otimes \rho_B)U^{\dagger}(P \otimes P)]. \tag{23}$$

The next fact we employ is that for any Gaussian operator with zero first moments, such as P, we have that

$$U^{\dagger}(P \otimes P)U = P \otimes P. \tag{24}$$

This equality is well known (see, e.g., Ref. [51]), but for completeness we give a proof in Appendix A. Hence we have

$$\tau' \propto \mathrm{tr}_B[U(P\rho_A P \otimes P\rho_B P)U^{\dagger}]. \tag{25}$$

Again using the shortened notation $\tau_A \propto P\rho_A P$ and $\tau_B \propto P\rho_B P$ gives

$$\tau' \propto \mathrm{tr}_B[U(\tau_A \otimes \tau_B)U^\dagger]. \tag{26}$$

By choosing Π, and equivalently P, as proportional to a Gaussian state, we have been able to exploit the symmetry of the problem to reach a greatly simplified expression. The characteristic function of this object is then

$$\chi_{\tau'}(\mathbf{r}) \propto \mathrm{tr}[(\mathbb{1} \otimes D_{\mathbf{r}})U(\tau_A \otimes \tau_B)U^\dagger]. \tag{27}$$

Conjugating $U^\dagger$ with the displacement operator gives

$$\begin{aligned} U^\dagger(\mathbb{1} \otimes D_{\mathbf{r}})U &= U^\dagger \exp[i(\mathbb{1} \otimes \mathbf{r}.\hat{\mathbf{Q}})]U \\ &= \exp[i\sqrt{T}(\mathbf{r}\cdot\hat{\mathbf{Q}} \otimes \mathbb{1}) + i\sqrt{R}(\mathbb{1} \otimes \mathbf{r}\cdot\hat{\mathbf{Q}})] \\ &= D_{\sqrt{T}\mathbf{r}} \otimes D_{\sqrt{R}\mathbf{r}}. \end{aligned} \tag{28}$$

Using this relation we deduce that

$$\begin{aligned} \chi_{\tau'}(\mathbf{r}) &\propto \mathrm{tr}[D_{\sqrt{T}\mathbf{r}}\tau_A \otimes D_{\sqrt{R}\mathbf{r}}\tau_B] \\ &\propto \mathrm{tr}[D_{\sqrt{T}\mathbf{r}}\tau_A]\mathrm{tr}[D_{\sqrt{R}\mathbf{r}}\tau_B]. \end{aligned} \tag{29}$$

However, these factors are simply the characteristic functions for τ_A and τ_B but with a modified value of $\mathbf{r}$, so

$$\chi_{\tau'}(\mathbf{r}) = \chi_{\tau_A}(\sqrt{T}\mathbf{r})\chi_{\tau_B}(\sqrt{R}\mathbf{r}). \tag{30}$$

We have shifted to equality, rather than proportionality, because the characteristic function of a unit trace object takes $\chi_\tau(0) = 1$. As promised, the effect of the protocol on the filtered τ objects is much more straightforward than for the actual density matrices. Note that if we considered non-Hermitian objects $\sigma_{A,B} \propto \rho_{A,B}\Pi$ and output $\sigma' = \rho'\Pi$, we would have similarly arrived at

$$\chi_{\sigma'}(\mathbf{r}) = \chi_{\sigma_A}(\sqrt{T}\mathbf{r})\chi_{\sigma_B}(\sqrt{R}\mathbf{r}). \tag{31}$$

These results generalize those of Ref. [44], where the input states were taken to be identical and reflectivity set to be 50/50 and only the non-Hermitian objects were considered. Later in this article, we find that working with Hermitian objects proves to be the more elegant approach.

Before proceeding we remark on the assumption that Π, and hence all P, are invertible. The assumption is required to ensure that τ' uniquely defines ρ'. All Gaussian operators, except projectors, are full rank and invertible so the assumption simply rules out projectors. However, we wish for our general analysis to encompass previous protocols [7,8,43] that prescribe projecting two modes onto the vacuum, where $\Pi = P = |0,0\rangle\langle 0,0|$, which is clearly not invertible. However, any realistic experiment will use detectors with some nonunit efficiency of photon detection. Indeed, often efficiency is significantly less than unity. Such inefficiencies can be modeled by placing a beam splitter ahead of the detector, and can be easily incorporated into our analysis. This modification results in a realistic filter that is still Gaussian but no longer a projector. As such, the assumption of invertible filters is always justified.

IV. PROTOCOLS

A. The recursive Gaussifer

The first class of protocols we review was originally introduced in Ref. [44], generalizing the proposals of

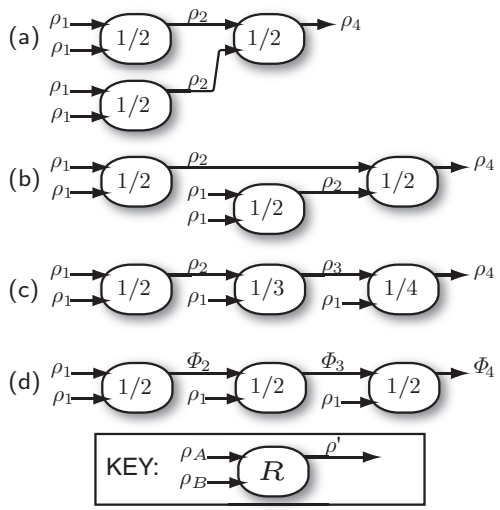

FIG. 2. Different protocols combining building blocks in ways. All building blocks use the same filter, Π, and are labeled with their beam-splitter reflectivity R. (a) The recursive-Gaussifier protocol; (b) the temporally reordered recursive-Gaussifier protocol; (c) the pumping-Gaussifier protocol; and (d) the compact distillery protocol. The key shows how the building block labels compare with the variables used in Sec. II.

Refs. [7,8]. We refer to the protocols considered here as recursive Gaussifers and the general structure is outlined in Fig. 2(a). All building blocks of the recursive protocol use the same filter Π, and set $R = T = 1/2$. In the first round of the protocol many copies of a raw state ρ_1 are taken and are simultaneously used as inputs to building blocks, with $\rho_A = \rho_B = \rho_1$. The successful outputs from these rounds are labeled ρ_2, and are used as the inputs into the building blocks for the next round. On the nth round, each building block takes two input states labeled ρ_{2^n} and outputs $\rho_{2^{n+1}}$. The subscript counts the number of raw copies so far consumed. Denoting $\tau_{2^n} \propto P\rho_{2^n}P$ and applying Eq. (31) we find that

$$\chi_{\tau_{2^{n+1}}}(\mathbf{r}) = \chi_{\tau_{2^n}}\left(\frac{\mathbf{r}}{\sqrt{2}}\right)^2, \tag{32}$$

which is easier to represent in terms of $N = 2^n$ so

$$\chi_{\tau_{2N}}(\mathbf{r}) = \chi_{\tau_N}\left(\frac{\mathbf{r}}{\sqrt{2}}\right)^2. \tag{33}$$

In terms of τ_1 we have

$$\chi_{\tau_N}(\mathbf{r}) = \chi_{\tau_1}\left(\frac{\mathbf{r}}{\sqrt{N}}\right)^N. \tag{34}$$

To reach n rounds, assuming every building block succeeds, we must have a memory capable of storing $N = 2^n$ copies of ρ_1 simultaneously. The exponential increase in memory is required because we have assumed simultaneous execution of all building blocks within a round. However, relaxing the simultaneity requirement and using a smart ordering—for instance as in Fig. 2(b)—the recursive protocol can implement n rounds with a storage capacity of $n + 1$ modes per location, albeit at the cost of increasing the number of time steps. A growing quantum memory seems unavoidable, but we will soon see how it can be circumvented. The sequence of

characteristic functions

$$\{\chi_{\tau_1}, \chi_{\tau_2}, \chi_{\tau_4}, \chi_{\tau_8}, \dots\} \tag{35}$$

is known to evolve toward a Gaussian with unchanged second moments by virtue of a central limit theorem. We will later review central limit theorems, providing extensions to make more direct statements about the physical state.

B. The pumping Gaussifier

We propose protocols that use a fixed initial state to repeatedly pump a target state, surprisingly resulting in the same output as an analogous recursive protocol. The building blocks that compose the pumping Gaussifier use two distinct input states in later rounds and also weaken the beam-splitter reflectivity with the number of steps. On the Nth step, we take a copy of ρ_N and a raw initial state ρ_1 and mix on a beam splitter of reflectivity $R_N = 1/(N+1)$ as shown in Fig. 2(c). The output is labeled ρ_{N+1} and in the phase space picture we have the iterative formula

$$\chi_{\tau_{N+1}}(\mathbf{r}) = \chi_{\tau_N}\left(\frac{\sqrt{N}}{\sqrt{N+1}}\mathbf{r}\right)\chi_{\tau_1}\left(\frac{1}{\sqrt{N+1}}\mathbf{r}\right). \tag{36}$$

We can verify that

$$\chi_{\tau_N}(\mathbf{r}) = \chi_{\tau_1}\left(\frac{\mathbf{r}}{\sqrt{N}}\right)^N \tag{37}$$

satisfies the iterative formula because

$$\begin{aligned}\chi_{\tau_{N+1}}(\mathbf{r}) &= \chi_{\tau_1}\left(\frac{\sqrt{N}}{\sqrt{N+1}}\frac{\mathbf{r}}{\sqrt{N}}\right)^N \chi_{\tau_1}\left(\frac{\mathbf{r}}{\sqrt{N+1}}\right)\\ &= \chi_{\tau_1}\left(\frac{\mathbf{r}}{\sqrt{N+1}}\right)^N \chi_{\tau_1}\left(\frac{\mathbf{r}}{\sqrt{N+1}}\right)\\ &= \chi_{\tau_1}\left(\frac{\mathbf{r}}{\sqrt{N+1}}\right)^{N+1}.\end{aligned} \tag{38}$$

The neat cancellation of $\sqrt{N}/\sqrt{N}$ only occurs because of our exact choice of beam-splitter reflectivity. After the Nth step, the characteristic function matches that of the recursive Gaussifier implemented to depth $n = \log_2(N)$. Furthermore, for successful implementations both protocols consume the same number of raw copies to achieve the same output. However, in the pumping protocol we also have the option of terminating after a number of steps not of the form $N = 2^n$.

C. The compact distillery

The compact distillery (CD) protocol [43] also repeatedly pumps with the same initial state, but it keeps a constant beam-splitter reflectivity of $R = 1/2$ as outlined in Fig. 2(d). The CD protocol is known to provide a very different evolution from both our Gaussifier protocols. To highlight that it produces different states from the Gaussifiers, we label the output of the Nth step as Φ_{N+1} and equate $\Phi_1 = \rho_1$ for the raw resource. Denoting $\phi_N \propto P\Phi_N P$ we have the iterative relation

$$\chi_{\phi_{N+1}}(\mathbf{r}) = \chi_{\phi_N}\left(\frac{\mathbf{r}}{\sqrt{2}}\right)\chi_{\phi_1}\left(\frac{\mathbf{r}}{\sqrt{2}}\right). \tag{39}$$

We can immediately deduce properties of ϕ_{N+1} from those of the initial operator ϕ_1. For instance, if the characteristic function χ_{ϕ_1} is zero at point $\mathbf{r}_0$, then the characteristic function $\chi_{\phi_{N+1}}$ is zero at $\sqrt{2}\mathbf{r}_0$ for all N. Hence, the characteristic function $\chi_{\phi_{N+1}}$ will not have a Gaussian shape and consequently the corresponding physical state Φ_{N+1} will also be non-Gaussian. If there exists a limiting characteristic function χ_{ϕ_∞} the same argument applies and so non-Gaussianity would persist even in the asymptotic limit of many iterations. Indeed, all the examples considered in Ref. [43] found that the protocol converges toward non-Gaussian states. Our phase space techniques provide a clear explanation of *why* non-Gaussianity persists in the compact distillery. This illustrates the merit of the phase space perspective, even for examining protocols that do not Gaussify.

The CD protocol was proposed as an alternative to recursive Gaussifiers to reduce the required quantum memory and bring protocols closer to experimental feasibility. However, we have seen that our pumping Gaussifier can also operate under these stringent memory constraints. We must then consider other figures of merit to compare these protocols. The authors of Ref. [43] showed that, when fed with weakly entangled photon subtracted states, a few rounds of the CD achieves a similar entanglement increase as a few rounds of the Gaussifier. However, the maximum achievable entanglement of the Gaussifer proved to be much higher, and so after only three to four rounds the advantage of the pumping Gaussifier can be significant. Of course, whether we desire the output state to be non-Gaussian or Gaussian depends on the context and what quantum information protocol the resource is subsequently used for.

V. CENTRAL LIMIT THEOREMS

A. Characteristic function convergence

Central limit theorems are results that tell us when a sequence of characteristic functions approaches a Gaussian function and in what way they converge. Throughout we are interested in sequences of characteristic functions output by the recursive and pumping Gaussifers.

Definition 1 (Central limit sequence). We say a sequence of Hermitian positive operators $\{\tau_N\}$ and associated characteristic functions $\{\chi_{\tau_N}\}$ is a central limit sequence if

$$\chi_{\tau_N}(\mathbf{r}) = \chi_{\tau_1}\left(\frac{\mathbf{r}}{\sqrt{N}}\right)^N, \tag{40}$$

where χ_{τ_1} has zero first moments and Γ_τ second moments.

For such a sequence, if τ_1 is Hermitian and positive, then the results of Refs. [44,45,52] govern its limiting behavior. More generally, if τ_1 is non-Hermitian, then recent results [44] give conditions under which it approaches a Gaussian. These latter techniques were used to demonstrate Gaussification of physical systems by considering $\tau_1 \propto \rho_1\Pi$. Here we consider the Hermitian $\tau_1 \propto P\rho_1 P$ for which the convergence properties are simpler to state.

Theorem 1 (General quantum central limit theorem). Consider a central limit sequence $\{\chi_{\tau_N}\}$. For any finite radius r_0 and any accuracy $\epsilon > 0$, there exists an N_ϵ such that for all

$N \geqslant N_\epsilon$ and all $|\mathbf{r}| \leqslant r_0$ we have

$$\left|\chi_{\tau_N}(\mathbf{r}) - \chi_{\tau_\infty}(\mathbf{r})\right| < \epsilon, \tag{41}$$

where $\chi_{\tau_\infty}(\mathbf{r})$ is a Gaussian with covariance matrix Γ_{τ_1}.

The theorem can be proven by taking a cross section of the characteristic function for a unit direction $\mathbf{r}$, such that

$$f_N(t) = \chi_{\tau_N}(t\mathbf{r}), \tag{42}$$

and proving convergence to a Gaussian function in phase space for all such cross sections. Each cross section is equivalent to a characteristic function for a classical probability distribution. We may proceed by following one of the numerous classical proofs, such as Ref. [53]. Central limit theorems are fundamental to our method and so for completeness we will provide a proof here.

From the definition of a characteristic function, it follows that it can be expanded as

$$f_1(t) = 1 - \frac{t^2}{2}\nu + C(t^2), \tag{43}$$

where ν is the second moment in direction $\mathbf{r}$, such that

$$\nu = 2\,\mathrm{tr}[(\mathbf{r}\cdot\hat{\mathbf{Q}})^2\rho], \tag{44}$$

and the higher order terms $C(x^2)$ can be shown [44,53] to satisfy $C(x^2)/x^2 \to 0$ as $x \to 0$. Hence, the Nth function in the sequence is

$$f_N(t) = \left(1 - \frac{t^2}{2N}\nu + C(t^2/N)\right)^N. \tag{45}$$

We wish to compare this with $\exp(-t^2\nu)$, and so the difference of these quantities is

$$\delta_N(t) = |f_N(t) - \exp(-t^2\nu)|. \tag{46}$$

We can approximate $\exp(-t^2\nu)$ with some $(1 - t^2\nu/N)^N$ to any accuracy $\epsilon/2 > 0$, such that there exists an N'_ϵ and for $N > N'_\epsilon$ we have

$$\delta_N(t) \leqslant \left|\left[1 - \frac{t^2\nu}{N} + C\left(\frac{t^2}{N}\right)\right]^N - \left(1 - \frac{\nu}{N}\right)^N\right| + \frac{\epsilon}{2}.$$

Next, we use that for any complex numbers a and b with $|a| \leqslant 1$ and $|b| \leqslant 1$ we know (see Appendix B) that $|a^N - b^N| \leqslant N|a - b|$ and applying this yields

$$\begin{aligned}\delta_N(t) &\leqslant N\left|\left[1 - \frac{t^2\nu}{N} + C\left(\frac{t^2}{N}\right)\right] + \left(1 - \frac{\nu}{N}\right)\right| + \frac{\epsilon}{2} \\ &= N|C(t^2/N)| + \frac{\epsilon}{2} = t^2|C(x^2)/x^2| + \frac{\epsilon}{2},\end{aligned} \tag{47}$$

where $x^2 = t^2/N$. For constant t, we can decrease x to any desired value by increasing N. Since $C(x^2)/x^2$ vanishes in this limit, for any desired $\eta = \epsilon/2t^2 > 0$ we can find a N''_ϵ such that for all $N > N''_\epsilon$ we have $|C(x^2)/x^2| \leqslant \eta$. Hence, we have

$$\delta_N(t) \leqslant t^2\eta + \epsilon/2 = \epsilon. \tag{48}$$

This final result holds for $N > \max(N'_\epsilon, N''_\epsilon) = N_\epsilon$. The above argument tells us how individual points evolve in N, but the result can be strengthened further for all points within a ball of finite radius r_0. This extension to finite regions of phase space is outlined in Appendix C. This result is stronger as the same error bound uniformly holds across a whole region simultaneously. The region has a finite area and extensions of this result to the whole of phase space do not hold. Indeed, central limit theorems are aptly named as they dictate the limiting behavior around the origin of phase space but *not* into the tails (see also the similar discussion related to noncommutative central limit theorems applied to grasping quantum many-body dynamics [54,55]).

B. Convergence of moments

Next, we present a second aspect of central limit theorems, which we use later, that quantifies the evolution of higher moments. We begin by generalizing the idea of a quadrature. Typically, quadratures are thought of as single-mode position or momentum operators, but we take quadratures to include all linear combinations of such operators, such that

$$H = \sum_j r_j \hat{Q}_j \tag{49}$$

is always a quadrature. The kth moment of such an operator, assuming first moments are zero, is the expectation value of H^k. More generally, we say an operator is a kth moment if it is a product of k, potentially distinct quadratures such that

$$H^{(k)} = \prod_{j=1}^{k} H_j, \tag{50}$$

where each H_j is linear in quadrature operators as in Eq. (49). Another result known as a central limit theorem is the following.

Theorem 2 (Convergence of moments). For any central limit sequence $\{\tau_N\}$ and any kth moment $H^{(k)}$ in the large N limit,

$$|\mathrm{tr}(H^{(k)}\tau_N) - \mathrm{tr}(H^{(k)}\tau_\infty)| \to 0. \tag{51}$$

A simplified proof of this result is presented in Appendix D, but more involved proofs of more general results can be found in Refs. [56,57]. The theorem can be easily extended to finite linear sums of moments as follows.

Corollary 1 (Finite sums of moments). Consider an operator H, which is a sum of finitely many terms, each a kth moment. The sequence of operators τ_N for increasing N obeys

$$|\mathrm{tr}(H\tau_N) - \mathrm{tr}(H\tau_\infty)| \to 0. \tag{52}$$

C. Matrix element convergence

The above theorems tell us about the evolution of the characteristic functions and moments but what can be said on the level of the density matrices τ_N? We have the following.

Theorem 3 (Pointwise convergence). Consider a central limit sequence $\{\tau_N\}$ and a pair of pure states $\{|\psi_k\rangle, |\psi_j\rangle\}$, in the limit of large N,

$$|\langle\psi_k|\tau_N|\psi_j\rangle - \langle\psi_k|\tau_\infty|\psi_j\rangle| \to 0. \tag{53}$$

This tells us that individual matrix elements converge toward a fixed value and we give a proof in Appendix E. This result informs us of the evolution of the filtered object $\tau_N = P\rho_N P/\mathrm{tr}(P\rho_N P)$. However, we really want to know about the physical state ρ_N, and this is the problem we turn to in the next section.

VI. CONVERGENCE OF PHYSICAL STATE

Knowing the filtered object obeys a central limit theorem, we can draw conclusions on the evolution of the actual physical state. Recall that earlier we demanded, without loss of generality, that P was an invertible matrix. This assumption allows us to conclude that there exists a unique operator,

$$\rho_N \propto P^{-1}\tau_N P^{-1}. \tag{54}$$

Concerning these states we shall show the following.

Theorem 4 (State convergence). Consider a central limit sequence $\{\tau_N\}$ with limiting Gaussian operator τ_∞ and covariance matrix Γ_τ. Denote γ as the covariance matrix of the CJ state [see Eq. (14)] isomorphic to the channel $\mathcal{P}$, such that $\mathcal{P}(\rho) = P\rho P$ for some Gaussian P. If the covariance matrix

$$\Gamma_{\rho_\infty} = \gamma_{AB}^T(\gamma_{AA} - \Gamma_{\tau_\infty})^{-1}\gamma_{AB} - \gamma_{BB} \tag{55}$$

exists and is physical, then $\rho_\infty \propto P^{-1}\tau_\infty P^{-1}$ exists and is a Gaussian state with covariance matrix Γ_{ρ_∞}. Furthermore, if $\{|\psi_k\rangle,|\psi_j\rangle\}$ are eigenvectors of P, then the sequence $\{\rho_N\}$ in the large N limit satisfies

$$\left|\frac{\langle\psi_k|\rho_N|\psi_j\rangle}{\mathrm{tr}(P\rho_N P)} - \frac{\langle\psi_k|\rho_\infty|\psi_j\rangle}{\mathrm{tr}(P\rho_\infty P)}\right| \to 0. \tag{56}$$

Above we define a limiting physical state and show a weak form of convergence of the density matrix elements up to a normalization factor. It is worth noting that most existing results in the literature only go this far, though we will be interested in going further.

Corollary 2 (Fidelity convergence). In addition to Theorem 4, if also in the large N limit we have $\mathrm{tr}(P\rho_N P) \to \mathrm{tr}(P\rho_\infty P)$, then also

$$F(\rho_N,\rho_\infty) \to 1, \tag{57}$$

where F is the fidelity between its arguments.

Let us prove this straightforward corollary. If $\mathrm{tr}(P\rho_N P)$ converges to $\mathrm{tr}(P\rho_\infty P)$, then we have that for increasing N,

$$|\langle\psi_k|\rho_N|\psi_j\rangle - \langle\psi_k|\rho_\infty|\psi_j\rangle| \to 0. \tag{58}$$

Furthermore, it is well known that for physical states elementwise convergence of the density matrix entails convergence in terms of fidelity and other measures of similarity such as trace norm distance [52]. However, the corollary rests upon an additional key assumption that is the focus of the next section.

To prove our state convergence theorem we first find Γ_{τ_∞} in terms of Γ_{ρ_∞}, under the assumption that ρ_∞ is Gaussian. Since P is invertible, there exists a unique physical state, defined by $P\rho_\infty P \propto \tau_\infty$. In light of this uniqueness, the Gaussianity of ρ_∞ is assured provided that a Gaussian solution to $P\rho_\infty P \propto \tau_\infty$ exists. The operators are related by a CP map, $A \mapsto \mathcal{P}(A) = PAP$ with Gaussian P, and so we can apply the results of Refs. [4–6] on Gaussian channels and the CJ isomorphism (reviewed earlier). This tells us that for channel $\mathcal{P}$ with the Gaussian CJ state acting on a Gaussian input state, the covariance matrices are related such that

$$\Gamma_{\tau_\infty} = \gamma_{AA} - \gamma_{AB}(\gamma_{BB} + \Gamma_{\rho_\infty})^{-1}\gamma_{AB}^T, \tag{59}$$

where γ is as defined in Eq. (14). To reach Eq. (55) we simply rearrange the above expression for Γ_{ρ_∞}.

Furthermore, denoting $\{|\psi_j\rangle\}$ as the eigenvectors of P with eigenvalue λ_j, we can apply Theorem 3 with respect to $\{|\psi_j\rangle,|\psi_k\rangle\}$. Consequently, for large enough N,

$$\left|\frac{\lambda_j\lambda_k\langle\psi_k|\rho_N|\psi_j\rangle}{\mathrm{tr}(P\rho_N P)} - \frac{\lambda_j\lambda_k\langle\psi_k|\rho_\infty|\psi_j\rangle}{\mathrm{tr}(P\rho_\infty P)}\right| \to 0. \tag{60}$$

After canceling the $\lambda_j\lambda_k$ factors we have proven Theorem 4.

A. Convergence in fidelity

In the previous section we made very general, but weak, predictions on the evolution of the physical state. In order to deduce stronger conclusions, as captured by Corollary 2, we need that $\mathrm{tr}(P\rho_N P)$ converges to the value $\mathrm{tr}(P\rho_\infty P)$. Whether our protocols work correctly rests on the validity of this assumption. The assumption appears fairly innocuous but is actually quite subtle, and surprisingly, instances exist where it fails. We remedy the neglect of this important assumption.

Some sufficient conditions have been found for this assumption [44]. We strengthen these results, providing the basis for studies in subsequent sections. Our result makes use of the idea of a reference state that we first define.

Definition 2 (Reference state). Consider an operator τ and a Gaussian filter $\Pi \propto \exp(-\sum_j \beta_j \hat{b}_j^\dagger \hat{b}_j)$, where $\hat{b}_j = V\hat{a}_j V^\dagger$ for some Gaussian unitary V. If τ_{ref} is a Gaussian state, we write $\tau \leqslant_\Pi \tau_{\mathrm{ref}}$ if both of the following are satisfied: (i) $|\mathrm{tr}(H^{(k)}\tau)| \leqslant |\mathrm{tr}(H^{(k)}\tau_{\mathrm{ref}})|$ and (ii) $|\mathrm{tr}(H^{(k)}\tau_{\mathrm{ref}})| = \mathrm{tr}(H^{(k)}\tau_{\mathrm{ref}})$ for all moments $H^{(k)}$ composed of finite products of $\{\hat{b}_j^\dagger,\hat{b}_j\}$. When $\tau \leqslant_\Pi \tau_{\mathrm{ref}}$ we say τ_{ref} is a reference state for τ with respect to Π.

The concept is especially useful when considering central limit sequences because of the following.

Lemma 1 (Persistence of reference state). Consider a central limit sequence $\{\tau_N\}$ and a Gaussian filter Π. If there exists a $\tau_j \in \{\tau_N\}$ and Gaussian τ_{ref} such that $\tau_j \leqslant_\Pi \tau_{\mathrm{ref}}$, then for all $N \geqslant j$ we have $\tau_N \leqslant_\Pi \tau_{\mathrm{ref}}$.

That the reference state remains good for all N can be proven iteratively. For any kth moment,

$$\mathrm{tr}(H^{(k)}\tau_{N+1}) = \mathrm{tr}[U^\dagger(\mathbb{1}\otimes H^{(k)})U(\tau_N\otimes\tau_1)]. \tag{61}$$

The conjugation of $H^{(k)}$ by U gives a sum of 2^k terms, each a product of $\{\hat{b}_j^\dagger,\hat{b}_j\}$ operators. We label each term by x, with it having the form $H_x^{(k-j_x)}\otimes H_x^{(j_x)}$ for some integer j_x that depends on x. In particular, for every j the binomial "k choose j" counts the multiplicity of x values for which $j_x = j$. In this notation

$$\begin{aligned}\mathrm{tr}(H^{(k)}\tau_{N+1}) &= \sum_x C_x \mathrm{tr}\big[\big(H_x^{(k-j_x)}\otimes H_x^{j_x}\big)(\tau_N\otimes\tau_1)\big]\\ &= \sum_x C_x \mathrm{tr}\big(H_x^{(k-j_x)}\tau_N\big)\mathrm{tr}\big(H_x^{(j_x)}\tau_1\big),\end{aligned}$$

where $C_x = T_N^{(k-j_x)/2}R_N^{j_x/2}$. Assuming that the properties of reference states hold for τ_N, we have for τ_N that

$$\begin{aligned}|\mathrm{tr}(H^{(k)}\tau_{N+1})| &\leqslant \sum_x C_x\big|\mathrm{tr}\big(H_x^{(k-j_x)}\tau_N\big)\big|\big|\mathrm{tr}\big(H_x^{(j_x)}\tau_1\big)\big|\\ &\leqslant \sum_x C_x \mathrm{tr}\big(H_x^{(k-j_x)}\tau_{\mathrm{ref}}\big)\mathrm{tr}\big(H_x^{(j_x)}\tau_{\mathrm{ref}}\big).\end{aligned}$$

Next we recall that Gaussian states are invariant under the beam-splitter unitary $U(\tau_{\rm ref}\otimes\tau_{\rm ref})U^\dagger=\tau_{\rm ref}\otimes\tau_{\rm ref}$, as was shown in Appendix A. Being invariant under beam splitters, Gaussian states must also be fixed points of the protocol and since $\tau_{\rm ref}$ is Gaussian we infer

$$\mathrm{tr}(H^{(k)}\tau_{\rm ref})=\sum_x C_x\,\mathrm{tr}\big(H_x^{(k-j_x)}\tau_{\rm ref}\big)\mathrm{tr}\big(H_x^{(j_x)}\tau_{\rm ref}\big). \tag{62}$$

Using this invariance and applying it to the problem at hand we conclude

$$|\mathrm{tr}(H^{(k)}\tau_{N+1})|\leqslant\mathrm{tr}(H^{(k)}\tau_{\rm ref}). \tag{63}$$

This proves, as claimed earlier, that when a reference state has the desired properties with respect to some τ_j, it automatically follows for all $\tau_{N\geqslant j}$. The concept of a reference state is fundamental to the following result.

Theorem 5 (Convergence in fidelity). Consider a central limit sequence $\{\tau_N\}$ and filter Π. If there exists a $\tau_j\in\{\tau_N\}$ and Gaussian $\tau_{\rm ref}$ such that $\tau_j\leqslant_\Pi\tau_{\rm ref}$ and $\mathrm{tr}(\Pi^{-1}\tau_{\rm ref})<\infty$, then

$$\mathrm{tr}(\Pi\rho_N)\to\mathrm{tr}(\Pi\rho_\infty), \tag{64}$$

where $\rho_N=P^{-1}\tau_N P^{-1}/\mathrm{tr}(P^{-1}\tau_N P^{-1})$. Furthermore, as N increases

$$F(\rho_N,\rho_\infty)\to 1. \tag{65}$$

This tells us that, assuming a suitable reference exists, the convergence behavior of the operators τ_N is inherited by the physical states ρ_N. In Ref. [44] a similar result for the case $\tau_{\rm ref}=\tau_\infty$ was shown. Although this is useful in some cases, often τ_∞ will not always satisfy the conditions for a reference state and so this result allows us to use another operator as a proxy.

Our approach to the proof is to find $\mathrm{tr}(\Pi\rho_N)$ by calculating the expectation value of τ_N with respect to Π^{-1}. These quantities are related by

$$\mathrm{tr}(\Pi^{-1}\tau_N)=\frac{\mathrm{tr}(\Pi\Pi^{-1}\rho_N)}{\mathrm{tr}(\Pi\rho_N)}=\frac{1}{\mathrm{tr}(\Pi\rho_N)}. \tag{66}$$

The Gaussian filter can always be written as the exponential of some Hamiltonian

$$H_\Pi=\sum_j\beta_j V\hat{a}_j^\dagger\hat{a}_j V^\dagger, \tag{67}$$

such that $\Pi=\exp(-H_\Pi)$, where H_Π is Hermitian and quadratic in annihilation/creation operators. The inverse filter is then $\Pi^{-1}=\exp(+H_\Pi)$ and

$$\mathrm{tr}(\Pi^{-1}\tau_N)=\mathrm{tr}\left(\sum_{k=0}^{\infty}\frac{H_\Pi^k}{k!}\tau_N\right). \tag{68}$$

Each term is a sum of moments of degree $2k$ so it is tempting to think that Theorem 2 can be directly applied. However, the whole sum has infinitely many terms so Theorem 2 is not applicable. Each $\mathrm{tr}(H_\Pi^k\tau_{\rm ref})$ is positive and, by assumption, the infinite sum gives a finite value. It follows that for any $\epsilon>0$ we can pick an integer k_c such that the truncation satisfies

$$\left|\mathrm{tr}\left(\sum_{k=k_c+1}^{\infty}\frac{H_\Pi^k}{k!}\tau_{\rm ref}\right)\right|<\epsilon, \tag{69}$$

for the reference state $\tau_{\rm ref}$. Furthermore, using this k_c, we can partition the summation for τ_N such that

$$\mathrm{tr}(\Pi^{-1}\tau_N)=\mathrm{tr}\left(\sum_{k=0}^{k_c}\frac{H_\Pi^k}{k!}\tau_N\right)+\mathrm{tr}\left(\sum_{k=k_c+1}^{\infty}\frac{H_\Pi^k}{k!}\tau_N\right). \tag{70}$$

Now, the first term is a finite sum and so the results of Theorem 2 do apply to this portion of the sum. Hence, for sufficiently large N

$$\begin{aligned}|\mathrm{tr}[\Pi^{-1}(\tau_N-\rho_\infty)]|&\leqslant\epsilon+\left|\mathrm{tr}\left(\sum_{k=k_c+1}^{\infty}\frac{H_\Pi^k}{k!}(\tau_N-\tau_\infty)\right)\right|,\\&\leqslant\epsilon+2\left|\mathrm{tr}\left(\sum_{k=k_c+1}^{\infty}\frac{H_\Pi^k}{k!}\tau_{\rm ref}\right)\right|,\end{aligned} \tag{71}$$

where in the last line we have used the properties of a reference state. Combining this with (69) we deduce that for large enough N,

$$|\mathrm{tr}[\Pi^{-1}(\tau_N-\rho_\infty)]|\leqslant 3\epsilon. \tag{72}$$

By taking longer truncations k_c and larger N, the value of ϵ can be made arbitrarily small. Therefore, we have that for increasing N,

$$\mathrm{tr}(\Pi^{-1}\tau_N)\to\mathrm{tr}(\Pi^{-1}\rho_\infty). \tag{73}$$

Consequently, $\mathrm{tr}(\Pi\rho_N)$ approaches $\mathrm{tr}(\Pi\rho_\infty)$ and the fidelity between these states approaches unity.

These techniques, in particular the use of reference states, give us a handle on this difficult part of the analysis. The central limit theorems ensure that the filtered operators converge to a Gaussian. However, alone, the central limit theorems provide no guarantees on the behavior of expectation values for unnormalizable operations like Π^{-1}. Indeed, it is easy to find central limit sequences for which $\mathrm{tr}(\Pi^{-1}\tau_N)$ diverges with N. In such pathological examples, the physical states ρ_N would also diverge with ever increasing energy. However, in light of the arguments presented, when a suitable reference state exists these pathologies cannot occur.

The limiting operator τ_∞ may sometimes be chosen as a reference state, but in some cases it is unsuitable. Now we will discuss a few facts that simplify the task of finding a suitable reference state. First, we note that if

$$|\mathrm{tr}(H^{(k)}\tau_{\rm ref})|=\mathrm{tr}(H^{(k)}\tau_{\rm ref}) \tag{74}$$

holds for all second moments, then it must hold for all higher moments also. By Wick's theorem (see Appendix D) the higher moments for Gaussian states are simply a positive polynomial in second moments. Consequently, positivity of higher moments is inherited from positivity of second moments, which simplifies the search for appropriate reference states.

For single-mode states there is one very simple class of potential reference states. Consider the pure squeezed states

$$|\psi_R\rangle=\sum_{n=0}^{\infty}\lambda^n|2n\rangle, \tag{75}$$

where $0<\lambda<1$ and $|\lambda|=\lambda$. Calculating $\langle\psi_R|\hat{a}^\dagger\hat{a}|\psi_R\rangle$ and $\langle\psi_R|\hat{a}\hat{a}|\psi_R\rangle$ we find they are real, positive, and increasing

with λ. These form a promising class of single-mode reference states, as for any τ_1 and any even moment $H^{(k)}$ we can find a large enough λ such that $|\mathrm{tr}(H^{(k)}\tau_1)| < \mathrm{tr}(H^{(k)}\tau_{\mathrm{ref}})$. However, Theorem 5 also requires that $\mathrm{tr}(\Pi^{-1}\tau_{\mathrm{ref}}) < \infty$, but there will be a critical value of λ at which this expectation value diverges. For many single-mode central limit sequences there will exist a choice of λ that satisfies both these requirements, though some counterexamples do exist. For multimode problems, pure squeezed or entangled states can make suitable reference states.

Above, we focused on the even moments of the Gaussian state. It is easy to check that all Gaussian states, with zero first moments, have vanishing odd moments. This seems to entail severe constraints on the odd moments of τ_1. However, this problem can be remedied by a physical procedure that is a CV version of twirling. The concept of twirling, arising also in entanglement distillation of finite-dimensional systems [58] and magic state distillation [59], generates a symmetry in the initial resource. This symmetry significantly simplifies the analysis of a protocol's convergence. The twirling map we prescribe here applies, with 50/50 probability, to either the identity or the local Gaussian unitary U_T that maps $\hat{a}_j \mapsto -\hat{a}_j$ for all j, such that

$$\mathcal{T}(\rho_1) = \tfrac{1}{2}(\rho_1 + U_T \rho_1 U_T^\dagger). \tag{76}$$

For such a twirled state, the odd moments have zero expectation value, whereas the even moments are unchanged. Furthermore, twirling the physical state also results in twirling on the level of the filtered object, since

$$\frac{P\mathcal{T}(\rho_1)P}{\mathrm{tr}[P\mathcal{T}(\rho_1)P]} = \frac{\mathcal{T}(P\rho_1 P)}{\mathrm{tr}[\mathcal{T}(P\rho_1 P)]} = \frac{\mathcal{T}(P\rho_1 P)}{\mathrm{tr}[P\rho_1 P]} = \mathcal{T}(\tau_1). \tag{77}$$

The above follows immediately from the observation that U_T commutes with P as it does not change second moments. Another consequence of twirling preserving second moments is that the central limit sequence evolves to the same τ_∞ independently of whether we twirled or not. However, having twirled and eliminated all odd moments makes it possible for good reference states to exist and for Theorem 5 to hold.

Finally, we give another remark on condition (i) of the definition of reference states. For brevity we stated that this must hold for all products of operators $\{\hat{b}_j^\dagger, \hat{b}_j\}$. However, we only need to verify that the condition is valid for all *normally* ordered operators. Recall that normally ordered operators have all $\hat{b}_j^\dagger$ operators to the left side of any $\hat{b}_j$ operators, so $\hat{b}_j^\dagger \hat{b}$ is normally ordered but $\hat{b}_j \hat{b}_j^\dagger$ is not. By using $\hat{b}_j \hat{b}_j^\dagger = \hat{b}_j^\dagger \hat{b}_j + \mathbb{1}$ it is easy to rewrite the relevant operators—those composed of products from the set $\{\hat{b}_j^\dagger, \hat{b}_j\}$—as a positive sum of normally ordered operators. Provided $|\mathrm{tr}(H^{(k)}\tau_1)| \leqslant \mathrm{tr}(H^{(k)}\tau_{\mathrm{ref}})$ for normally ordered operators, it follows that the same holds for positive sums of normally ordered operators. Again, this observation is useful for reducing the workload of verifying that a purported reference state indeed meets all the requirements.

VII. HYBRID (CONTINUOUS VARIABLE) QUANTUM REPEATERS

Quantum repeaters are one of the main applications of the various variants of entanglement distillation discussed and proposed here. The aim of quantum repeaters is to distribute entanglement, despite the presence of noise, over large distances. There are many variants of such schemes, though they all share the common feature of using entanglement swapping rounds that entangle pairs that have not interacted in the past and distillation to reduce noise. It is long established [60,61] that discrete variable repeater networks can achieve distances far beyond those feasible by direct transmission of quantum states. Despite considerable work on CV entanglement distillation, surprisingly, it has not been shown that CV repeater networks can outperform direct transmission. Here, we give evidence that CV repeater networks can outperform direct transmission, albeit under some idealized conditions. In particular, we do not compute rates of entanglement production as these calculations are very computationally intensive for CV systems and so beyond our scope.

A. Primitives

The primitives discussed and introduced here are useful in constructing CV quantum repeater schemes. It is beyond the scope of the present work to present a comprehensive study of the possible repeater schemes that can be devised based on these basic elements. Given the importance of this application, we however sketch what parameters may be varied in variants of such schemes.

(1) *Gaussification.* There are several conceivable ways of performing Gaussification, including a recursive Gaussifier, a temporally reordered recursive Gaussifer, a pumping Gaussifier, and others. Since convergence of these protocols is fast, and in order not to arrive at low rates, it seems advisable to perform very few steps in each instance. The resource requirements, in particular involving memory requirements, are different in these schemes. The framework developed here and in Ref. [44] allows for a trade-off between success probability and quality of the output, when projecting onto Gaussian states different from the vacuum.

(2) *Swapping.* The precise procedure of entanglement swapping may be varied, with the original nested scheme being only one possibility. For Gaussian states, the optimum Gaussian entanglement swapping scheme is known [30,62] and is used subsequently. But other swapping steps are conceivable as well, such as mixing inputs at a symmetric beam splitter and projecting the outputs onto certain photon number states.

(3) *Non-Gaussian operations.* Given a source of Gaussian entangled states some non-Gaussian operation will be required prior to Gaussification, which is said to de-Gaussify the initial state. There are many possible ways to perform non-Gaussian operations in the scheme, such as in particular, only at the beginning, or also in later steps of the protocol. Also, several kinds of non-Gaussian steps have been considered in the literature so far. This includes (i) a mixing of the signal at a beam splitter with a single photon state, followed by a measurement at one of the output ports [7,8]. We will

refer to this step as single-photon replacement since a single photon is both added and removed. (ii) One can think of photon subtraction schemes, again leading to non-Gaussian states [8,11,14–17]. (iii) Ref. [11] introduces a modified non-Gaussian operation that is experimentally more challenging, but suggests better purification.

(4) *Non-Gaussian inputs.* In order to arrive at reasonable success probabilities, it may also be advantageous to make use of non-Gaussian input states that have higher photon numbers suppressed by their very preparation mechanism. For example, using entangled pairs generated from quantum dots in bi-photon cascades (see, e.g., Ref. [22]).

These parameters can be altered in benchmarking the functioning of such protocols, along the lines as has recently been done for discrete-variable quantum repeater schemes [63]. Needless to say, in any such effort, not only the losses in transmission have to be taken into account, but also the impact of imperfect swappings and Gaussification as well as issues of mode matching. Symmetric entanglement distillation schemes may also be favorable compared to asymmetric schemes [64].

B. Our repeater network

Here, we introduce a concrete class of quantum repeaters that are analyzed in the next section. In these protocols, any covariance matrix of any two-mode Gaussian state ρ encountered at any step is of the form

$$\Gamma_\rho = \begin{pmatrix} C & 0 & S & 0 \\ 0 & C & 0 & -S \\ S & 0 & C & 0 \\ 0 & -S & 0 & C \end{pmatrix}, \tag{78}$$

where $C,S \geqslant 0$ with $C^2 \geqslant 1 + S^2$. For a pure two-mode squeezed state $C^2 = 1 + S^2$, this equality becomes an inequality in the case of mixed Gaussian states. The EPR uncertainty [65], which for a Gaussian state with a covariance matrix as in Eq. (78), takes the simple form

$$\Delta(\rho) = C - S. \tag{79}$$

Rates in CV key distribution schemes will, in particular, relate to the above quantity. Indeed, a Gaussian state ρ with a covariance matrix of the above form is entangled if and only if $\Delta(\rho) < 1$ (the implication still being valid in one direction for non-Gaussian states).

The numerics presented here are based on the following CV repeater protocol (also illustrated in Fig. 3):

(1) Each of the $m = 2^k$ sources repeatedly produce many copies of a pure two-mode squeezed state (squeezing parameter r).

(2) Each half of every entangled pair is transmitted a distance l to a repeater node, and so becomes noisy due to attenuation.

(3) Photon replacement is used to probabilistically de-Gaussify.

(4) The de-Gaussified states are now iteratively Gaussified.

(5) The Gaussified states are swapped k times until an entangled state is shared across the full distance $L = 2ml = 2^{k+1}l$.

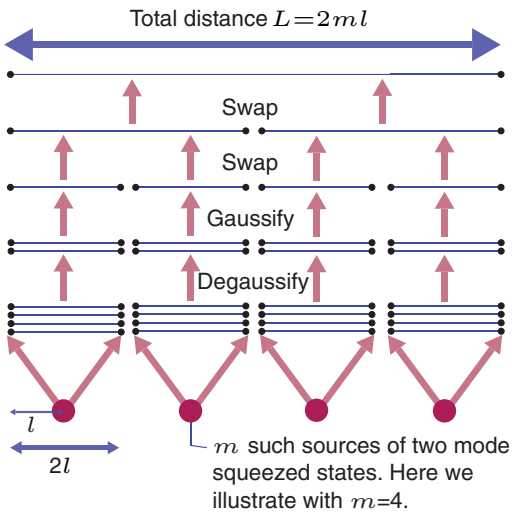

FIG. 3. (Color online) A schematic of the CV repeater network considered here. Sources are assumed to produce a pure two-mode squeezed state of some chosen squeezing. Channels are predominately affected by attenuation, but also a small amount of room-temperature thermal noise. De-Gaussification is performed by photon replacement. Gaussification is performed as described here and in previous work using measurements projecting onto the vacuum, using many copies of the de-Gaussified state and asymptotically approaching a Gaussian state. Entanglement swapping uses deterministic optimal continuous swapping protocol.

We require that the first step produces pure two-mode squeezed states of the form of Eq. (78). We set $C = \cosh(2r)$ and $S = \sinh(2r)$ and call $r > 0$ the squeezing parameter, for which we consider a range of possible values.

After the second step, the entangled pairs suffer noise from transmission over a lossy channel, becoming mixed states prior to distillation. For photons traveling in optical fiber the dominant noise source is attenuation through absorption, scattering, and mode mismatching. Indeed, attenuation is so dominant that previous analysis of CV distillation protocols has focused on pure attenuation noise channels. A solely attenuating channel will never completly eliminate the entanglement of a transmitted two-mode squeezed state. We consider Gaussian channels with a small contribution of additional noise, on top of attenuation, such that covariance matrices evolve as

$$\gamma \mapsto e^{-l/l_{\text{att}}}\gamma + (1 + 2n_{\text{th}})(1 - e^{-l/l_{\text{att}}})\mathbb{1}, \tag{80}$$

where l is is the distance (herein all distances in kilometers) traveled by each mode and l_{att} is the attenuation length of the fiber optic. In the infinite distance limit the state becomes thermal with an average photon number n_{th}. Applying such a noise model to the pure Gaussian state of Eq. (78) gives a mixed state of a similar form where

$$\begin{aligned} C &= e^{-l/l_{\text{att}}}\cosh(2r) + (1 + 2n_{\text{th}})(1 - e^{-l/l_{\text{att}}}), \\ S &= e^{-l/l_{\text{att}}}\sinh(2r). \end{aligned} \tag{81}$$

Herein we take $l_{\text{att}} = 22$ km as this is the state of the art for current fiber optic cable. For a pure attenuation channel $n_{\text{th}} = 0$, but we take $n_{\text{th}} = 10^{-8}$ as this corresponds to the thermal photon occupation at room temperature. The interesting feature of our analysis is that this modest additional noise source

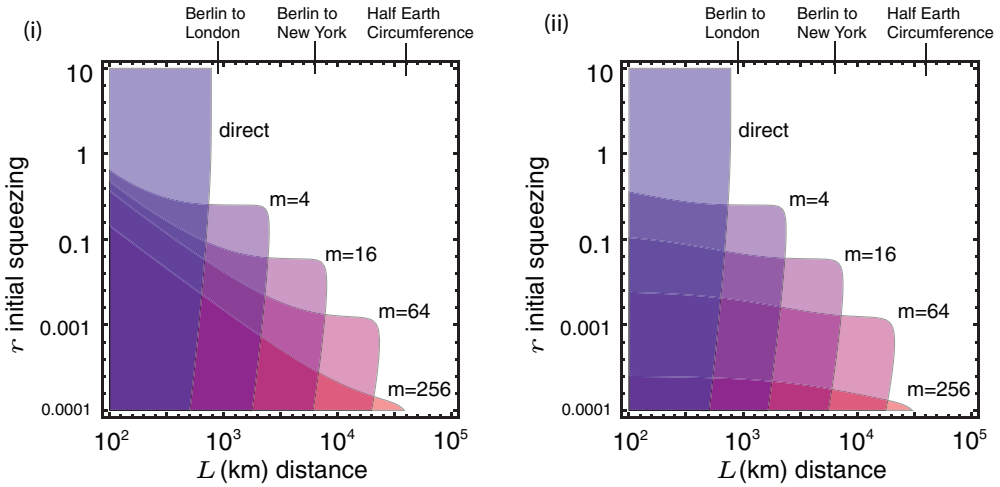

FIG. 4. (Color online) The maximum attainable distance L for a range of initial squeezing r for which entanglement can be distributed. In plot (i) noise arises wholly from transmission between repeater stations, whereas in plot (ii) we incur an additional 50% attenuation within each repeater station (see text for more details). Regions are shown for direct transmission and CV quantum repeater networks dividing the distance up into m intervals. Some noteworthy terrestrial scales are shown at the top.

is sufficient to put a hard cap on the distance at which various protocols can propagate entanglement. Assuming an initially pure two-mode squeezed state with squeezing parameter r, the maximum distance possible by direct transmission before the state is separable is easily found to be

$$l_{\max}(r) = 2l_{\text{att}} \ln\left(\frac{1 + 2n_{\text{th}} - \cosh(2r) + \sinh(2r)}{2n_{\text{th}}}\right).$$

This increases with r approaching the limiting value

$$\lim_{r\to\infty} l_{\max}(r) = 2l_{\text{att}} \ln\left(1 + \frac{1}{2n_{\text{th}}}\right), \tag{82}$$

which for our chosen parameters evaluates to 780 km. Recent continuous-variable experiments have achieved quantum cryptography, directly and without the aid of repeaters, at a distance of 80 km [66]. Our upper bound is of roughly the same order of magnitude, but larger as we take an optimistic noise model. We present results on two variants on the noise model. Analysis (i), as presented in Fig. 4(i), assumes that transmission noise dominates all other noise sources. Analysis (ii), as presented in Fig. 4(ii), is more pessimistic and assumes that an additional 50% photon loss occurs within the repeater station. This additional loss equates to over 15 km of optical fiber, but can also be attributed to other effects such as mode mismatching and detector inefficiencies.

On the third step of our repeater protocol we de-Gaussify by using symmetric photon replacement. The process begins with mixing a mode of the entangled pair on beam splitter of transmittivity $\eta^2 \in [0,1]$, where the second input mode contains a single photon. Next, the reflected signal mode is measured with a single photon resolving detector and we postselect on seeing a single photon. Such de-Gaussification procedures have been extensively studied [7,8,64] so we shall not repeat a full analysis here. However, it is informative to introduce the variable

$$\epsilon(\rho) = \frac{\langle 1,0|\rho|1,0\rangle}{\langle 1,1|\rho|0,0\rangle}, \tag{83}$$

which is meaningful because it is unchanged by photon replacement, or indeed any operation with Kraus operators diagonal in the Fock basis. In particular, for a symmetric Gaussian state of the form (78) we find

$$\epsilon(\rho) = \frac{C^2 - S^2 - 1}{2S}. \tag{84}$$

This variable is of interest as it cannot be increased either by Gaussification [67] or photon replacement. Indeed, ϵ remains unchanged by any local de-Gaussifying procedure resulting in Kraus operators diagonal in the Fock basis.

In a variation of the argument presented in Ref. [64] to accommodate for thermal noise, one obtains that the net effect of de-Gaussification and subsequent Gaussification, using $P = |0,0\rangle\langle 0,0|$, is that the state evolves to a Gaussian with

$$C = \frac{\Lambda^2(1-\epsilon^2)+1}{(1-\epsilon\Lambda)^2 - \Lambda^2}, \quad S = \frac{2\Lambda}{(1-\epsilon\Lambda)^2 - \Lambda^2}, \tag{85}$$

where ϵ depends on ρ after transmission through the noise channel and Λ can be tuned to any value in the interval $0 < \Lambda < (1+\epsilon)^{-1}$ by suitable choice of the beam-splitter transmittivity used in de-Gaussification. Larger values of Λ provide more entanglement in the final state, and we have numerically found that larger values also produce repeater networks capable of reaching larger distances. However, larger values of Λ also significantly reduce the success probability of de-Gaussification. Herein we assume that $\Lambda = 0.99/(1+\epsilon)$, as any further increase results in only a negligible increase in maximum repeater distance.

Having distributed entanglement and distilled at repeater stations, in the last step we perform swapping operations to generate entanglement between the most distant repeater nodes. In order to describe the optimum Gaussian

entanglement swapping [30,62] consider the function $g : \mathbb{R}^2 \to \mathbb{R}^2$, defined as

$$g(x,y) = \left(x - \frac{y^2}{2x}, \frac{y^2}{2x}\right). \tag{86}$$

Indeed, the covariance matrix before of the form (78) with $C, S \geqslant 0$ is mapped onto one of the same form with

$$(C', S') = g(C, S). \tag{87}$$

If $2l$ is the distance between the repeater stations, such a scheme would distribute an entangled state over a physical distance of $l2^{(k+1)}$ for k swaps. As such repeater networks are typically divided into $m = 2^k$ intervals for some integer k. This results in a mapping g^k.

C. Maximum distance of repeater networks

We now discuss the maximum distance that can be reached in the repeater scheme outlined in the previous section. We say a scheme achieves a distance L whenever it produces an entangled state, as verified by the Duan criteria $\Delta(\rho) < 1$ between the distant repeater nodes. These results are summarized by Fig. 4, where we show the achievable distances for different numbers of repeater stations and a range of initial squeezing parameters. For direct transmission—where no actual repeater techniques are exploited—we find performance is best in the large squeezing regime. However, we see that by using more repeater stations, and hence more intensive distillation, greater distances may be achieved. This provides the first evidence that CV techniques may achieve distances of a global scale, whereas direct transmission is incapable of achieving relatively short distances, such as Berlin to London. Comparing analyses (i) and (ii), the additional noise of the latter model does slightly reduce the maximum distance, but the decrease is very small. We also see that using more repeater stations typically requires a smaller initial squeezing, and this effect is more pronounced in analysis (ii) for short distances. This is consistent with observations made in Ref. [64] where they observed that distillation was more effective when combined with smaller initial squeezing. A possible explanation for this feature is that the more squeezed the initial state the more mixed the final state after suffering photon loss. Furthermore, the Gaussification process, while increasing entanglement, does not actually increase the purity, so limiting the impact of photon loss on purity is the key parameter to be optimized.

These results can be contrasted with those of Ref. [62]. Its authors compared the performance of direct transmission to the use of entanglement swapping, though without the benefits of any entanglement distillation, and found direct transmission to be preferable. Our noise model and figure of merit differ from those of Ref. [62], but our own numerics also found that entanglement swapping without distillation always achieves significantly inferior distances. Such behavior is a unique feature of CV protocols as the discrete variable protocol of Ref. [61] shows that swapping, albeit with some postselection, can be beneficial. We also considered some other variants of our repeater network. For instance, we considered several *nested* repeater schemes, where each entanglement swap is interleaved with distillation. Again, we found that these alternative protocols achieved shorter maximum distances compared to the protocol explicitly described in the previous section.

It seems that the results in Fig. 3, at least using the specific forms of Gaussification and de-Gaussification considered here, show the upper bounds of what is feasible with current technology. However, this does leave open the possibility of using alternative de-Gaussification procedures, such as that proposed in Ref. [11], or suitable deterministically prepared non-Gaussian states to start with. As commented earlier, the parameter ϵ is nonincreasing through our distillation techniques, though those techniques can vary this parameter potentially leading to an increase of the maximum attainable distance. However, to date such proposals are even more technologically challenging than replacement of a single photon. On the other hand, while CV systems pay a high price for de-Gaussification they can produce two-mode squeezed states at intrinsically higher rates than single-photon sources. They also benefit from the higher efficiency of homodyne detectors. In future work, a careful analysis of rates will be made, including also a comparison with common discrete variable schemes that weighs these relative merits.

VIII. SUMMARY AND CONCLUSION

In this work, we have further introduced and elaborated upon a formalism general enough to capture all of the known schemes of entanglement distillation leading to Gaussian quantum states, as well as to construct a plethora of new ones. The flexibility of the approach allows trading success probabilities against the quality of the resulting entangled states, or to realistically take memory requirements into account. As such, the formalism presented here provides a natural starting point for comprehensive comparisons of different entanglement distillation schemes in the continuous-variable setting. At the root of the formalism is a novel kind of noncommutative central limit theorem that is laid out in great detail. We also discuss the implications of the findings for devising novel schemes for quantum repeaters and highlight both potential and limitations. It is the hope that the general framework developed here gives a basis for assessing to what extent experimental large distance continuous-variable quantum communication is truly feasible.

ACKNOWLEDGMENTS

This work has been supported by the BMBF (QuOReP), the EU (Q-ESSENCE), the ERC (TAQ), the EURYI, and the EPSRC. We thank A. Serafini and M. Paris for interesting discussions that contributed ideas toward the design of the pumping Gaussifier.

APPENDIX A: CONJUGATION LEMMA

Here, we show that for any Gaussian operator P with zero first moments, we have $U(P \otimes P)U^\dagger = (P \otimes P)$, where U is a multilateral beam-splitter transformation. The Gaussian operator can be expressed as $P \otimes P = k \exp[-(H_A + H_B)]$ and

$$H_{X=A,B} = \sum_{i,j} h_{i,j} \hat{Q}_{Xi} \hat{Q}_{Xj}, \tag{A1}$$

for some $h_{i,j}$ and so

$$H_A + H_B = \sum_{i,j} h_{i,j}(\hat{Q}_{Ai}\hat{Q}_{Aj} + \hat{Q}_{Bi}\hat{Q}_{Bj}). \tag{A2}$$

The beam splitters cause

$$U(P \otimes P)U^\dagger = k\ \exp[-U(H_A + H_B)U^\dagger], \tag{A3}$$

and so we simply need to show that

$$U(\hat{Q}_{Ai}\hat{Q}_{Aj} + \hat{Q}_{Bi}\hat{Q}_{Bj})U^\dagger = \hat{Q}_{Ai}\hat{Q}_{Aj} + \hat{Q}_{Bi}\hat{Q}_{Bj}. \tag{A4}$$

Using the shorthand $J_{i,j}$ for the right-hand side and conjugating the quadrature operators with the unitary we have

$$\begin{aligned} U J_{i,j} U^\dagger = {} & (\sqrt{T}\hat{Q}_{Bi} + \sqrt{R}\hat{Q}_{Ai})(\sqrt{T}\hat{Q}_{Bj} + \sqrt{R}\hat{Q}_{Aj}) \\ & + (\sqrt{T}\hat{Q}_{Bi} - \sqrt{R}\hat{Q}_{Ai})(\sqrt{T}\hat{Q}_{Bj} - \sqrt{R}\hat{Q}_{Aj}). \end{aligned}$$

Expanding out, we find the cross terms ($\hat{Q}_{Ai}\hat{Q}_{Bj}$ and $\hat{Q}_{Bi}\hat{Q}_{Aj}$) cancel leaving only

$$U J_{i,j} U^\dagger = (R + T)(\hat{Q}_{Ai}\hat{Q}_{Aj} + \hat{Q}_{Bi}\hat{Q}_{Bj}). \tag{A5}$$

Recalling $R + T = 1$, we have $U J_{i,j} U^\dagger = J_{i,j}$, which in turn entails the result $U(P \otimes P)U^\dagger = (P \otimes P)$.

APPENDIX B: AN INEQUALITY

For any complex a and b satisfying $|a| \leqslant 1$ and $|b| \leqslant 1$ and any integer N, we have $|a^N - b^N| \leqslant N|a - b|$. For $N = 1$ it is trivial and for higher N it is proven iteratively,

$$\begin{aligned} |a^N - b^N| &= |(a - b)a^{N-1} + b(a^{N-1} - b^{N-1})| \\ &\leqslant |a - b| + |a^{N-1} - b^{N-1}|, \end{aligned} \tag{B1}$$

where we have used the triangle inequality and $|a^{N-1}| \leqslant 1$ and $|b| \leqslant 1$. Each unit increase in N contributes at most an additional $|a - b|$, and so we have the desired result.

APPENDIX C: UNIFORM CONVERGENCE

In the main text we prove Theorem 1 for an individual point of phase space. Here, we extend it to a uniform result over balls of finite radius. For any finite set of points $\mathcal{R}_{\text{finite}} = \{\mathbf{r}_1, \mathbf{r}_2, \dots\}$ convergence is uniform over that set as it is bounded by the point that converges slowest. For any small distance δ we can find a $\mathcal{R}_{\text{finite}}$ such that any point inside the ball is less than distance δ from some point in the finite set. All $\chi \in \{\chi_{\tau_N}\}$ are continuous and within the ball there is a maximum possible gradient. Hence, for every point in the ball we can approximate the characteristic function by a nearby point in the finite set $\text{Re}_{\text{finite}}$ and uniform convergence follows.

APPENDIX D: MOMENTS CONVERGENCE

Here, we present a proof of Theorem 2 that follows the combinatorial argument of Refs. [56,57]. Our proof is not as general, but benefits from requiring less mathematical background. We consider a single kth moment $H^{(k)} = \prod_j H_j$,

$$\begin{aligned} \text{tr}(H^{(k)}\tau_N) &= \text{tr}\left[\left(\prod_j H_j\right) U \tau_1^{\otimes N} U^\dagger\right] \\ &= \frac{1}{N^{k/2}}\text{tr}\left[\prod_j \left(\sum_{x=1,\dots,N} H_{j,x}\right) \tau_1^{\otimes N}\right], \end{aligned} \tag{D1}$$

where $H_{j,x}$ indicates the H_j operator but acting on the xth of the N systems. We need to expand out the brackets and some way of labeling terms. We have k different operators that can act on N different copies. Each possibility can be represented by a partition of k values into N bins. For example, for $k = 4$ and $N = 5$ a possible partition is $B = \{\{1,2\},\{\},\{3\},\{4\},\{\}\}$ with which we associate with a term $H_1 H_2 \otimes \mathbb{1} \otimes H_3 \otimes H_4 \otimes \mathbb{1}$. In general, for a partition $B = \{B_1, B_2, B_3, \dots, B_N\}$ we associate an operator

$$H_B = \otimes_{x=1}^N H_{B_x}, \tag{D2}$$

where

$$H_{B_x} = \prod_{j \in B_x} H_j, \tag{D3}$$

with the product over $j \in B_x$ always taken in order of smallest to largest value of j. In this notation

$$\begin{aligned} \text{tr}(H_B\tau_N) &= \frac{1}{N^{k/2}}\text{tr}\left[\left(\otimes_{x=1}^N H_{B_x}\right)\tau_1^{\otimes N}\right] \\ &= \frac{1}{N^{k/2}}\prod_{x=1}^N \text{tr}\left(H_{B_x}\tau_1\right). \end{aligned} \tag{D4}$$

The next key step of the proof is a smart way of collecting up terms with similar properties. We define $L(B)$ to be the number of nonempty bins in B and then collect terms with the same value.

$$\begin{aligned} \text{tr}(H^{(k)}\tau_N) &= \frac{1}{N^{k/2}}\sum_b \sum_{L(B)=b} \text{tr}\left[H_B\tau_1^{\otimes N}\right] \\ &= \frac{1}{N^{k/2}}\sum_b \sum_{L(B)=b} \prod_{x=1}^N \text{tr}\left(H_{B_x}\tau_1\right). \end{aligned} \tag{D5}$$

For any B there are $N!/[N - L(B)]! = N(N-1)\cdots[N - L(B) + 1]$ partitions that differ by only a permutation of whole bins. For instance, $B = \{\{1,2\},\{\},\{3\},\{4\},\{\}\}$ and $B' = \{\{\},\{3\},\{1,2\},\{4\},\{\}\}$ differ only by a permutation of whole bins, and so give the same expectation value. We can also choose a canonical set $\mathcal{B}$ such that for every B there exists a unique $B' \in \mathcal{B}$ such that B and B' differ only by a permutation of whole bins. By summing over just the canonical set we have

$$\text{tr}(H^{(k)}\tau_N) = \sum_b \frac{N!}{N^{k/2}(N-b)!} \sum_{\substack{L(B)=b;\\ B\in\mathcal{B}}} \prod_{x=1}^N \text{tr}(H_{B_x}\tau_1). \tag{D6}$$

We proceed by showing that terms with $b < k/2$ and $b > k/2$ are either zero or decreasing with N, and so only the $b = k/2$ terms persist in the large N limit.

When $L(B) > k/2$ there must exist at least one bin B_x that contains only one element, so $H_{B_x} = H_j$ for some x and j. This factor contributes $\mathrm{tr}(H_j\tau_1)$ to the product, but by assumption $\mathrm{tr}(H_j\tau_1) = 0$ and so all such terms vanish. As for the case with $L(B) < k/2$, we observe that as N increases,

$$\frac{N!}{N^{k/2}\,[N - L(B)]!} \to 0. \tag{D7}$$

Furthermore, for all $N > k$ the factor

$$\sum_{\substack{L(B)=b;\ x=1\\ B\in\mathcal{B}}}^{N} \prod \mathrm{tr}\big(H_{B_x}\tau_1\big) \tag{D8}$$

is constant with N as the number of canonical partitions stops increasing. Therefore, for any $b < k/2$ the product of these terms vanishes with N.

This leaves only $b = k/2$ terms as potentially nonvanishing. Note that, if k is an odd number there are no suitable integer b values and so all odd moments will vanish with increasing N. Assuming k is even, the only nonvanishing partitions consist of pairings, such that each bin contains either two elements or none. That is, nonvanishing B have $H_{B_x} = H_j H_k$ or $H_{B_x} = \mathbb{1}$ for all x. Putting these results together we have

$$\lim_{N\to\infty} \mathrm{tr}(\hat{Q}^k\tau_N) = \left(\lim_{N\to\infty} \frac{N!}{N^{k/2}(N - k/2)!}\right) \times \sum_{B\in\mathcal{B}_{\mathrm{pair}}} \mathrm{tr}\big(H_{B_x}\tau_1\big), \tag{D9}$$

where $\mathcal{B}_{\mathrm{pair}}$ is the set of canonical pairings. The expectation value only depends on the second moments of τ_1 and so we can replace τ_1 with the Gaussian state with the same second moments, namely, τ_∞. The combinatorial factor approaches 1 and so

$$\lim_{N\to\infty} \mathrm{tr}(H^{(k)}\tau_N) = \sum_{B\in\mathcal{B}_{\mathrm{pair}}} \mathrm{tr}\big(H_{B_x}\tau_\infty\big) = \mathrm{tr}(H^{(k)}\tau_\infty). \tag{D10}$$

In the simple case where the moment is a product of identical factors, so $H^{(k)} = H^k$, we have

$$\lim_{N\to\infty} \mathrm{tr}(H^k\tau_N) = |\mathcal{B}_{\mathrm{pair}}|\mathrm{tr}(H^2\tau_\infty)^{k/2}. \tag{D11}$$

The number of canonical (unordered) pairings of k numbers is simply $|\mathcal{B}_{\mathrm{pair}}| = (k-1)(k-3)\cdots 1$, which is known as a double factorial $(k-1)!!$. Consider the above results for when the input state is Gaussian, and so unchanging. This tells us that the higher moments of a Gaussian state are determined by its second moments, as captured by Eq. (D11), which is a well-known result called Wick's theorem.

APPENDIX E: MATRIX ELEMENT CONVERGENCE

This Appendix provides a proof of Theorem 3. We move from statements about characteristic functions to operators by recalling that for an operator $B = |\psi_j\rangle\langle\psi_k|$ acting on an m-mode Hilbert space we have

$$\mathrm{tr}(B\tau) = (2\pi)^{-m} \int \chi_B(\mathbf{r})\chi_\tau(\mathbf{r})d\mathbf{r}. \tag{E1}$$

Similar reasoning allows us to deduce that since $\mathrm{tr}(BB^\dagger) = 1$ and $\mathrm{tr}(\tau\tau^\dagger) \leqslant 1$, we know

$$(2\pi)^{-m} \int |\chi_B(\mathbf{r})|^2 = 1, \quad (2\pi)^{-m} \int |\chi_\varphi(\mathbf{r})|^2 \leqslant 1. \tag{E2}$$

The absolute difference in expectation values between τ_N and τ_∞ is

$$\begin{aligned} D_N &= |\mathrm{tr}(B\tau_N) - \mathrm{tr}(B\tau_\infty)|, \\ &= \frac{1}{(2\pi)^m} \int \chi_B(\mathbf{r}) \left[\chi_{\tau_N}(\mathbf{r}) - \chi_{\tau_\infty}(\mathbf{r})\right] d\mathbf{r}, \\ &= \frac{1}{(2\pi)^m} \int \chi_B(\mathbf{r})\Delta_N(\mathbf{r})d\mathbf{r}, \end{aligned} \tag{E3}$$

where $\Delta_N = \chi_{\tau_N}(\mathbf{r}) - \chi_{\tau_\infty}(\mathbf{r})$. The proof proceeds by splitting the integral up into two parts so $D_N = D'_N + D''_N$. We take D'_N to be an integral over a large but finite ball of radius R and D''_N over the complement. Over the complement we have that

$$\begin{aligned} D''_N &= \frac{1}{(2\pi)^m} \int_{|\mathbf{r}|>R} \chi_B(\mathbf{r})\Delta_N(\mathbf{r})d\mathbf{r}, \\ |D''_N| &\leqslant \frac{1}{(2\pi)^m} \left(\int_{|\mathbf{r}|>R} |\chi_B(\mathbf{r})|^2 d\mathbf{r} \int_{|\mathbf{r}|>R} |\Delta_N(\mathbf{r})|^2 d\mathbf{r}\right)^{1/2}, \end{aligned} \tag{E4}$$

where we have used the Cauchy-Schwarz inequality. From Eq. (E1) we can know $\int |\chi_B(\mathbf{r})|^2 = 1$ and so the integral over $|\mathbf{r}| > R$ can be made arbitrarily small by increasing R. Formally, for any $\epsilon' > 0$ we can find an R such that $\int_{|\mathbf{r}|>R} |\chi_B(\mathbf{r})|^2 \leqslant \epsilon'$. Furthermore, Eq. (E1) entails that the integration over $|\Delta_N(\mathbf{r})|^2$ must be less than 2. Hence, we deduce

$$|D''_N| \leqslant \frac{(2\epsilon')^{1/2}}{(2\pi)^m}, \tag{E5}$$

which holds for all N. As for the integral inside radius R we have

$$|D'_N| \leqslant (2\pi)^{-m} \int_{|\mathbf{r}|\leqslant R} |\chi_B(\mathbf{r})\Delta_N(\mathbf{r})|d\mathbf{r}. \tag{E6}$$

For all characteristic functions $|\chi_B(\mathbf{r})| \leqslant \mathrm{tr}(\sqrt{B^\dagger B})$ and so for $B = |\psi_j\rangle\langle\psi_k|$ we have $|\chi_B(\mathbf{r})| \leqslant 1$. Furthermore we know that within a finite ball $\Delta_N(\mathbf{r})$ vanishes uniformly, so for any $\epsilon' > 0$ there is a N_ϵ such that for all $N > N_{\epsilon'}$ we have

$$|D'_N| \leqslant (2\pi)^{-m} \int_{|\mathbf{r}|<R} \epsilon'\, d\mathbf{r} = \frac{\epsilon'' V}{(2\pi)^m}, \tag{E7}$$

where V is the volume of the ball. Combining these results we have, for $N > N_{\epsilon'}$, that

$$|\langle\psi_k|\tau_N|\psi_j\rangle - \langle\psi_k|\tau_\infty|\psi_j\rangle| < \frac{(2\epsilon')^{1/2} + \epsilon'' V}{(2\pi)^m}. \tag{E8}$$

Since ϵ' and ϵ'' can be made arbitrarily small, we have proven Theorem 3.

[1] P. Kok, W. J. Munro, K. Nemoto, T. C. Ralph, J. P. Dowling, and G. J. Milburn, Rev. Mod. Phys. **79**, 135 (2007).

[2] J. Eisert, S. Scheel, and M. B. Plenio, Phys. Rev. Lett. **89**, 137903 (2002).

[3] M. Horodecki, P. Horodecki, and R. Horodecki, Phys. Rev. Lett. **78**, 574 (1997).

[4] J. Fiurášek, Phys. Rev. Lett. **89**, 137904 (2002).

[5] J. Eisert and M. B. Plenio, Int. J. Quantum Inf. **1**, 479 (2003).

[6] G. Giedke and J. I. Cirac, Phys. Rev. A **66**, 032316 (2002).

[7] D. E. Browne, J. Eisert, S. Scheel, and M. B. Plenio, Phys. Rev. A **67**, 062320 (2003).

[8] J. Eisert, D. Browne, S. Scheel, and M. B. Plenio, Ann. Phys. **311**, 431 (2004).

[9] J. Eisert, M. B. Plenio, D. E. Browne, S. Scheel, and A. Feito, Opt. Spectrosc. **103**, 173 (2007).

[10] J. Fiurášek, P. Marek, R. Filip, and R. Schnabel, Phys. Rev. A **75**, 050302 (2007).

[11] J. Fiurášek, Phys. Rev. A **82**, 042331 (2010).

[12] B. Kraus, K. Hammerer, G. Giedke, and J. I. Cirac, Phys. Rev. A **67**, 042314 (2003).

[13] J. Heersink, C. Marquardt, R. Dong, R. Filip, S. Lorenz, G. Leuchs, and U. L. Andersen, Phys. Rev. Lett. **96**, 253601 (2006).

[14] A. Ourjoumtsev, A. Dantan, R. Tualle-Brouri, and P. Grangier, Phys. Rev. Lett. **98**, 030502 (2007).

[15] F. Dell'Anno, S. De Siena, G. Adesso, and F. Illuminati, Phys. Rev. A **82**, 062329 (2010).

[16] S. Olivares and M. G. A. Paris, Phys. Rev. A **70**, 032112 (2004).

[17] M. S. Kim, J. Phys. B **41**, 133001 (2008).

[18] T. Opatrný, G. Kurizki, and D.-G. Welsch, Phys. Rev. A **61**, 032302 (2000).

[19] P. T. Cochrane, T. C. Ralph, and G. J. Milburn, Phys. Rev. A **65**, 062306 (2002).

[20] S. Olivares, M. G. A. Paris, and R. Bonifacio, Phys. Rev. A **67**, 032314 (2003).

[21] F. Dell'Anno, S. De Siena, and F. Illuminati, Phys. Rev. A **81**, 012333 (2010).

[22] R. M. Stevenson, A. J. Hudson, R. J. Young, P. Atkinson, K. Cooper, D. A. Ritchie, and A. J. Shields, Opt. Express **15**, 6507 (2007).

[23] F. Dell'Anno, S. D. Siena, and F. Illuminati, Phys. Rep. **428**, 53 (2006).

[24] M. G. Genoni and M. G. A. Paris, Phys. Rev. A **82**, 052341 (2010).

[25] J. S. Bell, *Speakable and Unspeakable in Quantum Mechanics* (Cambridge University Press, Cambridge, 1987).

[26] K. Banaszek and K. Wódkiewicz, Acta Phys. Slov. **49**, 491 (1999).

[27] W. Son, J. Kofler, M. S. Kim, V. Vedral, and C. Brukner, Phys. Rev. Lett. **102**, 110404 (2009).

[28] S. D. Bartlett, H. de Guise, and B. C. Sanders, Phys. Rev. A **65**, 052316 (2002).

[29] S. D. Bartlett and B. C. Sanders, Phys. Rev. Lett. **89**, 207903 (2002).

[30] M. Ohliger, K. Kieling, and J. Eisert, Phys. Rev. A **82**, 042336 (2010).

[31] A. Mari and J. Eisert, Phys. Rev. Lett. **109**, 230503 (2012).

[32] V. Veitch, N. Wiebe, C. Ferrie, and J. Emerson, New J. Phys. **15**, 013037 (2013).

[33] A. Mari, K. Kieling, B. M. Nielsen, E. S. Polzik, and J. Eisert, Phys. Rev. Lett. **106**, 010403 (2011).

[34] W. Dür and H. J. Briegel, Rep. Prog. Phys. **70**, 1381 (2007).

[35] L. I. Childress, J. M. Taylor, A. Sørensen, and M. D. Lukin, Phys. Rev. A **72**, 052330 (2005).

[36] M. V. G. Dutt, L. Childress, L. Jiang, E. Togan, J. Maze, F. Jelezko, A. S. Zibrov, P. R. Hemmer, and M. D. Lukin, Science **316**, 1312 (2007).

[37] L. Jiang, J. M. Taylor, A. S. Sørensen, and M. D. Lukin, Phys. Rev. A **76**, 062323 (2007).

[38] E. T. Campbell and S. C. Benjamin, Phys. Rev. Lett. **101**, 130502 (2008).

[39] E. T. Campbell, Phys. Rev. A **76**, 040302(R) (2007).

[40] E. T. Campbell, Int. J. Quantum Inf. **8**, 161 (2010).

[41] Y. Li and S. C. Benjamin, New J. Phys. **14**, 093008 (2012).

[42] K. Fujii, T. Yamamoto, M. Koashi, and N. Imoto, arXiv:1202.6588.

[43] A. Datta, L. Zhang, J. Nunn, N. K. Langford, A. Feito, M. B. Plenio, and I. A. Walmsley, Phys. Rev. Lett. **108**, 060502 (2012).

[44] E. T. Campbell and J. Eisert, Phys. Rev. Lett. **108**, 020501 (2012).

[45] M. M. Wolf, G. Giedke, and J. I. Cirac, Phys. Rev. Lett. **96**, 080502 (2006).

[46] S. M. Barnett and P. M. Radmore, *Methods in Theoretical Quantum Optics* (Oxford Science Publications, Oxford University Press, Oxford, 2005).

[47] A. Jamiolkowski, Rep. Math. Phys. **3**, 275 (1972).

[48] K. Zyczkowski and I. Bengtsson, Open Syst. Inf. Dyn. **11**, 3 (2004).

[49] R. A. Horn and C. R. Johnson, *Matrix Analysis* (Cambridge University Press, Cambridge, UK, 2006).

[50] R. F. Werner and M. M. Wolf, Phys. Rev. Lett. **86**, 3658 (2001).

[51] S. Olivares and M. G. A. Paris, Phys. Rev. Lett. **107**, 170505 (2011).

[52] C. D. Cushen and R. L. Hudson, J. Appl. Probab. **8**, 454 (1971).

[53] P. A. P. Moran, *An Introduction to Probability Theory* (Clarendon Press, Oxford, 1968), p. 542.

[54] M. Cramer and J. Eisert, New J. Phys. **12**, 055020 (2010).

[55] M. Cramer, C. M. Dawson, J. Eisert, and T. J. Osborne, Phys. Rev. Lett. **100**, 030602 (2008).

[56] N. Giri and W. von Waldenfels, Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete **42**, 129 (1978).

[57] D. Petz, *An Invitation to the Algebra of Canonical Commutation Relations*, Leuven Notes in Mathematical and Theoretical Physics: Mathematical Physics (Leuven University Press, Leuven, 1990).

[58] C. H. Bennett, G. Brassard, S. Popescu, B. Schumacher, J. A. Smolin, and W. K. Wootters, Phys. Rev. Lett. **76**, 722 (1996).

[59] S. Bravyi and A. Kitaev, Phys. Rev. A **71**, 022316 (2005).

[60] W. Dür, H. J. Briegel, J. I. Cirac, and P. Zoller, Phys. Rev. A **59**, 169 (1999).

[61] L. M. Duan, M. D. Lukin, J. I. Cirac, and P. Zoller, Nature (London) **414**, 413 (2001).

[62] J. Hoelscher-Obermaier and P. van Loock, Phys. Rev. A **83**, 012319 (2011).

[63] S. Abruzzo, S. Bratzik, N. K. Bernardes, H. Kampermann, P. van Loock, and D. Bruss, arXiv:1208.2201.

[64] A. P. Lund and T. C. Ralph, Phys. Rev. A **80**, 032309 (2009).

[65] L.-M. Duan, G. Giedke, J. I. Cirac, and P. Zoller, Phys. Rev. Lett. **84**, 2722 (2000).

[66] P. Jouguet, S. Kunz-Jacques, A. Leverrier, P. Grangier, and E. Diamanti, arXiv:1210.6216.

[67] Here we assume Gaussification using $P = |0,0\rangle\langle 0,0|$, although the increase of ϵ by Gaussification using different choices of P.