Compression for quantum population coding

Yuxiang Yang, Ge Bai, Giulio Chiribella, and Masahito Hayashi *Fellow, IEEE*

***Abstract*—We study the compression of n quantum systems, each prepared in the same state belonging to a given parametric family of quantum states. For a family of states with f independent parameters, we devise an asymptotically faithful protocol that requires a hybrid memory of size $(f/2)\log n$, including both quantum and classical bits. Our construction uses a quantum version of local asymptotic normality and, as an intermediate step, solves the problem of compressing displaced thermal states of n identically prepared modes. In both cases, we show that $(f/2)\log n$ is the minimum amount of memory needed to achieve asymptotic faithfulness. In addition, we analyze how much of the memory needs to be quantum. We find that the ratio between quantum and classical bits can be made arbitrarily small, but cannot reach zero: unless all the quantum states in the family commute, no protocol using only classical bits can be faithful, even if it uses an arbitrarily large number of classical bits.**

***Index Terms*—Population coding, compression, quantum system, local asymptotic normality, identically prepared state**

I. Introduction

Many problems in quantum information theory involve a source that prepares multiple copies of the same quantum state. This is the case, for example, of quantum tomography [1], quantum cloning [2], [3], and quantum state discrimination [4]. The state prepared by the source is generally unknown to the agent who has to carry out the task. Instead, the agent knows that the state belongs to some parametric family of density matrices $\{\rho_\theta\}_{\theta\in\Theta}$, with the parameter θ varying in the set Θ. It is generally assumed that the source prepares each particle identically and independently: when the source is used n times, it generates n quantum particles in the tensor product state $\rho_\theta^{\otimes n}$.

A fundamental question is how much information is contained in the n-particle state $\rho_\theta^{\otimes n}$. One way to address the question is to quantify the minimum amount of memory needed to store the state, or equivalently, the minimum amount of communication needed to transfer the state from a sender to a receiver. Solving this problem requires an optimization over all possible compression protocols.

It is important to stress that the problem of storing the n-copy states $\{\rho_\theta^{\otimes n}\}_{\theta\in\Theta}$ in a quantum memory is different from the standard problem of quantum data compression [5], [6], [7]. In our scenario, the mixed state ρ_θ is not regarded as the average state of an information source, but, instead, as a physical encoding of the parameter θ. The goal of compression is to preserve the encoding of the parameter θ, by storing the state $\rho_\theta^{\otimes n}$ into a memory and retrieving it with high fidelity for all possible values of θ. To stress the difference with standard quantum compression, we refer to our scenario as *compression for quantum population coding*. The expression "quantum population coding" refers to the encoding of the parameter θ into the many-particle state $\rho_\theta^{\otimes n}$, viewed as the state of a "population" of quantum systems. We choose this expression in analogy with the classical notion of population coding, where a parameter θ is encoded into the population of n individuals [8]. The typical example of population coding arises in computational neuroscience, where the population consists of neurons and the parameter θ represents an external stimulus.

The compression for quantum population coding has been studied by Plesch and Bužek [9] in the case where ρ_θ is a pure qubit state and no error is tolerated (see also [10] for a prototype experimental implementation). A first extension to mixed states, higher dimensions, and non-zero error was proposed by some of us in [11]. The protocol therein was proven to be optimal under the assumption that the decoding operation must satisfy a suitable conservation law. Later, it was shown that, when the conservation law is lifted, a new protocol can achieve a better compression, reaching the ultimate information-theoretic bound set by Holevo's bound [12]. This result applies to two-dimensional quantum systems with completely unknown Bloch vector and/or completely unknown purity. The classical version of the compression for population coding was addressed in [13]. However, finding the optimal protocol for arbitrary parametric families and for quantum systems of arbitrary dimension has remained as an open problem so far.

In this paper, we provide a general theory of compression for quantum states of the form $\rho_\theta^{\otimes n}$. We consider two categories of states: *(i)* generic quantum states in finite dimensions, and *(ii)* displaced thermal states in infinite dimension. These two categories of states are connected by the quantum version of local asymptotic normality (Q-LAN) [14], [15], [16], [17], [18], which locally reduces the tensor product state $\rho_\theta^{\otimes n}$ to a displaced thermal state, regarded as the quantum version of the normal distribution.

We will discuss first the compression of displaced thermal states. Then, we will employ Q-LAN to reduce the problem of compressing generic finite-dimensional states to the problem of compressing displaced thermal states. In both cases, our compression protocol uses a hybrid memory, consisting both of classical and quantum bits. For a family of quantum states

Y. Yang (e-mail: yangyx09@connect.hku.hk) and G. Bai (e-mail: baige@connect.hku.hk) are with the Department of Computer Science, The University of Hong Kong, Pokfulam Road, Hong Kong, and with the HKU Shenzhen Institute of Research and Innovation, Kejizhong 2nd Road, Shenzhen, China.

G. Chiribella (e-mail: giulio.chiribella@cs.ox.ac.uk) is with the Department of Computer Science, The University of Oxford, Parks Road, Oxford, UK, with the Canadian Institute for Advanced Research, CIFAR Program in Quantum Information Science, and with the HKU Shenzhen Institute of Research and Innovation, Kejizhong 2nd Road, Shenzhen, China.

M. Hayashi is with the Graduate School of Mathematics, Nagoya University, Furocho, Chikusaku, Nagoya, 464-860, Japan, and Centre for Quantum Technologies, National University of Singapore, 3 Science Drive 2, Singapore 117542.

described by f independent parameters, the total size of the memory is $f/2\log n$ at the leading order, matching the ultimate limit set by Holevo's bound [19].

An intriguing feature of our compression protocol is that the ratio between the number of quantum bits and the number of classical bits can be made arbitrarily close to zero, but not exactly equal to zero. Such a feature is not an accident: we show that, unless the states commute, every asymptotically faithful protocol must use a non-zero amount of quantum memory. This result extends an observation made in [20] from certain families of pure states to generic families of states.

The paper is structured as follows. In section II we state the main results of the paper. In Section III we study the compression of displaced thermal states. In Section IV we provide the protocol for the compression of identically prepared finite-dimensional states. In Section V we show that every protocol achieving asymptotically faithful compression must use a quantum memory. Optimality of the protocols is proven later in Section VI. Finally, the conclusions are drawn in Section VII.

II. Main result.

The main result of this work is the optimal compression of identically prepared quantum states. We consider two categories of states: generic finite dimensional (i.e. qudit) states and infinite-dimensional displaced thermal states.

Let us start from the first category. For a quantum system of dimension $d < \infty$, also known as qudit, we consider *generic states* described by density matrices with full rank and non-degenerate spectrum. We parametrize the states of a d-dimensional quantum system as

$$\rho_\theta = U_\xi\, \rho_0(\mu)\, U_\xi^\dagger\,, \quad \theta = (\xi,\mu) \quad \xi \in \mathbb{R}^{d(d-1)} \quad \mu \in \mathbb{R}^{d-1} \tag{1}$$

where $\rho_0(\mu)$ is the fixed state

$$\rho_0(\mu) = \sum_{j=1}^{d} \mu_j\, |j\rangle\langle j| \qquad \mu_d := 1 - \sum_{k=1}^{d-1} \mu_k, \tag{2}$$

with spectrum ordered as $\mu_1 > \cdots > \mu_{d-1} > \mu_d > 0$, while U_ξ is the unitary matrix defined by

$$U_\xi = \exp\left[i \left(\sum_{1\le j<k\le d} \frac{\xi^{\rm I}_{j,k} T^{\rm I}_{j,k} + \xi^{\rm R}_{j,k} T^{\rm R}_{k,j}}{\sqrt{\mu_j - \mu_k}} \right) \right] \tag{3}$$

Here ξ is a vector of real parameters $(\xi^{\rm R}_{j,k}, \xi^{\rm I}_{j,k})_{1\le j<k\le d}$, and $T^{\rm I}$ ($T^{\rm R}$) is the matrix defined by $T^{\rm I}_{j,k} := iE_{j,k} - iE_{k,j}$ ($T^{\rm R}_{k,j} := E_{j,k} + E_{k,j}$), where $E_{j,k}$ is the $d\times d$ matrix with 1 in the entry (j,k) and 0 in all the other entries.

We consider n-copy qudit state families, denoted as $\{\rho_\theta^{\otimes n}\}_{\theta\in\Theta}$ where Θ is the set of possible vectors $\theta = (\xi,\mu)$. We call the components of the vector ξ *quantum parameters* and the components of the vector μ *classical parameters*. The classical parameters determine the eigenvalues of the density matrix ρ_θ, while the quantum parameters determine the eigenbasis.

We say that a parameter is *independent* if it can vary continuously while the other parameters are kept fixed. For a given family of states, we denote by f_c (f_q) the maximum number of independent classical (quantum) parameters describing states in the family. For example, the family of all diagonal density matrices $\rho_0(\mu)$ in Eq. (2) has $d-1$ independent parameters. The family of all quantum states in dimension d has d^2-1 independent parameters, of which $d-1$ are classical and $d(d-1)$ are quantum. In general, we will assume that the family $\{\rho_\theta^{\otimes n}\}_{\theta\in\Theta}$ is such that every component of the vector θ is either independent or fixed to a specific value.

Let us introduce now the second category of states that are relevant in this paper: the displaced thermal states [21], [22]. Displaced thermal states are a type of infinite-dimensional states frequently encountered in quantum optics [23]. Mathematically, they have the form

$$\rho_{\alpha,\beta} = D_\alpha\, \rho_\beta^{\rm thm} D_\alpha^\dagger \tag{4}$$

where $D_\alpha = \exp(\alpha\hat a^\dagger - \bar\alpha \hat a)$ is the *displacement operator*, defined in terms of a complex parameter $\alpha\in\mathbb{C}$ (the *displacement*), and $\hat a$ is the annihilation operator, satisfying the relation $[\hat a, \hat a^\dagger] = 1$, while $\rho_\beta^{\rm thm}$ is a *thermal state*, defined as

$$\rho_\beta^{\rm thm} := (1-\beta)\sum_{j=0}^{\infty} \beta^j\, |j\rangle\langle j|\,, \tag{5}$$

where $\beta\in[0,1)$ is a real parameter, here called the *thermal parameter*, and the basis $\{|j\rangle\}_{j\in\mathbb{N}}$ consists of the eigenvectors of $\hat a^\dagger \hat a$. For $\beta = 0$, the the displaced thermal states are pure. Specifically, the state $\rho_{\alpha,\beta=0}$ is the projector on the coherent state [24] $|\alpha\rangle := D_\alpha |0\rangle$.

In the context of quantum optics, infinite dimensional systems are often called *modes*. We will consider the compression of n modes, each prepared in the same displaced thermal state. We denote the n-mode states as $\{\rho_{\alpha,\beta}^{\otimes n}\}_{(\alpha,\beta)\in\Theta}$ with $\Theta = \Theta_\alpha \times \Theta_\beta$ being the parameter space. There are three real parameters for the displaced thermal state family: the thermal parameter β, the amount of displacement $|\alpha|$, and the phase $\varphi = \arg\alpha$. Here, β is a classical parameter, specifying the eigenvalues, while $|\alpha|$ and φ are quantum parameters, determining the eigenstates. We will assume that each of the three parameters $\beta, |\alpha|$ and φ is either independent, or fixed to a determinate value.

A compression protocol consists of two components: the encoder, which compresses the input state into a memory, and the decoder, which recovers the state from the memory. The compression protocol for n identically prepared systems is represented by a couple of quantum channels (completely positive trace-preserving linear maps) $(\mathcal{E}_n, \mathcal{D}_n)$ characterizing the encoder and the decoder, respectively. We focus on *asymptotically faithful* compression protocols, whose error vanishes in the large n limit. As a measure of error, we choose the supremum of the trace distance between the original state and the recovered state $\mathcal{D}_n\circ\mathcal{E}_n(\rho_\theta^{\otimes n})$

$$\epsilon := \sup_{\theta\in\Theta} \frac{1}{2} \left\| \rho_\theta^{\otimes n} - \mathcal{D}_n\circ\mathcal{E}_n(\rho_\theta^{\otimes n}) \right\|_1. \tag{6}$$

The main result of the paper is the following:

Theorem 1. *Let $\{\rho_\theta^{\otimes n}\}_{\theta=(\xi,\mu)\in\Theta}$ be a generic family of n-copy qudit states with f_c independent classical parameters and*

f_q independent quantum parameters. For any $\delta \in (0, 2/9)$, the states in the family can be compressed into $[(1/2+\delta)f_c + (1/2)f_q]\log n$ classical bits and $(f_q\delta)\log n$ qubits with an error $\epsilon = O\left(n^{-\kappa(\delta)}\right) + O\left(n^{-\delta/2}\right)$, where $\kappa(\delta)$ is the error exponent of Q-LAN [17] (cf. Eq. (44) in the following). The protocol is optimal, in the sense that any compression protocol using a memory of size $[(f_c + f_q)/2 - \delta']\log n$ with $\delta' > 0$ cannot be asymptotically faithful.

The same results hold for a family $\{\rho_{\alpha,\beta}^{\otimes n}\}_{(\alpha,\beta)\in\Theta}$ of displaced thermal states, except that in this case the error is only $\epsilon = O\left(n^{-\delta/2}\right)$.

Theorem 1 is a sort of "equipartition theorem", stating that each independent parameter requires a memory of size $(1/2+\delta)\log n$. When the parameter is classical, the required memory is fully classical; when the parameter is quantum, a quantum memory of $\delta \log n$ qubits is required.

III. Compression of displaced thermal states

In this section, we focus on the compression of identically prepared displaced thermal states. We separately treat eight possible cases, corresponding to the possible combinations where the three parameter β, $|\alpha|$ and φ are either independent or fixed. The total memory cost for each case is determined by Theorem 1 and is summarized in Table I. On the other hand, the errors for all cases satisfy the unified bound

$$\epsilon = O\left(n^{-\delta/2}\right). \tag{7}$$

A. Quantum optical techniques used in the compression protocols

To construct compression protocols for the displaced thermal states, we adopt several tools in quantum optics. As a preparation, we introduce three quantum optical tools which are key components of the compression protocols: the beam splitter, the heterodyne measurement, and the quantum amplifier.

- *Beam splitter.* A beam splitter is a linear optical device implementing the unitary gate

$$U_\tau = \exp\left[i\tau(\hat{a}_1^\dagger \hat{a}_2 + \hat{a}_1 \hat{a}_2^\dagger)\right], \tag{8}$$

  where τ is a real parameter, and $\hat{a}_1$ and $\hat{a}_2$ are the annihilation operators associated to the two systems.
  Beam splitters can be used to split or merge laser beams. For example, one can use a beam splitter with $\tau = \pi/4$ to merge two identical coherent states into a single coherent state with larger amplitude:

$$U_{\pi/4}|\alpha\rangle_1 \otimes |\alpha\rangle_2 = |\sqrt{2}\alpha\rangle_1 \otimes |0\rangle_2. \tag{9}$$

  We will use beam splitters to manipulate the information about the parameter α in the n-copy state $\rho_{\alpha,\beta}^{\otimes n}$. In particular, we will make frequent use of the beam splitter unitary U_τ that implements the transformation

$$U_\tau\left(\rho_{\alpha_0,\beta} \otimes \rho_{\alpha_1,\beta}\right) U_\tau^\dagger = \rho_{\alpha_2,\beta} \otimes \rho_\beta^{\rm thm} \tag{10}$$

  where the displaced thermal states satisfy the relation $\alpha_0 = |\alpha_0|e^{i\varphi}$, $\alpha_1 = |\alpha_1|e^{i\varphi}$, and $\alpha_2 = \sqrt{|\alpha_0|^2 + |\alpha_1|^2}e^{i\varphi}$.
- *Heterodyne measurement.* The heterodyne measurement (see e.g. Section 3.5.2 of [25]) is a common measurement in continuous variable quantum optics. Here, the measurement outcome is a complex number $\hat{\alpha}$ and the corresponding measurement operator is the projector on the coherent state $|\hat{\alpha}\rangle$. The heterodyne POVM $\{\frac{\mathrm{d}^2\hat{\alpha}}{\pi}|\hat{\alpha}\rangle\langle\hat{\alpha}|\}$ is normalized in such a way that the integral over the complex plane gives the identity operator, namely

$$\int \frac{\mathrm{d}^2\hat{\alpha}}{\pi}\,|\hat{\alpha}\rangle\langle\hat{\alpha}| = I\,. \tag{11}$$

  We will use heterodyne measurements to estimate the amount of displacement of a displaced thermal state. For a displaced thermal state $\rho_{\alpha,\beta}$, the conditional probability density of finding the outcome $\hat{\alpha}$ is

$$\begin{aligned} Q(\hat{\alpha}|\alpha,\beta) &= \frac{1}{\pi}\langle\hat{\alpha}|\rho_{\alpha,\beta}|\hat{\alpha}\rangle \\ &= \frac{1}{\pi}\langle\hat{\alpha}-\alpha|\rho_\beta^{\rm thm}|\hat{\alpha}-\alpha\rangle \\ &= \frac{(1-\beta)}{\pi}\exp[-(1-\beta)|\hat{\alpha}-\alpha|^2]. \end{aligned} \tag{12}$$

- *Quantum amplifier.* A quantum amplifier is a device that increases the intensity of quantum light while preserving its phase information, namely a device which approximately implements the process $|\alpha\rangle \to |\gamma\alpha\rangle$ for $\gamma > 1$. Quantum amplification is an analogue of approximate quantum cloning for finite-dimensional systems, since coherent or displaced thermal states can be merged and split in a reversible fashion [cf. Eqs. (9), (10)]. In this work, we use the following amplifier [26]:

$$\begin{aligned} \mathcal{A}^\gamma(\rho) = \mathrm{Tr}_B\Big[&e^{\cosh^{-1}(\sqrt{\gamma})(\hat{a}^\dagger\hat{b}^\dagger - \hat{a}\hat{b})}(\rho \otimes |0\rangle\langle 0|_B) \\ &\times e^{\cosh^{-1}(\sqrt{\gamma})(\hat{a}\hat{b} - \hat{a}^\dagger\hat{b}^\dagger)}\Big] \end{aligned} \tag{13}$$

  where $\hat{a}$ and $\hat{b}$ are the annihilation operators of the input mode and the ancillary mode B, and γ is the amplification factor.

We now show details of the compression protocol for each case.

B. Case 1: fixed α, independent β

Let us start from Case 1, where the thermal parameter β is the only independent parameter. Note that the input state can be regarded as the state of n optical modes with each mode in a displaced thermal state, and thus the compression protocol can be regarded as a sequence of operations on the n-mode system.

Since α is known, we can get rid of the displacement using a certain unitary and convert the displaced thermal states into (undisplaced) thermal states. For the compression of thermal states, we have the following lemma:

Lemma 1 (Compression of identically prepared thermal states)**.** *Let $\{(\rho_\beta^{\rm thm})^{\otimes n}\}_{\beta\in[\beta_{\min},\beta_{\max}]}$ be a family of n-copy*

TABLE I
COMPRESSION RATE FOR DIFFERENT STATE FAMILIES. HERE $\delta > 0$ IS AN ARBITRARY POSITIVE CONSTANT.

Case	displacement $\alpha = |\alpha|e^{i\varphi}$	thermal parameter β	quantum bits	classical bits
0	fixed	fixed	0	0
1	fixed	independent	0	$(1/2+\delta)\log n$
2	independent	fixed	$2\delta \log n$	$\log n$
3	independent	independent	$2\delta \log n$	$(3/2+\delta)\log n$
4	φ independent; $\|\alpha\|$ fixed	fixed	$\delta \log n$	$(1/2)\log n$
5	φ independent; $\|\alpha\|$ fixed	independent	$\delta \log n$	$(1+\delta)\log n$
6	$\|\alpha\|$ independent; φ fixed	fixed	$\delta \log n$	$(1/2)\log n$
7	$\|\alpha\|$ independent; φ fixed	independent	$\delta \log n$	$(1+\delta)\log n$

thermal states. For any $\delta > 0$, there exists a protocol $\left(\mathcal{E}^{\rm thm}_{n,\delta}, \mathcal{D}^{\rm thm}_{n,\delta}\right)$ that compresses n copies of a thermal state $\rho^{\rm thm}_\beta$ into $(1/2+\delta)\log n$ classical bits with error

$$\epsilon_{\rm thm} = O\left(n^{-\delta}\right). \tag{14}$$

The proof of the above lemma can be found in the appendix. Note that no quantum memory is required to encode thermal states, in agreement with the intuition that β is classical, because all the states $\rho^{\rm thm}_\beta$ are diagonal in the same basis.

For any $\delta > 0$, the compression protocol for Case 1 (fixed α, independent β) is constructed as follows:

- *Encoder.*
  1) Transform each input copy with the displacement operation $\mathcal{D}_{-\alpha}$, defined by $\mathcal{D}_{-\alpha}(\cdot) := D_{-\alpha} \cdot D^\dagger_{-\alpha}$, where $D_{-\alpha} = \exp(-\alpha\hat{a}^\dagger + \bar{\alpha}\hat{a})$ is the displacement operator. The displacement operation transforms each input copy $\rho_{\alpha,\beta}$ into the thermal state $\rho^{\rm thm}_\beta$.
  2) Apply the thermal state encoder $\mathcal{E}^{\rm thm}_{n,\delta}$ in Lemma 1 on the n-mode state and the outcome is encoded in a classical memory.
- *Decoder.*
  1) Use the thermal state decoder $\mathcal{D}^{\rm thm}_{n,\delta}$ in Lemma 1 to recover the n copies of the thermal state $\rho^{\rm thm}_\beta$ from the classical memory.
  2) Perform the displacement operation $\mathcal{D}_\alpha$ on each mode.

Obviously, the memory cost and the error of the above protocol are given by Lemma 1.

C. Case 2: fixed β, independent α

Next we study the case when the displacement α is independent, while the thermal parameter β is fixed. The heuristic idea of the compression protocol is to gently test the input state, in order to extract information about the parameter α. The "gentle test" is based on a heterodyne measurement, performed on a small fraction of the n input copies. The information gained by the measurement is then used to perform suitable encoding operations on the remaining copies.

Let us see the details of the compression protocol. The key observation is that n identically displaced thermal states are unitarily equivalent to a single displaced thermal state, with the displacement scaled up by a factor $\sqrt{n}$, times the product of $n-1$ undisplaced thermal states. In formula, one has [27], [28]

$$\rho^{\otimes n}_{\alpha,\beta} = U^\dagger_{\rm BS}\left(\rho_{\sqrt{n}\alpha,\beta} \otimes (\rho^{\rm thm}_\beta)^{\otimes(n-1)}\right) U_{\rm BS}\ , \tag{15}$$

where $U_{\rm BS}$ is a suitable unitary gate, realizable with a circuit of beam splitter gates as in Eq. (10). Using Eq. (15), we can construct a protocol that separately processes the displaced thermal state $\rho_{\sqrt{n}\alpha,\beta}$ and the $n-1$ thermal modes, up to a gentle testing of the input state.

For any $\delta > 0$, the protocol for Case 2 (fixed β, independent α) runs as follows (see also Fig 1 for a flowchart illustration):

- *Preprocessing.* A preprocessing procedure is needed in order to store the estimate of α: Divide the range of α into n intervals, each labeled by a point $\hat{\alpha}_i$ in it, so that $|\alpha' - \alpha''| = O(n^{-1/2})$ (note that α is complex) for any α', α'' in the same interval.
- *Encoder.*
  1) Perform the unitary channel $\mathcal{U}_{\rm BS}(\cdot) = U_{\rm BS} \cdot U^\dagger_{\rm BS}$ on the input state, where $U_{\rm BS}$ is the unitary defined by Eq. (15). The output state has the form $\rho_{\sqrt{n}\alpha,\beta} \otimes (\rho^{\rm thm}_\beta)^{\otimes(n-1)}$.
  2) Send the first and the last mode through a group of beam splitters (10) that implements the transformation $\rho_{\alpha,\beta} \otimes \rho^{\rm thm}_\beta \to \rho_{\sqrt{n-n^{1-\delta}}\alpha,\beta} \otimes \rho_{\sqrt{n^{1-\delta}}\alpha,\beta}$. The n-mode state is now $\rho_{\sqrt{n-n^{1-\delta}}\alpha,\beta} \otimes (\rho^{\rm thm}_\beta)^{\otimes(n-2)} \otimes \rho_{\sqrt{n^{1-\delta}}\alpha,\beta}$.
  3) Estimate α by performing the heterodyne measurement $\{\frac{{\rm d}^2\alpha'}{\pi}|\alpha'\rangle\langle\alpha'|\}$ on the last mode. If the measurement outcome is α', use the value $\hat{\alpha} = \alpha'/\sqrt{n^{1-\delta}}$ as an estimate for the displacement α. The conditional probability distribution of the estimate is $Q(\hat{\alpha}|\alpha,\beta) = \frac{(1-\beta)}{\pi}\exp[-(1-\beta)n^{1-\delta}|\hat{\alpha} - \alpha|^2]$ [cf. Eq. (12)].
  4) Encode the label $\hat{\alpha}^*$ of the interval containing $\hat{\alpha}$ in a classical memory.
  5) Displace the first mode with $\mathcal{D}_{-\sqrt{n-n^{1-\delta}}\hat{\alpha}^*}$.

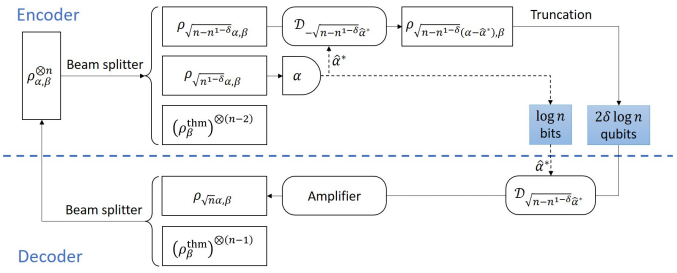


Fig. 1. **Compression protocol for displaced thermal states with fixed β and independent α.**

6) Truncate the state of the first mode in the photon number basis. The truncation is described by the channel $\mathcal{P}_{n^{2\delta}}$:

$$\mathcal{P}_{n^{2\delta}}(\rho) = P_{n^{2\delta}} \rho P_{n^{2\delta}} + (1 - \mathrm{Tr}[P_{n^{2\delta}} \rho])|0\rangle\langle 0| \tag{16}$$

where

$$P_{n^{2\delta}} = \sum_{m=0}^{n^{2\delta}} |m\rangle\langle m|. \tag{17}$$

The output state on the first mode is encoded in a quantum memory.

- *Decoder.*
  1) Read $\hat{\alpha}^*$ and perform the displacement operation $\mathcal{D}_{\sqrt{n-n^{1-\delta}}\hat{\alpha}^*}$ on the state of the quantum memory.
  2) Apply a quantum amplifier $\mathcal{A}^{\gamma_n}$ (13) with

$$\gamma_n = \frac{1}{1 - n^{-\delta}} \tag{18}$$

to the output.

  3) Prepare $(n-1)$ modes in the thermal state $\rho_\beta^{\mathrm{thm}}$, and perform on all the n modes the unitary channel $\mathcal{U}_{\mathrm{BS}}^{-1}$:

$$\mathcal{U}_{\mathrm{BS}}^{-1}(\rho) = U_{\mathrm{BS}}^\dagger \, \rho \, U_{\mathrm{BS}}. \tag{19}$$

The total memory cost consists of two parts: $\log n$ bits for encoding the (rounded) value $\hat{\alpha}^*$ of the estimate and $2\delta \log n$ qubits for encoding the first mode (in a displaced thermal state).

Let us analyze the error of the protocol. To upper bound the error, we first note that, with high probability, our estimate $\hat{\alpha}$ is close to the correct value, say $|\hat{\alpha} - \alpha| \leq f(n)$ for some function f vanishing for large n. When this happens, we can bound the error introduced by the truncation $\mathcal{P}_{n^{2\delta}}$ and by the amplification $\mathcal{A}^{\gamma_n}$. Otherwise, we just use the trivial error bound $\left\|\rho_{\alpha,\beta}^{\otimes n} - \mathcal{D}_n \circ \mathcal{E}_n(\rho_{\alpha,\beta}^{\otimes n})\right\|_1 \leq 2$. In this way, we obtain the bound

$$\begin{aligned}
\epsilon &= \sup_{\alpha,\beta} \frac{1}{2} \left\| \rho_{\alpha,\beta}^{\otimes n} - \mathcal{D}_n \circ \mathcal{E}_n \left(\rho_{\alpha,\beta}^{\otimes n} \right) \right\|_1 \\
&\leq \frac{1}{2} \sup_{\alpha,\beta} \Big\{ \sup_{\hat{\alpha}^* : |\hat{\alpha}^* - \alpha| \leq f(n)} \\
&\qquad \left\| \mathcal{A}^{\gamma_n} \circ \mathcal{D}_{\sqrt{n-n^{1-\delta}}\hat{\alpha}^*} \circ \mathcal{P}_{n^{2\delta}} \left(\rho_{\sqrt{n-n^{1-\delta}}(\alpha - \hat{\alpha}^*),\beta} \right) \right. \\
&\qquad \left. - \rho_{\sqrt{n}\alpha,\beta} \right\|_1 \Big\} + P(\alpha,\beta,n) \,,
\end{aligned} \tag{20}$$

where $P(\alpha,\beta,n)$ is the probability that $\hat{\alpha}$ deviates from α by more than $f(n)$, given by

$$\begin{aligned}
P(\alpha,\beta,n) &= \int_{|\hat{\alpha}-\alpha| > f(n)} \mathrm{d}^2\hat{\alpha} \, Q(\hat{\alpha}|\alpha,\beta) \\
&= \int_{|\hat{\alpha}-\alpha| > f(n)} \frac{\mathrm{d}^2\hat{\alpha}}{\pi} \, (1-\beta) \exp[-(1-\beta) n^{1-\delta} |\hat{\alpha} - \alpha|^2] \,,
\end{aligned} \tag{21}$$

having used Eq. (12) in the second equality. At this point, it is convenient to set $f(n) = n^{-1/2+3\delta/4}$, so that we obtain the relation

$$\begin{aligned}
&P(\alpha,\beta,n) \\
&= \int_{|\hat{\alpha}-\alpha| > n^{-1/2+3\delta/4}} \frac{\mathrm{d}^2\hat{\alpha}}{\pi} \, (1-\beta) \exp[-(1-\beta) n^{1-\delta} |\hat{\alpha} - \alpha|^2] \\
&= e^{-\Omega(n^{\delta/2})}.
\end{aligned} \tag{22}$$

Inserting this relation in Eq. (20), we obtain the bound

$$\begin{aligned}
\epsilon \leq \frac{1}{2} \sup_{\alpha,\beta} \Big\{ & \sup_{\hat{\alpha}^* : |\hat{\alpha}^* - \alpha| \leq f(n)} \\
& \left\| \mathcal{A}^{\gamma_n} \circ \mathcal{D}_{\sqrt{n-n^{1-\delta}}\hat{\alpha}^*} \circ \mathcal{P}_{n^{2\delta}} \left(\rho_{\sqrt{n-n^{1-\delta}}(\alpha - \hat{\alpha}^*),\beta} \right) \right. \\
& \left. - \rho_{\sqrt{n}\alpha,\beta} \right\|_1 \Big\} + e^{-\Omega(n^{\delta/2})} \,.
\end{aligned} \tag{23}$$

Now, we have to bound the first term in the right hand side. To this purpose, we split it into two terms, as follows

$$\begin{aligned}
&\left\| \mathcal{A}^{\gamma_n} \circ \mathcal{D}_{\sqrt{n-n^{1-\delta}}\hat{\alpha}^*} \circ \mathcal{P}_{n^{2\delta}} \left(\rho_{\sqrt{n-n^{1-\delta}}(\alpha - \hat{\alpha}^*),\beta} \right) - \rho_{\sqrt{n}\alpha,\beta} \right\|_1 \\
&\quad \leq \left\| \mathcal{A}^{\gamma_n} \circ \mathcal{D}_{\sqrt{n-n^{1-\delta}}\hat{\alpha}^*} \circ \mathcal{P}_{n^{2\delta}} \left(\rho_{\sqrt{n-n^{1-\delta}}(\alpha - \hat{\alpha}^*),\beta} \right) \right. \\
&\qquad \left. - \mathcal{A}^{\gamma_n} \circ \mathcal{D}_{\sqrt{n-n^{1-\delta}}\hat{\alpha}^*} \left(\rho_{\sqrt{n-n^{1-\delta}}(\alpha - \hat{\alpha}^*),\beta} \right) \right\|_1 \\
&\qquad + \left\| \mathcal{A}^{\gamma_n} \circ \mathcal{D}_{\sqrt{n-n^{1-\delta}}\hat{\alpha}^*} \left(\rho_{\sqrt{n-n^{1-\delta}}(\alpha - \hat{\alpha}^*),\beta} \right) - \rho_{\sqrt{n}\alpha,\beta} \right\|_1 \\
&\quad \leq \left\| \mathcal{P}_{n^{2\delta}} \left(\rho_{\sqrt{n-n^{1-\delta}}(\alpha - \hat{\alpha}^*),\beta} \right) - \rho_{\sqrt{n-n^{1-\delta}}(\alpha - \hat{\alpha}^*),\beta} \right\|_1 \\
&\qquad + \left\| \mathcal{A}^{\gamma_n} \left(\rho_{\sqrt{n-n^{1-\delta}}\alpha,\beta} \right) - \rho_{\sqrt{n}\alpha,\beta} \right\|_1 \,.
\end{aligned} \tag{24}$$

The two terms can be upper bounded individually. For the first term, we use the relations

$$\mathcal{A}^{\gamma} \left(\rho_{\alpha,\beta} \right) = \rho_{\sqrt{\gamma}\alpha,\beta'} \qquad \beta' = \frac{\beta + \gamma - 1}{\gamma} \tag{25}$$

and

$$\left\| \rho_{\beta'}^{\mathrm{thm}} - \rho_\beta^{\mathrm{thm}} \right\|_1 \leq \frac{2|\beta' - \beta|}{(1-\beta')^2} + O(|\beta' - \beta|^2) \,, \tag{26}$$

proven in Appendices B and C, respectively.

Using these two relations and Eq. (18), we obtain the bound

$$\begin{aligned}
&\frac{1}{2} \sup_{\alpha,\beta} \sup_{\hat{\alpha}^* : |\hat{\alpha}^* - \alpha| \leq n^{-1/2+3\delta/4}} \left\| \mathcal{A}^{\gamma_n} \left(\rho_{\sqrt{n-n^{1-\delta}}\alpha,\beta} \right) - \rho_{\sqrt{n}\alpha,\beta} \right\|_1 \\
&= O(n^{-\delta}) \,.
\end{aligned} \tag{27}$$

The first term in the right hand side of Eq. (24) can be bounded with the following lemma:

Lemma 2 (Photon number truncation of displaced thermal states.)**.** *Define the channel $\mathcal{P}_K$ as*

$$\mathcal{P}_K(\rho) = P_K \rho P_K + (1 - \mathrm{Tr}[P_K \rho])|0\rangle\langle 0| \tag{28}$$

where $P_K = \sum_{k=0}^K |k\rangle\langle k|$. When $K = \Omega\left(|\alpha|^{2+x}\right)$, $\mathcal{P}_K$ satisfies

$$\epsilon(\rho_{\alpha,\beta}) := \frac{1}{2} \left\| \mathcal{P}_K \left(\rho_{\alpha,\beta} \right) - \rho_{\alpha,\beta} \right\|_1 = \beta^{\Omega(K^{x/8})} + e^{-\Omega(K^{x/4})} \tag{29}$$

for any $0 \leq \beta < 1$.

See Appendix D for the proof.

In our case, we are using the projector $P_{n^{2\delta}}$ in Eq. (17), and the displacement is $\sqrt{n-n^{1-\delta}}(\alpha-\hat{\alpha}^*)$. Since $\sqrt{n-n^{1-\delta}}|\alpha-\hat{\alpha}^*| = O\left(n^{3\delta/4}\right)$, by Lemma 2 we obtain the bound

$$\begin{aligned}\frac{1}{2}\sup_{\alpha,\beta}\sup_{\hat{\alpha}^*:|\hat{\alpha}^*-\alpha|\le n^{-1/2+3\delta/4}}&\left\|\mathcal{P}_{n^{2\delta}}(\rho_{\sqrt{n-n^{1-\delta}}(\alpha-\hat{\alpha}^*),\beta})\right.\\&\left.-\rho_{\sqrt{n-n^{1-\delta}}(\alpha-\hat{\alpha}^*),\beta}\right\|_1\\&=\beta^{\Omega\left(n^{\delta/12}\right)}+e^{-\Omega(n^{\delta/6})}\,.\end{aligned}\tag{30}$$

Combining Eqs. (23), (27), and (30), we finally get the error bound

$$\begin{aligned}\epsilon&\le e^{-\Omega(n^{\delta/2})}+O\left(n^{-\delta}\right)+\beta^{\Omega\left(n^{\delta/12}\right)}+e^{-\Omega(n^{\delta/6})}\\&=O\left(n^{-\delta}\right).\end{aligned}\tag{31}$$

D. Case 3: independent α and β

Case 3 (independent α and β) can be treated in a similar way as Case 2. The main difference is that, since one mode is consumed in the estimation of α, we have to estimate also β to reconstruct this mode. Luckily, the thermal parameter β can be estimated freely (i.e. without disturbing the input state), and thus its estimation strategy is simpler than that of α.

For any $\delta > 0$ we can construct the protocol for Case 3 (independent α and β) as follows (see also Fig. 2 for a flowchart illustration):

- *Preprocessing.* Divide the range of α into n intervals, each labeled by a point $\hat{\alpha}_i$ in it, so that $|\alpha'-\alpha''| = O(n^{-1/2})$ for any α', α'' in the same interval.
- *Encoder.*
  1) Perform the unitary channel $\mathcal{U}_{\rm BS}(\cdot) = U_{\rm BS}\cdot U_{\rm BS}^\dagger$ on the input state, where $U_{\rm BS}$ is the unitary defined by Eq. (15). The output state has the form $\rho_{\sqrt{n}\alpha,\beta}\otimes(\rho_\beta^{\rm thm})^{\otimes(n-1)}$.
  2) Estimate β with the von Neumann measurement of the photon number on the $n-1$ copies of $\rho_\beta^{\rm thm}$ and denote by $\hat{\beta}$ the maximum likelihood estimate of β. Note that the $n-1$ copies will not be disturbed by the photon number measurement because they are diagonal in the photon number basis.
  3) Send the first and the last mode through a group of beam splitters (10) that implements the transformation $\rho_{\alpha,\beta}\otimes\rho_\beta^{\rm thm}\to\rho_{\sqrt{n-n^{1-\delta}}\alpha,\beta}\otimes\rho_{\sqrt{n^{1-\delta}}\alpha,\beta}$. The n-mode state is now $\rho_{\sqrt{n-n^{1-\delta}}\alpha,\beta}\otimes(\rho_\beta^{\rm thm})^{\otimes(n-2)}\otimes\rho_{\sqrt{n^{1-\delta}}\alpha,\beta}$.
  4) Estimate α by performing the heterodyne measurement $\{\frac{\mathrm{d}^2\alpha'}{\pi}|\alpha'\rangle\langle\alpha'|\}$ on the last mode, which yields an estimate $\hat{\alpha} = \alpha'/\sqrt{n^{1-\delta}}$ with the probability distribution $Q(\hat{\alpha}|\alpha,\beta)$ as in Eq. (12). Encode the label $\hat{\alpha}^*$ of the interval containing $\hat{\alpha}$ in a classical memory.
  5) Displace the first mode with $\mathcal{D}_{-\sqrt{n-n^{1-\delta}}\hat{\alpha}^*}$.
  6) Prepare the n-th mode in the thermal state $\rho_{\hat{\beta}}^{\rm thm}$. The n-mode state is now $\rho_{\sqrt{n-n^{1-\delta}}(\alpha-\hat{\alpha}^*),\beta}\otimes(\rho_\beta^{\rm thm})^{\otimes(n-2)}\otimes\rho_{\hat{\beta}}^{\rm thm}$.
  7) Truncate the state of the first mode, using the channel $\mathcal{P}_{n^{2\delta}}$ defined by Eq. (28). The output state is encoded in a quantum memory.
  8) Use the thermal state encoder $\mathcal{E}_{n-1,\delta}^{\rm thm}$ (see Lemma 1) to compress the remaining $n-1$ modes and encode the output state in a classical memory.
- *Decoder.*
  1) Read $\hat{\alpha}^*$ and perform the displacement $\mathcal{D}_{\sqrt{n-n^{1-\delta}}\hat{\alpha}^*}$ on the state of the quantum memory.
  2) Apply a quantum amplifier (13) $\mathcal{A}^{\gamma_n}$ with $\gamma_n = 1/(1-n^{-\delta})$ to the state.
  3) Use the thermal state decoder $\mathcal{D}_{n-1,\delta}^{\rm thm}$ to recover the other $(n-1)$ modes in the thermal state $\rho_\beta^{\rm thm}$ from the memory.
  4) Perform the channel $\mathcal{U}_{\rm BS}^{-1}$.

The memory cost of the protocol consists of three parts: $\log n$ bits for encoding the (rounded) value $\hat{\alpha}^*$ of the estimate, $2\delta\log n$ qubits for encoding the first mode (displaced thermal state), and $(1/2+\delta)\log n$ bits for encoding the other modes (thermal states). Overall, the protocol requires $2\delta\log n$ qubits and $(3/2+\delta)\log n$ classical bits.

On the other hand, the error of the protocol can be analyzed in a similar way as in Case 2, with the only difference that an extra error is introduced by estimating and compressing the thermal states. The state of the modes after the estimation step is

$$\rho_\beta^{\rm est} = \left(\int \mathrm{d}\hat{\beta}\, P(\hat{\beta}|\beta)\,\rho_{\hat{\beta}}^{\rm thm}\right)\otimes(\rho_\beta^{\rm thm})^{\otimes(n-2)}\,.\tag{32}$$

where $P(\hat{\beta}|\beta)$ is the probability density of estimating $\hat{\beta}$ when the true value is β. Applying the thermal state compression to this state, we obtain the output state $\mathcal{D}_{n,\delta}^{\rm thm}\circ\mathcal{E}_{n,\delta}^{\rm thm}(\rho_\beta^{\rm est})$, whose distance from the initial state can be bounded as

$$\begin{aligned}\epsilon_\beta &:= \frac{1}{2}\left\|\mathcal{D}_{n,\delta}^{\rm thm}\circ\mathcal{E}_{n,\delta}^{\rm thm}(\rho_\beta^{\rm est})-\left(\rho_\beta^{\rm thm}\right)^{\otimes(n-1)}\right\|_1\\&\le\frac{1}{2}\sup_\beta\left\{\left\|\mathcal{D}_{n,\delta}^{\rm thm}\circ\mathcal{E}_{n,\delta}^{\rm thm}\left(\rho_\beta^{\rm est}\right)\right.\right.\\&\quad\left.-\mathcal{D}_{n,\delta}^{\rm thm}\circ\mathcal{E}_{n,\delta}^{\rm thm}\left[(\rho_\beta^{\rm thm})^{\otimes(n-1)}\right]\right\|_1\\&\quad\left.+\left\|\mathcal{D}_{n,\delta}^{\rm thm}\circ\mathcal{E}_{n,\delta}^{\rm thm}\left[(\rho_\beta^{\rm thm})^{\otimes(n-1)}\right]-(\rho_\beta^{\rm thm})^{\otimes(n-1)}\right\|_1\right\}\\&\le\frac{1}{2}\sup_\beta\left\|\rho_\beta^{\rm est}-(\rho_\beta^{\rm thm})^{\otimes(n-1)}\right\|_1+O\left(n^{-\delta}\right),\end{aligned}\tag{33}$$

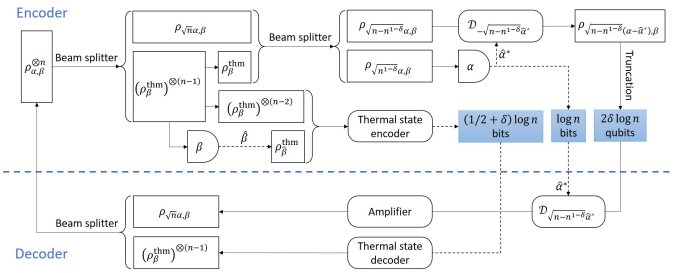


Fig. 2. **Compression protocol for displaced thermal states with independent α and β.**

having used Lemma 1 in the last inequality. The remaining term can be bounded as

$$\left\|\rho_\beta^{\rm est}-(\rho_\beta^{\rm thm})^{\otimes(n-1)}\right\|_1 \le \left\|\int {\rm d}\hat\beta P(\hat\beta|\beta)\left(\rho_{\hat\beta}^{\rm thm}-\rho_\beta^{\rm thm}\right)\right\|_1 . \tag{34}$$

Now, we split the integral in the right hand side of Eq. (34) into two terms, corresponding to the values of $\hat\beta$ in regions $\mathsf{R}_\le := \{\hat\beta\in\mathbb{C} \mid |\hat\beta-\beta|\le n^{-(1+\delta)/2}\}$ and $\mathsf{R}_> = \mathbb{C}\setminus\mathsf{R}_\le$. In this way, we obtain the bound

$$\begin{aligned}&\left\|\int {\rm d}\hat\beta P(\hat\beta|\beta)\,\rho_{\hat\beta}^{\rm thm}-\rho_\beta^{\rm thm}\right\|_1\\ \le&\sup_\beta\sup_{\hat\beta\in\mathsf{R}_\le}\left\|\rho_{\hat\beta}^{\rm thm}-\rho_\beta^{\rm thm}\right\|_1+2\int_{\mathsf{R}_>}{\rm d}\hat\beta P(\hat\beta|\beta)\,,\end{aligned}$$

having used the elementary inequality $\left\|\rho_{\hat\beta}^{\rm thm}-\rho_\beta^{\rm thm}\right\|_1 \le 2$. The first term in the right hand side is bounded by $O(n^{-(1+\delta)/2})$ using Eq. (26), while the second error term is bounded by the following property of the maximum likelihood estimate [29]

$$\int_{|\hat\beta-\beta|\ge l/\sqrt{nF_\beta}}{\rm d}\hat\beta P(\hat\beta|\beta)\le {\rm erfc}\left(\frac{l}{\sqrt{2}}\right) \tag{35}$$

where $F_\beta=(\beta^2+1)/[\beta(1-\beta)^3]$ is the Fisher information of β and ${\rm erfc}(x):=(2/\pi)\int_x^\infty e^{-s^2}{\rm d}s$ is the complementary error function. Picking $l=n^{-\delta/2}\sqrt{F_\beta}$, we have

$$\int_{\mathsf{R}_>}{\rm d}\hat\beta P(\hat\beta|\beta)\le {\rm erfc}\left(\frac{n^{-\delta/2}\sqrt{F_\beta}}{\sqrt{2}}\right)=e^{-\Omega(n^\delta)}. \tag{36}$$

In conclusion, ϵ_β can be bounded as

$$\begin{aligned}\epsilon_\beta &\le O(n^{-\delta})+O(n^{-(1+\delta)/2})+e^{-\Omega(n^\delta)}\\ &= O(n^{-\delta})\,.\end{aligned} \tag{37}$$

The remaining contribution to the error can be bounded as in Eq. (31), leading to an overall error of size $O(n^{-\delta})$.

E. Case 4 (independent φ, fixed $|\alpha|$ and β) and Case 6 (independent $|\alpha|$, fixed φ and β).

In Case 4 (independent φ, fixed $|\alpha|$ and β) and Case 6 (independent $|\alpha|$, fixed φ and β), the displacement α is partially known. Such a knowledge allows us to reduce the amount of memory.

The protocols for these two cases 4 and 6 are very similar. Let us start from Case 4, where the phase of the displacement is independent while the modulus is fixed. The protocol for Case 4 (independent φ, fixed $|\alpha|$ and β) runs as follows:

- *Preprocessing.* Divide the range of φ into $n^{1/2}$ intervals, each labeled by a point $\hat\varphi_i$ in it, so that $|\varphi'-\varphi''|=O(n^{-1/2})$ for any φ',φ'' in the same interval.
- *Encoder.*
  1) Perform the unitary channel $\mathcal{U}_{\rm BS}(\cdot)$ on the input state to transform it into $\rho_{\sqrt{n}\alpha,\beta}\otimes(\rho_\beta^{\rm thm})^{\otimes(n-1)}$.
  2) Send the first and the last mode through a group of beam splitters (10) that implements the transformation $\rho_{\alpha,\beta}\otimes\rho_\beta^{\rm thm}\to\rho_{\sqrt{n-n^{1-\delta/2}}\alpha,\beta}\otimes\rho_{\sqrt{n^{1-\delta/2}}\alpha,\beta}$.
  3) Estimate φ by the heterodyne measurement $\{\frac{{\rm d}^2\alpha'}{\pi}|\alpha'\rangle\langle\alpha'|\}$ on the last mode, which yields an estimate $\hat\varphi$ which is the phase of α'. Encode the label $\hat\varphi^*$ of the interval containing $\hat\varphi$ in a classical memory.
  4) Displace the first mode with $\mathcal{D}_{-\sqrt{n-n^{1-\delta/2}}\hat\alpha^*}$ with $\hat\alpha^*:=|\alpha|e^{i\hat\varphi^*}$.
  5) Send the state of the first mode through a truncation channel $\mathcal{P}_{n^\delta}$ defined in (28) and encode the output state in a quantum memory.
- *Decoder.*
  1) Read $\hat\alpha^*$ and perform the displacement $\mathcal{D}_{\sqrt{n-n^{1-\delta/2}}\hat\alpha^*}$ on the state of the quantum memory.
  2) Apply the quantum amplifier $\mathcal{A}^{\gamma_n}$ with $\gamma_n=1/(1-n^{-\delta/2})$.
  3) Prepare the other $(n-1)$ modes in the thermal state $\rho_\beta^{\rm thm}$.
  4) Perform $\mathcal{U}_{\rm BS}^{-1}$ on the thermal state $\rho_\beta^{\rm thm}$ and the quantum memory.

The protocol for Case 6 works in the same way except that $|\alpha|$ is estimated instead of φ. For both cases the memory cost consists of two parts: $(1/2)\log n$ bits for encoding the (rounded) value $\hat\alpha^*$ of the estimate and $\delta\log n$ qubits for encoding the first mode (displaced thermal state). The error can be bounded as previous as $\epsilon=O\left(n^{-\delta/2}\right)$.

F. Case 5 (fixed $|\alpha|$, independent φ and β) and Case 7 (fixed φ, independent $|\alpha|$ and β).

Case 5 (fixed $|\alpha|$, independent φ and β) and Case 7 (fixed φ, independent $|\alpha|$ and β) can be treated in the same way as Case 4 (independent φ, fixed $|\alpha|$ and β) and Case 6 (independent $|\alpha|$, fixed φ and β), except that the thermal parameter β is now independent. We illustrate only the protocol for Case 5 (fixed $|\alpha|$, independent φ and β) and the other naturally follows. The protocol runs as follows:

- *Preprocessing.* Divide the range of φ into $n^{1/2}$ intervals, each labeled by a point $\hat\varphi_i$ in it, so that $|\varphi'-\varphi''|=O(n^{-1/2})$ for any φ',φ'' in the same interval.
- *Encoder.*
  1) Perform the unitary channel $\mathcal{U}_{\rm BS}(\cdot)$ on the input state to transform it into $\rho_{\sqrt{n}\alpha,\beta}\otimes(\rho_\beta^{\rm thm})^{\otimes(n-1)}$.
  2) Estimate β with the von Neumann measurement of the photon number on the $n-1$ copies of $\rho_\beta^{\rm thm}$. Denote by $\hat\beta$ the maximum likelihood estimate of β.
  3) Send the first and the last mode through a group of beam splitters (10) that implements the transformation $\rho_{\alpha,\beta}\otimes\rho_\beta^{\rm thm}\to\rho_{\sqrt{n-n^{1-\delta/2}}\alpha,\beta}\otimes\rho_{\sqrt{n^{1-\delta/2}}\alpha,\beta}$.
  4) Estimate φ by the heterodyne measurement $\{\frac{{\rm d}^2\alpha'}{\pi}|\alpha'\rangle\langle\alpha'|\}$ on the last mode, which yields an estimate $\hat\varphi$ which is the phase of α'. Encode the label $\hat\varphi^*$ of the interval containing $\hat\varphi$ in a classical memory.
  5) Displace the first mode with $\mathcal{D}_{-\sqrt{n-n^{1-\delta/2}}\hat\alpha^*}$ with $\hat\alpha^*:=|\alpha|e^{i\hat\varphi^*}$.

6) Prepare the n-th mode in the thermal state $\rho^{\rm thm}_{\hat\beta}$. The n-mode state is now $\rho_{\sqrt{n-n^{1-\delta/2}}(\alpha-\hat\alpha^*),\beta}\otimes(\rho^{\rm thm}_\beta)^{\otimes(n-2)}\otimes\rho^{\rm thm}_{\hat\beta}$.
7) Send the state of the first mode through a truncation channel $\mathcal{P}_{n^\delta}$ defined in (28) and encode the output state in a quantum memory.
8) Use the thermal state encoder $\mathcal{E}^{\rm thm}_{n-1,\delta}$ (see Lemma 1) to compress the remaining $n-1$ modes and encode the output state in a classical memory.

- *Decoder.*
1) Read $\hat\alpha^*$ and perform the displacement $\mathcal{D}_{\sqrt{n-n^{1-\delta/2}}\hat\alpha^*}$ on the state of the quantum memory.
2) Apply the quantum amplifier $\mathcal{A}^{\gamma_n}$ with $\gamma_n=1/(1-n^{-\delta/2})$.
3) Use the thermal state decoder $\mathcal{D}^{\rm thm}_{n-1,\delta}$ to recover the other $(n-1)$ modes in the thermal state $\rho^{\rm thm}_\beta$ from the memory.
4) Perform $\mathcal{U}^{-1}_{\rm BS}$ on the output of $\mathcal{D}^{\rm thm}_{n-1,\delta}$ and the quantum memory.

The protocol for Case 7 works in the same way except that $|\alpha|$ is estimated instead of φ. For both cases the memory cost consists of three parts: $(1/2)\log n$ bits for encoding the (rounded) value $\hat\alpha^*$ of the estimate, $\delta\log n$ qubits for encoding the first mode (displaced thermal state), and $(1/2+\delta)\log n$ bits for encoding the other modes (thermal states). Overall, the protocol requires $\delta\log n$ qubits and $(1+2\delta)\log n$ classical bits. The error can be bounded as previous as $\epsilon=O\left(n^{-\delta/2}\right)$.

IV. Compression of identically prepared finite dimensional systems.

In this section, we study the compression of finite-dimensional non-degenerate quantum systems, using quantum local asymptotic normality and leveraging on our results on displaced thermal states. We show that, just as for displaced thermal states, each independent parameter of a qudit family requires $(1/2+\delta)\log n$ memory for any $\delta>0$. The compression protocol is introduced in the following.

A. The compression protocol

To construct a compression protocol, we will use the following techniques:

- *Quantum local asymptotic normality (Q-LAN).* The quantum version of local asymptotic normality has been derived in several different forms [15], [16], [17], [18]. Here we use the version of [17], which states that n identical copies of a qudit state can be locally approximated by a classical-quantum Gaussian state in the large n limit. Explicitly, for a fixed point $\theta_0=(\xi_0,\mu_0)$, one defines the neighborhood

$$\Theta_{n,x}(\theta_0)=\left\{\theta=\theta_0+\delta\theta/\sqrt{n},\ |\ \|\delta\theta\|_\infty\le n^{\frac{x}{2}}\right\}, \tag{38}$$

where $\|\delta\theta\|_\infty$ is the max vector norm $\|\delta\theta\|_\infty:=\max_i(\delta\theta)_i$ and $x\in(0,1)$. Q-LAN states that every n-fold product state $\rho_\theta^{\otimes n}$ with θ in the neighborhood $\Theta_{n,x}(\theta_0)$ can be approximated by a classical-quantum Gaussian state:

$$\begin{aligned}G_{n,\theta}&=N\left(\delta\mu,V_{\mu_0}\right)\otimes\Phi\left(\delta\xi,\mu_0\right)\\ \Phi\left(\delta\xi,\mu_0\right)&=\bigotimes_{1\le j<k\le d}\rho_{\alpha_{j,k},\beta_{j,k}},\end{aligned} \tag{39}$$

where $(\delta\xi,\delta\mu)=\sqrt{n}(\theta-\theta_0)$, $N\left(\delta\mu,V_{\mu_0}\right)$ is the multivariate normal distribution with mean $\delta\mu$ and covariance matrix V_{μ_0} (equal to the inverse of the Fisher information of the $(d-1)$-dimensional family of probability distributions $\{\mu\}$, evaluated at $\mu=\mu_0$) and $\rho_{\alpha_{j,k},\beta_{j,k}}$ is the displaced thermal state defined as

$$\rho_{\alpha_{j,k},\beta_{j,k}}=D_{\alpha_{j,k}}\rho^{\rm thm}_{\beta_{j,k}}D^\dagger_{\alpha_{j,k}} \tag{40}$$

$$\alpha_{j,k}=\delta\xi^I_{j,k}+i\delta\xi^R_{j,k}\qquad\beta_{j,k}=\frac{(\mu_0)_k}{(\mu_0)_j}. \tag{41}$$

where $\{(\mu_0)_j\}_{j=1}^{d-1}$ are components of μ_0 and $(\mu_0)_d:=1-\sum_{k=1}^{d-1}(\mu_0)_k$. In the neighborhood $\Theta_{n,x}(\theta_0)$ of θ_0, it is possible to construct two quantum channels $\mathcal{T}^{(n)}_{\theta_0}$ and $\mathcal{S}^{(n)}_{\theta_0}$, which depend on θ_0 and n (but not on the exact value of θ). Using these two channels, n-copy qudit states and Gaussian states can be interconverted with an error vanishing in n [17]. Explicitly, one has the following bounds

$$\sup_{\theta\in\Theta_{n,x}(\theta_0,c)}\left\|\mathcal{T}^{(n)}_{\theta_0}\left(\rho_\theta^{\otimes n}\right)-G_{n,\theta}\right\|_1=O\left(n^{-\kappa(x)}\right) \tag{42}$$

$$\sup_{\theta\in\Theta_{n,x}(\theta_0,c)}\left\|\rho_\theta^{\otimes n}-\mathcal{S}^{(n)}_{\theta_0}\left(G_{n,\theta}\right)\right\|_1=O\left(n^{-\kappa(x)}\right), \tag{43}$$

where $\|\cdot\|_1$ denotes the trace norm and $\kappa(x)$ is defined as

$$\kappa(x)=\min\left\{\frac{1-z-\eta}{2},\frac{1-3x}{4}-y,\frac{2-9\eta}{24}\right\} \tag{44}$$

where y,z,η can be freely chosen under the constraints $(1+x)/2<z<1$, $y>0$, $\eta>0$ and $\eta>x-y$. With proper values for y,z, and η, when $x\in[0,2/9)$, the exponent $\kappa(x)$ is a non-increasing function of x and falls within the interval $[0.027,0.084]$.

- *Quantum state tomography.* State tomography is an important technique used to determine the density matrix of an unknown quantum state. In our protocol, the role of tomography is to provide a rough estimate of θ_0 so that we can apply Q-LAN. We adopt the tomographic protocol proposed in [30], which provides an estimate $\rho_{\hat\theta}$ of a qudit state ρ_θ. When the protocol is carried out on n copies of the state ρ_θ, the estimate satisfies the bound

$$\mathbf{Prob}\left[\frac{1}{2}\|\rho_\theta-\rho_{\hat\theta}\|_1\le\varepsilon\right]\ge1-(n+1)^{3d^2}e^{-n\varepsilon^2} \tag{45}$$

using n copies of the state.

Our compression protocol is illustrated in Fig. 3. For any $\delta\in(0,2/9)$, the protocol consists of the following steps:

- *Preprocessing.* Divide the parameter space Θ into a lattice $\mathsf{L} := \{\theta \in \Theta \mid \theta_i = z_i/(2\sqrt{n}), z_i \in \mathbb{Z} \ \forall \ i\}$. The lattice has approximately $n^{(f_c+f_q)/2}$ points, which will be used to store the outcome of tomography.
- *Encoder.* The encoder of $\rho_\theta^{\otimes n}$ consists of five steps:
  1) *Tomography.* Use $n^{1-\delta/2}$ copies of ρ_θ for quantum tomography, which yields an estimate $\hat{\theta}$ of θ.
  2) *Storage of the estimate.* Encode the estimate $\hat{\theta}$ as a point in the lattice
  $$\mathsf{L} := \{\theta \in \Theta \mid \theta_i = z_i/(2\sqrt{n}), z_i \in \mathbb{Z} \ \forall \ i\}. \quad (46)$$
  Choose the lattice point θ_0 that is closest to $\hat{\theta}$, namely
  $$\theta_0 := \operatorname*{argmin}_{\theta' \in \mathsf{L}} \|\hat{\theta} - \theta'\|_\infty. \quad (47)$$
  3) *Q-LAN.* After the tomography step, we end up with $n - n^{1-\delta/2}$ copies of the state. Define
  $$\gamma_n' = \frac{1}{1 - n^{-\delta/2}}. \quad (48)$$
  so that the number of remaining copies is n/γ_n'. The n/γ_n' copies are sent through the channel $\mathcal{T}_{\theta_0}^{(n/\gamma_n')}$ (42) which outputs the Gaussian state $G_{(n/\gamma_n'),\theta}$ defined by Eq. (39).
  4) *Amplification.* To compensate the loss of copies in tomography, the state $\rho_{\alpha_{j,k},\beta_{j,k}}$ of each quantum mode is amplified by the amplifier defined in Eq. (13) with $\gamma = \gamma_n'$. The Gaussian distribution on the classical register is rescaled by a constant factor:
  $$\mathcal{A}_{\rm c}(\rho) := \sum_u \langle u|\rho|u\rangle \, |\sqrt{\gamma_n'}u\rangle\langle\sqrt{\gamma_n'}u| \quad (49)$$
  where $\{|u\rangle\langle u|\}$ denotes the basis of the classical register with $u \in \mathbb{R}^{f_c}$. Notice that, in practice, u can be approximated by a long sequence of classical bits with arbitrary high precision.
  The whole amplification process is described by the channel $\mathcal{A}_{\theta_0}^{(n/\gamma_n')\to n}$ defined as the following:
  $$\mathcal{A}_{\theta_0}^{(n/\gamma_n')\to n} = \mathcal{A}_{\rm c} \otimes \mathcal{A}_{\rm q} \qquad \mathcal{A}_{\rm q} = \bigotimes_{j<k} \mathcal{A}^{\gamma_n'}. \quad (50)$$
  where $\mathcal{A}_{\rm c}$ is the classical amplifier just defined in Eq. (49).

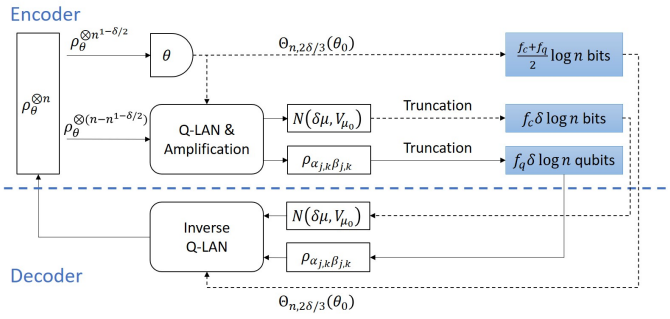


Fig. 3. **Compression protocol for qudit states.**

  5) *Gaussian state compression.* Each quantum mode of the amplified Gaussian state is then truncated by $\mathcal{P}_{n^\delta}$, defined by Eq. (28). The output state is then stored in a quantum memory of size $\delta \log n$ for each mode. The classical mode is compressed by a map $\mathcal{P}_{\rm c}$ that truncates the state into a $O(n^{\delta/2})$-hypercube centered around the mean of the Gaussian and rounds the continuous variable u into a discrete lattice. Explicitly, we have
  $$\mathcal{P}_{\rm c}(\rho) = \sum_{\|u\|_\infty \le n^{\delta/2}} \langle u|\rho|u\rangle \, |r_{n,\delta}(u)\rangle\langle r_{n,\delta}(u)| + \left[1 - \sum_{\|u'\|_\infty \le n^{\delta/2}} \langle u'|\rho|u'\rangle\right] |0\rangle\langle 0|, \quad (51)$$
  where $r_{n,\delta}(u)$ is the rounding function which maps $u \in \mathbb{R}^{f_c}$ to the closest point on the lattice $\left(\mathbb{Z}/n^{\delta/2}\right)^{f_c}$. The output of $\mathcal{P}_c$ is stored in classical memory. The memory size is determined by the number of lattice points covered by the range of truncation. The separation between lattice points is $n^{-\delta/2}$, and the range of truncation is $[-n^{\delta/2}, n^{\delta/2}]^{f_c}$, so it covers $O(n^{f_c\delta})$ points on the lattice. We therefore need $f_c\delta \log n$ bits.
  The whole process for Gaussian state compression is described by the channel
  $$\mathcal{P}_{\theta_0}^{(n)} = \mathcal{P}_{\rm c} \otimes \mathcal{P}_{\rm q} \qquad \mathcal{P}_{\rm q} = \bigotimes_{j<k} \mathcal{P}_{n^\delta}. \quad (52)$$

- *Decoder.* Read $|t\rangle\langle t| \in (\mathbb{Z}/n^{\delta/2})^{f_c}$ from the classical memory that stores the discretized Gaussian distribution and perform a uniform sampling in a shrinking hypercube containing t:
  $$\mathsf{C}_{n,\delta}(t) := \left\{\hat{u} \in \mathbb{R}^{f_c} \mid \|\hat{u} - t\|_\infty \le (1/2)n^{-\delta/2}\right\}. \quad (53)$$
  This step converts the discrete random variable t, whose distribution is a weighted sum of Dirac deltas, into a continuous random variable $\hat{u}$ with probability distribution that approximates $N(\delta\mu, V_{\mu_0})$. The output sample $\hat{u}$ together with the state of the quantum memory is sent through the channel $\mathcal{S}_{\theta_0}^{(n)}$ (43), which can be constructed from the outcome of tomography.

B. Error analysis.

To bound the error of our protocol, we need to specify a small neighborhood for discussion, which should contain the true value θ with high probability. A proper choice is the neighborhood $\Theta_{n,2\delta/3}(\theta_0)$ (38). Using the triangle inequality of trace distance, we split the overall error into four terms

$$\epsilon \le \epsilon_{\rm tomo} + \epsilon_{\rm amp} + \epsilon_{\rm G} + \epsilon_{\rm Q-LAN}, \quad (54)$$

where

$$\epsilon_{\rm tomo} = \mathbf{Prob}\left[\theta \notin \Theta_{n,2\delta/3}(\theta_0)\right] \tag{55}$$

$$\epsilon_{\rm amp} = \frac{1}{2}\sup_{\theta_0}\sup_{\theta\in\Theta_{n,2\delta/3}(\theta_0)} \left\| \mathcal{A}_{\theta_0}^{(n/\gamma_n')\to n}\left(G_{n/\gamma_n',\theta}\right) - G_{n,\theta}\right\|_1 \tag{56}$$

$$\epsilon_{\rm G} = \frac{1}{2}\sup_{\theta_0}\sup_{\theta\in\Theta_{n,2\delta/3}(\theta_0)} \left\| \mathcal{P}_{\theta_0}^{(n)}\left(G_{n,\theta}\right) - G_{n,\theta}\right\|_1 \tag{57}$$

$$\epsilon_{\rm Q-LAN} = \frac{1}{2}\sup_{\theta_0}\sup_{\theta\in\Theta_{n,2\delta/3}(\theta_0)} \left\{ \left\| \mathcal{T}_{\theta_0}^{(n/\gamma_n')}\left(\rho_\theta^{\otimes(n/\gamma_n')}\right) - G_{n/\gamma_n',\theta}\right\|_1 + \left\|\rho_\theta^{\otimes n} - \mathcal{S}_{\theta_0}^{(n)}\left(G_{n,\theta}\right)\right\|_1 \right\} \tag{58}$$

are the error terms of tomography, amplification, truncation, and Q-LAN, respectively. In the following, we will provide upper bounds for all four terms.

Let us start from the tomography error. By definition of the neighborhood $\Theta_{n,2\delta/3}(\theta_0)$ (38), we have

$$\begin{aligned}\epsilon_{\rm tomo} &= \mathbf{Prob}\left[\|\theta-\theta_0\|_\infty > n^{-1/2+\delta/3}\right]\\ &\le \mathbf{Prob}\left[\|\theta-\hat\theta\|_\infty + \|\hat\theta-\theta_0\|_\infty > n^{-1/2+\delta/3}\right]\\ &\le \mathbf{Prob}\left[\|\theta-\hat\theta\|_\infty > n^{-1/2+\delta/3}(1-O(n^{-\delta/3}))\right].\end{aligned} \tag{59}$$

The first inequality comes from triangle inequality and the second inequality holds since $\|\hat\theta-\theta_0\|_\infty \le 1/(2\sqrt{n})$ which is an immediate implication of Eq. (46) and Eq. (47).

To further bound the error, notice that the trace distance has a Euclidean expansion, of the form $\|\rho_\theta-\rho_{\theta'}\|_1 = C\|\theta-\theta'\|_\infty + O(\|\theta-\theta'\|_\infty^2)$, where $C>0$ is a suitable constant. Then we have

$$\epsilon_{\rm tomo} \le \mathbf{Prob}\left[\frac{1}{2}\|\rho_\theta-\rho_{\hat\theta}\|_1 > (C/4)n^{-1/2+\delta/3}\right]. \tag{60}$$

Substituting ε with $(C/4)n^{-1/2+\delta/3}$ and n with $n^{1-\delta/2}$ in Eq. (45) we have

$$\epsilon_{\rm tomo} = n^{-\Omega(n^{\delta/6})}. \tag{61}$$

Next, we look at the error of amplification. By Eqs. (39) and (50), the output of the amplifier can be expressed as

$$\begin{aligned}\mathcal{A}_{\theta_0}^{(n/\gamma_n')\to n}\left(G_{n/\gamma_n',\theta}\right) =& \mathcal{A}_{\rm c}\left(N\left(\delta\mu/\sqrt{\gamma_n'}, V_{\mu_0}\right)\right)\\ &\otimes \mathcal{A}_{\rm q}\left(\Phi\left(\delta\xi/\sqrt{\gamma_n'}, \mu_0\right)\right)\end{aligned} \tag{62}$$

where $(\delta\mu,\delta\xi) = \sqrt{n}(\theta-\theta_0)$. We then split the amplification error (56) into two terms: the term of the classical mode and the term of quantum mode amplification. We first analyze the classical mode, where the amplifier rescales the classical Gaussian distribution. The amplifier $\mathcal{A}_{\rm c}$ (49) amplifies the distribution by a factor of $\sqrt{\gamma_n'}$, shifts the center of the Gaussain distribution from $\delta\mu/\sqrt{\gamma_n'}$ to $\delta\mu$, and rescales the covariance matrix by a factor of γ_n', from V_{μ_0} to $\gamma_n' V_{\mu_0}$:

$$\mathcal{A}_{\rm c}\left(N\left(\delta\mu/\sqrt{\gamma_n'}, V_{\mu_0}\right)\right) = N\left(\delta\mu, \gamma_n' V_{\mu_0}\right). \tag{63}$$

The amplification error for the classical mode is:

$$\begin{aligned}\epsilon_{\rm classical} &= \|N\left(\delta\mu, \gamma_n' V_{\mu_0}\right) - N\left(\delta\mu, V_{\mu_0}\right)\|_1\\ &= \|N\left(0, \gamma_n' V_{\mu_0}\right) - N\left(0, V_{\mu_0}\right)\|_1.\end{aligned} \tag{64}$$

Writing explicitly the probability density functions of the Gaussian distributions, we have

$$\begin{aligned}\epsilon_{\rm classical} &= \int_{\mathbb{R}^d}\left|\frac{1}{\sqrt{2\pi|V_{\mu_0}|}}\exp\left(-\frac{\mathbf{x}^T V_{\mu_0}^{-1}\mathbf{x}}{2}\right)\right.\\ &\quad\left. - \frac{1}{\sqrt{2\pi\gamma_n'|V_{\mu_0}|}}\exp\left(-\frac{\mathbf{x}^T V_{\mu_0}^{-1}\mathbf{x}}{2\gamma_n'}\right)\right| d\mathbf{x}\\ &\le \frac{1}{\sqrt{2\pi|V_{\mu_0}|}}\int_{\mathbb{R}^d}\left|\exp\left(-\frac{\mathbf{x}^T V_{\mu_0}^{-1}\mathbf{x}}{2}\right)\right.\\ &\qquad\left. - \exp\left(-\frac{\mathbf{x}^T V_{\mu_0}^{-1}\mathbf{x}}{2\gamma_n'}\right)\right| d\mathbf{x}\\ &\quad + \frac{\left(1-(\gamma_n')^{-1/2}\right)}{\sqrt{2\pi|V_{\mu_0}|}}\int_{\mathbb{R}^d}\exp\left(-\frac{\mathbf{x}^T V_{\mu_0}^{-1}\mathbf{x}}{2\gamma_n'}\right) d\mathbf{x}\\ &\le \frac{\left(1-(\gamma_n')^{-1}\right)}{2\sqrt{2\pi|V_{\mu_0}|}}\int_{\mathbb{R}^d}\mathbf{x}^T V_{\mu_0}^{-1}\mathbf{x}\exp\left(-\frac{1}{2}\mathbf{x}^T V_{\mu_0}^{-1}\mathbf{x}\right) d\mathbf{x}\\ &\quad + O\left(n^{-\delta/2}\right)\\ &= O\left(n^{-\delta/2}\right).\end{aligned} \tag{65}$$

Note that $(1-(\gamma_n')^{-1})$ and $(1-(\gamma_n')^{-1/2})$ both have order $O(n^{-\delta/2})$.

Now we check the quantum term. On the quantum register, the amplifier acts independently on each mode as the displaced thermal state amplifier defined by Eq. (13). From a similar calculation as Eq. (25), we obtain the inequality

$$\epsilon_{\rm quantum} \le \frac{1}{2}\sum_{j<k}\left\|\mathcal{A}^{\gamma_n'}\left(\rho_{\alpha_{j,k},\beta_{j,k}}\right) - \rho_{\alpha_{j,k},\beta_{j,k}}\right\|_1, \tag{66}$$

where the error of each quantum amplifier is [cf. Eqs. (27) and (48)]

$$\frac{1}{2}\left\|\mathcal{A}^{\gamma_n'}\left(\rho_{\alpha_{j,k},\beta_{j,k}}\right) - \rho_{\alpha_{j,k},\beta_{j,k}}\right\|_1 = O\left(n^{-\delta/2}\right). \tag{67}$$

Therefore, we conclude that the amplification error (56) scales at most as

$$\epsilon_{\rm amp} \le \epsilon_{\rm classical} + \epsilon_{\rm quantum} = O\left(n^{-\delta/2}\right). \tag{68}$$

Let us now consider the error (57) of the Gaussian state compression, which can be upper bounded as

$$\begin{aligned}\epsilon_{\rm G} \le& \frac{1}{2}\left\|\mathcal{P}_{\rm c}\left[N(\delta\mu, V_{\mu_0})\right] - N\left(\delta\mu, V_{\mu_0}\right)\right\|_1\\ &+ \frac{1}{2}\sum_{j<k}\left\|\mathcal{P}_{n^\delta}\left(\rho_{\alpha_{j,k},\beta_{j,k}}\right) - \rho_{\alpha_{j,k},\beta_{j,k}}\right\|_1\end{aligned} \tag{69}$$

For the classical part, there are two sources of error: the error of rounding and the error of truncation. The former is simply $O(n^{-\delta/2})$, equal to the resolution of the rounding. The latter

can be bounded by noticing that $\|\delta\mu\|_\infty \leq n^{\delta/3}$, from which we have

$$\begin{aligned}&\left\|\mathcal{P}_{\rm c}\left[N(\delta\mu, V_{\mu_0})\right] - N\left(\delta\mu, V_{\mu_0}\right)\right\|_1 \\ &\leq \int_{\|u\|_\infty > n^{\delta/2}} N\left(\delta\mu, V_{\mu_0}\right)(\mathrm{d}u) \\ &\leq \int_{\|u-\delta\mu\|_\infty > n^{\delta/2}-n^{\delta/3}} N\left(\delta\mu, V_{\mu_0}\right)(\mathrm{d}u) \\ &= e^{-\Omega(n^\delta)}\end{aligned} \tag{70}$$

where $N\left(\delta\mu, V_{\mu_0}\right)(\mathrm{d}u)$ denotes the probability density function. For each of the quantum modes, employing Lemma 2, with K substituted by n^δ and $|\alpha_{j,k}| = O(n^{\delta/3})$, we have

$$\frac{1}{2}\left\|\mathcal{P}_{j,k}\left(\rho_{\alpha_{j,k},\beta_{j,k}}\right) - \rho_{\alpha_{j,k},\beta_{j,k}}\right\|_1 = \beta_{j,k}^{\Omega(n^{\delta/8})} + e^{-\Omega(n^{\delta/4})}. \tag{71}$$

Substituting Eqs. (70) and (71) into Eq. (69), we have

$$\epsilon_{\rm G} = \left(\max_{j<k}\beta_{j,k}\right)^{\Omega(n^{\delta/8})} + e^{-\Omega(n^{\delta/4})}. \tag{72}$$

Finally, we note that the error of the Q-LAN approximation, corresponding to the errors generated by the transformations between the input state and its Gaussian state approximation, is given by Eqs. (42) and (43) as $\epsilon_{\rm Q-LAN} = O\left(n^{-\kappa(2\delta/3)}\right)$. Since κ is non-increasing, the bound can be relaxed to

$$\epsilon_{\rm Q-LAN} = O\left(n^{-\kappa(\delta)}\right), \tag{73}$$

Summarizing the above bounds (61), (68), (72), (73) on each of the error terms, we conclude that the protocol generates an error which scales at most

$$\epsilon = O\left(n^{-\kappa(\delta)}\right) + O\left(n^{-\delta/2}\right). \tag{74}$$

C. Total memory cost.

The total memory cost consists of three contributions: a classical memory of $[(f_c+f_q)/2]\log n$ bits for the tomography outcome, a classical memory of $f_c\delta\log n$ bits for the classical part of the Gaussian state and a quantum memory of $f_q\delta\log n$ qubits for the quantum part of the Gaussian state. In short, it takes $(1/2+\delta)\log n$ bits to encode a classical independent parameter and $(1/2)\log n$ bits plus $\delta\log n$ qubits to encode a quantum independent parameter.

From the above discussion we can see that the ratio between the quantum memory cost and the classical memory cost is

$$R_{q/c} = \frac{\delta f_q}{(1/2+\delta)f_c + (1/2)f_q}, \tag{75}$$

which can be made close to zero by choosing δ close to zero. In conclusion, the size of the quantum memory can be made arbitrarily small compared to the classical memory.

V. Necessity of a quantum memory.

In the previous section, we showed that the ratio between the quantum and the classical memory cost can be made arbitrarily close to zero [see Eq. (75)]. It is then natural to ask whether the ratio can be exactly zero. The answer turns out to be negative. In fact, we prove an even stronger result: if a state family has at least one independent quantum parameter, then no protocol using a purely classical memory can be faithful, even if the amount of classical memory is arbitrarily large.

Theorem 2. *Let $\{\rho_\theta^{\otimes n}\}$ be a qudit state family with at least one independent quantum parameter, and let $(\mathcal{E}_n, \mathcal{D}_n)$ be generic compression protocol for $\{\rho_\theta^{\otimes n}\}$. If the protocol uses solely a classical memory, then the compression error will not vanish in the large n limit, no matter how large the memory is.*

The proof of Theorem 2 is based on the properties of two distance measures, known as the quantum Hellinger distance [31] (see also [32]) and the Bures distance [33], and defined as

$$d_{\rm H}(\rho_1,\rho_2) := \sqrt{2 - 2\,\mathrm{Tr}\left(\rho_1^{1/2}\rho_2^{1/2}\right)} \tag{76}$$

$$d_{\rm B}(\rho_1,\rho_2) := \sqrt{2 - 2\,\mathrm{Tr}\left|\rho_1^{1/2}\rho_2^{1/2}\right|}, \tag{77}$$

respectively.

The first property used in the proof of Theorem 2 is

Lemma 3. *For every pair of density matrices ρ_1 and ρ_2, one has the inequality $d_{\rm H}(\rho_1,\rho_2) \geq d_{\rm B}(\rho_1,\rho_2)$. The equality holds if and only if $[\rho_1,\rho_2]=0$.*

Proof. By definition, the condition $d_{\rm H}(\rho_1,\rho_2) \geq d_{\rm B}(\rho_1,\rho_2)$ is equivalent to the condition

$$\mathrm{Tr}\left|\rho_1^{1/2}\rho_2^{1/2}\right| \geq \mathrm{Tr}\left(\rho_1^{1/2}\rho_2^{1/2}\right).$$

The validity of this condition is immediate: for every square matrix A, one has $\mathrm{Tr}\,|A| \geq \mathrm{Tr}\,A$. The equality holds if and only if A is equal to $|A|$, meaning that A is positive semidefinite. For $A = \rho_1^{1/2}\rho_2^{1/2}$, the Hermiticity requirement $A = A^\dagger$ reads

$$\left(\rho_1^{1/2}\rho_2^{1/2}\right) = \left(\rho_1^{1/2}\rho_2^{1/2}\right)^\dagger = \left(\rho_2^{1/2}\rho_1^{1/2}\right),$$

which in turn is equivalent to the commutation relation $[\rho_1,\rho_2]=0$. □

The second property used in the proof of Theorem 2 is

Lemma 4. *Let $\mathcal{E}$ be a quantum channel sending states on $\mathcal{H}$ to states on $\mathcal{K}$ and let $\mathcal{D}$ be a quantum channel sending states on $\mathcal{K}$ to states on $\mathcal{H}$. Let ρ_1 and ρ_2 be two states on $\mathcal{H}$, satisfying the conditions*

$$[\mathcal{E}(\rho_1), \mathcal{E}(\rho_2)] = 0\,, \tag{78}$$

and $\frac{1}{2}\|\mathcal{D}\circ\mathcal{E}(\rho_i) - \rho_i\|_1 \leq \epsilon$ for $i \in \{1,2\}$. Then, the following inequality holds:

$$|d_{\rm H}(\rho_1,\rho_2) - d_{\rm B}(\rho_1,\rho_2)| \leq 2\sqrt{2\epsilon}. \tag{79}$$

Proof. Using the the triangle inequality for the quantum Hellinger distance, we obtain the upper bound

$$d_{\rm H}(\rho_1,\rho_2) \le d_{\rm H}\left(\rho_1, \mathcal{D}\circ\mathcal{E}(\rho_1)\right) + d_{\rm H}\left(\mathcal{D}\circ\mathcal{E}(\rho_1), \mathcal{D}\circ\mathcal{E}(\rho_2)\right) + d_{\rm H}\left(\mathcal{D}\circ\mathcal{E}(\rho_2), \rho_2\right) . \quad (80)$$

Now, for every pair of states ρ and σ, the quantum Hellinger distance and the trace distance are related by inequality $d_{\rm H}(\rho,\sigma) \le \sqrt{\|\rho-\sigma\|_1}$ [31]. Using this fact, the upper bound (80) becomes

$$d_{\rm H}(\rho_1,\rho_2) \le d_{\rm H}\left(\mathcal{D}\circ\mathcal{E}(\rho_1), \mathcal{D}\circ\mathcal{E}(\rho_2)\right) + 2\sqrt{2\epsilon}\,. \quad (81)$$

At this point, we use the fact that the quantum Hellinger (respectively, Bures) distance is non-increasing under the action of quantum channels [32] (respectively, [33]). In this way, we obtain the inequality

$$\begin{aligned} d_{\rm H}\left(\mathcal{D}\circ\mathcal{E}(\rho_1), \mathcal{D}\circ\mathcal{E}(\rho_2)\right) &\le d_{\rm H}\left(\mathcal{E}(\rho_1), \mathcal{E}(\rho_2)\right) \\ &= d_{\rm B}\left(\mathcal{E}(\rho_1), \mathcal{E}(\rho_2)\right) \\ &\le d_{\rm B}\left(\rho_1,\rho_2\right) , \end{aligned} \quad (82)$$

in which we used the relation

$$d_{\rm H}\left(\mathcal{E}(\rho_1), \mathcal{E}(\rho_2)\right) = d_{\rm B}\left(\mathcal{E}(\rho_1), \mathcal{E}(\rho_2)\right) , \quad (83)$$

following from Eq. (78) and Lemma 3. Combining Eqs. (81) and (82), we finally obtain the bound

$$d_{\rm H}(\rho_1,\rho_2) \le d_{\rm B}\left(\rho_1,\rho_2\right) + 2\sqrt{2\epsilon}\,. \quad (84)$$

Since the difference between the quantum Hellinger distance and the Bures distance is non-negative, the above inequality is exactly Eq. (79). □

Now we give the proof of Theorem 2.

Proof. Let $\{\rho_\theta^{\otimes n}\}$ be a qudit state family with at least one independent quantum parameter.

Pick two states $\rho_{\theta_0}^{\otimes n}$ and $\rho_\theta^{\otimes n}$, with θ of the form $\theta = \theta_0 + s/\sqrt{n}$ where all entries of the vector s are zero except for the independent quantum parameter. Applying Q-LAN to the neighborhood of θ_0, the two states $\rho_{\theta_0}^{\otimes n}$ and $\rho_\theta^{\otimes n}$ can be converted into two multi-mode Gaussian states G_{n,θ_0} and $G_{n,\theta}$ that differ from each other only in one mode. Explicitly, the two Gaussian states can be written as

$$G_{n,\theta_0} = G^{(-)}_{n,\theta_0} \otimes \rho_\beta^{\rm thm} \quad (85)$$

$$G_{n,\theta} = G^{(-)}_{n,\theta_0} \otimes \rho_{\alpha(s),\beta}\,. \quad (86)$$

where the thermal parameter β and the displacement $\alpha(s)$ are non-zero quantities depending only on s and θ_0 via Eq. (39), while $G^{(-)}_{n,\theta_0}$ is the state of the remaining modes.

Now, let $(\mathcal{E}_n, \mathcal{D}_n)$ be a compression protocol that uses a purely classical memory to compress the state family $\{\rho_\theta^{\otimes n}\}$. By Q-LAN, there exists a compression protocol $(\mathcal{E}'_n, \mathcal{D}'_n)$ that uses a purely classical memory to compress the states $\{G_{n,\theta_0}, G_{n,\theta}\}$. Explicitly, the encoder and the decoder are described by the channels

$$\mathcal{E}'_n := \mathcal{E}_n \circ \mathcal{S}^{(n)}_{\theta_0} \qquad \mathcal{D}'_n := \mathcal{T}^{(n)}_{\theta_0} \circ \mathcal{D}_n\,, \quad (87)$$

where $\mathcal{T}^{(n)}_{\theta_0}$ and $\mathcal{S}^{(n)}_{\theta_0}$ are the channels used for Q-LAN.

Since the protocol $(\mathcal{E}'_n, \mathcal{D}'_n)$ uses a purely classical memory, Lemma 4 implies the bound

$$\left| d_{\rm H}\left(G_{n,\theta_0}, G_{n,\theta}\right) - d_{\rm B}\left(G_{n,\theta_0}, G_{n,\theta}\right)\right| \le 2\sqrt{2\epsilon'_n}, \quad (88)$$

where ϵ'_n is the compression error for the states $\{G_{n,\theta_0}, G_{n,\theta}\}$.

On the other hand, the errors from Q-LAN vanish as $O\left(n^{-\kappa(x)}\right)$. Hence, we have the bound

$$\epsilon'_n \le \epsilon_n + O\left(n^{-\kappa(x)}\right) \quad (89)$$

Substituting Eqs. (85) and (86) into Eq. (88), we obtain the expression

$$\begin{aligned} \epsilon_n &\ge \epsilon'_n - O\left(n^{-\kappa(x)}\right) \\ &\ge \frac{1}{8}\left| d_{\rm H}\left(G_{n,\theta_0}, G_{n,\theta}\right) - d_{\rm B}\left(G_{n,\theta_0}, G_{n,\theta}\right)\right|^2 - O\left(n^{-\kappa(x)}\right) \\ &= \frac{1}{8}\left| d_{\rm H}\left(\rho_\beta^{\rm thm}, \rho_{\alpha(s),\beta}\right) - d_{\rm B}\left(\rho_\beta^{\rm thm}, \rho_{\alpha(s),\beta}\right)\right|^2 - O\left(n^{-\kappa(x)}\right). \end{aligned} \quad (90)$$

The quantum Hellinger and Bures distances between displaced thermal states can be computed using previous results. Using Eq. (16) of [34], we have

$$d_{\rm B}\left(\rho_\beta^{\rm thm}, \rho_{\alpha,\beta}\right) = \sqrt{2 - 2e^{-\frac{|\alpha|^2}{4\gamma_{\rm B}}}} \qquad \gamma_{\rm B} = \frac{1+\beta}{1-\beta}. \quad (91)$$

Using Eq. (3.18) of [35], we have

$$d_{\rm H}\left(\rho_\beta^{\rm thm}, \rho_{\alpha,\beta}\right) = \sqrt{2 - 2e^{-\frac{|\alpha|^2}{4\gamma_{\rm H}}}} \qquad \gamma_{\rm H} = \frac{(\sqrt{\beta}+1)^2}{2(1-\beta)}. \quad (92)$$

The difference between the two terms is strictly positive, except in the case when $\alpha = 0$. Therefore, the right hand side of Eq. (90) is strictly positive in the limit of $n \to \infty$, and thus $\lim_{n\to\infty} \epsilon_n > 0$. This concludes the proof. □

VI. Optimality of the compression

Here we prove that our compression protocol is asymptotically optimal in terms of total memory cost. Specifically, we show that every compression protocol with vanishing error must use an overall memory size of at least $(f/2)\log n$, where f is the number of independent parameters describing the input states.

The proof idea is to construct a communication protocol that transmits approximately $(f/2)\log n$ bits, using the compression protocol. Once this is done, the Holevo bound [19] implies that the overall amount of memory must be of at least $(f/2)\log n$ qubits/bits.

To construct the communication protocol, we define a mesh M_n on the parameter space Θ, by choosing a set of equally spaced points starting from a fixed point $\theta_0 \in \Theta$. Specifically, we define the mesh as

$$\mathsf{M}_n = \left\{\theta \in \Theta \mid |(\theta-\theta_0)_i| = z_i \cdot \log n/\sqrt{n}, z_i \in \mathbb{Z}\ \forall\ i\right\} . \quad (93)$$

The number of points in the mesh M_n satisfies the bound

$$|\mathsf{M}_{\mathsf{n}}| \ge T_\Theta \left[\frac{\sqrt{n}}{\log n}\right]^f , \quad (94)$$

where $T_\Theta > 0$ is a constant independent of n.

The next step is to define a finite set of states

$$\mathsf{S}_n = \left\{ \rho_\theta^{\otimes n} \mid \theta \in \mathsf{M}_n \right\} , \tag{95}$$

and to observe that they are almost perfectly distinguishable in the large n limit. One way to distinguish between the states in S_n is to use quantum tomography. Intuitively, since tomography provides an estimate of the state with error of order $1/\sqrt{n}$, the distance between two states in the set is large enough to make the states almost perfectly distinguishable. To make this argument rigorous, we describe the tomographic protocol using a POVM $P_n(\mathrm{d}\hat{\rho})$, where $\hat{\rho}$ is the estimate of the state. In particular, we use the POVM defined in Eq. (45), which has the property [30]

$$\int_{\|\hat{\rho}-\rho\|_1 > \varepsilon} \mathrm{Tr}[P_n(\mathrm{d}\hat{\rho})\, \rho^{\otimes n}] \leq (n+1)^{3d^2} e^{-n\varepsilon^2}. \tag{96}$$

The continuous POVM $P_n(\mathrm{d}\hat{\rho})$ can be used to distinguish between the states in the set S_n. To this purpose, we construct the discrete POVM with operators $\{Q_n(\theta)\}_{\theta\in\mathsf{M}_n} \cup \{Q_n(\mathrm{rest})\}$, where the operator $Q_n(\mathrm{rest})$ is defined as

$$Q_n(\mathrm{rest}) = I^{\otimes n} - \sum_{\theta\in\mathsf{M}_n} Q_n(\theta) \tag{97}$$

while the operator $Q_n(\theta)$ is defined as

$$Q_n(\theta) = \int_{\|\hat{\rho}-\rho_\theta\|_1 \leq \frac{\varepsilon_{\min}}{2}} P_n(\mathrm{d}\hat{\rho}) \,, \qquad \theta \in \mathsf{M}_n \,. \tag{98}$$

Here $\varepsilon_{\min}$ is the minimum distance between two distinct states in S_n, which can be quantified as

$$\varepsilon_{\min} = \min_{\theta,\theta'\in\mathsf{M}_n, \theta\neq\theta'} \frac{1}{2}\|\rho_\theta - \rho_{\theta'}\|_1 = \frac{C\log n}{2\sqrt{n}} + O\left(\frac{\log^2 n}{n}\right) , \tag{99}$$

having used the Euclidean expansion of trace distance, given by

$$\|\rho_\theta - \rho_{\theta'}\|_1 = C\|\theta - \theta'\|_\infty + O(\|\theta-\theta'\|_\infty^2) \,, \tag{100}$$

where $C > 0$ is a suitable constant. Hence, for any $\theta \in \mathsf{M}_n$ the probability of error for the state $\rho_\theta^{\otimes n}$ can be bounded as

$$\begin{aligned} P_{\mathrm{err},n}(\theta) &\leq \int_{\|\hat{\rho}-\rho_\theta\|_1 > \frac{\varepsilon_{\min}}{2}} \mathrm{Tr}\left[P_n(\mathrm{d}\hat{\rho})\rho_\theta^{\otimes n}\right] \\ &\leq (n+1)^{3d^2} e^{-\frac{C\log^2 n}{16}} \\ &= (n+1)^{3d^2} n^{-\frac{C\log n}{16\ln 2}} \\ &\leq n^{-\frac{C\log n}{16}}, \end{aligned} \tag{101}$$

where the last inequality holds for large enough n.

Using the results above, we can construct a communication protocol that communicates $(f/2)\log n$ bits given any compression protocol $(\mathcal{E}_n, \mathcal{D}_n)$. The protocol is defined as follows:

1) Both parties agree on a code that associates messages with points in the mesh M_n.
2) To communicate a certain message, the sender picks the corresponding point $\theta \in \mathsf{M}_n$ and prepares the state $\rho_\theta^{\otimes n}$.
3) The sender applies the encoder $\mathcal{E}_n$ and transmits $\mathcal{E}_n(\rho_\theta^{\otimes n})$ to the receiver.
4) The receiver applies the decoder $\mathcal{D}_n$.
5) The receiver measures the output state with the POVM $\{Q_n(\theta)\}_{\theta\in\mathsf{M}_n} \cup \{Q_n(\mathrm{rest})\}$.

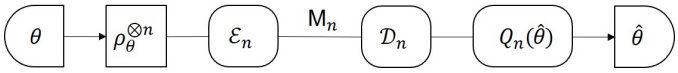


Fig. 4. **A protocol to communicate $\log|\mathsf{M}_n|$ bits of information.** Here $\mathcal{E}_n$ and $\mathcal{D}_n$ are the encoder and the decoder, $Q_n(\hat{\theta})$ is the POVM to recover the message, and M_n denotes the memory.

The protocol is illustrated in Fig. 4. A protocol can be constructed for displaced thermal states, following the steps 1) - 4) and replacing the POVM in step 5) of the above protocol by the heterodyne measurement of α and maximum likelihood estimation of β [27]. In this way, the proof here can be converted to a proof of optimality for displaced thermal states, which we omit for simplicity.

It is not hard to see that the protocol can communicate no less than $(f/2)\log n$ bits, with an error probability

$$P^*_{\mathrm{err},n} \leq P_{\mathrm{err},n} + \epsilon_n \tag{102}$$

where $P_{\mathrm{err},n} := \max_{\theta\in\mathsf{M}_n} P_{\mathrm{err},n}(\theta)$ and ϵ_n is the error of the compression protocol $(\mathcal{E}_n, \mathcal{D}_n)$. Consider the case when the messages are uniformly distributed. In this case, the number of transmitted bits can be bounded through Fano's inequality, which yields the bound

$$I(\Theta : \hat{\Theta}) \geq (1 - P^*_{\mathrm{err},n}) \log|\mathsf{M}_n| - h(P^*_{\mathrm{err},n}) \,, \tag{103}$$

with $h(x) = -x\log x$ and $h(0) := 0$. When the compression protocol has vanishing error, i.e. $\lim_{n\to\infty}\epsilon_n = 0$, using Eqs. (94), (101), and (102) we obtain the lower bound

$$I(\Theta : \hat{\Theta}) \geq \frac{f}{2}\log n - f\log\log n + o(1) \,. \tag{104}$$

Using the monotonicity of mutual information and the upper bound of entropy, the total number n_{enc} of memory bits/qubits is lower bounded as

$$\begin{aligned} n_{\mathrm{enc}} &\geq H\left[\mathcal{E}_n\left(|\mathsf{M}_n|^{-1}\sum_{\theta\in\mathsf{M}_n}\rho_\theta^{\otimes n}\right)\right] \\ &\geq I(\Theta : \mathsf{M}_n) \\ &\geq I(\Theta : \hat{\Theta}). \end{aligned} \tag{105}$$

Combining Eq. (105) with Eq. (104), we obtain that n_{enc} must be at least

$$n_{\mathrm{enc}} \geq \frac{f}{2}\log n - f\log\log n + o(1) \,. \tag{106}$$

This proves that $f/2\log n$ bits/qubits are necessary to achieve compression with vanishing error.

VII. Conclusion

In this work we addressed the problem of compressing identically prepared states of finite-dimensional quantum systems and identically prepared displaced thermal states. We showed that the total size of the required memory is approximately $(f/2)\log n$, where f is the number of independent parameters

of the state and n is the number of input copies. Moreover, we observed that the asymptotic ratio between the amount of quantum bits and the amount of classical bits can be set to an arbitrarily small constant. Still, a fully classical memory cannot faithfully encode genuine quantum states: only states that are jointly diagonal in fixed basis can be compressed into a purely classical memory.

A natural development of our work is the study of compression protocols for quantum population coding beyond the case of identically prepared states. Motivated by the existing literature on classical population coding [8], the idea is to consider families of states representing a population of quantum particles. At the most fundamental level, the indistinguishability of quantum particles leads to the Bose-Einstein [36], [37] and Fermi-Dirac statistics [38], [39], as well as to other intermediate statistics [40], [41]. Since the Hilbert space describing identical quantum particles is not the tensor product of single-particle Hilbert spaces, the compression of quantum populations of indistinguishable particles requires a non-trivial extension of our results. The optimal compression protocols are likely to shed light on how information is encoded into a broad range of real physical systems. In addition, the compression protocols will offer a tool to simulate large numbers of particles using quantum computers of relatively smaller size.

From the point of view of quantum simulations, it is also meaningful to consider the compression of tensor network states [42], [43] which provide a variational ansatz for a large number of quantum manybody systems. The extension of quantum compression to this scenario is appealing as a technique to reduce the number of qubits needed to simulate systems of distinguishable quantum particles, in the same spirit of the compressed simulations introduced by Kraus [44] and coauthors [45], [46]. In the long term, the information-theoretic study of manybody quantum systems may provide a new approach to the simulation of complex systems that are not efficiently simulatable on classical computers.

Acknowledgments

This work is supported by the Canadian Institute for Advanced Research (CIFAR), by the Hong Kong Research Grant Council through Grant No. 17326616, by National Science Foundation of China through Grant No. 11675136, and by the HKU Seed Funding for Basic Research, and by the Foundational Questions Institute through grant FQXi-RFP3-1325. Y. Y. is supported by a Hong Kong and China Gas Scholarship. MH was supported in part by a MEXT Grant-in-Aid for Scientific Research (B) No. 16KT0017, the Okawa Research Grant and Kayamori Foundation of Informational Science Advancement.

REFERENCES

[1] K. Banaszek, M. Cramer, and D. Gross, "Focus on quantum tomography," *New Journal of Physics*, vol. 15, no. 12, p. 125020, 2013.

[2] V. Scarani, S. Iblisdir, N. Gisin, and A. Acín, "Quantum cloning," *Reviews of Modern Physics*, vol. 77, pp. 1225–1256, Nov 2005. [Online]. Available: http://link.aps.org/doi/10.1103/RevModPhys.77.1225

[3] N. J. Cerf, A. Ipe, and X. Rottenberg, "Cloning of continuous quantum variables," *Physical Review Letters*, vol. 85, no. 8, p. 1754, 2000.

[4] S. M. Barnett and S. Croke, "Quantum state discrimination," *Advances in Optics and Photonics*, vol. 1, no. 2, pp. 238–278, 2009.

[5] B. Schumacher, "Quantum coding," *Physical Review A*, vol. 51, no. 4, p. 2738, 1995.

[6] R. Jozsa and B. Schumacher, "A new proof of the quantum noiseless coding theorem," *Journal of Modern Optics*, vol. 41, no. 12, pp. 2343–2349, 1994.

[7] H.-K. Lo, "Quantum coding theorem for mixed states," *Optics Communications*, vol. 119, no. 5-6, pp. 552–556, 1995.

[8] S. Wu, S.-i. Amari, and H. Nakahara, "Population coding and decoding in a neural field: a computational study," *Neural Computation*, vol. 14, no. 5, pp. 999–1026, 2002.

[9] M. Plesch and V. Bužek, "Efficient compression of quantum information," *Physical Review A*, vol. 81, no. 3, p. 032317, 2010.

[10] L. A. Rozema, D. H. Mahler, A. Hayat, P. S. Turner, and A. M. Steinberg, "Quantum data compression of a qubit ensemble," *Physical Review Letters*, vol. 113, p. 160504, Oct 2014. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRevLett.113.160504

[11] Y. Yang, G. Chiribella, and D. Ebler, "Efficient quantum compression for ensembles of identically prepared mixed states," *Physical Review Letters*, vol. 116, p. 080501, Feb 2016. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRevLett.116.080501

[12] Y. Yang, G. Chiribella, and M. Hayashi, "Optimal compression for identically prepared qubit states," *Physical Review Letters*, vol. 117, p. 090502, Aug 2016. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRevLett.117.090502

[13] M. Hayashi and V. Tan, "Minimum rates of approximate sufficient statistics," *IEEE Transactions on Information Theory, in press. arXiv: 1612.02542*, 2016.

[14] M. Hayashi and K. Matsumoto, "Asymptotic performance of optimal state estimation in qubit system," *Journal of Mathematical Physics*, vol. 49, no. 10, p. 102101, 2008.

[15] M. Guţă and J. Kahn, "Local asymptotic normality for qubit states," *Physical Review A*, vol. 73, no. 5, p. 052108, 2006.

[16] M. Guţă and A. Jenčová, "Local asymptotic normality in quantum statistics," *Communications in Mathematical Physics*, vol. 276, no. 2, pp. 341–379, 2007.

[17] J. Kahn, "Quantum local asymptotic normality and other questions of quantum statistics," Theses, Université Paris XI, Sep. 2009. [Online]. Available: https://hal.inria.fr/tel-01657373

[18] J. Kahn and M. Guţă, "Local asymptotic normality for finite dimensional quantum systems," *Communications in Mathematical Physics*, vol. 289, no. 2, pp. 597–652, 2009.

[19] A. S. Holevo, "Bounds for the quantity of information transmitted by a quantum communication channel," *Problemy Peredachi Informatsii*, vol. 9, no. 3, pp. 3–11, 1973.

[20] Y. Yang, G. Chiribella, and G. Adesso, "Certifying quantumness: Benchmarks for the optimal processing of generalized coherent and squeezed states," *Physical Review A*, vol. 90, no. 4, p. 042319, 2014.

[21] A. Vourdas, "Superposition of squeezed coherent states with thermal light," *Physical Review A*, vol. 34, no. 4, p. 3466, 1986.

[22] P. Marian and T. A. Marian, "Squeezed states with thermal noise. i. photon-number statistics," *Physical Review A*, vol. 47, no. 5, p. 4474, 1993.

[23] B. Saleh, *Photoelectron statistics: with applications to spectroscopy and optical communication*. Springer, 2013, vol. 6.

[24] R. J. Glauber, "Coherent and incoherent states of the radiation field," *Physical Review*, vol. 131, no. 6, p. 2766, 1963.

[25] P. Busch, M. Grabowski, and P. J. Lahti, *Operational quantum physics*. Springer Science & Business Media, 1997, vol. 31.

[26] C. M. Caves, "Quantum limits on noise in linear amplifiers," *Physical Review D*, vol. 26, no. 8, p. 1817, 1982.

[27] M. Hayashi, "Asymptotic quantum estimation theory for the thermal states family," in *Quantum Communication, Computing, and Measurement 2*. Springer, 2002, pp. 99–104.

[28] W. Kumagai and M. Hayashi, "Quantum hypothesis testing for gaussian states: Quantum analogues of χ^2, t-, and f-tests," *Communications in Mathematical Physics*, vol. 318, no. 2, pp. 535–574, 2013.

[29] A. W. Van der Vaart, *Asymptotic statistics*. Cambridge university press, 2000, vol. 3.

[30] J. Haah, A. W. Harrow, Z. Ji, X. Wu, and N. Yu, "Sample-optimal tomography of quantum states," in *Proceedings of the Forty-eighth Annual ACM Symposium on Theory of Computing*, ser. STOC '16. New York, NY, USA: ACM, 2016, pp. 913–925. [Online]. Available: http://doi.acm.org/10.1145/2897518.2897585

[31] A. S. Holevo, "On quasiequivalence of locally normal states," *Theoretical and Mathematical Physics*, vol. 13, no. 2, pp. 1071–1082, 1972.

[32] S. Luo and Q. Zhang, "Informational distance on quantum-state space," *Physical Review A*, vol. 69, no. 3, p. 032106, 2004.
[33] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information*. Cambridge University Press, 2000.
[34] H. Scutaru, "Fidelity for displaced squeezed thermal states and the oscillator semigroup," *Journal of Physics A: Mathematical and General*, vol. 31, no. 15, p. 3659, 1998.
[35] P. Marian and T. A. Marian, "Hellinger distance as a measure of gaussian discord," *Journal of Physics A: Mathematical and Theoretical*, vol. 48, no. 11, p. 115301, 2015.
[36] S. N. Bose, "Plancks gesetz und lichtquantenhypothese," *Z. phys*, vol. 26, no. 3, p. 178, 1924.
[37] A. Einstein, *Quantentheorie des einatomigen idealen Gases*. Akademie der Wissenshaften, in Kommission bei W. de Gruyter, 1924.
[38] E. Fermi, "Eine statistische methode zur bestimmung einiger eigenschaften des atoms und ihre anwendung auf die theorie des periodischen systems der elemente," *Zeitschrift für Physik*, vol. 48, no. 1-2, pp. 73–79, 1928.
[39] P. A. Dirac, "The quantum theory of the electron," in *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 117, no. 778. The Royal Society, 1928, pp. 610–624.
[40] J. M. Leinaas and J. Myrheim, "On the theory of identical particles," *Il Nuovo Cimento B (1971-1996)*, vol. 37, no. 1, pp. 1–23, 1977.
[41] F. Wilczek, "Quantum mechanics of fractional-spin particles," *Physical review letters*, vol. 49, no. 14, p. 957, 1982.
[42] N. Schuch, D. Pérez-García, and I. Cirac, "Classifying quantum phases using matrix product states and projected entangled pair states," *Physical Review B*, vol. 84, no. 16, p. 165139, 2011.
[43] N. Schuch, M. M. Wolf, F. Verstraete, and J. I. Cirac, "Entropy scaling and simulability by matrix product states," *Physical Review Letters*, vol. 100, no. 3, p. 030504, 2008.
[44] B. Kraus, "Compressed quantum simulation of the ising model," *Physical review letters*, vol. 107, no. 25, p. 250503, 2011.
[45] W. L. Boyajian, V. Murg, and B. Kraus, "Compressed simulation of evolutions of the x y model," *Physical Review A*, vol. 88, no. 5, p. 052329, 2013.
[46] W. L. Boyajian and B. Kraus, "Compressed simulation of thermal and excited states of the one-dimensional x y model," *Physical Review A*, vol. 92, no. 3, p. 032323, 2015.
[47] A. Winter, "Coding theorem and strong converse for quantum channels," *IEEE Transactions on Information Theory*, vol. 45, no. 7, p. 2481, 1999.
[48] M. Wilde, "The quantum data-compression theorem," in *Quantum information theory*. Cambridge University Press, 2013, ch. 18.
[49] F. De Oliveira, M. Kim, P. L. Knight, and V. Bužek, "Properties of displaced number states," *Physical Review A*, vol. 41, no. 5, p. 2645, 1990.

APPENDIX

A. Proof of Lemma 1.

To compress the n-fold thermal state $(\rho_\beta^{\rm thm})^{\otimes n}$ into classical memory, we can first do measurements on the state, and then encode the outcome using the smallest possible classical memory. The state $\rho_\beta^{\rm thm}$ is fully described by the thermal parameter β, or equivalently, the average photon number $\beta/(1-\beta)$. An estimate of $\beta/(1-\beta)$ can be obtained by measuring the photon number of the n modes independently and by computing the sum. The sum can be encoded into a classical memory. To this purpose, we divide the set of nonnegative integers into intervals, and only store the index of the interval the sum lies in.

For any $\delta > 0$, we define a series of $t+1 = \lfloor n^{1/2+\delta} \rfloor$ intervals $\mathsf{L}_0, \dots \mathsf{L}_t$ as

$$\begin{aligned}
&\mathsf{L}_0 = \{0\} \\
&\mathsf{L}_i = \{(i-1)\lfloor n^{(1-\delta)/2}\rfloor + 1, \dots, i\lfloor n^{(1-\delta)/2}\rfloor\} \quad 0 < i < t \\
&\mathsf{L}_t = \{(t-1)\lfloor n^{(1-\delta)/2}\rfloor + 1, \dots\}.
\end{aligned}$$

For any non-negative integer m, we denote by $i(m)$ the index of the interval containing m, i.e. $m \in \mathsf{L}_{i(m)}$.

To design the compression protocol, we notice that the n-fold thermal state can be written in the form

$$(\rho_\beta^{\rm thm})^{\otimes n} = (1-\beta)^n \sum_{\vec{m}} \beta^{|\vec{m}|} |\vec{m}\rangle\langle\vec{m}|, \tag{107}$$

where $|\vec{m}\rangle = |m_1\rangle \otimes \cdots \otimes |m_n\rangle$ is the photon number basis of n modes and $|\vec{m}| := m_1 + \cdots + m_n$. The compression protocol runs as follows:

- *Encoder.* First perform projective measurement in the photon number basis of n modes, which yields an n-dimensional vector $\vec{m}$. Then compute $i(|\vec{m}|)$ and encode it into a classical memory. The encoding channel can be represented as

$$\mathcal{E}_{n,\delta}^{\rm thm}(\rho) := \sum_{\vec{m}} \langle\vec{m}|\rho|\vec{m}\rangle \ |i(|\vec{m}|)\rangle\langle i(|\vec{m}|)|.$$

- *Decoder.* Read the integer i from the memory. If $i \geq t$ prepare a fixed state $|\vec{t}\rangle\langle\vec{t}|$ (defined below); if $i < t$ perform random sampling in the interval L_i. For each outcome $\hat{m}$ of the sampling, prepare the n-mode state

$$\binom{n+\hat{m}-1}{\hat{m}}^{-1} \sum_{\vec{m}:|\vec{m}|=\hat{m}} |\vec{m}\rangle\langle\vec{m}|.$$

Then the decoding channel can be represented as

$$\mathcal{D}_{n,\delta}^{\rm thm}(|i\rangle\langle i|) := \begin{cases} \sum_{\vec{m}:|\vec{m}|\in\mathsf{L}_i} \frac{|\vec{m}\rangle\langle\vec{m}|}{|\mathsf{L}_i|\binom{n+|\vec{m}|-1}{|\vec{m}|}} & i < t \\ |\vec{t}\rangle\langle\vec{t}| & i = t \end{cases}$$
$$\vec{t} = ((t-1)\lfloor n^{(1-\delta)/2}\rfloor + 1, \dots, (t-1)\lfloor n^{(1-\delta)/2}\rfloor + 1).$$

It is straightforward from definition that the protocol uses $\log(t+1) = (1/2+\delta)\log n + o(1)$ classical bits. What remains is to bound the error of the protocol. First, we notice that the recovered state is

$$\mathcal{D}_{n,\delta}^{\rm thm} \circ \mathcal{E}_{n,\delta}^{\rm thm}\left[(\rho_\beta^{\rm thm})^{\otimes n}\right] = \sum_{\vec{m}} (\rho_{\beta,n})_{\vec{m}} \ |\vec{m}\rangle\langle\vec{m}| \tag{108}$$

$$(\rho_{\beta,n})_{\vec{m}} = \begin{cases} (1-\beta)^n \beta^{|\vec{m}|} \sum_{m\in\mathsf{L}_{i(|\vec{m}|)}} \frac{\beta^{m-|\vec{m}|}\binom{n+m-1}{m}}{|\mathsf{L}_{i(|\vec{m}|)}|\binom{n+|\vec{m}|-1}{|\vec{m}|}} & i(|\vec{m}|) < t \\ \sum_{\vec{m}':i(|\vec{m}'|)=t} (1-\beta)^n \beta^{|\vec{m}'|} & \vec{m} = \vec{t} \\ 0 & \text{else.} \end{cases} \tag{109}$$

We choose S as the minimal set satisfying $i)$ S is a union of several intervals chosen from the set $\{\mathsf{L}_i\}$ and ii)

$$\mathsf{S} \supset \left[\frac{\beta n}{1-\beta} - n^{(1-\delta)/2}, \frac{\beta n}{1-\beta} + n^{(1-\delta)/2}\right]. \tag{110}$$

Apparently, $\mathsf{S} \subset [0, n^{1+\delta/2}]$ for large enough n. Then the error can be bounded as

$$\begin{aligned}
\epsilon_{\rm thm} &= \frac{1}{2}\left\| \mathcal{D}^{\rm thm}_{n,\delta} \circ \mathcal{E}^{\rm thm}_{n,\delta} \left[(\rho^{\rm thm}_\beta)^{\otimes n} \right] - (\rho^{\rm thm}_\beta)^{\otimes n} \right\|_1 \\
&= \frac{1}{2} \sum_{\vec{m}} \left| (1-\beta)^n \beta^{|\vec{m}|} - (\rho_{\beta,n})_{\vec{m}} \right| \\
&\le \sum_{\vec{m}:|\vec{m}|\notin \mathsf{S}} (1-\beta)^n \beta^{|\vec{m}|} + \frac{1}{2} \left[\sum_{\vec{m}:|\vec{m}|\in \mathsf{S}} (1-\beta)^n \beta^{|\vec{m}|} \right] \\
&\qquad \times \max_{\substack{m' \in \mathsf{L}_i \\ \mathsf{L}_i \cap \mathsf{S} \neq \emptyset}} \left| \sum_{m \in \mathsf{L}_i} \frac{\beta^m \binom{n+m-1}{m}}{|\mathsf{L}_i| \beta^{m'} \binom{n+m'-1}{m'}} - 1 \right| \\
&\qquad + \sum_{\vec{m}':i(|\vec{m}'|)=t} (1-\beta)^n \beta^{|\vec{m}'|} \\
&\le \sum_{\vec{m}:|\vec{m}|\notin \mathsf{S}} (1-\beta)^n \beta^{|\vec{m}|} + \frac{1}{2} \left[\sum_{\vec{m}:|\vec{m}|\in \mathsf{S}} (1-\beta)^n \beta^{|\vec{m}|} \right] \\
&\qquad \times \max_{\substack{m,\, m' \in \mathsf{L}_i \\ \mathsf{L}_i \cap \mathsf{S} \neq \emptyset}} \left| \frac{\beta^m \binom{n+m-1}{m}}{\beta^{m'} \binom{n+m'-1}{m'}} - 1 \right| \\
&\qquad + O\left((1-\beta)^n \beta^n \right) \\
&\le \sum_{\vec{m}:|\vec{m}|\notin \mathsf{S}} (1-\beta)^n \beta^{|\vec{m}|} \\
&\qquad + \frac{1}{2} \max_{\substack{m,\, m' \in \mathsf{L}_i \\ \mathsf{L}_i \cap \mathsf{S} \neq \emptyset}} \left| \frac{\beta^m \binom{n+m-1}{m}}{\beta^{m'} \binom{n+m'-1}{m'}} - 1 \right| \\
&\qquad + O\left((1-\beta)^n \beta^n \right).
\end{aligned}$$

On one hand, we notice that $|\vec{m}|$ is the sum of n i.i.d. random variables with geometric distribution $\{(1-\beta)\beta^i\}_{i=0}^\infty$ and thus, by Central Limit Theorem, the first error term scales as

$$\sum_{\vec{m}:|\vec{m}|\notin \mathsf{S}} (1-\beta)^n \beta^{|\vec{m}|} = O\left[\operatorname{erfc}\left(\frac{n^{\delta/2}(1-\beta)}{\sqrt{2\beta}} \right) \right] = e^{-\Omega(n^\delta)}$$

where $\operatorname{erfc}(\delta) := (2/\pi) \int_\delta^\infty e^{-s^2}\, {\rm d}s$ is the complementary error function. On the other hand, in the second error term, m and m' are in the same order as n, so the second term can be bounded as

$$\begin{aligned}
&\max_{\substack{m,\, m' \in \mathsf{L}_i \\ \mathsf{L}_i \cap \mathsf{S} \neq \emptyset}} \left| \frac{\beta^m \binom{n+m-1}{m}}{\beta^{m'} \binom{n+m'-1}{m'}} - 1 \right| \\
&\quad = \max_{\substack{m,\, m' \in \mathsf{L}_i \\ \mathsf{L}_i \cap \mathsf{S} \neq \emptyset}} \left| \beta^{m-m'} \frac{(n+m-1)\cdots(n+m')}{m \ldots (m'+1)} - 1 \right| \\
&= \max_{\substack{m,\, m' \in \mathsf{L}_i \\ \mathsf{L}_i \cap \mathsf{S} \neq \emptyset}} \left| \left(\frac{\beta m + \beta n}{m} \right)^{m-m'} \left[1 + O\left(\frac{|\mathsf{L}_i|^2}{n} \right) \right] - 1 \right| \\
&= \left| \left[1 + O\left(n^{-\delta}\right) \right] \left[1 + O\left(n^{-\delta}\right) \right] - 1 \right| \\
&= O\left(n^{-\delta}\right).
\end{aligned}$$

Therefore, we have proved Eq. (14).

B. Derivation of Eq. (25)

Eq. (25) is a standard result in quantum optics. Here we provide its derivation for the benefit of those readers who may be less familiar with this area.

Note that the amplifier of Eq. (13) can be represented as

$$\mathcal{A}^\gamma(\rho) = {\rm Tr}_B \left[S_{\cosh^{-1}(\sqrt{\gamma})} (\rho \otimes |0\rangle\langle 0|_B) S^\dagger_{\cosh^{-1}(\sqrt{\gamma})} \right] \quad (111)$$

with $S_r := e^{r(\hat{a}^\dagger \hat{b}^\dagger - \hat{a}\hat{b})}$. The unitary S_r satisfies $S_r \hat{a} S_r^\dagger = (\cosh r)\hat{a} - (\sinh r)\hat{b}^\dagger$ (cf. Eq. (B8) of [**?**]), which immediately implies the relation $S_r D_\alpha = D_{(\cosh r)\alpha} S_r$. Hence, we have

$$\mathcal{A}^\gamma \circ \mathcal{D}_\alpha = \mathcal{D}_{\sqrt{\gamma}\alpha} \circ \mathcal{A}^\gamma \quad (112)$$

for any $\alpha \in \mathbb{C}$ and $\gamma \ge 1$. In particular, when the amplifier is applied to displaced thermal states, one has the relation

$$\mathcal{A}^\gamma(\rho_{\alpha,\beta}) = \mathcal{D}_{\sqrt{\gamma}\alpha} \circ \mathcal{A}^\gamma(\rho^{\rm thm}_\beta). \quad (113)$$

To prove Eq. (25), it only remains to show the identity $\mathcal{A}^\gamma(\rho^{\rm thm}_\beta) = \rho^{\rm thm}_{\beta'}$ with β' as in Eq. (25). This equality, which is standard in quantum optics, can be proven by observing that every thermal state can be generated from the vacuum through the action of the Gaussian additive noise channel $\mathcal{N}_x$, defined as

$$\mathcal{N}_x(\rho) = \int {\rm d}^2\mu \left(\frac{x}{\pi} \right) e^{-x|\mu|^2} D_\mu \, \rho D_\mu^\dagger. \quad (114)$$

Specifically, it is easy to verify the relation

$$\rho^{\rm thm}_\beta = \mathcal{N}_{\frac{1-\beta}{\beta}}(|0\rangle\langle 0|), \quad (115)$$

valid for every $\beta \in (0,1]$. Then, using Eq. (113), we have

$$\mathcal{A}^\gamma(\rho^{\rm thm}_\beta) = \mathcal{A}^\gamma \circ \mathcal{N}_{\frac{1-\beta}{\beta}}(|0\rangle\langle 0|) = \mathcal{N}_{\frac{1-\beta}{\beta\gamma}} \circ \mathcal{A}^\gamma(|0\rangle\langle 0|). \quad (116)$$

It is straightforward to verify that $\mathcal{A}^\gamma(|0\rangle\langle 0|)$ is a thermal state; specifically, one has

$$\mathcal{N}_{\frac{1}{\gamma-1}}(|0\rangle\langle 0|). \quad (117)$$

Hence, we have

$$\begin{aligned}
\mathcal{A}^\gamma(\rho^{\rm thm}_\beta) &= \mathcal{N}_{\frac{1-\beta}{\beta\gamma}} \circ \mathcal{N}_{\frac{1}{\gamma-1}}(|0\rangle\langle 0|) \\
&= \mathcal{N}_{\frac{1-\beta}{\gamma-1+\beta}}(|0\rangle\langle 0|) \\
&= \rho^{\rm thm}_{\beta'} \qquad \beta' = \frac{\beta+\gamma-1}{\gamma}, \quad (118)
\end{aligned}$$

the second equality following from the composition property of the additive Gaussian noise

$$\mathcal{N}_x \circ \mathcal{N}_y = \mathcal{N}_{\frac{xy}{x+y}} \qquad \forall\, x, y > 0.$$

Indeed, the above equation can be derived using the definition of the additive Gaussian noise (114):

$$\begin{aligned}
\mathcal{N}_x \circ \mathcal{N}_y(\rho) &= \int {\rm d}^2\mu \int {\rm d}^2\nu \left(\frac{xy}{\pi^2} \right) e^{-x|\mu|^2 - y|\nu|^2} D_{\mu+\nu} \, \rho D^\dagger_{\mu+\nu} \\
&= \int {\rm d}^2\mu \int {\rm d}^2\alpha \left(\frac{xy}{\pi^2} \right) e^{-x|\mu|^2} \cdot e^{-y|\alpha-\mu|^2} D_\alpha \, \rho D_\alpha^\dagger \\
&\qquad \alpha := \mu + \nu \\
&= \int {\rm d}^2\mu \left(\frac{x+y}{\pi} \right) e^{-(x+y)\left|\mu - \frac{y}{x+y}\alpha\right|^2} \\
&\qquad \times \int {\rm d}^2\alpha \left[\frac{xy}{(x+y)\pi} \right] e^{-\frac{xy}{x+y}|\alpha|^2} D_\alpha \, \rho D_\alpha^\dagger \\
&= \mathcal{N}_{\frac{xy}{x+y}}(\rho).
\end{aligned}$$

Combining Eqs. (113) and (118) gives Eq. (25) as desired.

C. Proof of Eq. (26)

The distance between

$$\left\|\rho_{\beta'}^{\rm thm}-\rho_{\beta}^{\rm thm}\right\|_1 \le \frac{2|\beta'-\beta|}{(1-\beta')^2}+O(|\beta'-\beta|^2)\,.$$

By definition

$$\begin{aligned}\left\|\rho_{\beta'}^{\rm thm}-\rho_{\beta}^{\rm thm}\right\|_1 &= \sum_{j=0}^{\infty}|(1-\beta)\beta^j-(1-\beta')(\beta')^j| \\ &\le \sum_{j=0}^{\infty}\left[|\beta^j-(\beta')^j|+|\beta^{j+1}-(\beta')^{j+1}|\right] \\ &= 2\sum_{j=0}^{\infty}|\beta^j-(\beta')^j|+O(|\beta-\beta'|^2).\end{aligned}$$

Extracting $|\beta-\beta'|$ from the first term on the r.h.s. of the last equality and focusing on the first order, we get

$$\begin{aligned}\left\|\rho_{\beta'}^{\rm thm}-\rho_{\beta}^{\rm thm}\right\|_1 &\le 2|\beta-\beta'|\sum_{j=0}^{\infty}(j+1)\beta^j+O(|\beta-\beta'|^2) \\ &= \frac{2|\beta-\beta'|^2}{(1-\beta)^2}+O(|\beta-\beta'|^2)\end{aligned}$$

as desired.

D. Proof of Lemma 2.

For any input state ρ we have

$$\begin{aligned}\epsilon(\rho) &\le \frac{1}{2}\left\|P_K\rho P_K-\rho\right\|_1+\frac{1}{2}[1-{\rm Tr}(\rho P_K)] \\ &\le \frac{1}{2}\left[2\sqrt{[1-{\rm Tr}(\rho P_K)]}+1-{\rm Tr}(\rho P_K)\right] \\ &\le \frac{3}{2}\sqrt{1-{\rm Tr}(\rho P_K)}.\end{aligned}$$

The second inequality came from the gentle measurement lemma [47], [48]. Substituting $\rho_{\alpha,\beta}=D_\alpha\rho_\beta^{\rm thm}D_\alpha^\dagger$ into the above bound, we express the truncation error for $\rho_{\alpha,\beta}$ as

$$\epsilon(\rho_{\alpha,\beta})\le\frac{3}{2}\sqrt{1-{\rm Tr}\left[D_\alpha\rho_\beta^{\rm thm}D_\alpha^\dagger P_K\right]}. \qquad (119)$$

We now bound the trace part in the right hand side of the last inequality as

$$\begin{aligned}1-{\rm Tr}\left[D_\alpha\rho_\beta^{\rm thm}D_\alpha^\dagger P_\alpha\right] &= \sum_{k\le K}\sum_{l=0}^{\infty}(1-\beta)\beta^l|\langle l|D_\alpha|k\rangle|^2 \\ &\le \max_{l\le l_0}\sum_{k>K}|\langle l|D_\alpha|k\rangle|^2+\sum_{l\ge l_0}^{\infty}(1-\beta)\beta^l \\ &\le \max_{l\le l_0}\sum_{k>K}|\langle l|D_\alpha|k\rangle|^2+\beta^{l_0}\end{aligned} \qquad (120)$$

Here we set $l_0=K^{x/8}$. Notice that $|\langle l|D_\alpha|k\rangle|^2$ is the photon number distribution of a displaced number state [49], which can be bounded as

$$\begin{aligned}|\langle l|D_\alpha|k\rangle|^2 &= \frac{e^{-|\alpha|^2}(|\alpha|^2)^{k+l}}{k!l!}\left|\sum_{i=0}^{\min\{k,l\}}\frac{k!l!(-|\alpha|^2)^{-i}}{i!(k-i)!(l-i)!}\right|^2 \\ &\le \frac{e^{-|\alpha|^2}(|\alpha|^2)^{k+l}}{k!l!}\left|\sum_{i=0}^{k}\binom{k}{i}\left(\frac{l}{|\alpha|^2}\right)^i\right|^2 \\ &= \frac{e^{-|\alpha|^2}|\alpha|^{2l}}{k!l!}\left(\frac{l+|\alpha|^2}{|\alpha|}\right)^{2k}.\end{aligned}$$

Then we can bound the first term in (120) as

$$\begin{aligned}\max_{l\le l_0}\sum_{k>K}|\langle l|D_\alpha|k\rangle|^2 &\le \max_{l\le l_0}\frac{|\alpha|^{2l}}{l!}\sum_{k>K}\frac{e^{-|\alpha|^2}}{k!}\left(\frac{l+|\alpha|^2}{|\alpha|}\right)^{2k} \\ &= \max_{l\le l_0}\frac{e^{2l+\frac{l^2}{|\alpha|^2}}|\alpha|^{2l}}{l!}\sum_{k>K}\mathbf{Pois}_{\lambda_\alpha}(k),\end{aligned}$$

where $\mathbf{Pois}_{\lambda_\alpha}(k)$ is the Poisson distribution with mean $\lambda_\alpha=(l+|\alpha|^2)^2/|\alpha|^2$. Notice that $[\lambda_\alpha-K^{1/2+x/2},\lambda_\alpha+K^{1/2+x/2}]\subseteq[0,K]$ and thus we have

$$\begin{aligned}\max_{l\le l_0}\sum_{k>K}|\langle l|D_\alpha|k\rangle|^2 &\le \max_{l\le l_0}\frac{|\alpha|^{2l}e^{2l+\frac{l^2}{|\alpha|^2}}}{l!}\sum_{|k-\lambda_\alpha|>K^{1/2+x/2}}\mathbf{Pois}_{\lambda_\alpha}(k) \\ &= \max_{l\le|\alpha|^{x/4}}\frac{|\alpha|^{2l}e^{2l+\frac{l^2}{|\alpha|^2}}}{l!}e^{-\Omega(K^{x/2})} \\ &= e^{-\Omega(K^{x/4})}.\end{aligned}$$

having used the tail bound for Poisson distribution and $l_0=K^{x/8}$. Substituting the above bound into Eq. (120), we get

$$\sqrt{1-{\rm Tr}\left[D_\alpha\rho_\beta^{\rm thm}D_\alpha^\dagger P_\alpha\right]}\le\beta^{\Omega(K^{x/8})}+e^{-\Omega(K^{x/4})}.$$

Finally, substituting the above inequality into Eq. (119), we can bound the error as

$$\epsilon(\rho_{\alpha,\beta})=\beta^{\Omega(K^{x/8})}+e^{-\Omega(K^{x/4})}.$$