Quantum-inspired algorithm for estimating the permanent of positive semidefinite matrices

L. Chakhmakhchyan, N. J. Cerf, and R. Garcia-Patron
Centre for Quantum Information and Communication, Ecole polytechnique de Bruxelles, CP 165, Université libre de Bruxelles, 1050 Brussels, Belgium
(Received 24 September 2016; published 31 August 2017)

We construct a quantum-inspired classical algorithm for computing the permanent of Hermitian positive semidefinite matrices by exploiting a connection between these mathematical structures and the boson sampling model. Specifically, the permanent of a Hermitian positive semidefinite matrix can be expressed in terms of the expected value of a random variable, which stands for a specific photon-counting probability when measuring a linear-optically evolved random multimode coherent state. Our algorithm then approximates the matrix permanent from the corresponding sample mean and is shown to run in polynomial time for various sets of Hermitian positive semidefinite matrices, achieving a precision that improves over known techniques. This work illustrates how quantum optics may benefit algorithm development.

DOI: 10.1103/PhysRevA.96.022329

I. INTRODUCTION

Linear quantum optics, which deals with the scattering of photons in a linear interferometer, is a promising candidate for the implementation of universal quantum computing [1]. This capability, first proved by the seminal scheme proposed by Knill, Laflamme, and Milburn, makes use of passive linear-optics components, single-photon sources and detectors, as well as adaptive measurements [2]. Recent considerable advances in the fabrication of reconfigurable optical circuits, single-photon sources, and measurement devices of increasing reliability [3,4] will make possible the design of fully integrated quantum optical circuits of tens to hundreds of qubits in the near future. However, due to the requirement of measurement-induced circuit control and the need to bring in ancillary photonic modes, among others, the current proof-of-principle realizations of this universal optical scheme are still far from a level where quantum advantage could be demonstrated. Surprisingly, as shown by Aaronson and Arkhipov, these highly demanding requirements can be relaxed within the boson sampling paradigm, while the resulting problem remains intractable for a classical computer [5]. Namely, boson sampling is a task that consists of sampling from the probability distribution of detecting identical single photons at the output of a linear-optical circuit. Despite the seeming simplicity of this task, the photon-counting probabilities are proportional to the squared modulus of permanents of complex matrices [6], whose exact computation—and even approximate estimation—is, in general, an intractable (#P-hard) task [7,8]. This observation, along with several plausible complexity assumptions, is at the heart of the hardness proof of boson sampling. For this reason, boson sampling is now viewed as a very promising model to establish the advantage of a quantum computer over its classical counterpart, which has motivated a series of proof-of-principle experimental works [9].

Another striking feature of linear quantum optics, which we demonstrate in this paper, is its ability to inspire the construction of *efficient classical* algorithms. Specifically, building on the model of boson sampling with thermal states [10], we propose an algorithm for estimating the permanent of Hermitian positive semidefinite matrices (HPSMs). The algorithm exploits the optical equivalence theorem for multimode thermal states at the input of a linear-optical circuit. Thermal states can be represented as a geometric distribution over Fock states, connecting the permanent of HPSMs with the single-photon measurement probabilities on the given multimode thermal state, evolved through an optical circuit [10]. At the same time, thermal states are also represented as Gaussian mixtures of coherent states [11], an observation that allows us to construct an algorithm for approximating the permanent of HPSMs. Namely, up to a constant factor, our algorithm outputs the average (over a sample of polynomial size) of the probability of detecting the same single-photon pattern upon sending a Gaussian-distributed coherent state in a linear-optical circuit.

The problem of approximating the permanent of HPSMs to within a multiplicative error has recently attracted the attention of computer scientists [12]. In this regard, nothing is known as of today about the existence of polynomial-time techniques for approximating permanents of HPSMs with bounded relative errors [12]; apart from that, this can be achieved within the third level of the polynomial hierarchy (thus implying that the problem is not #P-hard, unless the polynomial hierarchy collapses to its third level or beyond) [10,12]. Further, we believe that, with the exception of Gurvits' seminal algorithm [13], which approximates the permanent of any given $M \times M$ matrix with an additive error, no algorithm especially tailored to approximate the permanent of HPSMs has previously been developed. In this paper, we present such an algorithm, which substantially improves over Gurvits' technique in terms of additive errors and even achieves "almost relative" error (i.e., proportional to the square root of the permanent itself) for some restricted set of HPSMs. Although our scheme does not resolve the open complexity-theoretic problem of approximating the permanent of HPSMs to within a multiplicative error [12], it further illustrates the potential of intertwining quantum optics with computer science. On a different note, let us point out that an algorithm for estimating permanents of HPSMs may find applications in the assessment of boson sampling itself, in terms of generalized bunching [14]. Namely, the latter allows one to efficiently certify that a physical device realizing boson sampling operates in the regime of full quantum coherence.

The rest of the paper is organized as follows. In the next section we recall the boson sampling model with thermal states

and provide the basis for our algorithm. In Sec. III we present the algorithm for approximating the permanent of Hermitian positive semidefinite matrices. Finally, in Sec. IV we draw our conclusions.

II. BOSON SAMPLING WITH THERMAL STATES

As input state, we consider a multimode thermal state $\rho_{\rm in}^{\rm th} = \bigotimes_{i=1}^{M} \rho_i^{\rm th}$. Each state $\rho_i^{\rm th}$ is characterized in terms of its average photon number $\langle n_i \rangle$ and can be expressed as an incoherent mixture of Fock states weighted by a geometric distribution with parameter $\tau_i = \langle n_i \rangle / (\langle n_i \rangle + 1)$; that is, $\rho_i^{\rm th} = (1-\tau_i) \sum_{n=0}^{\infty} \tau_i^n |n\rangle\langle n|$.

The set of thermal states is injected into an M-mode linear-optical network described by means of an $M \times M$ unitary matrix $\mathbf{U}$ that transforms the input-mode operators $a_i^\dagger$ onto the output-mode operators $b_i^\dagger$ (see also Fig. 1):

$$b_i^\dagger = \sum_{j=1}^{M} \mathbf{U}_{ij}\, a_j^\dagger. \tag{1}$$

Exploiting the natural homomorphism between the $M \times M$ unitary matrix $\mathbf{U}$ and the corresponding unitary transformation $\mathcal{U}$ in state space [8], the linear-optical evolution of $\rho_{\rm in}^{\rm th}$ is given by $\rho_{\rm out} = \mathcal{U} \rho_{\rm in}^{\rm th} \mathcal{U}^\dagger$, and the joint probability of detecting m_i photons on the ith output-mode reads

$$p^{\rm th}(\mathbf{m}) = {\rm Tr}[\rho_{\rm out} |\mathbf{m}\rangle\langle\mathbf{m}|], \tag{2}$$

where $|\mathbf{m}\rangle$ stands for a product of Fock states with $\mathbf{m} \equiv \{m_1, \ldots, m_M\}$. The boson sampling problem with thermal states consists of sampling from the probability distribution defined by Eq. (2), given the set of input states $\rho_i^{\rm th}$ and the transformation $\mathbf{U}$ [10].

We now restrict our analysis to a single element of the probability distribution $p^{\rm th}(\mathbf{m})$ in Eq. (2), corresponding to the detection of a single photon in each output mode (i.e., $m_i = 1$, $\forall\, i$), which reads (see Ref. [10] and the Appendix C)

$$p^{\rm th} \equiv p^{\rm th}(1, \ldots, 1) = \frac{{\rm Per}\, \mathbf{A}}{\prod_{i=1}^{M} (1 + \langle n_i \rangle)}, \tag{3}$$

where

$$\mathbf{A} = \mathbf{U} \mathbf{D} \mathbf{U}^\dagger, \tag{4}$$

$$\mathbf{D} = {\rm diag}\{\tau_1, \ldots, \tau_M\}, \tag{5}$$

and the eigenvalues satisfy

$$1 > \tau_i = \langle n_i \rangle / (\langle n_i \rangle + 1) \geqslant 0, \quad \forall\, i. \tag{6}$$

This connection between the probability of detecting a pattern of M single photons at the output of the linear-optical evolution of an M-mode thermal state $\rho_{\rm in}^{\rm th}$ and the permanent of an $M \times M$ Hermitian positive semidefinite matrix $\mathbf{A}$ of bounded eigenvalues ($\tau_i < 1$) is one of the two main ingredients in the construction of our algorithm. It is also worth pointing out that the unitary that diagonalizes $\mathbf{A}$ is precisely the one that describes the circuit itself, while the spectrum of $\mathbf{A}$ is determined in terms of the average photon numbers $\langle n_i \rangle$ of the input thermal states.

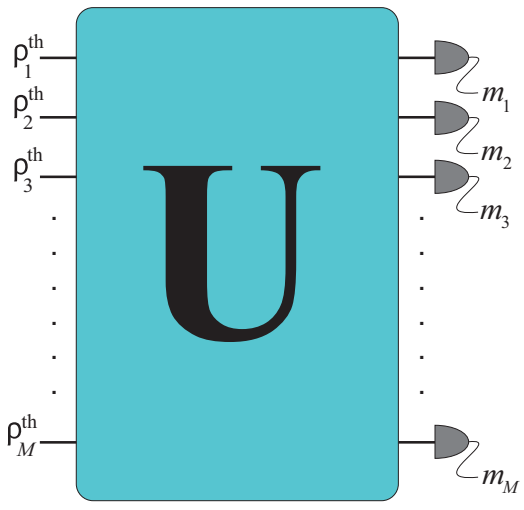

FIG. 1. Boson sampling setup with an M-mode thermal state $\rho_{\rm in}^{\rm th} = \bigotimes_{i=1}^{M} \rho_i^{\rm th}$ injected into an M-mode linear-optical network that is characterized by the unitary matrix $\mathbf{U}$ (acting on bosonic-mode operators), followed by the detection of a pattern of photons $\{m_1, \ldots, m_M\}$. The problem is to sample from the probability distribution given by Eq. (2).

III. ALGORITHM FOR ESTIMATING PERMANENTS OF HERMITIAN POSITIVE SEMIDEFINITE MATRICES

A. The main intuition behind our algorithm

The next key ingredient in the construction of our algorithm is to exploit the Glauber-Sudarshan P representation [11] to write down the M-mode input thermal state $\rho_{\rm in}^{\rm th}$ as a mixture of M-mode coherent states $|\boldsymbol{\alpha}\rangle \equiv |\alpha_1, \ldots, \alpha_M\rangle = \bigotimes_{i=1}^{M} |\alpha_i\rangle$ according to a Gaussian distribution

$$\rho_{\rm in}^{\rm th} = \int_{\mathbb{C}^M} \prod_{i=1}^{M} \left[\frac{d^2\alpha_i}{\pi \langle n_i \rangle} \exp\left(-\frac{|\alpha_i|^2}{\langle n_i \rangle} \right) \right] |\boldsymbol{\alpha}\rangle\langle\boldsymbol{\alpha}|. \tag{7}$$

Consequently, one can express the linear-optical evolution of the input state in terms of that of their component coherent states $|\boldsymbol{\alpha}\rangle$. Namely, $\mathcal{U}$ transforms a tensor product of coherent states $\bigotimes_{i=1}^{M} |\alpha_i\rangle$ into another tensor product of coherent states $\bigotimes_{i=1}^{M} |\beta_i\rangle$ with amplitudes

$$\beta_i = \sum_{j=1}^{M} \mathbf{U}_{ji} \alpha_j. \tag{8}$$

In other words, coherent states remain in a tensor product form while evolved through a linear-optical circuit. Thus, the joint probability $p^{\rm cs}(\boldsymbol{\alpha}) \equiv p^{\rm cs}(\alpha_1, \ldots, \alpha_M)$ of detecting a single photon at each output mode, with an M-mode coherent state $|\boldsymbol{\alpha}\rangle$ at the input, admits a simple product form,

$$p^{\rm cs}(\boldsymbol{\alpha}) = \prod_{i=1}^{M} e^{-|\beta_i|^2} |\beta_i|^2, \tag{9}$$

where the dependence on α_i's is implicit via Eq. (8). As a consequence, the probability $p^{\rm th}$ of Eq. (3) is alternatively

represented as

$$p^{\rm th} = \int_{\mathbb{C}^M} \prod_{i=1}^{M} \frac{d^2\alpha_i}{\pi \langle n_i \rangle} \exp\left(-\frac{|\alpha_i|^2}{\langle n_i \rangle}\right) p^{\rm cs}(\boldsymbol{\alpha}). \tag{10}$$

Therefore, we end up with the expected value of the function $p^{\rm cs}(\boldsymbol{\alpha})$ of random variables $\alpha_1, \ldots, \alpha_M$, which results from integrating $p^{\rm cs}(\boldsymbol{\alpha})$ over a complex Gaussian-distributed probability measure over the complex random variables $\alpha_i \in \mathcal{N}_{\mathbb{C}}(0, \langle n_i \rangle)$ [or for the sake of shortness, $\boldsymbol{\alpha} \in \mathcal{N}$, with $\mathcal{N}$ denoting the set $\{\alpha_i \in \mathcal{N}_{\mathbb{C}}(0, \langle n_i \rangle), i = 1, \ldots, M\}$], i.e., $p^{\rm th} = \mathbb{E}[p^{\rm cs}(\boldsymbol{\alpha})]_{\boldsymbol{\alpha} \in \mathcal{N}}$. Finally, exploiting the connection between $p^{\rm th}$ and Per $\mathbf{A}$ in Eq. (3), we rewrite the permanent of any HPSM $\mathbf{A}$ satisfying Eqs. (4)–(6) as

$$\mathrm{Per}\mathbf{A} = \frac{1}{\prod_{i=1}^{M}(1-\tau_i)} \mathbb{E}[p^{\rm cs}(\boldsymbol{\alpha})]_{\boldsymbol{\alpha} \in \mathcal{N}}. \tag{11}$$

A sampling algorithm approximating $\mathbb{E}[p^{\rm cs}(\boldsymbol{\alpha})]_{\boldsymbol{\alpha} \in \mathcal{N}}$ will thus immediately translate into an algorithm which, according to the law of large numbers, converges to Per $\mathbf{A}$.

B. The algorithm

In order to approximate the permanent of an arbitrary HPSM matrix Λ, we first diagonalize $\Lambda = \mathbf{U}\mathbf{D}\mathbf{U}^{\dagger}$, where the unitary matrix $\mathbf{U}$ encodes the eigenvectors and $\mathbf{D} = \mathrm{diag}(\lambda_1, \ldots, \lambda_M)$, the spectrum of $\mathbf{\Lambda}$. If the largest eigenvalue does not satisfy $\lambda_{\max} < 1$, we rescale the matrix, $\mathbf{A} = \mathbf{\Lambda}/(C\lambda_{\max})$, reducing the problem to finding the permanent of $\mathbf{A}$, knowing that $\mathrm{Per}\,\mathbf{\Lambda} = (C\lambda_{\max})^M \mathrm{Per}\,\mathbf{A}$. Here $C > 1$ is a specific constant that is necessary to avoid the divergence of the highest $\langle n_i \rangle = \tau_i/(1-\tau_i)$, while it also provides a certain tunability for the algorithm. Consequently, we can write $\mathrm{Per}\,\mathbf{\Lambda} = Z\mathbb{E}[p^{\rm cs}(\boldsymbol{\alpha})]_{\boldsymbol{\alpha} \in \mathcal{N}}$, with

$$Z = \frac{(C\lambda_{\max})^{2M}}{\prod_{i=1}^{M}(C\lambda_{\max} - \lambda_i)}, \tag{12}$$

where we expressed the parameters τ_i in terms of the eigenvalues of $\mathbf{\Lambda}$. The second step of the algorithm then consists in approximating $\mathbb{E}[p^{\rm cs}(\boldsymbol{\alpha})]_{\boldsymbol{\alpha} \in \mathcal{N}}$ as follows.

(i) For the selected error ϵ in the approximation of Per $\mathbf{\Lambda}$ and for a probability δ of failure of the algorithm, calculate the needed number of samples N, using Eq. (15) below.

(ii) Generate N samples $\boldsymbol{\alpha}^{(j)}$ $(j = 1, \ldots, N)$ of the random string $\boldsymbol{\alpha} = \{\alpha_1, \ldots, \alpha_M\}$, where each $\alpha_i^{(j)}$ is drawn at random from the Gaussian distribution $\mathcal{N}_{\mathbb{C}}(0, \langle n_i \rangle)$.

(iii) For each string $\boldsymbol{\alpha}^{(j)}$, by means of Eqs. (8) and (9), calculate $p^{\rm cs}[\boldsymbol{\alpha}^{(j)}]$.

(iv) Calculate the sample mean

$$\mu = \sum_{j=1}^{N} p^{\rm cs}[\boldsymbol{\alpha}^{(j)}]/N. \tag{13}$$

(v) Finally, output $Z\mu$.

Our algorithm involves several computational steps whose running time is polynomial in M. First, computing the eigenvalues and eigenvectors of the $M \times M$ HPSM can be done in time $O(M^3)$ with the traditional QR or divide-and-conquer algorithms [15]. Next, for each sample of $p^{\rm cs}(\boldsymbol{\alpha})$, one has to generate a random string $\boldsymbol{\alpha} = \{\alpha_1, \ldots, \alpha_M\}$ with $\alpha_i \in \mathcal{N}_{\mathbb{C}}(0, \langle n_i \rangle)$, for which efficient sampling techniques from Gaussian distributions are available [16]. Afterwards, one multiplies the matrix $\mathbf{U}$ with the column of α_i's yielding the amplitudes β_i's in time $O(M^2)$. If one generates N samples of $p^{\rm cs}(\boldsymbol{\alpha})$, then the overall running time of the algorithm scales as $O(M^2[M+N])$.

C. Error analysis and scaling of the algorithm

In order to determine the efficiency of our algorithm, as well as the scaling of its running time with respect to the matrix size M, we have to estimate the number of samples N, needed to reach a given precision in the approximation of $\mathrm{Per}\mathbf{\Lambda} = Z\mathbb{E}[p^{\rm cs}(\boldsymbol{\alpha})]_{\boldsymbol{\alpha} \in \mathcal{N}}$. Note that as far as relative-error analysis is concerned, the prefactor Z in front of $\mathbb{E}[p^{\rm cs}(\boldsymbol{\alpha})]_{\boldsymbol{\alpha} \in \mathcal{N}}$ [Eq. (12)] is irrelevant. In contrast, for the additive-error approximation, it should be carefully taken into account (see the Appendix A).

In order to estimate the sample size N needed, we make use of the Hoeffding inequality [17,18]. The latter provides an upper bound for the probability of the sample mean μ to deviate from the expected value $\mathbb{E}[p^{\rm cs}(\boldsymbol{\alpha})]_{\boldsymbol{\alpha} \in \mathcal{N}}$, given the sample size N and the constraint $0 \leqslant p^{\rm cs}(\boldsymbol{\alpha}) \leqslant e^{-M}$. Translated into the estimate $Z\mu$ of Per$\mathbf{\Lambda}$, Hoeffding inequality yields

$$\Pr(|\mathrm{Per}\,\mathbf{\Lambda} - Z\mu| \geqslant Z\epsilon) \leqslant \exp\left(-\frac{2N\epsilon^2}{e^{-2M}}\right). \tag{14}$$

Denoting by δ the failure probability of the algorithm, we find the sample size N that results in an error ϵ for Per $\mathbf{\Lambda}$,

$$N = \frac{Z^2 e^{-2M}}{2\epsilon^2} \ln\frac{1}{\delta}. \tag{15}$$

For the algorithm running time to scale polynomially, the sample size N should stay polynomial in the matrix size M, imposing conditions on the spectra of the HPSM, as we detail below.

To our knowledge, the only algorithm capable of approximating the permanent of HPSMs is Gurvits' algorithm, which is defined for general complex matrices. Gurvits' algorithm exploits the fact that the permanent of any matrix $\mathbf{X}$ can be written as the expected value of efficiently computable bounded random variables [19]. This makes it possible to approximate Per$\mathbf{X}$ in terms of the corresponding sample mean in time $O(M^2/\varepsilon^2)$, yielding an additive error $\pm\varepsilon\|\mathbf{X}\|^M$, where $\|\mathbf{X}\|$ denotes the trace norm of $\mathbf{X}$, which reads $\pm\varepsilon\lambda_{\max}^M$ for HPSMs. As we detail in the Appendix A, analyzing the additive error of our scheme, we find a set $(S1)$ of cases where our technique outperforms Gurvits' algorithm for HPSMs. Namely, we find a specific regime where the additive error of our algorithm decreases exponentially faster than that of Gurvits, at the price of a small polynomial overhead. More precisely, we achieve the error $\pm\varepsilon l^M \lambda_{\max}^M$ $(l \leqslant 1)$ for the set of matrices such that their spectra satisfy the following necessary and sufficient condition

$$\sqrt[M]{\prod_{i=1}^{M}\left(1 - \frac{\lambda_i}{C\lambda_{\max}}\right)} \geqslant \frac{C}{e}. \tag{16}$$

For yet another set $(S2)$ of HPSMs, satisfying similar constraints but also the condition $\lambda_{\max} > 1$, our scheme yields an additive error decreasing exponentially with M, where the

Gurvits' algorithm fails to do so. Interestingly, as we detail in the Appendix B, our derivation yields as a corollary an upper bound for the permanent of HPSMs in $S2$. It implies an exponential decrease of the permanent with M, where Glynn's formula fails to do so for matrices satisfying $\lambda_{\max} > 1$, as Glynn's formula leads to the upper bound $\mathrm{Per}\mathbf{\Lambda} \leqslant \lambda_{\max}^{M}$ [20].

Finally, we are also able to achieve an "almost-relative" error $\pm\varepsilon\sqrt{\mathrm{Per}\mathbf{\Lambda}}$ for a different restricted class ($S3$) of HPSMs. The corresponding condition relies again on the spectral properties of the matrix $\mathbf{\Lambda}$.

IV. CONCLUSION

We have presented a quantum-inspired algorithm, which exploits tools from quantum optics to address a classical computational problem: estimating the permanent of Hermitian positive semidefinite matrices. By use of a Monte Carlo-type technique, the permanent is approximated as the expected value of a random variable, up to a prefactor that depends only on the spectrum of the matrix. Interestingly, this random variable finds a natural physical interpretation as it stands for the joint probability of detecting a single photon at the output of a specific linear-optical circuit, injected with an M-mode coherent state of normally distributed random amplitudes. Additionally, the unitary defining the circuit is the one that diagonalizes the given Hermitian positive semidefinite matrix, and the eigenvalues are connected to the variance of the normal distribution of the M-mode coherent state.

The error analysis shows, for a specific set of Hermitian positive semidefinite matrices, that our polynomial-time algorithm yields better additive errors than Gurvits' technique. Moreover, for a restricted class of Hermitian positive semidefinite matrices, we are even able to achieve an "almost-relative" error, proportional to the square root of the permanent itself. We believe that the necessary conditions developed in the Appendix A indicate that these restricted sets of matrices do not reduce to computationally trivial classes (with respect to the permanent computation), but a full analysis should be carried out in order to confirm it. Whether these restrictions should be viewed as a caveat of the proposed algorithm is left for future work, but we stress that it is especially tailored to approximate the permanent of Hermitian positive semidefinite matrices.

We hope that this work will motivate further investigation to develop a multiplicative-error approximation algorithm of the permanent of Hermitian positive semidefinite matrices, a complexity-theoretic question that remains open. We also believe that our work highlights the benefits that exploiting the connection between the theory of computer science and quantum optics could bring to both communities.

ACKNOWLEDGMENTS

We thank Anthony Leverrier for useful discussions and comments. This work was supported by H2020-FETPROACT-2014 Grant QUCHIP (Quantum Simulation on a Photonic Chip; Grant Agreement No. 641039), http://www.quchip.eu). R.G.-P. acknowledges financial support as a research associate of the *Fonds de la Recherche Scientifique* (F.R.S.-FNRS, http://www.fnrs.be).

APPENDIX A: EFFICIENT REGIMES AND ERROR ANALYSIS OF THE ALGORITHM APPROXIMATING THE PERMANENT OF HERMITIAN POSITIVE SEMIDEFINITE MATRICES

In this section we detail the efficient regimes of the proposed algorithm and estimate its failure probability. As already mentioned in the main text, the running time of our scheme strongly depends on the sample size N, which approximates the expected value $\mathbb{E}[p^{\mathrm{cs}}(\boldsymbol{\alpha})]_{\boldsymbol{\alpha}\in\mathcal{N}}$ and thus the permanent of a given HPSM. In order to estimate N we make use of the Hoeffding inequality. It is applicable in our case, since the random variable $p^{\mathrm{cs}}(\boldsymbol{\alpha})$ is bounded. Namely, due to its definition $p^{\mathrm{cs}}(\boldsymbol{\alpha}) \leqslant \prod_{i=1}^{M} e^{-1} = e^{-M}$, which means that the random variable $p^{\mathrm{cs}}(\boldsymbol{\alpha})$ (and thus its expected value) lies within the interval $[0, e^{-M}]$. Thus, the Hoeffding inequality provides an upper bound for the probability of the approximant sample mean μ to be far from the expected value $\mathbb{E}[p^{\mathrm{cs}}(\boldsymbol{\alpha})]_{\boldsymbol{\alpha}\in\mathcal{N}}$, given the sample N and the fact that $0 \leqslant p^{\mathrm{cs}}(\boldsymbol{\alpha}) \leqslant e^{-M}$:

$$\Pr\{|\mathbb{E}[p^{\mathrm{cs}}(\boldsymbol{\alpha})]_{\boldsymbol{\alpha}\in\mathcal{N}} - \mu| \geqslant \epsilon\} \leqslant \exp\left(-\frac{2N\epsilon^2}{e^{-2M}}\right). \tag{A1}$$

Therefore, given the failure probability of the algorithm δ, from the above equation we find that the sample size N, which results in an error ϵ, is

$$N = \frac{e^{-2M}}{2\epsilon^2} \ln\frac{1}{\delta}. \tag{A2}$$

We also restate the relation between the expected value $\mathbb{E}[p^{\mathrm{cs}}(\boldsymbol{\alpha})]_{\boldsymbol{\alpha}\in\mathcal{N}}$ and the permanent of any given HPSM $\mathbf{\Lambda}$, in a form more suitable for the further analysis,

$$\mathrm{Per}\mathbf{\Lambda} = Z\mathbb{E}[p^{\mathrm{cs}}(\boldsymbol{\alpha})]_{\boldsymbol{\alpha}\in\mathcal{N}}, \tag{A3}$$

with

$$Z = \frac{C^{2M}\lambda_{\max}^{2M}}{a^M}, \tag{A4}$$

$$a = \sqrt[M]{\prod_{i=1}^{M}(C\lambda_{\max} - \lambda_i)}. \tag{A5}$$

In the above equation a is the geometric mean of the quantities $\{C\lambda_{\max} - \lambda_1, \ldots, C\lambda_{\max} - \lambda_M\}$, which, combined with the inequality of arithmetic and geometric means, satisfies

$$\lambda_{\max}(C-1) \leqslant a \leqslant C\lambda_{\max} - \bar{\lambda}, \tag{A6}$$

where

$$\bar{\lambda} = \frac{1}{M}\sum_{i=1}^{M}\lambda_i \tag{A7}$$

is the eigenvalue mean.

It is important to note here that since the permanent of the HPSM $\mathbf{\Lambda}$ is equal to the expected value of $p^{\mathrm{cs}}(\boldsymbol{\alpha})$ times the constant Z, the approximation of $\mathbb{E}[p^{\mathrm{cs}}(\boldsymbol{\alpha})]_{\boldsymbol{\alpha}\in\mathcal{N}}$ to within an additive error ϵ results in an error $Z\epsilon$ for $\mathrm{Per}\mathbf{\Lambda}$. We proceed now with the analysis of several regimes of our algorithm approximating $\mathrm{Per}\mathbf{\Lambda}$ and yielding distinct types of additive errors.

1. Additive error beating Gurvits' algorithm (set *S*1 of the main text)

First, we compare the error results provided by our algorithm to that of the Gurvits' one. The latter, having running time $O(M^2/\varepsilon^2)$, estimates the permanent of any $M\times M$ matrix $\mathbf{X}$, to within an additive error $\varepsilon\|\mathbf{X}\|^M$ ($\|\mathbf{X}\|$ is the spectral norm of $\mathbf{X}$). As for HPSMs $\|\mathbf{\Lambda}\| = \lambda_{\max}$, Gurvits' additive error reads $\varepsilon\lambda_{\max}^M$.

Consequently, the requirement that our algorithm results in an exponentially smaller error than that of Gurvits' (with a polynomial overhead), i.e., an additive error $\varepsilon(l\lambda_{\max})^M$ for the permanent approximation, imposes us to set $\epsilon = \varepsilon l^M \lambda_{\max}^M/Z$ in Eq. (A2),

$$N = \frac{1}{2\varepsilon^2}\ln\frac{1}{\delta}\left(\frac{\lambda_{\max}C^2}{lea}\right)^{2M}, \tag{A8}$$

together with the additional constraint $l \leqslant 1$. Now, our aim is to approximate the permanent of the matrix $\mathbf{\Lambda}$ to within a minimal additive error within this regime. Therefore, in order to avoid the exponential increase of the sample size N we impose the conditions

$$l \leqslant 1, \quad \frac{\lambda_{\max}C^2}{ea} \leqslant l, \tag{A9}$$

leading to the inequality

$$a \geqslant \frac{\lambda_{\max}C^2}{e}. \tag{A10}$$

This inequality defines a condition on the spectra of the HPSMs that is necessary and sufficient to guarantee the efficiency of our approximation algorithm [for achieving the additive error $\varepsilon(l\lambda_{\max})^M$, exponentially smaller in M than that of Gurvits' algorithm]. It is easy to see that conditions (A9) can be recovered from (A10) by setting $l = \lambda_{\max}C^2/(ea)$.

Meanwhile, combining the upper bound of a in Eq. (A6) with Eq. (A10), we find the necessary condition

$$\bar{\lambda} \leqslant \lambda_{\max}C\left(1-\frac{C}{e}\right) \leqslant \frac{e}{4} \approx 0.680, \tag{A11}$$

which provides some intuition on the regime of parameters where our algorithm improves over Gurvits' algorithm. Furthermore, since $\bar{\lambda} \geqslant 0$, we obtain the second necessary condition

$$C \leqslant e. \tag{A12}$$

It is also possible to bound l from below,

$$l \geqslant \frac{C^2\lambda_{\max}}{e(C\lambda_{\max}-\bar{\lambda})} \geqslant \frac{1}{e}, \tag{A13}$$

giving information on the minimal additive error that our scheme potentially provides within the regime discussed in this section.

Finally, it is worth noting that if the maximal eigenvalue of the given HPSM $\mathbf{\Lambda}$ is smaller than 1, $\lambda_{\max} < 1$, one does not require the rescaling of $\mathbf{\Lambda}$. In other words, there is no necessity to divide it by $C\lambda_{\max}$. This effectively corresponds to setting $C\lambda_{\max} = 1$ (or replacing C with $1/\lambda_{\max}$). Therefore, the conditions (A10)–(A12) can be readily applied for HPSMs with $\lambda_{\max} < 1$, by simply replacing C with $1/\lambda_{\max}$ [and doing that in the definition (A5) as well]. Note that when $\lambda_{\max} < 1$, the additive error $\varepsilon\lambda_{\max}$ of the Gurvits' algorithm itself is exponentially decreasing in M.

2. Exponentially decreasing additive error: $\lambda_{\max} \geqslant 1$ (set *S*2 of the main text)

In this section we show that our algorithm is capable of providing additive error results well beyond that of the Gurvits' scheme. Namely, we aim at achieving an additive error that decreases exponentially in M for HPSMs with $\lambda_{\max} \geqslant 1$, while Gurvits' error, $\varepsilon\lambda_{\max}$, is exponentially increasing in this case. In other words, we wish to guarantee an additive error εk^M ($k \leqslant 1$) for the permanent of the HPSM $\mathbf{\Lambda}$. Therefore, in Eq. (A2) we set $\epsilon = \varepsilon k^M/A$:

$$N = \frac{1}{2\varepsilon^2}\ln\frac{1}{\delta}\left(\frac{\lambda_{\max}^2C^2}{kea}\right)^{2M}. \tag{A14}$$

Our aim is then to approximate the permanent of the matrix $\mathbf{\Lambda}$ to within the minimal additive error, which decreases exponentially in M. Thus, in order to avoid the exponential increase of the sample size N, analogous to the previous subsection, we obtain

$$a \geqslant \frac{\lambda_{\max}^2C^2}{e}. \tag{A15}$$

This inequality defines a necessary and sufficient condition on the spectra of a given HPSM, which guarantees the efficiency of our approximation algorithm within the present regime.

On the other hand, due to Eq. (A6), $a \leqslant C\lambda_{\max} - \bar{\lambda}$ and thus along with Eq. (A15) we find the necessary condition

$$\bar{\lambda} \leqslant \lambda_{\max}C\left(1-\frac{\lambda_{\max}C}{e}\right). \tag{A16}$$

Meanwhile, since $\bar{\lambda} \geqslant 0$, we end up with another necessary condition,

$$\lambda_{\max} \leqslant \frac{e}{C}. \tag{A17}$$

The constraints (A16) and (A17) thus provide some intuition on the spectrum of HPSMs for which our algorithm yields an exponentially decreasing error in M (in polynomial time). As $\lambda_{\max} \geqslant 1$, we obtain the next necessary condition,

$$C \leqslant e. \tag{A18}$$

In other words, for the permanent of HPSMs with $1 \leqslant \lambda_{\max} < e$, we are potentially able to attain an exponentially decreasing error εk^M.

Finally, as in the previous section, we are able to bound k from below,

$$k \geqslant \frac{C^2\lambda_{\max}^2}{e(C\lambda_{\max}-\bar{\lambda})} \geqslant \frac{1}{e}, \tag{A19}$$

which gives information on the minimal additive error that our scheme potentially provides within the regime discussed here.

3. "Almost relative" error (set *S*3 of the main text)

In this section we analyze the efficiency of our scheme beyond approximations to within an additive error. Namely,

we consider the task of estimating the permanent of an HPSM $\mathbf{\Lambda}$ to within an error $\pm\varepsilon\sqrt{\mathrm{Per}\mathbf{\Lambda}}$. For that, in Eq. (A2) we set $\epsilon = \varepsilon\sqrt{\mathbb{E}[p^{\mathrm{cs}}(\boldsymbol{\alpha})]_{\boldsymbol{\alpha}\in\mathcal{N}}/Z}$, yielding

$$N = \frac{1}{2\varepsilon^2}\ln\frac{1}{\delta}\frac{Z}{e^{2M}\mathbb{E}[p^{\mathrm{cs}}(\boldsymbol{\alpha})]_{\boldsymbol{\alpha}\in\mathcal{N}}}. \tag{A20}$$

In order to reveal the efficient regimes of this specific case, we first provide a lower bound for $\mathbb{E}[p^{\mathrm{cs}}(\boldsymbol{\alpha})]_{\boldsymbol{\alpha}\in\mathcal{N}}$. Thus, we return to the definition of $p^{\mathrm{cs}}(\boldsymbol{\alpha})$ itself,

$$\mathbb{E}[p^{\mathrm{cs}}(\boldsymbol{\alpha})]_{\boldsymbol{\alpha}\in\mathcal{N}} = \int_{\mathbb{C}^M} d\boldsymbol{\alpha}\prod_{j=1}^{M} e^{-|\beta_j|^2}|\beta_j|^2\prod_{i=1}^{M}\left[\frac{1}{\pi\langle n_i\rangle}\exp\left(-\frac{|\alpha_i|^2}{\langle n_i\rangle}\right)\right] = \frac{1}{\pi^M\prod_{i=1}^{M}\langle n_i\rangle}\int_{\mathbb{C}^M} d\boldsymbol{\alpha}$$
$$\times\left[\prod_{j=1}^{M} e^{-|\beta_j|^2}|\beta_j|^2\exp\left(-\sum_{i=1}^{M}\frac{|\alpha_i|^2}{\langle n_i\rangle}\right)\right] \geqslant \frac{1}{\pi^M\prod_{i=1}^{M}\langle n_i\rangle}\int_{\mathbb{C}^M} d\boldsymbol{\alpha}\prod_{j=1}^{M} e^{-|\beta_j|^2}|\beta_j|^2\exp\left(-\frac{1}{\langle n_{\min}\rangle}\sum_{i=1}^{M}|\alpha_i|^2\right), \tag{A21}$$

where $d\boldsymbol{\alpha} \equiv d^2\alpha_1,\ldots,d^2\alpha_M$ and $\langle n_{\min}\rangle$ denotes the minimal $\langle n_i\rangle$:

$$\langle n_{\min}\rangle = \frac{\lambda_{\min}}{C\lambda_{\max}-\lambda_{\min}}. \tag{A22}$$

Next, taking into account that $\beta_i = \sum_{j=1}^{M}\mathbf{U}_{ji}\alpha_j$ and the fact that the matrix $\mathbf{U}$ is unitary, we find that $\sum_{i=1}^{M}|\alpha_i|^2 = \sum_{j=1}^{M}|\beta_j|^2$. Thus, Eq. (A21) reads

$$\mathbb{E}[p^{\mathrm{cs}}(\boldsymbol{\alpha})]_{\boldsymbol{\alpha}\in\mathcal{N}} \geqslant \frac{1}{\pi^M\prod_{i=1}^{M}\langle n_i\rangle}\int_{\mathbb{C}^M} d\boldsymbol{\beta}\prod_{i=1}^{M} e^{-|\beta_i|^2}|\beta_i|^2\exp\left(-\frac{1}{\langle n_{\min}\rangle}\sum_{i=1}^{M}|\beta_i|^2\right) = \frac{1}{\prod_{i=1}^{M}\langle n_i\rangle}\left(\frac{\langle n_{\min}\rangle}{1+\langle n_{\min}\rangle}\right)^{2M}. \tag{A23}$$

In the above equation we also made a change of integration variables (from $\{\alpha_1,\ldots,\alpha_M\}$ to $\{\beta_1,\ldots,\beta_M\}$), using the fact that the absolute value of the Jacobian determinant of the corresponding unitary transformation is unity ($|\det\mathbf{U}| = 1$) [21]. Finally, we rewrite the above lower bound in terms of the eigenvalues of $\mathbf{\Lambda}$, yielding

$$\mathbb{E}[p^{\mathrm{cs}}(\boldsymbol{\alpha})]_{\boldsymbol{\alpha}\in\mathcal{N}} \geqslant \frac{1}{\prod_{i=1}^{M}\langle n_i\rangle}\left(\frac{\lambda_{\min}}{C\lambda_{\max}}\right)^{2M} = \prod_{i=1}^{M}\frac{C\lambda_{\max}-\lambda_i}{\lambda_i}\left(\frac{\lambda_{\min}}{C\lambda_{\max}}\right)^{2M}. \tag{A24}$$

Substituting this expression into Eq. (A20) we find

$$N \leqslant \frac{1}{2\varepsilon^2}\ln\frac{1}{\delta}\frac{1}{e^{2M}}\frac{\lambda_{\max}^{2M}C^{2M}}{\prod_{i=1}^{M}(C\lambda_{\max}-\lambda_i)}\frac{\prod_{i=1}^{M}\lambda_i}{\prod_{i=1}^{M}(C\lambda_{\max}-\lambda_i)}\left(\frac{C\lambda_{\max}}{\lambda_{\min}}\right)^{2M} = \left(\frac{\lambda_{\max}^4C^4d}{\lambda_{\min}^2e^2}\right)^M, \tag{A25}$$

where

$$d = \sqrt[M]{\prod_{i=1}^{M}\frac{\lambda_i}{(C\lambda_{\max}-\lambda_i)^2}} \tag{A26}$$

is the geometric mean of the quantities $\{\lambda_1/(C\lambda_{\max}-\lambda_1)^2,\ldots,\lambda_M/(C\lambda_{\max}-\lambda_M)^2\}$. As in the previous cases, the algorithm is efficient if

$$\frac{\lambda_{\max}^4C^4d}{\lambda_{\min}^2e^2} \leqslant 1. \tag{A27}$$

The set of matrices defined by the above inequality is not empty; i.e., there exist HPSMs such that their spectrum satisfies the condition (A27), which, in turn, guarantees the additive error $\pm\varepsilon\sqrt{\mathrm{Per}\mathbf{\Lambda}}$ of our algorithm. This condition has to be checked for a given matrix $\mathbf{\Lambda}$.

As for the special case $\lambda_{\max} < 1$, as already mentioned above, we readily obtain the corresponding constraints by simply replacing C with $1/\lambda_{\max}$,

$$\frac{f}{\lambda_{\min}^2e^2} \leqslant 1, \tag{A28}$$

where

$$f = \sqrt[M]{\prod_{i=1}^{M}\frac{\lambda_i}{(1-\lambda_i)^2}}. \tag{A29}$$

As a conclusion to this section we note that, depending on the type of the error one wants to achieve, the corresponding conditions (A10), (A15), and (A27) [or (A28)] should be checked, before applying the steps of the algorithm outlined in the main text of the paper. If satisfied, one proceeds with the estimation scheme. Finally, we also emphasize that our method possesses certain tunability in terms of the parameter C, which appears in the corresponding conditions. It allows one to expand the applicability of the algorithm, as well as to optimize the resulting error. Nevertheless, as our analysis shows, C cannot be chosen arbitrarily large. Namely, in order

to avoid the exponential increase of the corresponding additive error, one has to set $C \leqslant e$.

APPENDIX B: PERMANENT UPPER AND LOWER BOUNDS

It is worth noting an important corollary of our results, which provides an exponentially decreasing upper bound for the permanent of HPSMs. Namely, combining Eqs. (A3) and (A4) with the fact that $p^{\mathrm{cs}}(\boldsymbol{\alpha}) \leqslant e^{-M}$, we find

$$\mathrm{Per}\boldsymbol{\Lambda} \leqslant \left(\frac{C^2\lambda_{\max}^2}{ae}\right)^M. \tag{B1}$$

Therefore, if the matrix $\boldsymbol{\Lambda}$ satisfies the conditions $\lambda_{\max} \geqslant 1$ and (A15), $\mathrm{Per}\boldsymbol{\Lambda}$ decreases exponentially with the size of the matrix M. In contrast, the standard upper bound for permanents, resulting from Glynn's formula, yields $\mathrm{Per}\boldsymbol{\Lambda} \leqslant \lambda_{\max}^M$, which *increases* exponentially in M if $\lambda_{\max} \geqslant 1$.

Additionally, we are also able to provide a lower bound for the permanent of HPSMs, expressed in terms their spectra, using the lower bound for $p^{\mathrm{cs}}(\boldsymbol{\alpha})$ of Eq. (A24):

$$\mathrm{Per}\boldsymbol{\Lambda} \geqslant \frac{\lambda_{\min}^{2M}}{\prod_{i=1}^M \lambda_i}. \tag{B2}$$

APPENDIX C: SINGLE-PHOTON MEASUREMENT PROBABILITY AND THE PERMANENT OF HERMITIAN POSITIVE SEMIDEFINITE MATRICES [PROOF OF EQ. (3) OF THE MAIN TEXT]

For completeness, in this section we outline the derivation of the relation between the joint single-photon measurement probability p^{th} at the output of a linear-optically evolved M-mode thermal state, and the permanent of Hermitian positive semidefinite matrices [Eq. (3) of the main text], following the corresponding proof of Ref. [10]. For this purpose we use the Husimi Q-function representation of the thermal state ρ_i^{th} [22,23],

$$\begin{aligned} Q_i^{\mathrm{th}}(\alpha_i) &= \frac{1}{\pi}\langle\alpha_i|\rho_i^{\mathrm{th}}|\alpha_i\rangle \\ &= \frac{1}{\pi(\langle n_i\rangle+1)}\exp\left(-\frac{|\alpha_i|^2}{\langle n_i\rangle+1}\right), \end{aligned} \tag{C1}$$

which, for an M-mode thermal state, $\bigotimes_{i=1}^M \rho_i^{\mathrm{th}}$, yields

$$Q_{\mathrm{in}}^{\mathrm{th}}(\boldsymbol{\alpha}) = \prod_{i=1}^M Q_i^{\mathrm{th}}(\alpha_i), \tag{C2}$$

where $\boldsymbol{\alpha}$ denotes the set of variables $\{\alpha_1,\ldots,\alpha_M\}$. On the other hand, a remarkable feature of the Husimi function is that for the state $\rho_{\mathrm{out}} = \mathcal{U}\rho_{\mathrm{in}}^{\mathrm{th}}\mathcal{U}^\dagger$ it is again a product of the input functions $Q_i^{\mathrm{th}}(\alpha_i)$, but of a different argument:

$$\begin{aligned} Q_{\mathrm{out}}(\boldsymbol{\alpha}) &= \frac{1}{\pi^M}\langle\boldsymbol{\alpha}|\rho_{\mathrm{out}}|\boldsymbol{\alpha}\rangle = \frac{1}{\pi^M}\langle\boldsymbol{\alpha}|\mathcal{U}\rho_{\mathrm{in}}^{\mathrm{th}}\mathcal{U}^\dagger|\boldsymbol{\alpha}\rangle \\ &= \frac{1}{\pi^M}\langle\boldsymbol{\eta}|\rho_{\mathrm{in}}^{\mathrm{th}}|\boldsymbol{\eta}\rangle = \prod_{i=1}^M Q_i^{\mathrm{th}}\left(\sum_{j=1}^M \bar{\mathbf{U}}_{ji}\alpha_j\right). \end{aligned} \tag{C3}$$

In the above equation $|\boldsymbol{\alpha}\rangle = \bigotimes_{i=1}^M |\alpha_i\rangle$, $|\boldsymbol{\eta}\rangle = \mathcal{U}^\dagger|\boldsymbol{\alpha}\rangle$, and $\bar{\mathbf{U}}$ stands for the complex conjugate of the unitary matrix $\mathbf{U}$. As a result,

$$Q_{\mathrm{out}}(\boldsymbol{\alpha}) = \frac{1}{\pi^M\prod_{i=1}^M(\langle n_i\rangle+1)}\exp(-\vec{\boldsymbol{\alpha}}\mathbf{B}\vec{\boldsymbol{\alpha}}^\dagger), \tag{C4}$$

where $\vec{\boldsymbol{\alpha}} = (\alpha_1,\ldots,\alpha_M)$ stands for the row of variables α_i, $\mathbf{B} = \mathbf{U}\boldsymbol{\zeta}\mathbf{U}^\dagger$, and $\boldsymbol{\zeta} = \mathrm{diag}[1/(\langle n_1\rangle+1),\ldots,1/(\langle n_M\rangle+1)]$. Meanwhile, the single-photon measurement probability p^{th} at the output of the linear-optical circuit can be also rewritten as

$$\begin{aligned} p^{\mathrm{th}} &= \mathrm{Tr}[\rho_{\mathrm{out}}|\mathbf{1}\rangle\langle\mathbf{1}|] \\ &= \pi^M\int_{\mathbb{C}^M} d\boldsymbol{\alpha}\, Q_{\mathrm{out}}(\boldsymbol{\alpha})P_{|\mathbf{1}\rangle\langle\mathbf{1}|}(\boldsymbol{\alpha}). \end{aligned} \tag{C5}$$

Here $|\mathbf{1}\rangle = \bigotimes_{i=1}^M |1\rangle$ and $P_{|\mathbf{1}\rangle\langle\mathbf{1}|}(\boldsymbol{\alpha})$ stands for the Glauber-Sudarshan P representation of the single-photon state projector $|\mathbf{1}\rangle\langle\mathbf{1}|$ [23],

$$P_{|\mathbf{1}\rangle\langle\mathbf{1}|}(\boldsymbol{\alpha}) = \prod_{i=1}^M e^{|\alpha_i|^2}\frac{\partial^2}{\partial\alpha_i\partial\bar{\alpha}_i}\delta^{(2)}(\alpha_i), \tag{C6}$$

where $\delta^{(2)}(\alpha_i) = \delta[\mathrm{Re}(\alpha_i)]\delta[\mathrm{Im}(\alpha_i)]$ is the two-dimensional Dirac δ function. Using Eq. (C6), we thus rewrite Eq. (C5) as

$$\begin{aligned} p^{\mathrm{th}} = \frac{1}{\prod_{i=1}^M(\langle n_i\rangle+1)}&\int_{\mathbb{C}^M} d\boldsymbol{\alpha}\exp(-\vec{\boldsymbol{\alpha}}\mathbf{B}\vec{\boldsymbol{\alpha}}^\dagger) \\ &\times\prod_{i=1}^M e^{|\alpha_i|^2}\frac{\partial^2}{\partial\alpha_i\partial\bar{\alpha}_i}\delta^{(2)}(\alpha_i). \end{aligned} \tag{C7}$$

Finally, integrating the last equation by parts we obtain the expression

$$p^{\mathrm{th}} = \frac{1}{\prod_{i=1}^M(\langle n_i\rangle+1)}\left[\prod_{i=1}^M\frac{\partial^2}{\partial\alpha_i\partial\bar{\alpha}_i}e^{F(\boldsymbol{\alpha})}\right]\Bigg|_{\alpha_i=0}, \tag{C8}$$

where

$$F(\boldsymbol{\alpha}) = \vec{\boldsymbol{\alpha}}\mathbf{D}\vec{\boldsymbol{\alpha}}^\dagger = \sum_{i,j=1}^M \mathbf{D}_{ij}\alpha_i\bar{\alpha}_j, \tag{C9}$$

with $\mathbf{D} = \mathbf{I} - \mathbf{B}$, $\mathbf{I}$ being the $M\times M$ identity matrix. Note that the function $F(\boldsymbol{\alpha})$ is a second-order polynomial in α_i and $\bar{\alpha}_j$, where every term is proportional to $\alpha_i\bar{\alpha}_j$, while the right-hand side of Eq. (C8) corresponds to a product of derivatives of a multivariate exponential function evaluated at $\boldsymbol{\alpha} = 0$ [$e^{F(\boldsymbol{\alpha}=0)} = 1$]. In full generality, this expression is written as a sum over products of combinations of partial derivates of the function $F(\boldsymbol{\alpha})$ with respect to the variables α_i and $\bar{\alpha}_j$. Due to the quadratic from of $F(\boldsymbol{\alpha})$ and the evaluation at $\boldsymbol{\alpha} = 0$, we thus observe that the only terms that contribute to the final result are products of M second-order derivatives of $F(\boldsymbol{\alpha})$, where we first derive over a variable α_i followed by a derivative over $\bar{\alpha}_j$. Next, every i and j can only appear once over each product of M second-order derivatives; i.e., we have instances of the form $\prod_{i=1}^M \partial^2 F(\boldsymbol{\alpha})/\partial\alpha_i\partial\bar{\alpha}_{\sigma(i)}$, where σ denotes a specific permutation of natural numbers $\{1,\ldots,M\}$. Because of the symmetry and the form of Eq. (C8) we can

see that the sum runs over all permutations S_M of the set $\{1,\ldots,M\}$, which leads to the following expression for p^{th}:

$$p^{\text{th}} = \frac{1}{\prod_{i=1}^{M}(\langle n_i\rangle + 1)} \sum_{\sigma\in S_M} \prod_{i=1}^{M} \frac{\partial^2 F(\boldsymbol{\alpha})}{\partial\alpha_i \partial\alpha_{\sigma(i)}}. \tag{C10}$$

Therefore, using Eq. (C9) we find that each term in the above sum represents a product of M elements of $\mathbf{D}$, $\prod_{i=1}^{M} \mathbf{D}_{i\sigma(i)}$. Hence, by the definition of the permanent, we conclude that

$$p^{\text{th}} = \frac{1}{\prod_{i=1}^{M}(\langle n_i\rangle + 1)} \text{Per}\mathbf{D}. \tag{C11}$$

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

[2] E. Knill, R. Laflamme, and G. J. Millburn, Nature (London) **409**, 46 (2001).

[3] X. Ding, Y. He, Z.-C. Duan, N. Gregersen, M.-C. Chen, S. Unsleber, S. Maier, C. Schneider, M. Kamp, S. Höfling, C.-Y. Lu, and J.-W. Pan, Phys. Rev. Lett. **116**, 020401 (2016).

[4] J. Carolan, C. Harrold, C. Sparrow *et al.*, Science **349**, 711 (2015).

[5] S. Aaronson and A. Arkhipov, Theor. Comput. **9**, 143 (2013).

[6] S. Scheel and S. Y. Buhmann, Acta Physica Slovaca **58**, 675 (2008).

[7] L. G. Valiant, Theor. Comput. Sci. **8**, 189 (1979).

[8] S. Aaronson, Proc. R. Soc. London, Ser. A **467**, 3393 (2011).

[9] J. B. Spring, B. J. Metcalf, P. C. Humphreys *et al.*, Science **339**, 798 (2013); M. A. Broome, A. Fedrizzi, S. Rahimi-Keshari *et al.*, *ibid.* **339**, 794 (2013); M. Tilmann, B. Dakić, R. Heilmann *et al.*, Nat. Photon. **7**, 540 (2013); A. Crespi, R. Osellame, R. Ramponi *et al.*, *ibid.* **7**, 545 (2013).

[10] S. Rahimi-Keshari, A. P. Lund, and T. C. Ralph, Phys. Rev. Lett. **114**, 060501 (2015).

[11] R. J. Glauber, Phys. Rev. Lett. **10**, 84 (1963); E. C. G. Sudarshan, *ibid.* **10**, 277 (1963).

[12] S. Aaronson, See point (4) of Comment #84 at http://www.scottaaronson.com/blog/?p=2408#comment-757410

[13] L. Gurvits, in *Mathematical Foundations of Computer Science* (Springer-Verlag, Berlin, 2005), pp. 447–458.

[14] V. S. Shchesnovich, Phys. Rev. Lett. **116**, 123601 (2016).

[15] J. W. Demmel, *Applied Numerical Linear Algebra* (SIAM, Philadelphia, 1997); L. N. Trefethen, and D. Bau, III, *Numerical Linear Algebra* (SIAM, Philadelphia, 1997).

[16] G. E. P. Box and M. E. Muller, Ann. Math. Stat. **29**, 610 (1958); A. J. Kinderman and J. F. Monahan, ACM Trans. Math. Softw. **3**, 257 (1977); C. F. F. Karney, *ibid.* **42**, 1 (2016); G. Marsaglia and W. W. Tsang, J. Stat. Softw. **5** (2000).

[17] W. Hoeffding, J. Am. Stat. Assoc. **58**, 13 (1963).

[18] H. Chernoff, Ann. Math. Stat. **23**, 493 (1952).

[19] H. Ryser, *Combinatorial Mathematics* (Wiley, New York, 1963).

[20] S. Aaronson and T. Hance, Quantum Inf. Comput. **14**, 541 (2014).

[21] W. Kaplan, *Advanced Calculus* (Addison-Wesley, Boston, 2002).

[22] G. S. Agarwal, *Quantum Opics* (Cambridge University Press, Cambridge, UK, 2013).

[23] U. Leonhardt, *Essential Quantum Optics* (Cambridge University Press, Cambridge, UK, 2010).