Sufficient condition for the mode mismatch of single photons for scalability of the boson-sampling computer

V. S. Shchesnovich

Centro de Ciências Naturais e Humanas, Universidade Federal do ABC, Santo André, São Paulo 09210-170, Brazil

(Received 26 November 2013; published 24 February 2014)

The boson sampler proposed by Aaronson and Arkhipov is a nonuniversal quantum computer, which can serve as evidence against the extended Church-Turing thesis. It samples the probability distribution at the output of a linear unitary optical network with indistinguishable single photons at the input. Four experimental groups have already tested their small-scale prototypes with up to four photons. A boson sampler with a few dozens of single photons is believed to be hard to simulate on a classical computer. For scalability of a realistic boson sampler with current technology it is necessary to know the effect of the photon mode mismatch on its operation. Here a nondeterministic model of the boson sampler is analyzed, which employs partially indistinguishable single photons emitted by identical sources. A sufficient condition on the average mutual fidelity $\langle \mathcal{F} \rangle$ of the single photons is found, which guarantees that the realistic boson sampler outperforms the classical computer. Moreover, the boson-sampler computer with partially indistinguishable single photons is scalable and has more power than classical computers when the single-photon mode mismatch $1 - \langle \mathcal{F} \rangle$ scales as $O(N^{-3/2})$ with the total number of photons N.

DOI: 10.1103/PhysRevA.89.022333 PACS number(s): 42.50.Ex, 03.67.Lx, 05.30.Jp, 42.50.Ar

I. INTRODUCTION

The boson-sampler (BS) computer proposed recently by Aaronson and Arkhipov [1] can serve as evidence against the extended Church-Turing (ECT) thesis which says that any physical device can be efficiently simulated on the probabilistic Turing machine. No interaction between bosons is required, and thus the BS computer can be built using only passive linear optical devices and emitters of indistinguishable single photons [2], i.e., the single photons that show the Hong-Ou-Mandel type of interference [3] (see also Refs. [4,5]). Whereas the universal quantum computer targets NP decision problems, widely believed to be classically hard, such as factoring large integers [6,7], the BS computer just samples the output probability distribution of an M-mode unitary network U with N identical bosons at its input. It is shown that simulation of the BS on a classical computer requires exponential resources in the number of bosons N (when $M \geqslant N$) [1], since bosonic amplitudes are given as the permanents (see Ref. [8] for the definition and properties) of complex $N \times N$ submatrices of U [9,10], whose computation is exponentially hard [11,12] [Ryser's algorithm [13], the fastest known, requires $O(N^2 2^N)$ flops]. On the conceptual side, a classical algorithm for the matrix permanent would provide also for solution of all problems in the complexity class #P, of a higher complexity than the NP class, which, in its turn, would imply dramatic theoretical consequences: collapse of the whole polynomial hierarchy of the computational complexity [1]. While a universal quantum computer can simulate the BS, the scalability of the BS beyond classical computational power is easier to achieve: already with $20 \leqslant N \leqslant 30$ photons it would outperform classical computers [1]. Four independent groups have already tested their prototypes of the BS on small networks with up to four input photons [14–17].

It is crucial that even an *approximate* simulation of the BS computer must be classically hard (at least when $M \gg N^2$) [1]; hence, the stringent fault tolerances required for the universal quantum computer [18–21] may be significantly relaxed for the BS computer. The necessary, though not sufficient, conditions for BS operation beyond the power of classical computers were analyzed in Refs. [22,23], supporting this view. It was even suggested [23] that scaling up helps to combat photon mode mismatch and losses. Recently, the effect of noise in the experimental realization of a unitary network on BS complexity was studied [24]. It was shown that even with fidelity of the optical elements of at least $1 - O(N^{-2})$ a noisy-network realization of the BS is still hard to simulate classically. These results suggest the experimental feasibility of the BS computer in the near future.

In practice, limitations on the indistinguishability of single photons from realistic sources will always be present. All four groups of Refs. [14–17] have tested their BS prototypes using so-called heralded single photons from parametric down-conversion, not free from multiphoton components and noise. It is clear that some amount of indistinguishability of single photons is essential for the BS computer (a large mode mismatch allows for an efficient simulation on a classical computer [1] by a probabilistic algorithm [25]; see also below). Recently a spatial multiplexing of heralded single-photon sources was proposed to enhance the relative yield of the single-photon component [26], but scalability is still out of reach. On the other hand, scalable single-photon sources with high photon antibunching can be based on individual emitters such as quantum dots [27–29], but they are inherently nondeterministic, since they are based on spontaneous emission or on spontaneous decay from a cavity. Could nondeterministic sources of single photons be employed to scale up the BS? Generally, what specific features of bosonic particles are necessary for the BS computer to outperform the classical computer? A related fundamental problem is that, to date, no *sufficient* bound is known on the mode mismatch of single photons for an experimentally realistic BS to serve as evidence against the ECT thesis.

Thus, it is of paramount importance for building a scalable BS device to establish the degree of distinguishability of

single photons for which the BS is still hard to simulate on a classical computer. This is the main focus of the present work. The analysis is concentrated on the effect of the photon mode mismatch and neglects two other sources of error, i.e., noise in experimental realization of an unitary network and photon losses. A sufficient bound on the mode mismatch is derived for the BS computer with partially indistinguishable single photons to outperform the classical computer. *For instance, the BS computer with partially indistinguishable single photons is scalable beyond the power of the classical computer if the mode mismatch* $1-\langle\mathcal{F}\rangle$*, where* $\langle\mathcal{F}\rangle$ *is the average single-photon fidelity, scales as* $O(N^{-3/2})$ *with the total number of photons* N. In derivation of the fidelity bound, the indistinguishability of N single photons in distinct modes is quantified by an N-vector parameter—an approach which might be useful in other problems.

The rest of the text is organized as follows. In Sec. II a nondeterministic boson-sampler (NDBS) model is formulated which captures the essential features of any nonideal BS computer with the single photons only partially indistinguishable. Section III is devoted to analyzing the conditions under which the NDBS performs a classically hard computational task. In Sec. IV a short summary of the results is given. Some inessential mathematical details of the derivations and other computational details are relegated to Appendixes A, B, and C.

II. THE NONDETERMINISTIC BOSON-SAMPLER MODEL

Consider N single photons emitted by identical sources and launched into distinct input modes $k_1,\ldots,k_N$ of an M-mode linear optical network given by a unitary matrix U: $a_k^\dagger(\omega)=\sum_{l=1}^M U_{kl}b_l^\dagger(\omega)$, where $a_k(\omega)$ and $b_k(\omega)$ are the input and output modes of frequency ω, respectively (see Fig. 1). The input state is given by a density matrix. Setting $\mathbf{x}$ to be a fluctuating vector parameter in the spectral function $\phi(\mathbf{x},\omega)$ of a single photon (for instance, the arrival time or phase) with the distribution $p(\mathbf{x})$, identical for each source, the density matrix reads $\rho^{(in)}=\int d\mathbf{x}_1\cdots\int d\mathbf{x}_N[\prod_{\alpha=1}^N p(\mathbf{x}_\alpha)]|\Psi(\mathbf{x}_1,\ldots,\mathbf{x}_N)\rangle\langle\Psi(\mathbf{x}_1,\ldots,\mathbf{x}_N)|$,

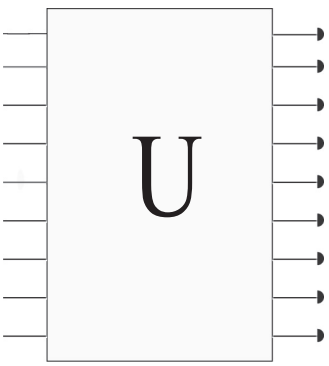

FIG. 1. Schematic (black-box) depiction of the NDBS setup with the network matrix U, where on the left are the input modes corresponding to the operators $a_k(\omega)$, linked to the identical photon sources, and on the right are the output modes corresponding to the operators $b_k(\omega)$ linked to the detectors.

where

$$|\Psi(\mathbf{x}_1,\ldots,\mathbf{x}_N)\rangle=\prod_{\alpha=1}^N\int_0^\infty d\omega_\alpha\phi(\mathbf{x}_\alpha,\omega_\alpha)a_{k_\alpha}^\dagger(\omega_\alpha)|0\rangle \tag{1}$$

is a Fock state of N photons at the input. This is a more general setup than in Ref. [1], which allows consideration of the effect of photon mode mismatch. The output probability of detecting $m_1,\ldots,m_M$ photons in modes $1,\ldots,M$ can be derived using quantum photon counting theory [30–32]. The result is that the probability is given by the following positive Hermitian operator (see Appendix A):

$$\Pi(m_1,\ldots,m_M)=\frac{1}{\prod_{l=1}^M m_l!}\int_0^\infty d\omega_1\cdots\int_0^\infty d\omega_N\prod_{\alpha=1}^N\Gamma(\omega_\alpha)\times\left[\prod_{\alpha=1}^N b_{l_\alpha}^\dagger(\omega_\alpha)\right]|0\rangle\langle 0|\left[\prod_{\alpha=1}^N b_{l_\alpha}(\omega_\alpha)\right], \tag{2}$$

where $(l_1,\ldots,l_N)\equiv\{1,\ldots,1,2,\ldots,2,\ldots,M,\ldots,M\}$, with index j appearing m_j times, and $\Gamma(\omega)\geqslant 0$ is the spectral function of the detector. The set of all such operators as in Eq. (2), after a suitable normalization (see below), constitutes the positive operator-valued measure (POVM) describing photon detection at the output modes. From Eqs. (1) and (2), the detection probability $P(m_1,\ldots,m_M|k_1,\ldots,k_N)=\mathrm{tr}\{\Pi(m_1,\ldots,m_M)\rho^{(in)}\}$ becomes

$$P(m_1,\ldots,m_M|k_1,\ldots,k_N)=\frac{1}{\prod_{l=1}^M m_l!}\sum_{\sigma_1}\sum_{\sigma_2}J\left(\sigma_2\sigma_1^{-1}\right)\prod_{\alpha=1}^N U^*_{k_{\sigma_1(\alpha)},l_\alpha}U_{k_{\sigma_2(\alpha)},l_\alpha} \tag{3}$$

with each sum running over all permutations of N indices $k_1,\ldots,k_N$ in an $N\times N$ submatrix of the network matrix U. In fact, Eq. (3) applies more generally, not necessarily with identical sources, when the network input consists of states with up to one photon per mode. In this general case, J depends only on the relative permutation $\sigma_{21}\equiv\sigma_2\sigma_1^{-1}$.[1] Evidently $J=\delta_{\sigma_1,\sigma_2}$ is the classical limit, whereas the ideal BS of Aaronson and Arkhipov has $J=1$ (independently of its argument). In our case, due to the sources are identical, J factorizes into a product of functions of cycles of the relative permutation, where cycles of the same length contribute the same factor.[2] Thus J is a function of the cyclic structure $C_1,\ldots,C_N$ of σ_{21} (C_k is the number of cycles of length k; $\sum kC_k=N$ [33]). In particular, we obtain (see Appendix A)

$$J(\sigma)=\prod_{k=2}^N g_k^{C_k(\sigma)}, \tag{4}$$

[1]Since J is independent of U, the substitution $\alpha=\sigma_1^{-1}(\beta)$ gives $U^*_{k_\beta,l'_\beta}U_{k_{\sigma_{21}(\beta)},l'_\beta}$, i.e., the transformed matrix U with the σ_1-permuted output indices $l'_\beta\equiv l_{\sigma_1^{-1}(\beta)}$. Hence J depends on σ_{21} only.

[2]Since the sources are identical, two cycles involving k sources can be mapped into each other by relabeling the sources.

where we have introduced

$$g_k = \prod_{\alpha=1}^{k} \int d\mathbf{x}_\alpha\, p(\mathbf{x}_\alpha) \int_0^\infty d\omega_\alpha\, \Phi(\mathbf{x}_\alpha,\omega_{\alpha-1})\Phi^*(\mathbf{x}_\alpha,\omega_\alpha) \quad (5)$$

with $\Phi(\mathbf{x},\omega) \equiv \sqrt{\Gamma(\omega)}\phi(\mathbf{x},\omega)$ (the product is a shortcut notation for the multiple integrals over $\mathbf{x}_\alpha$ and ω_α, where $\alpha = 0$ is the same as $\alpha = k$). For efficient broadband detectors a small percentage of losses can be dealt with by using postselection. In this case, normalizing the modified spectral function as $\int d\omega |\Phi(\mathbf{x},\omega)|^2 = 1$, we get for the probabilities of Eq. (3) $\sum_{\{m_j\}} P(m_1,\ldots,m_M|k_1,\ldots,k_N) = 1$, where the summation is constrained by $m_1 + \cdots + m_M = N$ [indeed, the described renormalization is equivalent to setting $\Gamma(\omega) = 1$, i.e., to the case of bandwidth-unlimited ideal detectors and single photons with the modified spectral function, where all photons are detected].

g_k has the physical meaning of the k-photon indistinguishability parameter defined for identical single-photon sources (in general, indistinguishability of single photons is described by Young diagrams [34]; for the general multiphoton case see Refs. [35,36]). In the ideal BS case all $g_k = 1$, whereas in the classical case $g_k = 0$, $k \geqslant 2$. The physical meaning of g_k requires that it is positive. This and other properties of g_k can be easily seen from the following representation. Introduce the following one-particle density matrix:

$$\rho \equiv \int d\mathbf{x}\, p(\mathbf{x}) |\Phi(\mathbf{x})\rangle\langle\Phi(\mathbf{x})| \quad (6)$$

with the vector $|\Phi(\mathbf{x})\rangle \in \mathcal{H}$ defined as $\langle\omega|\Phi(\mathbf{x})\rangle \equiv \Phi(\mathbf{x},\omega)$, where the Hilbert space $\mathcal{H}$ has a resolution of unity given by $\int_0^\infty d\omega |\omega\rangle\langle\omega| = \hat{1}$. Note that the above normalization of $\Phi(\mathbf{x},\omega)$ guarantees that $\mathrm{tr}(\rho) = 1$. Under these definitions, Eq. (5) can be cast in the form of a trace of a positive operator (by recognizing in the integrals the resolution of unity in $\mathcal{H}$)

$$g_n = \int_0^\infty d\omega_1 \cdots \int_0^\infty d\omega_n \prod_{j=1}^{n} \langle\omega_j|\rho|\omega_{j+1}\rangle$$
$$= \int_0^\infty d\omega_1 \langle\omega_1|\rho^n|\omega_1\rangle = \mathrm{tr}\rho^n. \quad (7)$$

Hence, $0 \leqslant g_n \leqslant 1$. Moreover, passing to the diagonal basis, we also obtain an important bound for higher indistinguishability parameters (setting also $g_1 = 1$, for convenience)

$$g_n = \mathrm{tr}(\rho^k \rho^{n-k}) \leqslant \mathrm{tr}(\rho^k)\mathrm{tr}(\rho^{n-k}) = g_k g_{n-k}. \quad (8)$$

For instance, $g_{n+1} \leqslant g_n$.

One general observation follows: Since the computational complexity of the NDBS decreases as $J(\sigma)$ deviates from its maximum[3] $J = 1$ (except for the identity permutation) and the indistinguishability parameters satisfy $g_{n+1} \leqslant g_n$, it is doubtful that scaling up to a higher number of single photons can help to combat the photon mode mismatch (as suggested in Ref. [23]). Below we derive a sufficient condition on the mode mismatch which has an inverse 3/2-power law scaling in the total number of photons.

Equations (3)–(5) are the basis of our consideration. Below we focus on the region of small mode mismatch. In this case the average mutual fidelity of the single photons (denoting the averaging over $\mathbf{x}$ by $\langle\cdots\rangle$)

$$\langle\mathcal{F}\rangle = \int d\mathbf{x}_1\, p(\mathbf{x}_1) \int d\mathbf{x}_2\, p(\mathbf{x}_2) |\langle\Phi(\mathbf{x}_1)|\Phi(\mathbf{x}_2)\rangle|$$
$$= \int d\mathbf{x}_1\, p(\mathbf{x}_1) \int d\mathbf{x}_2\, p(\mathbf{x}_2) \left| \int d\omega \Phi^*(\mathbf{x}_1,\omega)\Phi(\mathbf{x}_2,\omega) \right| \quad (9)$$

can be expanded in powers of the vector variable $\mathbf{x}$ (we set, for simplicity, $\langle\mathbf{x}\rangle = \mathbf{0}$). Indeed, from Eq. (9), using that $\mathbf{x}_{1,2}$ have identical distributions, we get

$$\langle\mathcal{F}\rangle = \left\langle \left[\int d\omega \Phi^*(\mathbf{x}_1,\omega)\Phi(\mathbf{x}_2,\omega) \right.\right.$$
$$\left.\left. \times \int d\omega' \Phi(\mathbf{x}_1,\omega')\Phi^*(\mathbf{x}_2,\omega') \right]^{1/2} \right\rangle$$
$$= 1 - \sum_{i,j} A_{ij}\langle x_i x_j\rangle + O(\langle\mathbf{x}^3\rangle), \quad (10)$$

where we have used that $\mathbf{x}$ is real and defined a symmetric (necessarily positive) matrix given by the photon sources:

$$A_{ij} = -\mathrm{Re}\left\{ \int d\omega \Phi^*(\mathbf{0},\omega) \frac{\partial^2\Phi(\mathbf{0},\omega)}{\partial x_i \partial x_j} \right.$$
$$\left. + \int d\omega \Phi(\mathbf{0},\omega) \frac{\partial\Phi^*(\mathbf{0},\omega)}{\partial x_i} \int d\omega' \Phi^*(\mathbf{0},\omega') \frac{\partial\Phi(\mathbf{0},\omega')}{\partial x_j} \right\}. \quad (11)$$

One important relation can be also established between g_k and $\langle\mathcal{F}\rangle$ for small mode mismatch. Indeed, the single-particle density matrix (6) has the following expansion in power series of $\mathbf{x}$:

$$\rho = |\Phi(\mathbf{0})\rangle\langle\Phi(\mathbf{0})| - \sum_{ij} \mathcal{A}_{ij}\langle x_i x_j\rangle + O(\langle\mathbf{x}^3\rangle), \quad (12)$$

where the operator $\mathcal{A}_{ij}$ reads

$$\mathcal{A}_{ij} = -\frac{1}{2}\left[|\Phi(\mathbf{0})\rangle \left\langle \frac{\partial^2\Phi(\mathbf{0})}{\partial x_i \partial x_j} \right| + \left| \frac{\partial^2\Phi(\mathbf{0})}{\partial x_i \partial x_j} \right\rangle \langle\Phi(\mathbf{0}) \right]$$
$$- \left| \frac{\partial\Phi(\mathbf{0})}{\partial x_i} \right\rangle \left\langle \frac{\partial\Phi(\mathbf{0})}{\partial x_j} \right|. \quad (13)$$

Then, utilizing Eq. (7), noticing that $\mathrm{Re}[\langle\Phi(\mathbf{0})|\mathcal{A}_{ij}|\Phi(\mathbf{0})\rangle] = A_{ij}$ defined in Eq. (11), and comparing with Eq. (10) the following important relation is established: $g_k = 1 - k(1 - \langle\mathcal{F}\rangle) + O(\langle\mathbf{x}^3\rangle)$, i.e., for a small mode mismatch, the k-photon distinguishability parameter $1 - g_k$ is k times the mode mismatch [defined here as the deviation of the average fidelity $\langle\mathcal{F}\rangle$ of Eq. (9) from 1].

One important model, considering nondeterministic sources, is of photons with random arrival times τ (equivalently, random phases), where $\Phi(\tau,\omega) = \phi(\omega)e^{i\omega\tau}$ (we

[3] As follows from Eq. (3), such a deviation inserts small coefficients at the products of quantum amplitudes $\prod_{\alpha=1}^{N} U^*_{k_\alpha,l_\alpha} U_{n_\alpha,l_\alpha}$ of the ideal BS, thus reducing the number of effectively contributing terms (for fixed $k_1,\ldots,k_N$), which varies from 1 ($n_\alpha = k_\alpha$, the classical case) to $N!$ (the ideal BS case).

set $\langle\tau\rangle = 0$). Let us denote the standard deviation (i.e., dispersion) of the arrival times by $\Delta\tau$, that of the frequency by $\Delta\omega$ [under the spectral density $|\phi(\omega)|^2$], and introduce the classicality parameter $\eta = \Delta\omega\Delta\tau$ (for $\eta = 0$ we recover the BS of Aaronson and Arkhipov, and for $\eta = \infty$ the classical case). Then we obtain $\langle\mathcal{F}\rangle = 1 - \eta^2 + O(\eta^4)$. Similarly, we also have $g_k(\eta) = 1 - k\eta^2 + O(k^2\eta^4)$, giving $J(\sigma) = 1 - [N - C_1(\sigma)]\eta^2 + O(N^2\eta^4)$. These expressions for a small mismatch follow also from the general case, where one can identify $\eta^2 = \sum_{i,j} A_{ij}\langle x_i x_j\rangle$ and A_{ij} defined in Eq. (11) [however, generally, the order of the next term is $O(\langle\mathbf{x}^3\rangle)$, whereas the absence of the third-order term for the random-arrival-times model is due to a single fluctuating parameter τ and the fact that $\langle\mathcal{F}\rangle$ and g_k are symmetric with respect to permutations of the integration variables τ_i and only their differences $\tau_i - \tau_j$ enter the definitions]. Thus one can think of $[\sum_{i,j} A_{ij}\langle x_i x_j\rangle]^{1/2}$ as an analog of the classicality parameter in the general case (at least for a small mode mismatch).

III. THE NONDETERMINISTIC BOSON SAMPLER AND A CLASSICALLY HARD COMPUTATIONAL TASK

The hardness result of Aaronson and Arkhipov [1] is formulated for the Haar random network matrix U in the dilute limit (defined here as $M \gg N^2$), assuring that the submatrices of such a random matrix are approximated by matrices with the elements being independent and identically distributed (i.i.d.) Gaussians with $\langle U_{kl}\rangle = 0$ and $\langle|U_{kl}|^2\rangle = \frac{1}{M}$ (since $\sum_{l=1}^{M}|U_{kl}|^2 = 1$). The distribution density of elements of U factorizes in this approximation and is given by[4]

$$p(U_{kl}) = \frac{M}{\pi}\exp\{-M|U_{kl}|^2\}. \tag{14}$$

The dilute limit is also essential for practical implementation, since one can use the simplest on-off (bucket) photon detectors, because of the vanishing probability of multiphoton detection at the output modes, due to the "boson birthday paradox" [1,37], now experimentally verified [38], which is similar to the classical birthday paradox. Therefore, we can restrict ourselves to the output occupation numbers $m_l \in \{0,1\}$, introducing $l_1,\ldots,l_N$ as the *distinct* output modes [denoting $\vec{l} \equiv (l_1,\ldots,l_N)$, etc.] and setting $P_\eta(\vec{l}|\vec{k})$ to be the corresponding output probability. Note that the sum of probabilities of the bunched outputs is small on average over the Haar measure, being on the order of $O(N^2/M)$ [1].

The main result of Aaronson and Arkhipov [1] states that approximation of the ideal BS cannot be performed on a classical computer with only polynomial resources in the total number of photons N and the inverse of the approximation error. The approximation error ε is the variational distance of the output distributions between the ideal BS case, $\mathcal{D}_0$, and the proposed approximation, $\mathcal{D}_1$. In our case, the above means that the NDBS is classically hard to simulate in polynomial time in $(N,1/\varepsilon)$ if, for a Haar random network matrix U, its output distribution $\mathcal{D}_\eta$ on the single-photon outputs is variationally close to that of the ideal BS, i.e.,

$$||\mathcal{D}_0 - \mathcal{D}_\eta||' \equiv \frac{1}{2}\sum_{\vec{l}}|P_0(\vec{l}|\vec{k}) - P_\eta(\vec{l}|\vec{k})| \leqslant c\varepsilon, \tag{15}$$

for some fixed constant c. Indeed, the (average in the Haar measure) probability of having a bunched output is vanishing as $O(N^2/M)$; thus the correction to the variational distance, i.e., the difference between the complete and the nonbunched outputs, satisfies (on average) $||\mathcal{D}_0 - \mathcal{D}_\eta|| - ||\mathcal{D}_0 - \mathcal{D}_\eta||' = O(N^2/M) \ll 1$.

The main point of the arguments in Ref. [1] is that an approximation of the BS computer as that described above also solves some computational task impossible to solve on a classical computer. Specifically, it was shown that such a classical simulation would imply also approximation of the permanents of matrices of Gaussian i.i.d. complex random variables with only polynomial resources, which is conjectured to be impossible (some numerical and other evidence is provided). Below, we will use one of the equivalent formulations of the latter computational task, namely, the problem of approximating the probability of the ideal BS to within an additive error $\pm\varepsilon\langle P_0(\vec{l}|\vec{k})\rangle = \pm\varepsilon\frac{N!}{M^N}$, where the average with respect to the Haar measure is computed using the Gaussian approximation (14) (under the Gaussian approximation, this problem is equivalent to $|\mathrm{GPE}|^2_\pm$ of Ref. [1]). Let us formulate it in precise terms.

$|BS|^2_\pm$ *problem*. For the *ideal* BS computer with a Haar random $(M \times M)$-dimensional unitary network matrix U and N single photons at the input, given small parameters ε and δ, simulate the output probability $P_0(\vec{l}|\vec{k})$ to within the additive error $\pm\varepsilon\frac{N!}{M^N}$, with success probability (in the Haar measure) at least $1-\delta$, in a polynomial in $(N,1/\varepsilon,1/\delta)$ time.

Using the Gaussian approximation and the boson birthday paradox we show below that, under a condition on the mode mismatch, the NDBS does exactly what is asked in the $|\mathrm{BS}|^2_\pm$-problem, i.e., what the classical computer cannot do. We employ Chebyshev's probability inequality [39], stating that for a random variable X with $\langle X\rangle = 0$, the probability $\mathcal{P}(|X|/\sqrt{\langle X^2\rangle} \geqslant 1/s) \leqslant s^2$, for any $s > 0$. Using the facts that the U_{kl} are i.i.d. random variables with the probability density (14), that $J(I) = 1$ (I is the identity permutation), and $\langle U_{kl}\rangle = 0$ we obtain from Eqs. (3) and (4)

$$\begin{aligned}\langle P_0 - P_\eta\rangle &= \sum_{\sigma_1,\sigma_2}[1 - J(\sigma_{21})]\left\langle\prod_{\alpha=1}^{N} U^*_{k_{\sigma_1(\alpha)},l_\alpha}U_{k_{\sigma_2(\alpha)},l_\alpha}\right\rangle \\ &= \sum_{\sigma_1,\sigma_2}[1 - J(\sigma_{21})]\delta_{\sigma_1,\sigma_2}\prod_{\alpha=1}^{N}\left\langle\left|U_{k_{\sigma_1(\alpha)},l_\alpha}\right|^2\right\rangle = 0.\end{aligned} \tag{16}$$

[4]It is proven that for $M \geqslant (N^5/\epsilon)\log^2(N/\epsilon)$ the Haar probability density p_H satisfies $p_H(X) \leqslant [1 + O(\epsilon)]p(X)$, where p is the probability density of Eq. (14), but a similar relation is expected to be valid for $M \gg N^2$ [1].

Similarly, after more involved calculations (see Appendix B), we get

$$\langle (P_0 - P_\eta)^2 \rangle = \sum_{\sigma_1,\sigma_2} \sum_{\sigma_1',\sigma_2'} [1 - J(\sigma_{21})][1 - J(\sigma_{21}')] \times \left\langle \prod_{\alpha=1}^{N} U^*_{k_{\sigma_1(\alpha)},l_\alpha} U_{k_{\sigma_2(\alpha)},l_\alpha} U^*_{k_{\sigma_1'(\alpha)},l_\alpha} U_{k_{\sigma_2'(\alpha)},l_\alpha} \right\rangle = \left(\frac{N!}{M^N}\right)^2 \frac{1}{N!} \sum_{\sigma} \chi(C_1(\sigma)) [1 - J(\sigma)]^2 , \tag{17}$$

where we have defined $\chi(n) = n! \sum_{k=0}^{n} \frac{1}{k!} = \int_1^\infty dz\, z^n e^{1-z}$. Let us introduce a rescaled variance

$$\mathcal{V}(N,\eta) = \frac{1}{N!} \sum_{\sigma} \chi(C_1(\sigma)) [1 - J(\sigma)]^2 . \tag{18}$$

Now, the inequality complementary to Chebyshev's reads (for $\varepsilon > 0$)

$$\mathcal{P}\left(|P_0 - P_\eta| < \varepsilon \frac{N!}{M^N} \right) > 1 - \frac{\mathcal{V}(N,\eta)}{\varepsilon^2}, \tag{19}$$

where Eqs. (17) and (18) were used. Equation (19) resembles the statement of the $|\mathrm{BS}|^2_\pm$ problem: If we are able to control the cycle sum $\mathcal{V}(N,\eta)$, e.g., by varying the classicality parameter η, such that the right-hand side in Eq. (19) stays close to 1 then the NDBS, with success probability close to 1, approximates the ideal BS of Aaronson and Arkhipov to within an additive error (in the required form). Let us now formalize this statement. *Given an error ε and a success probability $1 - \delta$, if the rescaled variance $\mathcal{V}(N,\eta)$ (18) observes the bound*

$$\mathcal{V}(N,\eta) \leqslant \varepsilon^2 \delta, \tag{20}$$

then the NDBS solves the $|BS|^2_\pm$ problem, i.e., it performs a computational task which cannot be simulated on a classical computer with only polynomial resources. Equation (20) is a sufficient condition which may not be necessary for the NDBS to outperform classical computers, since Chebyshev's inequality can be a crude approximation. However, it usually captures the scaling of the tail probability of a random variable in terms of its variance. Equation (20) states that the N scaling of the minimal approximation error with which the NDBS satisfies the $|\mathrm{BS}|^2_\pm$ problem is defined by the rescaled variance $\mathcal{V}(N,\eta)$.

Equation (20) involves the cycle sum (18), computable only numerically for each particular density matrix ρ, depending on the sources. Let us analyze in detail the model of single photons with random arrival times discussed above, taking both $\Phi(\tau,\omega) = \phi(\omega)e^{i\omega\tau}$ and $p(\tau)$ to be Gaussian distributions, e.g., spectrally shaped by the stimulated Raman technique of Ref. [40] with the Gaussian distributed random arrival times (centered at $\tau = 0$):

$$\Phi(\tau,\omega) = \frac{1}{\sqrt{2\pi}\Delta\omega} \exp\left(i\omega\tau - \frac{(\omega - \omega_0)^2}{2\Delta\omega^2} \right), \tag{21}$$

$$p(\tau) = \frac{1}{\sqrt{2\pi}\Delta\tau} \exp\left(-\frac{\tau^2}{2\Delta\tau^2} \right). \tag{22}$$

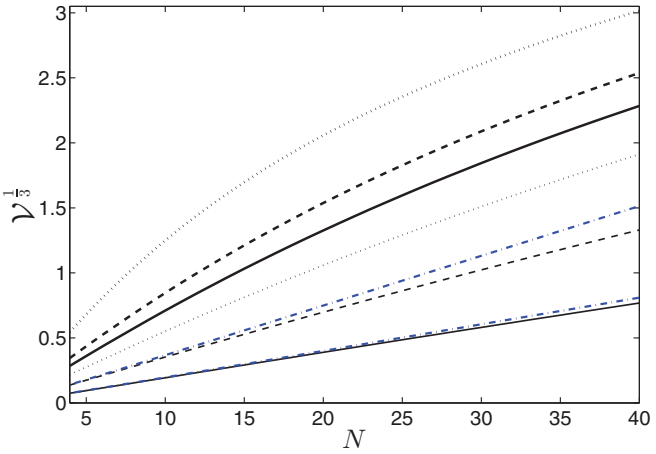

FIG. 2. (Color online) Behavior of the cubic root of the reduced variance $\mathcal{V}(N,\eta)$ for several values of the two-photon indistinguishability parameter g_2 (from bottom to top): $g_2 = 0.99$ (thin solid line), $g_2 = 0.975$ (thin dashed line), $g_2 = 0.95$ (thin dotted line), $g_2 = 0.925$ (thick solid line), $g_2 = 0.9$ (thick dashed line), and $g_2 = 0.8$ (thick dotted line). We have used the Gaussian model of single photons with random arrival times. The two dash-dotted lines give the approximation following from Eq. (25).

In this case, all integrals in Eq. (5) are Gaussian and can be evaluated. Such a model also is interesting from the point of view of practical optimality, since as shown in Ref. [41], the Gaussian-shaped form of single photons is optimal for interference experiments. Setting $\gamma = \frac{2\eta^2}{1+2\eta^2}$, we obtain g_k as a positive monotonically decreasing function of γ (and, hence, of η^2):

$$g_k = (1 - \gamma)^{k/2}(1 - \gamma^k)^{-1/2}. \tag{23}$$

Elementary algebra gives

$$J(\sigma) = (1 - \gamma)^{\frac{N}{2}} \prod_{k=1}^{N} (1 - \gamma^k)^{-C_k(\sigma)/2}. \tag{24}$$

In this case, one can also express g_k and, hence, J as functions of g_2 only, since g_2^2 and γ are Möbius transformations of each other. We have $\gamma = (1 - g_2^2)/(1 + g_2^2)$ [and $\eta^2 = (g_2^{-2} - 1)/2$]. Moreover, $g_2 = \langle \mathcal{F} \rangle / \sqrt{2 - \langle \mathcal{F} \rangle^2}$. For this model, the results are presented in Fig. 2, where we plot the cubic root of $\mathcal{V}(N,\eta)$.

For a small two-photon distinguishability $1 - g_2 \approx 2\eta^2 \ll 1$ (i.e., for a small mode mismatch), the dependence of $\mathcal{V}^{1/3}(N,\eta)$ on N in Fig. 2 is approximately a linear function. This is a general feature. Indeed, as shown above, $g_k(\eta) \approx 1 - k\eta^2$ for $\eta \ll 1$ and $J(\sigma) \approx 1 - \eta^2[N - C_1(\sigma)]^2$. Inserting this into the definition of $\mathcal{V}(N,\eta)$ and taking the integral over z in the resulting expression [coming from the integral representation of $\chi(C_1)$ in Eq. (18)] we get after elementary algebra

$$\mathcal{V}(N,\eta) \approx \eta^4 \left(\frac{N^3}{3} - \frac{N^2}{2} + \frac{7N}{6} - 1 \right). \tag{25}$$

Equation (25) for $N \gg 1$ reveals the scaling $\mathcal{V}(N,\eta) \approx \eta^4 N^3/3 \approx (1 - \langle \mathcal{F} \rangle)^2 N^3/3$. Therefore, the photon mode mismatch $(1 - \langle \mathcal{F} \rangle)$ must scale approximately as $N^{-3/2}$ in the

total number of photons, if the NDBS is to be scaled up while keeping the product $\varepsilon^2\delta$ constant (i.e., at the same level of hardness of the classical simulation). As seen from Fig. 2, the approximation (25) deviates from the exact result for sufficiently large N, where the contribution from the higher-order terms $\sim\eta^p$, $p > 2$, becomes important. Such higher-order terms are model specific and thus cannot be obtained in a general form. The optimality of the Gaussian model suggests that Fig. 2 shows the optimal instance of the bound (20).

Our main result (20) provides also a sufficient condition for approximation of the BS by the NDBS in the variational distance, i.e., as in Eq. (15), but for a fraction $1 - \frac{\mathcal{V}(N,\eta)}{4\varepsilon^2}$ of the network matrices U. Indeed, the one-norm (known in probability theory as the variational distance) is bounded as $||\mathcal{D}_0 - \mathcal{D}_\eta||^2 \leqslant \frac{1}{4}(\sum_{\vec{l}} 1)\sum_{\vec{l}}[P_0(\vec{l}|\vec{k}) - P_\eta(\vec{l}|\vec{k})]^2$. Using this upper bound and applying Chebyshev's inequality to $||\mathcal{D}_0 - \mathcal{D}_\eta||'$ of Eq. (15) considered as a random variable on the Haar measure, we get

$$\mathcal{P}(||\mathcal{D}_0 - \mathcal{D}_\eta||' < \varepsilon) > 1 - \frac{\mathcal{V}(N,\eta)}{4\varepsilon^2}. \tag{26}$$

An experimental demonstration of NDBS operation beyond the power of classical computers could proceed by showing that, for a randomly chosen network matrix, the NDBS with a fixed mode mismatch approximates the output probabilities of the ideal BS of Aaronson and Arkhipov to within an error $\pm\varepsilon\frac{N!}{M^N}$, i.e., it solves the computational task specified in the $|\mathrm{BS}|^2_{\pm}$ problem, where the product of the squared error ε^2 and the failure probability δ (i.e., the Haar measure of the excluded network matrices) is at least equal to the reduced variance $\mathcal{V}(N,\eta)$. The probabilities for the ideal BS computer can still be obtained for $N \sim 20$ by numerical simulations.

IV. CONCLUSION

In conclusion, we have considered a nondeterministic model of the BS computer, the NDBS, which generalizes the ideal BS computer of Aaronson and Arkhipov [1] and captures the essential features of a realistic BS device with only partially indistinguishable single photons at the input. If the average mutual fidelity of the single photons satisfies the derived N-dependent bound, the NDBS device cannot be efficiently simulated on a classical computer. The sufficient condition derived in this work may not be necessary for the NDBS to be hard to simulate classically; however, it reveals the inverse 3/2-power-law scaling of the photon mode mismatch on the total number of photons for scalability of the NDBS computer at the same level of hardness of the classical simulation (i.e., for the constant approximation error and fixed success probability with which the NDBS approximates the ideal BS in the variational distance). Moreover, the results are also applicable to any other realization of the BS with identical single-photon sources, for instance, with the Gaussian input states, proposed recently in Ref. [42], where the imperfect indistinguishability of heralded single photons can be treated in a similar way.

We have studied the so-called "dilute limit" of a unitary M-mode network with N bosons, i.e., with $M \gg N^2$, for which the classical hardness is established, and when the average probability (over the random network matrices in the Haar measure) of two bosons landing at the same output mode vanishes as $O(\frac{N^2}{M})$. One might wonder why then the output probability distribution of bosons is exponentially harder to compute than that of fermions in a similar setup. Since this question belongs to the field of computational complexity theory, the answer must be formulated in its terms: Bosonic amplitudes are given by matrix permanents, while fermonic ones are given by matrix determinants, where the permanent requires a computational time exponential in N, whereas the determinant is known to be polynomial in N.

However, a physicist can be left unsatisfied by the permanent vs determinant explanation, although it is absolutely correct, and try to inquire further: What specific feature of the bosonic statistics could be held responsible for this drastic difference, especially since the output rarely contains two bosons at the same mode? One plausible candidate is the very same bosonic bunching, which is unimportant at the output, but not *during* the propagation in the network. Indeed, let us compare bosonic and fermionic propagation through a unitary network, bringing the two cases to a common ground by decomposing the unitary map between the input and output Fock states into a product of infinitesimal unitary maps, i.e., using a Feynman-type sum over the paths, but now in the Fock space. Such an expansion involves summation over all intermediate occupation numbers and each term is a product of permanents (bosons) or determinants (fermions). In both cases, each factor in the product of amplitudes becomes easily computable for an infinitesimal unitary map (to a sufficient approximation) when the number of factors becomes sufficiently large. But, as soon as the number of infinitesimal maps in the product grows above the ratio M/N^2, it is necessary to sum over the multiple occupation numbers for bosons, i.e., bosonic bunching contributes in the intermediate Fock states, whereas in the fermionic case the occupation numbers remain bounded by 1. In the limit when the Feynman-type expansion becomes exact, one recovers full bosonic bunching as allowed by their statistics, while at the output it is still negligible. Therefore, reformulating Wigderson's famous joke slightly [1], we can conclude by saying that to arrive at the same output configuration as fermions, bosons have a much harder job indeed, since they must go along a much larger set of paths.

ACKNOWLEDGMENTS

This work was supported by the CNPq of Brazil. The author is indebted to Scott Aaronson for helpful comments.

APPENDIX A: DERIVATION OF THE PROBABILITY FORMULA

We consider the case of single photons which are emitted by identical photon sources and launched into distinct modes $k_1, \ldots, k_N$ of an M-mode linear optical network with the unitary matrix U relating the input $a_k(\omega)$ and output $b_l(\omega)$ modes of frequency ω, $a_k^\dagger(\omega) = \sum_{l=1}^M U_{kl} b_l^\dagger(\omega)$. The input state originating from a set of N independent identical sources

of single photons is given by the density matrix

$$\rho^{(in)} = \int d\mathbf{x}_1 \cdots \int d\mathbf{x}_N \prod_{\alpha=1}^{N} p(\mathbf{x}_\alpha) \times |\Psi(\mathbf{x}_1,\ldots,\mathbf{x}_N)\rangle\langle\Psi(\mathbf{x}_1,\ldots,\mathbf{x}_N)|, \quad \text{(A1)}$$

where the Fock state $|\Psi(\mathbf{x}_1,\ldots,\mathbf{x}_N)\rangle$ is given in Eq. (1) of Sec. II. The probability of detecting $m_1,\ldots,m_M$ photons in the output modes described by the annihilation operators $b_1(t),\ldots,b_M(t)$ can be derived by the standard quantum photon counting theory [30–32,43]. It is in the form of an average on the density matrix (A1)

$$p_{m_1,\ldots,m_M} = \frac{1}{\prod_{l=1}^{M} m_l!} \left\langle : \prod_{l=1}^{M} \mathcal{I}_l^{m_l} \exp\left\{ -\sum_{l=1}^{M} \mathcal{I}_l \right\} : \right\rangle, \quad \text{(A2)}$$

where the double dots denote the time and normal ordering of the creation and annihilation operators, and the detection operator reads

$$\mathcal{I}_l = \int_t^{t+\Delta t} d\tau \int_t^{t+\Delta t} d\tau' G(\tau-\tau') b_l^\dagger(\tau) b_l(\tau') \quad \text{(A3)}$$

with the detector efficiency described by the function $G(t)$. In our case, the initial state is a Fock state of N single photons in distinct modes and we postselect on the cases when all N photons are detected, $\sum m_l = N$. In this case the exponent in Eq. (A2) does not contribute. Substituting the Fourier expansions

$$b_l(t) = \int_0^\infty \frac{d\omega}{\sqrt{2\pi}} e^{-i\omega t} b_l(\omega) \quad \text{(A4)}$$

and (see, for instance, Ref. [43])

$$G(t) = \int_0^\infty \frac{d\omega}{2\pi} e^{-i\omega t} \Gamma(\omega), \quad \Gamma(\omega) > 0, \quad \text{(A5)}$$

in Eq. (A2), inserting the projector into the vacuum $|0\rangle\langle 0|$ between the creation and annihilation operators (since all photons are detected this changes nothing), and integrating over the times, we obtain that the probability is given by the average of the following operator:

$$\Pi(m_1,\ldots,m_M) = \frac{1}{\prod_{l=1}^{M} m_l!} \int_0^\infty d\omega_1 \cdots \int_0^\infty d\omega_N \times \prod_{l=1}^{N} \Gamma(\omega_l) \left[\prod_{\alpha=1}^{N} b_{l_\alpha}^\dagger(\omega_\alpha) \right] |0\rangle\langle 0| \left[\prod_{\alpha=1}^{N} b_{l_\alpha}(\omega_\alpha) \right], \quad \text{(A6)}$$

where the combined index $(l_1,\ldots,l_N)$ (the order being insignificant) is the set $\{1,\ldots,1,2,\ldots,2,\ldots,M,\ldots,M\}$ with index k appearing m_k times. Thus, the output probability of detecting $m_1,\ldots,m_M$ photons in modes $1,\ldots,M$ becomes

$$P(m_1,\ldots,m_M|k_1,\ldots,k_N) = \int d\mathbf{x}_1 \cdot \ldots \cdot \int d\mathbf{x}_N \prod_{\alpha=1}^{N} p(\mathbf{x}_\alpha) \times \langle\Psi(\mathbf{x}_1,\ldots,\mathbf{x}_N)|\Pi(m_1,\ldots,m_M)|\Psi(\mathbf{x}_1,\ldots,\mathbf{x}_N)\rangle. \quad \text{(A7)}$$

The operators $\Pi(m_1,\ldots,m_M)$ are positive Hermitian, but they generally do not sum up to the identity operator (more precisely, to the projector on the symmetric subspace of N bosons) for $m_1+\cdots+m_M = N$. However, for efficient detectors, when all output photons are detected, after a suitable normalization (see below) $\Pi(m_1,\ldots,m_M)$ become the POVM elements realizing the detection described above. In this case the probabilities in Eq. (A7) sum to 1 under the constraint $m_1+\cdots+m_M = N$. Using the evolution in the unitary network

$$a_k^\dagger(\omega) = \sum_{l=1}^{M} U_{kl} b_l^\dagger(\omega) \quad \text{(A8)}$$

and the identity

$$\langle 0| \left[\prod_{\alpha=1}^{N} b_{l'_\alpha}(\omega'_\alpha) \right] \left[\prod_{\alpha=1}^{N} b_{l_\alpha}^\dagger(\omega_\alpha) \right] |0\rangle = \sum_\sigma \delta_{l'_\alpha, l_{\sigma^{-1}(\alpha)}} \delta\left(\omega'_\alpha - \omega_{\sigma^{-1}(\alpha)}\right), \quad \text{(A9)}$$

where σ is a permutation, we obtain [transferring the permutations $\sigma_{1,2}$ from the two inner products, as in Eq. (A9), to the k indices]

$$P(m_1,\ldots,m_M|k_1,\ldots,k_N) = \frac{1}{\prod_{l=1}^{M} m_l!} \sum_{\sigma_1} \sum_{\sigma_2} J\left(\sigma_2\sigma_1^{-1}\right) \prod_{\alpha=1}^{N} U^*_{k_{\sigma_1(\alpha)}, l_\alpha} U_{k_{\sigma_2(\alpha)}, l_\alpha}, \quad \text{(A10)}$$

with J given as follows (the product is a shortcut notation for multiple integration over ω_α and $\mathbf{x}_\alpha$):

$$J(\sigma) = \prod_{\alpha=1}^{N} \int d\mathbf{x}_\alpha\, p(\mathbf{x}_\alpha) \int_0^\infty d\omega_\alpha \Gamma(\omega_\alpha) \times \phi^*(\mathbf{x}_\alpha,\omega_\alpha)\phi\left(\mathbf{x}_\alpha,\omega_{\sigma^{-1}(\alpha)}\right). \quad \text{(A11)}$$

Here we have used the symmetry of the multiple integral under permutation of the integration variables, reassigning the variables as $\omega_\alpha \equiv \omega_{\sigma_1^{-1}(\alpha)}$ and defining $\sigma \equiv \sigma_2\sigma_1^{-1}$.

The structure of the integrals in Eq. (A11) makes J factorize into a product of similar functions depending on the cycles from the cyclic decomposition of the permutation σ (since each of the two multiple integrals, one over ω_α and one over $\mathbf{x}_\alpha$, factorizes). Moreover, because of the above-mentioned permutational symmetry of the integration variables, cycles with the same number of elements contribute the same factor. Therefore we obtain

$$J(\sigma) = \prod_{k=2}^{N} g_k^{C_k(\sigma)}, \quad \text{(A12)}$$

$$g_k \equiv \prod_{\alpha=1}^{k} \int_0^\infty d\omega_\alpha \int d\mathbf{x}_\alpha\, p(\mathbf{x}_\alpha)\Phi(\mathbf{x}_\alpha,\omega_{\alpha-1})\Phi^*(\mathbf{x}_\alpha,\omega_\alpha), \quad \text{(A13)}$$

where the index α is cyclic ($\alpha = 0$ is $\alpha = k$), C_k is the number of cycles of length k, with $\sum kC_k = N$ [33], and $\Phi(\mathbf{x},\omega) \equiv \sqrt{\Gamma(\omega)}\phi(\mathbf{x},\omega)$.

APPENDIX B: DERIVATION OF THE EXPRESSION FOR THE VARIANCE OF $P_0 - P_\eta$

We have for the variance

$$\langle (P_0 - P_\eta)^2 \rangle = \sum_{\sigma,\tilde{\sigma}} \sum_{\sigma_R,\tilde{\sigma}_R} [1 - J(\tilde{\sigma}_R)][1 - J(\sigma_R)] \times \prod_{\alpha=1}^{N} \left\langle U^*_{k_{\sigma(\alpha)},l_\alpha} U_{k_{\sigma_R\sigma(\alpha)},l_\alpha} U^*_{k_{\tilde{\sigma}(\alpha)},l_\alpha} U_{k_{\tilde{\sigma}_R\tilde{\sigma}(\alpha)},l_\alpha} \right\rangle, \tag{B1}$$

where we have introduced the relative permutations σ_R and $\tilde{\sigma}_R$ and taken into account the mutual independence of U_{k_β,l_α} for the set of distinct indices $l_1,\ldots,l_N$. The nonzero terms in the sum over all permutations in Eq. (B1) occur under the condition that for any $\alpha \in \{1,\ldots,N\}$ either of the two sets of equations below is satisfied:

$$\sigma_R\sigma(\alpha) = \sigma(\alpha), \quad \tilde{\sigma}_R\tilde{\sigma}(\alpha) = \tilde{\sigma}(\alpha), \tag{B2}$$

$$\sigma_R\sigma(\alpha) = \tilde{\sigma}(\alpha), \quad \tilde{\sigma}_R\tilde{\sigma}(\alpha) = \sigma(\alpha). \tag{B3}$$

For each choice of the permutations $\{\sigma,\tilde{\sigma},\sigma_R,\tilde{\sigma}_R\}$ denote the ordered (in some way) set of all α satisfying Eq. (B2) as $\alpha^{(I)}$ and the ordered set of the rest of the indices as $\alpha^{(II)}$ [these satisfy Eq. (B3)] (the two ordered sets give an ordered partition of the set of all indices $\{1,\ldots,N\}$). Introduce also the ordered sets $\beta^{(I)}$ and $\beta^{(II)}$ and their versions with the tilde, $\tilde{\beta}^{(I)}$ and $\tilde{\beta}^{(II)}$, as the result of the action of σ (respectfully, $\tilde{\sigma}$) on the sets $\alpha^{(I)}$ and $\alpha^{(II)}$, i.e., by $\beta_j = \sigma(\alpha_j)$ and $\tilde{\beta}_j = \tilde{\sigma}(\alpha_j)$. Each β set and its version with the tilde is a permutation of the other: $\tilde{\beta}^{(I,II)}_j = \tilde{\sigma}\sigma^{-1}(\beta^{(I,II)}_j)$. Equation (B2) states that $\beta^{(I)}_j$ and $\tilde{\beta}^{(I)}_j$, $j = 1,\ldots,|\alpha^{(I)}|$, are fixed points (i.e., one-cycles) of the permutations σ_R and $\tilde{\sigma}_R$, respectfully (thus the sets of their fixed points coincide). Equation (B3) states that $\tilde{\sigma}_R$ is inverse to σ_R acting on $\beta^{(II)}$, i.e., $\sigma_R(\beta^{(II)}_j) = \tilde{\beta}^{(II)}_j$ and $\tilde{\sigma}_R(\tilde{\beta}^{(II)}_j) = \beta^{(II)}_j$, $j = 1,\ldots,|\alpha^{(II)}|$. From these facts the necessary conditions for nonzero contribution in Eq. (B1) follow:

$$\tilde{\sigma}_R = \sigma_R^{-1}, \quad \tilde{\sigma} = (\tau_1 \otimes I_2)\sigma_R\sigma, \tag{B4}$$

where τ_1 is an *arbitrary* permutation of the set $\beta^{(I)}$ and I_2 is the identity permutation of the set $\beta^{(II)}$. Note also that the number of all indices $\alpha^{(I)}$ satisfies $|\alpha^{(I)}| = C_1(\sigma_R)$, where C_1 is the number of one-cycles (fixed points) of the permutation. Let us now use Eqs. (B2), (B3), and (B4) in Eq. (B1). Under the Gaussian approximation in Eq. (14) of Sec. III $\langle |U_{kl}|^2 \rangle = 1/M$ and $\langle |U_{kl}|^4 \rangle = 2/M^2$. Hence, we obtain for $\alpha \in \alpha^{(I)}$:

$$\prod_{\alpha\in\alpha^{(I)}} \left\langle \left|U_{k\sigma(\alpha),l_\alpha}\right|^2 \left|U_{k\tilde{\sigma}(\alpha),l_\alpha}\right|^2 \right\rangle = \left(\frac{1}{M}\right)^{2[|\alpha^{(I)}| - C_1(\tilde{\sigma}\sigma^{-1})]} \left(\frac{2}{M^2}\right)^{C_1(\tilde{\sigma}\sigma^{-1})} = 2^{C_1(\tau_1)} \left(\frac{1}{M}\right)^{2C_1(\sigma_R)}, \tag{B5}$$

where we have taken into account that, since all fixed point of σ_R are in $\beta^{(I)}$, all fixed points of $\tilde{\sigma}\sigma^{-1}$ belong to the set $\beta^{(I)}$ and are also fixed points of τ_1. Hence, using that $|\alpha^{(II)}| = N - |\alpha^{(I)}| = N - C_1(\sigma_R)$, for $\alpha \in \alpha^{(II)}$ we obtain

$$\prod_{\alpha\in\alpha^{(II)}} \left\langle \left|U_{k\sigma(\alpha),l_\alpha}\right|^2 \left|U_{k\tilde{\sigma}(\alpha),l_\alpha}\right|^2 \right\rangle = \left(\frac{1}{M}\right)^{2[N - C_1(\sigma_R)]}. \tag{B6}$$

Inserting the results of Eqs. (B5) and (B6) into Eq. (B1), performing the summation over the independent (free) permutations σ, σ_R, and τ_1, and using that $J(\sigma^{-1}) = J(\sigma)$ (since the inverse permutation has the same cycle structure) we obtain the following expression for the variance:

$$\langle (P_0 - P_\eta)^2 \rangle = \left(\frac{N!}{M^N}\right)^2 \frac{1}{N!} \sum_{\sigma_R} \chi(C_1(\sigma_R)) [1 - J(\sigma_R)]^2. \tag{B7}$$

Here $\chi(n)$ is the cycle sum

$$\chi(n) \equiv \sum_{\tau} 2^{C_1(\tau)} = n! \sum_{k=0}^{n} \frac{1}{k!} = \int_1^\infty dt\, t^n e^{1-t}, \tag{B8}$$

where τ is a permutation of n elements (see, for instance, Ref. [33]).

APPENDIX C: THE CYCLE SUM $\mathcal{V}(N,\eta)$

One can express the summation over the permutations in the definition of $\mathcal{V}$ (note that there are, in total, $N!$ terms)

$$\mathcal{V}(N,\eta) = \frac{1}{N!} \sum_{\sigma} \chi(C_1(\sigma)) \left[1 - \prod_{k=2}^{N} g_k^{C_k(\sigma)}(\eta)\right]^2, \tag{C1}$$

as a summation over all partitions of N into a sum of positive integers. Indeed, there are $N!/(\prod_{k=1}^N k^{C_k} C_k!)$ permutations with the cycle structure $(C_1,\ldots,C_N)$ (see, for instance, Ref. [33]), the summation is over the integer partitions of N into a sum of integers, from 1 to N, where each integer k corresponds to a cycle length in the cyclic structure of the permutation, while the multiplicity is C_k. We get

$$\mathcal{V} = \sum_{(C_1,\ldots,C_N)} \frac{\chi(C_1)\left(1 - \prod_{k=2}^N g_k^{C_k}\right)^2}{\prod_{k=1}^N k^{C_k} C_k!}, \tag{C2}$$

where the summation is under the constraint that $\sum_{k=1}^N kC_k = N$. The sum in Eq. (C2) can be efficiently calculated numerically, if N is not very large.

[1] S. Aaronson and A. Arkhipov, Theory Comput. **9**, 143 (2013).
[2] See, for instance, the reviews: B. Lounis and M. Orrit, Rep. Prog. Phys. **68**, 1129 (2005); G. S. Buller and R. J. Collins, Meas. Sci. Technol. **21**, 012002 (2010); M. D. Eisaman, J. Fan, A. Migdall, and S. V. Polyakov, Rev. Sci. Instrum. **82**, 071101 (2011).
[3] C. K. Hong, Z. Y. Ou, and L. Mandel, Phys. Rev. Lett. **59**, 2044 (1987).
[4] Y. L. Lim and A. Beige, New J. Phys. **7**, 155 (2005).
[5] M. C. Tichy *et al.*, New J. Phys. **14**, 093015 (2012).
[6] P. Shor, in *Proceedings of the 35th Annual Symposium on Foundations of Computer Science* (IEEE Computer Society Press, Los Alamos, CA, 1994), p. 124; SIAM J. Comput. **26**, 1484 (1997).
[7] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, 2000).
[8] H. Minc, *Permanents, Encyclopedia of Mathematics and Its Applications* (Addison-Wesley, Reading, MA, 1978), Vol. 6.
[9] E. R. Caianiello, Nuovo Cimento **10**, 1634 (1953); *Combinatorics and Renormalization in Quantum Field Theory*, Frontiers in Physics, Lecture Note Series (W. A. Benjamin, Reading, MA, 1973).
[10] S. Scheel, arXiv:quant-ph/0406127.
[11] L. G. Valiant, Theor. Comput. Sci. **8**, 189 (1979).
[12] S. Aaronson, Proc. R. Soc. London, Ser. A **467**, 3393 (2008).
[13] H. Ryser, *Combinatorial Mathematics*, Carus Mathematical Monograph No. 14 (Wiley, New York, 1963).
[14] M. A. Broome *et al.*, Science **339**, 794 (2013).
[15] J. B. Spring *et al.*, Science **339**, 798 (2013).
[16] M. Tillmann *et al.*, Nat. Photonics **7**, 540 (2013).
[17] A. Crespi *et al.*, Nat. Photonics **7**, 545 (2013).
[18] D. Aharonov and M. Ben-Or, in *Proceedings of the ACM STOC*, p. 176 (ACM, New York, NY, 1997); D. S. Abrams and S. Lloyd, Phys. Rev. Lett. **79**, 2586 (1997).
[19] E. Knill, R. Laflamme, and W. Zurek, Science **279**, 342 (1998).
[20] A. M. Steane, Phys. Rev. A **68**, 042322 (2003).
[21] E. Knill, Nature (London) **434**, 39 (2005).
[22] P. P. Rohde and T. C. Ralph, Phys. Rev. A **85**, 022332 (2012).
[23] P. P. Rohde, Phys. Rev. A **86**, 052321 (2012).
[24] A. Leverrier and R. García-Patrón, arXiv:1309.4687 [quant-ph].
[25] M. Jerrum, A. Sinclair, and E. Vigoda, J. ACM **51**, 671 (2004).
[26] M. J. Collins *et al.*, Nat. Commun. **4**, 2582 (2013).
[27] E. B. Flagg, A. Muller, S. V. Polyakov, A. Ling, A. Migdall, and G. S. Solomon, Phys. Rev. Lett. **104**, 137401 (2010).
[28] R. B. Patel *et al.*, Nat. Photonics **4**, 632 (2010).
[29] Y.-M. He *et al.*, Nat. Nanotechnol. **8**, 213 (2013).
[30] L. Mandel, in *Progress in Optics*, edited by E. Wolf (North-Holland, Amsterdam, 1963), Vol. 2, p. 181.
[31] R. J. Glauber, *Optical Coherence and Photon Statistics* (Gordon & Breach, New York, 1965), p. 65.
[32] P. L. Kelley and W. H. Kleiner, Phys. Rev. **136**, A316 (1964).
[33] R. P. Stanley, *Enumerative Combinatorics*, 2nd ed. (Cambridge University Press, Cambridge, 2011), Vol. 1.
[34] R. A. Campos, B. E. A. Saleh, and M. C. Teich, Phys. Rev. A **40**, 1371 (1989); S.-H. Tan, Y. Y. Gao, H. de Guise, and B. C. Sanders, Phys. Rev. Lett. **110**, 113603 (2013).
[35] Z. Y. Ou, Phys. Rev. A **74**, 063808 (2006).
[36] P. P. Rohde, W. Mauerer, and C. Silberhorn, New J. Phys. **9**, 91 (2007).
[37] A. Arkhipov and G. Kuperberg, Geom. Topol. Monogr. **18**, 1 (2012).
[38] N. Spagnolo *et al.*, Phys. Rev. Lett. **111**, 130503 (2013).
[39] B. V. Gnedenko, *The Theory of Probability* (Mir Publishers, Moscow, 1978), p. 198 (English translation).
[40] A. Kuhn, M. Hennrich, and G. Rempe, Phys. Rev. Lett. **89**, 067901 (2002).
[41] P. P. Rohde, T. C. Ralph, and M. A. Nielsen, Phys. Rev. A **72**, 052332 (2005).
[42] A. P. Lund *et al.*, arXiv:1305.4346 [quant-ph].
[43] W. Vogel and D. G. Welsch, *Quantum Optics* (Wiley-VCH Verlag, Berlin, 2006), p. 173.