Non-linear Boson Sampling

Nicolò Spagnolo,1 Daniel J. Brod,2 Ernesto F. Galvão,3, 4 and Fabio Sciarrino1

1Dipartimento di Fisica, Sapienza Università di Roma, Piazzale Aldo Moro 5, I-00185 Roma, Italy
2Instituto de Física, Universidade Federal Fluminense, Niterói, RJ, 24210-340, Brazil
3International Iberian Nanotechnology Laboratory (INL), Ave. Mestre Jose Veiga, 4715-330, Braga, Portugal
4Instituto de Física, Universidade Federal Fluminense, Niterói, RJ, 24210-340, Brazil

Boson Sampling is a task that is conjectured to be computationally hard for a classical computer, but which can be efficiently solved by linear-optical interferometers with Fock state inputs. Significant advances have been reported in the last few years, with demonstrations of small- and medium-scale devices, as well as implementations of variants such as Gaussian Boson Sampling. Besides the relevance of this class of computational models in the quest for unambiguous experimental demonstrations of quantum advantage, recent results have also proposed first applications for hybrid quantum computing. Here, we introduce the adoption of non-linear photon-photon interactions in the Boson Sampling framework, and analyze the enhancement in complexity via an explicit linear-optical simulation scheme. By extending the computational expressivity of Boson Sampling, the introduction of non-linearities promises to disclose novel functionalities for this class of quantum devices. Hence, our results are expected to lead to new applications of near-term, restricted photonic quantum computers.

Introduction. – Quantum technologies promise to provide speed-up in several fields, ranging from intrinsically secure long-distance quantum communication [1] to a novel generation of high-precision sensors [2], and enhanced computational and simulation capabilities [3]. Among the currently developed experimental platforms, in the last few years photonic technologies have recently experienced a technological boost in all fundamental components, namely photon sources, manipulation and detection [4].

Recent studies have focused on identifying suitable dedicated classically-hard tasks, with the aim of leveraging the necessary technological resources and system size to reach the quantum advantage regime [5, 6]. Such regime corresponds to the scenario where a quantum device solves a given task faster than any classical counterpart. Within this context, a computational problem named Boson Sampling [7] has been defined as a promising approach. This problem, that consists in sampling from the output distribution of a system of $n$ non-interacting bosons undergoing linear evolution, is a classically-hard task (in $n$) while it can be naturally solved by a linear optical photonic system. Such sampling problem has also subsequently inspired other classes of sampling problems [8, 9] suitable to be solved with different quantum hardware [5, 10, 11].

Starting from the original proposal [7], several experimental implementations of Boson Sampling instances [12–23] and of recently proposed variants [24–28] have been reported [29]. In particular, hybrid algorithms based on Gaussian Boson Sampling have been proposed for various tasks: quantum simulation [27, 30–33], optimization problems [34], point processes [35] graph theory [36–38], and quantum optical neural networks [39].

Very recently, impressive experimental implementations of Gaussian Boson sampling have been reported [6, 40]. Besides the technological advances reported in the last few years, several studies have also focused on studying and improving classical simulation of Boson Sampling [41–46], and on defining the limits for simulability in the presence of imperfections, in particular losses and partial photon distinguishability [47–54]. All these studies aimed at establishing a classical benchmarking framework for Boson Sampling, and currently place the threshold for quantum advantage in such a system to $n \sim 50$ photons in a network composed by $m \sim n^2 = 2500$ optical modes. Recent improvements in photon sources [55–60] enabled first Boson Sampling experiments with a number of detected photons up to $n = 14$ [23]. However, reaching the quantum advantage regime with a photonic platform solving the original formulation of the task [7] still requires a technological leap to enhance single-photon generation rates and indistinguishability, and to reduce losses in the current platforms for linear-optical networks.

In this Letter, we introduce the adoption of non-linear interactions at the few-photon level within the Boson Sampling framework as a route to increase the complexity and reduce the threshold for the quantum advantage regime. This possibility is encouraged by recent advances showing the first experimental demonstrations of non-linear photonic processes within solid state devices [61]. We will first describe the introduction of non-linearities within the otherwise linear evolution. Then, we will provide an upper bound on the complexity of the enhanced devices via a simulation scheme based on auxiliary, linear-optical gadgets. We will discuss both the asymptotic and finite cases, leveraging results from the well-established linear Boson Sampling framework [7].

Boson Sampling. – Boson Sampling [7] is a computational task which corresponds to sampling from
the output distribution of \( n \) indistinguishable, non-interacting photons after evolution through a \( m \)-mode linear network [see Fig. 1(a)]. Given an interferometer described by a unitary matrix \( U \), the transition amplitude from input \( |S\rangle \) to output state \( |T\rangle \) can be written as:

\[
A_U(|S\rangle \rightarrow |T\rangle) = \langle S|\varphi(U)|T\rangle = \frac{\text{per}(U_{S,T})}{\sqrt{\prod_{i,j=1}^m s_i!t_j!}}, \tag{1}
\]

where \( \text{per}(A) = \sum_{\sigma \in S_n} \prod_{i=1}^n a_{\sigma(i)} \) is the matrix permanent, \( \{s_i\}\{t_i\} \) are the occupation numbers of states \( |S\rangle(|T\rangle) \), and \( U_{S,T} \) is the \( n \times n \) matrix obtained by selecting rows and columns of \( U \) according to the \( \{s_1, \ldots, s_m\} \) and \( \{t_1, \ldots, t_m\} \) respectively. Calculation of permanents of matrices with complex entries is in the \#P-hard computational complexity class [62]. In Eq. (1), \( \varphi(U) \) is the unitary transformation acting on the Hilbert space \( \mathcal{H}_{m,n} \) of \( n \) photons in \( m \) modes, that corresponds to the linear evolution \( U \) on the optical modes. Due to the linearity of the evolution, \( \varphi(U) \) is an homomorphism [63]. This means that, if a given evolution is the sequence of two linear networks \( W \) and \( V \), the overall evolution can be written in terms of permanents of submatrices of \( U = VW \).

In Ref. [7] it was shown that sampling (even approximately) from the output distribution of such a system is classically hard if (i) the input state \( |S\rangle = |s_1, \ldots, s_m\rangle \) has at most one photon per mode, (ii) \( U \) is drawn randomly from the uniform Haar measure, and (iii) the number of modes \( m \) and photons \( n \) satisfy \( m \gg n^6 \).

**Non-linear Boson Sampling.** – Let us now consider the scheme of Fig. 1(b). An input state \( |S\rangle = |s_1, \ldots, s_m\rangle \) of \( n \) indistinguishable photons undergoes a \( m \)-mode evolution divided in three steps. While steps 1 and 3 are linear evolutions \( V \) and \( W \) drawn from the Haar ensemble, the intermediate step 2 now consists of a non-linear evolution \( N \). This \( N \) transforms a state \( |R\rangle = |r_1, \ldots, r_m\rangle \) as \( |R\rangle \rightarrow \sum_{Q \in \Phi_m,n} N_{r_1 \ldots r_m}^{q_1 \ldots q_m} |Q\rangle \), where \( |Q\rangle = |q_1, \ldots, q_m\rangle \) and \( \Phi_m,n \) is the set of tuples corresponding to \( m \) photons in \( m \) modes. In this equation, function \( N_{r_1 \ldots r_m}^{q_1 \ldots q_m} \) represents the transition amplitude \( A_N(|R\rangle \rightarrow |Q\rangle) \) determined by the non-linear evolution. We assume \( N_{r_1 \ldots r_m}^{q_1 \ldots q_m} \) has an efficient classical description, e.g., it is given by the composition of a small number of few-mode non-linear transformations, or by a Hamiltonian with a simple form in terms of the field operators.

Let us now write the overall transformation of input state \( |S\rangle = |s_1, \ldots, s_m\rangle \) according to the three-step evolution \( W \rightarrow N \rightarrow V \), which includes linear transformations \( W,V \) of the form given by Eq. (1), and the non-linear \( N_{r_1 \ldots r_m}^{q_1 \ldots q_m} \):

\[
A_{W,N,V}(|S\rangle \rightarrow |T\rangle) = \sum_{R \in \Phi_m,n} \sum_{Q \in \Phi_m,n} \frac{\text{per}(W_{S,R})N_{r_1 \ldots r_m}^{q_1 \ldots q_m}\text{per}(V_{Q,T})}{\sqrt{\prod_{i,j,k,l=1}^m s_i!r_j!q_k!t_l!}}, \tag{2}
\]

This amplitude is written as a Feynman path sum over all possible basis states just before and after the non-linear evolution step. If the permanent distribution was peaked, it might be possible to obtain a good approximation to Eq. (2) by summing over only the dominant terms. Haar random matrices, however, display an anticoncentrated, relatively flat distribution [7]. In [64] we provide numerical evidence for this, showing that to account for (90\%, 95\%, 99\%) of the total probability mass function, we need to calculate the probabilities associated with respective fractions \( \sim (0.5, 0.6, 0.8) \) of all possible outcomes; moreover, this behavior is nearly independent of \( n \) and \( m \).

Of course, there may be computational shortcuts to evaluating Eq. (2), other than the explicit sum over paths. For example, if we replace the non-linear term \( N \) by a linear term, the amplitude can be evaluated as a single permanent. This motivates us to investigate different ways to assess the complexity of non-linear Boson Sampling.

**Single-mode non-linear phase shift gate.** – Let us proceed by studying a specific example of non-linear evolution \( N \) consisting of a single non-linear phase gate introduced in mode \( x \). The unitary operator describing this gate can be written as \( \hat{U}_{nlp} = \exp(-i\delta_{r,x}^2) \). Its action on a generic \( m \)-mode state \( |R\rangle \) leads to a function \( \mathcal{N}_{r_1 \ldots r_m}^{q_1 \ldots q_m} \equiv \exp(-i\delta_{r,x}^2)\prod_{i=1}^m \delta_{r_i,q_i} \). Inserting this choice of non-linear evolution into the general expression (2) we obtain:

\[
A_{W,N,V}^{nlp}(|S\rangle \rightarrow |T\rangle) = \sum_{R \in \Phi_m,n} \exp(-i\delta_{r,x}^2) \times \frac{\text{per}(W_{S,R})\text{per}(V_{R,T})}{\sqrt{\prod_{i,j=1}^m s_i!(\prod_{j=1}^m r_j!)^2\prod_{i=1}^m t_i!}}, \tag{3}
\]

Eq. (3) can be rearranged in the following form (see [64]):

\[
A_{W,N,V}^{nlp}(|S\rangle \rightarrow |T\rangle) = \sum_{R \in \Phi_m,n} \frac{\text{per}(\hat{U}_{S,T})}{\sqrt{\prod_{i,j=1}^m s_i!(\prod_{j=1}^m r_j!)^2\prod_{i=1}^m t_i!}} + \sum_{r_x>1} \frac{\text{per}(W_{S,R})\exp(-i\delta_{r,x}^2) - \exp(-i\delta_{r,x}^2)\text{per}(V_{R,T})}{\sqrt{\prod_{i,j,k,l=1}^m s_i!r_j!q_k!t_l!}}.
\]

Here, \( \hat{U} \) is a unitary transformation composed by the sequence \( W,F \) and \( V \), where \( F \) replaces the non-linear phase in layer \( N \) with a linear phase shift described by the operator \( \exp(-i\delta_{r,x}^2) \). Equation (4) clearly shows that the departure from linear evolution is due only to bunching terms, corresponding to more than a single photon in mode \( x \).

**Bounding complexity via linear-optical simulation using auxiliary photons.** – An upper bound on the complexity of non-linear Boson Sampling can be obtained by devising a specific linear-optical simulation algorithm, that we discuss below for the case of a single-mode non-linear phase.
The simulation is based on the results by Scheel et al. [65], describing how auxiliary photons and modes can be used, together with linear optics, to induce effective non-linear gates. In particular, given a single mode state in the photon number basis $|\psi_{in}\rangle = \sum_i c_i |i\rangle$, it is possible to apply a polynomial of degree $k$ in the photon number operator $P_k(\hat{n})$ to $|\psi_{in}\rangle$ by injecting the state in mode 1 of a suitably chosen $(k+1)$-mode linear-optical gadget described by unitary $U_{eff}$, where the auxiliary modes $j = 2, \ldots, k+1$ are injected with a single photon state $|1\rangle_j$. The desired output state $|\psi_{out}\rangle = P_k(\hat{n})|\psi_{in}\rangle$ is obtained upon conditional detection of a single photon on each of the auxiliary modes. If the input state has a maximum number of $l$ photons $|\chi_{in}\rangle = c_0 |0\rangle + \ldots + c_l |l\rangle$, a polynomial of degree $l$ in $\hat{n}$ is sufficient to obtain the general evolution from $|\chi_{in}\rangle$ to $|\chi_{out}\rangle = c_0 |0\rangle + \ldots + c'_l |l\rangle$ with arbitrary coefficients $\{c'_0, \ldots, c'_k\}$ [66]. The success probability of the operation is equal to $Pr_{\text{succ}} = |\text{per}(U_{eff}^{0,1,\ldots,1})|^2$ [65], where $U_{eff}^{0,1,\ldots,1}$ is the $k \times k$ submatrix of $U_{eff}$ obtained by removing row 1 and column 1 from the full matrix.

Finding the effective linear-optical simulation unitary $U_{eff}$ has been done previously only for a few types of gates and small $k$ [65, 67–69], as the computational effort seems to scale exponentially with $k$. Nevertheless, even limited non-linear gate simulations can be quite versatile, as it is known that almost any non-linear gate can be combined with linear optics to generate arbitrary non-linear gates [70] - for details, see [64].

In Fig. 2 we describe the linear-optical, postselection-based gadget that can be used to simulate single-mode non-linear gates. We see that the $k$-mode linear optical gadget (with $k \leq n$) replaces the single-mode non-linear gate. In the gadget, mode $x$ and the $k$ single photons undergo the effective unitary $U_{eff}$. This linear-optical simulation approximates the non-linear Boson Sampling evolution upon detection of $k$ photons at the auxiliary output modes.

Using the state-of-the-art weak classical simulation algorithm of Clifford and Clifford [42], we can simulate the enlarged $(n+k, m+k)$ linear optical system, postselecting only those events where a single photon is measured in each of the auxiliary modes $(m+1, \ldots, m+k)$ (see [64] for more details). This results in a classical simulation algorithm for the non-linear Boson Sampling experiment.

Let us now discuss some issues that arise when using this scheme to simulate non-linear Boson Sampling in either the asymptotic regime of large numbers of photons/modes, or in the finite setting.

**Non-linearities in the asymptotic setting.** Assuming uniformly drawn, Haar-random interferometer unitaries, it has been shown that the appropriate scaling between the original number of modes $m$ and number of photons $n$ will result in asymptotic suppression of multi-
FIG. 3. Analysis of total variation distance (TVD) between the non-linear process and its linear-optical simulation, and bunching probability at the non-linear phase-shift site for $n = 4$ photons. The TVD is calculated between the exact probability distribution [Eq. (4)], and the post-selected linear-optical simulation using $k < n$ auxiliary photons. In these plots, the exact distribution is calculated using the linear-optical simulation scheme with $n$ auxiliary photons. $P_{\text{bunch}}$ is the probability of having more than $k$ photons at the non-linearity site. (a) Parametric plot showing a correlation between TVD and $P_{\text{bunch}}$, averaged over 100 different evolutions $W$ and $V$ for fixed $\phi = \pi/2$. Each point corresponds to a different value of $m$, in the range $[5, 7, \ldots, 45]$. (b)-(e) Correlation between TVD and $P_{\text{bunch}}$ for fixed number of modes $m$. Here, each point corresponds to different evolutions $W$ and $V$ for fixed $\phi = \pi/2$. In all plots: blue points corresponds to $k = 1$, orange points to $k = 2$, and green points to $k = 3$.

 photon collisions. More precisely: if $m = O(n^{j/(j-1)})$, then $(j+1)$-fold collisions are suppressed, when $n, m$ go to infinity [71]. In particular, this will be true for the photon occupation numbers at the non-linear gates. So, by choosing $m = O(n^{k/(k+1)})$, at most $k$ photons will asymptotically be present at each non-linear gate, which means the linear-optical simulation (or classical simulation based on it) can be done with only $k$ auxiliary photons per non-linear gate. As we will soon show, such a simulation for small $k$, e.g. $k = 2, 3, 4$ can be readily obtained. These simulations using $k = 2$ are sufficient for an asymptotically perfect simulation for the usual Boson Sampling regime of $m = O(n^2)$. In other words, in this setting there is a precise correspondence between one single-mode non-linear phase gate and two extra auxiliary photons. More generally, the scaling of $m$ with $n$ dictates how many auxiliary photons are needed for asymptotically perfect simulation of a non-linear phase gate $N$.

**Non-linearities in the finite setting.** The setting with finite $n, m$ is experimentally relevant, and in this case there will be no strict suppression of multi-photon collisions at the non-linear gates. Setting $k = n$ results in an exact classical simulation of non-linear Boson Sampling. When $k < n$ we will have only an approximate simulation. As an example, when $m \sim n^2$ numerical results suggest that a number of auxiliary photons equal to $k = 2, 3$ should provide a sufficiently accurate simulation given the large effective suppression of bunching at the output of Haar-random unitaries (see [64]).

There are two main features that increase the simulation complexity. First, finding an effective unitary $U_{\text{eff}}$ that uses $k$ photons for a linear-optical simulation seems to require the computation of permanents of $k \times k$ matrices [64, 65], which results in a classical runtime that increases exponentially with $k$. The other cost incurred is the postselection overhead. From all the simulated events on the enlarged linear-optical set-up with $(n+k)$ photons, we only use events where the $k$ auxiliary photons were detected at the linear-optical simulation gadget. This will happen with a probability $Pr_{\text{succ}} = |\text{per}(U_{\text{eff}}^{0,1,\ldots,k})|^2 = p$. There is some evidence that the maximum value of $p$ tends to decrease as $k$ increases [72].

We have performed numerical simulations of non-linear Boson Sampling with a single non-linear phase in the finite setting, using the classical algorithm based on linear-optical simulation. The results are shown in Fig. 3 for $(n = 3, m = 5, 9, 16, 27)$ and $(n = 4, m = 16)$. As expected, having $k = n$ results in exact sampling from the non-linear process, and in fact (once the appropriate gadget $U_{\text{eff}}$ has been determined) is numerically found to be more computationally effective than directly using
Eq. (2). For fixed \( n, k \), note that the simulation error decreases with increasing \( m \), since bunching events become rarer. These results suggest that the crucial parameter for the simulation complexity is the scaling between \( n \) and \( m \) (see also [64]). The regime when \( m = O(n) \) is particularly interesting, as there is a trade-off between a faster classical simulation algorithm [46], and the increased complexity required to find the linear-optical gadget unitary \( U_{\text{eff}} \) for larger \( k \).

If the number of auxiliary photons \( k < n \), the simulation scheme based on linear-optical gadgets will be only approximate, due to non-linear dynamics of more than \( k \) photons. The key open point in this scenario is to quantify the simulation error incurred. In Fig. 3 we provide a numerical study on how the simulation error depends on \( n, m, k \), as quantified by the total variation distance (TVD) between the exact non-linear evolution and its simulation using \( k < n \) auxiliary photons. We observe a strong correlation between the TVD and the probability of bunching at the non-linearity. An open interesting research question is to obtain a quantitative description of this dependence between TVD and bunching, for instance by using bounds on bunching in the uniformly random, Haar ensemble of unitaries.

**Discussion.** We have proposed the adoption of non-linear gates within the framework of Boson Sampling as a way to increase the computational complexity of the model. We have shown how to quantify this complexity using a linear-optical simulation with postselection, which itself can be simulated classically. For large number \( m \) of modes and \( n \) of photons, suppressed bunching allows asymptotically perfect simulation at a cost of two extra photons per non-linear phase gate introduced, if we assume \( m = O(n^2) \). For finite \( m, n \) and single-mode non-linear phase gates, we identify the probability of bunching, governed by the scaling of \( m \) as a function of \( n \), as the key factor affecting the complexity of our proposed simulation scheme.

The non-linear Boson Sampling model we propose is inherently more expressive than linear Boson Sampling. In light of the recent developments regarding first application of Boson Sampling and its variants for hybrid quantum computational models, we expect that having access to increased functionalities enabled by non-linearities can be turned into useful advantage for tasks solvable with linear Boson Sampling, as well as propose altogether new tasks solvable by noisy, intermediate-scale quantum (NISQ) devices. In parallel, an important research direction regards development of more efficient simulation schemes for non-linear gates, which is directly relevant not only to the model we propose, but to general photonic quantum computation.

**Acknowledgments.** This work is supported by the ERC Advanced Grant QU-BOSS (QUQuantum advantage via non-linear BOSon Sampling, grant agreement no. 884676), by the European Union’s Horizon 2020 research and innovation program through the FET project PHOQUSSING (“PHOTonic Quantum Sampling” - Grant Agreement No. 899544) and by FCT – Fundação para Ciência e Tecnologia, via project CEECINST/00062/2018. We acknowledge useful comments from Scott Aaronson.

[1] Y. C. J. Yin and, Y.-H. Li, S.-K. Liao, J.-G. R. L. Zhang and, W.-Q. Cai, W.-Y. Liu, B. Li, H. Dai, G.-B. Li, Q.-M. Lu, Y.-H. Gong, Y. Xu, S.-L. Li, F.-Z. Li, Y.-Y. Yin, Z.-Q. Jiang, M. Li, J.-J. Jia, G. Ren, D. He, Y.-L. Zhou, X.-X. Zhang, N. Wang, X. Chang, Z.-C. Zhu, N.-L. Liu, Y.-A. Chen, C.-Y. Lu, R. Shu, C.-Z. Peng, J.-Y. Wang, and J.-W. Pan, Science **356**, 1140 (2017).
[2] V. Giovannetti, S. Lloyd, and L. Maccone, Nat. Photon. **5**, 222 (2011).
[3] M. Nielsen and I. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, 2010).
[4] F. Flamini, N. Spagnolo, and F. Sciarrino, Rep. Prog. Phys. **82**, 016001 (2019).
[5] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. S. L. Brandao, D. A. Buell, B. Burkett, Y. Chen, Z. Chen, B. Chiaro, R. Collins, W. Courtney, A. Dunsworm, E. Farhi, B. Foxen, A. Fowler, C. Girsend, M. Giustina, R. Graff, K. Guerin, S. Habegger, M. P. Harrigan, M. J. Hartmann, A. Ho, M. Hoffmann, T. Huang, T. S. Humble, S. V. Isakov, E. Jeffrey, Z. Jiang, D. Kafri, K. Kechedzhi, J. Kelly, P. V. Klimov, S. Krysh, A. Korotkov, F. Kostritsa, D. Landhuis, M. Lindmark, E. Lucero, D. Lyakh, S. Mandrá, J. R. McClean, M. McEwen, A. Megrant, X. Mi, K. Michielsen, M. Mohseni, J. Mutus, O. Naaman, M. Neeley, C. Neill, M. Y. Niu, E. Ostby, A. Petukhov, J. C. Platt, C. Quintana, E. G. R. P. Roushan, N. C. Rubin, D. Sank, K. J. Satzinger, V. Smelyanskiy, K. J. Sung, M. D. Trevitick, A. Vainschenker, B. Villalonga, T. White, Z. J. Yao, P. Yeh, A. Zalcman, H. Neven, and J. M. Martinis, Nature **574**, 505 (2019).
[6] H.-S. Zhong, H. Wang, Y.-H. Deng, M.-C. Chen, L.-C. Peng, Y.-H. Luo, J. Qin, D. Wu, X. Ding, Y. Hu, P. Hu, X.-Y. Yang, W.-J. Zhang, H. Li, X. J. Yuxuan Li, L. Gan, G. Yang, L. You, Z. Wang, L. Li, N.-L. Liu, C.-Y. Lu, and J.-W. Pan, Science **370**, 1460 (2020).
[7] S. Aaronson and A. Arkhipov, in *Proceedings of the 43rd annual ACM symposium on Theory of Computing*, edited by A. Press (2011) pp. 333–342.
[8] S. Boixo, S. V. Isakov, V. N. Smelyanskiy, R. Babbush, N. Ding, Z. Jiang, M. J. Bremner, J. M. Martinis, and H. Neven, Nat. Phys. **14**, 595 (2018).
[9] A. W. Harrow and A. Montanaro, Nature **549**, 203 (2017).
[10] H. Bernien, S. Schwartz, A. Keesling, H. Levine, A. Omran, H. Pichler, S. Choi, A. S. Zibrov, M. Endres, M. Greiner, V. Vuletic, and M. D. Lukin, Nature **551**, 579 (2017).
[11] K. K. C. Neill and, P. Roushan and, S. Boixo, S. V. Isakov, V. Smelyanskiy, A. Megrant, B. Chiaro, A. Dunsworm, K. Arya, R. Barends, B. Burkett,
[54] H. Qi, D. J. Brod, N. Quesada, and R. Garcia-Patron, Phys. Rev. Lett. 124, 100502 (2020).

[55] P. Michler, A. Kiraz, C. Becher, W. V. Schoenfeld, P. M. Petroff, L. Zhang, E. Hu, and A. Imamoglu, Science 290, 2282 (2000).

[56] 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).

[57] N. Somaschi, V. Giesz, L. De Santis, M. P. Loredo, J. C. Almeida, G. Hornecker, S. L. Portalupi, T. Grange, C. Anton, J. Demory, C. Gomez, I. Sagnes, N. D. Lanzillotti-Kimura, A. Lemaître, A. Auffèves, A. G. White, L. Lanco, and P. Senellart, Nat. Photon. 10, 340 (2016).

[58] P. Michler, ed., Quantum Dots for Quantum Information Technologies, Nano-Optics and Nanophotonics (Springer, Cham, 2017).

[59] D. Huber, M. Reindl, J. Aberl, A. Rastelli, and R. Trotta, J. Opt. 20, 073002 (2018).

[60] F. Basso Basset, M. B. Rota, C. Schimpf, D. Tedeschi, K. D. Zeuner, F. C. da Silva, M. Reindl, V. Zwiller, K. D. Jönns, A. Rastelli, and R. Trotta, Phys. Rev. Lett. 123, 160501 (2019).

[61] L. D. Santis, C. Anton, B. Reznychenko, N. Somaschi, G. Coppola, J. Senellart, C. Gomez, A. Lemaître, I. Sagnes, A. G. White, L. Lanco, A. Auffèves, and P. Senellart, Nat. Nanotechnol. 12, 663 (2017).

[62] L. Valiant, Theor. Comput. Sci. 8, 189 (1979).

[63] S. Aaronson, Proc. Roy. Soc. A 467, 3393 (2011).

[64] See Supplemental Material for more details on the theoretical derivations and on the performed numerical simulations.

[65] S. Scheel, K. Nemoto, W. J. Munro, and P. L. Knight, Phys. Rev. A 68, 032310 (2003).

[66] The effective evolution induced by this method must be some degree-l polynomial in n̂, though it remains an open question whether any degree-l polynomial can be implemented using only l auxiliary photons. This can be shown to hold for l = 2. Suppose we have some input |χin⟩ = c₀|0⟩ + c₁|1⟩ + c₂|2⟩. All cᵢ must be phases. The phase c₀ can be fixed by an overall global phase, and the phase c₁ can be fixed by applying a (linear) phase shifter. As shown in [64], there is a gadget that implements any chosen value of c₂. Combining these facts, it follows that any nonlinear phase acting on at most 2-photon states can be simulated by using 2 auxiliary photons.

[67] E. Knill, R. Laflamme, and G. J. Milburn, Nature 409, 46 (2001).

[68] D. B. Uskov, L. Kaplan, A. M. Smith, S. D. Huver, and J. P. Dowling, Phys. Rev. A 79, 042326 (2009).

[69] D. B. Uskov, A. M. Smith, and L. Kaplan, Phys. Rev. A 81, 012303 (2010).

[70] M. Oszmaniec and Z. Zimborás, Phys. Rev. Lett. 119, 220502 (2017).

[71] A. Arkhipov and G. Kuperberg, Geometry and Topology Monographs 18, 1 (2012).

[72] J. Eisert, Phys. Rev. Lett. 95, 040502 (2005).
I. TRANSITION AMPLITUDES FOR NON-LINEAR BOSON SAMPLING

Here we derive the transition amplitude for non-linear Boson Sampling, corresponding to Eq. (2) reported in the main text. For the case of linear dynamics, the transformation is represented by a unitary matrix $U$ which describes the evolution of creation operators according to $a_i^\dagger \rightarrow \sum_j U_{ij} a_j^\dagger$. A system of $n$ indistinguishable photons evolving via a linear transformation are characterized by input-output transition amplitudes of the form:

$$\mathcal{A}_U(|S\rangle \rightarrow |T\rangle) = \langle S|\varphi(U)|T\rangle = \frac{\text{per}(U_{S,T})}{\sqrt{\prod_{i=1}^m s_i! \prod_{j=1}^m t_j!}},$$  \hspace{1cm} (S1)

Here $|S\rangle = |s_1, \ldots, s_m\rangle$ is the input state configuration and $|T\rangle = |t_1, \ldots, t_m\rangle$ is the output state, where $\{s_i\}$ and $\{t_i\}$ are the respective lists of mode occupation numbers. Equation (S1) depends on the permanent of $U_{S,T}$, which is the $n \times n$ submatrix of $U$ obtained by choosing rows and columns according to $|S\rangle$ and $|T\rangle$ as described in Ref. [1]. The permanent is defined as:

$$\text{per}(A) = \sum_{\sigma \in S_n} \prod_{i=1}^n a_{i,\sigma(i)},$$  \hspace{1cm} (S2)

where the sum is extended over all permutations $\sigma$ in $S_n$.

Equation (S1) corresponds also to writing the output state as:

$$|S\rangle \xrightarrow{U} \sum_{T \in \Phi_{m,n}} \gamma^{U}_{S,T} |T\rangle,$$  \hspace{1cm} (S3)

where $\Phi_{m,n}$ is the set of tuples describing $n$ photons in $m$ modes, and $\gamma^{U}_{S,T} = \langle S|\varphi(U)|T\rangle$. Due to the linearity of the evolution, $\varphi(U)$ is an homomorphism [2]. Consider now an evolution $U$ that can be divided as the product of two matrices $U = VW$, thus corresponding to a sequence of two linear interferometers. Given the properties of $\varphi(U)$, the final output state can be written in two equivalent ways:

$$|S\rangle \xrightarrow{U} \sum_{T \in \Phi_{m,n}} \gamma^{U}_{S,T} |T\rangle = \sum_{T \in \Phi_{m,n}} \left( \sum_{R \in \Phi_{m,n}} \gamma^{W}_{S,R} \gamma^{V}_{R,T} \right) |T\rangle,$$  \hspace{1cm} (S4)

which corresponds to the following identity for matrix permanents when $U = VW$:

$$\text{per}(U_{S,T}) = \sum_{R \in \Phi_{m,n}} \frac{\text{per}(W_{S,R})\text{per}(V_{R,T})}{\prod_{i=1}^m r_i!}.$$  \hspace{1cm} (S5)

This expression is the starting point of the output state expansion in the presence of non-linearities.

Let us now consider the scenario where a non-linear layer $N$ is inserted between two linear evolutions $V$ and $W$. The non-linear evolution $N$ transforms a state $|R\rangle = |r_1, \ldots, r_m\rangle$ as:

$$|R\rangle \xrightarrow{N} \sum_{Q \in \Phi_{m,n}} \mathcal{N}_{r_1,\ldots,r_m}^{q_1,\ldots,q_m} |Q\rangle,$$  \hspace{1cm} (S6)

where $|Q\rangle = |q_1, \ldots, q_m\rangle$. The function $\mathcal{N}_{r_1,\ldots,r_m}^{q_1,\ldots,q_m}$ represents the transition amplitude $\mathcal{A}_N(|R\rangle \rightarrow |Q\rangle)$ due to the non-linear evolution. It must in general satisfy appropriate constraints to represent a physical evolution.
We now write the overall evolution of input state $|S\rangle = |s_1, \ldots, s_m\rangle$ according to the full evolution $W \rightarrow N \rightarrow V$. After the first (linear) transformation, the state evolves to:

$$|S\rangle \xrightarrow{W} \sum_{R \in \Phi_{m,n}} \gamma_{S,R}^W |R\rangle.$$  \hspace{2cm} (S7)

Then, after the non-linear term $N$, the state can be written as:

$$|S\rangle \xrightarrow{W,N} \sum_{R \in \Phi_{m,n}} \gamma_{S,R}^W \sum_{Q \in \Phi_{m,n}} \mathcal{N}_{q_{r_1}, \ldots q_{r_m}} |Q\rangle.$$  \hspace{2cm} (S8)

Finally, after the third step corresponding to the (linear) evolution $V$ we obtain the following expression:

$$|S\rangle \xrightarrow{W,N,V} \sum_{R \in \Phi_{m,n}} \gamma_{S,R}^W \sum_{Q \in \Phi_{m,n}} \mathcal{N}_{q_{r_1}, \ldots q_{r_m}} \sum_{T \in \Phi_{m,n}} \gamma_{Q,T}^V |T\rangle.$$  \hspace{2cm} (S9)

This equation can be rearranged as:

$$|S\rangle \xrightarrow{W,N,V} \sum_{T \in \Phi_{m,n}} \left( \sum_{R \in \Phi_{m,n}} \sum_{Q \in \Phi_{m,n}} \gamma_{S,R}^W \mathcal{N}_{q_{r_1}, \ldots q_{r_m}} \gamma_{Q,T}^V \right) |T\rangle.$$  \hspace{2cm} (S10)

Hence, the transition amplitude from input state $|S\rangle$ to $|T\rangle$ reads:

$$A_{W,N,V}(|S\rangle \rightarrow |T\rangle) = \sum_{R \in \Phi_{m,n}} \sum_{Q \in \Phi_{m,n}} \frac{\text{per}(W_{S,R}) \mathcal{N}_{q_{r_1}, \ldots q_{r_m}} \text{per}(V_{Q,T})}{\sqrt{\prod_{i=1}^{m} s_i! \prod_{j=1}^{m} t_j! \prod_{k=1}^{m} q_k! \prod_{l=1}^{m} r_l!}},$$  \hspace{2cm} (S11)

corresponding to Eq. (2) of main text.

**II. CUMULATIVE BOSON SAMPLING DISTRIBUTIONS**

As a first step to analyze the case where a non-linear evolution is introduced in the linear-optical system, we numerically investigate whether a small fraction of matrix permanents can provide a significant contribution to the full transition amplitude. If that were true, a good approximation to the exact amplitude might be achievable by including only a few terms in the sum of Eq. (S11). To that end we perform a numerical simulation by (i) sampling $N_{\text{unit}} = 1000$ linear transformations $U$ according to the Haar measure; (ii) evaluating the full probability distribution $P$ for each $U$; (iii) sorting each probability distribution in decreasing order to obtain $P_S$ as a function of the fraction $N/N_{\text{tot}}$ of output configurations. For each $P_S$ we evaluate its cumulative probability $C$ by summing over a given fraction of the combinations, obtained as $C(X) = \sum_{x \leq X} P_S(x)$. This cumulative probability provides information on the effective fraction of permanents that contribute to the overall probability mass up to a chosen threshold $p$. The results of a numerical simulation for $n = 3, 4, 5, 6$ and $m \in [5, 37]$ are shown in Fig. S1.

We observe that the sorted probability distributions $P_S$ present similar trends, and that the cumulative probabilities $C$ for fixed $m$ are almost independent on the number of photons $n$. An analysis of the fraction $N(C = p)/N_{\text{tot}}$ of configurations that contribute to the full distribution up to a threshold $p$ is then shown in the inset of Fig. S1. We find that fractions $\sim (0.5, 0.6, 0.8)$ of the output combinations are necessary to evaluate (90%, 95%, 99%) of the full probability mass, respectively, and that these values are almost independent of $n$ and $m$. This feature is related to the anti-concentration conjecture for complex Gaussian matrices [1], which is relevant for the complexity of computing permanents of random matrices.

**III. NON-LINEAR BOSON SAMPLING WITH A SINGLE-MODE NON-LINEAR PHASE**

Here we perform the explicit calculation for the scenario in which a non-linear phase shift is introduced in mode $x$, corresponding to Eq. (4) in main text. The evolution operator for this transformation can be written as $U_{\text{nlp}} = \exp(-i\hat{n}_x^2 \phi)$. Evaluating its action on a generic $m$-mode Fock state $|R\rangle$ reads:

$$|R\rangle \xrightarrow{\exp(-i\hat{n}_x^2 \phi)} e^{-i\hat{n}_x^2 \phi} |R\rangle.$$  \hspace{2cm} (S12)
thus corresponding to a function $N$ of the form:

$$N^{q_1 \ldots q_m} = \exp(-ir_x^2 \phi) \prod_{i=1}^{m} \delta_{r_x,q_i}. \tag{S13}$$

Substituting into the general expression (S11) we obtain:

$$\mathcal{A}_{W,N,V}^{nlp} \left( |S\rangle \rightarrow |T\rangle \right) = \sum_{R \in \Phi_{m,n}} \exp(-ir_x^2 \phi) \frac{\text{per}(W_{S,R})\text{per}(V_{R,T})}{\sqrt{\prod_{i=1}^{m} s_i!(\prod_{j=1}^{m} r_j!)^2 \prod_{l=1}^{m} t_l!}}. \tag{S14}$$

This expression can be written in a different form. We first observe that $\exp(-ir_x^2 \phi) = \exp(-ir_x \phi)$ for $r_x \in \{0, 1\}$.

Equation (S14) can then be rearranged as:

$$\mathcal{A}_{W,N,V}^{nlp} \left( |S\rangle \rightarrow |T\rangle \right) = \sum_{r_x=0,1} \exp(-ir_x \phi) \frac{\text{per}(W_{S,R})\text{per}(V_{R,T})}{\sqrt{\prod_{i=1}^{m} s_i!(\prod_{j=1}^{m} r_j!)^2 \prod_{l=1}^{m} t_l!}} + \sum_{r_x>1} \exp(-ir_x^2 \phi) \frac{\text{per}(W_{S,R})\text{per}(V_{R,T})}{\sqrt{\prod_{i=1}^{m} s_i!(\prod_{j=1}^{m} r_j!)^2 \prod_{l=1}^{m} t_l!}} = \tag{S15}$$

where the sums are extended over $R \in \Phi_{m,n}$ for those terms with $r_x = 0, 1$ and $r_x > 1$ respectively. By summing and subtracting the following term:

$$\sum_{r_x>1} \exp(-ir_x \phi) \frac{\text{per}(W_{S,R})\text{per}(V_{R,T})}{\sqrt{\prod_{i=1}^{m} s_i!(\prod_{j=1}^{m} r_j!)^2 \prod_{l=1}^{m} t_l!}}, \tag{S16}$$

the expression for the transition amplitude can be rearranged as:

$$\mathcal{A}_{W,N,V}^{nlp} \left( |S\rangle \rightarrow |T\rangle \right) = \sum_{R \in \Phi_{m,n}} \frac{\text{per}(W_{S,R})\text{per}(V_{R,T})}{\sqrt{\prod_{i=1}^{m} s_i!(\prod_{j=1}^{m} r_j!)^2 \prod_{l=1}^{m} t_l!}} + \sum_{r_x>1} \frac{\text{per}(W_{S,R})(\exp(-ir_x^2 \phi) - \exp(-ir_x \phi))\text{per}(V_{R,T})}{\sqrt{\prod_{i=1}^{m} s_i!(\prod_{j=1}^{m} r_j!)^2 \prod_{l=1}^{m} t_l!}} \tag{S17}$$
Let us now define $F$ as the unitary matrix associated to a linear transformation applying a phase shift on mode $x$: $F = \exp(-i \hat{n}_x \phi)$. We observe that the transition amplitude between an input state $|\hat{R}\rangle$ and an output state $|\hat{P}\rangle$ which evolve according to $F$ is written as:

$$A_F(|\hat{R}\rangle \rightarrow |\hat{P}\rangle) = \frac{\operatorname{per}(F_{R,P})}{\sqrt{\prod_{i=1}^{m} r_i! \prod_{j=1}^{m} p_j!}} = \exp(-ir_x \phi) \prod_y \delta_{y,p_y},$$

(S18)

This allows us to write the following chain of equalities for the first term of Eq. (S17)

$$\sum_{R \in \Phi_{m,n}} \exp(-ir_x \phi) \operatorname{per}(W_{S,R}) \operatorname{per}(V_{R,T}) = \sum_{R \in \Phi_{m,n}} \frac{\operatorname{per}(W_{S,R}) \exp(-ir_x \phi) \sqrt{\prod_{j=1}^{m} r_j! \prod_{k=1}^{m} p_k! (\prod_y \delta_{y,p_y}) \operatorname{per}(V_{R,T})}}{\sqrt{\prod_{i=1}^{m} s_i! \prod_{j=1}^{m} r_j! \prod_{k=1}^{m} p_k! \prod_{l=1}^{m} t_l!}} =$$

$$= \sum_{R,P \in \Phi_{m,n}} \frac{\operatorname{per}(W_{S,R}) \exp(-ir_x \phi) \sqrt{\prod_{i=1}^{m} s_i! \prod_{j=1}^{m} r_j! \prod_{k=1}^{m} p_k! \prod_{l=1}^{m} t_l!}}{\sqrt{\prod_{i=1}^{m} s_i! \prod_{j=1}^{m} r_j! \prod_{k=1}^{m} p_k! \prod_{l=1}^{m} t_l!}} = \frac{\operatorname{per}(\hat{U}_{S,T})}{\sqrt{\prod_i s_i! \prod_j t_j!}},$$

(S19)

where $\hat{U} = WFV$, and Eq. (S5) has been used to perform the last simplification. By replacing this into Eq. (S17) we obtain Eq. (4) of the main text:

$$A_{\text{nlp}}^{W,N,V}(|S\rangle \rightarrow |T\rangle) = \frac{\operatorname{per}(\hat{U}_{S,T})}{\sqrt{\prod_{i=1}^{m} s_i! \prod_{j=1}^{m} t_j!}} + \sum_{r_s > 1} \frac{\operatorname{per}(W_{S,R}) |\exp(-ir^2 \phi) - \exp(-ir_x \phi)| \operatorname{per}(V_{R,T})}{\sqrt{\prod_{i=1}^{m} s_i! \prod_{j=1}^{m} r_j! \prod_{k=1}^{m} p_k! \prod_{l=1}^{m} t_l!}}.$$  

(S20)

## IV. EVOLUTION WITH A NON-LINEAR SINGLE-MODE PHASE TERM

Here we present numerical simulations that quantify the perturbation, in the output distribution, induced by the introduction of a single-mode non-linear phase term $N$. We verify numerically the difference between two output distributions: the desired one (in the presence of $N$) and one obtained by replacing the non-linear phase shift with a linear one defined by the same phase $\phi$ (i.e., the overall transformation $\hat{U} = W^2FV^2$). This is suggested by the form of Eq. (S20), which shows that a single-mode non-linear phase introduce a departure from a linear phase for those paths where multiple photons propagate in mode $x$. Finally, we verify numerically that replacing the non-linear phase shift with a linear one represents the correct benchmark for the analysis presented in this Section. We find that, in most cases, the closest linear evolution to the non-linear one $W,N,V$ is that given by $\hat{U}$.

### A. Evolution in the finite setting

We now investigate the effect of introducing a non-linear phase term in a single mode. We begin by considering a scenario with $n = 3$ photons in $m = 9, 16$ modes according to the evolution corresponding to Eq. (S14). More specifically, we insert the non-linear phase in modes $x = 5$ and $x = 9$, respectively for $m = 9$ and $m = 16$, after a linear transformation $W$ and before the second linear evolution $V$.

As a first step, we study the total variation distance $\text{TVD} = 1/2 \sum_i |p_i - q_i|$ between the output distribution $\{p_i\}$, corresponding to $U = W$, and the output distribution $\{q_i\}$, which includes the non-linear phase between $W$ and $V$. The results of a numerical simulation with $N_{\text{unit}} = 1000$ sets of unitary transformations $\{W,V\}$ are shown in Fig. S2 (a-c). We observe that, as expected, the TVD increases for larger non-linear phases $\phi$. For $m = 9$ and $\phi = \pi$, the TVD reaches a value close to the one relative to a uniform distribution, thus showing that the insertion of a single non-linear phase introduces a significant change in the output distribution.

Having seen that a non-linear term introduces a significant change in the output distribution, as expected, we then analyze whether the evolution with the non-linear phase can be approximated by a linear transformation that is different from $\hat{U}$. The form of Eq. (S20) suggests an ansatz for a linear approximation for the non-linear dynamics, namely, replacing the non-linear term $N$ with a linear unitary $F = \exp(-i \hat{n}_x \phi)$, leading to an overall interferometer $\hat{U} = WFV$. In Fig. S2 (d-f) we report the TVD $= 1/2 \sum_i |q'_i - q_i|$ between the output distribution $\{q'_i\}$, with a linear phase term (i.e. evolution $\hat{U}$), and the output distribution $\{q_i\}$, in the presence of the non-linear phase (evolution
For phase values close to $\phi = 0$ and $\phi = \pi$, the two scenarios present a small TVD, meaning that the non-linear evolution can be approximated by a simple linear phase. Indeed, for $\phi = 0$ there is no phase term in both cases, whereas for $\phi = \pi$ the terms $\exp(-m\phi)$ and $\exp(-m^2\phi)$ provide the same phase factor for all values of $n$. Conversely, the maximum difference is obtained for $\phi = \pi/2$.

As a following step, we verify numerically that similar trends are obtained by directly looking at single amplitudes. In Fig. S3 we observe the same trend, as displayed by the TVD, but now for the differences in the moduli and in the argument of the amplitudes. More specifically, we consider the behavior of relative differences $\delta_{\text{abs}} = |(A_{W,N,V}^{\text{llp}} - A_{W,V})|/|A_{W,V}|$ and the absolute difference $\Delta_{\text{arg}} = |\arg(A_{W,N,V}^{\text{llp}}) - \arg(A_{W,V})|$. Furthermore, numerical simulations show that the average trend for these quantities can be calculated by sampling $N_{\text{amp}} = 10^4$ unitaries, and by calculating a single amplitude for each sampled unitary (see Fig. S4). This allows us to obtain an approximated estimate of $\delta_{\text{abs}}$ and $\Delta_{\text{arg}}$ without having to calculate the full distributions.

Following these results, we studied how the perturbation introduced by the non-linear phase scales with the number of photons $n$ and the number of modes $m$. The results are shown in Fig. S5 for $n = 3, 4, 5$, and different values of $m$. We observe that the change induced by the non-linear phase relative to the linear transformation $\mathcal{U} = \mathcal{W}FV$ decreases with the number of modes $m$. This trend is shown by dashed lines of Fig. S5 (g-h). This can be explained by the second term of Eq. (S20), which shows that the departure from the linear phase evolution depends on the weight of the bunching terms at the non-linearity. This is discussed in the main text, and will be also addressed in Sec. V.
FIG. S3. Average trends of (a) $\delta_{\text{abs}}$ and (b) $\Delta_{\text{arg}}$ as functions of the non-linear phase $\phi$, obtained by calculating all amplitudes for $N_{\text{unit}} = 1000$ unitaries. Solid lines: comparison between evolution $W, N, V$ and evolution $U = WV$. Dashed lines: comparison between evolution $W, N, V$ and the one obtained with evolution $U = WFV$. Shaded regions: 1$\sigma$ intervals obtained from the numerical simulation.

FIG. S4. Convergence of $\delta'_{\text{abs}}$ for (a) $n = 3, m = 9$ and (b) $n = 3, m = 16$. In each plot, the average behavior (red curves) obtained for all amplitudes corresponding to $N_{\text{unit}} = 1000$ unitaries are compared with the trend where only a single amplitude is calculated for $N_{\text{amp}}$ different unitaries. Shaded regions: 1$\sigma$ intervals for the simulation with $N_{\text{unit}}$ unitaries.

B. Approximating non-linear evolution with linear transformations

In the previous section, we employed the linear transformation $U$ as a benchmark to assess the disturbance induced by the non-linear phase shift. In particular, we reported numerical evidence that the insertion of $\exp(-i\hat{n}_2^2\phi)$ within linear evolutions $W$ and $V$ introduces a non-negligible departure from the linear behaviour. This choice of benchmark is suggested by the expression of the transition amplitude reported in Eq. (4) of the main text and Eq. (S20) above. There, we observe that the overall evolution can be written as a combination of a linear term, identified by unitary $U$, that includes a linear phase shift, and a second term which is present only when multiple photons propagate through the non-linearity.

Now we report on additional numerical analysis to support the choice of benchmark evolution $U$. More specifically, we verified whether a different effective linear transformation $H$ can have a smaller distance with respect to the full evolution $W, N, V$. We consider the case of $n = 2, 3$ photons and $m = 9$ modes, focusing on the $\phi = \pi/2$ case. We then sample two Haar random unitary matrices $W$ and $V$, and search numerically for a linear evolution $H$ whose output distribution has a smaller TVD relative to the output considering the non-linear evolution $W, N, V$. This search is
FIG. S5. Analysis of $\delta_{\text{abs}}$ and $\Delta_{\text{arg}}$ for different values of $n,m$, obtained by evaluating a single amplitude for $N_{\text{unit}} = 10^4$ unitaries. (a-f) Simulations as a function of the non-linear phase $\phi$. Red: $m = 5$. Blue: $m = 27$. The arrow shows the trend for increasing values of $m$. (g-h) Analysis for $\phi = \pi/2$ as a function of $m$, for $n = 3$ (green) and $n = 4$ (orange) and $n = 5$ (cyan). In all plots, solid lines are the comparison between evolution $W,N,V$ and the one obtained with evolution $U = WV$, while dashed lines are the comparison between evolution $W,N,V$ and evolution $U = WV$. 

performed by randomly sampling from the Haar measure.

In Fig. S6 (a) we show the progress of the best TVD as a function of the number of sampled unitaries $k$ up to $N_{\text{unit}} = 10^7$ sampled Haar-random matrices. The result suggests that the TVD approximately saturates after $N_{\text{unit}} = 10^6$ sampled matrices. We then repeated this numerical search for 100 different pairs $W,V$ [see Fig. S6 (b)]. We observe that, for most tested cases, the linear transformation $\overline{U}$ is closer to the non-linear evolution $W,N,V$ than the best Haar random matrix found via the brute force numerical search.

These numerical results then give supporting evidence that $\overline{U}$, obtained by replacing the non-linear phase shift with a linear one, can represent a suitable benchmark for the chosen non-linearity as per Eq. (S20).
V. SIMULATION OF NON-LINEAR BOSON SAMPLING WITH LINEAR OPTICS AND ANCILLARY PHOTONS

In this Section we give more details on the algorithm for simulating non-linear Boson Sampling using a linear-optical gadget acting on auxiliary modes and photons.

We begin by arguing that, at least in principle, there should always exist some linear-optical gadget that simulates a given non-linear transformation. There are likely to be decompositions that are much more efficient than the one we describe here in terms of required ancillary photons, success probability of the gadget, computational complexity of constructing the gadget and so on. We leave this optimization as an interesting avenue for future research, focusing on small-scale numerical investigations. We then provide some numerical insight on the role of the number of auxiliary photons, $k$, by looking at the contribution of bunching terms in Haar-random photon number distributions. Finally we provide the explicit construction of the simulation algorithm, and discuss the calculation of the effective linear-optical matrices required for the simulation. Finally we provide some numerical analysis to complement the discussion reported in the main text, and leverage the linear-optical simulation into a classical simulation algorithm.

A. Universality of non-linear gadgets

We begin by quoting a result due to Oszmaniec and Zimboras [3] regarding the ability of some non-linear gates to extend linear optics to full universality within the state space of $n$ photons in $m$ modes.

In particular, let $\mathcal{H}_{m,n}$ be the restriction of Fock space to $n$ photons in $m$ modes, and let $\text{LO}_b$ be set of all transformations generated by linear optics acting on $\mathcal{H}_{m,n}$. Let $V$ be some unitary gate that is not in $\text{LO}_b$ (hence, a non-linear gate), and let $\langle \text{LO}_b, V \rangle$ be the group of dynamics generated by sequences of elements from $\text{LO}_b \cup \{V\}$.

Theorem 1 in [3] guarantees that $\langle \text{LO}_b, V \rangle$ is the full unitary group acting on $\mathcal{H}_{m,n}$ if either:

(i) $m > 2$, or

(ii) $m = 2$ and $[V \otimes V, L] \neq 0$.

Here, $L = |\Psi\rangle \langle \Psi|$ is a project acting on two copies of $\mathcal{H}_{m,n}$ defined by

$$|\Psi\rangle = \sum_{l=0}^{n} (-1)^l |D_l\rangle |D_{n-l}\rangle,$$
and where \(|D_l\rangle\) is the two-mode state with \(l\) particles in the first mode and \(n-l\) particles in the second. In our description of the result of [3] we have also skipped one set of conditions which applies to fringe cases and do not concern us here.

This result allows us to check which conditions a particular (unitary) non-linear gate must satisfy in order to be able to approximately decompose any other non-linear gate. In other words, suppose we have some gate \(V\) that satisfies the above criteria. If our non-linear Boson Sampling setup includes some non-linearity \(N\), we know that it is possible, in principle, to approximate \(N\) by a sequence of linear-optical elements interspersed with some copies of \(V\).

By the result of [3], any non-linear unitary gate \(V\) suffices for the above purposes if \(m > 2\), which means that we can always replace \(N\) by a 3-mode circuit consisting of beam splitters and applications of \(V\). However, \(N\) is a single-mode non-linearity. Therefore it is likely that applying the above result for \(m = 3\) is less efficient (in the sense of number of applications of \(V\) required to obtain an approximation of \(N\) to desired accuracy) than applying it for \(m = 2\) whenever that option is available. Thus, we can test some reasonable non-linear gates to check whether they satisfy condition (ii) when \(m = 2\). The authors of [3] proved that for a two-mode cross-Kerr gate, given by \(\exp(-in_1n_2\phi)\), this is true for some values of \(\phi\) such as \(\pi/3\). It is easy to check that the following gates also satisfy condition (ii):

(i) The gate which we focus on in this paper, namely \(\hat{U}_{nlp} = \exp(-m^2\phi)\), for \(\phi = \pi/3\). If the number of photons \(n\) is odd, then this gate is also universal for \(\phi = \pi/2\).

(ii) The generalized non-linear sign shift gate \(NS_k\) [4], which leaves all states with up to \(k-1\) photons invariant and applies \(|k\rangle \rightarrow -|k\rangle\).

So far we considered the ability of some non-linear gates to simulate others, but our main goal is to map this into a post-selected linear-optical gadget. Two issues arise if we try to directly apply the result of [3] to this question. The first is that the result is not constructive. It provides a test to decide whether some non-linear gate is universal when supplemented with linear optics, but does not output a sequence that approximates a specific desired gate. This can be solved by invoking the standard Solovay-Kitaev theorem [5] (with the caveat that, besides \(m\), some \(n\) might suffice).

The second issue is that, given a maximum number of photons present in the input state, \(n\), we need there to always exist some unitary non-linear gate that can be implemented as a post-selected linear-optical gadget. To the best of our knowledge, this remains an open question. In subsequent sections we show that there exists a gadget for \(\hat{U}_{nlp}\) when \(\phi = \pi/2\) and for up to 4 photons, while previous work [4] has given evidence that the gate \(NS_k\) can be implemented for arbitrary \(k\) with success probability that decreases as \(1/k^2\).

We can combine the observations above into a procedure to obtain a post-selected gadget for any non-linearity \(N\). We begin by using the Solovay-Kitaev algorithm to provide a sequence that approximates \(N\) as a sequence of beam-splitters and some standard non-linear gate such as \(\hat{U}_{nlp}\) or \(NS_k\). Then, we replace every copy of that gate by a postselected gadget. We do not pursue this approach as a practical means to perform this task, only as an in-principle argument that it should always be possible. The circuit generated by this reasoning is likely to lead to a very inefficient gadget for a typical \(N\), and subsequently a very inefficient classical algorithm. However, we conjecture that is unnecessary, and that some gadget that uses only at most \(n\) auxiliary photons should exist for any \(n\)-photon single-mode non-linearity. As evidence, throughout this work we show that, for many different gates, a direct gadget exists that does not require a Solovay-Kitaev type decomposition.

B. Haar-random distribution with photon-number threshold

As a first step, we need to evaluate to evaluate the contribution of bunching terms in the photon number distribution after the first Haar-random interferometer. This is important because the measurement-induced scheme of Ref. [6] allows us to simulate single-mode non-linearities acting on up to \(n\) input photons by using a linear-optical gadget and \(n\) auxiliary photons. Therefore, if the rate bunching is not too large after the first interferometer, a smaller number of auxiliary photons might suffice.

In the asymptotic limit, the Bosonic Birthday Paradox holds, stating that bunching contributions are negligible for \(m \gg n^2\). To verify this limit for finite sizes, we performed numerical simulations that investigate the contribution of bunching terms at the output of Haar random matrices for \(n = 3, 4, 5, 6\) and \(7 \leq m \leq 37\). For each system size \((n, m)\) we numerically evaluated the overall average probability to obtain an output state with contributions having at most \(n_{max}\) photons for each output mode (i.e., excluding configurations where at least one output mode has \(n_{out} > n_{max}\) photons). The results are shown in Fig. S7.

We observe that, for \(m = n^2\), there is a significant portion (\(\sim 5\%\)) of the distribution probability where bunching terms with 3 photons on the same mode are occur, while higher order terms are almost negligible. This is a relevant aspect that must be taken into account for the simulation algorithm described below.
C. Simulating non-linearities with auxiliary photons and extended linear evolution

Simulation of non-linear Boson Sampling with single-mode non-linearities can be performed according to the approach reported in the main text. In particular, the scheme of Fig. 2 employs a set of ancillary photons and modes, and an auxiliary transformation $U_{\text{eff}}$. Such transformation depends on the actual non-linear transformation to be simulated (more details on the calculation of $U_{\text{eff}}$ can be found in Sec. V D). The non-linear evolution is successfully simulated conditioned to the detection of state $|1\rangle$ on each of the output auxiliary modes.

The scheme of Fig. 2 thus defines both a linear-optical simulation algorithm, which can be implemented by building the corresponding experimental apparatus and thus running the device, and a classical simulation algorithm. In the second case, sampling can be performed via the Clifford and Clifford algorithm [7], with the additional ingredient of discarding those events which do not correspond to a correct post-selected configuration.

We can now define below the classical simulation algorithm for non-linear Boson Sampling.

**Algorithm 1.** The following classical algorithm provides an approximate simulation of non-linear Boson Sampling with single-mode non-linearities:

(i) Compute the overall $m+k$-mode linear transformation resulting from the sequential applications of $W$, $U_{\text{eff}}$ and $V$.

(ii) Sample a single event from the overall transformation with an input state having single photons in modes $(1,\ldots,n)$ and $(m+1,\ldots,m+k)$, by using the exact algorithm due to Clifford and Clifford [7].

(iii) If one photon per mode is obtained for modes $(m+1,\ldots,m+k)$, retain the event as valid. Otherwise, discard the event and repeat point (ii).

(iv) Iterate the procedure until the number of required samples $N_{\text{sample}}$ is obtained.
D. Calculation of the effective unitary \( U_{\text{eff}} \)

A central step in the linear-optics based simulation algorithm is the calculation of the effective unitary \( U_{\text{eff}} \). This transformation depends on the employed non-linear evolution \( N \), and a recipe for obtaining it has been reported in Ref. [6].

Consider an input state composed by an arbitrary superposition of up to \( k \)-photon Fock terms:

\[
|\psi_{\text{in}}\rangle = c_0|0\rangle + c_1|1\rangle + c_2|2\rangle + \ldots + c_k|k\rangle. \tag{S21}
\]

We focus on the specific non-linear evolution \( N \) considered here, corresponding to a non-linear phase shift acting as \( \exp(-i\hat{n}^2\phi) \). The desired output state after transformation \( N \) is given by:

\[
|\psi_{\text{out}}\rangle = c_0|0\rangle + c_1e^{-i\phi}|1\rangle + c_2e^{-i4\phi}|2\rangle + \ldots + c_ke^{-i k^2\phi}|k\rangle. \tag{S22}
\]

![FIG. S8. Scheme of the linear-optics gadget for simulation of single-mode non-linearities with an input of up to \( k \) photons.](image)

The main idea is to consider the gadget shown in Fig. S8. It is composed by a \((k+1) \times (k+1)\) unitary transformation \( U_{\text{eff}} \). The input state \( |\psi_{\text{in}}\rangle \) is injected into mode 1, while a single photon is injected in each auxiliary mode (from 2 to \( k+1 \)). Conditioned on the detection a single photon in each output mode from 2 to \( k+1 \), the output state on mode 1 can be written as:

\[
|\psi'\rangle = c_0\text{per}(U^{0,1,\ldots,1}_{\text{eff}})|0\rangle + c_1\text{per}(U^{1,1,\ldots,1}_{\text{eff}})|1\rangle + c_2\text{per}(U^{2,1,\ldots,1}_{\text{eff}})|2\rangle + \ldots + c_k\text{per}(U^{k,1,\ldots,1}_{\text{eff}})|k\rangle. \tag{S23}
\]

Here, \( U^{l,1,\ldots,1}_{\text{eff}} \) is the \((l+k) \times (l+k)\) matrix obtained by repeating \( l \) times the first row and the first column of \( U_{\text{eff}} \), and by repeating a single time each of the remaining rows and columns of \( U_{\text{eff}} \). Assuming that \( \text{per}(U^{0,1,\ldots,1}_{\text{eff}}) \) is different from zero, Eq. (S23) can be rearranged as:

\[
|\psi'\rangle = \text{per}(U^{0,1,\ldots,1}_{\text{eff}})\left[c_0|0\rangle + \frac{c_1\text{per}(U^{1,1,\ldots,1}_{\text{eff}})}{\text{per}(U^{0,1,\ldots,1}_{\text{eff}})}|1\rangle + \frac{c_2\text{per}(U^{2,1,\ldots,1}_{\text{eff}})}{\text{per}(U^{0,1,\ldots,1}_{\text{eff}})}|2\rangle + \ldots + \frac{c_k\text{per}(U^{k,1,\ldots,1}_{\text{eff}})}{\text{per}(U^{0,1,\ldots,1}_{\text{eff}})}|k\rangle\right]. \tag{S24}
\]

In this expression, the common term \( \text{per}(U^{0,1,\ldots,1}_{\text{eff}}) \) is directly related to the success probability of the process, which is obtained as \( \text{Pr}_{\text{succ}} = |\text{per}(U^{0,1,\ldots,1}_{\text{eff}})|^2 \). The conditions that \( U_{\text{eff}} \) must satisfy in order to correspond to the desired phase shift are obtained by imposing that, up to the global term, coefficients for state \( |\psi'\rangle \) in Eq. (S22) are equal to those for state \( |\psi''\rangle \) in Eq. (S24). This procedure leads to the following set of equations:

\[
\begin{align*}
\text{per}(U^{1,1,\ldots,1}_{\text{eff}}) &= \text{per}(U^{0,1,\ldots,1}_{\text{eff}})e^{-i\phi} \\
\text{per}(U^{2,1,\ldots,1}_{\text{eff}}) &= \text{per}(U^{0,1,\ldots,1}_{\text{eff}})e^{-i4\phi} \\
&\quad \vdots \\
\text{per}(U^{k,1,\ldots,1}_{\text{eff}}) &= \text{per}(U^{0,1,\ldots,1}_{\text{eff}})e^{-ik^2\phi}
\end{align*} \tag{S25}
\]

Finding \( U_{\text{eff}} \) such that all conditions (S25) are satisfied is not a trivial task. Indeed, evaluation of these conditions involves matrix permanents of size up to \((2k) \times (2k)\). Here, to determine the matrices \( U_{\text{eff}} \) to simulate the non-linear phase shift evolution up to \( k = 4 \), we employed a numerical minimization approach. More specifically, we minimized the following quantity:

\[
D = \sum_{l=1}^{k} \left\{ \left[ \text{Re} \left( \text{per}(U^{l,1,\ldots,1}_{\text{eff}}) \right) - \text{Re} \left( \text{per}(U^{l,0,1,\ldots,1}_{\text{eff}})e^{-i l^2\phi} \right) \right]^2 + \left[ \text{Im} \left( \text{per}(U^{l,1,\ldots,1}_{\text{eff}}) \right) - \text{Im} \left( \text{per}(U^{l,0,1,\ldots,1}_{\text{eff}})e^{-i l^2\phi} \right) \right]^2 \right\}. \tag{S26}
\]
Parametrization of \( U_{\text{eff}} \) is obtained by using the Reck decomposition [8], which is employed as a mathematical tool to define the set of parameters (beam-splitter transmittivities and phase shifts) over which the minimization is performed.

Performing an unconstrained minimization, we found that the algorithm can collapse into a minimum (a value of \( U_{\text{eff}} \)) which provides a very small success probability \( P_{\text{succ}} \). To avoid this issue, we searched for \( U_{\text{eff}} \) by performing a constrained numerical optimization, imposing that \( P_{\text{succ}} \geq p_{\text{th}} \). Here, \( p_{\text{th}} \) is a threshold which has to be manually inserted before the minimization process. By trial and error, we found that the optimization process is sensitive to the value of \( p_{\text{th}} \), which thus should be carefully chosen.

By using this numerical approach, we calculated the effective unitaries \( U_{\text{eff}} \) for the linear-optics simulation gadget, for \( k = 2, 3, 4 \) and \( \phi = \pi/2 \). For \( k = 2 \) we found:

\[
\begin{bmatrix}
0. & -0.4574 & -0.8426 & 0.0223 & 0.2822 & -0.0261 \\
-0.0969 & -0.0943 & -0.1689 & 0.1830 & -0.6028 & 0.7458 \\
0.6775 & 0.5599 & -0.2940 & 0.3756 & 0. \\
\end{bmatrix}
\]

which leads to a success probability of \( P_{\text{succ}} \approx 0.209 \). For \( k = 3 \) we found:

\[
\begin{bmatrix}
0.0032 & 0.2218 & 0.0889 & -0.8075 & 0.1756 & -0.0772 & 0.1455 & -0.4826 \\
0.6671 & 0.1569 & 0.1942 & -0.2732 & 0.1871 & -0.0443 & -0.1623 & 0.5956 \\
0.0606 & -0.1733 & 0.2204 & -0.2742 & -0.8767 & +0.1685 & -0.2133 & -0.0099 \\
0.2418 & -0.6237 & 0.3134 & +0.0717 & 0.3240 & +0.1560 & -0.4179 & -0.3802 \\
\end{bmatrix}
\]

which leads to a success probability of \( P_{\text{succ}} \approx 0.04 \). Finally, for \( k = 4 \) we found that:

\[
\begin{bmatrix}
-0.0006 & -0.1994 & -0.5735 & +0.0763 & 0.0071 & -0.5055 & 0.2902 & -0.3501 & 0.1843 & -0.6181 \\
0.3200 & -0.2740 & 0.3072 & 0.5019 & -0.0749 & -0.0098 & -0.0761 & +0.3436 & 0.4135 & -0.4190 \\
0.4328 & -0.1960 & -0.0933 & +0.4354 & -0.1877 & +0.3742 & 0.4265 & -0.1473 & -0.0404 & +0.4421 \\
0.4356 & 0.6058 & 0.0671 & -0.2111 & 0.2851 & +0.0872 & -0.0559 & -0.5462 & -0.0572 & -0.0219 \\
0.0123 & 0.0017 & 0.2591 & 0.0667 & -0.3623 & -0.7721 & 0.2273 & -0.3355 & 0.1105 & +0.1560 \\
\end{bmatrix}
\]

which leads to a success probability of \( P_{\text{succ}} \approx 0.008 \).

We first observe that, in all cases, all matrices present a finite success probability. Furthermore, we observe a decreasing trend in \( P_{\text{succ}} \) with respect to the photon number threshold \( k \). We also find that, due to the presence of larger matrix permanents and to the increasing number of parameters, the computation becomes progressively more expensive for increasing \( k \).

As a further analysis, we have then focused on the \( k = 2 \) scenario, in particular with reference to the results of Ref. [9]. In that paper, upper bounds on the success probabilities for measurement-induced non-linear gates have been derived. For instance, in the case of a non-linear phase gate of the form:

\[
c_0|0\rangle + c_1|1\rangle + c_2|2\rangle \rightarrow c_0|0\rangle + c_1|1\rangle + c_2 e^{i\varphi}|2\rangle,
\]

the success probability is bounded by:

\[
P_{\text{succ}} \leq \left[3 - \cos(\pi - \varphi)\right]^2 / 16.
\]

This bound is derived by assuming unlimited resources for the measurement-induced scheme, that is, unbounded number of ancillary modes and photons.

This result can be applied directly to the non-linear phase shift \( \exp(-i\hat{n}^2\phi) \). More specifically, the output state of up to 2 photon terms after this non-linear evolution is

\[
\exp(-i\hat{n}^2\phi) (c_0|0\rangle + c_1|1\rangle + c_2|2\rangle) = c_0|0\rangle + c_1 e^{-i\phi}|1\rangle + c_2 e^{-i4\phi}|2\rangle
\]

The state of Eq. (S32) can be mapped to Eq. (S30) by means of a linear phase shift operator as:

\[
c_0|0\rangle + c_1 e^{-i\phi}|1\rangle + c_2 e^{-i4\phi}|2\rangle = \exp(-i\hat{n}\phi) (c_0|0\rangle + c_1|1\rangle + c_2 e^{-i2\phi}|2\rangle)
\]

Thus, the bound of Eq. (S31) applies to the non-linear phase shift considered here up to 2 input photons, by identifying \( \varphi = -2\phi \):

\[
P_{\text{succ}} \leq \left[3 - \cos(\pi + 2\phi)\right]^2 / 16
\]
As a first step, we focus on the $\phi = \pi/2$ scenario. In this case, the bound of Eq. (S34) gives $P_{\text{succ}} \leq 0.25$. The effective unitary $U_{\text{eff}}^{(2)}$ of Eq. (S27), found with the previous method, corresponds to $P_{\text{succ}} \simeq 0.209$, and thus is close to the bound of Ref. [9].

We then investigated whether it is possible to use the same approach to obtain effective unitaries with finite success probability for all values of $\phi \in [0, \pi/2]$. The results are shown in Fig. S9, and compared with the bound of Eq. (S34). We observe that, in fact, for all tested values of $\phi$, $P_{\text{succ}} \gtrsim 0.1$. Thus, it is possible to implement the simulation algorithm with non-negligible success probability. On the other hand, for small $\phi$ we also observe larger deviations from the bound of Eq. (S34). One reason could be the fact that this bound allows for measurement-induced schemes with unbounded resources, whereas the gadget we consider here uses a specific number of photons and modes (see Fig. S8). It is also relevant that at this stage we are only considering the maximum success probability for an exact implementation of the non-linear phase. Thus, even though there is a trivial implementation with $P_{\text{succ}} = 1$ for the specific value $\phi = 0$ (i.e. the identity matrix), as soon as $\phi$ deviates from 0 it might become necessary to have more modes and photons achieve success probability greater than $\approx 0.1$.

However, in the regime where $\phi$ is small, an exact implementation might not be the best approach. As discussed in Sec. IV B, in that regime it might be better to incur a small error in TVD and use a linear phase as approximation to the non-linear one (which would also lead to a smaller runtime in the simulation, since it would not require simulation via extra photons). We leave the investigation of the tradeoff between these figures of merit—runtime and approximation error—as direction for future work.

![Figure S9](image-url)

**FIG. S9.** Success probability for the effective unitaries, found with the approach discussed in the text, for $k = 2$ and different values of $\phi$. Blue solid line: bound for $P_{\text{succ}}$ obtained from Eq. (S34) as per Ref. [9]. Orange points: $P_{\text{succ}}$ found for the best unitaries obtained via the constrained numerical minimization of Eq. (S26).

### E. Correlations between bunching and TVD

In the main text we have shown that the bunching probability at the non-linear site is related to the TVD between the exact distribution and the one obtained with the simulation scheme based on linear optics and auxiliary photons and modes. In Fig. S10 we report a similar analysis to the one shown in Fig. 3 of the main text. However, here we replace the bunching probability at the non-linear site $P_{\text{bunch}}$ with the overall bunching probability $P'_{\text{bunch}}$ calculated for the full distribution.

In this scenario, we still observe an overall trend between the average TVD and the average bunching probability $P'_{\text{bunch}}$ for varying number of modes [see Fig. S10 (a)]. However, when looking at the distribution for fixed number of modes $m$, we observe no correlations between TVD and the bunching probability (over all modes). This highlights the fact that, as shown in the main text and in Eq. (S20), the TVD is related to the bunching probability restricted to the mode with the non-linearity.
FIG. S10. Analysis of total variation distance (TVD) and bunching probability, for \( n = 4 \) photons and different number of modes. The TVD is calculated between the exact probability distribution, whose single transition amplitudes are obtained from Eq. (S20), and the post-selected distribution corresponding to the linear optics measurement-based simulation algorithm for \( k < n \). In these plots, the exact distribution is calculated as the output one from the simulation scheme with \( n \) ancillary photons and modes. At variance with Fig. 3 of the main text, the bunching probability \( P'_{\text{bunch}} \) corresponds to the overall probability obtained by summing up all terms having more than \( k \) photons in the full distribution. (a) Parametric plot showing the relative trend between the TVD and the bunching probability \( P'_{\text{bunch}} \) at the non-linear site, averaged over 100 different evolutions \( W \) and \( V \) for fixed \( \phi = \pi/2 \). Each point corresponds to a different value of \( m \), in the range \([5, 7, \ldots, 45]\). (b)-(e) Correlation between TVD and \( P'_{\text{bunch}} \) for fixed number of modes \( m \). Here, each point corresponds to different evolutions \( W \) and \( V \) for fixed \( \phi = \pi/2 \). In all plots: blue points corresponds to \( k = 1 \), orange points to \( k = 2 \), and green points to \( k = 3 \).

### F. Classical simulation algorithm implementation with finite sample size

In this section we report on numerical simulations performed to test the implementation of Algorithm 1 as a classical simulation for non-linear Boson Sampling. The results of these numerical simulations are shown in Fig. S11 for different values of \( n \) and \( m \). In all cases, we considered a single non-linear phase positioned in the central mode between linear transformations \( W \) and \( V \) according to the scheme described in the main text. We then verified the convergence, in total variation distance, of a sample composed of \( N_{\text{sample}} \) events with respect to the exact output distribution.

More specifically, for each tested value of \( n \) and \( m \) we have drawn samples by using different approaches. These include brute-force sampling from the exact distribution obtained with the Feynman integral approach, and sampling with the Algorithm 1 for \( k = n \) and \( k < n \). From the plots of Fig. S11, we first observe that sampling from the exact distribution and from the linear-optics simulation algorithm with \( k = n \) provides the same trend as a function of \( N_{\text{sample}} \). This provides an additional numerical evidence that using Algorithm 1 with \( k = n \) leads to an exact simulation algorithm for non-linear Boson Sampling. Conversely, when \( k < n \), the simulation saturates to an asymptotic value provided by the post-selected distribution at the output of the linear-optics gadget.

---

[1] S. Aaronson and A. Arkhipov, in *Proceedings of the 43rd annual ACM symposium on Theory of Computing*, edited by A. Press (2011) pp. 333–342.
[2] S. Aaronson, Proc. Roy. Soc. A 467, 3393 (2011).
[3] M. Oszmaniec and D. J. Brod, New Journal of Physics 20, 092002 (2018).
[4] S. Scheel and K. M. Audenaert, New Journal of Physics 7, 149 (2005).
[5] M. Nielsen and I. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, 2010).
[6] S. Scheel, K. Nemoto, W. J. Munro, and P. L. Knight, Phys. Rev. A 68, 032310 (2003).
[7] P. Clifford and R. Clifford, in *SODA ’18: Proc. 29th ACM-SIAM Symp. on Discrete Algorithms* (2018) pp. 146–155.
FIG. S11. Classical simulation of non-linear Boson Sampling with a single non-linear phase $\phi = \pi/2$ via algorithm based on linear optics and ancillary photons. Quality of the simulation is measured via the total variation distance (TVD), defined as $\text{TVD} = 1/2 \sum_i |p_i - q_i|$ between the output distribution $p_i$ calculated from Eq. (S14) with $\phi = \pi/2$ and a sample $q_i$ drawn by a chosen sampling algorithm. (a-d) $n = 3$ with $m = 5, 9, 16, 27$ and (e) $n = 4, m = 16$. Red: TVD obtained from exact simulation. Blue: TVD obtained from the linear optics based algorithm with $k = 2$. Green: TVD obtained from the linear optics based algorithm with $k = 3$. Horizontal dashed lines correspond to the TVD in the asymptotic limit of very large sample size.

[8] M. Reck, A. Zeilinger, H. J. Bernstein, and P. Bertani, Phys. Rev. Lett. 73, 58 (1994).
[9] J. Eisert, Phys. Rev. Lett. 95, 040502 (2005).