UNIVERSITY OF BRISTOL

Peer reviewed version

Link to published version (if available):
10.1038/nphys4270

Link to publication record in Explore Bristol Research
PDF-document

University of Bristol - Explore Bristol Research
General rights

Classical boson sampling algorithms with superior performance to near-term experiments

Alex Neville[1], Chris Sparrow[1,2], Raphaël Clifford[3], Eric Johnston[1], Patrick M. Birchall[1], Ashley Montanaro[4] & Anthony Laing[1*]

[1]*Quantum Engineering and Technology Laboratories, School of Physics and Department of Electrical and Electronic Engineering, University of Bristol, UK*

[2]*Department of Physics, Imperial College London, UK*

[3]*Department of Computer Science, University of Bristol, UK*

[4]*School of Mathematics, University of Bristol, UK*

It is predicted that quantum computers will dramatically outperform their conventional counterparts. However, large-scale universal quantum computers are yet to be built. Boson sampling [1] is a rudimentary quantum algorithm tailored to the platform of linear optics, which has sparked interest as a rapid way to demonstrate this quantum supremacy [2–6]. Photon statistics are governed by intractable matrix functions, which suggests that sampling from the distribution obtained by injecting photons into a linear-optical network could be solved more quickly by a photonic experiment than by a classical computer. The apparently

low resource requirements for large boson sampling experiments have raised expectations of a near-term demonstration of quantum supremacy by boson sampling [7,8]. Here we present classical boson sampling algorithms and theoretical analyses of prospects for scaling boson sampling experiments, showing that near-term quantum supremacy via boson sampling is unlikely. Our classical algorithm, based on Metropolised independence sampling, allowed the boson sampling problem to be solved for 30 photons with standard computing hardware. Compared to current experiments, a demonstration of quantum supremacy over a successful implementation of these classical methods on a supercomputer would require the number of photons and experimental components to increase by orders of magnitude, while tackling exponentially scaling photon loss.

It is believed that new types of computing machines will be constructed to exploit quantum mechanics for an exponential speed advantage in solving certain problems compared with classical computers [9]. Recent large state and private investments in developing quantum technologies have increased interest in this challenge. However, it is not yet experimentally proven that a large computationally useful quantum system can be assembled, and such a task is highly non-trivial given the challenge of overcoming the effects of errors in these systems.

Boson sampling is a simple task which is native to linear optics and has captured the imagination of quantum scientists because it seems possible that the anticipated supremacy of quantum machines could be demonstrated by a near-term experiment. The advent of integrated quantum photonics [10] has enabled large, complex, stable and programmable optical circuitry [11,12], while

recent advances in photon generation [13–15] and detection [16,17] have also been impressive. The possibility to generate many photons, evolve them under a large linear optical unitary transformation, then detect them, seems feasible, so the role of a boson sampling machine as a rudimentary but legitimate computing device is particularly appealing. Compared to a universal digital quantum computer, the resources required for experimental boson sampling appear much less demanding. This approach of designing quantum algorithms to demonstrate computational supremacy with near-term experimental capabilities has inspired a raft of proposals suited to different hardware platforms [18–20].

Based on a simple architecture, the boson sampling problem is similarly straightforward to state. A number n of indistinguishable noninteracting bosons (e.g. photons) should be injected into n input ports of a circuit comprised of a number m of linearly coupled bosonic modes. The circuit should be configured so that the transformation between input and output ports is described by a uniformly ("Haar") random unitary matrix. The probability for the n bosons to be detected at given set of n output ports is equal to the square of the absolute value of the permanent of the transfer matrix that describes the transformation.

While choosing a number of modes $m \sim n^5 \log^2 n$ guarantees that the distribution of any $n \times n$ sub-matrix is approximately equal to that of a matrix of elements drawn independently from the complex normal distribution [1], the less impractical scaling of $m \sim n^2$ is typically targeted. This polynomial relation between n and m is also important because it ensures a not too large probability that two or more of the bosons arrive at the same output port, i.e. bunch; the conjectured hardness

only applies to collision-free events, i.e. no bunching. Because approximating the permanent of a random matrix is conjectured to be computationally hard [1], calculating any transition probability is intractable; the collection of all of the possible collision-free transition probabilities (m choose n) constitutes an exponentially large probability distribution, where each element is exponentially hard to calculate. Running an ideal boson sampler would solve the problem of producing samples from this distribution.

Importantly, a strong case for the classical hardness of boson sampling can be made even when the distribution being sampled from is only approximately correct [1]: assuming certain conjectures from computational complexity theory, there can be no efficient classical algorithm to sample from any distribution within a small distance from the boson sampling distribution.

Current estimates for the regime in which photonic experiments could achieve quantum supremacy have been between 20 and 30 photons [1,21,22] and, recently, as low as 7. [7] However, our classical algorithm, based on Metropolised independence sampling (MIS), while necessarily inefficient for large n, was able to output a sample for $n = 30$ bosons in half an hour on a standard laptop, and would enable an $n = 50$ sample to be produced in under 10 days on a supercomputer [23]. MIS [24] is a specific Markov Chain Monte Carlo (MCMC) method. For an instance of the problem, our aim is to construct a Markov chain where each state in the chain signifies a boson sampling detection event. New states in the chain are proposed from a classical mockup of the boson sampling distribution: the distribution of distinguishable particles, where probabilities of detection events are equal to permanents of real and positive transition matrices, and sampling becomes ef-

ficient [25]. Proposed states are then accepted or rejected by comparison with the actual probability to observe that event for indistinguishable bosons. This means that, at each step, only a relatively small number of permanents must be calculated; a calculation of the full and exponentially large boson sampling distribution is not required.

More precisely, let P_{D} be the distinguishable particle distribution with probability mass function $g(x)$, over the set of tuples of length m with elements in $\{0, 1\}$ which sum to n. And let P_{BS} be the boson sampling distribution over these tuples, with probability mass function $f(x)$. Then starting at a random (according to P_{D}) tuple x, propose a new random tuple x'. The transition from x to x' is accepted with probability

$$T\left(x'|x\right) = \min\left(1, \frac{f(x')}{f(x)} \frac{g(x)}{g(x')}\right). \tag{1}$$

Repeating this procedure generates a Markov chain, which will converge such that thereafter, the chain is sampling from P_{BS}.

Not all states in the Markov chain are retained as detection events. The time taken for the Markov chain to converge means that a number of tuples at the beginning of the chain must be discarded, known as the 'burn in' period, τ_{burn}. For the size of simulations covered here, empirical tests find that a burn in period of 100 is sufficient for convergence to have occurred. In addition, autocorrelation between states in the chain can occur, for example because two consecutive states in the chain will be identical whenever a proposed new state x' is not accepted. We empirically find for the size of problem we tackle that autocorrelation is suppressed with a thinning procedure that retains only every 100th state (see Supplementary Information). Generally, the burn in period

5

and thinning interval are reduced by a greater overlap between target and proposal distributions, as measured by (1), the transition probability. We find that a proposal distribution of the distinguishable particle transition probabilities has a high acceptance rate of $\sim 40\%$, a sign that the distributions overlap well. In Fig. 1 we present a schematic of the MIS based approach to boson sampling, alongside schematics of a quantum photonic approach and the boson sampling problem itself.

The classical resources used to produce the thinned Markov chain are far fewer than those required by the brute force approach of calculating all permanents in the full distribution. The asymptotically fastest known algorithm for exactly computing the permanent of an $n \times n$ matrix is Ryser's algorithm [26], whose runtime when implemented efficiently is $\mathcal{O}(n2^n)$. Generating the first tuple in a sample requires the computation of τ_{burn} $n \times n$ real valued matrix permanents and τ_{burn} $n \times n$ complex valued matrix permanents. Each subsequent sample requires τ_{thin} $n \times n$ real valued and complex valued matrix permanents to be computed. The relative scaling of the approaches to classical boson sampling using a standard laptop is shown in Fig. 2(b), setting both τ_{burn} and τ_{thin} equal to 100. The MIS sampler is over 50 orders of magnitude faster for the $n = 30$ case than the brute force computation of the entire distribution.

We used this algorithm on a standard laptop to produce samples of size 20,000 for up to 20 bosons, and used a local server, which allowed around 30 times more chains to be run in parallel, to produce 250 samples for 30 bosons in 900 modes in less than five hours. As in the experimental case, a central challenge is to provide evidence for sampling from the correct distribution. Here we

addressed this using standard statistical techniques. Testing for distinguishability among photons is one of the main verification tools used in experiments, while here, the distribution resulting from distinguishable photons has the desirable property of being somewhat correlated with the ideal boson sampling distribution. It is conceivable (though not necessary) that results from an inaccurate sampling algorithm could sometimes be closer to the distinguishable distribution than the ideal boson sampling distribution. This is especially relevant when the sampling algorithm itself uses samples from the distinguishable distribution. The likelihood ratio tests [27] in Fig. 2(a) show a rapid growth in confidence in the hypothesis that these samples are from the indistinguishable boson distribution rather than the distinguishable particle distribution for $n = 7, 12, 20$ and 30.

Further results for verification against the distinguishable distribution are shown in Fig. 2(c) for the case of 7 bosons in 49 modes. For each tuple in a sample size of 20,000 produced by our classical algorithm, we calculate $-\log(|\mathrm{Per}(A_S)|^2)$, where A_S is the matrix associated to each tuple, and produce a probability mass histogram. The same function and associated histogram is plotted for a sample of 20,000 tuples chosen from the distinguishable particle distribution (note that $\mathrm{Per}(|A_S|^2)$ gives the probability to observe a transition of distinguishable particles). Fig. 2(c) shows the clear difference between the two distributions, which we analyse with a (bootstrapped) 2-sample Kolmogorov-Smirnov (KS) test [28]. We are able to reject the null hypothesis, that the two samples are chosen from the same distribution, at a significance level of 0.001.

We implement stronger verification tests by comparing (up to computational limits) our MIS algorithm against algorithms that are known to be exact. The KS tests in Fig. 2(c) between distribu-

tions from our MIS algorithm and a rejection sampling algorithm, and between our MIS algorithm

and the brute force approach of calculating all permanents in the full distribution, both found large

p-values (see Supplementary Information). This striking level of consistency (including further

verification against rejection sampling for 12 photons - see Supplementary Information) is reassur-

ing evidence for the correct operation of our MIS algorithm.

We next compare our classical approach with plausible experimental parameters. It is worth

noting that asymptotically, experimental boson sampling will have a slower runtime than our al-

gorithm. This is because photon losses scale exponentially badly with n. [1,29] The runtime for an

experiment with a transmission probability resulting from fixed loss (generation, coupling, detec-

tion) η_f and a transmission probability η_0, resulting from loss per unit optical circuit depth scales

as $\mathcal{O}\left(\left(\frac{1}{\eta_f}\right)^n\left(\frac{1}{\eta_0}\right)^{dn}\right)$ for an optical circuit depth of d, which is worse than Ryser's algorithm if d

grows with n for any $\eta_f, \eta_0 < 1$. However, the region of interest for quantum supremacy is likely

to be restricted to $n < 100$, where low-loss experiments still have the potential to produce large

speedups. Assuming that our MIS sampler continues to perform equally well for larger instance

sizes, we can compare its runtime with current and future experiments. The classical and quantum

runtimes for an instance of size n bosons in $m = n^2$ modes can be estimated as

$$c_t(n) = a100n2^n \tag{2}$$

$$q_t(n, \eta) = \frac{e}{R\eta^n} \tag{3}$$

where a is the time scaling of the classical computer (for computing one real and one complex

permanent), the factor of e is an approximation to the probability of obtaining a collision-free

event [30], R is the experimental source repetition rate, and $\eta = \eta_{\mathrm{f}}\eta_0^d$ is the experimental transmission probability of a single photon including the efficiencies of photon generation, coupling, circuit propagation and detection (note that R and η will generally be a function of n). We define the *quantum advantage* (QA) as the improvement in quantum runtime versus classical runtime measured in orders of magnitude,

$$\mathrm{QA}(n, \eta) = \max\left[0, \log_{10}\left(\frac{c_t}{q_t}\right)\right]. \tag{4}$$

We now consider two plausible notions of quantum supremacy. First, we can define supremacy as a speedup so large that it is unlikely to be overcome by algorithmic or hardware improvements to the classical sampler, for which we choose a speedup of ten orders of magnitude. Secondly, we may wish to define supremacy as the point at which a computational task is performed in a practical runtime on a quantum device, for which we choose under a week, but in an impractical runtime on a classical device, for which we choose over a century.

These criteria can be summarised as

$$\mathrm{QS}_1 : \mathrm{QA} > 10 \tag{5}$$

$$\mathrm{QS}_2 : q_t < 1\ \text{week}, c_t > 100\ \text{years}. \tag{6}$$

In order to make concrete estimates of future runtimes, we need to fix a and R. Choosing $a = 3n \times 10^{-15}$s as the time scaling for computing one real and one complex matrix permanent recently reported for the supercomputer *Tianhe 2* [23] and $R = 10$GHz, which is faster than any experimentally demonstrated photon source to our knowledge, we can plot QA against n and η.

We first note that current approaches using spontaneous parametric down conversion (SPDC) photon pairs are generally inefficient with $\eta < 0.002$. [2–5,11,22,31,32] Recently, improved rates have been demonstrated with quantum dot photon sources [8,33,34]. The current leading experimental demonstration however is still restricted to $\eta \approx 0.08$ for $n = 5$ where $q_t \approx 10^9 c_t$. This calculation includes the rate used in the experiment $(76n^{-1}\text{MHz})$ and includes a suppression factor caused by a lower collision-free event rate using a linear instead of quadratic mode scaling. In Wang *et al.* [8] a number of realistic, near-term experimental improvements are suggested to reach 20 photon boson sampling. Using these projections we find that η is increased to ≈ 0.35, which would be a major experimental breakthrough. However, as shown in Fig. 2(d), in this case we predict that the classical runtime would still be over six orders of magnitude faster. Fig. 2(e) shows the regions of quantum advantage and quantum supremacy with current and projected experiments.

In Ref. 35, the authors showed that the boson sampling problem can be modified to allow for a fixed number of lost photons at the input of the circuit whilst retaining computational hardness. We show that if the overall losses in the experiment are path-independent then this is equivalent to loss at the input (see Methods). The MIS sampler can be readily adapted to this scenario by adding an initial step which generates a uniformly random input subset, followed by the usual MIS method for this input state. The dashed contours and lines in Fig. 2(e) and (d) take into account the adjusted classical and quantum runtimes when up to two lost photons are allowed. Although allowing loss helps the experiments to compete, the complexity of realistic experimental regimes such as losing a constant fraction of photons remains unknown and it is easy to see that losing too many photons eventually allows the problem to become efficiently solvable classically.

We have shown that the boson sampling problem can likely be solved for 50 photons and any given Haar random unitary matrix, in a feasible time frame using todays classical technology. While our classical methods for boson sampling are not optimized, they are roughly as effective as they could possibly be without violating standard complexity-theoretic hardness assumptions. And while likely incremental improvements would push the supremacy threshold further still from current experimental capabilities, our methods serve as the first rigorous benchmark for laboratory experiments. To solve the boson sampling problem at a scale that is beyond the capabilities of the best classical computers, experiments face the initial challenge of efficiently generating upwards of 50 photons in well defined modes. Low loss photon propagation is then required in circuitry of several thousands of modes, before arriving at several thousands of high efficiency detectors. Programming different Haar random unitary matrices over those modes requires the precise setting of millions of phase shifters [36].

Addressing these challenges could be more feasible through the use of other (non-spatial) encodings such as the time domain [37,38], though low loss programmability in other domains at this scale brings additional challenges.

Although the boson sampling algorithm could be run on a fault-tolerant quantum computer, this approach would lose the appealing simplicity of the original proposal. An interesting direction for boson sampling is the development of basic error correction techniques that are far short of what is required for universal digital quantum computing, yet sufficient to enable a demonstration of quantum supremacy [18]. Such error correction schemes might then be carried over into other

photonic quantum technologies. Finally, we note that our classical sampling methods could be more broadly applicable to other sampling algorithms, not least to the many variants of boson sampling that have arisen [39,40].

Acknowledgements

Author contributions

All authors contributed to discussions and project development. Simulations were carried out by AN. Data were analysed by AN, CS, RC, AM, and AL. The MIS algorithm and the Random Restart Hill Climbing routine used in our RS algorithm are by RC. Additional classical sampling algorithms were designed and implemented by AN and RC, with input from CS, AM, and AL. Code optimization was carried out by EJ. The proof that photon loss can be considered at the input is by PB. The manuscript was written by AL, AN, and CS, with input from RC and AM. The project was conceived and managed by AL.

Correspondence

Correspondence should be addressed to AL (email: anthony.laing@bristol.ac.uk).

References

1. Aaronson, S. & Arkhipov, A. The computational complexity of linear optics. In *Proc. 43rd Annual ACM Symp. Theory of Computing*, 333–342 (2011).

2. Broome, M. A. *et al.* Photonic boson sampling in a tunable circuit. *Science* **339**, 794–798 (2013).

3. Tillmann, M. *et al.* Experimental boson sampling. *Nat. Photon.* **7**, 540–544 (2013).

4. Spring, J. B. *et al.* Boson Sampling on a Photonic Chip. *Science* **339**, 798–801 (2013).

5. Crespi, A. *et al.* Integrated multimode interferometers with arbitrary designs for photonic boson sampling. *Nat. Photon.* **7**, 545–549 (2013).

6. Lund, A. P., Bremner, M. J. & Ralph, T. C. Quantum sampling problems, bosonsampling and quantum supremacy. *npj Quantum Information* **3**, 15 (2017).

7. Latmiral, L., Spagnolo, N. & Sciarrino, F. Towards quantum supremacy with lossy scattershot boson sampling. *New Journal of Physics* **18**, 113008 (2016).

8. Wang, H. *et al.* High-efficiency multiphoton boson sampling. *Nat. Photon.* **11**, 361–365 (2017).

9. Shor, P. W. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM Journal on Computing* **26**, 1484–1509 (1997).

10. Politi, A., Cryan, M. J., Rarity, J. G., Yu, S. & O'Brien, J. L. Silica-on-silicon waveguide quantum circuits. *Science* **320**, 646–649 (2008).

11. Carolan, J. *et al.* Universal linear optics. *Science* **349**, 711–716 (2015).

12. Harris, N. C. *et al.* Quantum transport simulations in a programmable nanophotonic processor. *Nat. Photon.* **11**, 447–452 (2017).

13. Ding, X. *et al.* On-demand single photons with high extraction efficiency and near-unity indistinguishability from a resonantly driven quantum dot in a micropillar. *Phys. Rev. Lett.* **116**, 020401 (2016).

14. Somaschi, N. *et al.* Near-optimal single-photon sources in the solid state. *Nat. Photon.* **10**, 340–345 (2016).

15. Spring, J. B. *et al.* Chip-based array of near-identical, pure, heralded single-photon sources. *Optica* **4**, 90–96 (2017).

16. Lita, A. E., Miller, A. J. & Nam, S. W. Counting near-infrared single-photons with 95% efficiency. *Opt. Express* **16**, 3032–3040 (2008).

17. Hadfield, R. H. Single-photon detectors for optical quantum information applications. *Nat. Photon.* **3**, 696–705 (2009).

18. Bremner, M. J., Montanaro, A. & Shepherd, D. J. Achieving quantum supremacy with sparse and noisy commuting quantum computations. *Quantum* **1**, 8 (2017).

19. Boixo, S. *et al.* Characterizing quantum supremacy in near-term devices. Preprint at `https://arxiv.org/abs/1608.00263` (2016).

20. Bermejo-Vega, J., Hangleiter, D., Schwarz, M., Raussendorf, R. & Eisert, J. Architectures for quantum simulation showing quantum supremacy. Preprint at `https://arxiv.org/abs/1703.00466` (2017).

21. Preskill, J. Quantum computing and the entanglement frontier. Preprint at `https://arxiv.org/abs/1203.5813` (2012).

22. Bentivegna, M. *et al.* Experimental scattershot boson sampling. *Science Advances* **1**, e1400255 (2015).

23. Wu, J. *et al.* Computing permanents for boson sampling on Tianhe-2 supercomputer. Preprint at `https://arxiv.org/abs/1606.05836` (2016).

24. Liu, J. Metropolized independent sampling with comparisons to rejection sampling and importance sampling. *Statistics and Computing* **6**, 113–119 (1996).

25. Aaronson, S. & Arkhipov, A. BosonSampling Is Far From Uniform. *Quant. Inf. Comp.* **14**, 1383–1432 (2014).

26. Ryser, H. J. *Combinatorial Mathematics*. Carus Mathematical Monographs, No. 14 (1963).

27. Bentivegna, M. *et al.* Bayesian approach to boson sampling validation. *Int. J. Quantum Inform.* **12**, 1560028 (2014).

28. Praestgaard, J. T. Permutation and bootstrap Kolmogorov-Smirnov tests for the equality of two distributions. *Scandinavian Journal of Statistics* **22**, 305–322 (1995).

29. Rohde, P. P., Motes, K. R., Knott, P. A. & Munro, W. J. Will boson-sampling ever disprove the extended church-turing thesis? Preprint at `https://arxiv.org/abs/1401.2199` (2014).

30. Arkhipov, A. & Kuperberg, G. The bosonic birthday paradox. *Geometry & Topology Monographs* **18**, 1–7 (2012).

31. Spagnolo, N. *et al.* Experimental validation of photonic boson sampling. *Nat. Photon.* **8**, 615–620 (2014).

32. Carolan, J. *et al.* On the experimental verification of quantum complexity in linear optics. *Nat. Photon.* **8**, 621–626 (2014).

33. Loredo, J. C. *et al.* Boson sampling with single-photon fock states from a bright solid-state source. *Phys. Rev. Lett.* **118**, 130503 (2017).

34. He, Y. *et al.* Time-bin-encoded boson sampling with a single-photon device. *Phys. Rev. Lett.* **118**, 190501 (2017).

35. Aaronson, S. & Brod, D. J. BosonSampling with lost photons. *Phys. Rev. A* **93**, 012335 (2016).

36. Russell, N. J., Chakhmakhchyan, L., OBrien, J. L. & Laing, A. Direct dialling of Haar random unitary matrices. *New Journal of Physics* **19**, 033007 (2017).

37. Motes, K. R., Gilchrist, A., Dowling, J. P. & Rohde, P. P. Scalable boson sampling with time-bin encoding using a loop-based architecture. *Phys. Rev. Lett.* **113**, 120501 (2014).

38. Pant, M. & Englund, D. High-dimensional unitary transformations and boson sampling on temporal modes using dispersive optics. *Phys. Rev. A* **93**, 043803 (2016).

39. Huh, J., Guerreschi, G. G., Peropadre, B., McClean, J. R. & Aspuru-Guzik, A. Boson sampling for molecular vibronic spectra. *Nat. Photon.* **9**, 615–620 (2015).

40. Hamilton, C. S. *et al.* Gaussian boson sampling. Preprint at `https://arxiv.org/abs/1612.01199` (2016).

Figures:

(a) Boson sampling problem

Input: U: Haar-random unitary

$$\begin{pmatrix} U_{11} & U_{12} & U_{13} & \dots & U_{1m} \\ U_{21} & U_{22} & U_{23} & \dots & U_{2m} \\ U_{31} & U_{32} & U_{33} & \dots & U_{3m} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ U_{m1} & U_{m2} & U_{m3} & \dots & U_{mm} \end{pmatrix}$$

$$A_S = \begin{pmatrix} U_{11} & U_{12} & U_{13} \\ U_{31} & U_{32} & U_{33} \\ U_{m1} & U_{m2} & U_{m3} \end{pmatrix}$$

$$\Pr_{P_{\text{BS}}}(S) = |\text{Per}(A_S)|^2$$

Output: sample from P_{BS}

(b) Quantum photonic solver

$|\Psi_{\text{in}}\rangle = |11\dots00\dots\rangle$

R: repetition rate

Input loss

Linear Optical Network

Transmission loss

Detector loss

(c) Classical MIS solver

P_{D}: distinguishable particle distribution
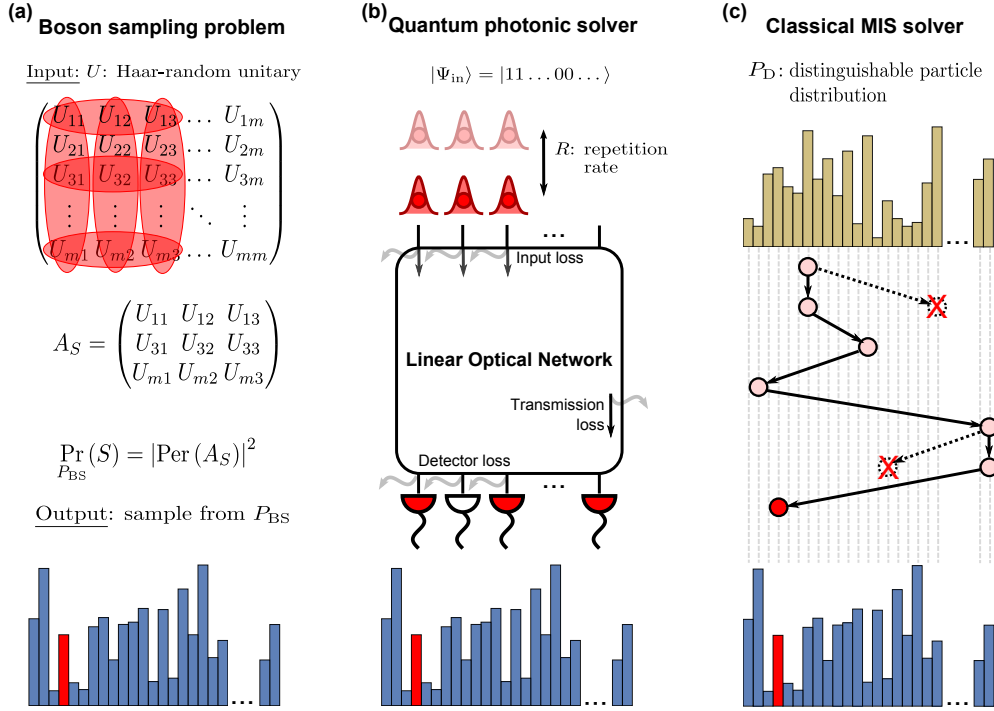
Figure 1: Photonic and classical approaches to the boson sampling problem. (a) Definition of the problem. Output a sample from the distribution defined by the modulus squared permanents of submatrices of a Haar-random unitary matrix U. (b) Photonic experiments solve the problem by propagating single photons through a linear optical network followed by single photon detection and can be broadly parameterised by R, the n-photon generation rate, and η, the transmission probability for a single photon taking into account input, coupling, transmission and detection losses. (c) A classical boson sampling algorithm based on Metropolised independence sampling using the distinguishable particles transition probabilities as the proposal distribution. The algorithm computes 100 complex and real permanents to produce a single output pattern, and enabled classical boson sampling for 30 bosons on a laptop.
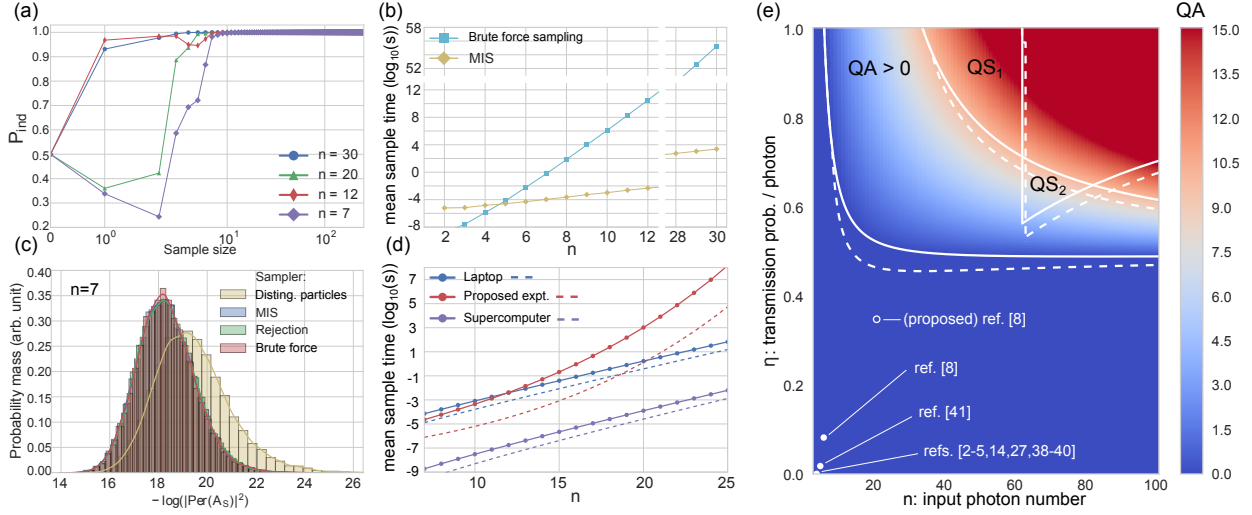
Figure 2: Results and projections for classical boson sampling. (a) A likelihood ratio test with the boson sampling and distinguishable particle distributions as the hypotheses for sample sizes of up to 250 for $n = 7, 12, 20$ and 30 bosons. P_{ind} is the probability that the data are drawn from the boson sampling and not the distinguishable particle distribution. (b) Mean time to get a sample value using a laptop via the MIS and brute force approach to classical boson sampling, averaged over samples of size 100. (c) Verification of sampler by comparing the distribution of $-\log(|\text{Per} A_S|^2)$ for a sample size of 20000 to other boson samplers (rejection sampler and brute force sampler) and a distinguishable particle sampler. (d) Mean time to get a sample using a laptop, supercomputer and the proposed experiment in Ref. 8. Dashed lines represent the time to get sample in a variant of boson sampling where 2 photons are lost. (e) Quantum advantage, QA, as a function of n and η assuming the classical time scaling of a supercomputer and an experimental rate $R = 10$GHz. Lines separate the regions of no quantum advantage, positive quantum advantage and quantum supremacy (as measured by criterion QS_1 or QS_2). Dashed lines demonstrate adjusted regions when up to 2 photons can be lost (optimised to maximise QA). A represents the proposed experiment in Ref. 8, B represents the experiment in Ref. 8, C represents the experiment in Ref. 34 and D represents the experiments in Refs. 2–$5, 11, 22, 31$–33.

Methods

The boson sampling problem [1]**.** Given $U \in SU(m)$, let A be the column-orthonormal, $m \times n$ matrix formed by taking the first n columns of U. Also let $\Phi_{m,n}$ be the set of all possible tuples of length m of non-negative integers whose elements sum to n.

For some tuple $S = (s_1, \ldots, s_m) \in \Phi_{m,n}$, let A_S be the sub-matrix of A with s_i copies of row i. Boson sampling is the problem of sampling from the probability distribution P_{BS} over $\Phi_{m,n}$, for a given input U, with probabilities defined in terms of permanents of submatrices as:

$$\Pr(S) = \frac{|\mathrm{Per}\,(A_S)|^2}{s_1! \ldots s_n!}. \tag{7}$$

In this work we restrict ourselves to the collision free subspace (CFS), meaning that $s_i \in \{0, 1\} \, \forall \, i$. See Supplementary Information for more detail.

Distinguishable particle distribution. The proposal used in the MIS algorithm is that which describes the same distribution for distinguishable particles (7):

$$\Pr_{P_{\mathrm{D}}}(S) = \frac{\mathrm{Per}\,(|A_S|^2)}{s_1! \ldots s_n!}, \tag{8}$$

where for a complex matrix A with elements A_{ij}, $|A|^2$ denotes the matrix with elements $|A_{ij}|^2$. However, there is a classical algorithm which can sample from this distribution in time $O(mn)$ [25].

Brute force exact sampling. This method computes all of the probabilities in the CFS and samples from the associated probability mass function, and is computationally demanding given the exponential scaling of the CFS with the size of the problem. This approach requires computing

$\binom{m}{n}$ permanents of $n \times n$ complex valued matrices before a single sample can be output. Using the lower bound $\binom{m}{n} \geq (m/n)^n$, for $m \geq n^2$ at least n^n permanents must be computed; and even for $n = 10$ we need to compute more than 17×10^{12} permanents. The computation quickly becomes swamped by the number of permanents to compute, rather than the complexity of computing the permanent itself.

Rejection sampling with a uniform proposal. Rejection sampling is a general approach for exactly sampling from a desired distribution P with probability mass function $f(x)$, given the ability to sample from a distribution Q with probability mass function $g(x)$, where $f(x) \leq \lambda g(x)$ for some λ and all x. The algorithm proceeds as follows:

1. Generate a sample x from Q.

2. With probability $\frac{f(x)}{\lambda g(x)}$, output x. Otherwise, go to step 1.

The probability that an accepted sample is generated is

$$\sum_x g(x) \frac{f(x)}{\lambda g(x)} = \frac{1}{\lambda}.$$

Here we take P to be the boson sampling distribution restricted to the CFS, and Q to be the uniform distribution on the CFS (so $N = \binom{m}{n}$). Note that P is subnormalised, so is not quite a probability distribution. However, the rejection sampling algorithm is blind to this subnormalisation (as this is effectively the same as increasing λ), so will generate samples from the renormalised distribution.

Each iteration of rejection sampling requires the computation of one permanent, corresponding to $f(x)$. In order to use rejection sampling to exactly sample from the target distribution with a uniform proposal most efficiently, it is required to know the maximum value of $f(x)$, which corresponds to the largest permanent of all $n \times n$ submatrices A of a given $m \times m$ unitary matrix U. Without any bound on this quantity, we would be forced to use the trivial bound $\mu = 1$, corresponding to $\binom{m}{n}$ permanent computations being required to obtain one sample from P.

As we are only attempting to perform *approximate* boson sampling, we only require a good estimate of $\mu = \max_x f(x)$. It was argued in 1 that sampling from a distribution within total variation distance ϵ of the real boson sampling distribution P_{BS} should be computationally difficult, for some small constant ϵ. Imagine that our guess $\widetilde{\mu}$ for μ is too small, such that $\sum_{x, f(x) > \widetilde{\mu}} f(x) = \epsilon > 0$. Then if x is sampled uniformly at random and $f(x) > \widetilde{\mu}$, step 2 of the rejection sampling algorithm will fail. If we modify the rejection sampling algorithm to simply produce a new uniform sample in this case and repeat, it is easy to see that we can view the modified algorithm as sampling from the truncated distribution P_{low} with probability mass function

$$
\widetilde{f}(x) =
\begin{cases}
\frac{f(x)}{\sum_{x, f(x) \leq \widetilde{\mu}} f(x)} & \text{if } f(x) \leq \widetilde{\mu} \\
0 & \text{if } f(x) > \widetilde{\mu}
\end{cases}.
$$

Then the total variation distance between P_{low} and P is

$$
\frac{1}{2} \sum_x |f(x) - \widetilde{f}(x)| = \sum_{x, f(x) > \widetilde{\mu}} f(x).
$$

So if the probability mass of P above μ is at most ϵ, we have sampled from a distribution within distance ϵ of P. We have found that we are able to use a simple random restart hill climbing algorithm to provide a suitable estimate of μ with $\mathcal{O}(m^2 n)$ computations of $n \times n$ matrix permanents.

Random restart hill climbing algorithm. We start by randomly sampling a submatrix, represented by the tuple S, from the uniform distribution. For one pass, we greedily try replacing each row in the sampled submatrix by each row from A in turn, accepting only if this increases $\Pr(S)$ while also making sure to avoid selecting the same row twice. We perform repeat passes until there is no improvement of the probability over a complete pass. At this point we randomly resample a new starting submatrix and repeat from the beginning. The total number of permanent calculations for one pass is $n(m - n)$.

This method is not guaranteed to find a global maximum. However, in our experiments for $n \leq 7$ where we can still compute the full probability mass function exactly, we found the estimates for the maximum probability to be exactly equal to the global maximum in the overwhelming majority of cases. In the range $8 \leq n \leq 12$ where we no longer are able to compute the exact maximum probability, the bounds from our hill-climbing algorithm also allowed us to sample using rejection sampling efficiently and then compare our results with our MIS sampler. This provided further evidence for both sampling techniques.

Scattershot boson sampling. We can straightforwardly modify our MIS-based method to carry out the scattershot boson sampling problem. For each sample that we wish to output, we can first sample (efficiently) from the uniform distribution on n-fold input modes (which fixes the columns of U contributing to submatrices), before running the algorithm in the way described above for a single sample. In this case, τ_{thin} becomes meaningless and we are only interested in τ_{burn}, as we start a new chain for each sample.

Alternatively, our proposal distribution can be changed to include the uniform distribution over n-fold input modes, meaning that each state in a given Markov chain can correspond to a different input configuration.

Events with collisions. Although in this work we do not examine the more general situation of boson sampling where there can be more than one boson in an output mode, we anticipate that relaxing the CFS restriction will not increase the run time of the MIS method. In fact, it is possible that the average run time could be decreased with this relaxation, as there exists an algorithm for computing the permanent which is exponential in matrix rank, rather than matrix size [41]. However, due to there usually existing very large permanents of sub-unitary matrices with many repeated rows, relaxing the CFS restraint has an adverse effect on the average run time of our rejection sampling method.

Likelihood ratio test. We perform a modified version of the likelihood ratio test described by Bentivegna et al. [27] (see Supplementary Information for more detail). Defining two hypotheses $\mathcal{Q}$ and $\mathcal{R}$, the indistinguishable boson hypothesis the distinguishable particle hypothesis respectively, and letting q_x be the probability of seeing the sampled event x according to hypothesis $\mathcal{Q}$, and r_x be the corresponding probability under hypothesis $\mathcal{R}$, we can write

$$\frac{P\left(\mathcal{Q}|N_{\text{events}}\right)}{P\left(\mathcal{R}|N_{\text{events}}\right)} = \prod_{x=1}^{N_{\text{events}}} \left(\frac{q_x}{r_x}\right) = \mathcal{X},$$ (9)

where N_{events} is the sample size. Normalising equation (9) such that $P\left(\mathcal{Q}|N_{\text{events}}\right) + P\left(\mathcal{R}|N_{\text{events}}\right) = 1$ yields

$$P_{\text{ind}} \equiv P\left(\mathcal{Q}|N_{\text{events}}\right) = \frac{1}{\mathcal{X}+1} \prod_{x=1}^{N_{\text{events}}} \left(\frac{q_x}{r_x}\right).$$ (10)

24

We note that q_x and r_x should not simply correspond to $|\mathrm{Per}\,(A_x)|^2$ and $\mathrm{Per}\,(|A_x|^2)$ respectively. The reason for this is that we are restricted to the CFS, and the probability of a sample being collision-free differs between indistinguishable bosons and distinguishable particles. So q_x and r_x must be normalised independently, such that they independently sum to 1 over all events x. Doing this exactly would require summing all probabilities in the CFS for the specific instance of the problem being considered. To approximate the normalisation of q_x efficiently, here we instead average the probability of the output being collision-free over the Haar measure [1,30]:

$$P_{\mathrm{CFS}} \approx \binom{m}{n} \Big/ \binom{m+n-1}{n} . \tag{11}$$

For the hypothesis $\mathcal{R}$, we can efficiently sample output tuples, and so the ratio of collision-free tuples to tuples with collisions in a large sample provides an approximate normalisation for r_x.

For each problem size, we use this likelihood ratio test to assess the performance of MIS samplers with different τ_{burn} and τ_{thin}. As the proposal distribution for the sampler is the distinguishable particle distribution, we might expect that if the chain has not converged to the target distribution, this will manifest itself as samples looking more like they are from the distinguishable particle distribution than they should. Also, we expect that samplers with larger τ_{burn} and τ_{thin} are more likely to sample from the target distribution. Because of this, we expect to be able to observe a point at which increasing τ_{burn} and τ_{thin} has, on average, no effect on the outcome of a likelihood ratio test between hypotheses $\mathcal{Q}$ and $\mathcal{R}$.

Lossy boson sampling. In this variant [35], we assume that $n - k$ photons are lost before entering the circuit enacting the linear optical transfer matrix, so k photons remain. Probabilities in this

setting are not determined directly by $|\mathrm{Per}\,(A_S)|^2$ for some m-tuple S, but by the average of this quantity over all possible ways of losing $n - k$ photons from n photons:

$$\mathrm{Pr}(S) = \frac{1}{|\Lambda_{n,k}|} \sum_{T \in \Lambda_{n,k}} |\mathrm{Per}\,(A_{S,T})|^2 \tag{12}$$

where $\Lambda_{n,k}$ is the set of k-subsets of $\{1, \ldots, n\}$ and $A_{S,T}$ is the $k \times k$ submatrix of A obtained by taking columns of A according to T and rows of A according to S, which remains a subset of $\{1, \ldots, m\}$. Note that once again we restrict to the collision-free subspace, making the assumption that the probability of a collision is low enough that this does not significantly affect the probabilities.

If, at most, a constant number of photons in total are lost before entering the circuit, the lossy boson sampling problem remains hard [35]. Open problems are whether this is true in the more realistic setting where a constant fraction of photons are lost, and to generalise the loss model to include loss within and after the linear optical circuit. Here we prove a slightly stronger result than that shown in 42; that as long as the overall transfer matrix is proportional to a unitary, loss can always be considered at the input even if the physical loss channels, wherever they occur, are not uniform (i.e are mode-dependent).

Consider a boson sampling device consisting of an ideal unitary linear optical transformation U on a set of m optical modes which is preceded or succeeded by path-independent loss. This loss can be modelled by considering a set of m additional virtual ancilla modes such that the optical transfer matrix on all $2m$ modes remains unitary. A uniform transmission probability of η can then be described by beamsplitters coupling each mode to its corresponding ancilla, resulting in the

transfer matrix

$$L = \begin{pmatrix} \sqrt{\eta} & \sqrt{1-\eta} \\ \sqrt{1-\eta} & -\sqrt{\eta} \end{pmatrix} \otimes \mathbb{1}_m \tag{13}$$

and so including the interferometer, the full transfer matrices for input and output losses are

$$M_I = (U \oplus \mathbb{1}_m)L = \begin{pmatrix} \sqrt{\eta}U & \sqrt{1-\eta}U \\ \sqrt{1-\eta}\mathbb{1} & -\sqrt{\eta}\mathbb{1} \end{pmatrix} \tag{14}$$

$$M_O = L(U \oplus \mathbb{1}_m) = \begin{pmatrix} \sqrt{\eta}U & \sqrt{1-\eta}\mathbb{1} \\ \sqrt{1-\eta}U & -\sqrt{\eta}\mathbb{1} \end{pmatrix}. \tag{15}$$

Any m-mode optical state can be expressed in a coherent state basis [43]:

$$\rho_m = \int \lambda(\boldsymbol{\alpha}, \boldsymbol{\beta}) |\boldsymbol{\alpha}\rangle\langle\boldsymbol{\beta}| d^m\boldsymbol{\alpha} d^m\boldsymbol{\beta} \tag{16}$$

where $|\boldsymbol{\alpha}\rangle = |(\alpha_1, \alpha_2, ..., \alpha_m)^T\rangle \equiv \bigotimes_{i=1}^m |\alpha_i\rangle$ is an m-mode coherent state and $\langle\boldsymbol{\beta}| = \langle(\beta_1, \beta_2, ...\beta_m)^T| =$

$\bigotimes_{i=1}^m \langle\beta_i|$. A coherent state evolves under a transfer matrix T as

$$\mathcal{U}(T)|\boldsymbol{\alpha}\rangle = \bigotimes_{i=1}^m \left| \sum_j T_{ij}\alpha_j \right\rangle = |T\boldsymbol{\alpha}\rangle \tag{17}$$

It can then be shown that when the initial state contains vacuum in all ancilla modes, $\rho = \rho_m \otimes$

$|0\rangle\langle 0|_{\text{an}}$, the same state is produced in the m system modes under the transformations M_O and M_I

$$
\begin{aligned}
\rho_O &= \text{Tr}_{\text{an}} \left[\mathcal{U}(M_O)\rho\mathcal{U}^\dagger(M_O) \right] \\
&= \text{Tr}_{\text{an}} \left[\int d^{2m}\boldsymbol{\alpha} d^{2m}\boldsymbol{\beta} \lambda(\boldsymbol{\alpha},\boldsymbol{\beta}) \left| \sqrt{\eta}U\boldsymbol{\alpha} \right\rangle\!\left\langle \sqrt{\eta}U\boldsymbol{\beta} \right| \otimes \left| \sqrt{1-\eta}U\boldsymbol{\alpha} \right\rangle\!\left\langle \sqrt{1-\eta}U\boldsymbol{\beta} \right| \right] \\
&= \int d^{2m}\boldsymbol{\alpha} d^{2m}\boldsymbol{\beta} \lambda(\boldsymbol{\alpha},\boldsymbol{\beta}) \left| \sqrt{\eta}U\boldsymbol{\alpha} \right\rangle\!\left\langle \sqrt{\eta}U\boldsymbol{\beta} \right| \text{Tr} \left[\left| \sqrt{1-\eta}U\boldsymbol{\alpha} \right\rangle\!\left\langle \sqrt{1-\eta}U\boldsymbol{\beta} \right| \right] \\
&= \int d^{2m}\boldsymbol{\alpha} d^{2m}\boldsymbol{\beta} \lambda(\boldsymbol{\alpha},\boldsymbol{\beta}) \left| \sqrt{\eta}U\boldsymbol{\alpha} \right\rangle\!\left\langle \sqrt{\eta}U\boldsymbol{\beta} \right| \text{Tr} \left[\mathcal{U}(U) \left| \sqrt{1-\eta}\boldsymbol{\alpha} \right\rangle\!\left\langle \sqrt{1-\eta}\boldsymbol{\beta} \right| \mathcal{U}^\dagger(U) \right] \\
&= \int d^{2m}\boldsymbol{\alpha} d^{2m}\boldsymbol{\beta} \lambda(\boldsymbol{\alpha},\boldsymbol{\beta}) \left| \sqrt{\eta}U\boldsymbol{\alpha} \right\rangle\!\left\langle \sqrt{\eta}U\boldsymbol{\beta} \right| \text{Tr} \left[\left| \sqrt{1-\eta}\boldsymbol{\alpha} \right\rangle\!\left\langle \sqrt{1-\eta}\boldsymbol{\beta} \right| \right] \\
&= \rho_I.
\end{aligned}
\tag{18}
$$

More generally, wherever loss occurs in the experiment, the overall transfer matrix K on the system modes can be efficiently characterised [44,45]. Since path-dependent loss is usually small in experiments, and can be mitigated by interferometer design [46], the matrix $K/||K||_2 \approx U$. The matrix K can then be embedded into a larger unitary matrix acting on additional modes as before. We note all unitary dilations of K, $M_K \in U(m+p)$ where $p \geq m$, can be parameterised using the Cosine-Sine decomposition as

$$
M_K = \begin{pmatrix} A & 0 \\ 0 & X \end{pmatrix} \left(\begin{array}{cc|c} \cos(\Theta) & -\sin(\Theta) & 0 \\ \hline \sin(\Theta) & \cos(\Theta) & 0 \\ 0 & 0 & \mathbb{1}_{p-m} \end{array} \right) \begin{pmatrix} B^\dagger & 0 \\ 0 & Y \end{pmatrix}
$$

where $K = A\cos(\Theta)B^\dagger$, with $A, B \in U(m)$ and $\cos(\Theta) = \text{diag}(\cos\theta_1, ..., \cos\theta_m)$ with $\theta_1 \leq \theta_2 \leq .. \leq \theta_m$, is a singular value decomposition of K and $X, Y \in U(p)$. In fact, all unitary dilations are related by the choice of X and Y [47,48]. Since $\mathcal{U}(Y)|0\rangle = |0\rangle$ and the choice of X

does not affect ρ_K using the cyclic property of the trace as above, setting $\eta = ||K||_2^2$, we see that

$\rho_K = \rho_I$. Moreover, we have shown that all unitary dilations of a transfer matrix produce the same

output state and therefore any boson sampling experiment with overall path-independent losses

is equivalent to introducing uniform loss channels with transmission probability η at the input,

followed by the ideal unitary evolution.

Our MIS method can readily be adapted to deal with loss at the input, by inserting an initial

step for each tuple to be output, which generates a uniformly random input subset T. This would be

followed by the usual MIS method with permanents of $k \times k$ submatrices computed. The core part

of the classical sampling procedure for both the lossy and scattershot variants therefore follows

precisely that of standard boson sampling. From our analysis of the required burn in period for

MIS (see Supplementary Information), we can see that the performance of our sampler will be

similar to the standard boson sampling case. That is, it is likely that lossy and scattershot boson

sampling is no more difficult classically than standard boson sampling.

Methods references

41. Barvinok, A. I. Two algorithmic results for the traveling salesman problem. *Mathematics of Operations Research* **21**, 65–84 (1996).

42. Berry, D. W. & Lvovsky, A. I. Linear-optical processing cannot increase photon efficiency. *Phys. Rev. Lett.* **105**, 203601 (2010).

43. Gerry, C. & Knight, P. *Introductory quantum optics* (Cambridge university press, 2005).

44. Laing, A. & O'Brien, J. L. Super-stable tomography of any linear optical device. Preprint at `https://arxiv.org/abs/1208.2868` (2012).

45. Rahimi-Keshari, S. *et al.* Direct characterization of linear-optical networks. *Opt. Express* **21**, 13450–13458 (2013).

46. Clements, W. R., Humphreys, P. C., Metcalf, B. J., Kolthammer, W. S. & Walmsley, I. A. Optimal design for universal multiport interferometers. *Optica* **3**, 1460–1465 (2016).

47. Horn, R. A. & Johnson, C. R. *Topics in Matrix Analysis* (Cambridge University Press, 1991).

48. Allen, J. & Arceo, D. Matrix dilations via cosine-sine decomposition. Tech. Rep., DTIC Document (2006).

Data availability

Classical boson sampling algorithms with superior performance to near-term experiments – supplementary information

Alex Neville[1], Chris Sparrow[1,2], Raphaël Clifford[3], Eric Johnston[1],
Patrick M. Birchall[1], Ashley Montanaro[4], and Anthony Laing[1]

[1] *Quantum Engineering and Technology Laboratories,*
School of Physics and Department of Electrical and Electronic Engineering, University of Bristol, UK
[2] *Department of Physics, Imperial College London, UK*
[3] *Department of Computer Science, University of Bristol, UK*
[4] *School of Mathematics, University of Bristol, UK*

1. INTRODUCTION

The boson sampling problem has, since its introduction by Aaronson and Arkhipov in 2011 [1], been the subject of great interest within theoretical and experimental quantum physics. Informally, boson sampling is the problem of sampling from the distribution obtained by transmitting photons through a linear-optical network. No efficient classical algorithm for this task is known, but by its nature, boson sampling is suited to being carried out on a specialised linear optical quantum device which could be substantially simpler than a fully universal linear-optical quantum computer. As such, boson sampling is considered a leading candidate for a problem which could see enhanced performance when solved by a quantum device compared with a classical device – so-called *quantum supremacy* [2, 3] – in the near future.

Estimates for the size of experiment required to demonstrate quantum supremacy via boson sampling have evolved over time. Aaronson and Arkhipov initially predicted this to be in the region of 20-30 photons in a 400-900 mode linear optical network [1]; Preskill suggested "about 30" photons [2]; and Bentivegna et al. [4] suggested that, by making some minor modifications to the problem, 30 photons and 100 modes would suffice. Recently Latmiral et al. [5] have reported that the number of photons and modes required to achieve quantum supremacy could be as small as 7 and 50 respectively.

This estimate is based on a comparison with the simplest possible classical technique for simulating boson sampling: evaluating the entire probability distribution. However, there exist well-known classical techniques for exact or approximate sampling that do not require the entire distribution of interest to be determined. Here we assess the requirements on a linear-optical implementation in order to achieve quantum supremacy in light of these.

We find that, using projected loss parameter estimates with a promising experimental boson sampling setup [6], quantum supremacy (via the standard boson sampling problem or in a modified version with loss [7]) is not achieved.

2. THE BOSON SAMPLING PROBLEM

The original boson sampling problem, as described by Aaronson and Arkhipov [1], is based on sampling from the probability mass function defined by the linear scattering of multiple bosons (in practice, photons) prepared in a Fock state and measured in the Fock basis.

Given $U \in SU(m)$, let A be the column-orthonormal, $m \times n$ matrix formed by taking the first n columns of U. Also let $\Phi_{m,n}$ be the set of all possible tuples of length m of non-negative integers whose elements sum to n.

For some tuple $S = (s_1, \ldots, s_m) \in \Phi_{m,n}$, let A_S be the sub-matrix of A with s_i copies of row i. Boson sampling is the problem of sampling from the probability distribution P_{BS} over $\Phi_{m,n}$, for a given input U, with probabilities defined as:

$$\Pr(S) = \frac{|\mathrm{Per}\,(A_S)|^2}{s_1! \ldots s_n!} \tag{1}$$

where $\mathrm{Per}(X)$ is the permanent of an $n \times n$ matrix $X = (x)_{ij}$, defined by:

$$\mathrm{Per}(X) = \sum_{\sigma \in \mathcal{S}_n} \prod_{i=1}^{n} x_{i,\sigma(i)} \tag{2}$$

where $\mathcal{S}_n$ is the group of permutations of integers 1 to n.

The matrix permanent is similar to the more commonly encountered matrix determinant, but the definition lacks an alternating sign. Although this difference is seemingly minor, computing the permanent of a complex-valued matrix falls in the #P-hard computational complexity class [8, 9], and is in general vastly more demanding than computing the determinant. Indeed, the fastest algorithms known for computing the permanent run in time exponential in n.

Using the fact that computing the permanent is a #P-hard problem, Aaronson and Arkhipov [1] have shown that the existence of an efficient, exact classical algorithm for boson sampling would collapse the polynomial hierarchy to its third level, a complexity-theoretic consequence considered extremely unlikely. However, the boson sampling distribution could (in principle) be sampled from using a simple linear-optical circuit with n photons in m modes, where the circuit corresponds to the desired unitary U. To resolve the apparent contradiction with #P-hardness, note that their result only shows that a classical sampler from the boson sampling distribution could be used to compute the permanent within the polynomial hierarchy, not that a quantum sampler could.

In addition to these results on the exact form of the problem, Aaronson and Arkhipov proposed two additional (yet plausible) conjectures which together would imply that an approximate form of boson sampling could not be carried out efficiently on a classical computer. In approximate boson sampling, the task is to output a sample from an arbitrary probability distribution P within total variation distance ϵ of the real boson sampling distribution P_{BS}, for some small constant ϵ. This is a particularly important result for experimental implementations of boson sampling, as it allows for the inevitable imperfections in a real boson sampling device, up to a point [10, 11].

An important feature of the approximate boson sampling problem is that the sub-unitary matrices A_S must look like Gaussian random matrices. In order for this to be true, it must hold that $m \gg n$ (the currently best proven bound is $m = \mathcal{O}(n^5 \log^2 n)$ [1], although it is generally accepted that $m = \mathcal{O}(n^2)$ should be sufficient). A side effect of this condition is that, due to the bosonic birthday paradox [12], $\Pr(s_i > 1)$ is small, and hence that one can consider the problem restricted to the "collision free subspace" (CFS) of the overall Hilbert space of the bosonic states, where there is no bunching at the output ($s_i \leq 1$ for all i). So $\Pr(S) = |\mathrm{Per}\,(A_S)|^2$. Then our task becomes to sample from the restriction of the boson sampling distribution to this subspace, suitably renormalised by a constant close to 1. This is the problem on which we focus in this work.

In the setting of linear optics, an instance of the boson sampling problem for random U can be implemented by injecting the n-photon Fock state

$$|\Psi_{in}\rangle = |\underbrace{1, \ldots, 1}_{n}, \underbrace{0, \ldots, 0}_{m-n}\rangle \tag{3}$$

into a linear optical network of beam-splitters and phase-shifters with transfer matrix chosen Haar-randomly, and a single-photon detector coupled to each output mode. These detectors need not be able to discern the number of output photons in each mode, as one can restrict the problem to the CFS. A sample is obtained by recording the n-fold coincident detection signals at the output.

Schemes for implementing arbitrary unitary linear optical transfer matrices using beam-splitters and phase-shifters are well known [13, 14] and have been realised successfully [15, 16]. The current record for implementing the "classic" version of boson sampling is 5 photons in 9 modes [6].

For large n, producing the input state $|\Psi_{in}\rangle$ is a major experimental challenge. A "scattershot" variant of boson sampling has therefore been proposed which is easier to implement experimentally, and yet satisfies similar complexity-theoretic properties to classic boson sampling [17, 18]. This approach is useful when using photon pair sources such as spontaneous parametric down conversion (SPDC) sources, where each source has a small probability of generating a photon pair per pulse of a pump laser. In the scattershot setup, m sources are pumped concurrently, and some subset of these sources may generate a photon pair. One photon in each pair is sent to a detector in order to herald the presence of the other, which is sent to an input mode of the linear optical circuit (which may now be any of the m modes, rather than one of the first n modes).

In principle, and with perfectly efficient single photon detectors, this allows one to select the instances where $2n$ detection events occur (n herald signals and n post-circuit signals) and generate samples from a boson sampling distribution with a known input configuration. The benefit of this approach is that, as the number of pairs generated is distributed binomially, one can tune the probability that each source generates a photon pair so as to maximise the probability of an n photon-pair outcome, and this is much greater than if one were limited to n sources. Scattershot boson sampling has been implemented with 6 sources and 13 modes [4].

From the point of view of classical simulation, the scattershot boson sampling problem is not significantly harder than the classic boson sampling problem. The only difference is that, rather than sampling from a distribution defined by a fixed $m \times n$ submatrix A of an $n \times n$ unitary matrix U, one first chooses A itself at random, according to a probability distribution which is easy to sample from classically. We therefore focus on classic boson sampling here, but stress that all our results also apply to scattershot boson sampling.

3. CLASSICAL SAMPLING TECHNIQUES FOR BOSON SAMPLING

We now describe some techniques that can be used to implement boson sampling on a classical computer. All the classical algorithms we consider will ultimately be based on computing permanents, and our goal will be to minimise the number of permanents computed per sample obtained. The asymptotically fastest known algorithm for exactly computing the permanent of an $n \times n$ matrix is Ryser's algorithm [19], whose runtime when implemented efficiently is $\mathcal{O}(n2^n)$. The Balasubramanian–Bax/Franklin–Glynn algorithm achieves the same asymptotic performance but may be preferred from the perspective of numerical stability [20]. For the values of n we consider here, numerical stability is unlikely to be a significant issue [20].

Although this exponential runtime may seem discouraging, note that this is substantially faster than the naïve approach of evaluating the sum (2) directly, which would take time $\mathcal{O}(n!)$. On a standard personal computer, the permanent of an $n \times n$ matrix can be computed in under 1 second for $n \approx 25$ (see [5, 20] and Sec. 4 4.1 below for more detailed numerical experiments), and it is predicted in [20] that a supercomputer could solve the case $n = 50$ in under 2 hours. In addition, some proposed verification techniques for boson sampling experiments (e.g. likelihood-ratio tests) rely themselves on computing permanents [1, 21].

3.1. Brute force exact sampling

Perhaps the most obvious way to tackle the boson sampling problem on a classical computer is to compute all of the probabilities in the CFS, and sample from the probability mass function associated with these probabilities [5]. It becomes apparent that this scheme is immensely computationally demanding when one considers how the dimension of the CFS scales with the size of the problem. Indeed, this approach requires computing $\binom{m}{n}$ permanents of $n \times n$ complex valued matrices before a single sample can be output. Using the lower bound $\binom{m}{n} \geq (m/n)^n$, for $m \geq n^2$ at least n^n permanents must be computed; and even for $n = 10$ we need to compute more than 17×10^{12} permanents. Thus our computation quickly becomes swamped by the number of permanents to compute, rather than the complexity of computing the permanent itself.

We therefore seek an approach which relies on computing a number of permanents which scales more favourably with n and m than $\binom{m}{n}$.

3.2. Rejection sampling with a uniform proposal

Rejection sampling is a general approach for exactly sampling from a desired distribution P with probability mass function $f(x)$, given the ability to sample from a distribution Q with probability mass function $g(x)$, where $f(x) \leq \lambda g(x)$ for some λ and all x. The algorithm proceeds as follows:

1. Generate a sample x from Q.

2. With probability $\frac{f(x)}{\lambda g(x)}$, output x. Otherwise, go to step 1.

It is easy to show that the sample eventually output by the rejection sampling algorithm is distributed precisely according to P. The probability that an accepted sample is generated is

$$\sum_x g(x) \frac{f(x)}{\lambda g(x)} = \frac{1}{\lambda},$$

so the expected number of samples from Q required to output a sample from P is just λ. The simplest case in which we can apply the algorithm is where Q is the uniform distribution on N elements, and we have the upper bound $f(x) \leq \mu$ for some μ and all x. Then the expected number of uniform samples required to obtain a sample from P is μN, which will be minimised when $\mu = \max_x f(x)$.

Here we take P to be the boson sampling distribution restricted to the CFS, and Q to be the uniform distribution on the CFS (so $N = \binom{m}{n}$). Note that P is subnormalised, so is not quite a probability distribution. However, the rejection sampling algorithm is blind to this subnormalisation (as this is effectively the same as increasing λ), so will generate samples from the renormalised distribution.

Each iteration of rejection sampling requires the computation of one permanent, corresponding to $f(x)$. However, we seem to have a problem: in order to use rejection sampling most efficiently, it is required to know the maximum value of $f(x)$, which corresponds to the largest permanent of all $n \times n$ submatrices A of a given $m \times m$ unitary matrix

U. One might expect that computing this would be computationally hard. Without any bound on this quantity, we would be forced to use the trivial bound $\mu = 1$, corresponding to $\binom{m}{n}$ permanent computations being required to obtain one sample from P.

Fortunately, as we are only attempting to perform *approximate* boson sampling, we only require a good estimate of $\mu = \max_x f(x)$. Recall that it was argued in [1] that sampling from a distribution within total variation distance ϵ of the real boson sampling distribution P_{BS} should be computationally difficult, for some small constant ϵ. Imagine that our guess $\widetilde{\mu}$ for μ is too small, such that $\sum_{x, f(x) > \widetilde{\mu}} f(x) = \epsilon > 0$. Then if x is sampled uniformly at random and $f(x) > \widetilde{\mu}$, step 2 of the rejection sampling algorithm will fail. If we modify the rejection sampling algorithm to simply produce a new uniform sample in this case and repeat, it is easy to see that we can view the modified algorithm as sampling from the truncated distribution P_{low} with probability mass function

$$\widetilde{f}(x) = \begin{cases} \frac{f(x)}{\sum_{x, f(x) \le \widetilde{\mu}} f(x)} & \text{if } f(x) \le \widetilde{\mu} \\ 0 & \text{if } f(x) > \widetilde{\mu} \end{cases}.$$

Then the total variation distance between P_{low} and P is

$$\frac{1}{2} \sum_x |f(x) - \widetilde{f}(x)| = \sum_{x, f(x) > \widetilde{\mu}} f(x).$$

So if the probability mass of P above μ is at most ϵ, we have sampled from a distribution within distance ϵ of P. We have found that we are able to use a simple random restart hill climbing algorithm to provide a suitable estimate of μ with $\mathcal{O}(m^2 n)$ computations of $n \times n$ matrix permanents.

Our random restart hill climbing algorithm works as follows. We start by randomly sampling a submatrix, represented by the tuple S, from the uniform distribution. For one pass, we greedily try replacing each row in the sampled submatrix by each row from A in turn, accepting only if this increases $\Pr(S)$ while also making sure to avoid selecting the same row twice. We perform repeat passes until there is no improvement of the probability over a complete pass. At this point we randomly resample a new starting submatrix and repeat from the beginning. The total number of permanent calculations for one pass is $n(m - n)$.

This method is not guaranteed to find a global maximum. However, in our experiments for $n \le 7$ where we can still compute the full probability mass function exactly, we found the estimates for the maximum probability to be exactly equal to the global maximum in the overwhelming majority of cases. In the range $8 \le n \le 12$ where we no longer are able to compute the exact maximum probability, the bounds from our hill-climbing algorithm also allowed us to sample using rejection sampling efficiently and then compare our results with our Metropolised independence sampler (qv). This provided further evidence for both sampling techniques.

3.3. Metropolised independence sampling

Markov Chain Monte Carlo (MCMC) methods have been long been the standard tool within statistics for the task of sampling from complicated probability distributions. We employ for our problem of boson sampling a special case of the MCMC procedure known as Metropolised independence sampling [22, 23] (MIS), which enables us to sample efficiently from the target distribution.

The general principle of MCMC methods is to construct a Markov chain which has the target distribution (in our case, the boson sampling distribution of tuples) as its stationary distribution. The overall MCMC method requires us only to specify a proposal probability distribution, in our case over tuples $(s_1, \ldots, s_m)$. At each turn we sample from this probability distribution and accept the new state according to the prespecified rules of the MCMC method. In the general case the proposal distribution can depend on the current state of the chain but in MIS these proposals are entirely independent. By choosing a proposal distribution which is not too far from our target boson sampling distribution, this simplification turns out to be particularly useful for us. Empirically we show that it not only ensures fast convergence but a simple and efficient thinning procedure almost completely eliminates dependence between successive samples.

Let Y be some proposal distribution over the tuples in $\Phi_{m,n}$ with probability mass function $g(x)$. As before, let P be an instance of the CFS restricted (subnormalised) boson sampling distribution over these tuples, with probability mass function $f(x)$. Starting at some random (according to Y) tuple x, propose a new random tuple x'. We accept the proposal and transition from x to x' with probability

$$T(x'|x) = \min\left(1, \frac{f(x')}{f(x)} \frac{g(x)}{g(x')}\right) \tag{4}$$

where we note that any normalisation factor multiplying probabilities would cancel at this stage. Repeating this generates a Markov chain with stationary distribution equal to the target distribution, P. If one can argue that the chain has converged then each sample from the generated chain after this point will be from P. Each state in the Markov chain requires generating a single tuple sample from Y, one evaluation of $f(x')$ and one evaluation of $g(x')$, which entails one computation of an $n \times n$ complex valued matrix permanent.

It should be noted that it takes some time for the Markov chain to converge to its stationary distribution, and so some number of tuples at the beginning of the chain must be discarded (referred to as "burn in") in order for samples to come from the target distribution. In addition to this, for example as there is a probability that the state of the Markov chain stays the same between iterations, there exists some autocorrelation in the chain. In order to generate independent samples from the Markov chain, one usually throws away a number of states (referred to as "thinning") at least equal to the autocorrelation period of the chain before harvesting a sample state. Both burn period τ_{burn} and thinning interval τ_{thin} depend both on the target distribution and the choice of proposal distribution Y. In general, the greater the overlap between target and proposal distributions (as measured by the acceptance probability (4)), the smaller τ_{burn} and τ_{thin} need be.

A candidate for the proposal distribution is, as above in the case of rejection sampling, the uniform distribution over the tuples. However, we have found that a more suitable proposal distribution is the distribution P_{D} of distinguishable particles for a given input U, as it has a greater overlap with the target distribution. Indeed, our experiments have found that for $n = 20$, the acceptance rate when using P_{D} as the proposal distribution is roughly 40%.

This distribution is tantalisingly similar to the usual boson sampling distribution (1):

$$\Pr_{P_{\text{D}}}(S) = \frac{\text{Per}\left(|A_S|^2\right)}{s_1! \ldots s_n!}, \tag{5}$$

where for a complex matrix A with elements A_{ij}, $|A|^2$ denotes the matrix with elements $|A_{ij}|^2$. However, there is a classical algorithm which can sample from this distribution in time $O(mn)$ [21].

Using P_{D} as the proposal distribution, one evaluation of $f(x')$ requires computing one $n \times n$ *real* valued matrix permanent. So, in order to generate the first tuple in a sample, τ_{burn} $n \times n$ real valued matrix permanents and τ_{burn} $n \times n$ complex valued matrix permanents must be computed. Each subsequent sample requires τ_{thin} $n \times n$ real valued and complex valued matrix permanents to be computed.

Testing (see Section 4 4.4) has suggested that, using the distinguishable particle proposal distribution, it is sufficient for τ_{burn} and τ_{thin} to grow very slowly with n, and for neither to grow above 100 for $n \leq 25$.

Adapting our MIS-based method to carry out the scattershot boson sampling problem is simple. For each sample that we wish to output, we can first sample (efficiently) from the uniform distribution on n-fold input modes (which fixes the columns of U contributing to submatrices), before running the algorithm in the way described above for a single sample. In this case, τ_{thin} becomes meaningless and we are only interested in τ_{burn}, as we start a new chain for each sample.

Alternatively, our proposal distribution can be changed to include the uniform distribution over n-fold input modes, meaning that each state in a given Markov chain can correspond to a different input configuration.

Although in this work we do not examine the more general situation of boson sampling where there can be more than one boson in an output mode, we anticipate that relaxing the CFS restriction will not increase the run time of the MIS method. In fact, it is possible that the average run time could be decreased with this relaxation, as there exists an algorithm for computing the permanent which is exponential in matrix rank, rather than matrix size [24]. However, due to there usually existing very large permanents of sub-unitary matrices with many repeated rows, relaxing the CFS restraint has an adverse effect on the average run time of our rejection sampling method.

4. NUMERICAL RESULTS

For concreteness and convenience, we will examine the performance of our samplers for the case where the number of modes scales exactly as the square of the number of photons, i.e. $m = n^2$. This regime is generally expected to be hard to simulate classically [1, 5].

4.1. Computing matrix permanents

In order to make scaling based arguments for the expected run time of our sampling methods, we need to both confirm that we are able to compute permanents of $n \times n$ matrices in $\mathcal{O}(n2^n)$ time and compute the associated constant of proportionality.
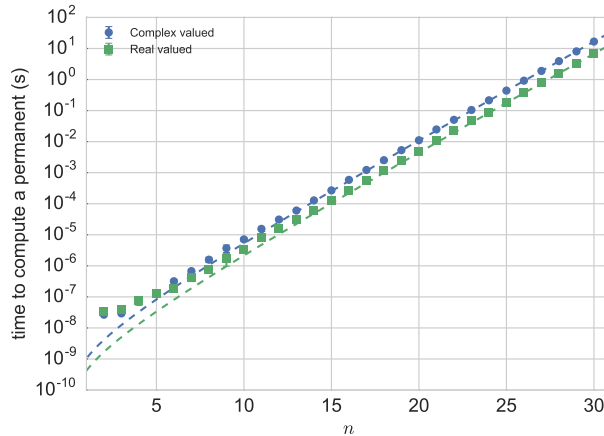
Supplementary Fig. 1. Numerical results for the mean time to compute $n \times n$ matrix permanents on a laptop with a 2.30GHz Intel(R) Core(TM) i5-5300U CPU. Blue points correspond to computing permanents of complex valued matrices and green squares to real valued matrices. Values correspond to the mean time recorded over 20 random matrices, with 12000000, 120000, 12000, 1200 and 12 repeats for values of n in ranges $2 \leq n \leq 3$, $4 \leq n \leq 7$, $8 \leq n \leq 15$, $16 \leq n \leq 20$ and $21 \leq n \leq 30$ respectively. Dashed lines correspond to a fit to the function $t(n) = cn2^n$, with c calculated as 5.180×10^{-10} (blue) and 2.106×10^{-10} (green). Error bars displayed (where visible) represent the standard error in the mean time.

Timing results are presented in fig. 1, where we see the expected scaling for large n. We attribute the difference between the fitting function $t(n) = cn2^n$ and the data points for small n to extra operations in our permanent computation code (checks, memory allocation etc.), on top of an implementation of Ryser's algorithm. From the fit we were able to calculate c for the case of computing permanents of complex and real valued matrices on a personal computer[1] finding that $c = 5.180 \times 10^{-10}$ and $c = 2.106 \times 10^{-10}$ respectively. In the case of computing permanents of real valued matrices we note that all values are in fact nonnegative, and therefore it may be beneficial to use the fully-polynomial randomized approximation scheme of Jerrum, Sinclair and Vigoda [25].

Importantly, these values (and approximate values for more powerful machines) allow us to predict the average run time for our sampling methods for any value of n.

### 4.2. Rejection sampling and random restart hill climbing

In addition to how long it takes to compute a single matrix permanent, we are also interested in how many matrix permanent evaluations are required before a sample value is output.

In the case of rejection sampling with a uniform proposal distribution, we must compute the upper bound μ using our random restart hill climbing algorithm before implementing the rejection sampling algorithm. We emphasise here that μ need only be computed once per input U, and so for large enough sample sizes, the computational contribution of computing μ gets washed out. For concreteness, we focus here on outputting a sample consisting of 100 tuples, and present the number of required permanent computations averaged over these 100 tuples.

In fig. 2 we see that, by using rejection sampling instead of brute force sampling, one can reduce the number of computations required at a rate which increases rapidly with n. In fact, a significant portion of the computation for all values of n investigated here is spent getting an estimate of μ. It should be noted, however, that this is an artefact of our choice of number of restarts implemented in our random restart hill climbing algorithm. We chose the number of restarts to be $4m$, based on empirical evidence that this results in the probability mass above this estimate being small. For 600 random instances of the problem with $2 \leq n \leq 7$, we found that the largest submatrix permanent was found 598 times. When the largest value was not found, the probability mass above the estimate was found to be small ($< 2 \times 10^{-5}$).

As an example for reference, on a personal computer the average time to produce one sample for $n = 12$ (averaged over 100 samples) was 2.03s.

---

[1] Specifically, numerical experiments in this paper were performed on a Dell Latitude E5450 laptop with 2.30GHz Intel Core i5-5300U CPU

(a)

(b)
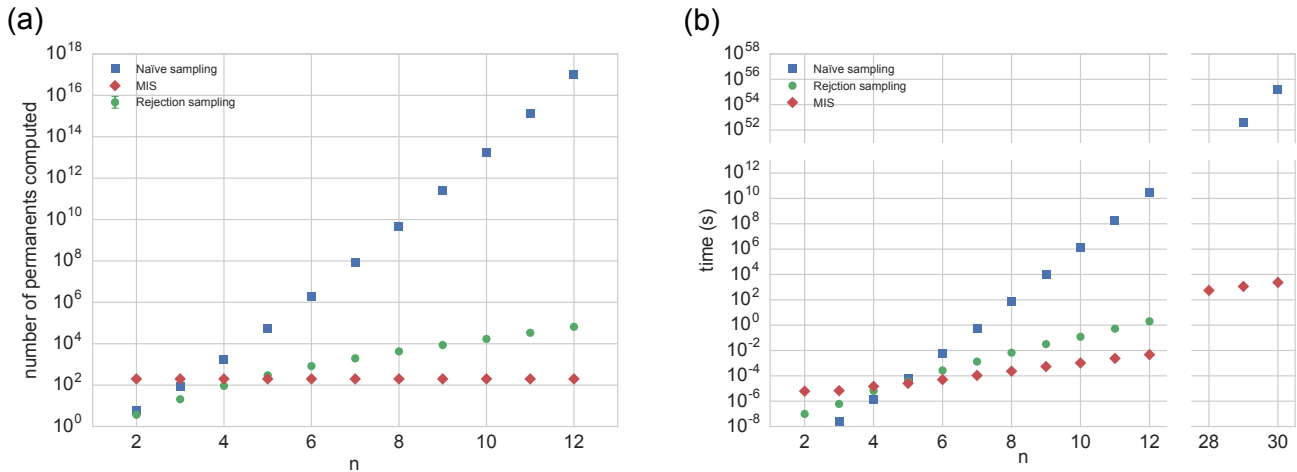
Supplementary Fig. 2. (a): Numerical results for the mean number of $n \times n$ permanent computations required to output a single tuple, averaged over a sample of 100 tuples, with $m = n^2$. Red diamonds correspond to MIS with $\tau_{\text{burn}} = \tau_{\text{thin}} = 100$, green circles to rejection sampling with μ (an approximation of $\max_S |\text{Per}(A_S)|^2$) computed using our random restart hill climbing algorithm with $4m$ restarts, and blue squares correspond to the number of permanent computations for brute force sampling (i.e. simply $\binom{n^2}{n}/100$). Each rejection sampling point is calculated by finding the mean number of permanents computed before accepting a proposed tuple over 10000 accepted tuples (100 repeats for 100 Haar-random input unitaries U), including those computed in performing the random restart hill climbing algorithm. Error bars, where visible, represent one standard error in the mean number. (b): These values converted to mean time (in seconds) to output a single tuple using timing data from fig. 1, and extended to $n = 30$ photons.

4.3. Metropolised independence sampling

Determining the theoretical runtime of MIS is straightforward: τ_{thin} and τ_{burn}, together with the time to compute a matrix permanent, completely determine the time to produce a sample, as discussed above. We include these bounds in fig. 2, taking $\tau_{\text{thin}} = \tau_{\text{burn}} = 100$, and amortised over 100 samples. Observe that for $n = 20$, each tuple can be output in 1.58s on a personal computer; even for $n = 25$, only 62.5s are required.

4.4. Assessment of our sampling methods

How can we confirm that our samplers are sampling from an approximately correct distribution? Boson sampling is believed to be computationally hard to verify [1, 21, 26], so, to date, experimental implementations of the problem have relied upon circumstantial evidence to support the claim that they are, in fact, sampling from the target distribution. This, coupled with the fact that reliability of MCMC methods is notoriously difficult to certify, requires us to proceed with caution.

However, the fact that we have 3 completely different samplers, each of which we can ask to provide us with a sample for the same instance of the problem, allows us to gather mutually supportive evidence for them all sampling from the correct target distribution; if we cannot reject that all pairs of samples come from the same underlying distribution, we can have some confidence that they are *all* in fact sampling from the target distribution.

A natural quantity to use to compare the distributions on matrices that we consider is $|\text{Per}(A_S)|^2$ of the sampled matrix A_S, which is equal to the probability of the sampled matrix in the boson sampling distribution; for convenience we actually use $\log |\text{Per}(A_S)|^2$.

We generated samples of size 20000 of tuples (assigning each of them an index of an integer value between 0 and $\binom{m}{n} - 1$), together with their corresponding $-\log(|\text{Per}(A_S)|^2)$, using various samplers and for problems of size $n = 3, 7, 12, 20$. The data are plotted in fig. 3.

We expect that the distribution of $-\log(|\text{Per}(A_S)|^2)$ differs between samples of indistinguishable bosons and distinguishable particles. This feature manifests itself clearly for each attempt and each method of attempted boson sampling; the distinguishable particle values (yellow) are distributed visibly differently to those of other sampling methods (all other colours). More rigorously, we performed 2-sample bootstrap Kolmogorov-Smirnov (KS) tests [27, 28] comparing the distribution of $-\log(|\text{Per}(A_S)|^2)$ values for tuples obtained by sampling from the distinguishable particle distribution with those obtained using our classical boson sampling methods. The result was that we could
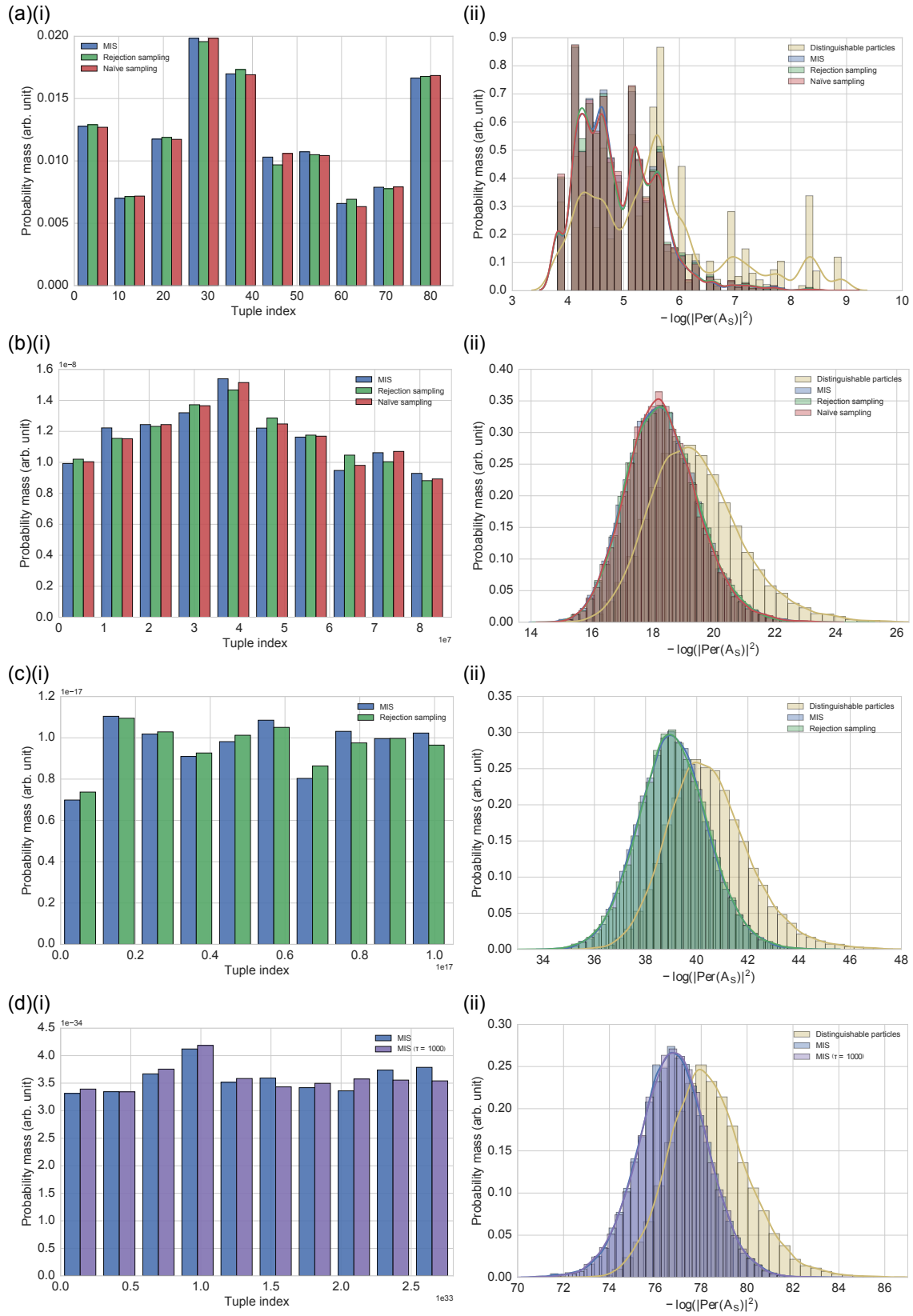
Supplementary Fig. 3. (a): (i): Distribution of 20000 indexed and binned tuples obtained using MIS (blue), rejection sampling (green) and naïve brute-force sampling (red) for a Haar-random instance of the boson sampling problem with $n = 3$ and $m = 9$. (ii): Distribution of $-\log(|\mathrm{Per}(A_S)|^2)$ for each tuple S sampled, along with the distribution of these values computed for a sample from the distinguishable particle distribution (yellow). Solid lines are obtained via kernel density estimation, and are included as a visual aid. (b): Distributions for a Haar-random instance of the boson sampling problem with $n = 12$ and $m = 144$ (colours as in A). (c): $n = 12$ and $m = 144$. Note that no sampling was done using the naïve brute-force sampling method, due to its inefficiency at this problem size. (d): $n = 20$ and $m = 400$ using MIS with $\tau = 100$ (blue) and $\tau = 1000$ (purple).

(a)

(b)

(c)

(d)
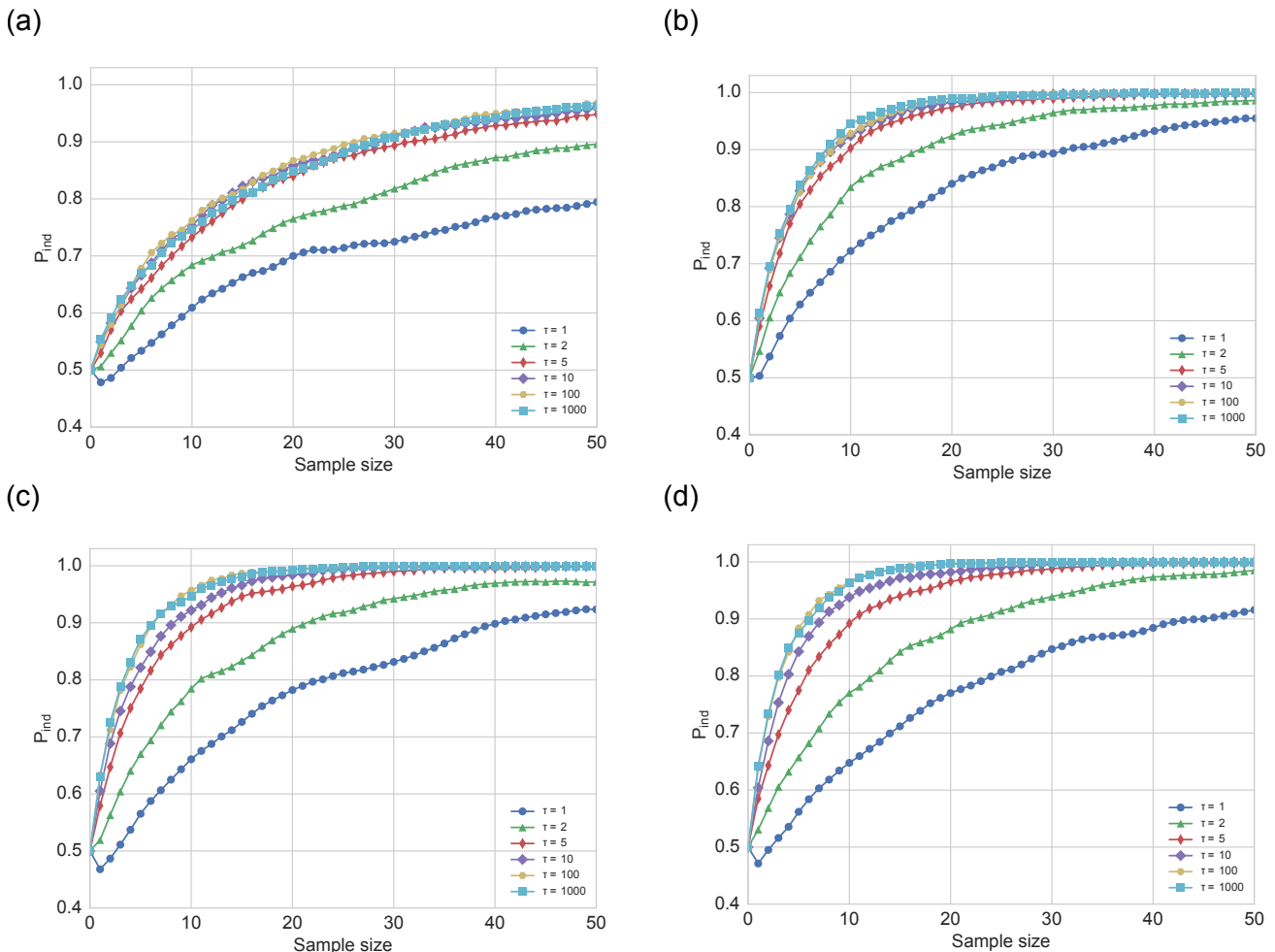
Supplementary Fig. 4. Results of likelihood ratio tests performed on samples generated using MIS with different values of τ. Here τ is shorthand for both τ_{burn} and τ_{thin}. All data points are the mean value of P_{ind} over 1000 independently generated samples. Confidence intervals have been omitted for ease of viewing. (a): Data for a Haar-random instance of the boson sampling problem with $n = 3$ and $m = 9$. (b): Similarly for $n = 7$ and $m = 49$. (c): $n = 12$ and $m = 144$. (d): $n = 20$ and $m = 400$.

reject the null hypothesis that this feature of the samples was distributed identically at the 0.1% significance level (i.e. $p < 0.001$ for all samplers and all problem sizes considered). This result was found to be consistent over 5 repetitions of taking a sample for each input unitary.

Not only do we see that our sampling methods give significantly differently distributed $-\log(|\mathrm{Per}(A_S)|^2)$ to the distinguishable particle sampler, we see *no* significant difference between these distributions for each of our sampling methods. Applying the same 2-sample bootstrap KS tests to compare these distributions returned that we cannot reject the null hypothesis that they are distributed identically at any reasonable significance level. We see this as evidence that all of our samplers are sampling from the target boson sampling distribution.

Due to some sampling methods being more efficient than others, we witness problem sizes beyond which certain methods become unreasonably slow. As a rough guide, on a personal computer these problem sizes are $n = 7$ for brute force sampling and $n = 12$ for rejection sampling. Thus, in the case of $n = 20$, we compare one MIS sampler (with $\tau_{\mathrm{burn}} = \tau_{\mathrm{thin}} = 100$) with another MIS sampler (with $\tau_{\mathrm{burn}} = \tau_{\mathrm{thin}} = 1000$).

In table I we present the p-values obtained when performing 2-sample bootstrap Kolmogorov-Smirnov tests on the values $-\log(|\mathrm{Per}(A_S)|^2)$ for sets of 20000 tuples obtained with various samplers. Independent samples were generated 5 times in order to repeat the tests. The null hypothesis in this test is that the values are identically distributed. As p-values are uniformly distributed when the null hypothesis is true, we see no significance in any relatively small values in table I; they are to be expected when quoting a large number of p-values.

Note that the assessment of MCMC algorithms is non-trivial: in particular, slow convergence to the target distribution and autocorrelation within the chain can result in an erroneous sample being output. We therefore also

(a)

(b)

(c)

(d)

(e)
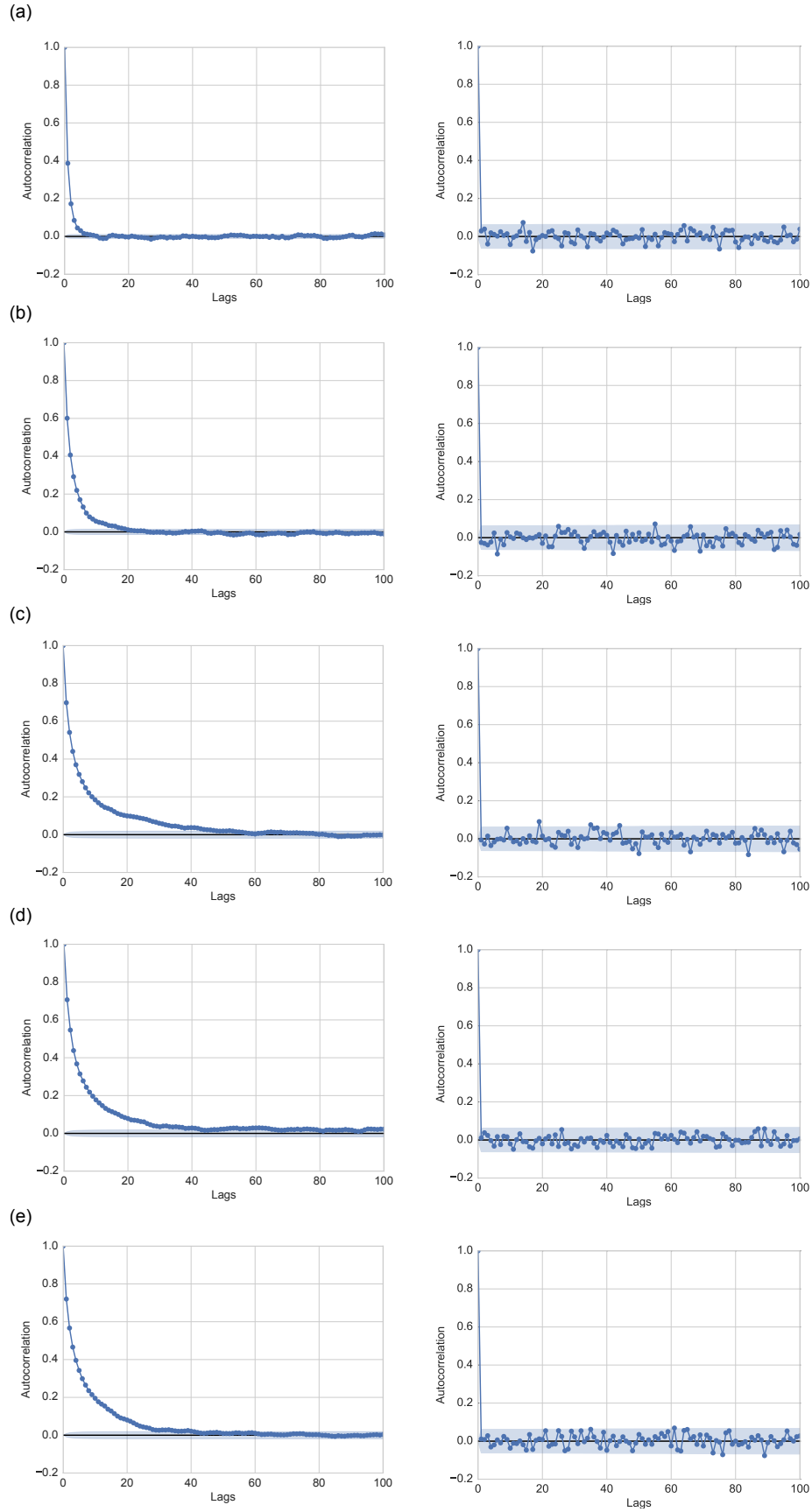
Supplementary Fig. 5. Examples of the autocorrelation of $-\log(|\mathrm{Per}(A_S)|^2)$ present at lags up to 100 in Markov chains before (left) and after (right) applying our burn in and thinning procedures. The un-thinned sample size was 50000 (reduced from 100100 by taking the first 50000 tuples), and the thinned sample size was 1000. Shaded regions represent a 95% confidence interval. (a): Data for a Haar-random instance of the boson sampling problem with $n = 3$ and $m = 9$. (b): Similarly for $n = 7$ and $m = 49$. (c): $n = 12$ and $m = 144$. (d): $n = 20$ and $m = 400$. (e): $n = 25$ and $m = 625$.

n	Sampler 1	Sampler 2	2-sample KS test p-values				
3	MIS	naïve brute-force	0.651	0.34	0.653	0.368	0.089
	MIS	rejection	0.209	0.415	0.679	0.053	0.167
	rejection	naïve brute-force	0.03	0.588	0.39	0.277	0.331
7	MIS	naïve brute-force	0.448	0.606	0.858	0.734	0.468
	MIS	rejection	0.628	0.615	0.246	0.581	0.775
	rejection	naïve brute-force	0.27	0.252	0.287	0.954	0.263
12	MIS	rejection	0.628	0.919	0.998	0.77	0.519
20	MIS ($\tau = 100$)	MIS ($\tau = 1000$)	0.548	0.258	0.832	0.483	0.585

TABLE I. p-value data for various problem sizes and sampler combinations. Here, τ is shorthand for both τ_{burn} and τ_{thin}.

performed a modified version of the likelihood ratio test described by Bentivegna et al. [29]. Following the notation of [29], the test works as follows. We define two hypotheses: $\mathcal{Q}$, the indistinguishable boson hypothesis, and $\mathcal{R}$, an alternative hypothesis. In our case, the alternative hypothesis $\mathcal{R}$ is the distinguishable particle hypothesis. Let q_x be the probability of seeing the sampled event x according to hypothesis $\mathcal{Q}$, and r_x be the corresponding probability under hypothesis $\mathcal{R}$. Assigning equal priors to each hypothesis, Bayes' theorem tells us that

$$\frac{P\left(\mathcal{Q}|N_{\text{events}}\right)}{P\left(\mathcal{R}|N_{\text{events}}\right)} = \prod_{x=1}^{N_{\text{events}}} \left(\frac{q_x}{r_x}\right) = \mathcal{X}, \tag{6}$$

where, for example, $P\left(\mathcal{Q}|N_{\text{events}}\right)$ is the probability of hypothesis $\mathcal{Q}$ being correct given that a sample of size N_{events} with events $\{k_x\}$ was obtained. A large value of $\mathcal{X}$ would correspond with us having a high degree of confidence that hypothesis $\mathcal{Q}$ provides a better description of the observed data than hypothesis $\mathcal{R}$.

It is convenient to rewrite equation (6) in terms of the probability assigned to hypothesis $\mathcal{Q}$,

$$P_{\text{ind}} \equiv P\left(\mathcal{Q}|N_{\text{events}}\right) = \frac{1}{\mathcal{X}+1} \prod_{x=1}^{N_{\text{events}}} \left(\frac{q_x}{r_x}\right) \tag{7}$$

which is now normalised such that $P\left(\mathcal{Q}|N_{\text{events}}\right) + P\left(\mathcal{R}|N_{\text{events}}\right) = 1$.

One, perhaps subtle, point here is that q_x and r_x do not simply correspond to $|\text{Per}\left(A_x\right)|^2$ and $\text{Per}\left(|A_x|^2\right)$ respectively. The reason for this is that we are restricted to the CFS, and the probability of a sample being collision-free differs between indistinguishable bosons and distinguishable particles. So q_x and r_x must be normalised independently, such that they independently sum to 1 over all events x. Unfortunately, doing this exactly would require summing all probabilities in the CFS for the specific instance of the problem being considered – a computationally daunting task, as previously discussed. To approximate the normalisation of q_x efficiently, here we instead average the probability of the output being collision-free over the Haar measure, as reported in [1, 12]:

$$P_{\text{CFS}} \approx \binom{m}{n} \Big/ \binom{m+n-1}{n}. \tag{8}$$

For the hypothesis $\mathcal{R}$, we can efficiently sample output tuples, and so the ratio of collision-free tuples to tuples with collisions in a large sample provides an approximate normalisation for r_x.

For each problem size, we use this likelihood ratio test to assess the performance of MIS samplers with different τ_{burn} and τ_{thin}. The reasoning is that, as the proposal distribution for the sampler is the distinguishable particle distribution, we might expect that if the chain has not converged to the target distribution, this will manifest itself as samples looking more like they are from the distinguishable particle distribution than they should. Also, we expect that samplers with larger τ_{burn} and τ_{thin} are more likely to sample from the target distribution. Because of this, we expect to be able to observe a point at which increasing τ_{burn} and τ_{thin} has, on average, no effect on the outcome of a likelihood ratio test between hypotheses $\mathcal{Q}$ and $\mathcal{R}$.

The data corresponding to this for $n = 3, 7, 12, 20$ are presented in fig. 4. We can see that, indeed, for a given problem size, increasing τ_{burn} and τ_{thin} beyond a point does not improve the likelihood ratio test results. Not only this, but we observe that this point gradually increases in size with the problem size n. For example, with problem size $n = 3$, $\tau_{\text{burn}} = \tau_{\text{thin}} = 5$ provides results comparable with $\tau_{\text{burn}} = \tau_{\text{thin}} = 1000$. However, with problem size $n = 20$, $\tau_{\text{burn}} = \tau_{\text{thin}} = 5$ provides markedly worse samples on average than $\tau_{\text{burn}} = \tau_{\text{thin}} = 1000$. In fact, $\tau_{\text{burn}} = \tau_{\text{thin}} = 100$ seems to be the point at which increasing these values provides no benefit.

Additionally, in fig. 5 we see that autocorrelation between $-\log(|\text{Per}(A_S)|^2)$ for 100 lags in an un-thinned chain is negligible for random instances of the problem for $n = 3, 7, 12, 20, 25$. This is further supported by the autocorrelation being negligible at *all* lags for chains generated by applying our thinning procedure to the un-thinned chains.

Based on this, we claim that $\tau_{\text{burn}} = \tau_{\text{thin}} = 100$ is sufficient for MIS for problems up to size $n = 20$, and is likely to be sufficient for values of n greater than 20. The consequence of this is that a sample from the boson sampling distribution can be output, using a classical computer, in roughly the time that it takes to compute 100 $n \times n$ real valued matrix permanents and 100 $n \times n$ complex valued matrix permanents. For $n = 20$ and $m = 400$, this equates to a computational saving of a factor of $\sim 10^{31}$ when compared with brute force sampling.

4.5. Moving to larger n

Although MIS is efficient in the number of matrix permanent computations required, it still requires the inefficient computation of matrix permanents by Ryser's algorithm [19]. As such, gathering a 20000 tuple sample as was obtained in fig. 3 is challenging for $n > 20$ using readily available hardware. We were able to perform boson sampling using MIS for $n = 25$ and $m = 625$, with a sample size of 1000 tuples for plotting the distributions of indexed tuples and $-\log(|\text{Per}(A_S)|^2)$ of these sampled tuples, alongside a sample from the distinguishable particle distribution. We also performed a likelihood-ratio test to compute the probability of the indistinguishable boson hypothesis against the distinguishable particle hypothesis, averaged over 100 samples of 25 tuples for MIS with different values of τ.

In addition to this, we performed boson sampling using MIS for $n = 30$ and $m = 900$, obtaining a 250 tuple sample with $\tau = 100$, and compared the convergence of P_{ind} with other values of τ.

All of the data for $n > 20$ was taken on a cluster of 4 servers[2] at the University of Bristol, allowing for 122 MIS instances (and therefore 122 Markov chains) run in parallel (vs a maximum of 4 on our personal computer). The data are presented in fig. 6. For $n = 25, 30$, we notice no change in the performance of MIS beyond having to compute permanents of larger matrices.

5. COMPARISON WITH EXPERIMENTAL IMPLEMENTATIONS

Limiting ourselves to the standard boson sampling problem for now, we will consider how the time to obtain a single tuple sample scales with experimental parameters for a physical boson sampling device.

Parameters which have an effect on the rate at which samples can be generated experimentally are: the n-photon generation rate $R(n)$, the single photon transmission probability η, and the Haar-averaged probability of the n-photon state being in the CFS given by eqn. (8), which we take to be an approximation of the relative size of the CFS for any boson sampling input U.

It is important to notice here that in fact η depends on n. This is because the loss is composed of fixed loss $1 - \eta_{\text{f}}$ (source efficiency, detection loss, insertion loss etc.) and loss within the circuit, $1 - \eta_c$. The circuit depth – and therefore the amount of material (or number of components) each photon must propagate through – necessarily grows with the size of the problem, as non-trivial interference between all photons and across all modes must occur.

Circuits generally considered for boson sampling have depth d which scales at least linearly with m [13, 14] and hence at least quadratically with n. However, Aaronson and Arkhipov have shown that boson sampling can be implemented using a different circuit construction with depth $d = \mathcal{O}(n \log m)$ [1] if the input modes are fixed, and recently a circuit and predictions of performance with a linear mode scaling of $m = 4n$ have been reported [6].

This might suggest that $d = \mathcal{O}(n)$ could also be sufficient, although note that when $m = \mathcal{O}(n)$ the probability of a sample being collision-free is exponentially small in n (see eqn. (8)), and no theoretical argument for computational hardness is known.

Here we compare the rates with which we expect to obtain samples from a boson sampling experiment for the two regimes $d = n^2$, $d = 4n$ where experimental attempts have been documented [4, 6, 15, 30–37]. The transmission probability of each photon through the circuit is $\eta_c = \eta_0{}^d$, where η_0 is the probability of a photon surviving per 2-mode coupling length (or component) in the interferometer. Using these parameters, we can write the time in which we expect to obtain a single sample from a quantum device as a function of the problem size:

$$q_t(n) = (R(n) P_{\text{CFS}} \eta^n)^{-1} \tag{9}$$

---

[2] 2 dual-socket Intel Xeon CPU E5-2697A v4 @ 2.60GHz and 2 dual-socket Intel Xeon CPU E5-2680 v3 @ 2.50GHz
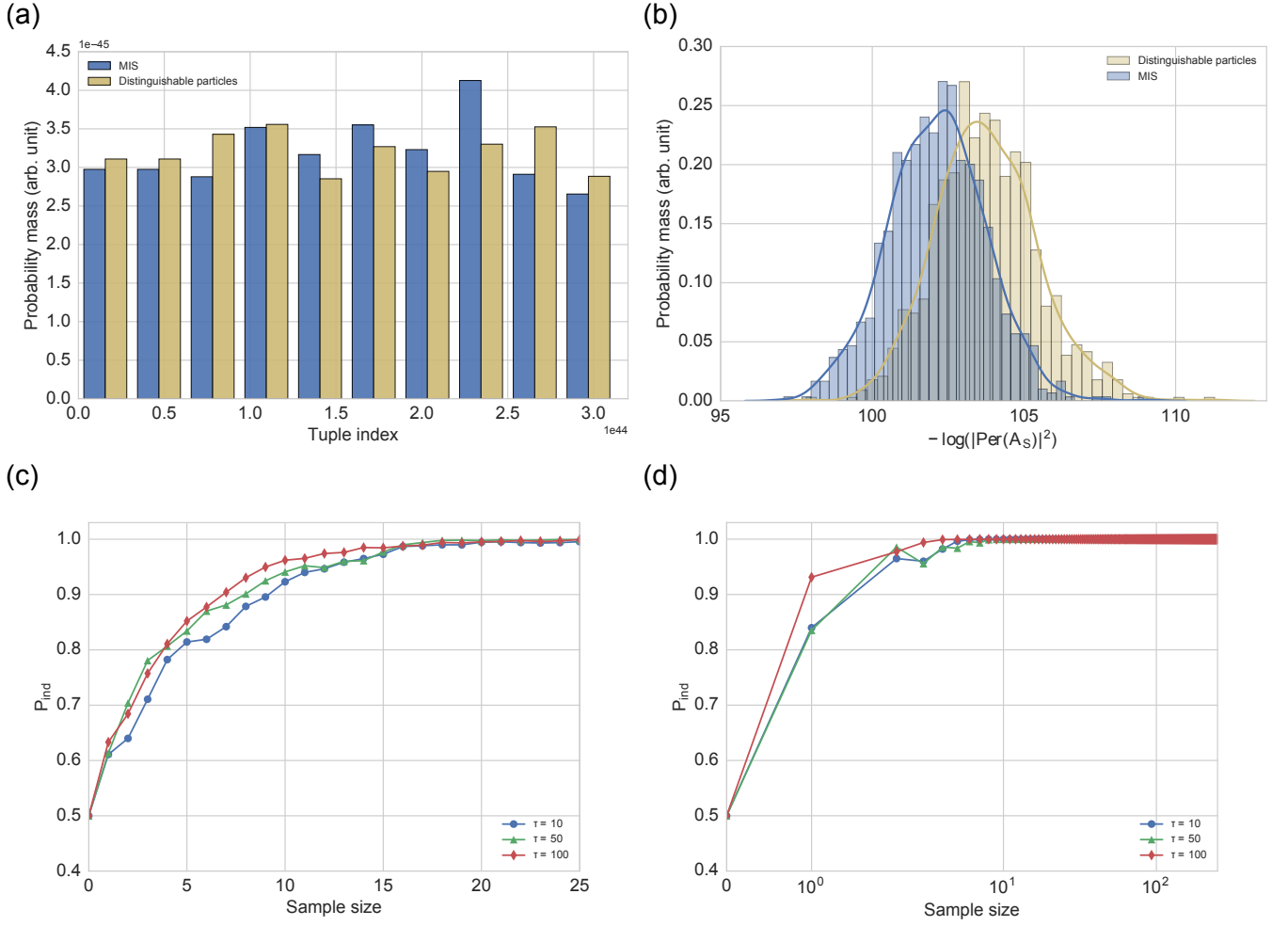
Supplementary Fig. 6. (a): Distribution of 1000 indexed and binned tuples obtained using MIS (blue) and a sampler sampling from the distinguishable particle distribution (yellow) for a Haar-random instance of the boson sampling problem with $n = 25$ and $m = 625$. (b): Distribution of $-\log(|\mathrm{Per}(A_S)|^2)$ for each tuple S sampled. Solid lines are obtained via kernel density estimation, and are included as a visual aid. (c): Results of likelihood ratio tests performed on samples generated using MIS for a Haar-random instance of the boson sampling problem with $n = 25$ and $m = 625$. (d): Results of a likelihood ratio test performed on a sample generated using MIS for a Haar-random instance of boson sampling with $n = 30$ and $m = 900$.

where $\eta = \eta_{\mathrm{f}} \eta_0^d$ Expanding η for linear and quadratic mode scaling respectively, we see that

$$q_t^{\mathrm{lin}}(n) = \left(R(n) \eta_{\mathrm{f}}{}^n \eta_0{}^{4n^2} P_{\mathrm{CFS}} \right)^{-1}$$

$$\sim \left(R(n) \left(\frac{4\eta_{\mathrm{f}}}{5} \right)^n \eta_0{}^{4n^2} \right)^{-1}$$

$$q_t^{\mathrm{quad}}(n) = \left(R(n) \eta_{\mathrm{f}}{}^n \eta_0{}^{n^3} P_{\mathrm{CFS}} \right)^{-1}$$

$$\sim \left(R(n) \eta_{\mathrm{f}}{}^n \left(\frac{1}{e} \right) \eta_0{}^{n^3} \right)^{-1}$$

where it should be emphasised that $\eta_{\mathrm{f}}, \eta_0 < 1$.

Using these bounds, we see that the time required to obtain a sample scales exponentially in either n^2 or n^3. However, classical computation of the permanent using Ryser's algorithm scales only exponentially in n. Therefore, for very large n and realistic loss parameters, the runtime scaling of a classical sampling algorithm would be favourable compared with the boson sampling experiment, even if the algorithm needed to compute exponentially many permanents to obtain one sample.

In order to provide a meaningful comparison of our classical boson sampling methods to an experimental implementation, we focus on the current leading experimental approach by Wang et al. [6]. This approach involves the use of a quantum dot based photon source, and has resulted in increased photon numbers and significantly higher rates compared to the current best SPDC based demonstrations (see Extended Data Table 1 in [6]). In fig. 7 (also see fig.2d in the main text), we consider the rate at which a similar experiment with projected near-future parameters [6] can generate sample values with both $m = 4n$ and $m = n^2$, and compare this to the average sample value generation rate for a personal computer running the MIS algorithm for boson sampling.

Specifically, in [6] Wang et al. report a quantum dot based photon source with $R(n) = 76n^{-1}$MHz. Also reported are loss parameter values predicted to be achievable in the near future:

$$\eta_{\mathrm{f}} = \eta_{\mathrm{QD}}\eta_{\mathrm{de}}\eta_{\mathrm{det}},$$

where η_{QD} is the single photon source end-user brightness ($= 0.74$), η_{de} is the demultiplexing efficiency for each channel ($= 0.845$) and η_{det} is the efficiency of each detector ($= 0.95$). The authors of [6] report the efficiency of their 9 mode interferometer to be $\sim 99\%$, and so we approximate the efficiency of a similar interferometer scaled to m modes as $\eta_{\mathrm{c}} = 0.99^{\frac{m}{9}}$ and hence $\eta_0 = 0.99^{1/9}$ as we are assuming $d = m$.

We see that, although the personal computer is outperformed for a small problem size, beyond $n \approx 14$ the personal computer outperforms the experimental boson sampler. Clearly, the point at which the classical device becomes dominant would increase if the experiment had access to an n-photon state generator with greater repetition rate. On the other hand, if the classical device were more powerful (i.e. could perform more floating point operations per second), it would start to become dominant at even smaller problem sizes. More important is that the rate falls off noticeably more rapidly for the experimental sampler with these loss parameters, for both regimes of scaling the number of modes with the number of photons.

To emphasise this point, we perform the following analysis. Assuming that our MIS sampler continues to perform equally well for larger instance sizes, we can compare its runtime with current and future experiments. The classical runtime for an instance of size n bosons in $m = n^2$ modes can be estimated as

$$c_t(n) = 3 \times 10^{-13} n^2 2^n$$

where $3 \times 10^{-15}s$ is the time scaling for computing one real and one complex permanent recently reported for the supercomputer *Tianhe 2* [20]. We note here that we do not assume that we can use the efficient (i.e Gray code ordered) version of Ryser's algorithm here, as we would need to parallelise the computation of the permanent to deal with large n problems. We define the *quantum advantage* (QA) as the order of magnitude improvement in quantum runtime versus classical runtime,

$$\mathrm{QA}(n, \eta) = \max\left[0, \log_{10}\left(\frac{c_t}{q_t}\right)\right]. \tag{10}$$

We now consider two possible definitions of quantum supremacy. First, we can define supremacy as a speedup so large that it is unlikely to be overcome by algorithmic or hardware improvements to the classical sampler, for which we choose a speedup of ten orders of magnitude. Secondly, we may wish to define supremacy as the point at which a computational task is performed in a practical runtime on a quantum device, for which we choose under a week, but in an impractical runtime on a classical device, for which we choose over a century.

These can be summarised as

$$\mathrm{QS}_1 : \mathrm{QA} > 10 \tag{11}$$
$$\mathrm{QS}_2 : q_t < 1 \text{ week}, c_t > 100 \text{ years}. \tag{12}$$

For a quadratic mode scaling and a photon source with rate $R(n) = 10$GHz (beyond the capabilities of the current best single photon sources), we plot QA against n and η in fig.1 of the main text, where q_t^{quad} is used. In order to achieve both quantum supremacy criteria in this regime, we see that it is required to have $\eta > 0.6$ and $n > 60$. We can perform the same analysis with q_t^{lin} and $R(n) = 76n^{-1}$MHz as per [6]. The data for this are plotted in fig. 7. Although using a linear mode scaling increases η, we see that the requirements for quantum supremacy become even more stringent. Indeed, in order to achieve both quantum supremacy criteria here we require $\eta > 0.9$ and $n > 70$.

6. LOSSY BOSON SAMPLING

We have seen that loss can cause significant difficulties in the setting of the standard boson sampling problem, leading to a classical sampler outperforming the quantum experiment. In this section, we explore whether modifying
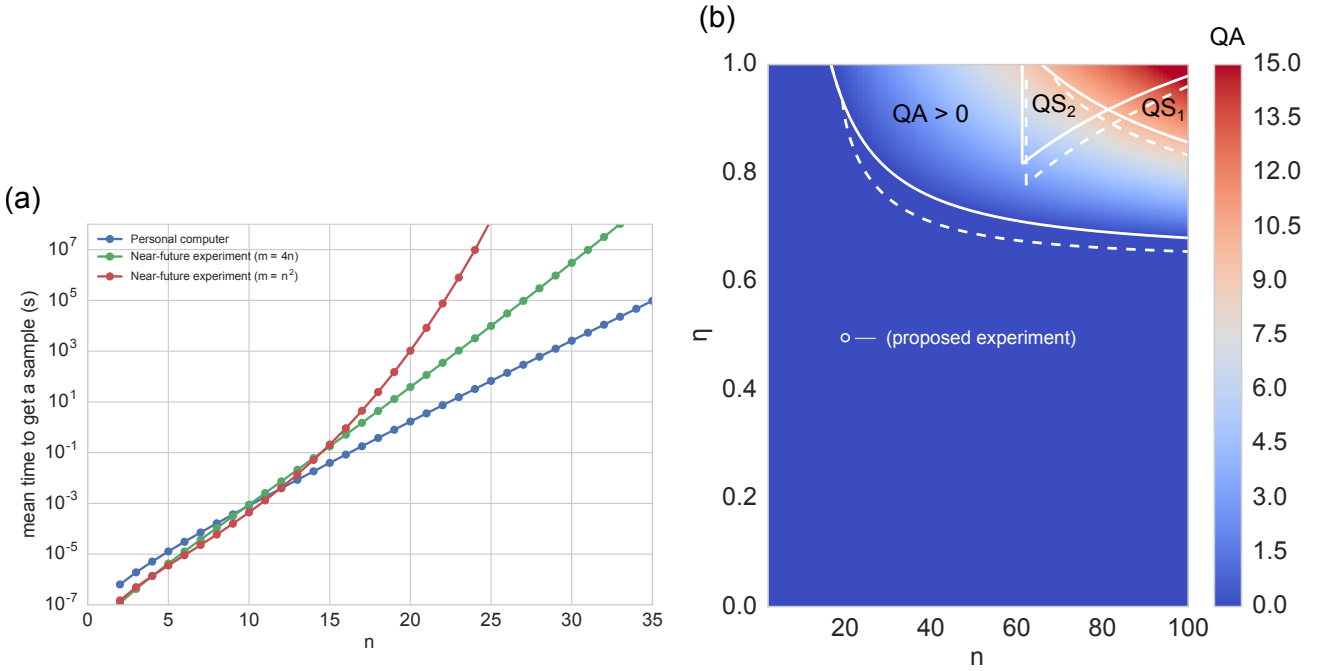
Supplementary Fig. 7. (a): Comparison of average time to output a single tuple sample from the boson sampling distribution experimentally (red: $m = n^2$, green: $m = 4n$) and using a personal computer running MIS (blue). Experimental points were estimated using projected future loss parameters and extrapolated from values stated in [6]. Exact parameter values are stated in Sec. 5. (b): Quantum advantage, QA, as a function of n and η assuming the classical time scaling of a supercomputer and an experimental rate $R(n) = 76n^{-1}$MHz. Lines separate the regions of no quantum advantage, positive quantum advantage and quantum supremacy (as measured by criteria QS_1 or QS_2 or both). Dashed lines demonstrate adjusted regions when up to 2 photons can be lost (optimised to maximise QA). The empty circle represents a proposed future experiment in ref. [6]

the problem to build loss in from the start could mitigate these issues. One might expect that losing photons would also make the boson sampling problem easier classically, because the matrices whose permanents are required to be computed would be smaller. However, it is conceivable that the quantum advantage would increase, enabling a demonstration of quantum supremacy.

A lossy variant of boson sampling was introduced by Aaronson and Brod [7]. In this variant, we assume that $n - k$ photons are lost before entering the circuit enacting the linear optical transfer matrix, so k photons remain. Probabilities in this setting are not determined directly by $|\text{Per}(A_S)|^2$ for some m-tuple S, but by the average of this quantity over all possible ways of losing $n - k$ photons from n photons:

$$\Pr(S) = \frac{1}{|\Lambda_{n,k}|} \sum_{T \in \Lambda_{n,k}} |\text{Per}(A_{S,T})|^2 \tag{13}$$

where $\Lambda_{n,k}$ is the set of k-subsets of $\{1, \ldots, n\}$ and $A_{S,T}$ is the $k \times k$ submatrix of A obtained by taking columns of A according to T and rows of A according to S, which remains a subset of $\{1, \ldots, m\}$. Note that once again we restrict to the collision-free subspace, making the assumption that the probability of a collision is low enough that this does not significantly affect the probabilities.

It was shown by Aaronson and Brod [7] that, if at most a constant number of photons in total are lost before entering the circuit, the lossy boson sampling problem remains hard, under the same assumptions as the original boson sampling problem. However, it was left as an open problem whether the same holds true in the more realistic setting where a constant fraction of photons are lost. Another open problem in [7] was to generalise the loss model to include loss within and after the linear optical circuit. In [38] it is shown that uniform (i.e. mode-independent) loss channels can commute with linear optics. Here we prove the slightly stronger result, that as long as the overall transfer matrix is proportional to a unitary, loss can always be considered at the input even if the physical loss channels, wherever they occur, are not uniform (i.e are mode-dependent).

Consider a boson sampling device consisting of an ideal unitary linear optical transformation U on a set of m optical modes which is preceded or succeeded by path-independent loss. This loss can be modelled by considering a set of m additional virtual ancilla modes such that the optical transfer matrix on all $2m$ modes remains unitary. A uniform

transmission probability of η can then be described by beamsplitters coupling each mode to its corresponding ancilla, resulting in the transfer matrix

$$L = \begin{pmatrix} \sqrt{\eta} & \sqrt{1-\eta} \\ \sqrt{1-\eta} & -\sqrt{\eta} \end{pmatrix} \otimes \mathbb{1}_m \tag{14}$$

and so including the interferometer, the full transfer matrices for input and output losses are

$$M_I = (U \oplus \mathbb{1}_m)L = \begin{pmatrix} \sqrt{\eta}U & \sqrt{1-\eta}U \\ \sqrt{1-\eta}\mathbb{1} & -\sqrt{\eta}\mathbb{1} \end{pmatrix} \tag{15}$$

$$M_O = L(U \oplus \mathbb{1}_m) = \begin{pmatrix} \sqrt{\eta}U & \sqrt{1-\eta}\mathbb{1} \\ \sqrt{1-\eta}U & -\sqrt{\eta}\mathbb{1} \end{pmatrix}. \tag{16}$$

Any m-mode optical state can be expressed in a coherent state basis [39]:

$$\rho_m = \int \lambda(\boldsymbol{\alpha}, \boldsymbol{\beta})|\boldsymbol{\alpha}\rangle\langle\boldsymbol{\beta}| d^m\boldsymbol{\alpha} d^m\boldsymbol{\beta} \tag{17}$$

where $|\boldsymbol{\alpha}\rangle = |(\alpha_1, \alpha_2, ..., \alpha_m)^T\rangle \equiv \bigotimes_{i=1}^m |\alpha_i\rangle$ is an m-mode coherent state and $\langle\boldsymbol{\beta}| = \langle(\beta_1, \beta_2, ...\beta_m)^T| = \bigotimes_{i=1}^m \langle\beta_i|$. A coherent state evolves under a transfer matrix T as

$$\mathcal{U}(T)|\boldsymbol{\alpha}\rangle = \bigotimes_{i=1}^m \left| \sum_j T_{ij}\alpha_j \right\rangle = |T\boldsymbol{\alpha}\rangle \tag{18}$$

It can then be shown that when the initial state contains vacuum in all ancilla modes, $\rho = \rho_m \otimes |\mathbf{0}\rangle\langle\mathbf{0}|_{\mathrm{an}}$, the same state is produced in the m system modes under the transformations M_O and M_I

$$\begin{aligned}
\rho_O &= \mathrm{Tr}_{\mathrm{an}}\left[\mathcal{U}(M_O)\rho\mathcal{U}^\dagger(M_O)\right] \\
&= \mathrm{Tr}_{\mathrm{an}}\left[\int d^{2m}\boldsymbol{\alpha}d^{2m}\boldsymbol{\beta}\lambda(\boldsymbol{\alpha}, \boldsymbol{\beta})\left|\sqrt{\eta}U\boldsymbol{\alpha}\right\rangle\left\langle\sqrt{\eta}U\boldsymbol{\beta}\right| \otimes \left|\sqrt{1-\eta}U\boldsymbol{\alpha}\right\rangle\left\langle\sqrt{1-\eta}U\boldsymbol{\beta}\right|\right] \\
&= \int d^{2m}\boldsymbol{\alpha}d^{2m}\boldsymbol{\beta}\lambda(\boldsymbol{\alpha}, \boldsymbol{\beta})\left|\sqrt{\eta}U\boldsymbol{\alpha}\right\rangle\left\langle\sqrt{\eta}U\boldsymbol{\beta}\right| \mathrm{Tr}\left[\left|\sqrt{1-\eta}U\boldsymbol{\alpha}\right\rangle\left\langle\sqrt{1-\eta}U\boldsymbol{\beta}\right|\right] \\
&= \int d^{2m}\boldsymbol{\alpha}d^{2m}\boldsymbol{\beta}\lambda(\boldsymbol{\alpha}, \boldsymbol{\beta})\left|\sqrt{\eta}U\boldsymbol{\alpha}\right\rangle\left\langle\sqrt{\eta}U\boldsymbol{\beta}\right| \mathrm{Tr}\left[\mathcal{U}(U)|\sqrt{1-\eta}\boldsymbol{\alpha}\rangle\langle\sqrt{1-\eta}\boldsymbol{\beta}|\mathcal{U}^\dagger(U)\right] \\
&= \int d^{2m}\boldsymbol{\alpha}d^{2m}\boldsymbol{\beta}\lambda(\boldsymbol{\alpha}, \boldsymbol{\beta})\left|\sqrt{\eta}U\boldsymbol{\alpha}\right\rangle\left\langle\sqrt{\eta}U\boldsymbol{\beta}\right| \mathrm{Tr}\left[|\sqrt{1-\eta}\boldsymbol{\alpha}\rangle\langle\sqrt{1-\eta}\boldsymbol{\beta}|\right] \\
&= \rho_I.
\end{aligned} \tag{19}$$

More generally, wherever loss occurs in the experiment, the overall transfer matrix K on the system modes can be efficiently characterised [40, 41]. Since path-dependent loss is usually small in experiments, and can be mitigated by interferometer design [14], the matrix $K/||K||_2 \approx U$. The matrix K can then be embedded into a larger unitary matrix acting on additional modes as before. We note all unitary dilations of K, $M_K \in U(m+p)$ where $p \geq m$, can be parameterised using the Cosine-Sine decomposition as

$$M_K = \begin{pmatrix} A & 0 \\ 0 & X \end{pmatrix} \left(\begin{array}{cc|c} \cos(\Theta) & -\sin(\Theta) & 0 \\ \sin(\Theta) & \cos(\Theta) & 0 \\ \hline 0 & 0 & \mathbb{1}_{p-m} \end{array} \right) \begin{pmatrix} B^\dagger & 0 \\ 0 & Y \end{pmatrix}$$

where $K = A\cos(\Theta)B^\dagger$, with $A, B \in U(m)$ and $\cos(\Theta) = \mathrm{diag}(\cos\theta_1, ..., \cos\theta_m)$ with $\theta_1 \leq \theta_2 \leq .. \leq \theta_m$, is a singular value decomposition of K and $X, Y \in U(p)$. In fact, all unitary dilations are related by the choice of X and Y [42, 43]. Since $\mathcal{U}(Y)|\mathbf{0}\rangle = |\mathbf{0}\rangle$ and the choice of X does not affect ρ_K using the cyclic property of the trace as above, setting $\eta = ||K||_2^2$, we see that $\rho_K = \rho_I$. Moreover, we have shown that all unitary dilations of a transfer matrix produce the same output state and therefore any boson sampling experiment with overall path-independent losses is equivalent to introducing uniform loss channels with transmission probability η at the input, followed by the ideal unitary evolution.

Our MIS method can readily be adapted to deal with loss at the input, by inserting an initial step for each tuple to be output, which generates a uniformly random input subset T. This would be followed by the usual MIS method with permanents of $k \times k$ submatrices computed. We described a similar adaptation in Sec. 4 4.3 to deal with scattershot boson sampling. The core part of the classical sampling procedure for both the lossy and scattershot variants therefore follows precisely that of standard boson sampling. From our analysis of the required burn in period for MIS, we can see that the performance of our sampler will be similar to the standard boson sampling case. That is, it is likely that lossy and scattershot boson sampling is no more difficult classically than standard boson sampling.

The modified MIS sampler can then be compared with experimental performance. Using the fixed loss regime of [7], we adjusted the quantum and classical runtimes and optimised QA for the case where up to 2 photons were allowed to be lost. This resulted in the dashed lines in fig.2d) and e) of the main text, and in fig. 7. We see that allowing this loss relaxes the requirements for quantum supremacy, but does not significantly change the experimental regimes required.

In current and future experiments, where $\eta \ll 1$, more photons than this are usually lost. Again using expected loss parameter values from [6], we find that the optimal number of photons to lose (in terms of performance enhancement relative to the classical sampler) increases with n. The estimated experimental time to produce a tuple when $n = 20$, $k = 12$ and $m = 80$ is $11.8\mu s$, where this time encompasses the expected number of repetitions required to detect exactly k photons. With the same parameter values and problem size, the estimated MIS runtime on a personal computer is 3.9ms. This amounts to a factor of 333 performance enhancement for the experiment over the personal computer.

However, the MIS sampler is not the only way in which one could attempt to sample from the lossy boson sampling distribution: indeed, it could even be the case that enough loss could render the boson sampling distribution easy to sample from classically. For example, it was shown in [44] that a class of quantum circuits whose output probability distributions are likely to be hard to sample from classically becomes easy in the presence of noise. We also remark that Rahimi-Keshari, Ralph and Caves [45] have proven classical simulability of boson sampling under various physically motivated models of errors (e.g. loss, mode mismatch, and dark counts), and that Aaronson and Brod have speculated [7] that each probability in the small-k lossy boson sampling distribution tends to the product of the squared 2-norms of the rows of A_S – an efficiently computable quantity.

Lossy boson sampling also suffers from an additional difficulty when one considers performing likelihood based verification techniques. In particular, computing the likelihood of a sample according to the indistinguishable boson hypothesis requires computing eqn. (13) for each tuple S in a sample. The sum over all $T \in \Lambda_{n,k}$ necessitates a factor of $\binom{n}{k}$ slowdown in computing the likelihood relative to a standard (i.e. lossless) k-photon boson sampling experiment. The worst case for this is when $k = n/2$ photons survive, amounting to $\sim 2^n / \sqrt{\pi n/2}$ matrix permanent computations per likelihood calculation. To avoid this exponential slowdown, one would be forced to use a nonstandard verification technique [21, 46], or devise a more efficient means of computing (13) than naïve evaluation of the sum.

7. DISCUSSION

We finish by discussing the limitations of our techniques, and prospects for future experimental and classical improvements.

7.1. Limitations

Our sampling methods have some limitations that we now set out. Our rejection sampling method which is applicable for n up to approximately 12, is guaranteed to give independent samples. Our approach to improving the efficiency of the method requires us first to estimate the maximum probability μ of the distribution. If this estimate is very poor, that is if there is a large proportion of the probability mass greater than μ, then there is a risk we will not sample from the boson sampling distribution correctly. From our computational experiments, we can see that this appears to be very unlikely to occur at least in the range we are able to test.

For our MIS method, we have to be concerned both with how long to wait until convergence and the related question of potential dependence between samples. It is unfortunately not possible to prove that the samples we take are entirely independent of each other or indeed of the initial state. Although this question of dependence in boson sampling is not one that has arisen previously, it is desirable for samples to be taken independently.

We have mitigated this problem in two ways. The first is by our choice of MIS as a sampling method, which means that samples from our proposal distribution are taken independently. The second is by our use of a thinning procedure. In fig. 5 we see that this thinning procedure successfully reduces the auto-correlation of the computed probabilities of the sampled submatrices to near zero. In addition to this, in fig. 4 we see that extending the thinning procedure

beyond a certain point causes no improvement in the results of a likelihood ratio test as outlined in [29]. However, it is still possible that there remains dependence on some other detailed property of the sampled submatrices which is not exposed by looking at the computed probabilities or likelihood ratios.

7.2. Prospects for experimental boson sampling

The fact that producing boson sampling machines requires low loss rates has been known to be a hurdle since the conception of the problem [1]. However, the extent of this hurdle has not been well understood, not least due to the lack of an optimised classical boson sampling algorithm.

Our results suggest that with loss parameters and photon numbers expected in the near future, no straightforward or significant quantum supremacy is achievable via boson sampling. This is *not* to say that it is impossible to build a boson sampling machine which demonstrates quantum supremacy. However, it is absolutely necessary to further decrease loss rates, or develop efficient fault-tolerance techniques targeted at linear optics in order to achieve this. A machine performing boson sampling with tens of photons in thousands of modes is likely required if we are to witness an experiment significantly outperforming a classical simulator. We note the engineering feat required to build an interferometer of this size alone may prohibit its realisation in the near future.

One approach to mitigating loss in an experiment would be to reduce circuit depth. In fig. 1e) of the main text and fig. 7 we have assumed a circuit depth of $d = m$ to infer η for the specific experimental points plotted (solid and open white points). Using more exotic structures such as the 3D platform used in [47], it may be possible to reduce this scaling, and hence increase η.

In addition to the problems caused by photon loss, it is also necessary to build reconfigurable linear optical circuits capable of performing high precision transformations on modes [48–50], to generate highly indistinguishable photons [51, 52] and to limit detector noise and higher order photon number terms in the input state [45] in order to perform a classically intractable task. We note that although there exist methods to aid in achieving these things – such as the spectral filtering of single photons or adding additional components to perform higher fidelity beamsplitter transformations [53, 54] – they typically result in an increased amount of photon loss.

There have also been proposals for performing boson sampling using other hardware platforms, such as trapped ions [55] and superconducting qubits [56]. Although currently experimentally untested, it is possible these approaches could provide better scalability. However, coherence and gate operation times are likely to limit the rates and sizes of implementations possible in these architectures, e.g. in ref. [56] current estimates suggest up to 20 photons.

7.3. Prospects for classical boson sampling

Our current approach to simulating a boson sampling experiment assumes a Fock state input of the form given in eqn. (3). Another proposal for boson sampling type experiments is based on Gaussian input states [17, 57]. In this case, Hamilton et al. [57] showed that the probabilities in the output distribution are related to the hafnian [58] of a submatrix, rather than the permanent. It has been shown by Björklund that hafnians can be computed using an algorithm with runtime matching Ryser's algorithm for the permanent, up to polynomial factors [59]. We are hopeful that the techniques we have presented in this paper can be carried over to Gaussian boson sampling, potentially allowing for the classical simulation of experiments designed to generate molecular vibronic spectra [60].

In terms of performance, it is possible that improvements can be made in a number of ways. Firstly, it is likely that MIS with the distinguishable particle proposal distribution is not optimal for classical boson sampling. Although computing matrix permanents will remain the bottleneck, requiring fewer of these computations per output tuple could result in up to a ~ 100 times speed-up. Secondly, the implementation could be scaled up to run on high performance computing hardware. This would require careful parallelisation of both MIS and Ryser's (or Balasubramanian–Bax/Franklin–Glynn) algorithm. We remark that parallelisation of the most computationally expensive parts of MIS is relatively trivial. As the proposal distribution is independent of the current state, we can distribute the permanent computations across multiple processors. In addition to this, it may be possible to build a more realistic model of the experiment (including, for example, partial distinguishability of photons) to allow for some approximations of matrix permanents to be made, or for permanents of smaller submatrices to be computed with some probability.

8. CONCLUSION

We have shown that boson sampling can be simulated classically for a range of parameters that was previously considered likely to be computationally intractable [1, 2, 4, 5]. Our results pose a challenge for both experimentalists and

theorists. The challenge for experimentalists is to build low-loss, large-scale linear optical circuits going substantially beyond current technology, while the challenge for theorists is either to develop efficient linear-optical loss-tolerance schemes, or to find some way to distinguish our classical boson sampling simulators from the real boson sampling distribution.

Note that one way of implementing a large-scale loss-free boson sampling experiment is simply to simulate boson sampling on a fault-tolerant universal quantum computer, perhaps based on some other technology than optics. However, in the near term this is unlikely to be a more efficient means of demonstrating quantum supremacy than an approach targeted directly at the underlying hardware [3, 61–63].

We close by highlighting another interpretation of our results: as providing an additional tool for classical verification of future boson sampling experiments. Due to the computational difficulty of simulating the boson sampling distribution, existing verification procedures have not been designed to take advantage of statistical information about the true boson sampling distribution. Indeed for $n > 7$ this information has simply not been available within a realistic time frame. Our new efficient computational sampling procedure allows us for the first time to test directly whether an experiment is sampling from the correct distribution. Setting the null hypothesis to be that we are sampling from the boson sampling distribution, we can sample directly from the true distribution using our MIS procedure and then perform standard statistical hypothesis tests on the results of any boson sampling experiment. MIS can also be used to certify lossy versions of boson sampling, where previous methods based on computing likelihoods are inefficient, as discussed in Section 6. This will potentially give a new level of confidence that was not previously available in the accuracy of any experimental design and implementation.

---

[1] Aaronson, S. & Arkhipov, A. The computational complexity of linear optics. In *Proc. 43rd Annual ACM Symp. Theory of Computing*, 333–342 (2011).

[2] Preskill, J. Quantum computing and the entanglement frontier. Preprint at `https://arxiv.org/abs/1203.5813` (2012).

[3] Lund, A. P., Bremner, M. J. & Ralph, T. C. Quantum sampling problems, bosonsampling and quantum supremacy. *npj Quantum Information* **3**, 15 (2017).

[4] Bentivegna, M. *et al.* Experimental scattershot boson sampling. *Science Advances* **1**, e1400255 (2015).

[5] Latmiral, L., Spagnolo, N. & Sciarrino, F. Towards quantum supremacy with lossy scattershot boson sampling. *New Journal of Physics* **18**, 113008 (2016).

[6] Wang, H. *et al.* High-efficiency multiphoton boson sampling. *Nat. Photon.* **11**, 361–365 (2017).

[7] Aaronson, S. & Brod, D. J. BosonSampling with lost photons. *Phys. Rev. A* **93**, 012335 (2016).

[8] Valiant, L. The complexity of computing the permanent. *Theoretical Computer Science* **8**, 189–201 (1979).

[9] Aaronson, S. A linear-optical proof that the permanent is #P-hard. *Proc. Roy. Soc. Ser. A* **467**, 3393–3405 (2011).

[10] Rohde, P. P. & Ralph, T. C. Error tolerance of the boson-sampling model for linear optics quantum computing. *Phys. Rev. A* **85**, 022332 (2012).

[11] Arkhipov, A. BosonSampling is robust against small errors in the network matrix. *Phys. Rev. A* **92**, 062326 (2015).

[12] Arkhipov, A. & Kuperberg, G. The bosonic birthday paradox. *Geometry & Topology Monographs* **18**, 1–7 (2012).

[13] Reck, M., Zeilinger, A., Bernstein, H. J. & Bertani, P. Experimental realization of any discrete unitary operator. *Phys. Rev. Lett.* **73**, 58–61 (1994).

[14] Clements, W. R., Humphreys, P. C., Metcalf, B. J., Kolthammer, W. S. & Walmsley, I. A. Optimal design for universal multiport interferometers. *Optica* **3**, 1460–1465 (2016).

[15] Carolan, J. *et al.* Universal linear optics. *Science* **349**, 711–716 (2015).

[16] Harris, N. C. *et al.* Quantum transport simulations in a programmable nanophotonic processor. *Nat. Photon.* **11**, 447–452 (2017).

[17] Lund, A. P. *et al.* Boson sampling from a Gaussian state. *Phys. Rev. Lett.* **113**, 100502 (2014). `arXiv:1305.4346`.

[18] Aaronson, S. Scattershot BosonSampling: A new approach to scalable BosonSampling experiments (2013). `http://www.scottaaronson.com/blog/?p=1579`.

[19] Ryser, H. J. *Combinatorial Mathematics*. Carus Mathematical Monographs, No. 14 (1963).

[20] Wu, J. *et al.* Computing permanents for boson sampling on Tianhe-2 supercomputer. Preprint at `https://arxiv.org/abs/1606.05836` (2016).

[21] Aaronson, S. & Arkhipov, A. BosonSampling Is Far From Uniform. *Quant. Inf. Comp.* **14**, 1383–1432 (2014).

[22] Liu, J. Metropolized independent sampling with comparisons to rejection sampling and importance sampling. *Statistics and Computing* **6**, 113–119 (1996).

[23] Liu, J. *Monte Carlo Strategies in Scientific Computing* (Springer, 2004).

[24] Barvinok, A. I. Two algorithmic results for the traveling salesman problem. *Mathematics of Operations Research* **21**, 65–84 (1996).

[25] Jerrum, M., Sinclair, A. & Vigoda, E. A polynomial-time approximation algorithm for the permanent of a matrix with nonnegative entries. *J. ACM* **51**, 671–697 (2004).

[26] Gogolin, C., Kliesch, M., Aolita, L. & Eisert, J. Boson-Sampling in the light of sample complexity. Preprint at `https://arxiv.org/abs/1306.3995` (2013).

[27] Praestgaard, J. T. Permutation and bootstrap Kolmogorov-Smirnov tests for the equality of two distributions. *Scandinavian Journal of Statistics* **22**, 305–322 (1995).

[28] Sekhon, J. Multivariate and propensity score matching software with automated balance optimization: The matching package for r. *Journal of Statistical Software* **42**, 1–52 (2011).

[29] Bentivegna, M. *et al.* Bayesian approach to boson sampling validation. *Int. J. Quantum Inform.* **12**, 1560028 (2014).

[30] Spring, J. B. *et al.* Boson Sampling on a Photonic Chip. *Science* **339**, 798–801 (2013).

[31] Crespi, A. *et al.* Integrated multimode interferometers with arbitrary designs for photonic boson sampling. *Nat. Photon.* **7**, 545–549 (2013).

[32] Tillmann, M. *et al.* Experimental boson sampling. *Nat. Photon.* **7**, 540–544 (2013).

[33] Broome, M. A. *et al.* Photonic boson sampling in a tunable circuit. *Science* **339**, 794–798 (2013).

[34] Carolan, J. *et al.* On the experimental verification of quantum complexity in linear optics. *Nat. Photon.* **8**, 621–626 (2014).

[35] Spagnolo, N. *et al.* Experimental validation of photonic boson sampling. *Nat. Photon.* **8**, 615–620 (2014).

[36] He, Y. *et al.* Time-bin-encoded boson sampling with a single-photon device. *Phys. Rev. Lett.* **118**, 190501 (2017).

[37] Loredo, J. C. *et al.* Boson sampling with single-photon fock states from a bright solid-state source. *Phys. Rev. Lett.* **118**, 130503 (2017).

[38] Berry, D. W. & Lvovsky, A. I. Linear-optical processing cannot increase photon efficiency. *Phys. Rev. Lett.* **105**, 203601 (2010).

[39] Gerry, C. & Knight, P. *Introductory quantum optics* (Cambridge university press, 2005).

[40] Laing, A. & O'Brien, J. L. Super-stable tomography of any linear optical device. Preprint at `https://arxiv.org/abs/1208.2868` (2012).

[41] Rahimi-Keshari, S. *et al.* Direct characterization of linear-optical networks. *Opt. Express* **21**, 13450–13458 (2013).

[42] Horn, R. A. & Johnson, C. R. *Topics in Matrix Analysis* (Cambridge University Press, 1991).

[43] Allen, J. & Arceo, D. Matrix dilations via cosine-sine decomposition. Tech. Rep., DTIC Document (2006).

[44] Bremner, M. J., Montanaro, A. & Shepherd, D. J. Achieving quantum supremacy with sparse and noisy commuting quantum computations. *Quantum* **1**, 8 (2017).

[45] Rahimi-Keshari, S., Ralph, T. & Caves, C. Sufficient conditions for efficient classical simulation of quantum optics. *Phys. Rev. X* **6**, 021039 (2016).

[46] Walschaers, M. *et al.* Statistical benchmark for bosonsampling. *New Journal of Physics* **18**, 032001 (2016).

[47] Crespi, A. *et al.* Suppression law of quantum states in a 3d photonic fast fourier transform chip. *Nature Communications* **7**, 10469 (2016).

[48] Leverrier, A. & García-Patrón, R. Analysis of circuit imperfections in BosonSampling. *Quantum Inf. Comput.* **15**, 0489–0512 (2015). `arXiv:1309.4687`.

[49] Kalai, G. & Kindler, G. Gaussian noise sensitivity and BosonSampling (2014). `arXiv:1409.3093`.

[50] Arkhipov, A. BosonSampling is robust against small errors in the network matrix. *Phys. Rev. A* **92**, 062326 (2015).

[51] Tichy, M. C. Sampling of partially distinguishable bosons and the relation to the multidimensional permanent. *Phys. Rev. A* **91**, 022316 (2015).

[52] Shchesnovich, V. S. Sufficient condition for the mode mismatch of single photons for scalability of the boson-sampling computer. *Phys. Rev. A* **89**, 022333 (2014).

[53] Miller, D. A. B. Perfect optics with imperfect components. *Optica* **2**, 747–750 (2015).

[54] Wilkes, C. M. *et al.* 60db high-extinction auto-configured mach–zehnder interferometer. *Opt. Lett.* **41**, 5318–5321 (2016).

[55] Shen, C., Zhang, Z. & Duan, L.-M. Scalable implementation of boson sampling with trapped ions. *Phys. Rev. Lett.* **112**, 050504 (2014).

[56] Peropadre, B., Guerreschi, G. G., Huh, J. & Aspuru-Guzik, A. Proposal for microwave boson sampling. *Phys. Rev. Lett.* **117**, 140505 (2016).

[57] Hamilton, C. S. *et al.* Gaussian boson sampling. Preprint at `https://arxiv.org/abs/1612.01199` (2016).

[58] Minc, H. *Permanents* (Addison-Wesley, 1978).

[59] Björklund, A. Counting perfect matchings as fast as Ryser. In *Proc. 23rd ACM-SIAM Symp. Discrete Algorithms*, 914–921 (2012). `arXiv:1107.4466`.

[60] Huh, J., Guerreschi, G. G., Peropadre, B., McClean, J. R. & Aspuru-Guzik, A. Boson sampling for molecular vibronic spectra. *Nat. Photon.* **9**, 615–620 (2015).

[61] Boixo, S. *et al.* Characterizing quantum supremacy in near-term devices. Preprint at `https://arxiv.org/abs/1608.00263` (2016).

[62] Bremner, M., Montanaro, A. & Shepherd, D. Average-case complexity versus approximate simulation of commuting quantum computations. *Phys. Rev. Lett.* **117**, 080501 (2016).

[63] Bermejo-Vega, J., Hangleiter, D., Schwarz, M., Raussendorf, R. & Eisert, J. Architectures for quantum simulation showing quantum supremacy. Preprint at `https://arxiv.org/abs/1703.00466` (2017).