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] and Anthony Laing[1]*

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 such 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 (for example, 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.

Although choosing a number of modes $m \sim n^5 \log^2 n$ guarantees that the distribution of any $n \times n$ submatrix 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, that is, bunch; the conjectured hardness applies only to collision-free events, that is, 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 (ref. 7). However, our classical algorithm, based on Metropolised independence sampling (MIS), although 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

[1]Quantum Engineering and Technology Laboratories, School of Physics and Department of Electrical and Electronic Engineering, University of Bristol, Bristol BS8 1TL, UK. [2]Department of Physics, Imperial College London, London SW7 2AZ, UK. [3]Department of Computer Science, University of Bristol, Bristol BS8 1UB, UK. [4]School of Mathematics, University of Bristol, Bristol BS8 1TW, UK. *e-mail: Anthony.Laing@bristol.ac.uk
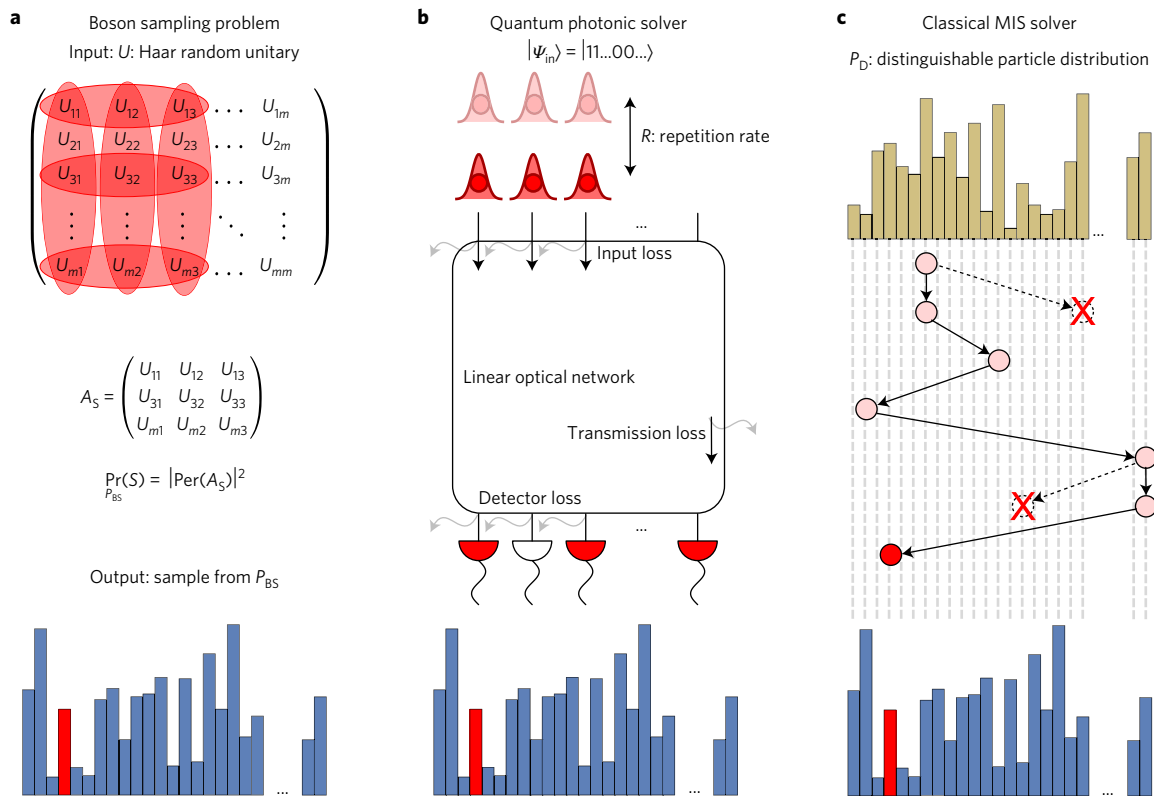
Figure 1 | Photonic and classical approaches to the boson sampling problem. a, Definition of the problem. Output a sample from the distribution defined by the modulus squared permanents of submatrices of a Haar random unitary matrix U. **b**, Photonic experiments solve the problem by propagating single photons through a linear optical network followed by single-photon detection and can be broadly parameterized 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.

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 efficient[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

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

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 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 $\tau_{\text{burn}} n \times n$ real-valued matrix permanents and $\tau_{\text{burn}} n \times n$ complex-valued matrix permanents. Each subsequent sample requires $\tau_{\text{thin}} n \times n$ 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. 2b, 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.
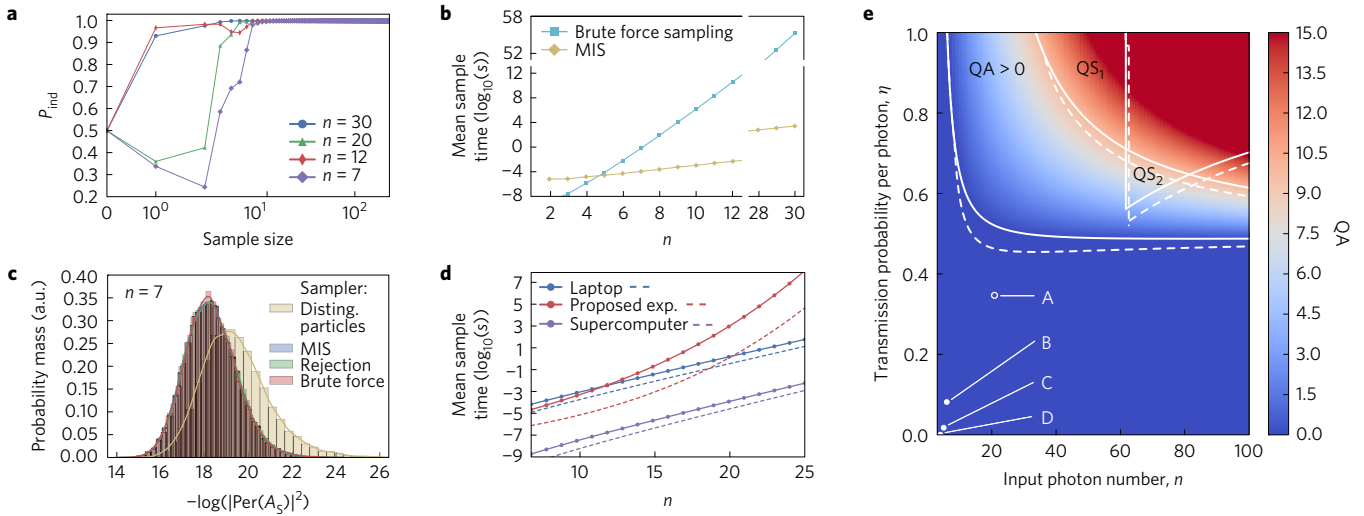
Figure 2 | Results and projections for classical boson sampling. a, A likelihood ratio test with the boson sampling and distinguishable particle distributions as the hypotheses for sample sizes of up to 250 for $n = 7, 12, 20$ and 30 bosons. P_{ind} is the probability that the data are drawn from the 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 20,000 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 two 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 two photons can be lost (optimized to maximize 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.

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 (although 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. 2a 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. 2c 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 $-\log(|\text{Per}(A_S)|^2)$, 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 $\text{Per}(|A_S|^2)$ gives the probability to observe a transition of distinguishable particles). Figure 2c 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. 2c between distributions 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 reassuring 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 algorithm. This is because photon losses scale exponentially badly with n (refs 1,29). The runtime for an experiment with a transmission probability resulting from fixed loss (generation, coupling, detection) η_f and a transmission probability η_0, resulting from loss per unit optical circuit depth scales as $\mathcal{O}((1/\eta_f)^n (1/\eta_0)^{dn})$ 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 $\eta = \eta_f \eta_0^d$ 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,

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

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 summarized as

$$QS_1 : QA > 10 \qquad (5)$$

$$QS_2 : q_t < 1 \text{ week}, c_t > 100 \text{ years} \qquad (6)$$

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 (ref. 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 $\eta < 0.002$ (refs 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 ($76n^{-1}$ 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 approximately 0.35, which would be a major experimental breakthrough. However, as shown in Fig. 2d, in this case we predict that the classical runtime would still be over six orders of magnitude faster. Fig. 2e 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. 2e 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 probably be solved for 50 photons and any given Haar random unitary matrix, in a feasible time frame using today's classical technology. Although 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 although 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], although 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].

Methods

Methods, including statements of data availability and any associated accession codes and references, are available in the online version of this paper.

References

1. Aaronson, S. & Arkhipov, A. *Proc. 43rd Annual ACM Symp. Theory Comput.* The computational complexity of linear optics. 333–342 (ACM, 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 Quant. Inf.* **3**, 15 (2017).
7. Latmiral, L., Spagnolo, N. & Sciarrino, F. Towards quantum supremacy with lossy scattershot boson sampling. *New J. Phys.* **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 J. Comput.* **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. *Sci. Adv.* **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. *Stat. Comput.* **6,** 113–119 (1996).
25. Aaronson, S. & Arkhipov, A. Boson sampling is far from uniform. *Quantum. 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 Inf.* **12,** 1560028 (2014).
28. Praestgaard, J. T. Permutation and bootstrap Kolmogorov–Smirnov tests for the equality of two distributions. *Scand. J. Stat.* **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. *Geom. Topol. Monogr.* **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. Boson sampling with lost photons. *Phys. Rev. A* **93,** 012335 (2016).
36. Russell, N. J., Chakhmakhchyan, L., O'Brien, J. L. & Laing, A. Direct dialling of Haar random unitary matrices. *New J. Phys.* **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).

Author contributions

All authors contributed to discussions and project development. Simulations were carried out by A.N. Data were analysed by A.N., C.S., R.C., A.M. and A.L. The MIS algorithm and the Random Restart Hill-Climbing routine used in our RS algorithm are by R.C. Additional classical sampling algorithms were designed and implemented by A.N. and R.C., with input from C.S., A.M. and A.L. Code optimization was carried out by E.J. The proof that photon loss can be considered at the input is by P.M.B. The manuscript was written by A.L., A.N. and C.S., with input from R.C. and A.M. The project was conceived and managed by A.L.

Additional information

Supplementary information is available in the online version of the paper. Reprints and permissions information is available online at www.nature.com/reprints. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to A.L.

Competing financial interests

The authors declare no competing financial interests.

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 submatrix 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:

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

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):

$$\Pr_{P_D}(S) = \frac{\mathrm{Per}\left(|A_S|^2\right)}{s_1! \ldots s_n!} \qquad (8)$$

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 $O(mn)$ (ref. 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: Step 1: Generate a sample x from Q. Step 2: With probability $f(x)/\lambda g(x)$, 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 subnormalized, so is not quite a probability distribution. However, the rejection sampling algorithm is blind to this subnormalization (as this is effectively the same as increasing λ), so will generate samples from the renormalized distribution.

Each iteration of rejection sampling requires the computation of one permanent, corresponding to $f(x)$. 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 ref. 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 interested only 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 and 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

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

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

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

We note that q_x and r_x should not simply correspond to $|\mathrm{Per}(A_x)|^2$ and $\mathrm{Per}\left(|A_x|^2\right)$, 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 normalized 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 normalization of q_x efficiently, here we instead average the probability of the output being collision-free over the Haar measure[1,30]:

$$P_{\mathrm{CFS}} \approx \binom{m}{n} \Big/ \binom{m+n-1}{n} \qquad (11)$$

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 normalization 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 $|\text{Per}(A_S)|^2$ for some m-tuple S, but by the average of this quantity over all possible ways of losing $n - k$ photons from n photons:

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

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 generalize the loss model to include loss within and after the linear optical circuit. Here we prove a slightly stronger result than that shown in ref. 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 (that is, 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

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

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

$$M_\text{I} = (U \oplus \mathbb{1}_m)L = \begin{pmatrix} \sqrt{\eta}U & \sqrt{1-\eta}U \\ \sqrt{1-\eta}\mathbb{1} & -\sqrt{\eta}\mathbb{1} \end{pmatrix} \qquad (14)$$

$$M_\text{O} = L(U \oplus \mathbb{1}_m) = \begin{pmatrix} \sqrt{\eta}U & \sqrt{1-\eta}\mathbb{1} \\ \sqrt{1-\eta}U & -\sqrt{\eta}\mathbb{1} \end{pmatrix} \qquad (15)$$

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

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

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

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

It can then be shown that when the initial state contains vacuum in all ancilla modes, $\rho = \rho_m \otimes |0\rangle\langle0|_\text{an}$, the same state is produced in the m system modes under the transformations M_O and M_I

$$\rho_\text{O} = \text{Tr}_\text{an}\left[\mathcal{U}(M_\text{O})\rho\,\mathcal{U}^\dagger(M_\text{O})\right]$$

$$= \text{Tr}_\text{an}\left[\int d^{2m}\boldsymbol{\alpha}\, d^{2m}\boldsymbol{\beta}\,\lambda(\boldsymbol{\alpha},\boldsymbol{\beta})\left|\sqrt{\eta}U\boldsymbol{\alpha}\right\rangle\left\langle\sqrt{\eta}U\boldsymbol{\beta}\right| \otimes |\sqrt{1-\eta}U\boldsymbol{\alpha}\rangle\langle\sqrt{1-\eta}U\boldsymbol{\beta}|\right]$$

$$= \int d^{2m}\boldsymbol{\alpha}\, d^{2m}\boldsymbol{\beta}\,\lambda(\boldsymbol{\alpha},\boldsymbol{\beta})\left|\sqrt{\eta}U\boldsymbol{\alpha}\right\rangle\left\langle\sqrt{\eta}U\boldsymbol{\beta}\right| \text{Tr}\left[|\sqrt{1-\eta}U\boldsymbol{\alpha}\rangle\langle\sqrt{1-\eta}U\boldsymbol{\beta}|\right]$$

$$= \int d^{2m}\boldsymbol{\alpha}\, d^{2m}\boldsymbol{\beta}\,\lambda(\boldsymbol{\alpha},\boldsymbol{\beta})\left|\sqrt{\eta}U\boldsymbol{\alpha}\right\rangle\left\langle\sqrt{\eta}U\boldsymbol{\beta}\right| \text{Tr}\left[\mathcal{U}(U)|\sqrt{1-\eta}\boldsymbol{\alpha}\rangle\langle\sqrt{1-\eta}\boldsymbol{\beta}|\mathcal{U}^\dagger(U)\right]$$

$$= \int d^{2m}\boldsymbol{\alpha}\, d^{2m}\boldsymbol{\beta}\,\lambda(\boldsymbol{\alpha},\boldsymbol{\beta})\left|\sqrt{\eta}U\boldsymbol{\alpha}\right\rangle\left\langle\sqrt{\eta}U\boldsymbol{\beta}\right| \text{Tr}\left[|\sqrt{1-\eta}\boldsymbol{\alpha}\rangle\langle\sqrt{1-\eta}\boldsymbol{\beta}|\right]$$

$$= \rho_\text{I} \qquad (18)$$

More generally, wherever loss occurs in the experiment, the overall transfer matrix K on the system modes can be efficiently characterized[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 parameterized 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 $\cos(\Theta) = \text{diag}(\cos\theta_1, \ldots, \cos\theta_m)$ with $\theta_1 \leq \theta_2 \leq \ldots \leq \theta_m$, 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 (refs 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 $\rho_K = \rho_\text{I}$. 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.

Data availability. The data that support the plots within this paper and other findings of this study are available at the University of Bristol data repository, data.bris, at https://doi.org/10.5523/bris.2ok605tzyel9o20lpmz2kky7wd.

References

41. Barvinok, A. I. Two algorithmic results for the traveling salesman problem. *Math. Oper. Res.* **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 Univ. 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 Univ. Press, 1991).
48. Allen, J. & Arceo, D. *Matrix Dilations via Cosine-Sine Decomposition* Tech. Rep., DTIC Document (SCC, 2006).