Science Advances
AAAS

Supplementary Materials for

Quantum computational advantage via high-dimensional Gaussian boson sampling

Abhinav Deshpande, Arthur Mehta, Trevor Vincent, Nicolás Quesada, Marcel Hinsche, Marios Ioannou, Lars Madsen, Jonathan Lavoie, Haoyu Qi, Jens Eisert, Dominik Hangleiter, Bill Fefferman, Ish Dhand*

*Corresponding author. Email: ishdhand@gmail.com

Published 5 January 2022, *Sci. Adv.* **8**, eabi7894 (2022)
DOI: 10.1126/sciadv.abi7894

This PDF file includes:

Supplementary materials

Evidence for hiding in GBS

In this section, we characterize the distribution of the symmetric product of $N \times K$ sub-matrices of $M \times M$ Haar-random unitaries. As described earlier, the Hafnian of such symmetric products determines the output distribution of GBS. Here, we give evidence that this distribution tends to the distribution of the symmetric product XX^T for X being an $N \times K$ Gaussian matrix. In GBS, this ensures the hiding property since a small $N \times N$ symmetric Gaussian matrix XX^T can be hidden in a large symmetric unitary matrix $UI_K U^T$ for any $K \geq N$. Since any particular sub-matrix cannot be distinguished from any other such sub-matrix of the same size, this enforces the constant error budget of an adversarial sampler to be roughly equally distributed across all outcomes.

In particular, we consider three regimes—with respect to the relations between the total number of photons at the output (N), the number of input squeezers (K), and the number of modes (M)—in order to provide evidence for Conjecture 1. This conjecture relates the following ensembles of random matrices.

1. $\mathcal{H}^M_{N,K}$: The ensemble of $N \times K$ sub-matrices of Haar-random unitaries $U \in U(M)$.
2. $\mathcal{G}_{N,K}(\mu, \sigma^2)$: The ensemble of $N \times K$ matrices with independent and identically distributed (i.i.d.) complex normal entries with mean μ and variance σ^2.
3. $\text{COE}^M_{N,K}$: The ensemble of matrices VV^T where $V \sim \mathcal{H}^M_{N,K}$.
4. $\mathcal{G}^{\text{sym}}_{N,K}(\mu, \sigma^2)$: The ensemble of matrices XX^T where $X \sim \mathcal{G}_{N,K}(\mu, \sigma^2)$.

As the conjecture might be interesting for random matrix theory in itself, we will abstract away the meaning of the parameters K, N, M.

Conjecture 1 (Hiding in GBS)**.** *For any K such that $N \leq K \leq M$ the following statements are true:*

1. *For $M \in O(N^{2+\epsilon})$ and $\epsilon \in (0, 1]$, $\mathit{COE}^M_{N,K}$ asymptotically approaches $\mathcal{G}^{\mathit{sym}}_{N,K}(0, 1/M)$ in probability in terms of the entrywise max-norm.*
2. *There exists a polynomial p such that for any $\delta > 0$ and $M \geq p(N)/\delta$, the total-variation distance $\|\cdot\|_{TV}$ between $\mathit{COE}^M_{N,K}$ and $\mathcal{G}^{\mathit{sym}}_{N,K}(0, 1/M)$satisfies*

$$\left\|\mathit{COE}^M_{N,K} - \mathcal{G}^{\mathit{sym}}_{N,K}(0, 1/M)\right\|_{TV} \in O(\delta). \tag{S1}$$

Here, we give analytical evidence that the characterization of Conjecture 1 holds true in the extreme cases of $K = N$ and $K = M$ for M growing fast enough with N and numerically show that it is true for any K such that $N \leq K \leq M$.

In the first regime we consider K is such that $M \in \Omega(K^5 \log^2 K)$ and $N = K$. This regime closely resembles the one in the original boson sampling proposal (thus we refer to it as the "AA regime") for which we will see that both parts 1. and 2. are provably true. In this regime, Aaronson and Arkhipov (*7*) have proven that all $N \times K$ sub-matrices of Haar-random linear-optical unitaries U, are approximately Gaussian distributed. In particular they show that $\mathcal{H}_{N,K}^{M}$ asymptotically approaches $\mathcal{G}_{N,K}(0, 1/M)$ as well as bounding the rate of convergence by showing that the total-variation distance satisfies

$$\|\mathcal{H}_{N,K}^{M} - \mathcal{G}_{N,K}(0, 1/M)\|_{\text{TV}} \in O(\delta) \tag{S2}$$

for $M \geq (N^5/\delta) \log^2(N/\delta)$ (*7*). Using this we can directly see that Conjecture 6 is also true in the "AA regime".

On the other end of the spectrum, we consider the regime in which $K = M$ where part 1. of the conjecture is provably true. For this case, Jiang (*22*) has shown that the distribution of $N \times N$ sub-matrices of $M \times M$ COE matrices for $M \in o(\sqrt{N}/\log N)$ asymptotically approaches the distribution of matrices XX^T, where $X \sim \mathcal{G}_{N,M}(0, 1/M)$.

Finally, there is the intermediate regime in which $M^{1/5} \lesssim K < M$. This regime interpolates between the two extreme regimes of very small, square sub-matrices of U and very short, fat sub-matrices of U. A priori, there is no reason to believe why the behaviour should differ from the extreme regimes. Indeed, for this regime we can provide numerical evidence for both parts of Conjecture 6.

We do so by comparing the singular-value spectra of matrices drawn according to $\text{COE}_{N,K}^{M}$ and $\mathcal{G}_{N,K}^{\text{sym}}(0, 1/M)$, respectively. Since both distributions $\text{COE}_{N,K}^{M}$ and $\mathcal{G}_{N,K}^{\text{sym}}(0, 1/M)$ over complex, symmetric $N \times N$ matrices are invariant under conjugation with $V \cdot V^T$ for any $N \times N$ unitary matrix V, the probability of drawing a particular matrix C from these distributions depends only on the singular values of the matrix C. Consequently, the distribution of singular values captures the essence of both distributions alike. Let $P(r)$ denote this distribution, that is, the distribution over singular values r of a matrix C drawn either from $\text{COE}_{N,K}^{M}$ and $\mathcal{G}_{N,K}^{\text{sym}}(0, 1/M)$.

In Fig. S1(a), we show the finite approximation to the distribution $P(r)$ for both ensembles under consideration for fixed values of M, K, N. While the distributions differ (as expected for any finite matrix size), they are already very close to each other for reasonably small matrices. In Figs. S1(b) and (c), we then further investigate the scaling of the total-variation distance between finite-bin approximations of $P(r)$ for $\text{COE}_{N,K}^{M}$ and $\mathcal{G}_{N,K}^{\text{sym}}(0, 1/M)$ in the size of the sub-matrices. In Fig. S1(b), we consider the scaling of the total-variation distance in the short side N of $N \times M$ sub-matrices, i.e., for the second regime where $K = M$. As expected, the total-variation distance increases with N but decreases as the relative size of N to M decreases, too. This provides evidence that the rigorous result about the asymptotic convergence of $\text{COE}_{N,K}$ and $\mathcal{G}_{N,K}^{\text{sym}}(0, 1/M)$ for $K = M$ due to Jiang (*22*) can be strengthened to an inverse polynomial total-variation distance bound (Conjecture 6.2). Finally, in Fig. S1(c), we show that the size of the long side K of the sub-matrices does not significantly affect the total-variation distance in the regime of $N \ll M$ (the collision-free regime. This constitute evidence that the value of

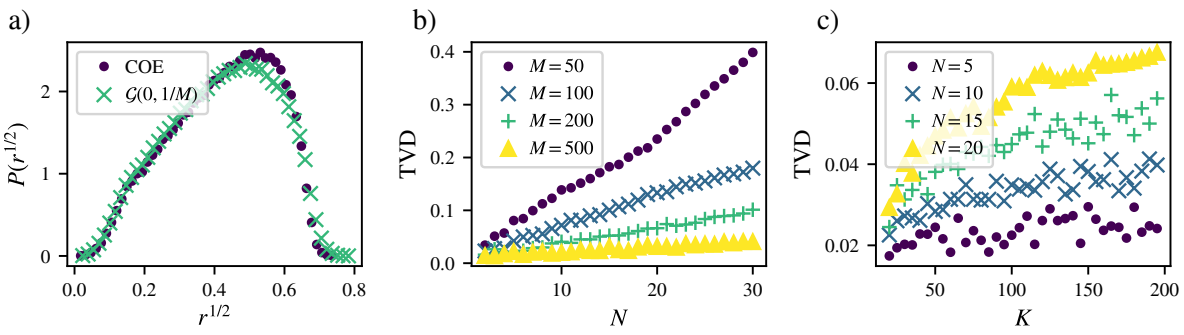

Figure S1: **Numerical evidence that the ensembles $\mathrm{COE}^{M}_{N,K}$ and $\mathcal{G}^{\mathrm{sym}}_{N,K}(0, 1/M)$ converge in total-variation distance for any $K \geq N$ so long as $N \in o(\sqrt{M})$.** a) The singular-value spectra of $\mathrm{COE}^{M}_{N,K}$ and $\mathcal{G}^{\mathrm{sym}}_{N,K}(0, 1/M)$ for $M = 200$, $K = 200$, and $N = 10$. b) Total variation distance between singular-value spectra of $\mathrm{COE}^{M}_{N,K}$ and $\mathcal{G}^{\mathrm{sym}}_{N,K}(0, 1/M)$ for different $M = K$ as a function of N. c) Total variation distance between singular-value spectra of $\mathrm{COE}^{M}_{N,K}$ and $\mathcal{G}^{\mathrm{sym}}_{N,K}(0, 1/M)$ for $M = 200$ and different N as a function of K.

$K \geq N$ does not make a significant difference to the closeness of the distributions $\mathrm{COE}_{N,K}$ and $\mathcal{G}^{\mathrm{sym}}_{N,K}(0, 1/M)$ of symmetric matrix products.

To summarize this section, we have formulated an interesting conjecture regarding the distribution of symmetric products of sub-matrices of Haar-random unitaries. In the main text, we argued that this conjecture captures the hiding property for Gaussian boson sampling. Here, we have provided analytical evidence for the conjecture in the two extremal regimes of $K = N$ (where we know both parts to be true) and $K = M$ (where we know part 1. to be true). We then provided numerical evidence for an inverse polynomial total-variation distance bound for any value of K such that $N \leq K \leq M$.

Let us note that – as in the case of standard boson sampling – our conjecture does not apply to the case in which $N \in \Omega(\sqrt{M})$. Indeed, for the case of $N = K \in \Omega(\sqrt{M})$ Ref. (*22*) shows that $\mathcal{H}^{M}_{N,N}$ and $\mathcal{G}_{N,N}(0, 1/M)$ are far from each other in total-variation distance. This indicates that the statement of our conjecture does not hold in this case since there is no 'short side' of $U_{\mathbf{n},1_K}$.

Average-case hardness of computing GBS output probabilities

In this section, we show average-case hardness of computing GBS output probabilities. As explained in the main text, this amounts to showing that the following problem is #P-hard.

(δ, ϵ)-SQUARED-HAFNIANS-OF-GAUSSIANS

Input A matrix XX^T with $X \sim \mathcal{G}_{N,K}(0, 1/M)$.

Output $\left|\mathrm{Haf}(XX^T)\right|^2$ to additive error ϵ, with probability $\geq \delta$ over the distribution $\mathcal{G}_{N,K}(0, 1/M)$.

The proof will proceed in two steps: First, we will show that an oracle for the (δ, ϵ)-SQUARED-HAFNIANS-OF-GAUSSIANS problem allows one to approximate $\left|\mathrm{Haf}(YY^T)\right|^2$ for arbitrary $Y \in \mathbb{C}^{2N\times 2K}$. This first part of the proof constitutes the worst-to-average-case reduction. Second, we will show that approximating $\left|\mathrm{Haf}(YY^T)\right|^2$ for arbitrary $Y \in \mathbb{C}^{2N\times 2K}$ is actually #P-hard in the worst-case. We show this by reducing the task of approximating the permanent of an arbitrary complex $N \times N$ matrix to the task of approximating $\left|\mathrm{Haf}(YY^T)\right|^2$.

Worst-case hardness Consider the following problem:

ϵ-SQUARED-HAFNIANS

Input A matrix YY^T with $Y \in \mathbb{C}^{N\times K}$ for $K \in \mathbb{N}$, $N \in 2\mathbb{N}$ such that the entries of Y are of the form $(x + iy)/\sqrt{M}$ for $|x|, |y|$ some $O(1)$-bounded integers and additive-error tolerance $\epsilon > 0$.

Output An estimate h s.t. $\left|h - \left|\mathrm{Haf}(YY^T)\right|^2\right| \leq \epsilon$.

We prove the following Lemma.

Lemma 2. *The problem* ϵ-SQUARED-HAFNIANS *is worst-case* #P*-hard for any additive error* $\epsilon \leq 1/(2M^N)$.

Proof. Without loss of generality, we restrict to $N \leq K$. We begin the proof by noting that the permanent of any square matrix G can be expressed as the Hafnian of a corresponding block matrix twice the size of G (*11*),

$$\mathrm{Per}(G) = \mathrm{Haf}\left[\begin{pmatrix} 0 & G \\ G^T & 0 \end{pmatrix}\right].$$

Hence, computing the squared permanent of any complex $N/2 \times N/2$ matrix $G \in \mathbb{C}^{N/2\times N/2}$ reduces to computing the squared Hafnian of a corresponding block matrix

$$B(G) = \begin{pmatrix} 0 & G \\ G^T & 0 \end{pmatrix}. \tag{S3}$$

Computing the squared permanent exactly is known to be worst-case #P-hard even over $0/1$-matrices (*7, 21*).

Next we note that any matrix $B(G)$ for $G \in \mathbb{C}^{N/2\times N/2}$ can be decomposed as XX^T in terms of some complex matrix $X \in \mathbb{C}^{N\times K}$. Indeed the block matrix $B(G)$ is a complex, symmetric matrix, so we can decompose it using the Takagi decomposition as WDW^T, where

$W \in U(N)$ is a unitary matrix and $D \in \mathbb{R}^{N\times N}$ is a nonnegative diagonal matrix. We now define $X' = (WD^{1/2})$ and X by appending $(K-N)$ all-0-columns to X'. This gives rise to a decomposition of $B(G) = XX^T$ with $X \in \mathbb{C}^{N\times K}$. Hence it is #P-hard to exactly compute the Hafnian of matrices of the form XX^T in the worst case. Additionally, since the Hafnian is a continuous function, we can compute $\mathrm{Haf}(XX^T)$ to an arbitrary level of precision by considering $\mathrm{Haf}(YY^T)$ with the entries of Y being of the form $x + iy$, with x and y integers (by suitably rescaling the entries of the matrix). Finally, we note that by normalization we can assume that the entries of the matrix Y are of the form $(x + iy)/\sqrt{M}$ with x and y $O(1)$ bounded integers. Then the squared Hafnian of YY^T is an integer multiple of $1/M^N$. Therefore, computing the Hafnian of YY^T up to additive error of $1/(2M^N)$ serves to compute the squared Hafnian of $B(G)$ exactly, which is #P-hard. This concludes the proof.

The proof holds equally for $N \in \mathsf{poly}(K)$: in this case we embed a square matrix in $\mathbb{C}^{K\times K}$ and append 0 rows instead of columns. □

Worst-to-average equivalence We now prove the average-case hardness of computing GBS output probabilities. That is, we prove the following Lemma:

Theorem 3 (Theorem 3 restated)**.** *The* (δ, ϵ)-SQUARED-HAFNIANS-OF-GAUSSIANS *problem is* #P*-hard under* PH *reductions for any* $\epsilon \leq O\left(\exp[-6N \log N - \Omega(N)]\right)$ *and any constant* $\delta > 3/4$.

We first sketch the proof idea and elaborate on the technique used. The overall idea is to give a worst-to-average-case reduction from the problem ϵ-SQUARED-HAFNIANS to the problem (δ, ϵ)-SQUARED-HAFNIANS-OF-GAUSSIANS. The worst-case #P-hardness of problem ϵ-SQUARED-HAFNIANS has already been established.

We use the same technique as Refs. (*7, 15*) to establish this reduction. Assume that we are given an oracle O that solves (δ, ϵ)-SQUARED-HAFNIANS-OF-GAUSSIANS, meaning that with probability at least δ over the input X, it outputs a squared Hafnian of XX^T to additive error ϵ. The rest of the time, it may output an incorrect value, with no guarantees whatsoever on how close the output is to the desired output. In the following, we will show how to use the oracle O to obtain the squared Hafnian of an arbitrary worst-case matrix YY^T with high probability (this latter probability is over the choice of the random variables instantiated in the algorithm).

The key idea is that for $X \in \mathbb{C}^{N\times K}$, the quantity $|\mathrm{Haf}(XX^T)|^2$ is a degree $2N$ polynomial over the entries of the matrix X. This allows for the use of polynomial interpolation to recover the squared Hafnian of an arbitrary worst-case matrix YY^T. An important technique we use in this proof is the robust Berlekamp-Welch algorithm due to Ref. (*15*), which is important for polynomial interpolation over $\mathbb{R}$ as opposed to a finite field. Polynomial interpolation over the reals is a technique often used for the problem of average-case hardness of computing output probabilities of random quantum circuits (*14, 26*). The Berlekamp-Welch algorithm cannot be used as is for the reals, and therefore, recent works (*14, 26*) use techniques like Lagrange interpolation. The new robust Berlekamp-Welch algorithm of Ref. (*15*) allows for improved robustness of the worst-to-average-case reduction.

As an example, in the context of random quantum circuits over n qubits and m gates, Lagrange interpolation can only give average-case #P-hardness of computing output probabilities to error $2^{-O(m^3)}$ rather than the $O(2^{-n})$ that suffices for proving the hardness of approximate sampling (see (*7, 26*)). The modified Berlekamp-Welch algorithm of Ref. (*15*), which is boosted with an NP oracle, can sidestep the need for Lagrange interpolation and obtain average-case #P-hardness with $2^{-O(m \log m)}$ error (see also, the recent work of Kondo *et al*. (*25*) which also obtains this robustness error).

Theorem 4 (Robust Berlekamp-Welch algorithm (*15*))**.** *Let p be a univariate polynomial of degree d over the reals. Suppose that we have $k \geq 100d^2$ points (x_i, y_i), with $\{x_i\}$ uniformly spaced in the interval $[0, \kappa]$ and obeying the promise*

$$\Pr[|y_i - p(x_i)| \geq \Delta] \leq \eta < \frac{1}{4}.$$

Then there is a P^{NP} algorithm that can estimate $p(1)$ to additive error $\Delta \exp[d \log \kappa^{-1} + O(d)]$ with probability at least 2/3.

Proof of Theorem 3. The polynomial interpolation procedure is as follows. Let $X(t)$ be the matrix obtained by drawing a random $X \sim \mathcal{G}_{N,K}(0, 1/M)$ and setting

$$X(t) := (1 - t)X + tY,$$

where Y is the matrix corresponding to the worst-case instance. Now, the quantity

$$p(t) := |\mathrm{Haf}(X(t)X^T(t))|^2$$

is a polynomial of degree $2N$ over the entries of $X(t)$, and consequently, over t itself. For t close to 0, $X(t)$ is close to Gaussian distributed, while when t is close to 1, the distribution is close to being deterministic. We select k points in the range $[0, \kappa]$ and query the oracle O for the value of $p(t)$ for these points. By the promise, the oracle outputs the correct value of $p(t)$ for most values of t with high probability. Conditioned on this event, the robust Berlekamp-Welch algorithm stated in Theorem 9 allows one to reconstruct the polynomial in the second level of the polynomial hierarchy. The polynomial can then be evaluated at the point $t = 1$ to obtain an estimate of the squared Hafnian of the worst-case matrix YY^T.

We now check that the conditions of Theorem 9 are met. We say that a call to the oracle O is successful if it outputs the squared Hafnian of a matrix to additive error ϵ. By assumption, for X drawn at random from $\mathcal{G}_{N,K}(0, 1/M)$, the oracle is successful with probability at least δ. Note however that the matrix $X(t) = (1 - t)X + tY$ is not exactly distributed according to $\mathcal{G}_{N,K}(0, 1/M)$. Instead, for small t, due to the rescaling by $(1 - t)$ and the shift by tY, $X(t)$ is distributed according to a slightly different distribution $\mathcal{G}'$. If we query the oracle for the value of $p(t)$ with matrices drawn from this different distribution $\mathcal{G}'$, the probability of success can, in the worst case, decrease. By definition, the success probability can decrease at most by the variation distance between the two distributions $\mathcal{G}_{N,K}(0, 1/M)$ and $\mathcal{G}'$, which is

$O(t\max(N,K)^2)$. Therefore, for $K \geq N$, the probability of success is at least $\delta - O(\kappa K^2)$. We choose κ to be $O(c/K^2)$ with some small enough c so that the success probability is at least $\delta - O(c) > 3/4$. This ensures that the conditions of the theorem are met.

We finally conclude by examining the additive error to which we can compute, using the $\mathsf{BPP}^{\mathsf{NP}}$ reduction, the squared Hafnian of the worst-case matrix YY^T. If the additive error for successful queries to the oracle is at most ϵ, Theorem 9 implies that the error in computing $p(1)$ is $\epsilon \exp[d \log \kappa^{-1} + O(d)]$. Plugging in $d = 2N$ and $\kappa = c/N^2$, we get the total additive error in estimating $p(1)$ to be $\epsilon \exp[4N \log N + O(N)]$. Finally, we note that the squared Hafnian is shown to be worst-case hard for additive error $O(1/M^N)$. Therefore, we make the choice

$$\epsilon \exp[4N \log N + O(N)] = O\left(\frac{1}{M^N}\right), \tag{S4}$$

or

$$\epsilon = O\left(\exp[-4N \log N - \Omega(N) - 2N \log N]\right) = O\left(\exp[-6N \log N - \Omega(N)]\right),$$

where we have assumed $M = \Theta(N^2)$. This choice ensures that we can, with probability at least 2/3, compute the squared Hafnian of an arbitrary matrix with bounded entries of the form YY^T to additive error $O(1/M^N)$. As shown in Lemma 7, this task is $\#\mathsf{P}$-hard. This completes our proof. □

Average-case hardness of computing noisy GBS output probabilities We argue here that computing the output probabilities for a *noisy* random GBS experiment is $\#\mathsf{P}$-hard on average. That is, we show the following lemma.

Lemma 5. *There exists a polynomial $p(N)$ and a loss threshold η_* such that (ϵ, η)-*NOISYGBS-PROBABILITY *with $\eta \leq \eta_*$, $\delta > 3/4$, and $\epsilon \leq 2^{-p(N)}$ is $\#\mathsf{P}$-hard under PH reductions.*

Proof. For worst-case hardness despite the presence of noise, we follow the proof technique in Refs. (*15, 27*). At a high level, the worst-case hardness follows from the error-detection property of the system. In particular, the error-detection property implies that as long as the noise η is smaller than a certain threshold η_*, there is a fixed outcome on a subset of the modes, say $\mathbf{m}$, such that conditioned on this outcome, the probability distribution on the rest of the modes is exponentially close to the target noiseless distribution. In other words, we have

$$\left| \Pr_{\text{noisy}}[\mathbf{n}|\mathbf{m}] - \Pr_{\text{ideal}}[\mathbf{n}] \right| \leq 2^{-\mathsf{poly}(N)}$$

for any desired polynomial on the right hand side. Since $\Pr_{\text{ideal}}[\mathbf{n}]$ is $\#\mathsf{P}$-hard to approximate in the worst case by virtue of Lemma 7, so is computing the conditional probability

$$\Pr_{\text{noisy}}[\mathbf{n}|\mathbf{m}] = \frac{\Pr_{\text{noisy}}[\mathbf{n}, \mathbf{m}]}{\Pr_{\text{noisy}}[\mathbf{m}]}.$$

The denominator here is the probability of seeing the outcome $\mathbf{m}$, which flags the no-error event. The probability of this can be exponentially small, and satisfies (*15, 27*)

$$\left| \Pr_{\text{noisy}}[\mathbf{m}] - (1-\eta)^{O(Md)} \right| \leq \Pr_{\text{noisy}}[\mathbf{m}] 2^{-\mathsf{poly}(N)},$$

where η is the maximum noise parameter as defined earlier in the main text. In other words, for an error-detected circuit, the probability that the outcome on the subset of heralding modes is in the state $\mathbf{m}$ is exponentially close to the probability that no error occurred, which is given by $(1-\eta)^{O(Md)}$.

Therefore, approximating $\Pr_{\text{noisy}}[\mathbf{n}, \mathbf{m}]$ is also #P-hard:

$$\left| \Pr_{\text{noisy}}[\mathbf{n}, \mathbf{m}] - \Pr_{\text{noisy}}[\mathbf{n}|\mathbf{m}](1-\eta)^{O(Md)} \right| \leq \Pr_{\text{noisy}}[\mathbf{n}, \mathbf{m}] 2^{-\mathsf{poly}(N)} \tag{S5}$$

$$\Rightarrow \left| \Pr_{\text{noisy}}[\mathbf{n}, \mathbf{m}] - \Pr_{\text{ideal}}[\mathbf{n}](1-\eta)^{O(Md)} \right| \leq \Pr_{\text{noisy}}[\mathbf{n}, \mathbf{m}] 2^{-\mathsf{poly}(N)} + 2^{-\mathsf{poly}(N)}.$$

Since computing $\Pr_{\text{ideal}}[\mathbf{n}]$ to additive error $\pm O(2^{-\mathsf{poly}(N)})$ is #P-hard, so is computing $\Pr_{\text{noisy}}[\mathbf{n}, \mathbf{m}]$ to additive error $O(2^{-\mathsf{poly}(N)})(1-\eta)^{O(Md)}$. A similar analysis in Ref. (*15*) shows that it is $\mathsf{coC_{=}P}$-hard to compute a noisy probability in the worst case to additive error $2^{-O(m \log m)}$ in the context of RCS. This proves the worst-case hardness.

For the worst-to-average-case reduction, we again use the technique of polynomial interpolation in conjunction with a robust Berlekamp-Welch algorithm. We observe that any noisy output probability for a local noise model can still be written as a polynomial in the gate entries of the circuit, using the Feynman sum-over-paths idea. As before, we perform interpolation from a random instance from the ensemble to the worst-case-hard instance. This is achieved now using the Cayley path interpolation technique of Ref. (*26*) instead of the direct interpolation between two matrices. This is because the noisy output probability is no longer a simple function of only the linear-optical unitary (like the Hafnian), but is also a function of the circuit implementation. The full interpolation involves interpolating every gate of a circuit implementation from the average-case instance A_i to the worst-case instance W_i along the Cayley path

$$C_i(t) = \left(t\mathbb{1} + (2-t)A_i W_i^{-1}\right)\left((2-t)\mathbb{1} + tA_i W_i^{-1}\right)^{-1} \cdot W_i,$$

which satisfies $C_i(0) = A_i$ and $C_i(1) = W_i$. Using this interpolation and the fact that any local noise can be "purified" gate-wise by introducing ancillary systems of finite dimension, we can again write the noisy probability $\Pr_{\text{noisy}}[\mathbf{n}, \mathbf{m}][t]$ as a polynomial in t. The rest of the proof follows from before. □

Average-case hardness of computing noisy probabilities in high-dimensional GBS For the worst-case hardness of computing noisy probabilities of the high-dimensional GBS architecture, we mainly use the previous results on error-detection of noise. The additional ingredient used

is the fact that a constant-depth linear-optical architecture in two dimensions (and higher) has been shown by Brod (*45*) to be hard to exactly sample from.

The proof of Ref. (*45*) uses post-selection to argue for exact sampling hardness. Note that the post-selection result does not, by itself, imply the #P-hardness of computing output probabilities: it implies the PP-hardness of strong simulation, which involves computing both the output probabilities and the marginals. However, we note that the post-selection proof can often be "opened up" in order to directly argue about the hardness of computing output probabilities. This is done by giving an amplitude-preserving reduction from a BQP circuit to the circuit family in question (here, high-dimensional GBS). Since computing output amplitudes of BQP circuits is #P-hard, so is computing that of the circuit family in question. Using the results from earlier, so is computing the *noisy* output probability in the worst case for an error-detected circuit as long as the noise level is smaller than some (constant) threshold η_*.

The average case hardness again essentially follows by observing that there is a polynomial structure in the output probability, to prove Theorem 5. We again use the Cayley technique of Ref. (*26*) to set up the polynomial interpolation in this case, and use results from Ref. (*15*) to strengthen it, such as using a variable rescaling and applying a robust version of the Berlekamp-Welch algorithm (Theorem 9).

Total photon number distribution

For pure state GBS, the total photon number distribution can be obtained efficiently by simply convolving the photon number distributions of the individual modes going into the interferometer (*58*). In the case where M identical squeezed states (with squeezing parameter r) are sent into an interferometer and undergo uniform loss by transmission parameter η, the probability of obtaining n photons is given by

$$\Pr(n) = \begin{cases} \eta^n \binom{\frac{M}{2}+\frac{n}{2}-1}{\frac{n}{2}} \operatorname{sech}^M r \tanh^n(r) {}_2F_1\left(\frac{n}{2}+\frac{1}{2}, \frac{M}{2}+\frac{n}{2}; \frac{1}{2}; (1-\eta)^2 \tanh^2 r\right) \text{ if } n \text{ is even,} \\ (1-\eta)(n+1)\eta^n \binom{(M+n-1)/2}{(n+1)/2} \operatorname{sech}^M r \times \\ \tanh^{n+1}(r) {}_2F_1\left(\frac{n+2}{2}, \frac{1}{2}(M+n+1); \frac{3}{2}; (1-\eta)^2 \tanh^2 r\right) \text{ if odd.} \end{cases}$$

where ${}_2F_1(a,b,c;z)$ is a hypergeometric function. This equation reduces to the well-known lossless limit (*10*) when $\eta \to 1$, since in that case ${}_2F_1\left(\frac{n}{2}+\frac{1}{2}, \frac{M}{2}+\frac{n}{2}; \frac{1}{2}; 0\right) = 1$ and the probabilities for all odd photon numbers become zero since they are proportional to $1-\eta$. This distribution has the following moments

$$\mathbb{E}(n) = \eta M \sinh^2 r, \text{Var}(n) = \eta M \sinh^2 r[1+\eta(1+2\sinh^2 r)].$$

Note that even if the losses are not uniform, one can still calculate in polynomial time the moments of the random variable n (*52, 58*).