

Exponential improvements to the average-case hardness of BosonSampling

Abstract

Establishing quantum advantage is a central problem in the theory of quantum computing. Average-case sampling problems like BosonSampling and Random Circuit Sampling have risen to prominence, both as a theoretical tool for separating quantum and classical computation, and as an experimental means of demonstrating quantum speedups. While the worst-case hardness of these sampling problems was established 15 years ago, average-case hardness of sampling remains open for all quantum advantage proposals. This is because the prior works show that average-case hardness of sampling can be derived from certain unproven conjectures about the hardness of computing output probabilities, such as the Permanent-of-Gaussians Conjecture (PGC), which states that $e^{-n \log n - n - O(\log n)}$ additive-error estimates to the output probability of most random BosonSampling experiments are $\#P$ -hard. Prior works have shown weaker forms of average-case hardness for these problems, but these do not imply average-case sampling statements. Proving these conjectures has since become a central question in the theory of quantum advantage.

In this work we make substantial progress towards proving these conjectures. In particular, we show that $e^{-n \log n - n - O(n^\delta)}$ additive-error estimates to output probabilities of most random BosonSampling experiments are $\#P$ -hard, for any constant $\delta > 0$, exponentially improving on prior work. In the process, we circumvent all known barrier results for proving PGC. This is nearly the robustness needed to prove PGC—the remaining hurdle is now “merely” to show that the $O(n^\delta)$ term in the exponent can be improved to $O(\log n)$. We also obtain an analogous result for Random Circuit Sampling.

We then show, for the first time, a hardness of average-case classical *sampling* result for BosonSampling, under an anticoncentration conjecture. Specifically, we prove the impossibility of multiplicative-error sampling from random BosonSampling experiments with probability $1 - 2^{-\tilde{O}(N^{1/3})}$ for input size N , unless the Polynomial Hierarchy collapses. This exponentially improves upon the state-of-the-art. To do this, we introduce new proof techniques which tolerate exponential loss in the worst-to-average-case reduction. This opens the possibility to show hardness of average-case sampling without ever proving PGC.

Contents

1	Introduction	1
1.1	Our results	3
1.2	Proof techniques	4
1.2.1	What controls robustness in the standard worst-to-average-case reduction? .	4
1.2.2	Coefficient extraction: a new way to encode the permanent	6
1.2.3	Overcoming the convexity barrier: square method and magnification lemma .	7
1.2.4	Average-case hardness of sampling using random rare events lemmas	9
1.3	Discussion and open problems	12
2	Background	12
3	Techniques	14
3.1	Coefficient extraction	14
3.2	Square method	15
3.3	Worst case magnification	16
3.4	Rare events lemma I: going out of distribution	19
3.5	Rare events lemma II: tail probabilities for orthogonal submatrices and i.i.d. Gaussians	20
4	Hardness of computing output probabilities, Theorem 1	21
5	Hardness of sampling, Theorem 3	22
A	How far can you shift and scale i.i.d. Gaussian matrices?	26
B	Corollaries for Random Circuit Sampling	26
C	Rare events for orthogonal submatrices and i.i.d. Gaussians	29
D	Numerical evidence for Conjecture 6	35
E	Square method for complex polynomials	36

1 Introduction

What makes quantum mechanics hard to simulate classically? This has been the central question of quantum computation since it was first proposed [Ben80, Man80, Fey82]. The need to answer this question has been made even more urgent by recent experiments claiming to solve certain problems much faster than by any classical computer [Aru19, Mor24, Wu21, Zho20, Zho21, Den23, YGE⁺24, Mad22, DHL⁺24]. These claims represent the first *experimental* violations of the Extended Church-Turing Thesis, the belief that all physically realizable models of computation are efficiently simulable by randomized Turing machines, and thus deserves careful scrutiny.

While the physics of these experiments differs dramatically, from a computational standpoint they all solve random sampling problems that have three components: (i) initialize a fiducial starting state (such as $|0^n\rangle$), (ii) evolve by a random quantum circuit drawn from some particular distribution, and (iii) measure to draw a sample. Seminal results due to Terhal and DiVincenzo [TD04] and concurrent works of Aaronson and Arkhipov [AA13] and Bremner, Jozsa, and Shepherd [BJS10] gave evidence that for these distributional sampling problems, even sub-universal quantum computation can outperform any efficient classical algorithm in the worst case. Fascinatingly this only assumes the non-collapse of the PH.

However, the shortcoming of these statements is that they are *brittle*, pertaining only to exact sampling in the worst case. Thus the important open problem is to make these separations more robust, so as to close the gaps between theory and experiment.

Foremost among these gaps is to prove classical hardness of sampling from an *average-case* experiment, i.e., to extend the previous worst-case sampling results to prove that sampling is hard for a *randomly chosen* or typical circuit from a given ensemble. This random choice of circuit is crucial in quantum advantage experiments. For one thing, such randomness gives a hard candidate distribution to test with a quantum device. Additionally, randomness plays an important role in classical verification, e.g., by benchmarking tests such as Linear Cross-Entropy which make use of specific properties of random circuits. It has also been shown that average-case hardness allows one to derive cryptographic primitives (see e.g., [KT24]). Thus for reasons both practical and fundamental, it behooves us to study the average-case hardness of sampling—which, in contrast to the worst case, remains an *open problem* for *all* quantum advantage proposals.

Why should sampling from random quantum circuits be intractable for classical computers? The first evidence for this came from Aaronson and Arkhipov, in two parts. First, they showed a reduction from the problem of *sampling* from random circuits to the problem of *approximately computing* output probabilities of random circuits, via Stockmeyer counting [Sto83]. Second, they showed that computing an output probability of a random BosonSampling circuit is just as hard as in the worst case, i.e. #P-hard, using the connection between bosons and the matrix permanent. This was extended to random circuits on qubits [BFNV19] and subsequently improved and generalized [Mov23, HHB⁺19]. However, these two parts do not connect with one another to establish hardness of sampling. The key issue is that existing average-case hardness of computing proofs are not *error tolerant* enough to prove hardness of sampling. That is, to show hardness of sampling we want to show it is #P-hard to estimate output probabilities to additive error ϵ , but so far we have only proven it is #P-hard to estimate them to additive error $\epsilon' \ll \epsilon$. We call this gap between ϵ and ϵ' the “robustness gap”, and it remains open for all quantum advantage proposals. Hardness of average-case sampling has therefore only been established under unproven conjectures.

Consequently, the focus of this work is to close this robustness gap. While our results are broadly applicable to many random sampling experiments, we will primarily focus on BosonSampling. The goal is to prove the *Permanent-of-Gaussians Conjecture* (PGC), the statement that the following problem, known as Gaussian Permanent Estimation, GPE_{\pm} , is #P-hard: estimate the output

probability of a random BosonSampling experiment to within additive error $\pm e^{-n \log n - n - O(\log n)}$. Aaronson and Arkhipov showed this conjecture suffices to show hardness of sampling from average-case BosonSampling experiments [AA13].¹

In the last decade, progress has been made toward proving PGC [AA13, BFL22, Kro23]. While Aaronson and Arkhipov’s initial work showed computing additive error estimates of $e^{-O(n^4)}$ to the output probability of most BosonSampling experiments is #P-hard [AB16], this error tolerance was subsequently improved to $e^{-6n \log n - O(n)}$ by Bouland, Fefferman, Landau, and Liu [BFL22], and then to $e^{-4n \log n - O(n)}$ in unpublished work of Krovi [Kro23]. Therefore the remaining gap to establish the hardness of BosonSampling is to improve the robustness of this result by a constant factor in the exponent. This seems tantalizingly “close” to the target in additive terms yet *exponentially far away* in relative terms. We note the analogous conjectures for all other quantum advantage experiments remain open as well, such as Random Circuit Sampling [BIS⁺18], despite much progress in the area [BFNV19, Mov23, BFL22, KMM22, ODMZ22, Kro22].

Why has it been so difficult to improve the robustness of output probability estimation and prove the classical hardness of BosonSampling or any other quantum advantage experiment? One of the major reasons is that there are a number of proof barriers that have been identified, indicating that to prove hardness of sampling, new techniques are required:²

- *Convexity barrier.* Noted in [AA13], the basic idea is that worst-to-average-case reductions for the permanent are based on polynomial extrapolation, following Lipton’s proof [Lip91]. Polynomial extrapolation is in general exponentially ill-conditioned, i.e. an error γ in a degree- d polynomial $p(t)$ near $t = 0$ becomes error $\sim 2^d \gamma$ near $t = 1$. Moreover, one can show this is necessary even for the set of polynomials corresponding to valid matrix permanents, which is a convex set. Thus any worst-to-average case reduction for the permanent based on polynomial extrapolation will introduce exponential relative error. A special case was referred to as the “noise barrier” of [BFL22].
- *“Jerrum-Sinclair-Vigoda” barrier for BosonSampling.* This barrier is inspired by a landmark result of [JSV04] giving an efficient classical algorithm to estimate the permanent of a non-negative matrix to $1/\text{poly}(n)$ relative error. This algorithm tells us that any technique used to prove PGC must fundamentally make use of the fact that i.i.d. Gaussian matrices have negative as well as positive entries. By contrast, all existing worst-to-average-case reductions for Gaussian permanents work equally well for permanents of nonnegative matrices, and therefore cannot possibly prove PGC. In other words, to show hardness of sampling, we will need a proof which uses a special property of matrices with negative entries that does not hold for nonnegative matrices, such as like multiplicative hardness in the worst case.
- *Depth and “Born-rule” barriers for Random Circuit Sampling.* [NLPD⁺22] gives a classical algorithm that approximately samples from the output distribution of a particular ensemble of constant depth RCS experiments. On the other hand, the existing techniques for proving hardness of computing output probabilities work with respect to circuits of any depth. Therefore, if we are to prove hardness of sampling, we need to find a proof technique that is sensitive to circuit depth and only works to prove hardness for sufficiently deep circuits.

The “Born-rule” barrier identified by Krovi [Kro22] is that the additive error needed to prove the hardness of average-case sampling (2^{-n}) is already larger than the additive error known

¹We note [AA13] also conjecture a certain “flatness” property about the output distribution known as anticoncentration (which we also assume in this work). This allows them to convert average-case additive estimates to relative-error estimates.

to be hard in the worst case (2^{-2n} , which is derived from the Born rule by squaring the output amplitude of a Quantum Fourier Sampling circuit). How can we ever hope to prove a worst-to-average case reduction in which the additive error in the average case *is larger* than the additive error we need to obtain in the worst case?

There has also been work aiming to falsify variants of PGC. For example, Eldar and Mhriban showed there is a quasipolynomial-time classical algorithm to multiplicatively estimate random Gaussian permanents if the means are non-zero but asymptotically slowly vanishing, despite being $\#P$ -hard to compute exactly on average [EM18]. Thus the dividing line between classically easy and classically hard is very narrow—making it yet more difficult to furnish a proof of PGC.

1.1 Our results

In this work we introduce a new suite of tools which allow us to exponentially improve on the state-of-the-art hardness results for BosonSampling. In particular, we invent new techniques that overcome *all of the barriers* described above.

Our first result makes progress towards proving the Permanent-of-Gaussians Conjecture (PGC). We show a new worst-to-average-case reduction for computing Gaussian permanents whose additive error tolerance exponentially improves on the state-of-the-art. Our error tolerance for the first time matches to leading order that of the Permanent-of-Gaussians Conjecture (PGC).

Theorem 1 (Hardness of computing output probabilities). *For any $\delta > 0$, it is $\#P$ -hard under a BPP^{NP} reduction to approximate output probabilities of an n -photon, $O(n^2)$ -mode BosonSampling experiment to additive error $\exp(-n \log n - n - O(n^\delta))$ with success probability at least $2/3$, assuming the Permanent Anticoncentration Conjecture 9.*

This is nearly the additive error tolerance needed to prove PGC, $e^{-n \log n - n - O(\log n)}$. In particular, all that remains is “merely” to improve the $O(n^\delta)$ term in the exponent to $O(\log n)$. In order to prove this result, we give a new worst-to-average-case reduction for BosonSampling which replaces polynomial extrapolation with polynomial *coefficient extraction*. This allows us to use a technique we call “*dilution*” to lessen the degree of the polynomial involved and hence reduce the error blowup of the worst-to-average-case reduction.

Crucially, our proof surpasses the Jerrum-Sinclair-Vigoda barrier as it requires that the worst-case matrix contain both positive and negative entries. This is because our worst-to-average-case reduction derives the worst-case permanent value to within small relative error, which is only $\#P$ -hard with mixed signs. This is an essential feature of any proof that might solve PGC, and a feature which was missing from all prior proofs of average-case hardness for the permanent [AA13, BFL22, Kro23].

We also show this idea can be ported to other quantum advantage experiments, like RCS:

Corollary 2. *For any $\delta > 0$, it is $\#P$ -hard to approximate the output probabilities of n -qubit Random Circuit Sampling experiments of $\Omega(\log n)$ depth to additive error $2^{-n - O(n^\delta)}$.*

Just as with BosonSampling, this exponentially improves over prior work [BFNV19, Mov23, BFL22, KMM22, Kro22], and obtains hardness which is within an n^δ factor of what is needed for hardness of sampling. Applied to RCS, our techniques overcome the depth barrier by requiring anticoncentration, and the Born-rule barrier by “diluting” the worst-case instance to be polynomially smaller than the average-case instance. We prove Corollary 2 in Sec. B.

²There is also a relativization barrier to proving hardness of average-case sampling to small ℓ_1 error [AC17]. However, here we focus on average-case multiplicative-error sampling (Def. 7) to which no relativization barrier applies.

Our second major contribution is to develop a suite of techniques that taken together allow us to show, for the first time, a hardness of average-case sampling theorem.

Theorem 3 (Hardness of sampling). *There does not exist a multiplicative-error classical sampler (see Def. 7) from the output distribution of an n -photon, $O(n^2)$ -mode real BosonSampling experiment that succeeds with probability at least $1 - \exp(-O(n))$ over the choice of experiment, assuming PH does not collapse and a slight generalization of Permanent Anticoncentration, Conjecture 6.*

This theorem exponentially improves upon the trivial hardness of sampling statement. In particular, if the sampling algorithm succeeds with probability $1 - 2^{-\tilde{O}(n^3)}$, then the algorithm directly estimates the value of the worst case as the input size³ is $\tilde{O}(n^3)$ (see Lemma 19). We note another exponential improvement would be required to show the desired hardness of sampling for $1 - 1/\text{poly}(n)$ fraction of experiments. This is the first hardness result for average-case multiplicative-error sampling. This had been open for all quantum advantage proposals, as prior hardness results for *computing* output probabilities do not imply average-case *sampling* hardness (even for exact sampling) due to the losses in the Stockmeyer reduction from sampling to computing.

In order to show this result, we develop a suite of new techniques that allow us to tolerate an exponential error blowup in the worst-to-average-case reduction, overcoming the convexity barrier of [AA13]. This is achieved by “magnifying” the worst-case permanent value to tolerate more error in the reduction, among other improvements.

This still falls short of proving PGC—the bottleneck is that the average-case algorithm can only compute permanents of matrices which are close in total variation distance to i.i.d. Gaussian, which limits the error tolerance. To overcome this bottleneck, we show that if an average-case algorithm works with sufficiently high probability, then it can also compute permanents “out of distribution” in TV distance. This uses special properties of the Gaussian measure and also requires proving new results in random matrix theory regarding submatrices of Haar-random orthogonals. Our work opens up the possibility that one could prove the classical hardness of sampling, even *without* proving PGC, by improving some of the parameters of these new tools. Interestingly, this result pertains only to real BosonSampling, and extending to the complex case requires solving an open problem in complex analysis—see Appendix E.

1.2 Proof techniques

1.2.1 What controls robustness in the standard worst-to-average-case reduction?

To explain our proof, it is helpful to briefly recall the average-case hardness proofs of [AA13] and its subsequent improvements [BFL22, Kro23]. The basic idea is to use polynomial extrapolation to show the squared permanent is hard to compute on average, following Lipton [Lip91]. Suppose we wish to compute the squared permanent of an arbitrary (worst-case) matrix $W \in \{0, \pm 1\}^{n \times n}$ using only the ability to compute most Gaussian permanents R drawn from $\mathcal{N}(0, 1)^{n \times n}$. We define a univariate family of matrices interpolating between W and a single random choice of Gaussian matrix R :

$$A(t) = (1 - t)R + tW$$

This family has three nice properties that enable the reduction: (i) $|\text{Per}(A(t))|^2$ is a degree $2n$ polynomial in t , (ii) for small values of t , $A(t)$ is close to i.i.d. Gaussian in total variation distance, and (iii) $|\text{Per}(A(1))|^2 = |\text{Per}(W)|^2$. This motivates a worst-to-average-case reduction whereby one computes $|\text{Per}(W)|^2$ by computing the average-case permanents $|\text{Per}(A(t))|^2$ at many small

³Here n is the number of photons, and the input is an $n \times n$ matrix of reals specified to $\tilde{O}(n)$ bits of precision.

values of t , inferring the polynomial in t , and extrapolating it to $t = 1$. This shows that computing average-case permanents, namely estimating the polynomial close to $t = 0$, is as hard as computing a worst-case permanent, the polynomial at $t = 1$.

What controls robustness, i.e. the additive error tolerance, in this worst-to-average-case reduction? In other words, what are the largest error bars we can tolerate on our estimates to the polynomial close to $t = 0$, and how do these errors accrue under polynomial extrapolation?

Polynomial extrapolation is ill-conditioned, in the sense that errors in the values of the polynomial close to $t = 0$ blow up exponentially under extrapolation to $t = 1$. Formally, we can quantify the error blowup using a discrete version of the Remez inequality that we prove in this work.⁴

Lemma 4 (Discrete Remez inequality). *Let $\{x_j\}_{j=0}^d \subset [0, 1]$ be a δ -separated set of points, meaning that $|x_i - x_j| \geq \delta$ for $i \neq j$. Then if p is a degree- d polynomial and $L \geq 1$,*

$$|p(L)| \leq (e^2(\delta d)^{-1}L)^d \max_{0 \leq j \leq d} |p(x_j)|.$$

Lemma 4 has a very simple interpretation. If we take $p(t)$ to be the difference between the true polynomial $|\text{Per}(A(t))|^2$ and the approximate polynomial obtained from estimates of the permanent at $\{t_j\}$, then $\max_{0 \leq j \leq d} |p(t_j)| =: \gamma$ is precisely the robustness, i.e. the maximum additive error tolerance on average-case values of the permanent. Then Lemma 4 tells us that the error blowup $|p(1)|$ is bounded above by $\gamma(\delta d)^{-d}$, which for δ -separated points in the interval $[0, \Delta]$ where $0 < \Delta < 1$ is $\gamma(1/\Delta)^d$. We will refer to Δ as the “box size,” which is determined by the largest value of t such that the total variation distance between $A(t)$ and i.i.d. Gaussian is, say, 0.01. In short, estimating a degree d polynomial to within $\pm\gamma$ at points $t \in [0, \Delta]$ incurs a blowup at $t = 1$ of $\gamma(1/\Delta)^d$.

Posed in this way, we see that to increase the robustness of our worst-to-average-case reduction we need to reduce our effective polynomial degree d or increase the box size Δ over which we estimate average-case values. In both the proofs of [BFL22] and [Kro23] the main improvement was in reducing the distance of extrapolation, while keeping the same degree of polynomial ($2n$ for a squared permanent). In particular in [BFL22] the distance was reduced to $\Delta = O(1/n^2)$ by introducing a robust version of Berlekamp-Welch over the complex numbers.⁵ In [Kro23] the box size was improved to $\Delta = O(1/n)$ by a more sophisticated calculation of the total variation distance between $A(t)$ and Gaussian, which saves a factor of $n^{2n} = e^{2n \log n}$.

⁴The Remez inequality is more commonly shown in a continuous form to bound sup norms of polynomials defined over measurable sets [Rem36]. We discretize the inequality so that it is more natural for a computer science setting and in particular our worst-to-average-case reduction.

⁵We note similar results for BosonSampling could be obtained by the techniques of [KMM22].

Technique	Polynomial degree	Box size Δ
Robust Berlekamp-Welch [BFL22]	$2n$	$1/n^2$
Tighter TVD analysis [Kro23]	$2n$	$1/n$
(This work) Square method, Lemma 13	n	$1/n$
(This work) Rare events lemmas 16 & 17	n	$1/\sqrt{n}$
(This work) Dilution via coefficient extraction, Thm. 1	$n^\delta \forall \text{ const. } \delta > 0$	$1/n^\delta \forall \text{ const. } \delta > 0$

Table 1: Lemma 4 tells us that estimating a degree d polynomial to within $\pm\gamma$ at points $t \in [0, \Delta]$ incurs a blowup at $t = 1$ of $\gamma(1/\Delta)^d$. Our work introduces a suite of techniques, shown here, that decrease d and increase Δ .

In this work, we will introduce a suite of techniques that improve both the polynomial degree d and the “box size” Δ . We tabulate these techniques and their improvements in Table 1.

1.2.2 Coefficient extraction: a new way to encode the permanent

A natural approach to try to improve the robustness of this argument is to reduce the degree of the polynomial involved. A simple observation is that for any $\varepsilon > 0$, it is $\#P$ -hard to compute the permanent of an $n^\varepsilon \times n^\varepsilon$ matrix W as well—this is simply polynomially shrinking the input size. Therefore a natural way to improve the robustness is to try to make W smaller, an idea we henceforth refer to as “*dilution*.” Using standard polynomial extrapolation arguments, this doesn’t yield much progress. That’s because if we set W to have small support—say with only $O(n^\varepsilon)$ nonzero entries—then $\text{Per}(W) = 0$. Trivially, a matrix must have at least n non-zero entries for its permanent to be non-zero. This lower bounds how much one could gain by such arguments using extrapolation, and the best one can obtain by dilution is $e^{-3n \log n - O(n)}$ robustness⁶—which sits right at the convexity barrier.

To get around this obstacle, our first step is to change the worst-to-average-case reduction from a problem about polynomial extrapolation to a problem about polynomial *coefficient extraction*. We consider a one-parameter family of matrices

$$A(t) = R + tW_{\text{dilute}}$$

and consider the case that W_{dilute} consists of a tiny n^ε -sized worst case matrix W' in direct sum with the all 0’s matrix on the remaining $n - n^\varepsilon$ dimensions. The key point of this construction is, even though the value of $|\text{Per}(A(1))|^2$ is not what we want (as $A(1) = R + W_{\text{dilute}}$), the coefficients of the polynomial $|\text{Per}(A(t))|^2$ do encode information about $\text{Per}(W')$. In particular, the degree of the polynomial $|\text{Per}(A(t))|^2$ is now $n^{2\varepsilon}$, and the top coefficient is $|\text{Per}(W')|^2 |\text{Per}(R_D)|^2$, where R_D is the bottom righthand minor of R of dimension $n - n^\varepsilon$ (see Figure 1):

$$|\text{Per}(A(t))|^2 = |\text{Per}(W')|^2 |\text{Per}(R_D)|^2 t^{2n^\varepsilon} + \sum_{\ell=0}^{n^{2\varepsilon}-1} c_\ell t^\ell$$

where the c_ℓ are some other coefficients which depend (in some complicated manner) on the entries of R and W . To see this, simply note that any term in the permanent which picks up all possible factors of t must take all of its entries in the first n^ε rows from the upper left submatrix.

With this insight in hand, we can now give a new worst-to-average-case reduction for the permanent based on coefficient extraction: to compute $|\text{Per}(W')|^2$ for some worst-case matrix $W' \in \{0, \pm 1\}^{n^\varepsilon \times n^\varepsilon}$, pick many small values of t ($t = O(1/n^\varepsilon)$ suffices by prior arguments) and compute $|\text{Per}(A(t))|^2$ using our average-case algorithm. Then ask the NP oracle to give us a polynomial of degree $2n^\varepsilon$ which approximately matches these values. Now look at the top coefficient of that polynomial, and divide by the value of $|\text{Per}(R_D)|^2$. Crucially, we can estimate the value of $|\text{Per}(R_D)|^2$ to small multiplicative error, as this is another average-case instance. As multiplicative error only adds under division, this now gives us a *multiplicative* estimate for $|\text{Per}(W')|^2$. In other words, our algorithm translates relative error in the average case to relative error in the worst case.

We show that the overall robustness of this algorithm is merely $O(n^\delta)$ far in the exponent from showing quantum advantage, for any $\delta > 0$ (Theorem 1). The key point is that our polynomial now has degree $2n^\varepsilon$ rather than $2n$, and as such polynomial coefficient extraction incurs exponentially

⁶This is obtained by setting W to be a (tiny) arbitrary matrix of size $n^\varepsilon \times n^\varepsilon$ in direct sum with an identity on the remaining $n - n^\varepsilon$ dimensions.

$$\left| \text{Per} \left(\underbrace{\begin{pmatrix} R_A & R_B \\ R_C & R_D \end{pmatrix}}_{R \sim \mathcal{N}(0,1)^{n \times n}} + t \underbrace{\begin{pmatrix} W' & 0 \\ 0 & 0 \end{pmatrix}}_{W_{\text{dilute}}}, \right) \right|$$

$W' \in \{0, \pm 1\}^{n^\varepsilon \times n^\varepsilon}$

Figure 1: In Theorem 1, we extract the coefficient of the polynomial $|\text{Per}(R + tW_{\text{dilute}})|$, where R is a matrix of standard normals and W_{dilute} has a worst-case matrix in its upper left block of size $n^\varepsilon \times n^\varepsilon$ for any constant $\varepsilon > 0$, with all other matrix entries being 0. The top coefficient of this polynomial is $|\text{Per } W'| |\text{Per } R_D|$, where R_D is the complementary minor to W' .

less error blowup. As noted earlier, our proof crosses the Jerrum-Sinclair-Vigoda barrier as this argument intrinsically requires that W have mixed signs.

The corollary for RCS follows by a similar dilution argument—one simply picks a worst case random circuit which is a concatenation of an n^ε qubit worst case instance with an $(n - n^\varepsilon)$ -sized random instance, and applies prior worst-to-average-case reductions [Mov23, BFLL22, KMM22]. See Appendix B for details.

1.2.3 Overcoming the convexity barrier: square method and magnification lemma

While this first result exponentially improves on prior work, it is natural to ask how much closer we are to proving the Permanent-of-Gaussians Conjecture, or more generally to establishing hardness of sampling. The above results are obtained by diluting the worst case instance size so as to lessen the error incurred by coefficient extraction. However, the amount of error blowup relative to the worst case instance size has not improved. At a deeper level, despite crossing the Jerrum-Sinclair-Vigoda barrier, the proof still does not imply hardness of sampling from Stockmeyer counting. This is because Stockmeyer counting gives BPP^{NP} algorithm for approximating these squared permanents to $1/\text{poly}(n)$ multiplicative error, but the worst-to-average-case reduction then blows up this error exponentially. There is no compensating factor in the reduction to “fight against” this exponential loss. In other words, we have not yet crossed the convexity barrier.

In our next set of results, we extend the coefficient extraction technique to cross the convexity barrier. In particular we prove a new worst-to-average-case reduction for the permanent that can tolerate exponential losses from coefficient extraction, by developing two new techniques: the “*square method*” and “*magnification*.”

To do this, it is helpful to take a step back to examine what happens with *dense* worst case matrices with our new coefficient extraction approach. We apply two new modifications to coefficient extraction which improve the robustness of the dense case from $e^{-4n \log n - O(n)}$ [Kro23] to $e^{-1.5n \log n - O(n)}$. While these modifications appear simple at first glance, we will see they introduce a term which we can use to combat the error blowup from coefficient extraction. This dense result may at first look like a step backwards, but we will later show that this result is strong enough to

imply a hardness of sampling theorem.

The first idea to improve robustness in the dense case, which we call the “*square method*,” is to simply use the fact that $|\text{Per}(A(t))|^2$ is the square of a polynomial to reduce the degree of coefficient extraction. Suppose our worst-case matrix W is dense and define $A(t) := R + tW$ as before. While $|\text{Per}(A(t))|^2$ is a degree $2n$ polynomial, trivially we have that

$$|\text{Per}(A(t))|^2 = p(t)^2$$

for some degree- n polynomial $p(t)$. In our reduction, after (approximately) computing $p(t)^2$ at many values of t using our average-case algorithm, we can ask the NP oracle to give us the underlying degree n polynomial $p(t)$ which squares to the correct value (up to the error tolerance in the average-case computation). For real-value matrices, $p(t)$ is real, so is uniquely defined up to a sign. Again the highest coefficient of this polynomial (now the coefficient of t^n) contains the value of $\text{Per}(W)$ that we wish to compute.

One might a priori guess this simple change merely reduces the effective polynomial degree from $2n$ to n . Surprisingly, it has more benefit than that! In particular, suppose our average-case algorithm computes $p(t)^2$ to additive error $\pm\gamma$ at the points t near 0. How much error is induced on $p(t)$ itself? It turns out, $p(t)$ is estimated to *less error* than γ . Suppose our NP oracle gives us a polynomial $\tilde{p}(t) = p(t) + e(t)$ where $e(t)$ is some error polynomial. Then trivially we have

$$p(t)^2 \pm \gamma = (p(t) + e(t))^2 = p(t)^2 + 2p(t)e(t) + e(t)^2$$

As our errors are vanishingly small in relative terms, the cross error term dominates, and we see

$$|e(t)| \leq \frac{\gamma}{p(t)}$$

at points t near 0. In other words, we get to divide our error by the average-case value of the permanent, before we propagate the error through coefficient extraction. By assuming the Permanent Anticoncentration Conjecture 9, this value is $\sqrt{n!}$ to leading order, saving us an additional $\exp(\frac{n \log n}{2})$ beyond what we might have otherwise expected to gain in additive terms. This observation gets more interesting if we view it in relative terms. This correction factor can be seen as ensuring the relative error on $p(t)$ is the same (up to a constant factor of 2) as the relative error on $p(t)^2$, as relative error is preserved (up to constants) under taking powers.

Observe that degree reduction via the square method kept our error constant in relative terms on our underlying polynomial. On the other hand, polynomial coefficient extraction is naturally sensitive to error in additive terms. Our next observation is that we can use this mismatch to *reduce* the coefficient extraction error blowup in relative terms, by an exponential amount. The basic idea is to now consider a worst case matrix with two components: first, a smaller and possibly negative-entry matrix W' in the upper left hand corner of size n^ε , in direct sum with a larger matrix of all 1's of dimension $n - n^\varepsilon$ (see Figure 2).

Interestingly, including this large-permanent submatrix in our worst case actually *improves* our robustness in the worst-to-average-case reduction! This is because for this scheme, the top coefficient of the polynomial $p(t) = \text{Per}(A(t))$ is equal to $\text{Per}(W')(n - n^\varepsilon)!$, where this *magnification factor*, $(n - n^\varepsilon)!$, comes from the value of the permanent of the bottom right hand submatrix. Therefore, to compute $\text{Per}(W')$ it suffices to estimate this top coefficient to additive error $\frac{1}{3}(n - n^\varepsilon)!$ because $\text{Per}(W')$ is integer-valued, so this error is removed by rounding to the nearest integer multiple of $(n - n^\varepsilon)!$. In other words, the fact that this permanent of the all 1s submatrix is big allows for more error tolerance in the reduction, overall improving the robustness. We show this trick can be generalized to the more general formula:

$$\left| \text{Per} \left(\begin{array}{c} \boxed{R \sim \mathcal{N}(0, 1)^{n \times n}} \\ + \quad t \quad \underbrace{\begin{array}{c|c} \boxed{0} & \begin{array}{ccccc} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{array} \end{array} }_W \end{array} \right) \right|$$

$W' \in \{0, \pm 1\}^{n^\varepsilon \times n^\varepsilon}$

Figure 2: Polynomial $|\text{Per}(R+tW)|$, whose top coefficient is $|\text{Per } W'| (n - n^\varepsilon)!$. This is the ensemble under consideration in Thm. 18 where we coefficient-extract the unsquared permanent via the square method and use worst case magnification by padding W with a matrix of 1s.

Lemma 5 (Magnification of robustness in worst-to-average-case reductions). *It is #P-hard to compute random Gaussian permanents on average to within relative error*

$$\gamma_{\text{rel}} \leq \frac{(n - n^\varepsilon)!}{|\text{Per } R|} \cdot \Delta^n \cdot 2^{-O(n)}$$

for any constant $\varepsilon > 0$, where $R \sim \mathcal{N}(0, 1)^{n \times n}$, and where Δ is the “box size” as in Sec. 1.2.1, assuming a slight generalization of Permanent Anticoncentration, Conjecture 6.

In particular this worst-to-average-case reduction now has an exponential term—namely the ratio of the magnification factor to the average-case permanent—fighting against the exponential error blowup of polynomial coefficient extraction. For BosonSampling, this ratio is roughly $n!/\sqrt{n!} \approx \exp(\frac{n \log n}{2})$ which fights against a coefficient extraction error of $e^{-n \log n - O(n)}$, resulting in a net relative error of $\approx \exp(-\frac{n \log n}{2})$ (to leading order) needed in the average case to show hardness of sampling. To show hardness of sampling in the average case, this means we “merely” need to reduce the exponential loss of coefficient extraction to a weaker exponential, or increase the value of the worst-case matrix (now all 1s) by an exponential factor. This is not an easy problem—these terms are interrelated, so say simply boosting the norm of the all 1s matrix simultaneously improves the magnification-to-average-case ratio and worsens the coefficient extraction error, and does not show hardness of sampling. However, we now finally have a term fighting *against* coefficient extraction loss. We note a similar lemma can be shown for RCS as well—in particular for a real version of RCS with random orthogonal gates (see Section 3.3)—but does not yield any hardness of sampling results (see Discussion 1.3).

1.2.4 Average-case hardness of sampling using random rare events lemmas

Finally, we apply this new worst-to-average-case reduction to obtain the first nontrivial hardness of average-case sampling for BosonSampling. This uses techniques specific to BosonSampling, which to the best of our knowledge do not carry over to other quantum advantage schemes.

To show this, we consider our new worst-to-average-case reduction, whose relative error robustness is given by Lemma 5. To show an average-case hardness of sampling result via Stockmeyer,

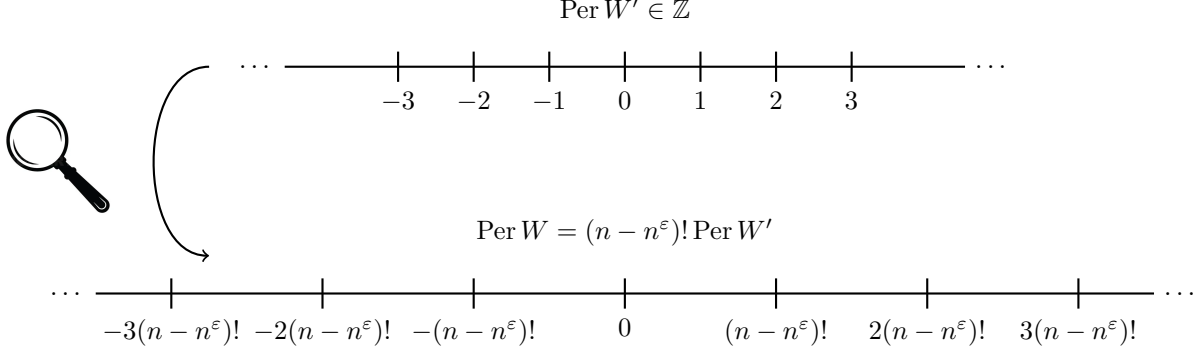


Figure 3: Magnification: variation between the possible values of $\text{Per } W'$, which is integer-valued, are *magnified* by a factor of $(n - n^\epsilon)!$ in $\text{Per } W$ for any $\epsilon > 0$ we choose. We can instantiate this with e.g. the matrix W in Fig. 2. Since the gradations in $\text{Per } W$ have been made so much wider, we can sustain more error while still computing $\text{Per } W$ precisely (see Lemma 5).

we need our relative error tolerance for $\#P$ -hardness to be inverse polynomial. Our compensating ratio of the magnification factor to the average-case permanent is $\exp(\frac{n \log n}{2})$, so we can only afford this much error from coefficient extraction. Unfortunately this is not enough of a loss budget to do a standard worst-to-average-case reduction. This is because in these reductions, we compute values of $|\text{Per}(A(t))|^2$ for values of t which are small enough so that $A(t)$ is distributed close in total variation distance to Gaussian, to ensure our average-case algorithm correctly computes $A(t)$ with high probability. To ensure closeness of total variation distance to constant error, t must be $O(1/n)$ —this calculation (due to Krovı [Kro23]) is optimal. Recalling from Sec. 1.2.1 the discrete Remez inequality, Lemma 4, this yields an error blowup of $\sim n^n = e^{n \log n}$. There is no hope of closing this gap with a standard total variation distance approach.

To get around this issue, our key idea is to go *out of distribution*. That is, what if we query points $A(t)$ which are *far* from Gaussian distributed? Clearly if our average-case algorithm could successfully compute the permanent of these matrices, then this would improve our robustness, as it would allow us to query points at much larger values of t , and hence reduce our error blowup. For example, if we could successfully compute $|\text{Per}(A(t))|^2$ for points $t = O(1/\sqrt{n})$, our coefficient extraction error would be halved in the exponent, and we could show hardness of average-case sampling! However, the issue is these matrices $A(t)$ at large values of t are far in total variation distance from Gaussian, so there is no trivial guarantee our algorithm works here. In fact total variation distance arguments are useless here; the TV distance between $A(t)$ and Gaussian is of the form $1 - \delta$ for a small value of δ . Even if we assume our average-case algorithm works perfectly, a TV distance argument would only say it must work with probability at least δ on these points. This is insufficient for our polynomial coefficient extraction techniques.

Instead, in our proof we go beyond total variation distance analysis to show that we can successfully query points $A(t)$ at high values of t , so long as our average-case algorithm works with very high probability. The basic idea is this: suppose our average-case algorithm works near perfectly, say with probability $1 - \delta$ over the choice of Gaussian matrix. We want to show it also works if we query it on these points $A(t)$ which are far from Gaussian. A basic observation is that these $A(t)$ are also Gaussian distributed, but with a shifted mean. We prove a simple lemma, Lemma 16, showing that rare events under one Gaussian distribution remain rare under another Gaussian, so long as their probability is less than e^{-d^2} where d is the distance between the means. Intuitively this is because if an event is extremely far from the mean of a Gaussian G_1 (much further than

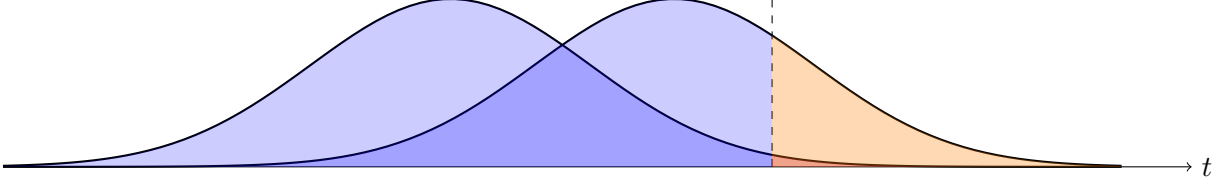


Figure 4: “Rare events” lemma 16 shows that a function that computes permanents of $\mathcal{N}(0, 1)^{n \times n}$ matrices with $1 - \exp(-O(n))$ probability also computes permanents of $\mathcal{N}(t, 1)^{n \times n}$ matrices with $1 - 1/\text{poly}(n)$ probability for $t = O(1/\sqrt{n})$. That is, an algorithm that works very often over a Gaussian distribution will also work reasonably often on a shifted Gaussian distribution. The figure depicts that events deep in the tail of one Gaussian are still tail events for a shifted Gaussian, with successful events colored blue and failure events colored orange.

the distance to the mean of G_2) it is also far from the mean of G_2 as well, and hence rare under G_2 (see Figure 4). We then apply this lemma to the event that the average-case algorithm fails under the standard Gaussian. If this is sufficiently rare for the average case, this is also rare for the distribution of $A(t)$, and hence the algorithm works with high probability to compute $A(t)$ as well. There is a loss in this argument which forces δ to be exponentially small. However, the key point is that if our average-case algorithm works with extremely high probability, then it can also evaluate these points $A(t)$ at high values of t , and hence lessen the coefficient extraction error in our reduction.

We show this can be leveraged to show a nontrivial hardness of sampling result for an exact (i.e. multiplicative-error) average-case sampler, following the outline above—but the proof requires several additional technical innovations. First, an average-case sampler that works with very high probability $1 - \delta$ over the choice of BosonSampling experiment *does not* immediately imply (by Stockmeyer counting) a BPP^{NP} algorithm for computing Gaussian permanents with probability $1 - \delta$. The issue is that submatrices of Haar random orthogonal matrices are not known to be exponentially close to Gaussian in TV distance, but rather have only been shown to be inverse polynomially close [JM19]. Thus setting the sampler success probability to $1 - \delta$ where $\delta = 2^{-O(n)}$ does not automatically yield a correspondingly good algorithm for computing Gaussian permanents.

To fix this we prove yet another “rare events lemma,” Proposition 17, that allows us to transfer our algorithm for Haar submatrices to Gaussian matrices. The proof, which may be of independent interest, requires showing new results in random matrix theory, exploiting properties of the probability densities and spectra of i.i.d. Gaussian matrices and submatrices of Haar orthogonals.

Second, for our algorithm to work we require $\text{Per}(A(t))$ to anticoncentrate. This is not guaranteed by the standard Permanent Anticoncentration Conjecture 9 as these matrices are out of distribution. We instead formulate a more general conjecture that shifted mean Gaussian permanents anticoncentrate:

Conjecture 6 (Anticoncentration of gently perturbed Gaussian permanents). *There exists a polynomial f such that for all n and $\epsilon > 0$,*

$$\mathbf{P}_{R \sim \mathcal{N}(0, 1)^{n \times n}} \left[|\text{Per}(R + tW)| < \frac{\sqrt{n!}}{f(n, 1/\epsilon)} \right] < \epsilon,$$

for arbitrary matrix W with entries bounded by 1 and $t = O(\frac{1}{\sqrt{n}})$.

We provide numerical evidence in support of Conjecture 6 in Appendix D. Moreover, we note a special case of this conjecture has already been proven for $\mathcal{N}(1/\text{poly} \log n, 1)$ matrices by [JLL21],

improving on work of Eldar and Mehraban [EM18]—and $N(0, 1)$ matrices are the subject of standard anticoncentration—so our conjecture is in some sense interpolating between these proven statements and conjectures to matrices with entries like $\mathcal{N}(1/\sqrt{n}, 1)$. See Figure 5 for a schematic.

1.3 Discussion and open problems

In this work we have exponentially improved over the best-known hardness results for BosonSampling, proving a robust worst-to-average-case reduction and showing the first non-trivial average-case multiplicative-error sampling result for (orthogonal) BosonSampling. It is natural to ask if our techniques can be pushed further to prove PGC and show hardness of BosonSampling in the general case. We note that further reductions in our coefficient extraction error could possibly yield intermediate results in this direction, in particular improving our average-case success probability of the sampler to be closer to $1 - 1/\text{poly}(n)$. In terms of pushing our results towards approximate average-case approximate sampling (i.e., from a distribution close in total variation distance), an important question is if our techniques relativize, as we know non-relativizing techniques will be required to show hardness of approximate sampling [AC17]. Interestingly Marshall, Aaronson and Djunko [MAD24] recently introduced new techniques that do not relativize. Of course the Permanent Anticoncentration Conjecture 9 remains open as well, and is assumed in our work.

Another natural question is if we can show any hardness of sampling for RCS similar to Theorem 3. Here the principal challenge is that the state-of-the-art of average-case hardness for RCS is substantially farther from the goal than for BosonSampling [BFNV19, Mov23, BFLL22, KMM22, Kro22]. While we show one can utilize the schemes of Lemma 5 for a real variant of RCS to obtain a magnification-to-average-case ratio which fights against extrapolation loss (see Sec. 3.3), this gain is at most 2^n for RCS, while existing worst-to-average-case reductions have much larger robustness losses. We leave this as an open problem.

We note a number of related works have studied the complexity of quantum advantage schemes under various forms of noise in the experiment, e.g., [ABOIN96, KK14, GD18, BFLL22, AGL⁺23, DNS⁺22, SYGY24, OLA⁺24, DHJB24, FGG⁺24, BBC⁺22, VNL⁺21] which can make the problems asymptotically easier in certain scenarios. In contrast our work is studying the complexity of near-noiseless variants of BosonSampling or RCS.

Finally, it remains open if our proofs can be extended from real (i.e. orthogonal) BosonSampling to complex (i.e. unitary) BosonSampling. The part of our proof that breaks here is the statement that, if you have evaluations of the square of a polynomial $|p(t)|^2$, that you can infer the underlying polynomial up to phase. While this is trivial in the real case (the phase is ± 1 , which is trivially disambiguated in the proof), in the complex case it is open if this approximately defines $p(t)$ up to a complex phase, and this appears to be an open problem in complex analysis [Her22]. We explain this in more detail in Appendix E.

2 Background

In this section, we record some background used throughout the paper.

In this work we will work to show hardness of *exact* (often called *multiplicative*)⁷ sampling of *average-case* BosonSampling. In BosonSampling the input is a Haar-random $m \times m$ unitary matrix U , describing a linear optical inteferometer on m modes, and a number n of photons. The goal

⁷We note that in the literature, it is common to use “*exact*” and “*multiplicative-error*” sampling synonymously, simply because the techniques that demonstrate exact hardness typically extend to multiplicative-error hardness automatically. We will also adopt this convention, using the two terms interchangeably.

is to output a sample of the probability distribution obtained by passing those n photons through the interferometer U and measuring in the photon number basis. We will work towards showing a classical algorithm cannot perform this task on average over the choice of U . We define an exact average-case sampler to be the following:

Definition 7 (Multiplicative-error average-case sampler). *A multiplicative-error average-case sampler for BosonSampling that succeeds with probability $1 - \alpha$ is an efficient classical probabilistic algorithm that, given a random $m \times m$ matrix U and error $\eta \geq 0$, outputs a sample y from a distribution \mathcal{P}'_U in time polynomial in n and $1/\eta$ such that, with probability at least $1 - \alpha$ over the choice of U ,*

$$(1 - \eta)P_U(y) < P'_U(y) \leq (1 + \eta)P_U(y) \quad (2.1)$$

for every possible output y , where P_U is the output distribution of the BosonSampling experiment.

We note that this notion of sampling to small multiplicative error has been previously studied in the literature, e.g. in [TD04, BJS10, AA13], where it has been shown that *worst-case* multiplicative-error sampling is hard assuming PH does not collapse. In this work we prove a similar statement for *average-case* multiplicative-error sampling.

Aaronson and Arkhipov gave a well-known reduction from classical sampling to approximate computing of output probabilities that uses Stockmeyer's approximate counting algorithm [Sto83], which runs in BPP^{NP} [AA13]. The idea of Stockmeyer's algorithm is to estimate the probability of any outcome by estimating the number of random strings that cause the sampler to output that outcome. This uses that a classical randomized algorithm can be treated as a deterministic algorithm that takes a random input. They then use this to show an efficient classical sampler cannot exist. The basic idea is that if approximately computing output probabilities is $\#\text{P}$ -hard, then it cannot lie in BPP^{NP} by Toda's theorem [Tod91]. Thus to show hardness of sampling, it suffices to conjecture that it is hard to compute the output probabilities of experiments.

Aaronson and Arkhipov showed that the Permanent-of-Gaussians Conjecture (PGC) suffices to prove hardness of average-case approximate sampling. PGC states that the following problem is $\#\text{P}$ -hard:

Definition 8 (GPE_{\pm}). *Given as input $R \sim \mathcal{N}(0, 1)^{n \times n}$ and error parameters $\epsilon, \delta > 0$, estimate $|\text{Per } R|^2$ to within additive error $\pm \epsilon \cdot n!$ with probability at least $1 - \delta$ over R , in $\text{poly}(n, 1/\epsilon, 1/\delta)$ time.*

This conjecture is natural in the context of BosonSampling as the output probabilities of these experiments correspond to matrix permanents of submatrices of the input U . For Haar-random unitaries of sufficient large dimension m , the submatrices are close to Gaussian, so this conjecture is referring to the complexity of computing output probabilities of the experiment.

Aaronson and Arkhipov also assume the Permanent Anticoncentration Conjecture, which posits a lower bound on the typical value of these permanents.

Conjecture 9 (Permanent Anticoncentration Conjecture (PACC) [AA13]). *There exists a polynomial f such that for all n and $\epsilon > 0$,*

$$\mathbf{P}_{R \sim \mathcal{N}(0, 1)^{n \times n}} \left[|\text{Per } R| < \frac{\sqrt{n!}}{f(n, 1/\epsilon)} \right] < \epsilon.$$

Assuming PACC, GPE_{\pm} and estimating Gaussian permanents to $1/\text{poly}(n)$ relative error, a problem known as GPE_{\times} , are polynomial-time equivalent. Thus if Conjecture 9 holds, then it suffices to show that GPE_{\times} is $\#\text{P}$ -hard.

Finally, we note that in standard BosonSampling U is assumed to be a Haar random unitary matrix. In this work all many of our results (all except the coefficient extraction and dilution arguments of Theorem 1, which apply to unitaries) pertain only to the case that U is a Haar random orthogonal matrix, whose $n \times n$ submatrices are distributed as $\mathcal{N}(0, 1)^{n \times n}$, i.e. our matrix entries are real i.i.d. standard normals. This is because the square method, developed in Section 3.2, applies to real polynomials. In Appendix E, we explain why the generalization to complex polynomials relies on a difficult question in complex analysis. Another advantage of random $m \times m$ orthogonal linear optical transformations is that there is a proof that $n \times n$ submatrices are close in total variation distance to i.i.d. Gaussian matrices for $m = \Omega(n^2)$ [JM19]. It is widely conjectured in the BosonSampling literature that the same holds for random unitaries. Our focus on orthogonal matrices means that we sidestep this conjecture altogether.

3 Techniques

3.1 Coefficient extraction

The overall scheme of past worst-to-average-case reductions for BosonSampling is an interpolation argument inspired by Lipton’s self-reducibility of the permanent, which exploits its polynomial structure to show that average-case instances are as hard as in the worst case [Lip91]. In particular, by taking a convex combination in variable t of an average-case instance and a worst-case instance, the permanent is a univariate polynomial in t . Then, by estimating values of the polynomial for small t by the average-case algorithm, one can extrapolate to $t = 1$, the permanents of which are $\#P$ -hard.

In this way, prior work has used the polynomial $\text{Per}((1 - t)R + tW)$ where R is a Gaussian random matrix and W is a worst-case matrix. On the other hand, the polynomial $\text{Per}(R + tW)$ also records information about $\text{Per}(W)$ as the highest order term is $t^n \text{Per}(W)$. We use this observation to provide an alternative way to perform a worst-to-average-case reduction for computation of the permanent. We can sample the values of this polynomial up to $t = O(n^{-1})$ because translation does not change the probability distribution quickly (see Appendix A).

Our main new technical ingredient is a way to recover the top coefficient of a polynomial from its values on an interval.

Lemma 10. *Let $p(x) = \sum_{j=0}^d p_j x^j$ be a polynomial of degree d satisfying*

$$\sup_{x \in [-\ell, \ell]} |p(x)| \leq \alpha.$$

Then

$$|p_d| \leq 2^{d+1} \ell^{-d} \alpha.$$

One should think of $p(x)$ as the difference between the true permanent polynomial and the approximate polynomial provided by the NP oracle in the reduction. The following lemma then provides a bound on the error in the approximate polynomial’s top coefficient, which encodes the worst-case permanent, i.e. $\text{Per}(W)$ above. This induces an additive error bound on the worst-case permanent.

Proof. By rescaling the inputs, it suffices to prove the result with $\ell = 1$.

Let $T_n(x)$ be the n -th Chebyshev polynomial. These polynomials satisfy the orthogonality relation

$$\int_{-1}^1 T_n(x)T_m(x) \frac{dx}{\sqrt{1-x^2}} = \begin{cases} 0, & n \neq m \\ \pi, & n = m = 0 \\ \frac{\pi}{2}, & n = m \neq 0. \end{cases}$$

Since $\text{span}\{1, x, \dots, x^n\} = \text{span}\{T_0, T_1, \dots, T_n\}$, it also follows that

$$\int_{-1}^1 q(x)T_d(x) \frac{dx}{\sqrt{1-x^2}} = 0$$

whenever q is a polynomial of degree at most $d-1$. Since the coefficient of x^n in T_n is 2^n , the polynomial $p - 2^{-d}p_d T_d$ is a polynomial of degree $d-1$ so that

$$\int_{-1}^1 (p(x) - 2^{-d}p_d T_d(x))T_d(x) \frac{dx}{\sqrt{1-x^2}} = 0.$$

Rearranging and using the orthogonality relations above, this becomes

$$2^{-d-1}\pi p_d = \int_{-1}^1 p(x)T_d(x) \frac{dx}{\sqrt{1-x^2}}.$$

On the other hand, using the uniform bound on $p(x)$ we can bound the latter integral as follows:

$$\left| \int_{-1}^1 p(x)T_d(x) \frac{dx}{\sqrt{1-x^2}} \right| \leq \alpha \int_{-1}^1 \frac{dx}{\sqrt{1-x^2}} = \pi\alpha.$$

□

3.2 Square method

Observe that the polynomial $|\text{Per}(R_t)|^2 := |\text{Per}(R + tW)|^2$ is not merely a polynomial of degree $2n$, but rather the square of a polynomial of degree n . It is natural to then ask whether this observation can be used to reduce the effective degree of the extrapolation to n instead of $2n$.

We suppose that we have some approximate values of a square polynomial p^2 , and we use an NP oracle to find some square q^2 that agrees with the approximate values. Then one expects that either $p \approx +q$ or $p \approx -q$ on these values. Notably in the case of complex polynomials, rather than a sign ambiguity there is a phase ambiguity. The consequences of this are explored in Appendix E. Throughout the main body of the text, we restrict our attention to real polynomials.

Before we prove Lemma 13 we record an elementary fact.

Lemma 11. *Let $p, q \in \mathbf{R}$ be real numbers satisfying*

$$|p^2 - q^2| < \delta.$$

Then $||p| - |q|| < |p|^{-1}\delta$.

Proof. We can assume without loss of generality that p and q are positive. Then $|p + q| > |p|$, so

$$|p - q| \leq |p|^{-1}|p - q||p + q| = |p|^{-1}|p^2 - q^2| < |p|^{-1}\delta.$$

□

Here we introduce the square method in the setting of coefficient extraction. To do so, we need the following discrete Remez inequality. It is proved in Appendix B, as is the square method for extrapolation.

Lemma 12 (Discrete Remez inequality). *Let $\{x_j\}_{j=0}^d \subset [-\ell, \ell]$ be a δ -separated set of points, meaning that $|x_i - x_j| \geq \delta$ for $i \neq j$. Then if p is a degree- d polynomial*

$$\sup_{[-\ell, \ell]} |p(x)| \leq (2e^2(\delta d)^{-1}\ell)^d \max_{0 \leq j \leq d} |p(x_j)|.$$

Lemma 13 (The square method for coefficient extraction). *Let p and q be real-valued polynomials of degree d and let S be a δ -separated collection of points in $[-\ell, \ell]$ with $|S| \geq 2d + 1$. Suppose moreover that*

$$\sup_{x \in S} |p^2(x) - q^2(x)| \leq \gamma$$

and $\inf_{x \in S} |q(x)| \geq K$. Then

$$||p_d| - |q_d|| \leq 2^{2d+1} e^{2d} (d\delta)^{-d} K^{-1} \gamma. \quad (3.1)$$

Proof of Lemma 13. By Lemma 11, we can conclude that

$$||p(x)| - |q(x)|| \leq K^{-1} |p(x)^2 - q(x)^2|$$

for all $x \in S$. In particular, for each $x \in S$ there exists a sign $\sigma_x \in \{\pm 1\}$ such that

$$|p(x) - \sigma_x q(x)| \leq K^{-1} |p(x)^2 - q(x)^2|.$$

Let $S^+ = \{x \in [-\ell, \ell] \mid \sigma_x = 1\}$ and $S^- = \{x \in [-\ell, \ell] \mid \sigma_x = -1\}$. At least one of these sets must contain more than $d + 1$ points, so without loss of generality suppose that $|S^+| \geq d + 1$. Then S^+ is also a δ -separated set of points, so by Lemma 12,

$$\sup_{x \in [-\ell, \ell]} |p(x) - q(x)| \leq 2^d e^{2d} (d\delta)^{-d} \ell^d K^{-1} \gamma$$

By Lemma 10 we have

$$|p_d - q_d| \leq 2^{2d+1} e^{2d} (d\delta)^{-d} K^{-1} \gamma.$$

□

We develop the square method for Random Circuit Sampling in Appendix B, where we use extrapolation rather than coefficient extraction. There, we can extrapolate the values of p itself rather than p^2 , and use the discrete Remez inequality to bound the extrapolation blowup induced on p^2 .

3.3 Worst case magnification

Building on the square method in Sec. 3.2, in this section we develop the method of worst case magnification.

Lemma 5 (Magnification of robustness in worst-to-average-case reductions). *It is #P-hard to compute random Gaussian permanents on average to within relative error*

$$\gamma_{\text{rel}} \leq \frac{(n - n^\varepsilon)!}{|\text{Per } R|} \cdot \Delta^n \cdot 2^{-O(n)}$$

for any constant $\varepsilon > 0$, where $R \sim \mathcal{N}(0, 1)^{n \times n}$, and where Δ is the “box size” as in Sec. 1.2.1, assuming a slight generalization of Permanent Anticoncentration, Conjecture 6.

Remark 14 (Overcoming the convexity barrier of [AA13]). The “convexity” barrier identified in [AA13] observes that worst-to-average-case reductions based on numerical tasks such as polynomial extrapolation or coefficient extraction cannot exactly estimate a worst-case (i.e., arbitrary) permanent given only $1/\text{poly}(n)$ relative error on random instances in the average-case, simply because the reduction suffers from an exponential loss without a factor to compensate. Lemma 5 surmounts the convexity barrier by demonstrating that one can magnify the reduction’s error tolerance by an exponentially large factor $(n - n^\varepsilon)!/|\text{Per } R|$, fighting against the exponential loss $\Delta^n 2^{-O(n)} = O(2^{-n \log n})$. Box size Δ is explicitly calculated in Appendix A to be $O(1/n)$.

Proof. Let us in particular take W to be a block-diagonal matrix consisting of a (small) block $W' \in \{0, \pm 1\}^{n^\varepsilon \times n^\varepsilon}$ for some constant $\varepsilon > 0$, in direct sum with a $(n - n^\varepsilon) \times (n - n^\varepsilon)$ block of the all 1s matrix, as illustrated in Fig. 2. Observe that $\text{Per } W' \in \mathbb{Z}$ and that the all 1s matrix has a permanent of $(n - n^\varepsilon)!$. Furthermore, observe that the leading-order coefficient $q_n = \text{Per } W = \text{Per } W' \cdot (n - n^\varepsilon)!$. Therefore see that the values of $q_n = \text{Per } W$ are *magnified* by quantity $(n - n^\varepsilon)!$, namely that q_n can take on values $\dots - 2(n - n^\varepsilon), -(n - n^\varepsilon), 0, (n - n^\varepsilon), 2(n - n^\varepsilon) \dots$ ⁸ This is illustrated in Fig. 3.

This motivates the following worst-to-average-case reduction in BPP^{NP} : ask the NP oracle for a degree n polynomial p such that $\sup_{x \in S} |p^2(x) - q^2(x)| \leq \gamma$ for a set S of $2n + 1$ evenly-spaced points in $[-1/n, 1/n]$. This has an efficient certificate because by assumption, we can evaluate average-case permanents, specifically $\text{Per}(R + tW)$ for $t \in \Delta = O(1/n)$, and simply check that for all $x \in S$, p^2 and q^2 are $\pm\gamma$ -close. Then we may simply output the leading-order (degree n) coefficient p_n of p .

The *key idea* is that by magnifying $\text{Per } W' \in \mathbb{Z}$ by a factor of $(n - n^\varepsilon)!$, it suffices for $|p_n|$ to be $\pm \frac{1}{3}(n - n^\varepsilon)!$ close to $|q_n|$ to compute $\text{Per } W'$ exactly.⁹ This would imply that computing $|q(0)|^2 = |\text{Per } R|^2$ to within $\pm\gamma$ is $\#\text{P}$ -hard under BPP^{NP} , i.e. that the average case is as hard as the worst case.

To complete the proof, we upper bound γ by recalling Lemma 13 from the previous page: for real-valued polynomials p and q of degree d and for a δ -separated collection of points S in $[-\ell, \ell]$ for which $|S| \geq 2d + 1$, if

$$\sup_{x \in S} |p^2(x) - q^2(x)| \leq \gamma$$

and

$$\inf_{x \in S} |q(x)| \geq K,$$

then

$$||p_d| - |q_d|| \leq 2^{2d+1} e^{2d} (d\delta)^{-d} K^{-1} \gamma, \quad (3.2)$$

where p_d and q_d denote the leading-order coefficients of p and q .

To apply this lemma, we instantiate p, q, d, δ, S , and K as follows.

- Take polynomial $q(t) = \text{Per}(R + tW)$, where $R \sim \mathcal{N}(0, 1)^{n \times n}$ and W is a matrix whose permanent is $\#\text{P}$ -hard. This is a degree $d = n$ polynomial.

⁸We note independent work of [BGHS25] used a related concept known as “granularity” in a quantum advantage setting. While superficially similar, these ideas are different, as in our setting we are *artificially* making a *worst-case* more “granular” to magnify our error tolerance, whereas in their setting the granularity naturally arises in their *average-case* distribution from the integrality of representation-theoretic quantities.

- S is a finite set of points $\{t_i\}$ in $[-\Delta, \Delta]$ for which box size $\Delta = O(1/n)$ and $|S| \geq 2n + 1$. Therefore, S is a collection of $\delta = O(1/n^2)$ points, and the expression $(d\delta)^{-d}$ in Eq. 3.2 simplifies to $O(\Delta)^{-n} = \Delta^{-n} \cdot 2^{-O(n)}$.
- p is a polynomial of the same degree as q , whose squared values are $\pm\gamma$ -close to q^2 for $x \in S$.
- Conjecture 6 implies $|q(x)| = |\text{Per}(R + xW)| \geq K = \sqrt{n!}/\text{poly}(n)$ for all $x \in S$ with probability $1 - 1/\text{poly}(n)$ by applying a union bound.

Using Eq. 3.2 and the bulleted substitutions above gives

$$||p_n| - |q_n|| \leq 2(2e)^{2n} O(\Delta)^{-n} (\sqrt{n!}/\text{poly}(n))^{-1} \gamma \quad (3.3)$$

$$\leq \frac{(n - n^\varepsilon)!}{3}. \quad (3.4)$$

Re-arranging,

$$\gamma \leq \frac{(n - n^\varepsilon)!}{3} \cdot \frac{\sqrt{n!}}{\text{poly}(n)} \cdot O(\Delta)^n \cdot \frac{(2e)^{-2n}}{2} \quad (3.5)$$

$$= (n - n^\varepsilon)! \cdot \frac{\sqrt{n!}}{\text{poly}(n)} \cdot \Delta^n \cdot 2^{-O(n)}. \quad (3.6)$$

Next we ask, how well does the value of $|p(0)|^2$ approximate that of $|q(0)|^2 = |\text{Per } R|^2$? To answer this question, we define

$$\gamma_{rel} := \frac{||p(0)|^2 - |q(0)|^2|}{|q(0)|^2} = \frac{||p(0)|^2 - |\text{Per } R|^2|}{|\text{Per } R|^2} \quad (3.7)$$

$$\leq \frac{\gamma}{|\text{Per } R|^2}, \quad (3.8)$$

recalling that γ is the maximum additive error between p^2 and q^2 for $x \in [-\Delta, \Delta]$.

Using Eq. 3.6, we can suggestively bound γ_{rel} :

$$\gamma_{rel} \leq \frac{\gamma}{|\text{Per } R|^2} = (n - n^\varepsilon)! \cdot \frac{\sqrt{n!}/\text{poly}(n)}{|\text{Per } R|^2} \cdot \Delta^n \cdot 2^{-O(n)} \quad (3.9)$$

$$= \frac{(n - n^\varepsilon)!}{|\text{Per } R|} \cdot \Delta^n \cdot 2^{-O(n)}, \quad (3.10)$$

where in the final line we invoke Permanent Anticoncentration Conjecture 9. □

⁹The astute reader will notice that in this reduction we obtain $\text{Per } W$ exactly, rather than to within some amount of relative error as we did in Theorem 1. Consequently, this argument does not cross the Jerrum-Sinclair-Vigoda (JSV) barrier. However, a simple modification to the proof crosses the convexity and JSV barriers simultaneously: Take as W a block-diagonal matrix comprising the following three matrices in direct sum: $W' \in \{0, \pm 1\}^{n^\varepsilon \times n^\varepsilon}$, the all 0s matrix of size $n^\varepsilon \times n^\varepsilon$, and the all 1s matrix of size $(n - 2n^\varepsilon) \times (n - 2n^\varepsilon)$. Just as in the proof of Theorem 1, the introduction of a random minor into the leading-order coefficient of $\text{Per}(R + tW)$ makes it so that the reduction obtains a relative error approximation to $\text{Per } W$.

3.4 Rare events lemma I: going out of distribution

Even with the exponential gains in error tolerance made using the square method and magnification, they still do not yet show the hardness of average-case sampling. The limitation comes from total variation distance analysis in the standard worst-to-average-case reduction, where we evaluate permanents drawn from a distribution close in TVD to i.i.d. Gaussian (see Sec. 1.2.1).

Our next technical innovation is to go out of distribution. Namely, we prove what we call a “rare events lemma,” which show that an algorithm that computes Gaussian permanents with sufficiently high probability can also compute permanents distributed *far* in total variation distance from i.i.d. Gaussian reasonably well. As these results exploit the specific structure of the Gaussian measure, they pertain only to BosonSampling.

Lemma 15. *Let $S \subset \mathbf{R}^N$ be a measurable set and let*

$$\delta := (2\pi)^{-N/2} \int_S e^{-\|x\|^2/2} dx.$$

Then for all $v \in \mathbf{R}^N$,

$$(2\pi)^{-N/2} \int_S e^{-\|x-v\|^2/2} dx \leq e^{\|v\|^2/2} \delta^{1/2}.$$

Proof. Let χ_S be the indicator function for the set S . We compute

$$\begin{aligned} (2\pi)^{-N/2} \int_S e^{-\|x-v\|^2/2} dx &= (2\pi)^{-N/2} \int e^{-\|x-v\|^2/2} \chi_S(x) dx \\ &= (2\pi)^{-N/2} e^{-\|v\|^2/2} \int e^{-\|x\|^2/2} e^{v \cdot x} \chi_S(x) dx \\ &\leq e^{-\|v\|^2/2} \left((2\pi)^{-N/2} \int e^{-\|x\|^2/2} e^{2v \cdot x} dx \right)^{1/2} \\ &\quad \left((2\pi)^{-N/2} \int e^{-\|x\|^2/2} \chi_S(x) dx \right)^{1/2} \\ &= e^{-\|v\|^2/2} \left(e^{2\|v\|^2} (2\pi)^{-N/2} \int e^{-\|x-2v\|^2/2} dx \right)^{1/2} \\ &\quad \left((2\pi)^{-N/2} \int e^{-\|x\|^2/2} \chi_S(x) dx \right)^{1/2} \\ &= e^{\|v\|^2/2} \delta^{1/2}. \end{aligned} \tag{3.11}$$

In the inequality above we applied Cauchy-Schwartz by writing

$$e^{-\|x\|^2/2} e^{v \cdot x} \chi_S(x) = (e^{-\|x\|^2/4} e^{v \cdot x}) (e^{-\|x\|^2/4} \chi_S(x)).$$

□

Next we prove our first rare events lemma, Lemma 16, depicted in Fig. 4.

Lemma 16 (Rare events lemma I). *Take $A \sim \mathcal{N}(0, 1)^{n \times n}$. Let $g : \mathbf{R}^{n \times n} \rightarrow \mathbf{R}$ be a function such that*

$$|\text{Per}(A)|^2 - g(A) \leq \varepsilon$$

holds with probability $1 - \delta$. Let B be an arbitrary matrix with entries $|b_{ij}| \leq 1$. Then

$$|\text{Per}(A + tB)|^2 - g(A + tB) \leq \varepsilon$$

holds with probability at least $1 - \sqrt{e^{\|tB\|^2} \cdot \delta}$, where $\|\cdot\|$ is the Hilbert-Schmidt norm.

Proof. Let $S \subset \mathbf{R}^{n \times n}$ be the set

$$S := \{A \mid \mathbf{R}^{n \times n} \mid ||\text{Per}(A)|^2 - g(A)| > \varepsilon\}.$$

We apply Lemma 15 with the set S above on $\mathbf{R}^N = \mathbf{R}^{n \times n}$, where $\mathbf{P}(A \in S) = \delta$ and thus the lemma shows that $\mathbf{P}(A + tB \in S) \leq e^{\|tB\|^2/2} \delta^{1/2}$ as desired. \square

We will leverage this lemma, in combination with the second rare events lemma in the next section, to prove the first nontrivial average-case harness of sampling theorem in Sec. 5.

3.5 Rare events lemma II: tail probabilities for orthogonal submatrices and i.i.d. Gaussians

To prove our hardness of sampling Theorem 3, further technical innovations are required beyond our first rare events Lemma 16. A key issue that remains is that an average-case sampler that works with very high probability $1 - \delta$ over the choice of BosonSampling experiment does *not* immediately imply (by Stockmeyer counting) a BPP^{NP} algorithm for computing Gaussian permanents with probability $1 - \delta$. The issue is that submatrices of Haar random orthogonal matrices are not known to be exponentially close to Gaussian in TV distance, but rather have only been shown to be inverse polynomially close [JM19]. Thus setting the sampler success probability to $1 - \delta$ where $\delta = 2^{-O(n)}$ does not automatically yield a correspondingly good algorithm for computing Gaussian permanents.

To fix this, we prove yet another “rare events” lemma, Proposition 17, that allows us to transfer our high probability algorithm for Haar submatrices to Gaussian matrices. The proof requires some highly nontrivial random matrix theory, exploiting properties of the probability densities and spectra of i.i.d. Gaussian matrices and submatrices of Haar orthogonals, and may be of independent interest.

Formally, we consider two models of $n \times n$ random matrices. The first is a Gaussian matrix X^n with independent (real) entries of variance n^{-1} . It has a probability density given by

$$p_G(X) = Z_G^{-1}(n) \left(\prod_{i \in [n]} \exp(-n\lambda_i(X^T X)/2) \right)$$

where $\lambda_i(A)$ is the i -th eigenvalue of A . The factor of n comes from the normalization we apply, and $Z_G^{-1}(n)$ is a normalization constant so that

$$\int_{\mathbf{R}^{n \times n}} p_G(X) dX = 1.$$

The second model is that of a $n \times n$ submatrix of a Haar-random $m \times m$ orthogonal matrix. We rescale by $\sqrt{m/n}$ so that the individual entries have variance n^{-1} . Then for $m \geq 2n$ the probability density takes the form (see [JM19], Lemma 2.1)

$$p_S(X) = Z_S^{-1}(n, m) \prod_{i \in [n]} (1 - n\lambda_i(X^T X)/m)^{(m-2n)/2} \mathbf{1}_{\lambda_i \leq m/n}.$$

Our main result in this section is the following:

Proposition 17 (Rare events lemma II). *Let $E \subset \mathbf{R}^{n \times n}$ be a measurable subset of matrices, and suppose that*

$$\mathbf{P}_S(E) \leq \delta$$

when E is sampled as the $n \times n$ submatrix of a Haar-random $m \times m$ orthogonal matrix, scaled by $\sqrt{m/n}$ so that each entry has variance n^{-1} . Let $0 < \alpha \leq 1$ and suppose that $n > C$ and $m > Cn^2$ for some absolute constant C . Then

$$\mathbf{P}_G(E) \leq 3 \exp(-n^\alpha) + 10 \exp(n^{\alpha/2})\delta,$$

where \mathbf{P}_G indicates that E is sampled with independent Gaussian entries of variance n^{-1} .

Proposition 17 has a highly nontrivial proof that we give in Appendix C, using analytic and random matrix theory techniques.

4 Hardness of computing output probabilities, Theorem 1

Our first result makes progress towards proving the Permanent-of-Gaussians Conjecture (PGC). Theorem 1 gives a new worst-to-average-case reduction for computing Gaussian permanents whose additive error tolerance exponentially improves on the state-of-the-art. For the first time, our error tolerance matches to leading order that of the Permanent-of-Gaussians Conjecture (PGC), $\exp(-n \log n - n - O(n^\delta))$ compared to the goal of $\exp(-n \log n - n - O(\log n))$. All that remains is “merely” to improve the $O(n^\delta)$ term in the exponent to $O(\log n)$.

Theorem 1 (Hardness of computing output probabilities). *For any $\delta > 0$, it is $\#P$ -hard under a BPP^{NP} reduction to approximate output probabilities of an n -photon, $O(n^2)$ -mode BosonSampling experiment to additive error $\exp(-n \log n - n - O(n^\delta))$ with success probability at least $2/3$, assuming the Permanent Anticoncentration Conjecture 9.*

Proof. Take any arbitrary constants $\delta > \varepsilon > 0$ and for ease of notation, define $k := \lfloor n^\varepsilon \rfloor$.

Recall that the output probability of a BosonSampling experiment is

$$p_R := \frac{|\text{Per } R|^2}{m^n} = \frac{|\text{Per } R|^2}{n^{2n}},$$

where $R \sim \mathcal{N}(0, 1)^{n \times n}$ and the number of modes $m = \Theta(n^2)$. Let \mathcal{A} be an algorithm that given as input R approximates p_R up to additive error γ , with success probability at least $1 - \eta$ over the choice of R for some constant $\eta < 1/4$. Additionally, consider a “worst-case” matrix W_{dilute} consisting of an upper-left block $W' \in \{0, \pm 1\}^{k \times k}$ with all other entries being 0.

We will show that then there exists a BPP^{NP^A} procedure that given as input *any* matrix W_{dilute} , approximates $|\text{Per } W'|$ up to small relative error for $\gamma = \exp(-n \log n - n - O(n^\delta))$, with constant success probability $1 - \eta'$ for η' slightly $> \eta$. The theorem statement follows immediately from the $\#P$ -hardness of computing even a multiplicative approximation to the permanent of a $\{0, \pm 1\}$ matrix.

Define the polynomial

$$|\text{Per}(A(t))| := |\text{Per}(A(0) + tW_{\text{dilute}})|, \quad (4.1)$$

where $A(0) \sim \mathcal{N}(0, 1)^{n \times n}$ and W_{dilute} is as above. Then $|\text{Per}(A(t))|$ is a degree k polynomial in t whose leading coefficient is $|\text{Per } W'| |\text{Per } R'|$, where R' is the complementary minor to W' . This polynomial is illustrated in Fig. 1.

As computed in Lemma 20, the total variation distance between the distributions of $A(t)$ and $A(0)$ is $O(kt)$. This follows from the KL divergence between two translated Gaussians and an application of Pinsker’s inequality.

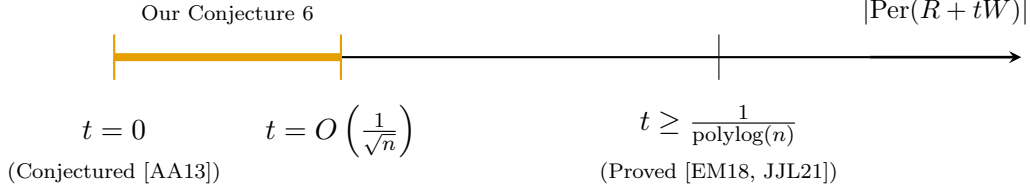


Figure 5: Conjecture 6 is that permanents of the form $|\text{Per}(R + tW)|$ are $\geq (n!/\text{poly}(n))^{-1}$ for matrices W whose entries are bounded by 1 and for $t = O(\frac{1}{\sqrt{n}})$. This interpolates between $t = 0$, i.e. PACC [AA13], and $t \geq \frac{1}{\text{polylog}(n)}$, where anticoncentration is proved unconditionally by [JLL21], improving upon [EM18]. In other words, we conjecture that permanents along the thickened orange line are at least as anticoncentrated as $|\text{Per } R|$ at $t = 0$.

Consider $O(k)$ equally spaced points $\{t_i\}$ in the interval $[0, \Delta]$ for $\Delta = O(1/k)$. For suitable choice of constants, we can ensure that for each t_i ,

$$\Pr \left[\left| \mathcal{A}(A(t_i)) - \frac{|\text{Per}(A(t_i))|^2}{n^{2n}} \right| \geq \gamma \right] \leq \eta + O(k\Delta) \leq \eta'$$

for some slightly larger constant η' . Then the BPP^{NP^A} procedure is as follows: query the NP oracle for a degree k polynomial q such that $|q(t_i)|^2/n^{2n}$ is $\pm\gamma$ -close to the value obtained by \mathcal{A} for at least half of the points $\{t_i\}$. This admits a certificate that can be efficiently verified by checking each point $\{t_i\}$ for agreement between \mathcal{A} and $|q(t_i)|^2/n^{2n}$. Return as output $|q_k|/|\text{Per } R'|$.

Finally, we will use Lemma 13 to guarantee that additive error $\gamma = \exp(-n \log n - n - O(n^\delta))$ gives a good relative error estimate of $|\text{Per } W'|$. Call S the subset of points $\{t_i\}$ at which \mathcal{A} and $|q(t_i)|^2/n^{2n}$ agree and observe that the points are $O(1/k^2)$ -separated. Moreover by permanent anticoncentration, $\inf_{t \in S} |q(t)| \geq \sqrt{n!}/\text{poly}(n)$ with probability at least $1 - 1/\text{poly}(n)$. Recalling that the leading-order coefficient of Eq. 4.1 is $|\text{Per } W'| |\text{Per } R'|$ and that $|\text{Per } R'| \geq \sqrt{n!} e^{-O(k \log n)}$, Lemma 13 gives that $||\text{Per } W'| - \frac{|q_k|}{|\text{Per } R'|}| \ll |\text{Per } W'|$ if $\gamma/|\text{Per } R'|^2 = e^{-O(k \log n)} = e^{-O(n^\delta)}$.

Overall, we have a BPP^{NP^A} procedure to multiplicatively estimate $|\text{Per } W'|$ if $\gamma = \exp(-n \log n - n - O(n^\delta))$, which concludes the proof. \square

This proof technique carries over to Random Circuit Sampling, which we show in Appendix B.

5 Hardness of sampling, Theorem 3

The goal of this section is to prove the following theorem, which closes the robustness gap for the first time at the expense of winnowing the failure probability to which we can prove hardness from $1/\text{poly}(n)$ to $1/\exp(O(n))$.

Theorem 3 (Hardness of sampling). *There does not exist a multiplicative-error classical sampler (see Def. 7) from the output distribution of an n -photon, $O(n^2)$ -mode real BosonSampling experiment that succeeds with probability at least $1 - \exp(-O(n))$ over the choice of experiment, assuming PH does not collapse and a slight generalization of Permanent Anticoncentration, Conjecture 6.*

To prove Theorem 3, we will assume the following anticoncentration conjecture, illustrated in Fig. 5. We provide numerical evidence for Conjecture 6 in Appendix D.

Conjecture 6 (Anticoncentration of gently perturbed Gaussian permanents). *There exists a polynomial f such that for all n and $\epsilon > 0$,*

$$\mathbf{P}_{R \sim \mathcal{N}(0,1)^{n \times n}} \left[|\text{Per}(R + tW)| < \frac{\sqrt{n!}}{f(n, 1/\epsilon)} \right] < \epsilon,$$

for arbitrary matrix W with entries bounded by 1 and $t = O(\frac{1}{\sqrt{n}})$.

Intuitively, the statement is that permanents of nonzero-mean Gaussian matrices are at least as anticoncentrated as are zero-mean Gaussians. In fact, the only setting in which there exists a proof of anticoncentration for Gaussian permanents¹⁰ is in the case of nonzero mean, in particular for $\mathcal{N}(t, 1)^{n \times n}$ matrices with t at least $1/\text{poly} \log n$ [JLL21, EM18]. With this exception, all forms of anticoncentration for BosonSampling remain open to date, to the authors' knowledge.

In order to prove Theorem 3, we prove a robust worst-to-average-case reduction that synthesizes the techniques developed earlier: coefficient extraction, the square method, magnification, and the first rare events lemma, Lemma 16.

Theorem 18. *It is $\#P$ -hard to compute $|\text{Per } R|^2$ for $R \sim \mathcal{N}(0, 1)^{n \times n}$ to $1/\text{poly}(n)$ relative error, with probability at least $1 - \exp(-O(n))$ over the choice of R , assuming Conjecture 6.*

Proof of Thm. 18. Take an arbitrary constant $\epsilon > 0$ and for ease of notation, define $k := \lfloor n^\epsilon \rfloor$.

Let \mathcal{A} be an algorithm that given as input $R \sim \mathcal{N}(0, 1)^{n \times n}$, approximates $|\text{Per } R|^2$ to within $1/\text{poly}(n)$ relative error, with probability at least $1 - \exp(-O(n))$ over the choice of R . Additionally, consider a “worst-case” block-diagonal matrix W with an upper-left block $W' \in \{0, \pm 1\}^{k \times k}$, and a lower-right $(n - k) \times (n - k)$ block of the all 1s matrix.

We will show that then there exists a $\text{BPP}^{\text{NP}^{\mathcal{A}}}$ procedure that given as input *any* such matrix W , approximates $|\text{Per } W'|$ to within small relative error, with success probability at least $\frac{2}{3}$. The theorem statement follows immediately from the $\#P$ -hardness of computing a multiplicative approximation to the permanent of a $\{0, \pm 1\}$ matrix.

Define the polynomial

$$|\text{Per } A(t)| := |\text{Per}(A(0) + tW)|, \tag{5.1}$$

where $A(0) \sim \mathcal{N}(0, 1)^{n \times n}$ and W is as above. Then $|\text{Per } A(t)|$ is a degree n polynomial in t whose leading coefficient is $|\text{Per } W| = (n - k)! |\text{Per } W'|$. This polynomial is illustrated in Fig. 2.

By Lemma 16, if \mathcal{A} computes a $\pm\gamma$ -approximation to $|\text{Per } R|^2$ with probability at least $1 - \beta$, then it computes a $\pm\gamma$ -approximation to $|\text{Per } A(t)|^2$ with probability at least $1 - \sqrt{\beta} \cdot e^{t^2 n^2}$. As in the theorem statement, we take $\beta = \exp(-O(n))$ so that \mathcal{A} has at least $1 - 1/\text{poly}(n)$ probability to correctly compute $|\text{Per } A(t^*)|^2$ where $t^* = O(1/\sqrt{n})$. In particular, we will take $t^* = 4e^{2.5} \cdot n^{k-1}/\sqrt{n} = 4e^{2.5}(1 + o(1))/\sqrt{n}$, and $\beta = \exp(-16e^5 n - O(\log n))$.

Consider $O(n)$ equally spaced $\{t_i\}$ in the interval $[0, \Delta]$ for $\Delta = O(1/\sqrt{n})$. By a union bound, all the points are correct to within $\pm\gamma$ with probability at least $1 - 1/\text{poly}(n)$. Then the $\text{BPP}^{\text{NP}^{\mathcal{A}}}$ procedure is as follows: query the NP oracle for a degree n polynomial q such that $|q(t_i)|^2$ is $\pm\gamma$ -close to the value obtained by \mathcal{A} for at least half of the points $\{t_i\}$. This admits a certificate that can be efficiently verified by checking each point $\{t_i\}$ for agreement between \mathcal{A} and $|q(t_i)|^2$. Return as output $|q_n|$.

¹⁰Although there are proofs of so-called “weak” anticoncentration, these do not imply the stronger form of anticoncentration necessary for the reductions made throughout the BosonSampling literature.

Finally, Lemma 13 guarantees that $|q_n|$ is a good multiplicative estimator of $|\text{Per } W|$. As \mathcal{A} obtains a $1/\text{poly}(n)$ relative error approximation to $|\text{Per } R|^2$, we have $\gamma = n!/\text{poly}(n)$. Assuming Conjecture 6, $|\text{Per } A(t_i)| \geq \sqrt{n!}/\text{poly}(n)$ on the set of points at which the NP oracle and \mathcal{A} agree. By construction, $|\text{Per } W| = (n-k)! |\text{Per } W'| = n! \exp(-k \log n + O(k \log k))$. Recalling from above that $t^* = 4e^{2.5} \cdot n^{k-1}/\sqrt{n} = 4e^{2.5}(1+o(1))/\sqrt{n}$ and substituting all these values into Eq. 3.1 of Lemma 13, we find at last that $||p_n| - |\text{Per } W|| \leq \frac{1}{\text{poly}(n)} |\text{Per } W|$. Overall, we have a $\text{BPP}^{\text{NP}^{\mathcal{A}}}$ procedure to multiplicatively estimate $|\text{Per } W|$, which concludes the proof. \square

We have now developed the machinery to prove the main result of this section, Theorem 3. Our robust worst-to-average-case reduction in Theorem 18, combined with the second rare events lemma, Lemma 17, allows us to prove the first nontrivial hardness of *sampling* result for average-case BosonSampling.

Theorem 3 (Hardness of sampling). *There does not exist a multiplicative-error classical sampler (see Def. 7) from the output distribution of an n -photon, $O(n^2)$ -mode real BosonSampling experiment that succeeds with probability at least $1 - \exp(-O(n))$ over the choice of experiment, assuming PH does not collapse and a slight generalization of Permanent Anticoncentration, Conjecture 6.*

Proof of Theorem 3. Suppose such a sampler exists. Then, given as input a Haar-random orthogonal matrix, to within $1/\text{poly}(n)$ relative error one can compute the squared permanent of the submatrix corresponding to a given output probability in BPP^{NP} via Stockmeyer’s approximate counting algorithm [Sto83]. Next we invoke Proposition 17 (proved in Appendix C) which says that “rare event” E sampled as the $n \times n$ submatrix of an $m \times m$ Haar-random orthogonal matrix that occurs with probability $\mathbf{P}_S(E) \leq \delta$, occurs with probability $\mathbf{P}_G(E) \leq \delta \cdot \exp(O(\sqrt{n})) + O(\exp(-n))$ if E is instead sampled from the i.i.d. Gaussian measure. Consequently $\mathbf{P}_S(E) \leq \exp(-O(n))$ implies

$$\begin{aligned} \mathbf{P}_G(E) &\leq \exp(-O(n)) \cdot \exp(O(\sqrt{n})) + O(\exp(-n)) \\ &\leq \exp(-O(n)). \end{aligned}$$

This suffices to show that the sampler of the theorem statement likewise has $1 - \exp(-O(n))$ success probability to correctly compute the squared permanents a matrix drawn from the $n \times n$ i.i.d. Gaussian matrices to within $1/\text{poly}(n)$ relative error. By Theorem 18, doing so is $\#\text{P}$ -hard. Finally by Toda’s theorem, this collapses PH. \square

Finally, we show that the failure probability in Theorems 3 and 18 exponentially improves upon the “trivial” algorithm that computes $\text{Per } W$ directly. The intuition is that an algorithm to compute $\text{Per}((1-t)R + tW)$ for t very close to 1 would need failure probability at most $\exp(-\tilde{O}(n^3))$, as it takes $\tilde{O}(n^3)$ bits to specify W , an $n \times n$ matrix of reals specified to $\tilde{O}(n)$ bits of precision. By comparison, Theorem 18 pertains to a sampler that fails with probability at most $\exp(-O(n))$.

We formalize this intuition as follows: Lemma 19 shows that if matrices A and $A + \delta B$ with entries bounded by 1 are sufficiently close, i.e. for sufficiently small δ , then their permanents are also close. As we show δ to be $\exp(-\tilde{O}(n^3))$, an algorithm that computes the permanents of all but this tiny fraction of matrices correctly is guaranteed to approximate the permanents of the remaining matrices, as well.

Lemma 19. *Suppose that A, B are matrices with entries $|a_{ij}|, |b_{ij}| \leq 1$ and $\delta < \frac{1}{100}n^{-1}(n!)^{-1}$. Then*

$$|\text{Per}(A) - \text{Per}(A + \delta B)| \leq 1.$$

Proof.

$$\begin{aligned}
|\text{Per}(A + \delta B) - \text{Per}(A)| &= \left| \sum_{\pi} \left[\prod_{i=1}^n a_{i\pi(i)} - \prod_{i=1}^n (a_{i\pi(i)} + \delta b_{i\pi(i)}) \right] \right| \\
&\leq \sum_{\pi} \left| \prod_{i=1}^n (1 + \delta b_{i\pi(i)}) - 1 \right|
\end{aligned}$$

Each term in the above sum is bounded by $(1 + \delta)^N - 1$. Thus, as there are $n!$ terms we compute

$$\begin{aligned}
|\text{Per}(A + \delta B) - \text{Per}(A)| &\leq n!((1 + \delta)^N - 1) \\
&\leq n!(|\exp(N\delta) - 1| + |\exp(N\delta) - (1 + \delta)^N|) \\
&< 1.
\end{aligned}$$

□

The key point is that if g is an approximation to the permanent that is wrong on $\frac{1}{2}$ of the cube of width $n^{-1}(n!)^{-1}$ centered at some matrix A_0 , then it is in particular wrong on a set of volume $(n^{-1}(n!)^{-1})^{n^2}$, which is to say $\exp^{-Cn^3 \log n}$. That means there is a trivial answer to the question only for error probabilities like $\exp(-\tilde{O}(n^3))$ rather than $\exp(-O(n))$.

Acknowledgements

Omitted for double blind review.

Appendices

A How far can you shift and scale i.i.d. Gaussian matrices?

In this section, we quantify how much an i.i.d. Gaussian matrix R is perturbed under “shift” and “scale,” namely dilation by $(1-t)$ and translation by tW for $t \in [0, 1]$ and worst-case matrix W . In other words, for what values of t is the distribution over $(1-t)R + tW$ a constant total variation distance from that of unperturbed distribution over R ?

The proof proceeds by an explicit calculation of the KL divergence between two Gaussians followed by Pinsker’s inequality. This is observed in [Kro22], with similar calculations appearing in [JM19] and [CDM22]. For completeness, we give the proof here.

It follows immediately from the proof that distributions that are only shifted, not scaled, likewise give $O(nt)$ total variation distance—this is the case for coefficient extraction. For Gaussians under shifts only, [AA13] (Lemma 48) also calculates a total variation distance of $O(nt)$ but by a different method.

Lemma 20 (Autocorrelation of Gaussian distribution). $\|\mathcal{D}_{(1-t)R+tW} - \mathcal{D}_R\|_{\text{TVD}} \leq O(nt)$.

Proof. We obtain an upper bound on total variation distance via Pinsker’s inequality:

$$\sqrt{2}\|\mathcal{D}_{(1-t)R+tW} - \mathcal{D}_R\|_{\text{TVD}} \leq \sqrt{D_{\text{KL}}(\mathcal{D}_{(1-t)R+tW}, \mathcal{D}_R)}, \quad (\text{A.1})$$

where on the right we have the KL divergence. By definition $R_{ij} \sim \mathcal{N}(0, 1)$, so $((1-t)R + tW)_{ij} \sim \mathcal{N}(tw_{ij}, (1-t)^2)$. The KL divergence between two Gaussians is

$$D_{\text{KL}}(\mathcal{N}(\mu_0, \sigma_0), \mathcal{N}(\mu_1, \sigma_1)) = \frac{(\mu_0 - \mu_1)^2 + \sigma_0^2}{2\sigma_1^2} + \log \frac{\sigma_1}{\sigma_0} - \frac{1}{2}. \quad (\text{A.2})$$

So

$$D_{\text{KL}}(\mathcal{N}(tw_{ij}, (1-t)^2), \mathcal{N}(0, 1)) = O(t^2), \quad (\text{A.3})$$

as $w_{ij} = O(1)$. Note that KL divergence is not symmetric so the order above matters.

Recalling that the KL divergence is additive for independent distributions, the RHS of Eq. A.1 is

$$\sqrt{D_{\text{KL}}(\mathcal{D}_{(1-t)R+tW}, \mathcal{D}_R)} = O(nt). \quad (\text{A.4})$$

□

Notably, in sharp contrast to the simple bound above, the analogous bounds for the shift-and-scale behavior of distributions in the “low-mode” or saturated limit regime of BosonSampling are highly nontrivial and are detailed in [BBD⁺23].

B Corollaries for Random Circuit Sampling

In this section we describe corollaries of our techniques for Random Circuit Sampling. First, we prove the discrete Remez inequality and give a much simpler proof of Robust Berlekamp-Welch, introduced in [BFL22].

Discrete Remez Inequality and Robust Berlekamp-Welch

Powering both extrapolation and coefficient extraction is the discrete Remez inequality, proved in this section.

Lemma 4 (Discrete Remez inequality). *Let $\{x_j\}_{j=0}^d \subset [0, 1]$ be a δ -separated set of points, meaning that $|x_i - x_j| \geq \delta$ for $i \neq j$. Then if p is a degree- d polynomial and $L \geq 1$,*

$$|p(L)| \leq (e^2(\delta d)^{-1}L)^d \max_{0 \leq j \leq d} |p(x_j)|.$$

Proof. Using Lagrange interpolation, we can write

$$p(x) = \sum_{j=0}^d p(x_j) \frac{\prod_{k \neq j} (x - x_k)}{\prod_{k \neq j} (x_j - x_k)}.$$

To see that this identity holds, observe that it holds at any x_j and that both sides are polynomials of degree d . Substituting $x = L$ and observing $|L - x_k| \leq L$, we obtain the bound

$$|p(L)| \leq L^d \max_{0 \leq j \leq d} |p(x_j)| \max_j \prod_{k \neq j} |x_j - x_k|^{-1}.$$

It remains to show that

$$\max_j \prod_{k \neq j} |x_j - x_k|^{-1} \leq e^{2d} (\delta d)^{-d}, \tag{B.1}$$

which by taking logarithms is equivalent to

$$\max_j \sum_{k \neq j} \log |x_j - x_k|^{-1} \leq 2d + d \log(d\delta)^{-1}.$$

We use the layer-cake formula to estimate the sum, writing

$$\begin{aligned} \sum_{k \neq j} \log |x_j - x_k|^{-1} &= \sum_{k \neq j} \int_0^{\log |x_j - x_k|^{-1}} dt \\ &= \int_0^\infty \#\{k \mid \log |x_j - x_k|^{-1} > t\} dt \\ &= \int_0^1 s^{-1} \#\{k \mid |x_j - x_k| < s\} ds. \end{aligned}$$

The second step follows from Fubini's theorem, and the last step from the change of variables $s = e^{-t}$. The δ -separated hypothesis on x_j implies

$$\#\{k \mid |x_j - x_k| < s\} \leq \begin{cases} 0, & s < \delta \\ 2\delta^{-1}s, & \delta \leq s \leq d\delta \\ d, & s > d\delta. \end{cases}$$

Therefore

$$\sum_{k \neq j} \log |x_j - x_k|^{-1} \leq \int_\delta^{d\delta} 2\delta^{-1} ds + \int_{d\delta}^1 ds^{-1} ds = 2d + d \log(d\delta)^{-1},$$

which concludes our proof of (B.1). \square

As a consequence of Lemma 4, we obtain a simpler proof of Robust Berlekamp-Welch, which was initially developed in [BFL22].

Theorem 21 (Robust Berlekamp-Welch bound). *Let $D = \{(x_i, y_i)\}_{i=1}^M$ be a set of $2(d+1) < M < 100d$ data points with x_i evenly spaced on the interval $[0, \Delta]$. Suppose that P_1 and P_2 are degree- d polynomials which satisfy*

$$\#\{j \mid |P_a(x_j) - y_j| \geq \delta\} < M/4 \quad (\text{B.2})$$

for $a = 1, 2$. Then

$$|P_1(1) - P_2(1)| \leq (C\Delta^{-1})^d \delta. \quad (\text{B.3})$$

Proof. The set on which P_1 and P_2 agree has at least $M/2 > (d+1)$ points. These points are $O(\Delta d^{-1})$ -separated. The conclusion follows from an application of Lemma 4. \square

The key observation is that Theorem 21 can be turned into an algorithm in BPP^{NP} that carries out extrapolation. In the reduction, P_2 is supplied by an NP oracle, where Eq. B.2 is the efficiently-verifiable predicate. Thus the algorithmic interpretation of Theorem 21 is that in P^{NP} , one can estimate at $t = 1$ a polynomial given faraway points close to $t = 0$, even when a constant fraction of the points are utterly corrupted. This is admissible because our use of Stockmeyer approximate counting already necessitates a BPP^{NP} reduction. The point is that because the worst-case is $\#\text{P}$ -hard, a reduction at any finite level of PH induces its collapse.

Square method for extrapolation

Lemma 22 (The square method for extrapolation). *Let p and q be real-valued polynomials of degree d , and let $S \subset [0, 1]$ be a δ -separated set of points with $|S| = 2d + 1$. Then,*

$$|p^2(1) - q^2(1)| \leq E|p(1)| + E^2 \quad (\text{B.4})$$

where

$$E = (e^2(d\delta)^{-1})^d \max_{x \in S} \frac{|p^2(x) - q^2(x)|}{|p(x)|}.$$

Proof of Lemma 22. By Lemma 11, we can conclude that

$$||p(x)| - |q(x)|| \leq |p(x)|^{-1} |p(x)^2 - q(x)^2|$$

for each data point j . In particular, for each $x \in S$ there exists a sign $\sigma_x \in \{\pm 1\}$ such that

$$|p(x) - \sigma_x q(x)| \leq |p(x)|^{-1} |p(x)^2 - q(x)^2|.$$

Let $S^+ = \{x \mid \sigma_x = 1\}$ and $S^- = \{x \mid \sigma_x = -1\}$. Since $|S^+ \cup S^-| = 2d + 1$, it follows that either $|S^+| \geq d + 1$ or $|S^-| \geq d + 1$. Without loss of generality suppose that $|S^+| \geq d + 1$. Then S^+ is also a δ -separated set of points, so by Lemma 4,

$$|p(1) - q(1)| \leq (e^2 d \delta)^{-d} \max_{x \in S^+} \frac{|p^2(x) - q^2(x)|}{|p(x)|} \leq E$$

To obtain (B.4) we use the triangle inequality to bound $|q(1)| \leq |p(1)| + |p(1) - q(1)|$ and write

$$|p^2(1) - q^2(1)| = |p(1) - q(1)| |p(1) + q(1)| \leq |p(1) - q(1)| (|p(1)| + |p(1) - q(1)|).$$

\square

Dilution for Random Circuit Sampling: Corollary 2

In this section, we show that the argument for Thm. 1 can be adapted to Random Circuit Sampling, as well. The proof follows readily from the dilution argument illustrated in Fig. 6 combined with well-established machinery from [Mov19, KMM22, BFL22, Mov23].

Corollary 2. *For any $\delta > 0$, it is $\#P$ -hard to approximate the output probabilities of n -qubit Random Circuit Sampling experiments of $\Omega(\log n)$ depth to additive error $2^{-n-O(n^\delta)}$.*

Proof sketch. As in the proof of Theorem 1, we note that output probabilities of a circuit have a polynomial structure, in this case coming from the Feynman path integral. Depicted in Fig. 1, we take a random circuit supported on n qubits, calling the circuit supported on the first n^ε qubits R_A and on the latter $n - n^\varepsilon$ qubits R_B .

Then, we perturb only the gates supported on the first n^ε qubits, circuit R_A , to a worst-case circuit W_A by the Cayley transform parametrized by θ [Mov19, Mov23]. Notably, two conditions hold: output probabilities of the θ -perturbed random circuit family are a low-degree rational function in θ , in particular with degree $(O(n^\varepsilon), O(n^\varepsilon))$, and moreover the total variation distance between the initial and θ -perturbed distributions is $O(k\theta)$.

By Theorem 21, an algorithm \mathcal{A} to compute output probabilities from the unperturbed circuit up to additive error γ can be in $\text{BPP}^{\text{NP}^\mathcal{A}}$ converted into Robust Berlekamp-Welch extrapolation that computes output probabilities of *any* circuit, e.g., a Fourier Sampling circuit, up to additive error $\gamma \cdot 2^{n+O(n^\delta)}$. This is $\#P$ -hard, completing the proof. \square

Remark 23 (Depth barrier). Referring to Lemma 22, our extrapolation bounds make use of lower bounds on the polynomial close to $\theta = 0$. In this way we invoke anticoncentration, proved for random circuits at log depth [DHB20]. In doing so, Corollary 2 is depth-sensitive, i.e. requires sufficiently deep circuits, and thus overcomes the depth barrier described in the Introduction.

Remark 24 (Born-rule barrier). Recall the “Born-rule” barrier identified by Krovi [Kro22], namely that the additive error needed to prove the hardness of average-case sampling (2^{-n}) is already larger than the additive error known to be hard in the worst case (2^{-2n} , which is derived from the Born rule by squaring the output amplitude of a Quantum Fourier Sampling circuit). Without dilution, it is seemingly impossible to prove a worst-to-average case reduction in which the additive error in the average case *is larger* than the additive error we need to obtain in the worst case. Corollary 2 overcomes the Born-rule barrier by simply “diluting” the worst-case instance to be polynomially smaller than the average-case instance.

C Rare events for orthogonal submatrices and i.i.d. Gaussians

We consider two models of $n \times n$ random matrices. The first is a Gaussian matrix X^n with independent (real) entries of variance n^{-1} . It has a probability density given by

$$p_G(X) = Z_G^{-1}(n) \left(\prod_{i \in [n]} \exp(-n\lambda_i(X^T X)/2) \right)$$

where $\lambda_i(A)$ is the i -th eigenvalue of A . The factor of n comes from the normalization we apply, and $Z_G^{-1}(n)$ is a normalization constant so that

$$\int_{\mathbf{R}^{n \times n}} p_G(X) dX = 1.$$

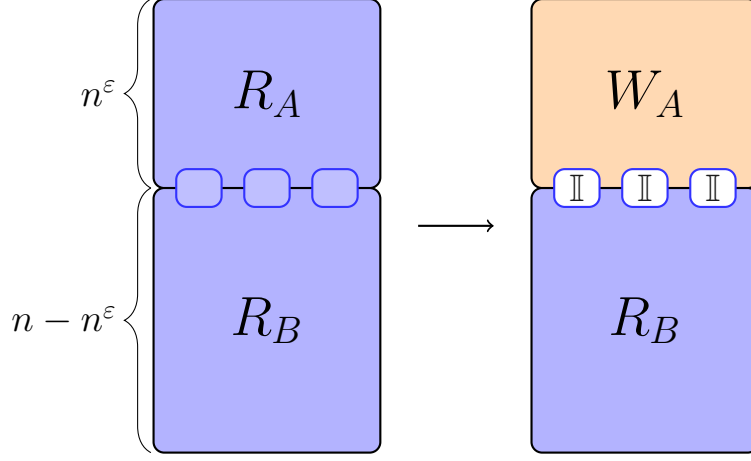


Figure 6: In Corollary 2, we take a random circuit supported on n qubits and perturb only the circuit supported on the first n^ε qubits, R_A , to a worst-case circuit W_A by the Cayley transform. The “bridge” gates on the interface between R_A and R_B transform into Identity gates so that the output probability on the righthand circuit factorizes.

The second model is that of a $n \times n$ submatrix of a Haar-random $m \times m$ orthogonal matrix. We rescale by $\sqrt{m/n}$ so that the individual entries have variance n^{-1} . Then for $m \geq 2n$ the probability density takes the form (see [JM19], Lemma 2.1)

$$p_S(X) = Z_S^{-1}(n, m) \prod_{i \in [n]} (1 - n\lambda_i(X^T X)/m)^{(m-2n)/2} \mathbf{1}_{\lambda_i \leq m/n}.$$

Our main result in this section is the following:

Proposition 17 (Rare events lemma II). *Let $E \subset \mathbf{R}^{n \times n}$ be a measurable subset of matrices, and suppose that*

$$\mathbf{P}_S(E) \leq \delta$$

when E is sampled as the $n \times n$ submatrix of a Haar-random $m \times m$ orthogonal matrix, scaled by $\sqrt{m/n}$ so that each entry has variance n^{-1} . Let $0 < \alpha \leq 1$ and suppose that $n > C$ and $m > Cn^2$ for some absolute constant C . Then

$$\mathbf{P}_G(E) \leq 3 \exp(-n^\alpha) + 10 \exp(n^{\alpha/2}) \delta,$$

where \mathbf{P}_G indicates that E is sampled with independent Gaussian entries of variance n^{-1} .

The proposition will follow from three main facts. The first relates the normalization constants c_G to c_S :

Lemma 25. *There exists a constant C such that for $n > C$ and $m > Cn^2$, the normalization constants $Z_G(n)$ and $Z_S(n, m)$ satisfy*

$$\frac{1}{10} \leq Z_G(n)/Z_S(n, m) \leq 10.$$

The second fact relates the densities p_G and p_S directly, up to the normalization c_G/c_S :

Lemma 26. Suppose that $\lambda_{\max}(X^T X) \leq K \leq \frac{m}{10n}$. Then

$$|\log(\frac{p_G(X)}{p_S(X)}) - \log(Z_S(n, m)/Z_G(n))| \leq K^2 \frac{n^3}{m^2} + K^3 \frac{n^4}{m^3} + \frac{n^2}{m} |\text{tr}[(X^T X)^2 - 2X^T X]|.$$

The third fact we need is that the right hand side above is often small for X sampled from the independent Gaussian distribution.

Lemma 27. let X be a Gaussian $n \times n$ matrix with independent entries of variance n^{-1} . Then:

$$\mathbf{P}(\lambda_{\max}(X^T X) > 3 + t) \leq \exp(-nt^2/2) \quad (\text{C.1})$$

Moreover,

$$\mathbf{P}(|\text{tr}[(X^T X)^2 - 2X^T X]| > 100\sqrt{t}) \leq \exp(-t) + \exp(-n) \quad (\text{C.2})$$

Before we proceed to the proofs of these lemmas we show how to combine them to deduce Proposition 17.

Proof of Proposition 17 using Lemmas 25-27. We write

$$E \subset (E \cap \Omega_{\text{good}}) \cup \Omega_{\text{bad}},$$

where we set $\Omega_{\text{bad}} = \Omega_{\text{good}}^c$ and Ω_{good} is the set of matrices satisfying

$$\Omega_{\text{good}} := \{\lambda_{\max}(X) \leq 4\} \cap \{|\text{tr}[(X^T X)^2 - 2X^T X]| \leq 100n^{\alpha/2}\}.$$

Then by Lemma 27 and the fact that $\alpha \leq 1$ we have

$$\mathbf{P}_G(\Omega_{\text{bad}}) \leq 2\exp(-n) + \exp(-n^\alpha) \leq 3\exp(-n^\alpha),$$

so using a union bound we have

$$\mathbf{P}_G(E) \leq \mathbf{P}_G(E \cap \Omega_{\text{good}}) + 3\exp(-n^\alpha).$$

Next we estimate $\mathbf{P}_G(E \cap \Omega_{\text{good}})$ by reweighting the probability measure:

$$\begin{aligned} \mathbf{P}_G(E \cap \Omega_{\text{good}}) &= \int_{E \cap \Omega_{\text{good}}} p_G(X) \, dX \\ &= \int_{E \cap \Omega_{\text{good}}} \frac{p_G(X)}{p_S(X)} p_S(X) \, dX. \end{aligned}$$

For $X \in \Omega_{\text{good}}$ we have by Lemma 26 the inequality

$$\frac{p_G(X)}{p_S(X)} \leq \frac{c_G}{c_S} \exp(16 \frac{n^3}{m^2} + 64 \frac{n^4}{m^3} + 100 \frac{n^2}{m} n^{\alpha/2}).$$

By Lemma 25 we have $\frac{c_G}{c_S} \leq 10$, and then using that $m > Cn^2$ and $n > C$ is large we have that for sufficiently large n ,

$$\frac{p_G(X)}{p_S(X)} \leq 10 \exp(n^{\alpha/2}).$$

Therefore we can conclude

$$\mathbf{P}_G(E \cap \Omega_{\text{good}}) \leq 10 \exp(n^{\alpha/2}) \int_E p_S(X) \, dX \leq 10 \exp(n^{\alpha/2}) \delta,$$

as desired. □

We now go through the proofs of the lemmas, in reverse order.

Proof of Lemma 27

The key ingredient in the proof of Lemma 27 is the following classical concentration inequality.

Lemma 28. *Let $f : \mathbf{R}^d \rightarrow \mathbf{R}$ be a Lipschitz-continuous function, that is one satisfying*

$$|f(x) - f(y)| \leq L\|x - y\|,$$

where the norm used above is the Euclidean one. Let X be a vector of independent standard Gaussians, and set $\bar{f} = \mathbf{E} f(X)$. Then

$$\mathbf{P}(|f(X) - \mathbf{E} f(X)| \geq t) \leq \exp(-t^2/(2L^2))$$

Now we can prove the proposition.

Proof of Lemma 27 using Lemma 28. First we observe that $\lambda_{\max}(X^T X)$ satisfies

$$\sqrt{\lambda_{\max}(X^T X)} = \sup_{\|u\|=\|v\|=1} u^T X v.$$

We can think of $\sqrt{\lambda_{\max}(X^T X)}$ as a function of n^2 independent Gaussian inputs, and the Lipschitz constant is equal to the maximum Lipschitz constant of the functions $u^T X v$. This latter Lipschitz constant is given by

$$\left(\sum_{i,j=1}^n n^{-1} |u_i v_j|^2 \right)^{1/2} = n^{-1/2} \|u\| \|v\| = n^{-1/2}.$$

Thus $L \leq n^{-1/2}$. Moreover for large enough n , $\mathbf{E} \lambda_{\max}(X^T X) \leq 3$ (in fact, $\lim_{n \rightarrow \infty} \mathbf{E} \lambda_{\max}(X^T X) = 2$). Therefore

$$\mathbf{P}(\sqrt{\lambda_{\max}(X^T X)} \geq 3 + t) \leq \exp(-nt^2/2).$$

This concludes the proof of (C.1).

Now we prove (C.2). Let $g(t)$ be the function

$$g(t) = \begin{cases} t^2 - 2t, & |t| \leq 4 \\ 8, & t > 4 \\ 24, & t < -4. \end{cases}$$

Then for X such that $\lambda_{\max}(X^T X) \leq 4$ (which by the above occurs with probability at least $1 - \exp(-n/2)$),

$$\phi(X) = \text{tr}((X^T X)^2 - 2(X^T X)) = \text{tr}(g(X^T X)) = \sum_i g(\lambda_i(X^T X)).$$

Now let E_{ij} be the matrix with a 1 in the (i, j) coordinate and 0's elsewhere. Then the Lipschitz constant of $\phi(X)$ (as a function of the independent Gaussian matrix entries) is

$$L = n^{-1/2} \left(\sum_{ij} \left(\frac{d}{dt} \phi(X + tE_{ij})|_{t=0} \right)^2 \right)^{1/2}$$

But the derivative in this trace is given by

$$\frac{d}{dt} \phi(X + tE_{ij})|_{t=0} = \text{tr}(g'(X^T X)(X^T E_{ij} + E_{ji} X)) = (g'(X^T X) X^T)_{ij} + (X g'(X^T X))_{ji}.$$

Thus

$$L \leq 2n^{-1/2}(\text{tr}[X^T X g'(X^T X)^2])^{1/2}.$$

Since g' itself is bounded by 24, this means $\|g'(X^T X)\| \leq 24$, so

$$L \leq 48n^{-1/2}(\text{tr}[X^T X])^{1/2}.$$

If $\lambda_{\max}(X^T X) \leq 4$, then $\text{tr}[X^T X] \leq 4n$, so this becomes

$$L \leq 96 \leq 100.$$

And now (C.2) follows. □

Proof of Lemma 26

The proof of Lemma 26 is a relatively simple calculation.

For X satisfying $\lambda_{\max}(X) \leq m/n$,

$$\log\left(\frac{p_G(X)}{p_S(X)}\right) = \log(Z_S(n, m)/Z_G(n)) + \frac{1}{2} \sum_{i \in [n]} [(2n - m) \log(1 - n\lambda_i/m) - n\lambda_i]$$

To simplify this further we use the Taylor approximation

$$|\log(1 + t) - (t - \frac{1}{2}t^2)| \leq |t|^3,$$

valid for $|t| \leq \frac{1}{10}$, which holds for $t = n\lambda_i/m$ when $\lambda_i \leq 2$ and $m > 20n$. Letting $E := |\log(\frac{p_G(X)}{p_S(X)}) - \log(c_G/c_S)|$, we can rearrange and cancel terms to obtain

$$\begin{aligned} |E| &= \left| \sum_{i \in [n]} [(2n - m)(-n\lambda_i/m - n^2\lambda_i^2/m^2 + O(n^3\lambda_i^3/m^3)) - n\lambda_i] \right| \\ &\leq \left| \sum_{i \in [n]} n^2\lambda_i^2/m - 2n^2\lambda_i/m \right| + \sum_{i \in [n]} |2n^2\lambda_i^2/m^2| + \sum_{i \in [n]} |n^3\lambda_i^3/m^3| \\ &\leq \frac{n^2}{m} |\text{tr}[(X^T X)^2 - 2X^T X]| + 8\frac{n^3}{m^2} + 8\frac{n^4}{m^3} \\ &\leq \sqrt{n} + O(1). \end{aligned}$$

To get to the last line we used that $\sum_{i \in [n]} \lambda_i^k = \text{tr}[(X^T X)^k]$ and the following inequalities which hold for “good” matrices:

$$\begin{aligned} \lambda_i &\leq 2 \\ |\text{tr}[(X^T X)^2 - 2X^T X]| &\leq \sqrt{n}. \end{aligned}$$

Proof of Lemma 25

First we establish some facts about submatrices of Haar-random orthogonal matrices. First we need a calculation for the moments of such matrices.

Lemma 29 (Lemma 2.5 of Jiang-Ma). *Letting X be an $n \times n$ submatrix of an $m \times m$ orthogonal matrix, scaled so that the entries have variance n^{-1} , we have*

$$\begin{aligned} \mathbf{E} \text{tr}[X^T X] &= n \\ \mathbf{E} \text{tr}[(X^T X)^2] &= \frac{m}{m+2} [2n + 1 - \frac{(n-1)^2}{(m-1)}]. \end{aligned}$$

We combine this with the following concentration inequality on the orthogonal group. The inequality below follows from the fact that the orthogonal group $SO(m)$ has Ricci curvature $\frac{m-2}{4}$ which by the Bakry-Emery argument (see [BGL⁺14], Theorem 2.1) shows that it has a log-Sobolev inequality with constant $\frac{4}{m-2}$, and therefore Gaussian concentration for Lipschitz functions.

Lemma 30. *Let $f : \mathbf{R}^{m^2} \rightarrow \mathbf{R}$ be a function taking as input $m \times m$ matrices, and suppose that f has Lipschitz constant L , that is*

$$|f(X) - f(Y)| \leq L\|X - Y\|,$$

where the norm used is the Hilbert-Schmidt norm, $\|X\|^2 = \sum_{ij} |x_{ij}|^2 = \text{tr}[X^T X]$. Let $\bar{f} = \mathbf{E} f(X)$ where the expectation is over $X \in SO(m)$ sampled uniformly from the Haar measure. Then also over this probability measure we have

$$\mathbf{P}(|f - \bar{f}| \geq t) \leq \exp(-mt^2/(8L^2)).$$

Next we need to know some facts about submatrices of typical Haar-random matrices.

Lemma 31. *Let $X = \sqrt{m/n}Y$ where Y is an $n \times n$ submatrix of a Haar-random orthogonal $m \times m$ matrix. There exist n_0 such that for $n > n_0$ and $m > n^2$, with probability at least $\frac{1}{2}$, X satisfies both $\lambda_{\max}(X^T X) \leq 5$ and $\text{tr}[(X^T X)^2 - 2X^T X] \leq 100$.*

Proof. We use the fact that, for any sequence $m(n)$ satisfying $m(n) \geq n^2$, we have

$$\lim_{n \rightarrow \infty} \mathbf{E} \lambda_{\max}(X^T X)^2 = 4.$$

In particular for some $n_0 > 0$ we have for any $n > n_0$ and $m > n^2$ the inequality

$$\mathbf{E} \lambda_{\max}(X^T X)^2 \leq 5.$$

Thus we conclude using Markov's inequality that

$$\mathbf{P}(\lambda_{\max}(X^T X) \geq 5) \leq \frac{1}{25} \mathbf{E} (\lambda_{\max}(X^T X)^2) \leq \frac{1}{5},$$

Next, by Lemma 29 we have

$$|\mathbf{E} \text{tr}[(X^T X)^2 - 2X^T X]| \leq 1.$$

Now let $g(t)$ be the same truncated version of $t^2 - 2t$ as in the proof of Lemma 27. The same argument as in there, combined with the concentration inequality of Lemma 30 implies that

$$\mathbf{P}(|\text{tr}[(X^T X)^2 - 2X^T X]| \geq 100) \leq \frac{1}{4}.$$

Combining these with a union bound proves the result. \square

We are finally ready to prove Lemma 25, and thus conclude the proof of Proposition 17.

Proof of Lemma 25. Let $A \subset \mathbf{R}^{n \times n}$ be the set of matrices that are “typical” both for the Gaussian distribution and as submatrices of orthogonal matrices:

$$A := \{\lambda_{\max}(X^T X) \leq 5\} \cap \{|\text{tr}[(X^T X)^2 - 2X^T X]| \leq 100\}.$$

Then by Lemma 31 and also Lemma 27 we have

$$\frac{1}{2} \leq \mathbf{P}_G(A) \leq 1$$

and also

$$\frac{1}{2} \leq \mathbf{P}_S(A) \leq 1.$$

In particular,

$$\frac{1}{2} \leq \frac{\mathbf{P}_G(A)}{\mathbf{P}_S(A)} \leq 2.$$

Moreover, by Lemma 26 we have for $X \in A$ that

$$|\log(p_G(X)/p_S(X)) - \log(c_G/c_S)| \leq 25\frac{n^3}{m^2} + 125\frac{n^4}{m^3} + 100\frac{n^2}{m} \leq 1$$

for $m > 101n^2$ and n sufficiently large. Therefore

$$\begin{aligned} \frac{\mathbf{P}_G(A)}{\mathbf{P}_S(A)} &= \frac{\int_A p_G(X) dX}{\int_A p_S(X) dX} \\ &= \frac{\int_A \frac{p_G(X)}{p_S(X)} p_S(X) dX}{\int_A p_S(X) dX} \\ &= \frac{\int_A \frac{p_G(X)}{p_S(X)} p_S(X) dX}{\int_A p_S(X) dX} \\ &= \frac{c_G}{c_S} \sup_{X \in A} \frac{c_S p_G(X)}{c_G p_S(X)} \leq e \frac{c_G}{c_S}. \end{aligned}$$

Thus $\frac{c_G}{c_S} \geq \frac{1}{2e}$. The argument also works to show that $\frac{c_S}{c_G} \geq \frac{1}{2e}$, so the proof follows from $2e < 10$. \square

D Numerical evidence for Conjecture 6

In this section, we provide brief numerical evidence for Conjecture 6:

Conjecture 6 (Anticoncentration of gently perturbed Gaussian permanents). *There exists a polynomial f such that for all n and $\epsilon > 0$,*

$$\mathbf{P}_{R \sim \mathcal{N}(0,1)^{n \times n}} \left[|\text{Per}(R + tW)| < \frac{\sqrt{n!}}{f(n, 1/\epsilon)} \right] < \epsilon,$$

for arbitrary matrix W with entries bounded by 1 and $t = O(\frac{1}{\sqrt{n}})$.

In particular, we compute permanents of several ensembles of Gaussian matrices with varying means, and plot their distribution. We observe similar distributions on the minimum non-zero permanent observed for all means tested.

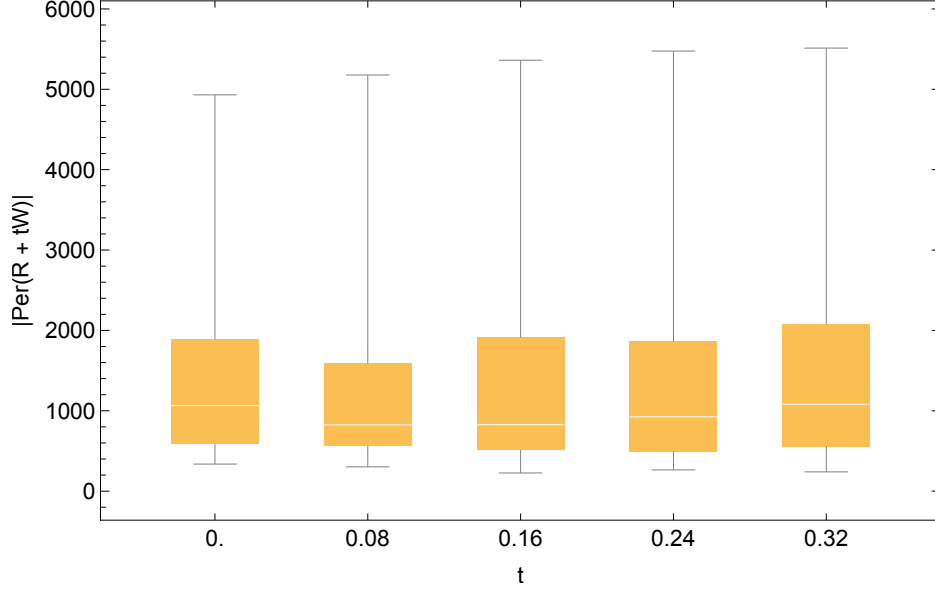


Figure 7: Box plots for the distribution of $|\text{Per}(R + tW)|$ of the form depicted in Fig. 2 for $n = 10$ and $k = n^\varepsilon = 5$. For five equally spaced values of $t \in [0, 1/\sqrt{n}]$, we randomly generate 30 such R and W . Notably, the box plots show remarkably little variation for increasing t in the relevant range, and in particular the lower bound for $t = 0$ holds for shifted t , as conjectured.

E Square method for complex polynomials

In this section we quickly discuss the difficulties involved in proving variants of the square trick (Lemma 13 in the context of coefficient extraction, and Lemma 22 in the context of extrapolation) in the case that p and q are *complex* valued polynomials. This is relevant if one wants to obtain hardness results for BosonSampling with a unitary random matrices (as opposed to orthogonal).

The complex case is significantly different from the real case because now one must recover a complex phase from the unit circle in \mathbb{C} rather than simply a sign ± 1 (for which there are only two possibilities – this is used in Lemma 13 for example).

This can be seen in the following example. Let $q(t) = 1$ be the constant polynomial, and let

$$p_d(t) = \sum_{j=0}^d \frac{(it/2)^j}{j!}$$

be the Taylor truncation of order d of the exponential $e^{it/2}$. Then by the Taylor remainder formula,

$$|p_d(t) - e^{it/2}| \leq \frac{2^{-d}}{d!}$$

on the interval $[-1, 1]$. Therefore $||p_d| - 1| \leq 2^{-d}(d!)^{-1}$ but also for any phase $e^{i\theta}$ there exists $t \in [0, 1]$ such that $|p_d - e^{i\theta}| > 1/4$.

This example shows that there is no complex analogue of Lemmas 13 or 22 if we only compare the values of p and q on some real interval. We can however perform an extrapolation by considering the values of the polynomial on the unit disk on the complex plane instead of the real interval $[-1, 1]$. To see that this has a hope of succeeding one can see that

$$\sup_{z \in B_1} ||p_d(z)| - 1| \gtrsim 1.$$

We are unfortunately unable to provide a complete proof of a variant of Lemma 13 in the complex case. What we are missing is an interesting and seemingly difficult question in complex analysis [Her22].

Conjecture 32. *For any complex-analytic functions f and g on the unit disk, there exists $\omega \in \mathbb{C}$, $|\omega| = 1$ such that $|\omega| = 1$ such that*

$$\max_{|z| \leq \frac{1}{4}} |f(z) - \omega g(z)| \leq C \max_{|z| \leq 1} ||f(z)| - |g(z)||.$$

Above D_r is the complex disk of radius r centered at the origin.

Note that the restriction to $D_{1/4}$ (or at least some D_r with $r < 1$) is necessary for the conjecture to hold. For example if $f = z^n$ and $g = z^{n+1}$ then for any $\omega = e^{i\phi}$, $\phi \in [0, 2\pi]$ one has

$$\max_{|z| \leq 1} |z^n - \omega z^{n+1}| \geq \max_{\theta \in [0, 2\pi]} |e^{in\theta} - e^{i\phi} e^{i(n+1)\theta}| = \max_{\theta \in [0, 2\pi]} |e^{i\theta} - e^{i\phi}| = 2.$$

On the other hand

$$\max_{|z| \leq 1} ||z|^n - |z|^{n+1}| \leq \max_{0 \leq r \leq 1} r^n(1 - r) \leq Cn^{-1}.$$

We are not claiming that Conjecture 32 is *sufficient* to transfer our results to Haar-random unitary matrices, only that it seems to be necessary to overcome this obstacle before one can transfer our techniques to that setting.

References

- [AA13] S. Aaronson and A. Arkhipov. The computational complexity of linear optics. *Theory of Computing*, 4:143–252, 2013.
- [AB16] Scott Aaronson and Daniel J. Brod. BosonSampling with lost photons. *Phys. Rev. A*, 93:012335, 2016.
- [ABOIN96] Dorit Aharonov, Michael Ben-Or, Russell Impagliazzo, and Noam Nisan. Limitations of noisy reversible computation. *arXiv preprint quant-ph/9611028*, 1996.
- [AC17] Scott Aaronson and Lijie Chen. Complexity-theoretic foundations of quantum supremacy experiments. In *Proceedings of the 32nd Computational Complexity Conference, CCC '17*, Dagstuhl, DEU, 2017. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.
- [AGL⁺23] Dorit Aharonov, Xun Gao, Zeph Landau, Yunchao Liu, and Umesh Vazirani. A polynomial-time classical algorithm for noisy random circuit sampling. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing*, pages 945–957, 2023.
- [Aru19] Frank Arute, et al. Quantum supremacy using a programmable superconducting processor. *Nature*, 574(7779):505–510, 2019.
- [BBC⁺22] Jacob FF Bulmer, Bryn A Bell, Rachel S Chadwick, Alex E Jones, Diana Moise, Alessandro Rigazzi, Jan Thorbecke, Utz-Uwe Haus, Thomas Van Vaerenbergh, Raj B Patel, et al. The boundary for quantum advantage in Gaussian boson sampling. *Science Advances*, 8(4):eabl9236, 2022.
- [BBD⁺23] Adam Bouland, Daniel Brod, Ishaun Datta, Bill Fefferman, Daniel Grier, Felipe Hernandez, and Michal Oszmaniec. Complexity-theoretic foundations of bosonsampling with a linear number of modes. *arXiv preprint arXiv:2312.00286*, 2023.
- [Ben80] Paul Benioff. The computer as a physical system: A microscopic quantum mechanical hamiltonian model of computers as represented by turing machines. *Journal of statistical physics*, 22:563–591, 1980.
- [BFLL22] Adam Bouland, Bill Fefferman, Zeph Landau, and Yunchao Liu. Noise and the frontier of quantum supremacy. In *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1308–1317, 2022.
- [BFNV19] Adam Bouland, Bill Fefferman, Chinmay Nirkhe, and Umesh Vazirani. On the complexity and verification of quantum random circuit sampling. *Nature Physics*, 15(2):159–163, Feb 2019.
- [BGHS25] Sergey Bravyi, David Gosset, Vojtech Havlicek, and Louis Schatzki. Classical and quantum algorithms for characters of the symmetric group. *arXiv preprint arXiv:2501.12579*, 2025.
- [BGL⁺14] Dominique Bakry, Ivan Gentil, Michel Ledoux, et al. *Analysis and geometry of Markov diffusion operators*, volume 103. Springer, 2014.

- [BIS⁺18] Sergio Boixo, Sergei V. Isakov, Vadim N. Smelyanskiy, Ryan Babbush, Nan Ding, Zhang Jiang, Michael J. Bremner, John M. Martinis, and Hartmut Neven. Characterizing Quantum Supremacy in Near-Term Devices. *Nature Physics*, 14:595–600, Jul 2018.
- [BJS10] Michael J. Bremner, Richard Jozsa, and Dan J. Shepherd. Classical simulation of commuting quantum computations implies collapse of the polynomial hierarchy. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 2010.
- [CDM22] Ulysse Chabaud, Abhinav Deshpande, and Saeed Mehraban. Quantum-inspired permanent identities. *Quantum*, 6:877, December 2022.
- [Den23] Yu-Hao Deng, et al. Gaussian boson sampling with pseudo-photon-number-resolving detectors and quantum computational advantage. *Phys. Rev. Lett.*, 131:150601, Oct 2023.
- [DHB20] Alexander M. Dalzell, Nicholas Hunter-Jones, and Fernando G. S. L. Brandão. Random quantum circuits anti-concentrate in log depth. *arXiv e-prints*, page arXiv:2011.12277, November 2020.
- [DHJB24] Alexander M Dalzell, Nicholas Hunter-Jones, and Fernando GSL Brandão. Random quantum circuits transform local noise into global white noise. *Communications in Mathematical Physics*, 405(3):78, 2024.
- [DHL⁺24] Matthew DeCross, Reza Haghshenas, Minzhao Liu, Yuri Alexeev, Charles H Baldwin, John P Bartolotta, Matthew Bohn, Eli Chertkov, Jonhas Colina, Davide DelVento, et al. The computational power of random quantum circuits in arbitrary geometries. *arXiv preprint arXiv:2406.02501*, 2024.
- [DNS⁺22] Abhinav Deshpande, Pradeep Niroula, Oles Shtanko, Alexey V Gorshkov, Bill Fefferman, and Michael J Gullans. Tight bounds on the convergence of noisy random circuits to the uniform distribution. *PRX Quantum*, 3(4):040329, 2022.
- [EM18] Lior Eldar and Saeed Mehraban. Approximating the permanent of a random matrix with vanishing mean. In *2018 IEEE 59th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 23–34, 2018.
- [Fey82] Richard P Feynman. Simulating physics with computers. *International journal of theoretical physics*, 21(6/7):467–488, 1982.
- [FGG⁺24] Bill Fefferman, Soumik Ghosh, Michael Gullans, Kohdai Kuroiwa, and Kunal Sharma. Effect of nonunitary noise on random-circuit sampling. *PRX Quantum*, 5:030317, Jul 2024.
- [GD18] Xun Gao and Luming Duan. Efficient classical simulation of noisy quantum computation. *arXiv preprint arXiv:1810.03176*, 2018.
- [Her22] Felipe Hernandez. Mathoverflow post: Complex functions with similar magnitude. <https://math.stackexchange.com/questions/4388057/complex-functions-with-similar-magnitude>, 2022. Accessed: 2024-11-4.

- [HHB⁺19] Jonas Haferkamp, Dominik Hangleiter, Adam Bouland, Bill Fefferman, Jens Eisert, and Juani Bermejo-Vega. Closing gaps of a quantum advantage with short-time Hamiltonian dynamics. *arXiv e-prints*, page arXiv:1908.08069, August 2019.
- [JL21] Zhengfeng Ji, Zhihan Jin, and Pinyan Lu. Approximating permanent of random matrices with vanishing mean: made better and simpler. In *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 959–975. SIAM, 2021.
- [JM19] Tiefeng Jiang and Yutao Ma. Distances between random orthogonal matrices and independent normals. *Transactions of the American Mathematical Society*, 372(3):1509–1553, 2019.
- [JSV04] Mark Jerrum, Alistair Sinclair, and Eric Vigoda. A polynomial-time approximation algorithm for the permanent of a matrix with nonnegative entries. *J. ACM*, 51(4):671–697, jul 2004.
- [KK14] Gil Kalai and Guy Kindler. Gaussian noise sensitivity and bosonsampling. *arXiv preprint arXiv:1409.3093*, 2014.
- [KMM22] Yasuhiro Kondo, Ryuhei Mori, and Ramis Movassagh. Quantum supremacy and hardness of estimating output probabilities of quantum circuits. In *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1296–1307. IEEE, 2022.
- [Kro22] Hari Krovi. Average-case hardness of estimating probabilities of random quantum circuits with a linear scaling in the error exponent. *arXiv preprint arXiv:2206.05642*, 2022.
- [Kro23] Hari Krovi. November 2023. Personal communication.
- [KT24] Dakshita Khurana and Kabir Tomer. Founding quantum cryptography on quantum advantage, or, towards cryptography from $\#P$ -hardness, 2024.
- [Lip91] Richard J. Lipton. New directions in testing. In Joan Feigenbaum and Michael J. Merritt, editors, *Distributed computing and cryptography*, volume 2, pages 191–202. American Mathematical Society, Oxford, 1991.
- [Mad22] Lars S. Madsen, et al. Quantum computational advantage with a programmable photonic processor. *Nature*, 606(7912):75–81, 2022.
- [MAD24] Simon C Marshall, Scott Aaronson, and Vedran Dunjko. Improved separation between quantum and classical computers for sampling and functional tasks. *arXiv preprint arXiv:2410.20935*, 2024.
- [Man80] Yuri Manin. Computable and uncomputable. *Sovetskoye Radio, Moscow*, 128:15, 1980.
- [Mor24] A. Morvan, et al. Phase transitions in random circuit sampling. *Nature*, 634(8033):328–333, 2024.
- [Mov19] Ramis Movassagh. Quantum supremacy and random circuits. *arXiv e-prints*, page arXiv:1909.06210, Sep 2019.
- [Mov23] Ramis Movassagh. The hardness of random quantum circuits. *Nature Physics*, pages 1–6, 2023.

- [NLPD⁺22] John C Napp, Rolando L La Placa, Alexander M Dalzell, Fernando GSL Brandao, and Aram W Harrow. Efficient classical simulation of random shallow 2D quantum circuits. *Physical Review X*, 12(2):021021, 2022.
- [ODMZ22] Michał Oszmaniec, Ninnat Dangniam, Mauro E.S. Morales, and Zoltán Zimborás. Fermion sampling: A robust quantum computational advantage scheme using fermionic linear optics and magic input states. *PRX Quantum*, 3:020328, May 2022.
- [OLA⁺24] Changhun Oh, Minzhao Liu, Yuri Alexeev, Bill Fefferman, and Liang Jiang. Classical algorithm for simulating experimental gaussian boson sampling. *Nature Physics*, pages 1–8, 2024.
- [Rem36] EJ Remez. Sur une propriété des polynômes de tchebycheff. *Comm. Inst. Sci. Kharkow*, 13:93–95, 1936.
- [Sto83] Larry Stockmeyer. The complexity of approximate counting. In *Proceedings of the Fifteenth Annual ACM Symposium on Theory of Computing*, STOC '83, page 118–126. Association for Computing Machinery, 1983.
- [SYGY24] Thomas Schuster, Chao Yin, Xun Gao, and Norman Y Yao. A polynomial-time classical algorithm for noisy quantum circuits. *arXiv preprint arXiv:2407.12768*, 2024.
- [TD04] Barbara M Terhal and David P DiVincenzo. Adaptive quantum computation, constant depth quantum circuits and Arthur-Merlin games. *Quantum Information & Computation*, 4(2):134–145, 2004.
- [Tod91] Seinosuke Toda. PP is as hard as the polynomial-time hierarchy. *SIAM Journal on Computing*, 20(5):865–877, October 1991.
- [VNL⁺21] Benjamin Villalonga, Murphy Yuezhen Niu, Li Li, Hartmut Neven, John C Platt, Vadim N Smelyanskiy, and Sergio Boixo. Efficient approximation of experimental gaussian boson sampling. *arXiv preprint arXiv:2109.11525*, 2021.
- [Wu21] Yulin Wu, et al. Strong quantum computational advantage using a superconducting quantum processor. *Phys. Rev. Lett.*, 127:180501, Oct 2021.
- [YGE⁺24] Aaron W Young, Shawn Geller, William J Eckner, Nathan Schine, Scott Glancy, Emanuel Knill, and Adam M Kaufman. An atomic boson sampler. *Nature*, 629(8011):311–316, 2024.
- [Zho20] Han-Sen Zhong, et al. Quantum computational advantage using photons. *Science*, 370(6523):1460–1463, Dec 2020.
- [Zho21] Han-Sen Zhong, et al. Phase-programmable gaussian boson sampling using stimulated squeezed light. *Physical Review Letters*, 127(18), Oct 2021.