BosonSampling Is Far From Uniform

Scott Aaronson* Alex Arkhipov†

Abstract

BosonSampling, which we proposed three years ago, is a scheme for using linear-optical networks to solve sampling problems that appear to be intractable for a classical computer. In a recent manuscript, Gogolin et al. claimed that even an ideal BosonSampling device's output would be "operationally indistinguishable" from a uniform random outcome, at least "without detailed a priori knowledge"; or at any rate, that telling the two apart might itself be a hard problem. We first answer these claims—explaining why the first is based on a definition of "a priori knowledge" so strange that, were it adopted, almost no quantum algorithm could be distinguished from a pure random-number source; while the second is neither new nor a practical obstacle to interesting BosonSampling experiments. However, we then go further, and address some interesting research questions inspired by Gogolin et al.'s mistaken arguments. We prove that, with high probability over a Haar-random matrix A, the BosonSampling distribution induced by A is far from the uniform distribution in total variation distance. More surprisingly, and directly counter to Gogolin et al., we give an efficient algorithm that distinguishes these two distributions with constant bias. Finally, we offer three "bonus" results about BosonSampling. First, we report an observation of Fernando Brandao: that one can efficiently sample a distribution that has large entropy and that's indistinguishable from a BosonSampling distribution by any circuit of fixed polynomial size. Second, we show that BosonSampling distributions can be efficiently distinguished from uniform even with photon losses and for general initial states. Third, we offer the simplest known proof that *Fermion*Sampling is solvable in classical polynomial time, and we reuse techniques from our BosonSampling analysis to characterize random FermionSampling distributions.

1 Background

BosonSampling [1] can be defined as the following computational problem. We are given as input an $m \times n$ complex matrix $A \in \mathbb{C}^{m \times n}$ ($m \geq n$), whose n columns are orthonormal vectors in $\mathbb{C}^m$. Let $\Phi_{m,n}$ be the set of all lists $S = (s_1, \ldots, s_m)$ of m nonnegative integers summing to n (we call these lists "experimental outcomes"); note that $|\Phi_{m,n}| = \binom{m+n-1}{n}$. For each outcome $S \in \Phi_{m,n}$, let $A_S \in \mathbb{C}^{n \times n}$ be the $n \times n$ matrix that consists of s_1 copies of A's first row, s_2 copies of A's second row, and so on. Then let $\mathcal{D}_A$ be the following probability distribution over $\Phi_{m,n}$:

$$\Pr_{\mathcal{D}_A}[S] = \frac{|\mathrm{Per}(A_S)|^2}{s_1! \cdots s_m!}, \tag{1}$$

*MIT. Email: aaronson@csail.mit.edu. This material is based upon work supported by the National Science Foundation under Grant No. 0844626, an Alan T. Waterman Award, a Sloan Fellowship, a TIBCO Chair, and an NSF STC on Science of Information.

†MIT. Email: arkhipov@mit.edu. Supported by an NSF Graduate Fellowship.

where Per represents the matrix permanent.[1] The BosonSampling problem is to sample from $\mathcal{D}_A$, either exactly or approximately. Ideally, we want to do so in time polynomial in n and m.

The BosonSampling problem has no known applications to cryptography or anything else. Nevertheless, it has two remarkable properties that motivate its study:

(1) BosonSampling is easy to solve using a quantum computer. Indeed, it is solvable by an especially simple *kind* of quantum computer: one that consists entirely of a network of beamsplitters, through which identical single photons are sent and then nonadaptively measured.[2] The reason for this is trivial: while our definition of BosonSampling was purely mathematical, the problem directly models the physics of identical, non-interacting photons (or in general, any bosonic particles). We merely need to interpret n as the number of photons (one for each input location or "mode"), m as the number of output modes, A as the transition matrix between the input and output modes (as determined by the beamsplitter network), and $S = (s_1, \ldots, s_m)$ as a possible output configuration consisting of s_i photons in the i^{th} mode for each i. Then according to quantum mechanics, the final amplitude for the basis state $|S\rangle$ is $\mathrm{Per}\,(A_S)/\sqrt{s_1! \cdots s_m!}$, so the probability of observing $|S\rangle$ on measuring is given by the formula (1).

(2) By contrast, it is *not* known how to solve BosonSampling efficiently using a classical computer. Indeed, one can say something much stronger: we showed in [1] that, if there *is* a polynomial-time classical algorithm for exact BosonSampling, then $\mathsf{P}^{\#\mathsf{P}} = \mathsf{BPP}^{\mathsf{NP}}$ (and hence the polynomial hierarchy collapses), which is considered vanishingly unlikely. The proof made indirect use of the #P-completeness of the permanent. Even for *approximate* BosonSampling, we proved that a fast classical algorithm would imply a $\mathsf{BPP}^{\mathsf{NP}}$ algorithm to approximate the permanent of an $n \times n$ matrix X of independent $\mathcal{N}(0,1)_{\mathbb{C}}$ Gaussian entries, with high probability over X. If (as we conjecture) this approximation problem is already #P-complete, then a $\mathsf{BPP}^{\mathsf{NP}}$ algorithm for it would be essentially ruled out as well. In summary, BosonSampling lets us base our belief in "Feynman's Conjecture"—the conjecture that quantum mechanics is exponentially hard to simulate by classical computers—on assumptions that seem much more "generic" than (say) the classical hardness of factoring integers.

One can study BosonSampling, as we did at first, purely from a theoretical computer science standpoint. However, BosonSampling can also be seen as an implicit proposal for a physics experiment—and perhaps not surprisingly, that is what has led to most of the interest in it.

In an ideal BosonSampling experiment, one would simultaneously generate n identical single photons, one in each of n input modes. One would then send the photons through a large network of beamsplitters, with the beamsplitter angles "random" and "arbitrary" but known to the experimenter in advance. Finally, one would measure the number of photons in each of m output modes, and check (after sufficiently many repetitions) whether the probability distribution over outputs $S \in \Phi_{m,n}$ was consistent with equation (1)—or in other words, with the prediction

[1] This is indeed a normalized probability distribution; see for example [1] for a proof.

[2] This type of quantum computer is not believed to be universal for quantum computation, or for that matter, even for *classical* computation! On the other hand, if we throw in one additional resource—namely *adaptivity*, or the ability to condition later beamsplitter settings on the outcomes of earlier photon-number measurements—then Knill, Laflamme, and Milburn [11] famously showed that we *do* get the capacity for universal quantum computation.

of quantum mechanics.[3] Assuming our complexity-theoretic conjectures, as n and m increased, those predictions would rapidly get harder to reproduce by any simulation running on a classical computer.

In this way, one might hope to get "experimental evidence" against the *Extended Church-Turing Thesis*: i.e., the thesis that all physical processes can be simulated by a classical computer with polynomial overhead. Furthermore, one might hope to get such evidence more easily than by building a universal quantum computer.

Last year, four independent groups (based in Brisbane [2], Oxford [14], Vienna [17], and Rome [4]) reported the first experiments more-or-less along the above lines. In these experiments, n (the number of photons) was generally 3,[4] while m (the number of output modes) was 5 or 6. The experiments directly confirmed, apparently for the first time, the prediction of quantum mechanics that the amplitudes of 3-photon processes are given by permanents of 3×3 complex matrices.

Obviously, these experiments do not yet provide any speedup over classical computing, nor are their results surprising: at some level they merely confirm quantum mechanics! But these are just the first steps. The eventual goal would be to demonstrate BosonSampling with (say) $n = 20$ or $n = 30$ photons: a regime where the quantum experiment probably *would* outperform its fastest classical simulation, if not by an astronomical amount. In our view, this would be an exciting proof-of-principle for quantum computation.

Scaling up BosonSampling to larger n remains a nontrivial experimental challenge. If it's possible at all, it will likely require optical technologies (especially single-photon sources) much more reliable than those that exist today. Indeed, we regard it as an open question whether BosonSampling experiments *can* be scaled to a "computationally interesting regime," without the use of quantum fault-tolerance. And presumably, if one can implement quantum fault-tolerance, then one might as well just skip BosonSampling and build a universal quantum computer!

2 The Claims of Gogolin et al.

The above issues with BosonSampling—the lack of a known practical motivation for it, the difficulties in scaling it up, etc.—are real and well-known. We have tried to be clear about them from the outset. However, in a recent preprint entitled "Boson-Sampling in the light of sample complexity," Gogolin et al. [7] criticize BosonSampling on different and much more theoretical grounds. Namely, they claim that the output of even an *ideal* BosonSampling device would be "operationally indistinguishable" from the uniform distribution. Indeed, they prove a theorem, which they interpret to mean that under "reasonable assumptions," a classical skeptic could never tell whether a claimed BosonSampling device was simply outputting uniformly random noise.

Gogolin et al. add that "it is important to note that our findings do not contradict the results of [Aaronson and Arkhipov [1]]." Yet despite this disclaimer, they strongly imply that [1] overlooked an elementary point, one that severely undermines the prospect of using BosonSampling to probe the Extended Church-Turing Thesis.

In Sections 5 and 6, we will explain in detail why Gogolin et al. are wrong. First, in Section

[3] To be clear, one would *not* try to estimate $\Pr[S]$ for each of the exponentially-many possible outputs S, since even for (say) $n = 10, m = 40$, that would require an impractical amount of data-collection. Instead, one would simply verify that the histogram of $\frac{|\mathrm{Per}(A_S)|^2}{s_1! \cdots s_m!}$ for the S's that *were* sampled was consistent with equation (1).

[4] Spring et al. [14] also managed to test $n = 4$, but for input states consisting of two modes with two photons each, rather than four modes with one photon each.

5, we consider their observation that so-called "symmetric algorithms" require exponentially many samples to distinguish a Haar-random BosonSampling distribution $\mathcal{D}_A$ from the uniform distribution. We explain why their restriction to "symmetric algorithms" is absurd: if one makes it, then countless other distributions become "indistinguishable from uniform," even though they are trivial to distinguish from uniform in reality!

Next, in Section 6, we consider Gogolin et al.'s "fallback position": that, even if one allows non-symmetric algorithms, distinguishing $\mathcal{D}_A$ from the uniform distribution could still be a hard *computational* problem. We point out that we made exactly the same observation in [1]—but that we also explained in [1] why the asymptotic hardness of verification will *not* be the limiting factor, in practice, for interesting BosonSampling experiments (with, say, $n = 30$ photons) designed to probe the Extended Church-Turing Thesis. (Unfortunately, Gogolin et al. never acknowledged or engaged with this point.)

3 Our Results

Even though we believe that the above completely suffices to answer Gogolin et al., in Sections 7 and 8 we go further, and address some interesting technical questions raised by their work. In particular, once we get over the confusion about "symmetric algorithms," it's clear on numerical and heuristic grounds that a generic BosonSampling distribution $\mathcal{D}_A$ is *not* close to the uniform distribution. But can we rigorously *prove* that $\mathcal{D}_A$ is not close to uniform? (This, of course, is necessary though not sufficient to prove that sampling from $\mathcal{D}_A$ is computationally intractable.) Also, is there a polynomial-time classical algorithm to *distinguish* $\mathcal{D}_A$ from the uniform distribution? What about from any efficiently-samplable distribution? Finally, what can we say about FermionSampling (defined in terms of the determinant rather than the permanent), whose statistical properties seem easier to understand?

Our results are as follows. In Section 7, we prove that a generic BosonSampling distribution is *not* close in variation distance to the uniform distribution. We get this as a consequence of a simple but nice fact, which could have independent applications to BosonSampling: that, if X is an iid Gaussian matrix, then $|\mathrm{Per}\,(X)|^2$ is a mixture of exponentially-distributed random variables. Then in Section 8, we describe a simple estimator R^* (the squared product of row-norms, scaled so that $\mathrm{E}\,[R^*] = 1$), which we prove can distinguish a generic BosonSampling distribution $\mathcal{D}_A$ from the uniform distribution with constant bias and in classical polynomial time. Let us state our result formally:

Theorem 1 *Let $A \in \mathbb{C}^{m \times n}$ be a Haar-random* BosonSampling *matrix with $m \geq n^{5.1}/\delta$. Let $\mathcal{U}$ be the uniform distribution over all experimental outcomes $S \in \Phi_{m,n}$ (or over all "collision-free" outcomes, i.e., those with $s_i \in \{0,1\}$ for all i), and let $\mathcal{D}_A$ be the* BosonSampling *distribution corresponding to A. There exists a linear-time computable estimator R^* such that, for sufficiently large n, and with probability $1 - O(\delta)$ over A, we have*

$$\Pr_{S \sim \mathcal{D}_A}\left[R^*\left(A_S\right) \geq 1\right] - \Pr_{S \sim \mathcal{U}}\left[R^*\left(A_S\right) \geq 1\right] \geq \frac{1}{9}.$$

In particular, this implies that, with $1 - O(\delta)$ probability, $\mathcal{D}_A$ and $\mathcal{U}$ have $\Omega(1)$ variation distance.

To clarify, the estimator R^* does not distinguish $\mathcal{D}_A$ from *any* efficiently-samplable distribution; indeed, we show in Section 8.1 that there are even natural "classical" models that produce the

same statistics for R^* as correct BosonSampling. However, R^* does confirm that the output of a purported BosonSampling device has nontrivial dependence on the beamsplitter settings, of a sort consistent with its working correctly. So, this could be combined with other evidence to build up a circumstantial case that a purported BosonSampling device works, even with (say) 100 or 1000 photons.

Thus, in Appendix 11, we study the broader question of BosonSampling versus *any* efficiently-samplable distribution. We first observe that, for any *fixed* k, it is easy to construct an efficiently-samplable distribution that is indistinguishable from a BosonSampling distribution $\mathcal{D}_A$—unconditionally!—by any circuit of size at most n^k. Indeed, this observation has nothing to do with BosonSampling: it follows from a Chernoff bound, and holds for any target distribution whatsoever. On the other hand, the "mockup" distribution thus constructed has only $O(\log n)$ entropy. So one could ask whether such a mockup distribution exists that *also* has large entropy. Here we report an observation due to Brandao (personal communication): namely, that for every k, a general theorem of Trevisan, Tulsiani, and Vadhan [18] can be used to construct an efficiently-samplable distribution that is indistinguishable from a generic BosonSampling distribution $\mathcal{D}_A$ by circuits of size at most n^k, and that *also* has $n - O(\log n)$ entropy. Of course, all of this leaves open the crucial question of whether or not there is a *single* efficiently-samplable distribution that cannot be distinguished from $\mathcal{D}_A$ by any polynomial-time algorithm.

Next, in Appendix 12, we sketch an argument that the estimator R^* works to distinguish a BosonSampling distribution from uniform, given *any* initial state (pure or mixed) with all photons concentrated in the first $n \ll m$ modes, and which has a non-negligible probability of a nonzero number of photons much less than m. In particular, this implies that R^* is "robust": it still works even if a large fraction of photons are randomly lost to the environment, and even if the inputs are (say) coherent or Gaussian states rather than single-photon Fock states.

Finally, Appendix 13 presents some results about the related problem of FermionSampling. In particular, we give a self-contained proof that FermionSampling is solvable in classical polynomial time. This was shown previously by Terhal and DiVincenzo [16] and by Knill [10] (and was implicit in work of Valiant [19]). However, our algorithm, which runs in $O\left(mn^2\right)$ time, is both simpler and faster than any previously-published FermionSampling algorithm, and seems like an obvious choice for implementations. The existence of this algorithm underscores that neither the "quantum" nature of BosonSampling, nor its exponentially-large Hilbert space, nor its n-particle interference can possibly suffice for computational hardness. This is why, contrary to the claims of, e.g., Gard et al. [5], we do not think it is possible to explain convincingly why BosonSampling should be a hard problem without using tools from computational complexity theory, as we did in [1].

In Appendix 13, we also reuse techniques from Section 8 to understand the statistical properties of Haar-random FermionSampling distributions. This turns out to be relatively easy, owing to the fact—which we prove for completeness—that $\left|\mathrm{Det}\left(X\right)\right|^2$ converges at an $O\left(\log^{-3/2} n\right)$ rate to a lognormal random variable, given a matrix $X \in \mathbb{C}^{n \times n}$ of iid Gaussians. The convergence of $\left|\mathrm{Det}\left(X\right)\right|^2$ to lognormal was previously shown by Girko [6] and by Costello and Vu [3], but for real X and without bounding the convergence rate. Note that numerically, the pdfs for $\left|\mathrm{Det}\left(X\right)\right|^2$ and $\left|\mathrm{Per}\left(X\right)\right|^2$ look nearly identical (see Figure 1). Thus, we conjecture that $\left|\mathrm{Per}\left(X\right)\right|^2$ converges to lognormal as well; if true, this would give us a much more detailed statistical understanding of Haar-random BosonSampling distributions.

4 Preliminaries

We use $[n]$ to denote $\{1,\ldots,n\}$. Given two probability distributions $\mathcal{D}_1 = \{p_x\}_x$ and $\mathcal{D}_2 = \{q_x\}_x$, the *variation distance*

$$\|\mathcal{D}_1 - \mathcal{D}_2\| := \frac{1}{2}\sum_x |p_x - q_x|$$

captures the maximum bias with which a sample from $\mathcal{D}_1$ can be distinguished from a sample from $\mathcal{D}_2$.

We already, in Section 1, defined the BosonSampling problem and most of the notation we will use in discussing it. However, one issue we need to get out of the way is that of multiple photons in the same mode: something that, from our perspective, is mostly an inconvenience that can be made irrelevant by taking sufficiently many modes. Formally, call an experimental outcome $S = (s_1,\ldots,s_m) \in \Phi_{m,n}$ *collision-free* if each s_i is either 0 or 1—so that A_S is simply an $n \times n$ submatrix of A, and $\Pr_{\mathcal{D}_A}[S]$ is simply $|\mathrm{Per}(A_S)|^2$. Also, let $\Lambda_{m,n} \subseteq \Phi_{m,n}$ be the set of all collision-free S. Note that $|\Lambda_{m,n}| = \binom{m}{n}$, which means that

$$|\Lambda_{m,n}| \geq \left(1 - \frac{n^2}{m}\right)|\Phi_{m,n}|. \tag{2}$$

In this paper, we will typically assume that $m \gg n^2$ (or, for technical reasons, even larger lower bounds on m), in which case (2) tells us that *most* outcomes are collision-free. Moreover, in the case that A is Haar-random, the following result from [1] justifies restricting our attention to the collision-free outcomes $S \in \Lambda_{m,n}$ only:

Theorem 2 ([1]) *Let $A \in \mathbb{C}^{m\times n}$ be a Haar-random* BosonSampling *matrix. Then*

$$\mathop{\mathrm{E}}_A\left[\Pr_{S\sim\mathcal{D}_A}[S \notin \Lambda_{m,n}]\right] < \frac{2n^2}{m}.$$

5 Limitations of "Symmetric Algorithms"

Suppose we want to verify that the output of a BosonSampling device matches the predictions of quantum mechanics (that is, equation (1)). Then given the matrix $A \in \mathbb{C}^{m\times n}$, our task can be abstracted as that of designing a *verification test*, $V_A : \Phi_{m,n}^k \rightarrow \{0,1\}$, that satisfies the following two constraints:

- **Efficiency.** $V_A(S_1,\ldots,S_k)$ can be computed classically in time polynomial in m, n, and k.
- **Usefulness.** $V_A(S_1,\ldots,S_k)$ is usually 1 if the outcomes $S_1,\ldots,S_k$ are drawn from $\mathcal{D}_A$, but usually 0 if $S_1,\ldots,S_k$ are generated in various "fake" ways (with the relevant "fake" ways depending on exactly what we are trying to verify).

In this paper, we will typically assume two additional properties:

- **Uniformity.** The polynomial-time algorithm to compute $V_A(S_1,\ldots,S_k)$ takes A as part of its input, rather than being a different algorithm for each A.

- **Properness.** $V_A(S_1, \ldots, S_k)$ distinguishes $\mathcal{D}_A$ from the "fake" distributions, even if we consider V_A's behavior only in the case where no S_i contains collisions—i.e., where $S_i \in \Lambda_{m,n}$ for all $i \in [k]$.

The motivation for the properness constraint is that our hardness results in [1] used the collision-free outcomes $S \in \Lambda_{m,n}$ only, so one might demand that one *also* restrict oneself to $\Lambda_{m,n}$ only when verifying that a hard problem is being solved.

Now, Gogolin et al. [7] insist on a further constraint on V_A, which they call "symmetry."

- **"Symmetry."** $V_A(S_1, \ldots, S_k)$ is a function only of the list of *multiplicities* of S_i's within $(S_1, \ldots, S_k)$, and not of any other information about the S_i's. In other words, $V_A(S_1, \ldots, S_k)$ is invariant under arbitrary permutations of the exponentially-large set $\Lambda_{m,n}$.

To illustrate, suppose that an alleged BosonSampling device with $n = 2$ photons and $m = 3$ modes were run $k = 3$ times, and suppose the collision-free outcomes were:

$$|1,1,0\rangle, \quad |1,1,0\rangle, \quad |0,1,1\rangle.$$

Then a symmetric verifier would be told only that one outcome occurred twice and that one outcome occurred once. From that information alone—together with knowledge of $A \in \mathbb{C}^{m \times n}$—the verifier would have to decide whether the device was sampling from the BosonSampling distribution $\mathcal{D}_A$, or (the "null hypothesis") whether it was just sampling from $\mathcal{U}$, the uniform distribution over $S \in \Lambda_{m,n}$.

Most of Gogolin et al.'s paper [7] is devoted to proving that, *if* A is drawn from the Haar measure, and k (the number of experimental runs) is less than exponential in n, then with high probability over A, the symmetric algorithm's task is information-theoretically impossible. We believe their proof of this theorem to be correct. The intuition is extremely simple: let $\tau(A)$ be the expected number of runs until *any* outcome $S \in \Lambda_{m,n}$ is observed more than once. Then we will have $\tau(A) = 2^{\Omega(n)}$, with overwhelming probability over the choice of A! This is just because $|\Lambda_{m,n}| = \binom{m}{n}$ is exponentially large—and while $\mathcal{D}_A$ is *not* close in variation distance to $\mathcal{U}$, neither is it concentrated on some tiny subset of size $2^{o(n)}$. Thus, regardless of whether the "true" distribution is $\mathcal{D}_A$ or $\mathcal{U}$, and notwithstanding the quadratic "speedup" obtained from the Birthday Paradox, after $k = 2^{o(n)}$ runs, a symmetric algorithm is overwhelmingly likely to have only the useless information, "no sample $S \in \Lambda_{m,n}$ was observed more than once so far." Or as Gogolin et al. put it:

> with probability exponentially close to one in the number of bosons, no symmetric algorithm can distinguish the Boson-Sampling distribution from the uniform one from fewer than exponentially many samples. This means that the two distributions are operationally indistinguishable without detailed a priori knowledge ... The realistic situation, at least as far as known certification methods are concerned, much more closely resembles the black box setting ... In this setting the certifier has no a priori knowledge about the output distribution. It is hence reasonable to demand that its decision should be independent of which particular samples he receives and only depend on how often he receives them. That is to say, knowing nothing about the probability distribution, the labels of the collected samples don't mean anything to the certifier, hence they should not influence his decision ... [O]ur findings imply that in the black box

> setting distinguishing the Boson-Sampling distribution from the uniform one requires exponentially many samples ... Colloquially speaking, our results on this problem give rise to a rather ironic situation: Instead of building a device that implements Boson-Sampling, for example by means of a quantum optical experiment, one could instead simply program a classical computer to efficiently sample from the uniform distribution over [outputs] and claim that the device samples from the post-selected Boson-Sampling distribution [for some unitary U]. If one chooses U from the Haar measure the chances of being caught cheating becomes significantly large only after one was asked for exponentially many samples. This implies that the findings of any experimental realisation of Boson-Sampling have to be interpreted with great care, as far as the notion "quantum supremacy" [sic] is concerned.

Our response to these claims can be summed up in one sentence: *there is no reason whatsoever to restrict the verifier to symmetric algorithms only.* The verifier's goal is to check whether the sampled distribution is a good match to the ideal BosonSampling distribution $\mathcal{D}_A$. Moreover, *even under Gogolin et al.'s assumptions,* the verifier knows the matrix A. And that makes sense: after all, A is not secret, but simply the input to the BosonSampling problem—just like some particular positive integer is the input to Factoring, or some particular graph is the input to Hamilton Cycle. So it seems bizarre to throw away the information about the actual identities of the observed outcomes $S \in \Lambda_{m,n}$, since the whole point is to compare those outcomes to $\mathcal{D}_A$, given knowledge of A.

By analogy, supposing we were testing an algorithm for factoring a composite number N into its prime factors p and q. Would anyone demand that our verdict be independent of *whether or not $p \times q$ actually equalled N?* Would anyone say that our decision should depend only on the shape of the histogram of probabilities for various output pairs (p, q), and be insensitive to the actual identities of the (p, q) pairs themselves? If not, then why impose such a strange constraint on BosonSampling verifiers?

As a side note, suppose we decided, for some reason, that the verifier's output had to be invariant under arbitrary relabelings of the *input and output modes*, though not necessarily of the entire set $\Lambda_{m,n}$ (call a verifier "weakly symmetric" if it has this property). Even then, we will show in Theorem 12 that $\text{poly}(m, n)$ samples information-theoretically suffice to distinguish $\mathcal{D}_A$ from $\mathcal{U}$, with high probability over A. The intuitive reason is that there are "only" $m!n!$ possible relabelings of the input and output modes, compared to $\binom{m}{n}!$ relabelings of the outcomes $S \in \Lambda_{m,n}$. So let V_A be a verifier that errs on input $\mathcal{U}$ with probability $\varepsilon \ll \frac{1}{m!n!}$—something we can easily achieve with $\text{poly}(m, n)$ samples, using amplification. Then a weakly symmetric verifier can simply run V_A for all $m!n!$ possible mode-relabelings, and check whether *any* of them yield a good match between $\mathcal{D}_A$ and the experimental results.

In truth, though, there is no reason to restrict to weakly symmetric verifiers either. Instead, the verifier should just get the raw list of experimental outcomes, $S_1, \ldots, S_k \in \Lambda_{m,n}$. In that case, Gogolin et al. themselves state in their Theorem 3 that $O\left(n^3\right)$ samples suffice to distinguish $\mathcal{D}_A$ from $\mathcal{U}$ information-theoretically, assuming the variation distance $\epsilon = \|\mathcal{D}_A - \mathcal{U}\|$ is a constant. This is true, but much too weak: in fact $O\left(1/\epsilon^2\right)$ samples suffice to distinguish *any* two probability distributions with variation distance ϵ. So the real issue is just to show that $\epsilon = \Omega(1)$ with high probability over A. That is what we will do in Section 7.

6 Intractability of Verification

Gogolin et al.'s second criticism is that, even if the verification algorithm is given access to the input matrix A (which they strangely call "side information"), it *still* can't verify in polynomial time that a claimed BosonSampling device is working directly. Indeed, this is the sole argument they offer for restricting attention to "symmetric" algorithms.[5] As they write:

> due to the very fact that Boson-Sampling is believed to be hard, efficient classical certification of Boson-Sampling devices seems to be out of reach ... The complexity theoretic conjecture under which Boson-Sampling is a hard sampling problem, namely that it is expected to be #P hard to approximate the permanent, implies that approximating the probabilities of the individual outputs of a Boson-Sampling device is also computationally hard. A classical certifier with limited computational power will hence have only very limited knowledge about the ideal output distribution of a supposed Boson-Sampling device ... it is certainly unreasonable to assume that the [certifier] has full knowledge of the ideal Boson-Sampling distribution. After all, it is the very point of Boson-Sampling that approximating the probabilities of individual outcomes is a computationally hard problem ... Our results indicate that even though, unquestionably, the Boson-Sampling distribution has an intricate structure that makes sampling from it a classically hard problem, this structure seems inaccessible by classical means.

While Gogolin et al. never mention this, we raised in the same point in [1, Section 1.3]:

> [U]nlike with Factoring, we do not believe there is any NP witness for BosonSampling. In other words, if n is large enough that a classical computer cannot solve BosonSampling, then n is probably *also* large enough that a classical computer cannot even verify that a quantum computer is solving BosonSampling correctly.

And again, in [1, Section 6.1]:

> Unlike with Factoring, we do not know of any *witness* for BosonSampling that a classical computer can efficiently verify, much less a witness that a boson computer can produce. This means that, when n is very large (say, more than 100), even if a linear-optics device is correctly solving BosonSampling, there might be no feasible way to prove this without presupposing the truth of the physical laws being tested!

Note that Gogolin et al. do not offer any formal argument for why BosonSampling verification is intractable, and neither did we. As we'll see later, there are many different things one can *mean* by verification, and some of them give rise to fascinating technical questions, on which we make some initial progress in this paper. In general, though, even if we assume our hardness conjectures from [1], we still lack a satisfying picture of which types of BosonSampling verification can and can't be done in classical polynomial time.

We did make the following observation in [1, footnote 23]:

[5] We still find the argument bizarre: if verifying a BosonSampling device is computationally hard, then why make the problem *even harder*, by throwing away highly-relevant information? The idea seems to be that, if non-symmetric algorithms need exponential computation time (albeit very few samples), then at least no one can complain that, by restricting to symmetric algorithms, we are prohibiting a known polynomial-time algorithm. As we will explain, the trouble with this argument is that, even *assuming* we need $\exp(n)$ computation time, an algorithm that needs only $\text{poly}(n)$ experimental samples is vastly preferable in practice to one that needs $\exp(n)$ samples.

> [G]iven a matrix $X \in \mathbb{C}^{n \times n}$, there *cannot* in general be an NP witness proving the value of $\mathrm{Per}(X)$, unless $\mathsf{P}^{\#\mathsf{P}} = \mathsf{P}^{\mathsf{NP}}$ and the polynomial hierarchy collapses. Nor, under our conjectures, can there even be such a witness for *most* Gaussian matrices X.

However, one lesson of this paper is that the above does not suffice to show that verification is hard. For it is possible that one can "verify" a BosonSampling distribution $\mathcal{D}_A$ (in the sense of distinguishing $\mathcal{D}_A$ from a large and interesting collection of "null hypotheses"), without having to verify the value of $\mathrm{Per}(X)$ for any particular matrix X.

Having said all that, let's suppose, for the sake of argument, that it *is* computationally intractable to verify a claimed BosonSampling device, for some reasonable definition of "verify." If so, then what is our response to that objection? Here, we'll simply quote the response we gave in [1, Section 1.3]:

> While [the intractability of verification] sounds discouraging, it is not really an issue from the perspective of near-term experiments. For the foreseeable future, n being *too large* is likely to be the least of one's problems! If one could implement our experiment with (say) $20 \leq n \leq 30$, then certainly a classical computer could verify the answers—but at the same time, one would be getting direct evidence that a quantum computer could efficiently solve an "interestingly difficult" problem, one for which the best-known classical algorithms require many millions of operations.

And again, in [1, Section 6.1]:

> [F]or experimental purposes, the most useful values of n are presumably those for which a classical computer has some difficulty computing an $n \times n$ permanent, but can nevertheless do so in order to confirm the results.

In other words: yes, there might be no way to verify a BosonSampling device when $n = 100$. But 100-photon BosonSampling is probably neither experimentally feasible nor conceptually necessary *anyway*! The only known "application" of BosonSampling is as a proof-of-principle: showing that a quantum system can dramatically outperform its fastest classical simulation on some task, under plausible complexity conjectures. And for that application, all that matters is that there be *some* range of n's for which

(1) linear-optical devices and classical computers can both solve BosonSampling in "non-astronomical" amounts of time, but

(2) the linear-optical devices solve the problem noticeably faster.

More concretely, as we pointed out in [1], the fastest-known algorithm for general $n \times n$ permanents, called *Ryser's algorithm*, uses about $2^{n+1}n$ floating-point operations. If (say) $n = 30$, then $2^{n+1}n$ is about 64 billion, which is large but perfectly within the capacity of today's computers. A classical computer therefore could feasibly check that $|\mathrm{Per}(A_{S_1})|^2, \ldots, |\mathrm{Per}(A_{S_k})|^2$ satisfied the expected statistics, where $S_1, \ldots, S_k$ were the outputs of an alleged 30-photon BosonSampling device. At the same time, the device would presumably sample the S_i's *faster* than any known classical method, for any reasonable definition of the word "faster." If so, then BosonSampling would have achieved its intended purpose: in our view, one would have done an experiment that was harder than any previous experiment for a believer in the Extended Church-Turing Thesis to explain.

7 Deviation from Uniformity

Even if experimenters can live with 64 billion floating-point operations, they certainly *can't* live with 64 billion laboratory experiments! So we still have the burden of showing that few experiments suffice to distinguish a BosonSampling distribution $\mathcal{D}_A$ from "trivial" alternatives, and in particular from the uniform distribution $\mathcal{U}$ over $\Lambda_{m,n}$. That is what we will do in this section. In particular, we will prove that $\|\mathcal{D}_A - \mathcal{U}\| = \Omega(1)$, with $1 - o(1)$ probability over a Haar-random $A \in \mathbb{C}^{m \times n}$. Or equivalently, that the number of samples needed to distinguish $\mathcal{D}_A$ from $\mathcal{U}$ (information-theoretically and with constant bias) is a constant, independent of n.

Let $X = (x_{ij}) \in \mathbb{C}^{n \times n}$ be a matrix of iid Gaussians, drawn from $\mathcal{N} = \mathcal{N}(0,1)_{\mathbb{C}}^{n \times n}$. Thus, $\mathcal{N}$ has the pdf

$$f_{\mathcal{N}}(X) = \frac{1}{\pi^{n^2}} \exp\left(-\sum_{ij} |x_{ij}|^2\right).$$

Also, let

$$P = P(X) = \frac{|\text{Per}(X)|^2}{n!},$$

so that $\text{E}[P] = 1$, and let f_P be the pdf of P. (We will typically suppress dependence on n.) Our goal, in this section, is to understand some basic facts about the function f_P, ignoring issues of computational complexity, and then use those facts to prove that $\mathcal{D}_A$ is not close to the uniform distribution. For a plot of f_P and the corresponding pdf for the determinant in the case $n = 6$ (reproduced from [1]), see Figure 1.

The first step is to give a characterization of f_P that, while easy to prove (it involves considering only the topmost row of the matrix X), will provide a surprising amount of leverage.

Lemma 3 *There exists a random variable $c > 0$ such that*

$$f_P(x) = \text{E}_c\left[ce^{-cx}\right].$$

In other words, $P(X)$ is a (possibly continuous) mixture of exponentially-distributed random variables.

Proof. Given $X = (x_{ij}) \in \mathbb{C}^{n \times n}$, let $X_1, \ldots, X_n$ be the bottom $(n-1) \times (n-1)$ minors. Then by definition,

$$\text{Per}(X) = x_{11}\,\text{Per}(X_1) + \cdots + x_{1n}\,\text{Per}(X_n).$$

Now, $\text{Per}(X_1), \ldots, \text{Per}(X_n)$ have some (correlated) probability measure, call it $\mathcal{C}$. Then we can think of $\text{Per}(X)$ as simply $c_1 x_1 + \cdots + c_n x_n$, where $x_1, \ldots, x_n$ are independent $\mathcal{N}(0,1)_{\mathbb{C}}$ Gaussians, and $(c_1, \ldots, c_n)$ is drawn from $\mathcal{C}$. This, in turn, is a complex Gaussian with mean 0 and variance $|c_1|^2 + \cdots + |c_n|^2$. Therefore the pdf of $\text{Per}(X)$ must be a convex combination of complex Gaussians, each with mean 0 (but with different variances). So the pdf of $|\text{Per}(X)|^2$ is a convex combination of absolute squares of complex Gaussians, which are exponentially-distributed random variables. ■

As a side note, since $1/c$ is the expectation of the random variable with pdf ce^{-cx}, we have

$$\text{E}\left[\frac{1}{c}\right] = \text{E}_c\left[\int_0^\infty ce^{-cx} x dx\right] = \int_0^\infty f_P(x)\, x dx = 1.$$

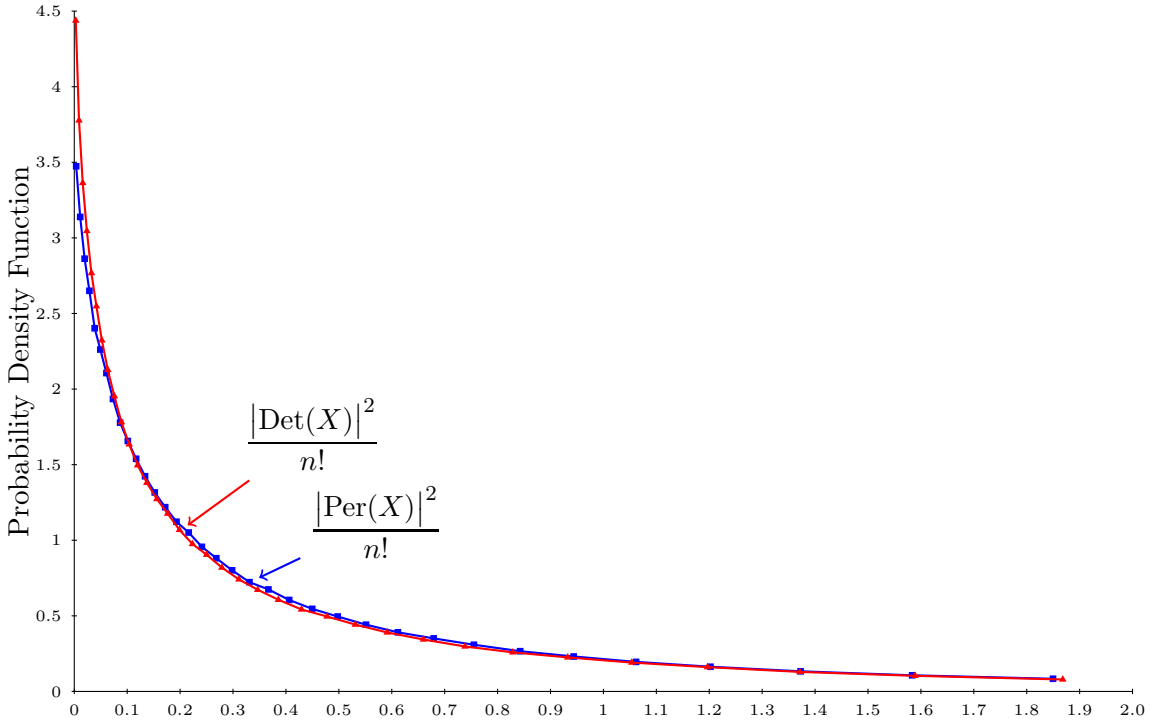

Figure 1: Probability density functions of the random variables $D = |\mathrm{Det}(X)|^2/n!$ and $P = |\mathrm{Per}(X)|^2/n!$, where $X \sim \mathcal{N}(0,1)_{\mathbb{C}}^{n\times n}$ is a complex Gaussian random matrix, in the case $n = 6$ (reproduced from [1]). Note that $\mathrm{E}[D] = \mathrm{E}[P] = 1$. As n increases, the bends on the left become steeper.

To illustrate the usefulness of Lemma 3, we now apply it to deduce various interesting facts about f_P, explaining what were left in [1] as empirical observations. The following theorem is not needed for later results in this section, but is included for completeness.

Theorem 4 *f_P is monotonically decreasing, finite, positive, and smooth, except that it might diverge at $x = 0$.*

Proof. For monotonically decreasing and positive, it suffices to note that f_P is a convex combination of functions with those properties. For finite,

$$f_P(x) = \mathrm{E}_c\left[ce^{-cx}\right] \leq \sup_{c>0} ce^{-cx} \leq \frac{1}{ex}$$

for $x > 0$. Likewise for continuous,

$$\begin{aligned}
\frac{df_P(x)}{dx} &= \lim_{\varepsilon\to 0} \frac{f_P(x+\varepsilon) - f_P(x-\varepsilon)}{2\varepsilon} \\
&= \lim_{\varepsilon\to 0} \frac{\mathrm{E}_c\left[ce^{-c(x+\varepsilon)} - ce^{-c(x-\varepsilon)}\right]}{2\varepsilon} \\
&= \lim_{\varepsilon\to 0} \mathrm{E}_c\left[\frac{ce^{-cx}\left(e^{-c\varepsilon} - e^{c\varepsilon}\right)}{2\varepsilon}\right] \\
&= \mathrm{E}_c\left[-c^2e^{-cx}\right]
\end{aligned}$$

which is finite for $x > 0$, and a similar calculation can be repeated for the higher derivatives. ■

We now use Lemma 3 to show, in two senses, that f_P is far from the point distribution concentrated on $P = 1$.

Lemma 5 *We have*

$$\frac{1}{2}\,\mathrm{E}\left[|P-1|\right] > 0.313, \qquad \Pr\left[|P-1| \geq \frac{1}{2}\right] > 0.615.$$

Proof. For the first part: since f_P is a mixture of pdf's of the form ce^{-cx}, by convexity it suffices to lower-bound

$$\inf_{c>0} \frac{1}{2} \int_0^\infty ce^{-cx}\,|x-1|\,dx.$$

The expression inside the inf can be evaluated as

$$\frac{1}{2} - \frac{1}{2c} + \frac{1}{ce^c},$$

for which a minimum of ~ 0.313 at $c \approx 1.678$ can be obtained numerically.

For the second part: again appealing to convexity, it suffices to lower-bound

$$\inf_{c>0} \left(1 - \int_{1/2}^{3/2} ce^{-cx}dx\right),$$

for which a minimum of $1 - \frac{2\sqrt{3}}{9} > 0.615$ is obtained at $c = \ln 3$. ■

Next, let $\mathcal{H}$ be the distribution over $X \in \mathbb{C}^{n \times n}$ defined by the pdf

$$f_{\mathcal{H}}(X) = f_{\mathcal{N}}(X) P(X).$$

In other words, $\mathcal{H}$ is the distribution over $n \times n$ matrices obtained by starting from the Gaussian distribution $\mathcal{N}$, then rescaling all the probabilities by $|\mathrm{Per}(X)|^2$. Then as an immediate corollary of Lemma 5, we find that

$$\|\mathcal{H} - \mathcal{N}\| = \frac{1}{2} \int_X |f_{\mathcal{H}}(X) - f_{\mathcal{N}}(X)| \, dX = \frac{1}{2} \mathrm{E}[|P - 1|] > 0.313. \tag{3}$$

Now let $A \in \mathbb{C}^{m \times n}$ be a Haar-random column-orthonormal matrix, and recall that A_S is the $n \times n$ submatrix of A corresponding to experimental outcome $S \in \Lambda_{m,n}$. Let $\mathcal{H}'$ be the distribution over $n \times n$ matrices $X = \sqrt{m} A_S$, where S was drawn from the BosonSampling distribution $\mathcal{D}_A$ (averaged over all A). Likewise, let $\mathcal{N}'$ be the distribution over $X = \sqrt{m} A_S$ where S was drawn from the uniform distribution $\mathcal{U}$. (Note that by symmetry, we could have just as well set $X = \sqrt{m} A_S$ for any *fixed* $S \in \Lambda_{m,n}$—say, the lexicographically first S.)

Then based on the results in [1], we might guess that $\mathcal{N}' \approx \mathcal{N}$ and $\mathcal{H}' \approx \mathcal{H}$: that is, a random $n \times n$ submatrix of a Haar-random A should look close to an iid Gaussian matrix, while a random submatrix whose probability was scaled by $|\mathrm{Per}|^2$ should look close to a $|\mathrm{Per}|^2$-scaled Gaussian matrix. Thus, line (3) strongly suggests that

$$\left\|\mathcal{H}' - \mathcal{N}'\right\| \approx \|\mathcal{H} - \mathcal{N}\| = \Omega(1).$$

Or in words: it should be easy to tell (information-theoretically and with constant bias) whether an $n \times n$ submatrix A_S was uniformly sampled or "BosonSampled" from among the $\binom{m}{n}$ submatrices of a Haar-random A, *just by examining A_S itself* (and not even knowing S or the rest of A). Intuitively, this is because a BosonSampled A_S will tend to "stick out" by having an unusually large $|\mathrm{Per}(A_S)|^2$.

As a consequence, one also expects that, with high probability over A, the BosonSampling distribution $\mathcal{D}_A$ should have $\Omega(1)$ variation distance from the uniform distribution $\mathcal{U}$. For once we decide whether the submatrix A_S was drawn from $\mathcal{H}'$ or from $\mathcal{N}'$, that should then tell us whether S itself was drawn from $\mathcal{D}_A$ or $\mathcal{U}$.

Unfortunately, there are two technical difficulties in formalizing the above. The first is that, if f_P (i.e., the pdf of $|\mathrm{Per}(X)|^2 / n!$) were *extremely* heavy-tailed—and we can't currently prove that it isn't—then $\mathcal{H}'$ wouldn't need to be close to $\mathcal{H}$ in variation distance. Intuitively, the rare matrices X such that $f_{\mathcal{N}'}(X)$ was far from $f_{\mathcal{N}}(X)$ could "acquire an outsized importance" for the scaled distributions $\mathcal{H}'$ and $\mathcal{H}$, if such X's had enormous permanents. The second difficulty is that, if we want to show that $\mathcal{D}_A$ is far from $\mathcal{U}$ with $1 - o(1)$ probability over A (rather than merely $\Omega(1)$ probability), then we need to say something about the lack of strong correlations between *different* submatrices $A_S, A_{S'}$ of the same A. For otherwise, it might be that all the A_S's that caused $\mathcal{H}'$ and $\mathcal{N}'$ to have constant variation distance from each other, were concentrated within (say) 50% of the A's.

Fortunately, we can overcome these difficulties, with help from a result proved in [1].

Theorem 6 ([1]) *Let $m \geq \frac{n^5}{\delta} \log^2 \frac{n}{\delta}$ for any $\delta > 0$. Then $\|\mathcal{N}' - \mathcal{N}\| = O(\delta)$.*

As discussed in [1], we are confident that the n^5 in Theorem 6 is purely an artifact of the proof, and that it can be improved to n^2. (Indeed, Jiang [9] *did* prove an analogue of Theorem 6 assuming only $m \gg n^2$, except for real orthogonal matrices rather than unitary matrices, and without the explicit dependence on δ.)

In any case, by combining Theorem 6 with the second part of Lemma 5, we immediately obtain the following.

Corollary 7 *Assume $m \geq \frac{n^5}{\delta} \log^2 \frac{n}{\delta}$. Then*

$$\Pr_{X \sim \mathcal{N}'} \left[|P(X) - 1| \geq \frac{1}{2} \right] \geq 0.615 - O(\delta).$$

Given a BosonSampling matrix $A \in \mathbb{C}^{m \times n}$, let $\mathcal{B}_A$ be the (discrete) distribution over matrices $X \in \mathbb{C}^{n \times n}$ obtained by first drawing S from $\mathcal{D}_A$, and then setting $X := \sqrt{m} A_S$. Then next we need a lemma that upper-bounds the probability, over A, that $\mathcal{B}_A$ looks very different from $\mathcal{H}'$ with respect to a certain statistical test.

Lemma 8 *Let $A \in \mathbb{C}^{m \times n}$ be a Haar-random* BosonSampling *matrix with $m \geq n^{5.1}/\delta$. Also, let*

$$W_A = \Pr_{S \in \Lambda_{m,n}} \left[|m^n P(A_S) - 1| \geq \frac{1}{2} \right]$$

(recalling that $P(X) = |\mathrm{Per}(X)|^2 / n!$). Then

$$\Pr_A \left[W_A \leq \frac{1}{2} \right] = O\left(\frac{1}{n} \right).$$

Or rewriting, we have

$$\Pr_{S \in \Lambda_{m,n}} \left[\left| \frac{m^n}{n!} |\mathrm{Per}(A_S)|^2 - 1 \right| \geq \frac{1}{2} \right] \geq \frac{1}{2}$$

with probability $1 - O(\delta)$ over A.

Proof. For all $S \in \Lambda_{m,n}$, let $w_S := 1$ if

$$|m^n P(A_S) - 1| \geq \frac{1}{2}$$

and $w_S := 0$ otherwise. Then clearly

$$W_A = \frac{1}{|\Lambda_{m,n}|} \sum_{S \in \Lambda_{m,n}} w_S.$$

Hence

$$\begin{aligned} \mathop{\mathrm{E}}_A [W_A] &= \frac{1}{|\Lambda_{m,n}|} \sum_{S \in \Lambda_{m,n}} \mathop{\mathrm{E}}_A [w_S] \\ &= \frac{1}{|\Lambda_{m,n}|} \sum_{S \in \Lambda_{m,n}} \Pr_{X \sim \mathcal{N}'} \left[|P(X) - 1| \geq \frac{1}{2} \right] \\ &\geq 0.615 - O(\delta), \end{aligned}$$

where the second line uses the definition of $\mathcal{N}'$ and the third line uses Corollary 7.

Now consider *two* experimental outcomes, $S, T \in \Lambda_{m,n}$. Notice that, if S and T are disjoint ($S \cap T = \varnothing$), then $A_{S \cup T}$ is simply a $2n \times n$ submatrix of A. So, if we think of A as an $m \times n$ submatrix of a Haar-random $m \times 2n$ matrix A', then $A_{S \cup T}$ is a submatrix of a $2n \times 2n$ submatrix of A'. But this means that, if we set $n' := 2n$ (which has no effect on the asymptotics), then we can apply Theorem 6 to $A_{S \cup T}$ exactly as if it were an $n \times n$ submatrix. So in particular, $A_{S \cup T}$ will be $O(\delta)$-close in variation distance to a $2n \times n$ matrix of iid Gaussians with mean 0 and variance $1/m$—or in other words, to two independent samples from $\mathcal{N}'$.

We can use the above considerations to upper-bound the variance of W_A:

$$\begin{aligned}
\operatorname*{Var}_A [W_A] &= \frac{1}{|\Lambda_{m,n}|^2} \sum_{S,T \in \Lambda_{m,n}} \left(\operatorname*{E}_A [w_S w_T] - \operatorname*{E}_A [w_S] \operatorname*{E}_A [w_T] \right) \\
&\leq O(\delta) + \frac{1}{|\Lambda_{m,n}|^2} \sum_{S,T \in \Lambda_{m,n} : S \cap T \neq \varnothing} 1 \\
&\leq \frac{n^2}{m} + O(\delta).
\end{aligned}$$

Here we used the facts that $\operatorname{E}_A [w_S w_T] \leq 1$, and that $\Pr_{S,T \in \Lambda_{m,n}} [S \cap T \neq \varnothing] \leq n^2/m$ by the union bound.

Combining and using Chebyshev's inequality,

$$\begin{aligned}
\Pr_A \left[W_A < \frac{1}{2} \right] &< \frac{\operatorname{Var}_A [W_A]}{\left(\operatorname{E}_A [W_A] - 1/2 \right)^2} \\
&\leq \frac{n^2/m + O(\delta)}{\left(0.615 - O(\delta) - 1/2 \right)^2} \\
&= O\left(\frac{n^2}{m} + \delta \right) \\
&= O(\delta),
\end{aligned}$$

where the last line used that $m \geq n^{5.1}/\delta$. ■

Once again, we are confident that the condition $m \geq n^{5.1}/\delta$ in Lemma 8 is not tight; indeed, one should be able to get something whenever $m \geq n^2/\delta$.

The final ingredient in the proof is a simple fact about variation distance.

Lemma 9 *Let $\mathcal{U}$ be the uniform distribution over a finite set X, and let $\mathcal{D} = \{p_x\}_{x \in X}$ be some other distribution. Also, let $Z \subseteq X$ satisfy $|Z| \geq (1-\alpha)|X|$, and let $M \in [(1-\beta)|X|, |X|]$. Then*

$$\|\mathcal{D} - \mathcal{U}\| \geq \frac{1-\alpha}{4} \Pr_{x \in Z} \left[|M p_x - 1| \geq \frac{1}{2} \right] - \frac{\beta}{2 - 2\beta}.$$

Proof. We have

$$\begin{aligned}
\|\mathcal{D}-\mathcal{U}\| &= \frac{1}{2}\sum_{x\in X}\left|p_x-\frac{1}{|X|}\right| \\
&\geq \frac{1}{2}\sum_{x\in Z}\left|p_x-\frac{1}{|X|}\right| \\
&\geq \frac{1}{2}\sum_{x\in Z}\left(\left|p_x-\frac{1}{M}\right|-\left(\frac{1}{M}-\frac{1}{|X|}\right)\right) \\
&= \left(\frac{1}{2M}\sum_{x\in Z}\left|Mp_x-1\right|\right)-\frac{\beta}{1-\beta}\left(\frac{|Z|}{2\,|X|}\right) \\
&\geq \frac{|Z|}{4M}\Pr_{x\in Z}\left[|Mp_x-1|\geq\frac{1}{2}\right]-\frac{\beta}{1-\beta}\left(\frac{|Z|}{2\,|X|}\right) \\
&\geq \frac{1-\alpha}{4}\Pr_{x\in Z}\left[|Mp_x-1|\geq\frac{1}{2}\right]-\frac{\beta}{2-2\beta}
\end{aligned}$$

■

Combining Lemmas 8 and 9 now yields the main result of the section.

Theorem 10 *Let $A\in\mathbb{C}^{m\times n}$ be a Haar-random* BosonSampling *matrix with $m\geq n^{5.1}/\delta$. Let $\mathcal{U}$ be the uniform distribution over $\Lambda_{m,n}$, and let $\mathcal{D}_A$ be the* BosonSampling *distribution corresponding to A. Then for sufficiently large n and with probability $1-O(\delta)$ over A,*

$$\|\mathcal{D}_A-\mathcal{U}\|\geq\frac{1}{9}.$$

Proof. In Lemma 9, set

$$X=\Phi_{m,n},\qquad Z=\Lambda_{m,n},\qquad M=\frac{m^n}{n!}.$$

Then $|X|=\binom{m+n-1}{n}$ and $|Z|=\binom{m}{n}$, which means that we can set $\alpha:=n^2/m$ and $\beta:=n^2/m$. So using Lemma 9 and then Lemma 8, we find that, with probability $1-O(\delta)$ over A:

$$\begin{aligned}
\|\mathcal{D}_A-\mathcal{U}\| &\geq \frac{1-\alpha}{4}\Pr_{x\in Z}\left[|Mp_x-1|\geq\frac{1}{2}\right]-\frac{\beta}{2-2\beta} \\
&= \frac{1-n^2/m}{4}\Pr_{S\in\Lambda_{m,n}}\left[\left|\frac{m^n}{n!}\left|\mathrm{Per}\left(A_S\right)\right|^2-1\right|\geq\frac{1}{2}\right]-\frac{n^2/m}{2-2n^2/m} \\
&\geq \frac{1-n^2/m}{8}-\frac{n^2/m}{2-2n^2/m} \\
&\geq \frac{1}{8}-O\left(\frac{n^2}{m}\right) \\
&\geq \frac{1}{9}
\end{aligned}$$

for sufficiently large n. ■

7.1 Weakly-Symmetric Verifiers

In this subsection, we use Theorem 10 to justify a claim made in Section 5: namely, that it's possible to distinguish a generic BosonSampling distribution $\mathcal{D}_A$ from the uniform distribution $\mathcal{U}$, even if we restrict ourselves to "weakly-symmetric" verifiers (which don't know the labels of the input and output modes).

The first step is to observe the following corollary of Theorem 10.

Corollary 11 *Let $A \in \mathbb{C}^{m \times n}$ be a Haar-random* BosonSampling *matrix with $m \geq n^{5.1}/\delta$. Then there exists a proper verifier V_A (not necessarily computationally efficient) and a constant $c > 1$ such that, with probability $1 - O(\delta)$ over A, the following holds:*

$$\Pr_{S_1,\ldots,S_k \sim \mathcal{D}_A} [V_A(S_1,\ldots,S_k) \ \ accepts] \geq 1 - \frac{1}{c^k}, \tag{4}$$

$$\Pr_{S_1,\ldots,S_k \sim \mathcal{U}} [V_A(S_1,\ldots,S_k) \ \ accepts] \leq \frac{1}{c^k}. \tag{5}$$

Moreover, if (4) and (5) hold for a given A, then they also hold for any $P_\sigma A P_\tau$, where P_σ and P_τ are the permutation matrices corresponding to $\sigma \in S_m$ and $\tau \in S_n$ respectively.

Proof. The inequalities (4) and (5) follow immediately from Theorem 10, together with a basic "amplification" property of variation distance: namely that, by a Chernoff bound,

$$\left\| \mathcal{D}_1^{\otimes k} - \mathcal{D}_2^{\otimes k} \right\| > 1 - \frac{1}{\exp\left(\Omega\left(k \cdot \|\mathcal{D}_1 - \mathcal{D}_2\|_1^2\right)\right)}.$$

If one wishes to be more explicit, the verifier $V_A(S_1,\ldots,S_k)$ should accept if and only if

$$\prod_{i=1}^{k} |\mathrm{Per}(A_{S_i})|^2 \geq \left(\frac{n!}{m^n}\right)^k.$$

For the last part, simply observe that, if we take V_A as above, then $\Pr_{S_1,\ldots,S_k \sim \mathcal{D}_A}[V_A \text{ accepts}]$ and $\Pr_{S_1,\ldots,S_k \sim \mathcal{U}}[V_A \text{ accepts}]$ are both completely unaffected by permutations of the rows and columns of A. ■

Using Corollary 11, we can easily construct a weakly-symmetric verifier.

Theorem 12 *Let $A \in \mathbb{C}^{m \times n}$ be a Haar-random* BosonSampling *matrix with $m \geq n^{5.1}/\delta$, and let $k = O\left(\log \frac{n!m!}{\Delta}\right)$. Then there exists a* ***weakly-symmetric*** *proper verifier V_A^* (not necessarily computationally efficient) such that, with probability $1 - O(\delta)$ over A, the following holds:*

$$\Pr_{S_1,\ldots,S_k \sim \mathcal{D}_A} [V_A^*(S_1,\ldots,S_k) \ \ accepts] \geq 1 - \Delta,$$

$$\Pr_{S_1,\ldots,S_k \sim \mathcal{U}} [V_A^*(S_1,\ldots,S_k) \ \ accepts] \leq \Delta.$$

Proof. Let $V_A^*(S_1,\ldots,S_k)$ accept if and only if there *exist* permutations $\sigma \in S_m$ and $\tau \in S_n$ that cause $V_{P_\sigma A P_\tau}(S_1,\ldots,S_k)$ to accept, where V_A is the verifier from Corollary 11. Then clearly V_A^*

is weakly-symmetric: i.e., it behaves identically on A and $P_\sigma A P_\tau$, for any pair $(\sigma, \delta) \in S_m \times S_n$. Moreover, by Corollary 11 together with the union bound, we have

$$\Pr_{S_1,\ldots,S_k \sim \mathcal{D}_A} [V_A^* (S_1, \ldots, S_k) \text{ rejects}] \leq n!m! \left(\frac{1}{c^k}\right) \leq \Delta,$$

$$\Pr_{S_1,\ldots,S_k \sim \mathcal{U}} [V_A^* (S_1, \ldots, S_k) \text{ accepts}] \leq n!m! \left(\frac{1}{c^k}\right) \leq \Delta$$

with probability $1 - O(\delta)$ over A. ■

8 Detecting Deviations from Uniformity

In Section 7, we showed that most BosonSampling distributions $\mathcal{D}_A$ have constant variation distance from the uniform distribution $\mathcal{U}$ over experimental outcomes S (and furthermore, that this is true even if we restrict to collision-free outcomes $S \in \Lambda_{m,n}$). However, we did not show how to distinguish $\mathcal{D}_A$ from $\mathcal{U}$ using a polynomial-time algorithm. Moreover, the distinguishing procedure V_A from Corollary 11 involved computing $|\text{Per}(A_S)|^2$ for each experimental outcome S. And just as Gogolin et al. [7] asserted, we do *not* expect $|\text{Per}(A_S)|^2$ to computable (or even closely approximable) in polynomial time—at least, to whatever extent we expect BosonSampling to be a hard problem in the first place.[6]

Nevertheless, in this section we will show that $\mathcal{D}_A$ *can* be distinguished in polynomial time from $\mathcal{U}$—and moreover, quite easily. As we will discuss in Section 8.1, one interpretation of this result is that $\mathcal{D}_A$ versus $\mathcal{U}$ was simply the "wrong comparison": instead, one should compare $\mathcal{D}_A$ to various more interesting "mockup distributions," which are easy to sample classically but at least encode *some* information about A. On the other hand, our result suffices to refute the claim of Gogolin et al. [7], about $\mathcal{D}_A$ looking just like the uniform distribution to any polynomial-time algorithm.

Given a matrix $X \in \mathbb{C}^{n \times n}$, let $R_i(X)$ be the squared 2-norm of X's i^{th} row:

$$R_i = R_i(X) = |x_{i1}|^2 + \cdots + |x_{in}|^2.$$

Clearly $\text{E}[R_i] = \text{Var}[R_i] = n$ if $X \sim \mathcal{N}$ is Gaussian. Indeed, it is not hard to see that R_i is a χ^2 random variable with $2n$ degrees of freedom, multiplied by the scalar $1/2$. Its pdf is

$$f_{R_i}(r) = \frac{e^{-r} r^{n-1}}{(n-1)!}.$$

Throughout this paper, we will refer to such a random variable as a *complex χ^2 variable with n degrees of freedom*.

Next, let R be the product of the squared row-norms:

$$R = R(X) = R_1 \cdots R_n.$$

Note that if $X \sim \mathcal{N}$ is Gaussian, then by the independence of the rows,

$$\text{E}[R] = \text{E}[R_1] \cdots \text{E}[R_n] = n^n.$$

[6] Indeed, we even conjecture that this problem is *nonuniformly* hard: that is, for a Haar-random $A \in \mathbb{C}^{m \times n}$, we conjecture that there is no polynomial-size circuit C_A that computes or closely approximates $|\text{Per}(A_S)|^2$ given $S \in \Lambda_{m,n}$ as input.

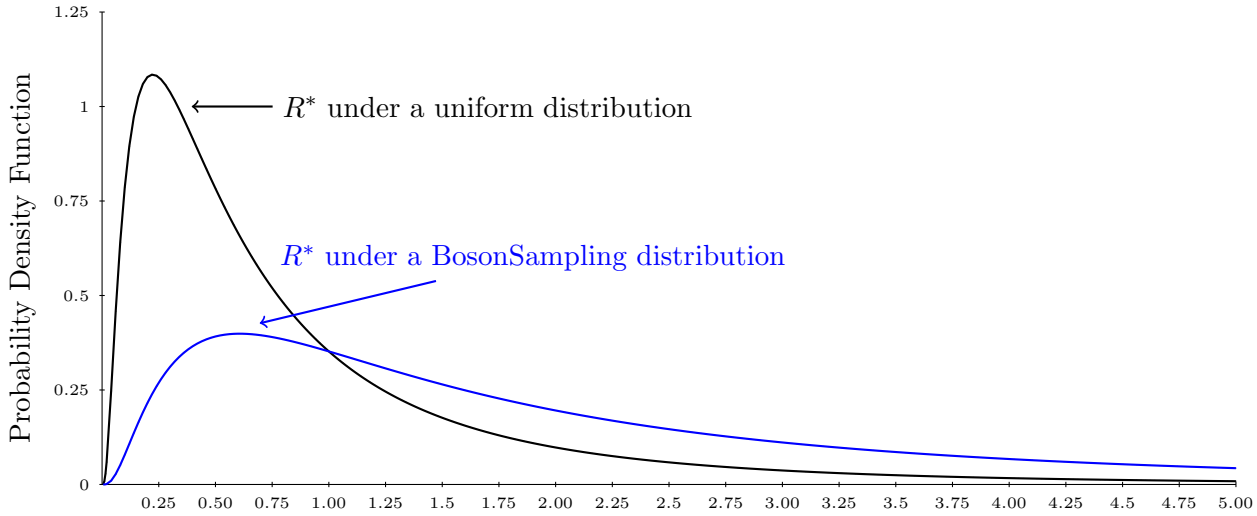

Figure 2: Probability density functions for the row-norm estimator $R^*(A_S)$, when S is drawn either from the uniform distribution $\mathcal{U}$ (in which case $R^*(A_S)$ quickly converges to a lognormal random variable), or from a Haar-random BosonSampling distribution $\mathcal{D}_A$, in the limits $n \rightarrow \infty$ and $m/n \rightarrow \infty$. One can see that R^* is typically larger under $\mathcal{D}_A$, and that R^* suffices to distinguish $\mathcal{D}_A$ from $\mathcal{U}$ with constant bias.

Thus, it will be convenient to define

$$R^* = \frac{R}{n^n}$$

so that $\mathrm{E}[R^*] = 1$. Of course, R^* can be computed in linear (i.e., $O(n^2)$) time given $X \in \mathbb{C}^{n \times n}$ as input.

Our claim is that *computing $R^*(A_S)$, for a few experimental outcomes $S \in \Lambda_{m,n}$, already suffices to tell whether S was drawn from $\mathcal{D}_A$ or from $\mathcal{U}$, with high probability.* The intuition for this claim is simple: first, each R_i is at least *slightly* correlated with $|\mathrm{Per}(X)|^2$, since multiplying the i^{th} row of X by any scalar c also multiplies $\mathrm{Per}(X)$ by c. Second, if X is Gaussian, then $R_i(X)/n$ will typically be $1 \pm \Theta(1/\sqrt{n})$. This suggests that the *product* of the $R_i(X)/n$'s, over all $i \in [n]$, will typically be $1 \pm \Theta(1)$: or in other words, that R^* will have constant-sized fluctuations. If so, then $R^*(X)$ should be an easy-to-compute random variable, whose fluctuations nontrivially correlate with the fluctuations in $|\mathrm{Per}(X)|^2$. Therefore $R^*(A_S)$ should be larger, in expectation, if S was drawn from $\mathcal{D}_A$ than if S was drawn from $\mathcal{U}$. (See Figure 2 for plots of expected distribution over R^*, both for a Haar-random $\mathcal{D}_A$ and for $\mathcal{U}$.)

Before formalizing these intuitions, we make two remarks. First, $R^*(X)$ does not yield a *good* approximation to $|\mathrm{Per}(X)|^2$—nor should we expect it to, if we believe (following [1]) that approximating $|\mathrm{Per}(X)|^2$, with $1 - 1/\mathrm{poly}(n)$ probability and to within $\pm n!/\mathrm{poly}(n)$ error, is a #P-hard problem! Instead, $R^*(X)$ is "just barely" correlated with $|\mathrm{Per}(X)|^2$. Still, we will show that this correlation is already enough to distinguish $\mathcal{D}_A$ from $\mathcal{U}$ with constant bias. Second, if we were satisfied to distinguish $\mathcal{D}_A$ from $\mathcal{U}$ with $1/\mathrm{poly}(n)$ bias, then it would even suffice to look at $R_i(X)$, for any single row i. Indeed, we conjecture that it would even suffice to look at $|x_{ij}|^2$, for any single *entry* x_{ij} of X! However, looking at the product of the R_i's seems necessary if we want a constant-bias distinguisher.

We now prove that looking at $R^*(A_S)$ indeed distinguishes $\mathcal{D}_A$ from $\mathcal{U}$, with high probability over $A \in \mathbb{C}^{m \times n}$. As in Section 7, let

$$P = P(X) = \frac{|\mathrm{Per}(X)|^2}{n!}.$$

Then let

$$Q = Q(X) = \frac{P(X)}{R^*(X)}.$$

The following proposition, though trivial to prove, is crucial to what follows.

Proposition 13 *$Q(X)$ and $R^*(X)$ are independent random variables, if $X \sim \mathcal{N}(0,1)_{\mathbb{C}}^{n \times n}$ is Gaussian. Indeed, we can think of Q as simply $|\mathrm{Per}(Y)|^2/n!$, where $Y \in \mathbb{C}^{n \times n}$ is a matrix each of whose rows is an independent Haar-random vector, with squared 2-norm equal to n. So in particular, $\mathrm{E}[Q] = 1$.*

Proof. Suppose we first choose Y as above, and then choose $R_1, \ldots, R_n$ as independent complex χ^2 random variables with n degrees of freedom. Then X, a Gaussian random matrix, can be obtained from Y by simply multiplying the i^{th} row of Y by $\sqrt{R_i/n}$, for all $i \in [n]$. So we can decompose P as

$$P = \frac{|\mathrm{Per}(X)|^2}{n!} = \frac{|\mathrm{Per}(Y)|^2}{n!} \frac{R_1 \cdots R_n}{n^n} = QR^*$$

where $Q = |\mathrm{Per}(Y)|^2/n!$ and $R^* = R_1 \cdots R_n/n^n$ are independent by construction. ■

As in Section 7, let $\mathcal{N} = \mathcal{N}(0,1)_{\mathbb{C}}^{n \times n}$ be the Gaussian distribution over $X \in \mathbb{C}^{n \times n}$, let $\mathcal{H}$ be the Gaussian distribution scaled by $|\mathrm{Per}(X)|^2$, and let $f_{\mathcal{N}}(X)$ and $f_{\mathcal{H}}(X) = f_{\mathcal{N}}(X) P(X)$ be the pdfs of $\mathcal{N}$ and $\mathcal{H}$ respectively. Then the following lemma says that, provided $\mathrm{E}_{\mathcal{N}}[|R^*(X) - 1|]$ is large, we can indeed use $R^*(X)$ to distinguish the case $X \sim \mathcal{H}$ from the case $X \sim \mathcal{N}$:

Lemma 14

$$\Pr_{\mathcal{H}}[R^* \geq 1] - \Pr_{\mathcal{N}}[R^* \geq 1] = \frac{1}{2} \mathrm{E}_{\mathcal{N}}[|R^* - 1|].$$

Proof. We have

$$\begin{aligned}
\Pr_{\mathcal{H}}[R^*(X) \geq 1] - \Pr_{\mathcal{N}}[R^*(X) \geq 1] &= \int_{X : R^*(X) \geq 1} (f_{\mathcal{H}}(X) - f_{\mathcal{N}}(X))\, dX \\
&= \int_{X : R^*(X) \geq 1} (f_{\mathcal{N}}(X) P(X) - f_{\mathcal{N}}(X))\, dX \\
&= \int_{X : R^*(X) \geq 1} f_{\mathcal{N}}(X) (Q(X) R^*(X) - 1)\, dX \\
&= \int_{X : R^*(X) \geq 1} f_{\mathcal{N}}(X) (R^*(X) - 1)\, dX. \qquad (6)
\end{aligned}$$

Here line (6) used the fact that Q and R^* are independent, and hence that

$$\mathrm{E}[Q | R^* = r] = \mathrm{E}[Q] = 1$$

regardless of the value of R^*.

Now, since $\mathrm{E}[R^*] = 1$, the above also equals

$$\int_{X:R^*(X)<1} f_{\mathcal{N}}(X)\left(1 - R^*(X)\right) dX.$$

But this means that it's simply

$$\frac{1}{2}\int_X f_{\mathcal{N}}(X)\left|R^*(X) - 1\right| dX = \frac{1}{2}\mathop{\mathrm{E}}_{\mathcal{N}}\left[\left|R^* - 1\right|\right].$$

■

Thus, we arrive the remarkable result that, to understand the ability of our row-norm estimator R^* to distinguish the distributions $\mathcal{H}$ and $\mathcal{N}$, *we need only understand the intrinsic variation of* R^*. All other information about $\mathrm{Per}(X)$ is irrelevant. (Of course, this result depends crucially on the independence of Q and R^*.)

Better yet, $R = R_1 \cdots R_n$ is a product of n iid random variables—which means that, with a little work, we can indeed understand the distribution of R. As a first step, let

$$L := \ln R.$$

Then L, of course, is a *sum* of n iid random variables,

$$L = \ln R_1 + \cdots + \ln R_n.$$

Indeed, let ℓ_n be the log of a complex χ^2 variable with n degrees of freedom: that is,

$$\ell_n = \ln\left(|x_1|^2 + \cdots + |x_n|^2\right),$$

where $x_j \sim \mathcal{N}(0,1)_{\mathbb{C}}$ for all $j \in [n]$. Then each $\ln R_i$ is distributed as ℓ_n, so L is simply the sum of n independent ℓ_n random variables. As such, we might expect L to be close to Gaussian. The Berry-Esseen Theorem (which we state for completeness) assures us that this is indeed the case, provided the ℓ_n's satisfy certain conditions.

Theorem 15 (Berry-Esseen Theorem) *Let $Z_1, \ldots, Z_n$ be real iid random variables satisfying*

$$\begin{aligned}\mathrm{E}[Z_i] &= \upsilon,\\ \mathrm{E}\left[(Z_i - \upsilon)^2\right] &= \sigma^2 > 0,\\ \mathrm{E}\left[|Z_i - \upsilon|^3\right] &= \rho < \infty.\end{aligned}$$

Then let

$$Z := Z_1 + \cdots + Z_n,$$

and let $W \sim \mathcal{N}\left(\upsilon n, \sigma^2 n\right)_{\mathbb{R}}$ be a real Gaussian with mean υn and variance $\sigma^2 n$. Then for all $x \in \mathbb{R}$,

$$\left|\Pr[Z \leq x] - \Pr[W \leq x]\right| \leq \frac{C\rho}{\sigma^3\sqrt{n}},$$

where C is some universal constant.

Thus, it remains only to check that the random variables ℓ_n satisfy the conditions for Theorem 15. We can do so using the following striking relations:

Lemma 16 *We have*

$$\begin{aligned}
\mathrm{E}\left[\ell_n\right] &= -\gamma + \sum_{j=1}^{n-1} \frac{1}{j} = \ln n - O\left(\frac{1}{n}\right), \\
\mathrm{Var}\left[\ell_n\right] &= \frac{\pi^2}{6} - \sum_{j=1}^{n-1} \frac{1}{j^2} = \frac{1+o\left(1\right)}{n}, \\
\mathrm{E}\left[\left(\ell_n - \mathrm{E}\left[\ell_n\right]\right)^4\right] &= 6\left(\frac{\pi^4}{90} - \sum_{j=1}^{n-1} \frac{1}{j^4}\right) + 3\,\mathrm{Var}\left[\ell_n\right]^2 = \frac{3+o\left(1\right)}{n^2},
\end{aligned}$$

where $\gamma = 0.577\ldots$ *is the Euler-Mascheroni constant.*

Proof Sketch. The first step is to rewrite the moments in terms of the so-called *cumulants* κ_k of ℓ_n, which are the coefficients of the log of ℓ_n's moment generating function:

$$\sum_{k=1}^{\infty} \kappa_k \frac{t^k}{k!} = \ln \mathrm{E}\left[e^{\ell_n t}\right] = \ln \mathrm{E}\left[\left(\left|x_1\right|^2 + \cdots + \left|x_n\right|^2\right)^t\right].$$

By standard facts about cumulants, we have

$$\begin{aligned}
\mathrm{E}\left[\ell_n\right] &= \kappa_1, \\
\mathrm{Var}\left[\ell_n\right] &= \kappa_2, \\
\mathrm{E}\left[\left(\ell_n - \mathrm{E}\left[\ell_n\right]\right)^4\right] &= \kappa_4 + 3\kappa_2^2.
\end{aligned}$$

The second step is to express the cumulants in terms of the so-called *polygamma function* $\phi^{(k)}$:

$$\phi^{(k)}\left(z\right) := \left(\frac{d}{dz}\right)^{k+1} \left(\ln \Gamma\left(z\right)\right).$$

One can show that

$$\kappa_k = \phi^{(k-1)}\left(n\right).$$

The third step is to note that, for positive integers n and $k \geq 2$, we have

$$\phi^{(k-1)}\left(n\right) = \left(-1\right)^k \left(k-1\right)! \sum_{j=n}^{\infty} \frac{1}{j^k} = \left(-1\right)^k \left(k-1\right)! \left(\zeta\left(k\right) - \sum_{j=1}^{n-1} \frac{1}{j^k}\right),$$

where $\zeta\left(k\right) = \sum_{j=1}^{\infty} \frac{1}{j^k}$ is the Riemann zeta function (which satisfies $\zeta\left(2\right) = \pi^2/6$ and $\zeta\left(4\right) = \pi^4/90$). On the other hand, noting that $\zeta\left(1\right)$ diverges, when $k = 1$ we instead have

$$\phi^{(0)}\left(n\right) = -\gamma + \sum_{j=1}^{n-1} \frac{1}{j}.$$

■

Note that, for any real random variable X, by convexity we have

$$\mathrm{E}\left[|X-\mathrm{E}[X]|^3\right] \le \mathrm{E}\left[(X-\mathrm{E}[X])^4\right]^{3/4}.$$

So one corollary of Lemma 16 is

$$\mathrm{E}\left[|\ell_n-\mathrm{E}[\ell_n]|^3\right] \le \frac{3^{3/4}+o(1)}{n^{3/2}}.$$

Thus, in the notation of Theorem 15, we have

$$\frac{\rho}{\sigma^3\sqrt{n}} = O\left(\frac{1/n^{3/2}}{(1/\sqrt{n})^3\sqrt{n}}\right) = O\left(\frac{1}{\sqrt{n}}\right).$$

This yields the following:

Theorem 17 *Let $L = \ln R$, and let L' be a real Gaussian with mean $\mathrm{E}[\ell_n]\cdot n$ and variance $\mathrm{Var}[\ell_n]\cdot n$. Then for all $x\in\mathbb{R}$,*

$$\left|\Pr[L\le x]-\Pr\left[L'\le x\right]\right| = O\left(\frac{1}{\sqrt{n}}\right).$$

Theorem 17 has the following corollary.

Corollary 18 *We have*

$$\Pr_{\mathcal{N}}\left[|R^*-1|\ge\frac{1}{2}\right] \ge 0.586 - O\left(\frac{1}{\sqrt{n}}\right), \qquad \frac{1}{2}\mathop{\mathrm{E}}_{\mathcal{N}}[|R^*-1|] \ge 0.146 - O\left(\frac{1}{\sqrt{n}}\right).$$

Proof. Letting L' be the real Gaussian from Theorem 17, observe that

$$\mathrm{E}\left[L'\right] = \mathrm{E}[\ell_n]\cdot n = \left(-\gamma+\sum_{j=1}^{n-1}\frac{1}{j}\right)\cdot n \in \left[n\ln n-\gamma, n\ln n-\frac{1}{2}\right]$$

and

$$\mathrm{Var}\left[L'\right] = \mathrm{Var}[\ell_n]\cdot n = \left(\frac{\pi^2}{6}-\sum_{j=1}^{n-1}\frac{1}{j^2}\right)\cdot n \ge 1.$$

So recalling that $R = n^n R^*$ and $L = \ln R$, we have

$$\begin{aligned}
\Pr_{\mathcal{N}}\left[|R^*-1|\ge\frac{1}{2}\right] &= \Pr_{\mathcal{N}}\left[R\le\frac{n^n}{2}\right] + \Pr_{\mathcal{N}}\left[R\ge\frac{3n^n}{2}\right] \\
&= \Pr_{\mathcal{N}}[L\le n\ln n-\ln 2] + \Pr_{\mathcal{N}}\left[L\ge n\ln n+\ln\frac{3}{2}\right] \\
&\ge \Pr\left[L'\le n\ln n-\ln 2\right] + \Pr\left[L'\ge n\ln n+\ln\frac{3}{2}\right] - O\left(\frac{1}{\sqrt{n}}\right) \\
&\ge \int_{-\infty}^{-\ln 2+1/2}\frac{e^{-x^2/2}}{\sqrt{2\pi}}dx + \int_{\ln 3/2+\gamma}^{\infty}\frac{e^{-x^2/2}}{\sqrt{2\pi}}dx - O\left(\frac{1}{\sqrt{n}}\right) \\
&\ge 0.586 - O\left(\frac{1}{\sqrt{n}}\right)
\end{aligned}$$

and

$$\frac{1}{2}\mathop{\mathrm{E}}_{\mathcal{N}}\left[\left|R^*-1\right|\right] \geq \frac{1}{4}\Pr_{\mathcal{N}}\left[\left|R^*-1\right| \geq \frac{1}{2}\right] \geq 0.146 - O\left(\frac{1}{\sqrt{n}}\right).$$

■

In particular, combining Lemma 14 with Corollary 18 gives

$$\Pr_{\mathcal{H}}\left[R^* \geq 1\right] - \Pr_{\mathcal{N}}\left[R^* \geq 1\right] \geq 0.146 - O\left(\frac{1}{\sqrt{n}}\right),$$

meaning that $R^*(X)$ does indeed distinguish the case $X \sim \mathcal{H}$ from the case $X \sim \mathcal{N}$ with constant bias.

Having established that, the last step is to deduce that, if A is a Haar-random BosonSampling matrix with $m \gg n$, then with high probability over A, the row-norm estimator $R^*(X)$ *also* distinguishes the case $S \sim \mathcal{D}_A$ from the case $S \sim \mathcal{U}$ with constant bias. However, this step is precisely parallel to the analogous step in Section 7: once again, we use the fact that an $n \times n$ submatrix of such an A is close in variation distance to an iid Gaussian matrix. Again, the main technical complications are that we want to restrict attention to collision-free S's only, and that we need to argue that different $n \times n$ submatrices of A are close to independent, in order to show that quantities such as $\Pr_{S \sim \mathcal{D}_A}\left[R^*(A_S) \geq 1\right]$ have small variances. If these complications are handled in precisely the same way as in Section 7, then we immediately obtain Theorem 1 from Section 3: namely, that asking whether $R^*(A_S) \geq 1$ suffices to distinguish $S \sim \mathcal{D}_A$ from $S \sim \mathcal{U}$ with constant bias, with probability $1 - O(\delta)$ over a Haar-random $A \in \mathbb{C}^{m \times n}$ ($m \geq n^{5.1}/\delta$) and for sufficiently large n.

8.1 Classical Mockup Distribution

We showed that the row-norm estimator R^* can distinguish a generic BosonSampling distribution $\mathcal{D}_A$ from the uniform distribution $\mathcal{U}$—but not, of course, that R^* distinguishes $\mathcal{D}_A$ from *any* classically-samplable distribution. And indeed, in this subsection we point out that there *are* natural distributions that are easy to sample classically, but that R^* fails to distinguish from $\mathcal{D}_A$.

Given a BosonSampling matrix $A \in \mathbb{C}^{m \times n}$, let $A^{\#}$ be the $m \times n$ matrix whose (i,j) entry is $|a_{ij}|^2$. Also, given $S = (s_1, \ldots, s_m) \in \Phi_{m,n}$, let $A_S^{\#}$ be the $n \times n$ matrix consisting of s_1 copies of $A^{\#}$'s first row, s_2 copies of $A^{\#}$'s second row, and so on (precisely analogous to A_S). Finally, let $\mathcal{M}_A$, the "classical mockup distribution for A," be the distribution over $\Phi_{m,n}$ defined by

$$\Pr_{\mathcal{M}_A}[S] = \frac{\mathrm{Per}\left(A_S^{\#}\right)}{s_1! \cdots s_m!}.$$

We claim that $\mathcal{M}_A$ is easy to sample in classical polynomial time (in fact, in $O(mn)$ time). To do so, for each $j := 1$ to n, just sample an index $h_j \in [m]$ from the probability distribution

$$\left(|a_{1j}|^2, \ldots, |a_{mj}|^2\right).$$

Then for all $i \in [m]$, let $s_i := \left|\{j : h_j = i\}\right|$, and output $(s_1, \ldots, s_m)$ as S. It is not hard to see that this algorithm will output a given $S = (s_1, \ldots, s_m)$ with probability exactly equal to $\mathrm{Per}\left(A_S^{\#}\right)/s_1! \cdots s_m!$ (the numerator comes from summing over all $n!$ possible permutations of h_j's

that yield S, while the denominator is the size of S's automorphism group). This also implies that $\mathcal{M}_A$ was a normalized probability distribution in the first place.

Note that there is a clear "physical" reason why $\mathcal{M}_A$ is samplable in classical polynomial time. Namely, $\mathcal{M}_A$ is just the distribution output by a BosonSampling device with matrix A, if the n input photons are all *distinguishable*—or equivalently, if they behave as classical particles rather than as bosons. In other words, $\mathcal{M}_A$ simply models n balls being thrown independently into m bins (possibly with a different distribution for each ball).

On the other hand, we now observe that *our row-norm estimator, $R^*(A_S)$, fails completely to distinguish $\mathcal{D}_A$ from its "classical mockup" $\mathcal{M}_A$.* The reason is just that $\Pr_{\mathcal{M}_A}[S]$ is correlated with $R^*(A_S)$ in exactly the same way that

$$\Pr_{\mathcal{D}_A}[S] = \frac{|\mathrm{Per}(A_S)|^2}{s_1! \cdots s_m!}$$

is correlated with $R^*(A_S)$. Indeed, *both* of these probabilities can be written as the product of $R^*(A_S)$ with a random variable (the permanent or absolute-squared permanent of a row-normalized matrix, respectively) that is independent of $R^*(A_S)$. As a result, any verification test only involving R^*—for example, accepting S if and only if $R^*(A_S) \geq 1$—will accept with exactly the same probability for $S \sim \mathcal{M}_A$ as for $S \sim \mathcal{D}_A$.

Note that there are other distributions, besides $\mathcal{M}_A$, with the same two properties: that they can be sampled in classical polynomial time, but that the row-norm estimator R^* fails completely to distinguish them from $\mathcal{D}_A$. One nice example is the FermionSampling distribution $\mathcal{F}_A$: see Appendix 13 for a detailed definition, as well as an $O\left(mn^2\right)$ classical sampling algorithm. In $\mathcal{F}_A$, the probabilities of collision-free outcomes $S \in \Lambda_{m,n}$ are given by $|\mathrm{Det}(A_S)|^2$ rather than $|\mathrm{Per}(A_S)|^2$. But since the determinant is affected by scaling of rows in exactly the same way as the permanent, it follows that $\mathcal{F}_A$ must satisfy the same row-norm statistics as $\mathcal{D}_A$.

Yet another example—and perhaps the simplest—is the distribution $\mathcal{B}_A$ obtained by sampling n rows $h_1, \ldots, h_n \in [m]$ independently, with each row h_j drawn from the same distribution

$$\Pr[h_j = h] = \frac{|a_{h1}|^2 + \cdots + |a_{hn}|^2}{n},$$

and then outputting $s_i := |\{j : h_j = i\}|$ as $S = (s_1, \ldots, s_m)$. Like $\mathcal{M}_A$, the distribution $\mathcal{B}_A$ is classically samplable in $O(mn)$ time. But again, the probabilities in $\mathcal{B}_A$ are affected by row scaling in exactly the same way as the probabilities in $\mathcal{D}_A$, $\mathcal{M}_A$, and $\mathcal{F}_A$.

What are we to make of the above observations? Arguably, they merely underscore what we said from the beginning: that the row-norm estimator cannot prove, by itself, that BosonSampling is being solved. Indeed, it can't even be used to prove the presence of quantum interference in an alleged BosonSampling device. If $R^*(A_S)$ satisfies the expected statistics, then we know that the device's output is *not* uniform random noise—and moreover, that the device samples from *some* distribution that depends nontrivially on the actual entries of A, in way consistent with correct BosonSampling. If this were combined with other evidence—e.g., verification with smaller numbers of photons, verification that the multi-photon collisions satisfy the expected statistics, and direct ruling out of alternatives such as $\mathcal{M}_A$—it would arguably provide circumstantial evidence that the device was working properly, even with hundreds or thousands of photons.

Along those lines, we now observe that, if we only want to verify that a BosonSampling device is *not* sampling from $\mathcal{M}_A$ (or from any distribution close to $\mathcal{M}_A$ in variation distance), then almost

certainly this can be done in classical polynomial time. The reason is that the probabilities in $\mathcal{M}_A$ are given by permanents of $n \times n$ *nonnegative* matrices—but such permanents can be approximated to within ε multiplicative error in $\text{poly}(n, 1/\varepsilon)$ time, using the famous randomized algorithm of Jerrum, Sinclair, and Vigoda [8]. Thus, given experimental outcomes $S_1, \ldots, S_k \in \Lambda_{m,n}$, in classical polynomial time we can approximate $\text{Per}\left(A_{S_1}^{\#}\right), \ldots, \text{Per}\left(A_{S_k}^{\#}\right)$, then check whether they satisfy the statistics that we expect if the S_i's were drawn from $\mathcal{M}_A$. For similar but even simpler reasons, we can almost certainly rule out, in classical polynomial time, that a BosonSampling device is sampling from the particular "mockup" distributions $\mathcal{F}_A$ or $\mathcal{B}_A$.

Admittedly, to make the argument completely rigorous, we would need to prove that, with high probability over a Haar-random A, the BosonSampling distribution $\mathcal{D}_A$ does *not* give rise to the same statistics for $\text{Per}\left(A_S^{\#}\right)$ as $\mathcal{M}_A$ does, or the same statistics for $|\text{Det}(A_S)|^2$ as $\mathcal{F}_A$ does, or the same statistics for $\Pr_{\mathcal{B}_A}[S]$ as $\mathcal{B}_A$ does. These statements are presumably true, but we leave their proofs to future work.

9 Summary and Open Problems

We began this paper by considering certain claims about BosonSampling made by Gogolin et al. [7]. We found those claims to be misleading on at least three different levels. First, when testing a BosonSampling device's output against the theoretical predictions, there is not the slightest reason to ignore the labels of the modes—and once the mode labels are accounted for, Gogolin et al.'s entire argument for the "near-uniformity" of the output distribution collapses. Second, the observation that certifying a BosonSampling distribution might be classically hard is not new; we made it in [1]. And third, a BosonSampling device can easily be *faster* than its fastest classical certification, without classical certification being practically impossible: indeed there exists a regime (around $n = 30$ photons) where that is precisely what one would expect. Moreover, we did not need any nontrivial technical work to reach these conclusions: had we been content to refute Gogolin et al., this paper could have been extremely short.

Having said that, Gogolin et al.'s paper does have the virtue that it suggests interesting questions about the statistical aspects of BosonSampling. So in the remainder of the paper, we took the opportunity to address some of those questions. First, we proved what had previously been known only heuristically: that given a Haar-random $A \in \mathbb{C}^{m \times n}$, with high probability the BosonSampling distribution $\mathcal{D}_A$ will have noticeable "fluctuations" in the probabilities (some outcomes being more likely, others less), which easily distinguish $\mathcal{D}_A$ from the uniform distribution $\mathcal{U}$. More surprisingly, we showed that $\mathcal{D}_A$ can even be distinguished from $\mathcal{U}$ in *classical polynomial time*, using a simple row-norm estimator. As we pointed out in Section 6, BosonSampling experiments with (say) $n = 30$ photons could still be feasible and interesting, even if distinguishing $\mathcal{D}_A$ from $\mathcal{U}$ were asymptotically hard—but, as it turns out, it isn't hard.

Needless to say, many open problems remain.

(1) Given a Gaussian matrix $X \sim \mathcal{N}(0,1)_{\mathbb{C}}^{n \times n}$, we showed that, in polynomial time, one can compute a quantity $R(X)$ that slightly but nontrivially correlates with $|\text{Per}(X)|^2$. This raises the obvious question: how well *can* one approximate $|\text{Per}(X)|^2$ in polynomial time, for a large fraction of Gaussian X's? Our "Permanent of Gaussians Conjecture" (PGC), from [1], said that approximating $|\text{Per}(X)|^2$ to within $\pm \varepsilon n!$ additive error, for a $1 - \delta$ fraction of X's, should be #P-hard, if we take $n + 1/\varepsilon + 1/\delta$ as the "input length." But there remains an

enormous gap in parameters between that hardness conjecture and the weak approximation algorithm given here. So in particular, even if we assume the PGC, it's perfectly conceivable that much better approximation algorithms for $|\mathrm{Per}\,(X)|^2$ exist.

As a simple example, Linial, Samorodnitsky, and Wigderson [12] proposed an approximation algorithm for the permanent that works by first normalizing all of the rows to 1, then normalizing all of the columns to 1, then normalizing the rows, and so on iteratively until convergence; then using the final product of row- and column-multipliers to produce an estimate for $\mathrm{Per}\,(X)$.[7] How good of an approximation does their algorithm produce, for a Gaussian matrix $X \sim \mathcal{N}\,(0,1)_{\mathbb{C}}^{n\times n}$? Can we show that, with high probability, it produces a *better* approximation than our algorithm from Section 8, which normalized the rows only? Note that the techniques of Section 8 no longer work, since one *cannot* decompose $\mathrm{Per}\,(X)$ as a product of Linial et al.'s estimator and an independent random variable.

(2) Can we prove that a generic BosonSampling distribution $\mathcal{D}_A$ can be distinguished, in classical polynomial time, from the classical "mockup" distribution $\mathcal{M}_A$ defined in Section 8.1? In Section 8.1, we sketched a polynomial-time estimator to distinguish $\mathcal{D}_A$ from $\mathcal{M}_A$ (based on the famous permanent approximation algorithm of Jerrum, Sinclair, and Vigoda [8]), but it remains to prove that our estimator works—or even, for that matter, that $\|\mathcal{D}_A - \mathcal{M}_A\| = \Omega\,(1)$ with high probability over A. One can ask similar questions about distinguishing $\mathcal{D}_A$ from $\mathcal{F}_A$ and from $\mathcal{B}_A$.

(3) Of course, the broader question is whether there exists an efficiently-samplable mockup distribution, call it $\mathcal{M}'_A$, that cannot be distinguished from $\mathcal{D}_A$ by *any* classical polynomial-time algorithm. In Appendix 11, we report an observation of Fernando Brandao (personal communication): that, for any *fixed* k, there exists an efficiently-samplable distribution that has large min-entropy, but that cannot be distinguished from (the collision-free part of) $\mathcal{D}_A$ by any algorithm running in $O\left((m+n)^k\right)$ time. However, while this sounds striking, it says very little about BosonSampling specifically. Indeed, if we removed the requirement of large min-entropy, then as we point out in Appendix 11, it would be trivial to construct such a "mockup" for *any* distribution $\mathcal{D}$ whatsoever! So in our view, the "real" question is whether or not there exists a *single* efficiently-samplable mockup distribution that works against *all* polynomial-time distinguishers.

(4) Everything in (1)-(3) can be asked anew, when we consider various "experimentally realistic" variations of BosonSampling. For example, what happens if some random subset of the photons gets lost between the sources and the detectors? Or what happens if the inputs are Gaussian states, rather than single-photon Fock states? In Appendix 12, we explain a simple reason why even in these cases, the row-norm estimator R still easily distinguishes the BosonSampling distribution from the uniform distribution $\mathcal{U}$. On the other hand, it is possible that the distinguishability between $\mathcal{D}_A$ and $\mathcal{M}_A$, or between $\mathcal{D}_A$ and various other "mockup" distributions, decreases when these or other experimental errors are incorporated.

[7] Actually, Linial et al.'s algorithm is for approximating $\mathrm{Per}\,(X)$ where X is a *nonnegative* matrix, and it proceeds by iteratively normalizing the 1-norms of the rows and columns (i.e., the sums of the entries). However, one could easily adapt their algorithm to attempt to approximate $|\mathrm{Per}\,(X)|^2$, where X is an arbitrary matrix, by instead normalizing the 2-norms of the rows and columns.

(5) For most of this paper, we restricted attention to collision-free outcomes $S \in \Lambda_{m,n}$. There were several reasons for this: first, it greatly simplified the calculations; second, it mirrored the hardness results of [1], which also used collision-free outcomes only; and third, any positive results about, say, the distinguishability of $\mathcal{D}_A$ and $\mathcal{U}$ only become *stronger* under such a restriction. However, what happens when we "put multi-photon collisions back in"? For example, can we give stronger distinguishing and verification algorithms, by taking advantage of the collision outcomes $S \in \Phi_{m,n} \setminus \Lambda_{m,n}$ that we previously discarded?

(6) All of our results in this paper hold with high probability for a Haar-random BosonSampling matrix $A \in \mathbb{C}^{m \times n}$, where (say) $m \geq n^{5.1}$. What can we say if A is arbitrary rather than random? What about if m is smaller—say, $O\left(n^2\right)$ or even $O\left(n\right)$? Finally, by choosing $m \geq n^{5.1} p\left(n\right)$, all of our results can be made to hold with probability at least $1 - 1/p\left(n\right)$ over a Haar-random A, where p is any desired polynomial. Can we show that they hold even with probability $1 - 1/\exp\left(n\right)$ over A, and even for fixed m?

(7) We showed that $\|\mathcal{D}_A - \mathcal{U}\| = \Omega\left(1\right)$ with high probability over A. Can we show the stronger result that $\|\mathcal{D}_A - \mathcal{U}\| = 1 - o\left(1\right)$?

(8) We show in Appendix 13 that, if $X \sim \mathcal{N}$ is Gaussian, then $\left|\mathrm{Det}\left(X\right)\right|^2$ is $O\left(1/\log^{3/2} n\right)$-close to a lognormal random variable. However, we know that this approximation must break down when $\left|\mathrm{Det}\left(X\right)\right|^2$ is either very small or very large. So it would be nice to have a more detailed understanding of the pdf for $\left|\mathrm{Det}\left(X\right)\right|^2$—among other things, because such an understanding might provide clues about the pdf for $\left|\mathrm{Per}\left(X\right)\right|^2$, and hence about how to prove the Permanent Anti-Concentration Conjecture (PACC) from [1].

10 Acknowledgments

We thank Shalev Ben-David and Charles Xu for helpful discussions, and especially Fernando Brandao for allowing us to include the results in Appendix 11, and Salil Vadhan and Luca Trevisan for clarifying aspects of [18]. We also thank Oded Regev and John Watrous for code used in generating Figures 1 and 2.

References

[1] S. Aaronson and A. Arkhipov. The computational complexity of linear optics. *Theory of Computing*, 9(4):143–252, 2013. Conference version in Proceedings of ACM STOC'2011. ECCC TR10-170, arXiv:1011.3245.

[2] M. A. Broome, A. Fedrizzi, S. Rahimi-Keshari, J. Dove, S. Aaronson, T. C. Ralph, and A. G. White. Photonic boson sampling in a tunable circuit. *Science*, (10.1126/science.1231440), 2012. arXiv:1212.2234.

[3] K. P. Costello and V. H. Vu. Concentration of random determinants and permanent estimators. *SIAM J. Discrete Math*, 23(3).

[4] A. Crespi, R. Osellame, R. Ramponi, D. J. Brod, E. F. Galv ao, N. Spagnolo, C. Vitelli, E. Maiorino, P. Mataloni, and F. Sciarrino. Experimental boson sampling in arbitrary integrated photonic circuits. *Nature Photonics*, 7:545–549, 2013. arXiv:1212.2783.

[5] B. T. Gard, R. M. Cross, M. B. Kim, H. Lee, and J. P. Dowling. Classical computers very likely can not efficiently simulate multimode linear optical interferometers with arbitrary Fock-state inputs - an elementary argument. arXiv:1304.4206, 2013.

[6] V. L. Girko. A refinement of the Central Limit Theorem for random determinants. *Teor. Veroyatnost. i Primenen*, 42:63–73, 1997. Translation in Theory Probab. Appl 42 (1998), 121-129.

[7] C. Gogolin, M. Kliesch, L. Aolita, and J. Eisert. Boson-Sampling in the light of sample complexity. arXiv:1306.3995, 2013.

[8] M. Jerrum, A. Sinclair, and E. Vigoda. A polynomial-time approximation algorithm for the permanent of a matrix with non-negative entries. *J. ACM*, 51(4):671–697, 2004. Earlier version in STOC'2001.

[9] T. Jiang. How many entries of a typical orthogonal matrix can be approximated by independent normals? *Ann. Probab.*, 34(4):1497–1529, 2006.

[10] E. Knill. Fermionic linear optics and matchgates. quant-ph/0108033, 2001.

[11] E. Knill, R. Laflamme, and G. J. Milburn. A scheme for efficient quantum computation with linear optics. *Nature*, 409:46–52, 2001. See also quant-ph/0006088.

[12] N. Linial, A. Samorodnitsky, and A. Wigderson. A deterministic strongly polynomial algorithm for matrix scaling and approximate permanents. *Combinatorica*, 20(4):545–568, 2000.

[13] R. Renner and S. Wolf. Smooth rényi entropy and applications. In *Proc. IEEE International Symposium on Information Theory (ISIT)*, page 233, 2004.

[14] J. B. Spring, B. J. Metcalf, P. C. Humphreys, W. S. Kolthammer, X.-M. Jin, M. Barbieri, A. Datta, N. Thomas-Peter, N. K. Langford, D. Kundys, J. C. Gates, B. J. Smith, P. G. R. Smith, and I. A. Walmsley. Boson sampling on a photonic chip. *Science*, (10.1126/science.1231692), 2012. arXiv:1212.2622.

[15] T. Tao. Determinantal processes, 2009. Blog post. http://terrytao.wordpress.com/2009/08/23/determinantal-processes/.

[16] B. M. Terhal and D. P. DiVincenzo. Classical simulation of noninteracting-fermion quantum circuits. *Phys. Rev. A*, 65(032325), 2002. quant-ph/0108010.

[17] M. Tillmann, B. Dakić, R. Heilmann, S. Nolte, A. Szameit, and P. Walther. Experimental boson sampling. *Nature Photonics*, 7:540–544, 2013. arXiv:1212.2240.

[18] L. Trevisan, M. Tulsiani, and S. Vadhan. Regularity, boosting, and efficiently simulating every high-entropy distribution. In *Proc. IEEE Conference on Computational Complexity*, pages 126–136, 2009. ECCC TR08-103.

[19] L. G. Valiant. Quantum circuits that can be simulated classically in polynomial time. *SIAM J. Comput.*, 31(4):1229–1254, 2002. Earlier version in STOC'2001.

11 Appendix: Fixed Polynomial-Size Distinguishers

In this appendix, we report a result due to Fernando Brandao (with Brandao's kind permission): namely, that one can apply a recent theorem of Trevisan et al. [18] to construct an efficiently-samplable distribution that

(i) has large entropy (in fact, large min-entropy), and

(ii) is indistinguishable from (the collision-free part of) a generic BosonSampling distribution, by circuits of any *fixed* polynomial size.

We first observe that, if condition (i) is dropped, then condition (ii) is trivial to satisfy for *any* distribution whatsoever: there is nothing specific about BosonSampling here![8] To see this: first, recall that two distributions $\mathcal{D}$ and $\mathcal{D}'$ are called *ε-indistinguishable* with respect to a circuit C if

$$\left| \Pr_{x \sim \mathcal{D}} [C(x) \text{ accepts}] - \Pr_{x \sim \mathcal{D}'} [C(x) \text{ accepts}] \right| \leq \varepsilon,$$

and are ε-indistinguishable with respect to a *class* of circuits $\mathcal{C}$ if they are ε-indistinguishable with respect to every $C \in \mathcal{C}$. Now let $\mathcal{D}$ be any distribution over $\{0,1\}^n$. Then choose w elements independently with replacement from $\mathcal{D}$, and let $\mathcal{D}'$ be the uniform distribution over the resulting multiset (so in particular, $H(\mathcal{D}') \leq \log_2 n$). Certainly $\mathcal{D}'$ can be sampled by a circuit of size $O(nw)$: just hardwire the elements.

Now, by a Chernoff bound, for any *fixed* circuit C, clearly $\mathcal{D}$ and $\mathcal{D}'$ are ε-indistinguishable with respect to C, with probability at least $1 - \exp\left(-\varepsilon^2 w\right)$ over the choice of $\mathcal{D}'$. But there are "only" $n^{O(n^k)}$ Boolean circuits of size at most n^k. So by the union bound, by simply choosing $w = \Omega\left(\frac{n^k \log n}{\varepsilon^2}\right)$, we can ensure that $\mathcal{D}$ and $\mathcal{D}'$ are ε-indistinguishable with respect to *all* circuits of size at most n^k, with high probability over $\mathcal{D}'$.

Thus, we see that the one nontrivial aspect of Brandao's observation is that, in the case of BosonSampling, the "mockup" distribution $\mathcal{D}'$ can itself have large entropy. Even for this, however, we will need very little that is specific to BosonSampling: only that a generic BosonSampling distribution $\mathcal{D}_A$ is close to a distribution with large min-entropy.

Before proving Brandao's observation, we need some more definitions. Given a probability distribution $\mathcal{D} = \{p_x\}_x$ over $[N]$, recall that the *min-entropy* of $\mathcal{D}$ is

$$H_{\min}(\mathcal{D}) := \min_{x \in [N]} \log_2 \frac{1}{p_x},$$

while the 2-*entropy* is

$$H_2(\mathcal{D}) := -\log_2 \sum_{x \in [N]} p_x^2.$$

One can show that $H_{\min}(\mathcal{D}) \leq H_2(\mathcal{D}) \leq H(\mathcal{D})$ for all $\mathcal{D}$, where $H(\mathcal{D})$ is the ordinary Shannon entropy. The following useful lemma, which we prove for completeness, provides a partial converse to the inequality $H_{\min}(\mathcal{D}) \leq H_2(\mathcal{D})$.

[8] We thank Salil Vadhan for this observation.

Lemma 19 (Renner and Wolf [13]) *For any distribution $\mathcal{D}$ over $[N]$ and $\varepsilon > 0$, there exists a distribution $\mathcal{D}'$ over $[N]$ such that $\|\mathcal{D} - \mathcal{D}'\| \leq \varepsilon$ and $H_{\min}(\mathcal{D}') \geq H_2(\mathcal{D}) - \log_2 \frac{1}{\varepsilon}$.*

Proof. Let $p \geq 1/N$ be a "cutoff" to be determined later. We define $\mathcal{D}' = \{q_x\}_x$ by starting from $\mathcal{D} = \{p_x\}_x$, then cutting all probabilities p_x such that $p_x > p$ down to p, and finally adding back the probability mass that we removed (i.e., increasing the q_x's) in any way that maintains the property $q_x \leq p$ for all $x \in [N]$. The result clearly satisfies

$$H_{\min}(\mathcal{D}') \geq \log_2 \frac{1}{p}$$

and also

$$\left\|\mathcal{D} - \mathcal{D}'\right\| = \frac{1}{2} \sum_{x \in [N]} |p_x - q_x| \leq \sum_{x : p_x > p} (p_x - p).$$

Indeed, let us simply choose p such that

$$\sum_{x : p_x > p} (p_x - p) = \varepsilon.$$

By continuity, it is clear that such a p exists for every $\varepsilon > 0$.

Now notice that

$$\frac{1}{2^{H_2(\mathcal{D})}} = \sum_{x \in [N]} p_x^2 \geq \sum_{x : p_x > p} p_x^2 > p \sum_{x : p_x > p} p_x \geq p\varepsilon \geq \frac{\varepsilon}{2^{H_{\min}(\mathcal{D}')}},$$

or rearranging,

$$H_{\min}(\mathcal{D}') \geq H_2(\mathcal{D}) - \log_2 \frac{1}{\varepsilon}.$$

■

We now state the result of Trevisan et al. [18] that we will use.

Theorem 20 (Trevisan-Tulsiani-Vadhan [18]) *Let $\mathcal{D}$ be any distribution over $\{0,1\}^n$ such that $H_{\min}(\mathcal{D}) \geq n - k$. Then for every T and $\varepsilon > 0$, there exists a circuit of size $(T + n) \operatorname{poly}\left(2^k, 1/\varepsilon\right)$ that samples a distribution $\mathcal{D}'$ that has $H_{\min}(\mathcal{D}') \geq n - k$ and that is ε-indistinguishable from $\mathcal{D}$ by circuits of size at most T.*

Let $\mathcal{D}_A^*$ be the BosonSampling distribution $\mathcal{D}_A$ conditioned on a collision-free outcome (that is, on $S \in \Lambda_{m,n}$). Then it remains to show that $\mathcal{D}_A^*$ has large min-entropy, with high probability over A. To do this, we first recall two results proved in [1] (Theorem 5.2 and Lemma 8.8 respectively there).

Theorem 21 ([1]) *Let $m \geq \frac{n^5}{\delta} \log^2 \frac{n}{\delta}$ for any $\delta > 0$, let $\mathcal{N}$ and $\mathcal{N}'$ be as defined in Section 7, and let $f_{\mathcal{N}}(X)$ and $f_{\mathcal{N}'}(X)$ be the pdfs of $\mathcal{N}$ and $\mathcal{N}'$ respectively. Then for all $X \in \mathbb{C}^{n \times n}$,*

$$f_{\mathcal{N}'}(X) \leq (1 + O(\delta)) f_{\mathcal{N}}(X).$$

In particular, if (say) $m \geq n^{5.1}$ and n is sufficiently large, then Theorem 21 implies that $f_{\mathcal{N}'}(X) \leq 2 f_{\mathcal{N}}(X)$.

Lemma 22 ([1]) $\mathrm{E}_{X\sim\mathcal{N}}\left[\left|\mathrm{Per}\left(X\right)\right|^4\right] = \left(n+1\right)\left(n!\right)^2$.

Combining Theorem 21 with Lemma 22, Brandao observed the following.

Lemma 23 (Brandao) *Let $m \geq n^{5.1}$, and let $A \in \mathbb{C}^{m\times n}$ be a Haar-random* BosonSampling *matrix. Then for sufficiently large n and for all $\delta > 0$, with probability at least $1-\delta$ over A we have*

$$\sum_{S\in\Lambda_{m,n}} \left|\mathrm{Per}\left(A_S\right)\right|^4 \leq \frac{2\left(n+1\right)!}{\delta m^n}.$$

Proof. For any fixed $S \in \Lambda_{m,n}$, we have

$$\begin{aligned}
\mathop{\mathrm{E}}_{A}\left[\left|\mathrm{Per}\left(A_S\right)\right|^4\right] &= \frac{1}{m^{2n}} \mathop{\mathrm{E}}_{X\sim\mathcal{N}'}\left[\left|\mathrm{Per}\left(X\right)\right|^4\right] \\
&\leq \frac{2}{m^{2n}} \mathop{\mathrm{E}}_{X\sim\mathcal{N}}\left[\left|\mathrm{Per}\left(X\right)\right|^4\right] \\
&= \frac{2\left(n+1\right)\left(n!\right)^2}{m^{2n}},
\end{aligned}$$

where the second line uses Theorem 21 and the third uses Lemma 22. Hence

$$\mathop{\mathrm{E}}_{A}\left[\sum_{S\in\Lambda_{m,n}} \left|\mathrm{Per}\left(A_S\right)\right|^4\right] \leq \binom{m}{n}\frac{2\left(n+1\right)\left(n!\right)^2}{m^{2n}} \leq \frac{2\left(n+1\right)!}{m^n}.$$

So by Markov's inequality,

$$\Pr_{A}\left[\sum_{S\in\Lambda_{m,n}} \left|\mathrm{Per}\left(A_S\right)\right|^4 > \frac{2\left(n+1\right)!}{\delta m^n}\right] < \delta.$$

■

Combining Lemma 23 with Lemma 19 now yields the following corollary.

Corollary 24 (Brandao) *Let $m \geq n^{5.1}$, and let $A \in \mathbb{C}^{m\times n}$ be a Haar-random* BosonSampling *matrix. Then for all $\varepsilon, \delta > 0$, with probability at least $1-\delta$ over A, there exists a distribution $\mathcal{D}'$ over $\Lambda_{m,n}$ such that $\|\mathcal{D}' - \mathcal{D}_A^*\| \leq \varepsilon$ and*

$$H_{\min}\left(\mathcal{D}'\right) \geq \log_2\binom{m}{n} - \log_2\frac{n}{\varepsilon\delta} - O\left(1\right).$$

Proof. By Lemma 23, for sufficiently large n and all $\delta > 0$, with probability at least $1 - \delta$ over A we have

$$\begin{aligned} H_2\left(\mathcal{D}_A^*\right) &= -\log_2 \sum_{S \in \Lambda_{m,n}} \Pr_{\mathcal{D}_A^*}\left[S\right]^2 \\ &\geq -\log_2 \sum_{S \in \Lambda_{m,n}} \left(2 \Pr_{\mathcal{D}_A}\left[S\right]\right)^2 \\ &= -2 - \log_2 \sum_{S \in \Lambda_{m,n}} \left|\mathrm{Per}\left(A_S\right)\right|^4 \\ &\geq -2 - \log_2 \frac{2\left(n+1\right)!}{\delta m^n} \\ &= \log_2 \frac{m^n}{n!} - \log_2\left(n+1\right) - \log_2 \frac{8}{\delta} \\ &\geq \log_2 \binom{m}{n} - \log_2 n - \log_2 \frac{9}{\delta}. \end{aligned}$$

So suppose the above inequality holds. Then by Lemma 19, for every $\varepsilon > 0$ there exists a distribution $\mathcal{D}'$ over $\Lambda_{m,n}$ such that $\left\|\mathcal{D}' - \mathcal{D}_A^*\right\| \leq \varepsilon$ and

$$\begin{aligned} H_{\min}\left(\mathcal{D}'\right) &\geq H_2\left(\mathcal{D}_A^*\right) - \log_2 \frac{1}{\varepsilon} \\ &\geq \log_2 \binom{m}{n} - \log_2 \frac{n}{\varepsilon\delta} - O\left(1\right). \end{aligned}$$

■

It remains only to combine Corollary 24 with Theorem 20.

Theorem 25 (Brandao) *Let $m \geq n^{5.1}$, and let $A \in \mathbb{C}^{m \times n}$ be a Haar-random* BosonSampling *matrix. Then with probability at least $1 - \delta$ over A, for every T and $\varepsilon > 0$, there exists a circuit of size $T\,\mathrm{poly}\left(n, 1/\varepsilon, 1/\delta\right)$ that samples a distribution $\mathcal{D}'$ that has*

$$H_{\min}\left(\mathcal{D}'\right) \geq \log_2 \binom{m}{n} - \log_2 \frac{n}{\varepsilon\delta} - O\left(1\right),$$

and that is ε-indistinguishable from $\mathcal{D}_A^*$ by circuits of size at most T.

12 Appendix: Arbitrary Initial States

Suppose the input to a BosonSampling device does *not* consist of a single-photon Fock state in each mode, but of some "messier" pure or mixed state. For example, perhaps some subset of photons are randomly lost to the environment, or perhaps the input involves coherent states or other Gaussian states. One might wonder: *then* does the device's output distribution "flatten out" and become nearly uniform, as Gogolin et al. [7] claimed?

In this appendix, we briefly explain why the answer is still no. We will argue that, for essentially *any* initial state, the row-norm estimator R^* from Section 8 still "works," in the sense that it still distinguishes the output distribution from the uniform distribution with non-negligible (and typically constant) bias. Our discussion will be at a "physics level of rigor."

Let k be the total number of photons. Then an arbitrary initial pure state might involve a superposition of different k's, like so:

$$|\psi\rangle = \sum_{k=0}^{\infty} \sum_{S \in \Phi_{n,k}} \alpha_S |S\rangle . \tag{7}$$

(Later we will generalize to mixed states.) We make the following assumptions:

(a) $m \gg n^2$: i.e., there are many more output modes than input modes, as assumed throughout this paper. Alternatively, if we identify input and output modes, then we assume that all photons are initially concentrated in the first n of m modes.

(b) With non-negligible probability, $m \gg k^2$ (i.e., there are many more output modes than photons).

(c) With non-negligible probability, $k > 0$ (i.e., at least one photon gets detected).

Assumption (c) is obviously necessary, but we believe that a more sophisticated analysis would let us weaken assumption (a) and remove assumption (b).[9]

If we want to model photon losses in this framework, we can do so by simply letting our initial state be a mixture of different $|\psi\rangle$'s, corresponding to different combinations of photons that are "destined to get lost." We do not attempt to model more complicated loss processes (e.g., processes that depend on the unitary), or other sources of error such as mode mismatch. Extending the statistical analysis of BosonSampling to those cases is an interesting challenge for future work.

In what follows, we assume some familiarity with *Fock polynomials*, i.e., polynomials in photon creation operators (see [1, Section 3.2] for a computer-science-friendly introduction). In particular, the Fock polynomial associated to the state $|\psi\rangle$ from (7) is

$$p_\psi(x) = \sum_{k=0}^{\infty} \sum_{S \in \Phi_{n,k}} \frac{\alpha_S x^S}{\sqrt{S!}},$$

where $x = (x_1, \ldots, x_m)$ and $S = (s_1, \ldots, s_m)$ (in this particular case, $s_i = 0$ if $i > n$). Here we use the shorthands $x^S = x_1^{s_1} \cdots x_m^{s_m}$ and $S! = s_1! \cdots s_m!$. Applying a unitary transformation U results in the rotated state $p_\psi(Ux)$. So the probability of measuring some particular k-photon outcome $S \in \Phi_{m,k}$ is then

$$\Pr[S] = \frac{1}{S!} \left| \left\langle x^S, p_\psi(Ux) \right\rangle \right|^2,$$

where $\langle , \rangle$ represents the Fock inner product. We can now use the adjoint property of the Fock product (Theorem 3.4 in [1]) to move U to the left-hand side:

$$\begin{aligned} \Pr[S] &= \frac{1}{S!} \left| \left\langle \left(U^\dagger x\right)^S, p_\psi(x) \right\rangle \right|^2 \\ &= \frac{1}{S!} \left| \left\langle \prod_{i=1}^{m} \left(U^\dagger x_i\right)^{s_i}, p_\psi(x) \right\rangle \right|^2. \end{aligned}$$

[9] On the other hand, we observe that assumption (a) cannot be removed entirely. For suppose the initial state was the k-photon maximally mixed state $I_{m,k}$ (for any k), which assigns equal probability to each element of $\Phi_{m,k}$. Then one can show that, *regardless* of what unitary transformation U was applied, the final state would again be $I_{m,k}$. And therefore, neither R^* nor any other estimator could distinguish the output distribution from uniform.

Because $|\psi\rangle$ has support only on the first n modes, the right-hand side uses only the variables $x_1, \ldots, x_n$, so only the first $n \ll \sqrt{m}$ rows of U affect $\Pr[S]$. Moreover, since only i's such that $s_i > 0$ contribute to the contribute to the left-hand side, only those columns of U matter, of which there are at most $k \ll \sqrt{m}$ (since S is a k-photon outcome).

So by Theorem 6, we can achieve a good approximation to $\Pr[S]$ by replacing $U^\dagger$ with an $n \times n$ scaled Gaussian matrix, $G/\sqrt{m}$ where $G \sim \mathcal{N}(0,1)_{\mathbb{C}}^{n \times n}$. In that case,

$$\Pr[S] \approx \frac{1}{S!m^k} \left| \left\langle \prod_{i=1}^{m} (Gx_i)^{s_i}, p_\psi(x) \right\rangle \right|^2.$$

We can then factor out the squared 2-norms of the rows of G (call them $R_1, \ldots, R_n$), resulting in a random $n \times n$ matrix G' remaining, each of whose rows is an independent, Haar-random unit vector:

$$\Pr[S] \approx \left(\prod_{i=1}^{n} R_i^{s_i} \right) \cdot \frac{1}{S!m^k} \left| \left\langle \prod_{i=1}^{m} \left(G'x_i\right)^{s_i}, p_\psi(x) \right\rangle \right|^2. \tag{8}$$

We have thus approximated $\Pr[S]$ by the product of two *independent* random variables, the first of which is precisely the row-norm estimator,

$$R_S := \prod_{i=1}^{n} R_i^{s_i}.$$

Let

$$R_S^* = \prod_{i=1}^{n} \frac{R_i^{s_i}}{n(n+1)\cdots(n+s_i-1)}$$

be the scaled version of R_S, so that $\mathrm{E}[R_S^*] = 1$. Also, let $\mathcal{U}$ be the uniform distribution over $S \in \Phi_{m,k}$, and let $\mathcal{D}$ be the distribution over $S \in \Phi_{m,k}$ induced by BosonSampling with unitary transformation U, and then conditioning on a k-photon outcome. Then just like in Lemma 14 of Section 8, an immediate consequence of the decomposition (8) is that

$$\Pr_{S \sim \mathcal{D}}[R_S^* \geq 1] - \Pr_{S \sim \mathcal{U}}[R_S^* \geq 1] \approx \frac{1}{2} \mathop{\mathrm{E}}_{S \sim \mathcal{U}}[|R_S^* - 1|].$$

Or in other words, R_S^* will distinguish the case $S \sim \mathcal{D}$ from the case $S \sim \mathcal{U}$ with non-negligible bias, *if and only if* there is non-negligible "intrinsic variation" in R_S^* itself when $S \sim \mathcal{U}$.

Now, at least when S is collision-free (i.e., $s_i \in \{0,1\}$ for all $i \in [m]$), our analysis in Section 8 implies that R_S^* converges to a lognormal random variable with

$$\mathop{\mathrm{E}}_{S \sim \mathcal{U}}[|R_S^* - 1|] = \Omega\left(\frac{k}{n}\right).$$

This is non-negligible whenever $k > 0$, and constant whenever $k = \Omega(n)$. Moreover, one can show that if S contains collisions, then for a fixed k, the variation $\mathrm{E}[|R_S^* - 1|]$ only becomes *larger* (we omit the proof).

We can extend the above analysis to a mixed initial state ρ by observing that, in the mixed case, $\Pr[S]$ is just a convex combination of the probabilities arising from the pure states $|\psi\rangle$ comprising ρ. So an identical term R_S^* can be factored out from all of those probabilities.

13 Appendix: FermionSampling

In this appendix, we prove two results about FermionSampling: the "easier cousin" of BosonSampling, involving determinants rather than permanents. The first result is that, in sharp contrast to what we conjecture for the bosonic case, FermionSampling can be solved in classical polynomial time (indeed, $O\left(mn^2\right)$ time), and by a particularly simple algorithm. It was already known, from work by Valiant [19], Terhal and DiVincenzo [16], and Knill [10], that FermionSampling is efficiently solvable classically. However, our algorithm—which is basically identical to an algorithm discovered independently in the field of *determinantal point processes* (see Tao [15] for example)—is both simpler and faster than the previous FermionSampling algorithms.

The second result is that, if $X \sim \mathcal{N}\left(0,1\right)_{\mathbb{C}}^{n \times n}$ is Gaussian, then $\left|\mathrm{Det}\left(X\right)\right|^2$ converges at a $1/\log^{3/2} n$ rate to a lognormal random variable. Again, this essentially follows from earlier results by Girko [6] and Costello and Vu [3], except that they considered real matrices only and did not bound the convergence rate. An implication is that, as $n \rightarrow \infty$, the histogram of outcome probabilities $\left\{\Pr\left[S\right]\right\}_{S \in \Lambda_{m,n}}$ for a Haar-random FermionSampling distribution converges to lognormal.

What does any of this have to do with the rest of the paper? There are three connections. First, as discussed in Section 8.1, FermionSampling is extremely interesting as a "mockup" of BosonSampling: something that satisfies the same row-norm statistics (and, indeed, also involves quantum interference among n identical particles), but is nevertheless easy for a classical computer. So it might be illuminating to see explicitly just *why* FermionSampling is so easy. Second, we will see how the mathematical techniques developed in Sections 7 and 8 for BosonSampling, can easily be reapplied to FermionSampling. And third, recall that the pdfs for $\left|\mathrm{Det}\left(X\right)\right|^2$ and $\left|\mathrm{Per}\left(X\right)\right|^2$ look almost identical when plotted (see Figure 1). For this reason, FermionSampling can serve as a useful "laboratory" or "model system": something whose statistical properties are almost certainly similar to those of BosonSampling, but that's easier to understand rigorously.

Before going further, we should define FermionSampling formally. Let $A \in \mathbb{C}^{m \times n}$ $(m \geq n)$ be a column-orthonormal matrix. Then the FermionSampling distribution $\mathcal{F}_A$ ranges over collision-free outcomes $S \in \Lambda_{m,n}$ (or equivalently, subsets of $[m]$ of size n), and is given by

$$\Pr_{\mathcal{F}_A}\left[S\right] = \left|\mathrm{Det}\left(A_S\right)\right|^2,$$

where A_S is the $n \times n$ submatrix of A corresponding to S. Note that, unlike with BosonSampling, we do not need to worry about outcomes with collisions. For if $S \notin \Lambda_{m,n}$, then A_S has a repeated row, and therefore $\mathrm{Det}\left(A_S\right) = 0$ (physically, this just reflects the famous *Pauli exclusion principle*, that no two fermions can occupy the same state at the same time).

We now show another difference between BosonSampling and FermionSampling: namely, that the latter is easy to solve classically. As we said, this fact was known [19, 16, 10], but our proof is shorter than previous ones and leads to a faster algorithm.

Theorem 26 *There exists a probabilistic classical algorithm that, given $A \in \mathbb{C}^{m \times n}$ as input, samples from $\mathcal{F}_A$ in $O\left(mn^2\right)$ time.*

Proof. The algorithm is the following:

(1) Let $v_{1,i} := \left(a_{i1}, \ldots, a_{in}\right)^T$ be the column vector in $\mathbb{C}^n$ obtained by transposing the i^{th} row of A.

(2) For $t = 1$ to n:

- For all $i \in [m]$, let
$$p_{t,i} := \frac{\|v_{t,i}\|_2^2}{n-t+1}.$$
- Sample $h_t \in [m]$ from the probability distribution $(p_{t,1}, \ldots, p_{t,m})$.
- For all $i \in [m]$, set $v_{t+1,i}$ equal to the projection of $v_{t,i}$ onto the orthogonal complement of v_{t,h_t}:
$$v_{t+1,i} := v_{t,i} - \frac{v_{t,h_t}^\dagger v_{t,i}}{\|v_{t,h_t}\|_2^2} v_{t,h_t}.$$
(So in particular, v_{t+1,h_t} is set to the all-0 vector.)

(3) Output $S = \{h_1, \ldots, h_n\}$ as a sample from $\mathcal{F}_A$.

Clearly the algorithm runs in $O\left(mn^2\right)$ time. To see that the probability distributions are normalized, observe that
$$\sum_{i=1}^m \|v_{1,i}\|_2^2 = \sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2 = n,$$
and that each iteration decreases $\sum_{i=1}^m \|v_{t,i}\|_2^2$ by exactly 1. Furthermore, if h_t is sampled, then
$$\|v_{t,h_t}\|_2^2 = (n-t+1)\, p_{t,h_t} > 0,$$
so the projection $v_{t+1,i}$ is well-defined.

To see that the algorithm is correct, recall that, if $X \in \mathbb{C}^n$ is an $n \times n$ matrix, then one way to *compute* $|\mathrm{Det}(X)|^2$ is to project each row x_t of X onto the orthogonal complement of the subspace spanned by all the rows $x_1, \ldots, x_{t-1}$ above x_t, and then take the product of the squared row-norms of the matrix X' that results. But this is precisely what our sampling procedure does. More concretely, let $H = (h_1, \ldots, h_n) \in [m]^n$ be an ordered sequence of rows, and let $S = \{h_1, \ldots, h_n\}$. Then H gets sampled with probability equal to
$$p_{1,h_n} \cdots p_{n,h_n} = \frac{\|v_{1,h_1}\|_2^2}{n} \frac{\|v_{2,h_2}\|_2^2}{n-1} \cdots \frac{\|v_{n,h_n}\|_2^2}{1} = \frac{|\mathrm{Det}(A_S)|^2}{n!}.$$
So the probability that $h_1, \ldots, h_n$ get sampled in *any* order is simply $n!$ times the above, or $|\mathrm{Det}(A_S)|^2$. ■

We now turn to understanding the pdf of $|\mathrm{Det}(X)|^2$, where $X \sim \mathcal{N}(0,1)_{\mathbb{C}}^{n \times n}$ is an iid Gaussian matrix. The reason we are interested in this pdf is that, by Theorem 6, an $n \times n$ submatrix of a Haar-random $A \in \mathbb{C}^{m \times n}$ will be close in variation distance to an iid Gaussian matrix if $m \gg n^2$. And therefore, the pdf of $|\mathrm{Det}(X)|^2$ controls the behavior of a Haar-random FermionSampling distribution, in exactly the same way that we saw that the pdf of $|\mathrm{Per}(X)|^2$ controls the behavior of a Haar-random BosonSampling distribution. (However, we will not discuss explicitly how to move from statements about $|\mathrm{Det}(X)|^2$ to statements about FermionSampling, since it is precisely analogous to how we moved from statements about $|\mathrm{Per}(X)|^2$ to statements about BosonSampling in Section 7.)

The key to understanding the pdf of $|\mathrm{Det}(X)|^2$ is the following proposition, which is noted for example in Costello and Vu [3, Appendix], and which we also used in [1].

Proposition 27 *If $\mathcal{N}(0,1)_{\mathbb{C}}^{n\times n}$, then $|\mathrm{Det}(X)|^2$ is distributed precisely as*

$$|y_{11}|^2\left(|y_{21}|^2+|y_{22}|^2\right)\cdots\left(|y_{n1}|^2+\cdots+|y_{nn}|^2\right),$$

where each y_{ij} is an independent $\mathcal{N}(0,1)_{\mathbb{C}}$ Gaussian.

Proof Sketch. Let $v_1,\ldots,v_n$ be the rows of X, and let w_t be the projection of v_t onto the orthogonal complement of the subspace spanned by $v_1,\ldots,v_{t-1}$. Then just like in the proof of Theorem 26, we can write $|\mathrm{Det}(X)|^2$ as $\|w_1\|_2^2\cdots\|w_n\|_2^2$. Now, clearly $\|w_1\|_2^2=\|v_1\|_2^2$ is distributed as a complex χ^2 random variable with n degrees of freedom. But it follows, from the rotational symmetry of the Gaussian measure, that $\|w_2\|_2^2$ is distributed as a complex χ^2 with $n-1$ degrees of freedom, $\|w_3\|_2^2$ is distributed as a complex χ^2 with $n-2$ degrees of freedom, and so on. Moreover these χ^2's are all independent of each other, since the v_t's are. ∎

Using Proposition 27, Girko [6] and Costello and Vu [3, Appendix] showed that, if $X\sim\mathcal{N}(0,1)_{\mathbb{R}}^{n\times n}$, then

$$\frac{\ln|\mathrm{Det}(X)|-\ln\sqrt{(n-1)!}}{\sqrt{\frac{\ln n}{2}}}$$

converges weakly to the standard normal distribution $\mathcal{N}(0,1)_{\mathbb{R}}$. This result falls short of what we want in two minor respects: it's for real rather than complex X, and it doesn't quantitatively bound the convergence rate. For completeness, we now fill these gaps. We will do so using the Berry-Esseen Theorem (Theorem 15), but in a variant for sums of non-iid random variables.

Theorem 28 (Berry-Esseen, non-iid case) *Let $Z_1,\ldots,Z_n$ be real iid random variables satisfying*

$$\begin{aligned}\mathrm{E}[Z_i]&=\upsilon_i,\\ \mathrm{E}\left[(Z_i-\upsilon_i)^2\right]&=\sigma_i^2>0,\\ \mathrm{E}\left[|Z_i-\upsilon_i|^3\right]&=\rho_i<\infty.\end{aligned}$$

Then let

$$Z:=Z_1+\cdots+Z_n,$$

and let W be a real Gaussian with mean $\upsilon_1+\cdots+\upsilon_n$ and variance $\sigma_1^2+\cdots+\sigma_n^2$. Then for all $x\in\mathbb{R}$,

$$|\Pr[Z\leq x]-\Pr[W\leq x]|\leq C\frac{\rho_1+\cdots+\rho_n}{\left(\sigma_1^2+\cdots+\sigma_n^2\right)^{3/2}},$$

where C is some universal constant.

By combining Theorem 28 with Lemma 16 and Proposition 27, we can now upper-bound the variation distance between $|\mathrm{Det}(X)|^2$ and a lognormal random variable.

Theorem 29 *Let $X\sim\mathcal{N}(0,1)_{\mathbb{C}}^{n\times n}$ be Gaussian. Then for all $x\in\mathbb{R}$,*

$$\Pr_X\left[\frac{|\mathrm{Det}(X)|^2}{n!}\leq\sqrt{\frac{e}{2\pi n}}e^{x\sqrt{\ln n+1+\gamma}}\right]=\int_{-\infty}^{x}\frac{e^{-y^2/2}}{\sqrt{2\pi}}dy\pm O\left(\frac{1}{\log^{3/2}n}\right).$$

In other words, the cdf of $|\mathrm{Det}\,(X)|^2$ *is pointwise* $O\left(\log^{-3/2} n\right)$*-close to the cdf of a lognormal random variable.*

Proof. We have

$$\begin{aligned}\ln\left(|\mathrm{Det}\,(X)|^2\right) &= \ln\left(|y_{11}|^2\right) + \ln\left(|y_{21}|^2 + |y_{22}|^2\right) + \cdots + \ln\left(|y_{n1}|^2 + \cdots + |y_{nn}|^2\right) \\ &= \ell_1 + \cdots + \ell_n,\end{aligned}$$

where the first line uses the notation of Proposition 27 (each y_{ij} being an independent $\mathcal{N}(0,1)_{\mathbb{C}}$ Gaussian), and the second line uses the notation of Section 8. Thus, $\ln\left(|\mathrm{Det}\,(X)|^2\right)$ is similar to the random variable $L = \ln R$ studied in Section 8, except that, whereas L was a sum of n *identical* ℓ_n random variables, $\ln\left(|\mathrm{Det}\,(X)|^2\right)$ is a sum of n non-identical variables $\ell_1, \ldots, \ell_n$.

Still, if we let

$$\begin{aligned}\upsilon_t &= \mathrm{E}\left[\ell_t\right], \\ \sigma_t^2 &= \mathrm{Var}\left[\ell_t\right], \\ \rho_t &= \mathrm{E}\left[\left|\ell_n - \upsilon_t\right|^3\right],\end{aligned}$$

then by Lemma 16,

$$\sigma_t^2 = \frac{1 + o(1)}{t}, \quad \rho_t = \frac{3^{3/4} + o(1)}{t^{3/2}}.$$

So let W be a real Gaussian with mean $\upsilon_1 + \cdots + \upsilon_n$ and variance $\sigma_1^2 + \cdots + \sigma_n^2$. Then by Theorem 28, for all $x \in \mathbb{R}$,

$$\left|\Pr\left[\ln\left(|\mathrm{Det}\,(X)|^2\right) \leq x\right] - \Pr\left[W \leq x\right]\right| \leq C\frac{\rho_1 + \cdots + \rho_n}{\left(\sigma_1^2 + \cdots + \sigma_n^2\right)^{3/2}} = O\left(\frac{1}{\log^{3/2} n}\right).$$

It remains only to estimate the mean and variance of $\ln\left(|\mathrm{Det}\,(X)|^2\right)$. Using Lemma 16, it is not hard to show the following:

$$\begin{aligned}\mathrm{E}\left[\ln\left(|\mathrm{Det}\,(X)|^2\right)\right] &= \upsilon_1 + \cdots + \upsilon_n = n\ln n - n + \frac{1}{2} - O\left(\frac{1}{n}\right), \\ \mathrm{Var}\left[\ln\left(|\mathrm{Det}\,(X)|^2\right)\right] &= \sigma_1^2 + \cdots + \sigma_n^2 = \ln n + 1 + \gamma + O\left(\frac{1}{n^2}\right).\end{aligned}$$

Since the $O(1/n)$ and $O\left(1/n^2\right)$ error terms are "swamped" by the $O\left(\log^{-3/2} n\right)$, this gives us that for all $x \in \mathbb{R}$,

$$\Pr_X\left[\frac{\ln\left(|\mathrm{Det}\,(X)|^2\right) - (n\ln n - n + 1/2)}{\sqrt{\ln n + 1 + \gamma}} \leq x\right] = \int_{-\infty}^{x} \frac{e^{-y^2/2}}{\sqrt{2\pi}} dy \pm O\left(\frac{1}{\log^{3/2} n}\right).$$

Rearranging and applying Stirling's approximation now yields the theorem. ■

As a counterpoint to Theorem 29, let us give a simple argument for why $|\mathrm{Det}\,(X)|^2$ cannot be *exactly* lognormal, for any fixed n. Recall, from Section 7, that the pdf of $|\mathrm{Per}\,(X)|^2$ is a mixture

of exponential random variables (Lemma 3) and is therefore monotonically decreasing (Theorem 4). It is easy to see that the same is true for $|\mathrm{Det}\,(X)|^2$, since the proof of Lemma 3 works just as well for the determinant as for the permanent. By contrast, the pdf of a lognormal random variable is *not* monotonically decreasing, but is 0 at the origin.