Freie Universität Berlin

Title: **Sample Complexity of Device-Independently Certified “Quantum Supremacy”**

Author(s): Hangleiter, D., Kliesch, M., Eisert, J., & Gogolin, C.

Document type: Postprint

Terms of Use: Copyright applies. A non-exclusive, non-transferable and limited right to use is granted. This document is intended solely for personal, non-commercial use.

Citation: Hangleiter, D., Kliesch, M., Eisert, J., & Gogolin, C. (2019). Sample Complexity of Device-Independently Certified “Quantum Supremacy.” Physical Review Letters, 122(21). https://doi.org/10.1103/PhysRevLett.122.210502

arXiv:1812.01023v2 [quant-ph] 30 May 2019

Sample complexity of device-independently certified "quantum supremacy"

Dominik Hangleiter,[1] Martin Kliesch,[2] Jens Eisert,[1,3] and Christian Gogolin[4,5,6]

[1]*Dahlem Center for Complex Quantum Systems, Freie Universität Berlin, 14195 Berlin, Germany*
[2]*Institute for Theoretical Physics, Heinrich Heine University Düsseldorf, 40225 Düsseldorf, Germany*
[3]*Department of Mathematics and Computer Science, Freie Universität Berlin, 14195 Berlin, Germany*
[4]*ICFO-Institut de Ciencies Fotoniques, The Barcelona Institute of Science and Technology, 08860 Castelldefels (Barcelona), Spain*
[5]*Institute for Theoretical Physics, University of Cologne, 50937 Köln, Germany*
[6]*Xanadu, 372 Richmond St W, Toronto, M5V 1X6, Canada*

Results on the hardness of approximate sampling are seen as important stepping stones towards a convincing demonstration of the superior computational power of quantum devices. The most prominent suggestions for such experiments include boson sampling, IQP circuit sampling, and universal random circuit sampling. A key challenge for any such demonstration is to certify the correct implementation. For all these examples, and in fact for all sufficiently flat distributions, we show that any non-interactive certification from classical samples and a description of the target distribution requires exponentially many uses of the device. Our proofs rely on the same property that is a central ingredient for the approximate hardness results: namely, that the sampling distributions, as random variables depending on the random unitaries defining the problem instances, have small second moments.

I. INTRODUCTION

Quantum sampling devices have been hailed as promising candidates for the demonstration of "quantum (computational) supremacy"[1] [1]. The goal of any such experiment is to unambiguously demonstrate that quantum devices can solve some tasks both faster and with a more favourable scaling of the computational effort than any classical machine. At the same time, in the near term it is bound to use those small and computationally restricted quantum devices that are available before the arrival of universal, scalable, and fault-tolerant quantum computers. This challenge has sparked a flurry of experimental activity [2–7] and prompted the development of better classical sampling schemes for exact [8, 9] and imperfect realizations [10–13]. Due to the reality of experimental imperfections, the key theoretical challenge — achieved in Refs. [14–22] using Stockmeyer's approximate counting algorithm [23] — is to prove that even *approximately* sampling from the output distribution of the quantum device is classically hard.

In any such demonstration, the issue of certification is of outstanding importance [10, 17, 24–27]: To demonstrate something non-trivial, one not only needs to build a device that is designed to sample approximately from a classically hard distribution but at the same time, one needs to ensure from a feasible number of uses of the device (or its parts) that it actually achieves the targeted task. How can one convince a skeptical certifier that a quantum device, which supposedly does something no classical machine can do, actually samples from a distribution that is close enough to the ideal target distribution?

The arguably most elegant and most convincing certification would be one based on purely classical data, ideally only the samples produced by the device and a description of the target distribution. Such certification would be free of additional complexity-theoretic assumptions and device-independent, in that it would be agnostic to all implementation details of the device and would directly certify that the classically defined sampling problem was solved.

In this work, we rigorously prove for a broad range of sampling problems, specifically for boson sampling [14], universal random circuit sampling [15, 17], IQP circuit sampling [16, 24], and sampling from post-selected-universal 2-designs [20–22, 28, 29] that they cannot be efficiently certified from classical samples and a description of the target probability distribution. Ironically, it turns out that the same property of a distribution that allows to prove the known approximate-hardness results also forbids their non-interactive sample-efficient device independent certification, to the effect that with the known proof methods both properties cannot be achieved simultaneously in such schemes. We directly bound the sample complexity of certification, which means that we automatically also lower bound the computational complexity

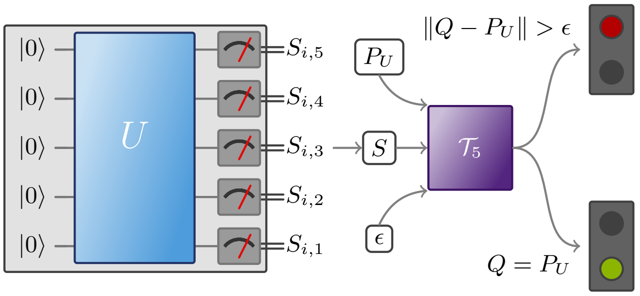

Figure 1. We consider the problem of certifying probability distributions of the form $P_U(S) = |\langle S|U|S_0\rangle|^2$ with an input state $|S_0\rangle = |0\rangle^{\otimes n}$ and a unitary $U \sim \mu_n$ drawn from some measure μ_n. Given $\epsilon > 0$ and access to an arbitrary-precision description of the target distribution P_U, the test $\mathcal{T}_n$ treats the sampler as a black box and receives a sequence $\mathcal{S} = (S_i)_{i=1}^s \sim Q$ of s samples from an unknown distribution Q. Given S the test is asked to output "Accept" if $Q = P_U$ and "Reject" if $\|Q - P_U\|_1 > \epsilon$ with high probability.

[1] Acknowledging the recent debate, we use the term "quantum (computational) supremacy" strictly in its established technical meaning [1].

and that our results cannot be circumvented by increasing the classical computational power of the certifier[2].

The specific question of certification we focus on here is (see Figure 1): Given unlimited computational power and a full description of the target distribution, how many samples from an unknown distribution are required to guarantee that this distribution is either identical to the target distribution or at least some preset distance away from it? This problem of distinguishing one (target) distribution from all sufficiently different alternatives is known as *identity testing* [31] in the property testing literature. Identity testing is an easier task than its robust version in which the certifier is moreover required to accept a constant-size region around the target distribution [26, 32]. At the same time, it is much harder than mere *state-discrimination*, where the task is to differentiate between two fixed distributions.

Lower bounds on the sample complexity of restricted state-discrimination scenarios [10] prompted the development of schemes [25] that allow to corroborate and build trust in experiments [6, 7, 33]. This helped spark interest in the problem of device-independent certification — on which there had not been much progress since [24]. In contrast to previous work [10], here, the certifier is given a full description of the target distribution[3] and unlimited computational power.

Our proofmakes use of a key property for the proof of hardness of approximate sampling, namely an upper bound on the second moments of the output probabilities with respect to the choice of a random unitary specifying the instance of the sampling problem. The bound on the second moments implies that the probabilities are concentrated around the uniform distribution and hence an anti-concentration property. This anti-concentration allows lifting results on the hardness of approximate sampling up to relative errors to ones for additive errors — provided relative-error approximation of the output probabilities is hard *on average*. It is thus a key property to prove hardness in the physically relevant case of approximate sampling that prevents a purely classical non-interactive certification of the output distribution, see Figure 2.

A central ingredient to our proof is a recent result by Valiant and Valiant [36] specifying the optimal sample complexity of certifying a known target distribution P. It can be stated as follows. Fix a preset distance $\epsilon > 0$ up to which we want to certify. Now, suppose we receive samples from a device that samples from an unknown probability distribution Q. Then — for some constants c_1, c_2 — it requires at least

$$c_1 \cdot \max\left\{\frac{1}{\epsilon}, \frac{1}{\epsilon^2}\|P_{-2\epsilon}^{-\max}\|_{2/3}\right\} \tag{1}$$

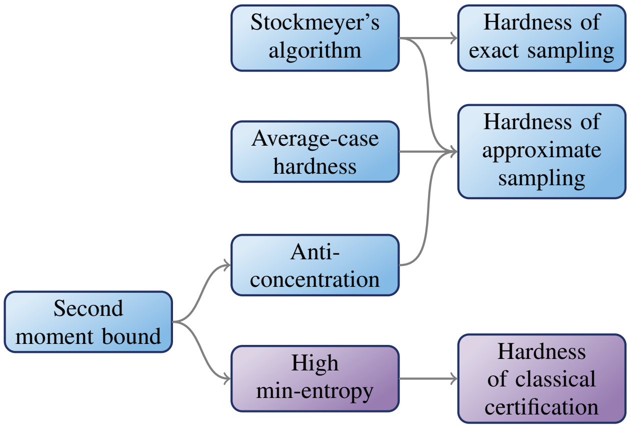

Figure 2. A high level overview of the approximate sampling "quantum supremacy" proofs of Refs. [14–16, 18–20, 34] using Stockmeyer's algorithm [23]. Invoking a worst-case hardness result for the calculation of the output probabilities of some circuit family, Stockmeyer's algorithm can be used to prove the hardness of exact sampling. The key properties of the output probabilities that allows to prove hardness of *approximate* sampling are that computing these probabilities is even hard *on average* and that the distribution anti-concentrates. We show that the same property that is essential to arrive at a hardness result for approximate sampling via anti-concentration also makes it hard to certify from classical samples and a complete description of the target distribution, even with unbounded *computational power*.

and at most

$$c_2 \cdot \max\left\{\frac{1}{\epsilon}, \frac{1}{\epsilon^2}\|P_{-\epsilon/16}^{-\max}\|_{2/3}\right\} \tag{2}$$

many samples to distinguish the case $P = Q$ from the case $\|P - Q\|_1 \geq \epsilon$ with high probability. Here $\|\cdot\|_1$ denotes the ℓ_1-norm reflecting the total-variation distance. The central quantity determining the sample complexity of certification is thus the quasi-norm $\|P_{-\epsilon}^{-\max}\|_{2/3}$ which is defined as follows. First, find the truncated distribution $P_{-\epsilon}^{-\max}$ by removing the tail of the target distribution P with weight at most ϵ as well as its largest entry, see Figure 3. Then, take the $\ell_{2/3}$-norm as given by $\|x\|_{2/3} = (\sum_i |x_i|^{2/3})^{3/2}$ for a vector x with entries x_i.

We now proceed in two steps. First, we show lower and upper bounds on the quantity $\|P_{-\epsilon}^{-\max}\|_{2/3}$ in terms of the largest

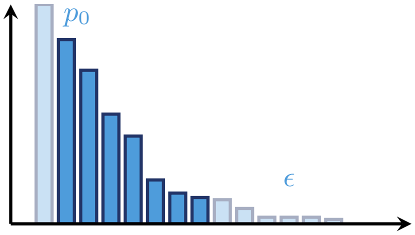

Figure 3. The vector $P_{-\epsilon}^{-\max}$ is obtained from P by removing the largest element p_0 of P as well as the smallest probabilities that accumulate to a total weight bounded by ϵ.

[2] This makes our results conceptually different from the observation of Brandão. This observation is based on a result by Trevisan, Tulsiani, and Vadhan [30], was reported by Aaronson and Arkhipov [25] and shows the following: For most unitaries U drawn from the Haar measure, and any fixed circuit size T, there exists a classical "cheating" circuit of size polynomially larger than T, whose output distribution can not be distinguished from the corresponding boson sampling distribution by any "distinguisher" circuit of size T.

[3] In particular, the certifier is given the value of all target probabilities to arbitrary precision.

probability p_0 occurring in P and its support $\|P_{-\epsilon}^{-\max}\|_0$ as given by

$$p_0^{-\frac{1}{2}}\left(1-\epsilon-p_0\right)^{3/2} \leq \|P_{-\epsilon}^{-\max}\|_{2/3} \\ \leq \left(1-p_0\right)\|P_{-\epsilon}^{-\max}\|_0^{\frac{1}{2}}. \tag{3}$$

Then it follows from Eqs. (1) and (3) that the sample complexity of certifying a distribution P up to a constant total-variation distance ϵ is essentially lower bounded by $1/\sqrt{p_0}$. Hence, if P is exponentially flat in the sense that the largest probability is exponentially small in the problem size (here, the number of particles), ϵ-certification requires exponentially many samples. Conversely, if $P_{-\epsilon/16}^{-\max}$ is supported on polynomially many outcomes only, sample-efficient certification is possible by the converse bound (2).

Second, we connect this result to the output distributions of "quantum supremacy" schemes. In all schemes that rely on the Stockmeyer argument, the problem instances are defined in terms of a unitary that is *randomly chosen* from some restricted family, e.g., linear optical circuits in the case of boson sampling [14] or random universal circuits [15, 17] in a qubit architecture . Specifically, we prove that with high probability over the choice of the random unitary, the distribution over outputs associated with this unitary is exponentially flat.

Putting everything together we obtain lower bounds on the sample complexity of certification for boson sampling, IQP circuit sampling and random universal circuit sampling with (sufficiently many) n particles. In all of these cases, the sample complexity scales at least as fast as

$$\frac{1}{\epsilon^2}\left(2^n\delta\right)^{1/4}, \tag{4}$$

with probability at least $1-\delta$ over the random choice of the unitary.

The upshot is: a key ingredient of the proof of approximate sampling hardness as effected by the random choice of the unitary prohibits sample-efficient certification.

We show that one cannot hope for purely classical, non-interactive, device-independent certification of the proposed quantum sampling problems. This highlights the importance of devising alternative schemes of certification, or improved hardness results for more peaked distributions. We hope to stimulate research in such directions.

A particularly promising avenue of this type of certification has been pioneered by Shepherd and Bremner [24]: By allowing the certifier to choose the classical input to the sampling device rather than drawing it fully at random, it is under some plausible cryptographic assumptions possible to efficiently certify the correct implementation of a quantum sampler from its classical outcomes. This is facilitated by checking a previously hidden bias in the obtained samples and has been achieved for a certain family of IQP circuits [24]. However, in contrast to Ref. [16], there is no approximate sampling hardness result for this family.

Focusing on so-called *relational problems* as opposed to sampling problems, it has been argued via new complexity-theoretic conjectures that the task *HOG* of outputting the *heavy outcomes* of a quantum circuit (those outcomes with probability weight larger than the median of its output distribution) is classically intractable [37]. Clearly, this task is sample-efficiently checkable via its in-built bias, but still requires exponential classical computation to determine the probabilities of the obtained samples, which are compared to the median.

Taking a pragmatic stance, one can make additional assumptions on the device. In fact, only recently has it been shown [17] that cross-entropy measures [15] provide direct bounds on the total-variation distance provided the entropy of the real distribution is larger than that of the target distribution. One might also be content with weaker notions of certification in total-variation distance such as the certification of a coarse-grained version of the full output distribution [38]. Coarse-graining procedures are practically useful as corroboration schemes when distinguishing against plausible alternative distributions such as the uniform distribution, but of course fail to certify against adversarial distributions on the full sample space. All such approaches yield sample-efficient certificates that require exponential computational effort, rendering them feasible at least for intermediate-scale devices.

Another way to certify a sampling device is the certification of the entire machine from its components. However, such schemes need to make assumptions about the absence of unwanted influences between the components such as crosstalk. In a similar vein, one can make use of implementation details and give the certifier some quantum capabilities such as access to a small quantum computer [39], the ability to manipulate single qubits [40], or to measure the output quantum state in different bases with trusted quantum detectors [27, 41] to devise certificates even in non-iid. settings [42]. In this way, one can partially trade-in the simplicity of sampling schemes for better certifiability.

It is interesting to note the connection of our result with results on classical simulation. Similarly to our findings for the case of certification, Schwarz and Van den Nest [35] find that for certain natural families of quantum circuits (including IQP circuits) classical simulation is possible for highly concentrated distributions, but impossible for flat ones, see Figure 4. This again gives substance to the interesting connection between superior computational power, the flatness of the distribution and the impossibility of an efficient certification.

Curiously, at the same time, the property that prohibits sample-efficient certification is by no means due to the hardness of the distribution. It is merely the flatness of the distribution on an exponential-size sample space as effected by the random choice of the unitary that is required for the approximate hardness argument via Stockmeyer's algorithm and standard conjectures. The uniform distribution on an exponentially large sample space, which is classically efficiently samplable, can also not be sample-efficiently certified.

A further noteworthy connection is that to Shor's algorithm. The output distribution of the quantum part of Shor's algorithm is typically spread out over super-polynomially many outcomes and can hence neither be efficiently simulated via the algorithm of Schwarz and Van den Nest [35], nor certified as we show here. However, after the classical post-processing,

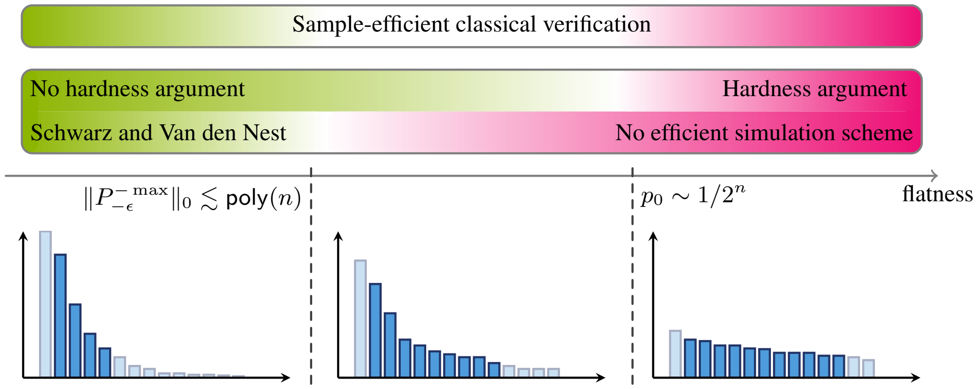
Sample-efficient classical verification
No hardness argument
Hardness argument
Schwarz and Van den Nest
No efficient simulation scheme
$\|P_{-\epsilon}^{-\max}\|_0 \lesssim \mathsf{poly}(n)$
$p_0 \sim 1/2^n$
flatness

Figure 4. Hardness and certification in terms of the flatness of $P_{-\epsilon}^{-\max}$ for the example of IQP circuits [16, 24] on n qubits as obtained from the present result and the classical simulation algorithm of Schwarz and Van den Nest [35]. There, it is shown that a certain natural family of quantum circuits (including IQP circuits) can be efficiently simulated on a classical computer if the output distribution is essentially concentrated on a polynomial number of outcomes only. In this case, i.e., for $\|P_{-\epsilon}^{-\max}\|_0 \lesssim \mathsf{poly}(n)$, the output distribution is also sample-efficiently certifiable as the bounds (2) and (3) show. Their classical simulation algorithm breaks down if the distribution is essentially spread out over more than polynomially many outcomes, and we even have a rigorous hardness argument by Bremner *et al.* [16] for exponentially flat distributions. Conversely, the number of samples required for certification becomes prohibitively large if the distribution is exponeentially spread out, as measured by the $\ell_{2/3}$-norm (1). Nevertheless, as we illustrate here, there could be "room in the middle" where, for reasonably but not exponentially flat distributions, one may hope to find tasks that are both classically intractable and sample-efficiently certifiable in a device-independent fashion.

the output distribution is strongly concentrated on few outcomes — the factors — from which one can verify the correct working of the algorithm. A certification of the intermediate distribution is simply not necessary to demonstrate a quantum speedup in Shor's algorithm, as its speedup is derived from it solving a problem in NP and not from it sampling close to a hard distribution. This shows that while intermediate steps of a computation might not be certifiable, the final outcome may well be. Whether this is enough to demonstrate a speedup depends on the nature of the hardness argument. In fact, the abovementioned task HOG [37] bears many similarities to factoring and its certifiability from the outcomes of the algorithm.

We hope that our result will stimulate research into new ways of proving hardness of approximate sampling tasks that are more robust than those based on anti-concentration, as well as into devising alternative verification schemes possibly based on mild and physically reasonable assumptions on the sampling device or the verifier.

II. SETUP AND DEFINITIONS

Let us begin the technical part of this work by setting the notation. We use the Landau symbols O and Ω for asymptotic upper and lower bounds and Θ for their conjunction. For any $j \in \mathbb{Z}^+$ we employ the short hand notation $[j] := \{1, \dots, j\}$ for the range. By $\log$ we denote the logarithm to basis 2. For any vector $x \in \mathbb{R}^n$ we define $\|x\|_\infty := \max_{i\in[n]} |x_i|$ and $\|x\|_p := \left(\sum_{i=1}^n |x_i|^p\right)^{1/p}$ for $0 < p < \infty$ and take $\|x\|_0 := |\{i \in [n] : x_i \neq 0\}|$ to denote the number of non-zero elements of x. Thus, $\|\cdot\|_p$ is the standard ℓ_p-norm whenever $p \geq 1$. For $p \in (0,1)$, $\|\cdot\|_p$ no longer satisfies the triangle inequality, but it is obviously absolutely homogeneous and still defines a quasinorm. We will also make use of the α-Rényi entropies, which for any probability vector $P = (p_1, \dots, p_n)$, $p_i \geq 0$, $\sum_i p_i = 1$ and $0 \leq \alpha \leq \infty, \alpha \neq 1$ are defined to be

$$H_\alpha(P) := \frac{\alpha}{1-\alpha} \log \|P\|_\alpha. \tag{5}$$

We refer to $H_\infty(P) = -\log \max_{i\in[n]} p_i$ as the *min-entropy* of P.

We are now in the position to formalize the notion of a test that certifies that a given device indeed samples from a distribution sufficiently close to a given target distribution. More precisely, we consider a family of sample spaces $\mathcal{E}_n$ with $n \in \mathbb{Z}^+$. The parameter n will be the natural problem size in the concrete examples below. The object of interest is a classical algorithm $\mathcal{T}_n$ which receives a description of a target distribution P_n over $\mathcal{E}_n$, and a sequence $\mathcal{S} \sim Q^s$ of s samples $S_1, \dots, S_s \in \mathcal{E}_n$ that have been drawn i.i.d. from some distribution Q over $\mathcal{E}_n$ and must output 1 or 0 for "accept" or "reject", respectively. We illustrate this notion of certification in Figure 1.

Definition 1 (Certification test)**.** *For any n let P be a (target) probability distribution on a sample space $\mathcal{E}$. We call $\mathcal{T} : \mathcal{E}^s \to \{0,1\}$ an* ϵ-certification test *of P from s samples if the following completeness and soundness conditions are satisfied for any distribution Q over $\mathcal{E}$:*

$$Q = P \;\Rightarrow\; \Pr_{\mathcal{S}\sim Q^s}[\mathcal{T}(\mathcal{S}) = 1] \geq \frac{2}{3}, \tag{6}$$

$$\|P - Q\|_1 > \epsilon \;\Rightarrow\; \Pr_{\mathcal{S}\sim Q^s}[\mathcal{T}(\mathcal{S}) = 1] < \frac{1}{3}. \tag{7}$$

For a family $\{P_n\}$ of probability distributions we call a family of tests $\{\mathcal{T}_n\}$ a sample-efficient ϵ-certification test *if for every n $\mathcal{T}_n$ is an ϵ-certification test from $s \in O(\mathsf{poly}(n, 1/\epsilon))$ samples.*

Our notion of certification is *device-independent* in the sense that it does not assume anything about the internal working of the sampler (not even whether it is quantum or classical), but uses only the classical samples it outputs and a classical description of the target distribution. Among such device-independent certification scenarios, our scenario is the most general one in the sense that the certifier is given *all* the information contained in the target distribution. In particular, it is crucial that we explicitly allow the certification test $\mathcal{T}_n$ to depend on all details of the target distribution P_n.

As we are not concerned with the computational complexity of the test, but only its sample complexity, we allow the certification algorithm *unlimited* computational power. In particular, it does not matter how exactly $\mathcal{T}_n$ is given access to a description of P_n, but for the sake of concreteness $\mathcal{T}_n$ can be thought of as having access to an oracle that provides the probabilities $P_n(S)$ of all $S \in \mathcal{E}_n$ up to arbitrary precision. Sample-efficiency is clearly a necessary requirement for computational efficiency of a test, as any test takes at least the time it needs to read in the required number of samples, so that lower bounds on the sample complexity are stronger than such on the computational complexity.

We note that our notion of certification corresponds to what in the literature on property testing [31] is called *identity testing* with a fixed target distribution. It stands in contrast to the previously considered task of *state discrimination* [10, 25], where the task is to decide from which of two given distributions P or Q a device samples. ϵ-certification in the sense of Definition 1 is more demanding in the sense that P has to be distinguished from *all* distributions Q such that $\|P - Q\|_1 \geq \epsilon$. It is precisely this type of certification that is necessary to convince a skeptic of "quantum supremacy" via, say boson sampling, as the hardness results on approximate boson sampling only cover distributions within a small ball in ℓ_1-norm around the ideal target distribution. A device sampling from a distribution further away from the ideal distribution, might still be doing something classically intractable, but this cannot be concluded from the hardness of approximate boson sampling.

III. NO CERTIFICATION OF FLAT DISTRIBUTIONS

This section is concerned with the question of whether distributions with a high *min-entropy* can be certified in a sample-efficient way. The main insights into this question come from a work by Valiant and Valiant [36] on property testing, which gives a *sample-optimal* certification test (up to constant factors) for any fixed distribution P, as well as a lower bound on the sample complexity of certification. The result is stated in terms of an $\ell_{2/3}$-norm of a vector obtained from the distribution. Our main technical contribution is to find bounds on these quasi-norms that are relevant in the context of certifying "quantum supremacy" distributions.

To state the main result of Ref. [36], we adapt their following notation and illustrate it in Figure 3. For any vector of non-negative numbers P,

(i) let $P^{-\max}$ be the vector obtained from P by setting the largest entry to zero, and

(ii) let $P_{-\epsilon}$ be the vector obtained from P by iteratively setting the smallest entries to zero, while the sum of the removed entries remains upper bounded by $\epsilon > 0$.

It turns out that the optimal sample complexity for ϵ-certifying any distribution P is essentially given by $\frac{1}{\epsilon^2}\|P^{-\max}_{-\epsilon}\|_{2/3}$. The intuition is that any ϵ deviation from P that is contained in either the largest probability or the tail of the distribution is easily detected. Intuitively, this is because a constant deviation in these parts of the distribution will be visible in the samples obtained with high probability [36]. More precisely, the following upper and lower bounds on the sample complexity of certification hold:

Theorem 2 (Optimal certification tests [36])**.** *There exist constants $c_1, c_2 > 0$ such that for any $\epsilon > 0$ and any target distribution P, there exists an ϵ-certification test from $c_1 \max\{\frac{1}{\epsilon}, \frac{1}{\epsilon^2}\|P^{-\max}_{-\epsilon/16}\|_{2/3}\}$ many samples, but there exists no ϵ-certification test from fewer than $c_2 \max\{\frac{1}{\epsilon}, \frac{1}{\epsilon^2}\|P^{-\max}_{-2\epsilon}\|_{2/3}\}$ samples.*

We note that $\|P^{-\max}_{-\epsilon}\|_{2/3} \leq \|P\|_{2/3}$ for any P, and in many cases the former is only a constant factor away from the latter. We obtain the following general bounds on $\|P^{-\max}_{-\epsilon}\|_{2/3}$ in terms of the min-entropy and support of P:

Lemma 3 (Bounds on $\|P^{-\max}_{-\epsilon}\|_{2/3}$)**.**

$$\begin{aligned} 2^{\frac{1}{2}H_\infty(P)}\left(1 - \epsilon - 2^{-H_\infty(P)}\right)^{3/2} &\leq \|P^{-\max}_{-\epsilon}\|_{2/3} \\ &\leq \left(1 - 2^{-H_\infty(P)}\right)\|P^{-\max}_{-\epsilon}\|_0^{\frac{1}{2}}. \end{aligned} \tag{8}$$

To get a feeling for what these bounds imply, let us consider two special cases and sufficiently small ϵ. If for some constant κ it holds that $H_\infty(P) = \log(\kappa\,|\mathcal{E}_n|)$, they imply the following lower bound on the required minimal number of samples, $s_{\min}$:

$$s_{\min}^2 \geq c_2^2 \kappa \frac{|\mathcal{E}_n|}{\epsilon^4}\left(1 - 2\epsilon - \frac{1}{\kappa\,|\mathcal{E}_n|}\right)^3. \tag{9}$$

For all distributions whose min-entropy is essentially given by the logarithm of the size $|\mathcal{E}_n|$ of the sample space, the sample complexity for certification thus scales at least as the square root of that size. If, on the contrary, $P_{-\epsilon/16}$ has support on at most $s \geq \|P_{-\epsilon/16}\|_0$ many probabilities we have the following upper bound

$$s_{\text{suf}} \leq c_1 \frac{1 - \frac{\epsilon}{16}}{\epsilon^2}\sqrt{s} \tag{10}$$

on the number of samples s_{suf} that is sufficient for ϵ-certification. This bound implies that distributions supported only on polynomially many outcomes can be certified from polynomially many samples.

Proof of Lemma 3. For the lower bound, we use that concavity of the function $x \mapsto x^{2/3}$ implies that for any fixed $x^* > 0$ and any $0 \leq x \leq x^*$ we have

$$x^{2/3} \geq \frac{{x^*}^{2/3}}{x^*}\, x = {x^*}^{-1/3}\, x \tag{11}$$

and thus for any (not necessarily normalized) $\tilde{P} := (\tilde{p}_1, \dots, \tilde{p}_{\tilde{n}})$ with $\tilde{p}_i \geq 0$

$$\begin{aligned} \|\tilde{P}\|_{2/3}^{2/3} &= \sum_{i=1}^{\tilde{n}} \tilde{p}_i^{2/3} \geq \sum_{i=1}^{\tilde{n}} (\|\tilde{P}\|_\infty^{-1/3}\, \tilde{p}_i) & (12)\\ &= \|\tilde{P}\|_\infty^{-1/3}\, \|\tilde{P}\|_1 . & (13) \end{aligned}$$

Using this for $\tilde{P} = P^{-\max}_{-\epsilon}$ and that both $\|P^{-\max}_{-\epsilon}\|_\infty \leq \|P\|_\infty$ and $\|P^{-\max}_{-\epsilon}\|_1 \geq 1 - \epsilon - \|P\|_\infty$ finally implies the lower bound.

For the upper bound, we use that for any vector v and $0 < p < q \leq \infty$ (see, e.g., Ref. [43, Eq. (A.3)])

$$\|v\|_p \leq s^{\frac{1}{p} - \frac{1}{q}} \|v\|_q , \tag{14}$$

where $s \geq \|v\|_0$. Inserting $p = 2/3$ and $q = 1$, one obtains for $v = P^{-\max}_{-\epsilon}$

$$\|P^{-\max}_{-\epsilon}\|_{2/3} \leq \|P^{-\max}_{-\epsilon}\|_0^{\frac{1}{2}}\, \|P^{-\max}_{-\epsilon}\|_1 \, . \tag{15}$$

□

Valiant and Valiant's result [36] also has immediate consequences on the certifiability of post-selected probability distributions, such as those arising in boson sampling [14]. A certification algorithm has to distinguish the target distribution P from all probability distributions that are at least ϵ-far away from P. That is true, in particular, for distributions that differ from P by at least ϵ in ℓ_1-norm only on some part $\mathcal{F}$ of the sample space, but are identical with P on its complement $\mathcal{F}^c$. Intuitively one can expect that to distinguish such distributions, samples from $\mathcal{F}^c$ do not help. One might hence expect that it should be possible to lower bound the sample complexity of certifying the full distribution by the sample complexity of the post-selected distribution on some subspace $\mathcal{F}$ of the sample space, at least as long as the post-selection probability is not too low.

To make this intuition precise, define for any probability distribution P and any subset $\mathcal{F} \subset \mathcal{E}$ the restriction $P_{\upharpoonright \mathcal{F}} := (p_i)_{i \in \mathcal{F}}$ of P to $\mathcal{F}$ (no longer normalized), as well as the post-selected probability distribution $P_{\mathcal{F}} := P_{\upharpoonright \mathcal{F}} / P(\mathcal{F})$, with post-selection probability $P(\mathcal{F}) := \|P_{\upharpoonright \mathcal{F}}\|_1$.

Lemma 4 (Lower bounds with post-selected distributions)**.** *Let P be a probability distribution on $\mathcal{E}$. Then with c_2 the constant from Theorem 2 and for any $\epsilon > 0$ and $\mathcal{F} \subset \mathcal{E}$, there exists no ϵ-certification test of P from fewer than $c_2 \max\{\frac{1}{\epsilon}, \frac{1}{\epsilon^2} P(\mathcal{F})\, \|(P_{\mathcal{F}})^{-\max}_{-2\epsilon/P(\mathcal{F})}\|_{2/3}\}$ samples.*

Proof of Lemma 4. For any $\mathcal{F} \subset \mathcal{E}$ we have

$$\begin{aligned} \|P^{-\max}_{-\epsilon}\|_{2/3} &\geq \|(P^{-\max}_{-\epsilon})_{\upharpoonright \mathcal{F}}\|_{2/3} & (16)\\ &\geq \|(P_{\upharpoonright \mathcal{F}})^{-\max}_{-\epsilon}\|_{2/3} & (17)\\ &= P(\mathcal{F}) \|(P_{\upharpoonright \mathcal{F}})^{-\max}_{-\epsilon} / P(\mathcal{F})\|_{2/3} & (18)\\ &= P(\mathcal{F}) \|(P_{\mathcal{F}})^{-\max}_{-\epsilon/P(\mathcal{F})}\|_{2/3} . & (19) \end{aligned}$$

Here, the first inequality becomes an equality in case $\mathcal{F}$ contains the support of $P^{-\max}_{-\epsilon}$. The second inequality becomes an equality whenever the smallest probabilities with weight not exceeding ϵ as well as the largest probability lie inside of $\mathcal{F}$. Finally, the last equality follows from the fact that when renormalizing $P_{\upharpoonright \mathcal{F}}$ we also need to renormalize the subtracted total weight ϵ by the same factor. The claim then straightforwardly follows from Theorem 2. □

A non-trivial bound for the sample complexity is therefore achieved only in case the post-selected subspace has at least weight $P(\mathcal{F}) > 2\epsilon$. This is due to the strength of Valiant and Valiant's result [36] in that a part of the distribution with total weight 2ϵ does not influence the minimally required sample complexity of ϵ-certification and this part might just be supported on $\mathcal{F}$.

IV. "QUANTUM SUPREMACY" DISTRIBUTIONS CANNOT BE CERTIFIED

We will now apply the result of the previous section to the case of certifying "quantum supremacy" distributions. As a result, we find that prominent schemes aimed at demonstrating "quantum supremacy", most importantly boson sampling, cannot be certified from polynomially many classical samples and a description of the target distribution alone. To be more concrete, in the context of "quantum supremacy", there has recently been an enormous activity aiming to devise simple sampling schemes, which show a super-polynomial speedup over any classical algorithm even if the sampling is correct only up to a constant ℓ_1-norm error [14–16, 18–21, 34, 44]. The method used to prove all the aforementioned speedups is the proof technique pioneered by Terhal and DiVincenzo [45] for the case of exact sampling, which is based on an application of Stockmeyer's approximate counting algorithm [23]. Stockmeyer's algorithm is used to prove that, conditioned on a conjecture on the average-case hardness of certain problems, the polynomial hierarchy would collapse if the respective distribution could be sampled efficiently classically. To extend this proof technique to the case of approximate sampling up to an (additive) ℓ_1-norm error requires an additional property on the sampled distribution, namely, *anti-concentration* [14, 16].

More precisely, the aforementioned tasks all fit the following schema: Given the problem size n, start from a reference state vector $|S_0\rangle$ from a Hilbert space $\mathcal{H}_n$ and apply a unitary U drawn with respect to some measure μ_n on the corresponding unitary group. The resulting state is then measured in the computational basis, thereby resulting in outcome S with probability $P_U(S) := |\langle S|U|S_0\rangle|^2$. One then says that the distribution over P_U induced by this procedure *anti-concentrates* if

$$\exists\, \alpha,\gamma > 0: \ \forall n \in \mathbb{Z}_+ \text{ and } \forall S \in \mathcal{E}_n: \\ \Pr_{U\sim\mu_n}\left(P_U(S) \geq \frac{\alpha}{|\mathcal{E}_n|}\right) \geq \gamma\,. \tag{20}$$

In words this roughly means: For any n the induced distribution over P_U has the property that for any fixed outcome S it does not become too unlikely that the probability of getting that outcome is much smaller than it would be for the uniform distribution.

It is intuitive that due to normalization, anti-concentration also implies that not too much of the probability weight can be concentrated in few outcomes, hence the name. This is however slightly misleading, as anti-concentration does not in itself imply a that P_U also needs to have a high min-entropy with high probability. To see this, take any $\alpha, \gamma(\alpha)$-anti-concentrating scheme of drawing probability distributions P_U. Construct $\tilde{P}_U$ from P_U by dividing all probabilities in half and adding their joint weight to $P_U(0)$. The resulting scheme to construct $\tilde{P}_U$ is now still $\alpha/2, \gamma(\alpha)$-anti-concentrating, but has min-entropy $H_\infty(\tilde{P}_U) \leq \log(2)$ with probability one.

However, most known proofs of anti-concentration [11, 16, 20, 25] (with the notable exception of Morimae's hardness result [44][4]) rely on the Paley-Zygmund inequality

$$\Pr(Z > a\mathbb{E}[Z]) \geq (1-a)^2 \frac{\mathbb{E}[Z]^2}{\mathbb{E}[Z^2]}\,, \tag{21}$$

for a random variable $Z \geq 0$ with finite variance and $0 \leq a \leq 1$. Anti-concentration is then proved by deriving a bound on the second moment of the distribution $\{P_U(S)\}_{U\sim\mu_n}$, and, as we will see in the next section, the same second moment bound that is used to derive anti-concentration implies a high min-entropy. We lay out the proof structure and the role that second moments play in Figure 2.

A. Second moments bound the min-entropy

We now turn to showing that with high probability over the choice of U, the second moment of the distribution $\{P_U(S)\}_{U\sim\mu_n}$ (for fixed S) — implying the anti-concentration property (20) — yields a lower bound on the min-entropy of the output distribution P_U of any fixed unitary U with high probability. Lemma 3 then implies that distributions with exponentially (in n) small second moments cannot be certified from polynomially many samples with high probability. Thus, the *very property* that implies sampling hardness of a distribution P_U up to an additive total-variation distance error also implies that the same distribution cannot be efficiently certified from classical samples only.

Lemma 5 (Tail bound for the min-entropy)**.** *For any $n \in \mathbb{Z}^+$, let P_U be a distribution on $\mathcal{E}_n$ induced via $P_U(S) = |\langle S|U|S_0\rangle|^2$, $U \sim \mu_n$ by a corresponding measure μ_n on the unitary group. Then, with probability at least $1-\delta$ over the choice of $U \sim \mu_n$,*

$$H_\infty(P_U) \geq \frac{1}{2}\left(\log\delta - \log \sum_{S\in\mathcal{E}_n} \mathbb{E}_{U\sim\mu_n}[P_U(S)^2]\right). \tag{22}$$

The following arguments used to derive the lemma, in fact, hold for more general families of probability distributions P_U where U need not be unitary without any scaling in n.

Proof of Lemma 5. We proceed as follows: First, we prove a lower bound on the typical Rényi-2-entropy of P_U using the second moment of $\{P_U\}_{U\sim\mu_n}$, and then use equivalence of the α-Rényi entropies for $\alpha > 1$.

Analogously to Ref. [25, App. 11], we use Markov's inequality to obtain that with probability at least $1-\delta$ over the choice of U, we have

$$H_2(P_U) := -\log \sum_{S\in\mathcal{E}_n} P_U(S)^2 \tag{23}$$

$$\geq -\log\left(\frac{1}{\delta}\cdot \mathbb{E}_{U\sim\mu_n}\left[\sum_{S\in\mathcal{E}_n} P_U(S)^2\right]\right). \tag{24}$$

What is more, one can show that all α-Rényi entropies for $\alpha > 1$ are essentially equivalent and in particular [46]

$$H_\alpha(P) \geq H_\infty(P) \geq \frac{\alpha-1}{\alpha} H_\alpha(P), \tag{25}$$

for any distribution P on $\mathcal{E}_n$ from which the claim follows[5]. □

We note that, indeed, the notion of anti-concentration as formalized in Eq. (20) in itself does not necessarily imply that the output distribution of every (or most) fixed unitaries have high min-entropy. This is because anti-concentration merely requires that the tails of the distribution have sufficient (constant) weight, while allowing for few large probabilities. Nevertheless, in prominent cases, an anti-concentration result derives from bounds on the 2-Rényi entropy.

B. "Quantum supremacy" distributions are flat

We now apply our results to the most prominent examples of "quantum supremacy" schemes — boson sampling [14], IQP circuits [16], and universal random circuits [15]. We will conclude from Lemmas 3–5 that these schemes cannot be efficiently certified from polynomially many samples only. In the following, we show that for all of these schemes with output distribution P_U we have that $H_\infty(P_U) \in \Omega(n)$ and hence

[4] There, Morimae proves anti-concentration for the output distribution so-called one-clean qubit model (DQC1) in a direct manner.

[5] We show Eq. (25) as well as an alternative proof for Lemma 5 in Section A.

that the minimal sample complexity for certification scales exponentially in n. More precisely, both for boson sampling and the qubit-based schemes mentioned above all of which we precisely define below in Sections V A–V C, we obtain the following lower bounds.

Theorem 6 (Lower bounds on certifying boson sampling)**.** *Let $0 < \epsilon < 1/2$, $n \in \mathbb{Z}_+$ sufficiently large and $m \in \Theta(n^\nu)$. Under the conditions on ν used in Ref. [14] to prove the hardness of approximate boson sampling, and with high probability over the random choice of the unitary, there exists no ϵ-certification test of boson sampling with n photons in m modes from $s < s_{\min}$ many samples, where*

$$s_{\min} \in \Omega\left(2^n/\epsilon^2\right) . \tag{26}$$

In Section V C, we discuss in detail the conditions under which Aaronson and Arkhipov's hardness argument [14] holds and provide a full version of the theorem as Theorem 9. The key ingredient for this to be the case is the closeness of the measure obtained by taking $n \times n$-submatrices of Haar random unitaries $U \in U(m)$ and the Gaussian measure on $n \times n$-matrices. This is provably the case for $\nu > 5$, but is conjectured to hold even for $\nu > 2$ [14]. Our bound on $s_{\min}$ (see Theorem 9) holds with exponentially high probability (in n) for $\nu > 3$. In the case $\nu > 2$ our result holds only with polynomially high probability and fails to cover a small set of the instances. The argument proving Theorem 6 easily extends to certain variants of *quantum Fourier sampling* [47], the output probabilities of which are also given by permanents of nearly Gaussian matrices.

Theorem 7 (Lower bounds on certifying random qubit schemes)**.** *For $0 < \epsilon < 1/2$ and sufficiently large n, with probability at least $1 - \delta$, there exists no ϵ-certification test from $s < s_{\min}$ many samples for*

a. IQP circuit sampling on n qubits, where

$$s_{\min} \in \Omega\left(2^{n/4}\delta^{1/4}/\epsilon^2\right) . \tag{27}$$

b. $\tilde{\varepsilon}$-approximate spherical 2-design sampling on n qubits, and in particular, depth-$(O(n^2) + O(n \log 1/\tilde{\varepsilon}))$ local random universal circuits, where

$$s_{\min} \in \Omega\left(\frac{2^{n/4}\delta^{1/4}}{\epsilon^2(1+\tilde{\varepsilon})^{1/4}}\right) . \tag{28}$$

The result of Theorem 7 applies to any circuit family $\mathcal{U}$ such that $\{U|S_0\rangle\}_{U\sim\mathcal{U}}$ forms a relative $\tilde{\varepsilon}$-approximate spherical 2-design, for which the second moments are upper bounded as in Eq. (33). This applies, in particular, to the random universal circuits of Refs. [15, 17, 48, 49] as well as other families of random circuits that have been proposed for the demonstration of "quantum supremacy" such as Clifford circuits with magic-state inputs [20, 22], diagonal unitaries [20, 28] and conjugated Clifford circuits [21].

Proofs of Theorems 6 and 7. We use Theorem 2 and Lemmas 3–5 as well as the lower bounds (50), (31), and (34) on the min-entropy of the respective output distributions as given in the following sections. What is more, we use that for $0 < \epsilon < 1/2$ and sufficiently large n the term $(1 - 2\epsilon - 2^{-H_\infty(P_U)})^{3/2}$ can be lower-bounded by a constant and, hence, be dropped inside the Ω. □

V. DETAILS ON RANDOM SAMPLING SCHEMES AND PROOFS OF THEOREMS 6 AND 7

We now turn to describing details on distributions arising from boson sampling, IQP circuits and universal random circuits, and present the proofs of the aforementioned theorems.

A. IQP circuits

An IQP circuit [24] is a quantum circuit of commuting gates that is drawn uniformly at random from the family $\mathcal{U}_{n,\mathrm{IQP}}$ on n qubits. The sample space is therefore given by $\mathcal{E}_n = \{0,1\}^n$. This family as formulated by Bremner, Montanaro, and Shepherd [16] is defined by a set of angles A, e.g., $A = \{0, \pi/8, \ldots, 7\pi/8\}$. An instance $U_W \in \mathcal{U}_{n,\mathrm{IQP}}$ with $W := (w_{i,j})_{i,j=1,\ldots,n}$ and $w_{i,j} \in A$ drawn uniformly at random, is then given by the following prescription

$$U_W = \exp\left[\mathrm{i}\left(\sum_{i<j} w_{i,j} X_i X_j + \sum_i w_{i,i} X_i\right)\right] , \tag{29}$$

where X_i is the Pauli-X matrix acting on site i. In other words, on every edge (i,j) of the complete graph on n qubits a gate $\exp(\mathrm{i} w_{i,j} X_i X_j)$ with edge weight $w_{i,j}$ and on every vertex i a gate $\exp(\mathrm{i} w_{i,i} X_i)$ with vertex weight $w_{i,i}$ is performed.

For the output distribution of IQP circuits, Bremner *et al.* [16, Appendix F] prove the second-moment bound

$$\mathbb{E}_W[|\langle S|U_W|0\rangle|^4] \leq 3 \cdot 2^{-2n} . \tag{30}$$

By Lemma 5, this implies the following min-entropy bound

$$H_\infty(P_{U_W}) \geq \frac{1}{2}\left(n + \log\frac{\delta}{3}\right) , \tag{31}$$

which holds with probability at least $1 - \delta$ over the choice of U_W.

B. Universal random circuits and spherical 2-designs

A universal random circuit on n qubits is defined by a universal gate set $\mathcal{G}$ comprising one- and two-qubit gates which give rise to the depth-N family $\mathcal{U}_{\mathcal{G},N}$. A circuit $U \in \mathcal{U}_{\mathcal{G},N}$ is then constructed according to the standard prescription of choosing one- or two-qubit gates $G \in \mathcal{G}$ and the qubits they are applied to at random [48], or according to some more specific prescription such as the one of Boixo *et al.* [15].

For the case of the random universal circuits of Ref. [15] there is evidence that the output distribution of fixed instances is essentially given by an exponential (Porter-Thomas) distribution P_{PT} whose second moment is given by [20, Eq. (8)]

$$\mathbb{E}_{p\sim P_{\mathrm{PT}}}[p^2] = \frac{2}{|\mathcal{E}_n|(|\mathcal{E}_n|+1)}\,. \tag{32}$$

This is provably true for the local random universal circuits investigated by Brandão, Harrow, and Horodecki [48] by the fact that the resulting circuit family forms a relative $\tilde{\varepsilon}$-approximate unitary 2-design μ in depth $O(n^2) + O(n\log 1/\tilde{\varepsilon})$ [48] so that

$$\mathbb{E}_{U\sim\mu}[|\langle S|U|S_0\rangle|^4] \leq \frac{2(1+\tilde{\varepsilon})}{|\mathcal{E}_n|(|\mathcal{E}_n|+1)}. \tag{33}$$

Likewise, for any circuit family $\mathcal{U}_n$ on n qubits such that $\{U|0\rangle\}_{U\sim\mathcal{U}_n}$ forms a relative $\tilde{\varepsilon}$-approximate spherical 2-design, the second moments are bounded as in Eq. (33). For all such circuit families, using Lemma 5, we thus obtain the min-entropy bound

$$H_\infty(P_U) \geq \frac{1}{2}\left(n + \log\frac{\delta}{2(1+\tilde{\varepsilon})}\right)\,, \tag{34}$$

which holds with probability at least $1-\delta$ over the choice of U.

C. Boson sampling

In the boson sampling problem $n \geq 1$ photons are injected into the first n of $m \in \mathsf{poly}(n)$ modes which are transformed in a linear-optical network via a mode transformation given by a Haar-random unitary $U \in U(m)$ and then measured in the Fock basis. The sample space of boson sampling is given by

$$\mathcal{E}_n := \Phi_{m,n} := \left\{(s_1,\dots,s_m) : \sum_{j=1}^{m} s_j = n\right\}, \tag{35}$$

i.e., the set of all sequences of non-negative integers of length m which sum to n. Its output distribution $P_{\mathrm{bs},U}$ is

$$P_{\mathrm{bs},U}(S) := |\langle S|\varphi(U)|1_n\rangle|^2. \tag{36}$$

Here, the state vector $|S\rangle$ is the Fock space vector corresponding to a measurement outcome $S \in \Phi_{m,n}$, $|1_n\rangle$ is the initial state vector with $1_n := (1,\dots,1,0,\dots,0)$, and $\varphi(U)$ the Fock space (metaplectic) representation of the implemented mode transformation U.

The distribution $P_{\mathrm{bs},U}$ can be expressed [14, 50] as

$$P_{\mathrm{bs},U}(S) = \frac{|\operatorname{Perm}(U_S)|^2}{\prod_{j=1}^{m}(s_j!)}, \tag{37}$$

in terms of the permanent of the matrix $U_S \in \mathbb{C}^{n\times n}$ constructed from U by discarding all but the first n columns of U and then, for all $j \in [m]$, taking s_j copies of the j^{th} row of that matrix (deviating from Aaronson and Arkhipov's notation [14]). Here, the permanent for a matrix $X = (x_{j,k}) \in \mathbb{C}^{n\times n}$ is defined similarly to the determinant but without the negative signs as

$$\operatorname{Perm}(X) := \sum_{\tau\in\operatorname{Sym}([n])}\prod_{j=1}^{n} x_{j,\tau(j)}, \tag{38}$$

where $\operatorname{Sym}([n])$ is the symmetric group acting on $[n]$. It is a known fact that calculating the permanent of a matrix to high precision is a problem that is #P-hard [51], while its close cousin, the determinant, is computable in polynomial time. In fact, computing the permanent exactly (or with exponential precision) is also #P-hard *on average* for randomly chosen Gaussian matrices [14, 52]. In Ref. [14] this connection is exploited to show that, up to plausible complexity-theoretic conjectures, approximately sampling from the boson sampling distribution is classically intractable with high probability over the choice of U if m is scaled appropriately with n.

The main part of the hardness proof of Ref. [14] is to prove the classical hardness of sampling from the *post-selected boson sampling distribution* $P^*_{\mathrm{bs},U}$. The post-selected distribution $P^*_{\mathrm{bs},U}$ is obtained from $P_{\mathrm{bs},U}$ by discarding all output sequences S with more than one boson per mode, i.e., all S which are not in the set of *collision-free* sequences

$$\Phi^*_{m,n} := \Big\{S \in \Phi_{m,n} : \forall s \in S : s \in \{0,1\}\Big\}. \tag{39}$$

The hardness of sampling from the full boson sampling distribution follows from the fact that for the relevant scalings of m with n the post-selection can be done efficiently in the sense that on average at least a constant fraction of the outcome sequences is collision-free (Theorem 13.4 in Ref. [14]).

More precisely, the actual result proved in Ref. [14, Theorem 1.3] states that unless certain complexity-theoretic conjectures fail, there exists no classical algorithm that can sample from a distribution Q satisfying $\|Q - P_{\mathrm{bs},U}\|_1 \leq \epsilon$ in time $\mathsf{poly}(n, 1/\epsilon)$. This result requires that $m \in \Omega(n^5\log(n)^2)$, but it is conjectured that m growing slightly faster than $\Omega(n^2)$ is sufficient for hardness. In fact, at the same time, a faster than quadratic scaling is necessary for the proof strategy to work.

The key technical ingredient in the proof strategy underlying these requirements is the following result: if m grows sufficiently fast with n, the measure induced on $U \sim \mu_H$ by the map $g_S = (U \mapsto U_S)$ for collision-free $S \in \Phi^*_{m,n}$, i.e., the measure induced by taking $n\times n$-submatrices of unitaries $U \in U(m)$ chosen with respect to the Haar measure μ_H is close to the complex Gaussian measure $\mu_G(\sigma)$ with mean zero and standard deviation $\sigma = 1/\sqrt{m}$ on $n\times n$-matrices. Given this result, Stockmeyer's algorithm could be applied to the samples obtained from $P^*_{\mathrm{bs},U}$ in order to infer the probabilities $P^*_{\mathrm{bs},U}(S)$ and thus solve a #P-hard problem, as these probabilities can be expressed as the permanent of a Gaussian matrix. Since the closeness of those measures is the essential ingredient, also suitably large scaling of m with n is crucial for the hardness argument.

The formal statement of closeness of measures proved in Ref. [14] implies the following:

Lemma 8 (implied by Ref. [14, Theorem 5.2])**.** *There exists a constant $C > 0$ such that for every $\nu > 5$ and every measurable $f : \mathbb{C}^{n\times n} \to [0,1]$ and every $m \in \Omega(m^\nu)$ it holds that for all $S \in \Phi^*_{m,n}$*

$$\mathbb{E}_{U\sim\mu_H}\, f(U_S) \leq (1+C)\; \mathbb{E}_{X\sim\mu_{G(1/\sqrt{m})}}\, f(X). \qquad (40)$$

At the same time, it is known from Ref. [53] (see also Ref. [14, Section 5.1 and 6.2]) that if $m \leq c\,n^\nu$ with $\nu \leq 2$ and $c \in O(1)$ the two measures $\mu_H \circ g_S^{-1}$ and $\mu_{G(1/\sqrt{m})}$ are no longer close for large n. One may hope [14] that there exists a constant $c > 0$ such that Theorem 5.2 in Ref. [14] and hence their hardness result as well as our Lemma 8 hold for any $m \geq c\,n^\nu$ with $\nu > 2$. What we show is that *even under this optimistic assumption* efficient certification from classical samples is impossible, if the post-selection probability is large enough. This rules out many further cases for which one can hope to prove a hardness result by the same method.

Theorem 9 (Lower bounds on certifying boson sampling (full version))**.** *Let $\nu > 2$, $0 < \epsilon < 1/2$, $n \in \mathbb{Z}_+$ sufficiently large and $m \in \Theta(n^\nu)$. Assume there exists a constant $C > 0$ such that the assertion (40) of Lemma 8 holds. Then:*

a. With probability at least $1-\delta-2n^2/(m\zeta)$ over the choice of Haar-random unitaries $U \sim \mu_H$ there exists no ϵ-certification test for boson sampling with n photons in m modes, from $s < s_{\min}$ many samples, where

$$s_{\min} \in \Omega\left(n^{cn(\nu-1)/4}\delta^{1/4}\left(1-\zeta-2\epsilon\right)^{3/2}/\epsilon^2\right), \qquad (41)$$

and $c > 0$ is the implicit constant in (48).

b. For $\nu > 3$, with probability at least $1-\exp(-\Omega(n^{\nu-2-1/n}))$ over the Haar-random choice of $U \sim \mu_H$, there exists no ϵ-certification test for boson sampling with n photons in m modes, where

$$s_{\min} \in \Omega\left(2^n/\epsilon^2\right). \qquad (42)$$

We remark that our results for the boson sampling distribution leave open the possiblity of sample-efficient ϵ-certification for those instances of boson sampling with $2 < \nu \leq 3$ in the regime in which the probability weight of the collision-free subspace is very small. For instance, this is the case whenever $1/\mathsf{poly}(n) \leq P_{\mathrm{bs},U}(\Phi^*_{m,n}) \leq 2\epsilon$. This is because the bound (41) becomes trivial for $1-\zeta \leq 2\epsilon$.

However, our result fully covers the regime in which boson sampling is provably hard as shown in Ref. [14].

Proof of Theorem 9a. The proof proceeds along the same lines as the proofs of Theorem 7 and is based on direct applications of Lemma 4 to the collision-free subspace and the min-entropy bound (22) from Lemma 5 to the post-selected boson sampling distribution $P^*_{\mathrm{bs},U}$ with post-selection onto the collision-free subspace $\Phi^*_{m,n} \subset \Phi_{m,n}$.

To apply Lemmas 4 and 5 simultaneously we need to account both for the probability weight of the collision-free subspace and large probabilities, however. To account for the probability weight of the collision-free subspace we use a simple application of Markov's inequality to [14, Theorem 13.4] (restated as Lemma 10 in Section B),

$$\Pr_{U\sim\mu_H}\left[P_{\mathrm{bs},U}[\Phi_{m,n}\setminus\Phi^*_{m,n}] > \zeta\right] < \frac{2n^2}{\zeta m}. \qquad (43)$$

This shows that the total probability weight of the collision-free subspace is at least $1-\zeta$ with probability at least $1-2n^2/(m\zeta)$. We then apply a union bound argument to obtain

$$\begin{aligned}\Pr_{U\sim\mu_H}\Big[&\{P_{\mathrm{bs},U}\text{ does not satisfy (22)}\}\\ &\cup\left\{P_{\mathrm{bs},U}(\Phi_{m,n}\setminus\Phi^*_{m,n}) > \zeta\right\}\Big] \leq \delta + \frac{2n^2}{\zeta m}.\end{aligned} \qquad (44)$$

In the next step, we use that the distribution of post-selected boson sampling is given by $P^*_{\mathrm{bs},U} = (P_{\mathrm{bs},U})_{\restriction\Phi^*_{m,n}}/P_{\mathrm{bs},U}(\Phi^*_{m,n})$. Consequently, with probability at least $1-\delta-2n^2/(\zeta m)$ the boson sampling distribution $P_{\mathrm{bs},U}$ restricted to the collision-free subspace has both of the desired properties – a large min-entropy and a probability weight of at least $1-\zeta$ of the collision-free subspace.

Let us now compute the min-entropy for the collision-free subspace. For all samples $S \in \Phi^*_{m,n}$, Ref. [14, Lemma 8.8] implies that there exists $C > 0$ such that for $m \in \Theta(n^\nu)$ with any $\nu > 2$ for which the assertion of Lemma 8 holds, the following second moment bound also holds[6]:

$$\mathbb{E}_{U_S\sim\mu_H}[|\operatorname{Perm}(U_S)|^4] \leq (1{+}C)(n!)^2\,(n{+}1)\,m^{-2n}. \quad (45)$$

To obtain a lower bound on the min-entropy of the distribution $P^*_{\mathrm{bs},U}$ on the collision-free subspace we use that

$$H_\infty(P^*_{\mathrm{bs},U}) = \log P_{\mathrm{bs},U}(\Phi^*_{m,n}) + H_\infty((P_{\mathrm{bs},U})_{\restriction\Phi^*_{m,n}}). \qquad (46)$$

Applying Lemma 5 together with the second moment bound (45), the union bound (44), the bound

$$|\Phi^*_{m,n}| = \binom{m}{n} = \frac{m(m-1)\cdots(m-n+1)}{n!} \leq \frac{m^n}{n!}, \qquad (47)$$

on the size of the collision-free subspace and Stirling's formula yields

$$\begin{aligned}2H_\infty(P^*_{\mathrm{bs},U}) &\geq 2\log(1-\zeta)+\log\delta\\ &\quad-\log\left(\frac{m^n}{n!}(1+C)(n!)^2(n+1)m^{-2n}\right)\\ &\in\Omega\left((\nu-1)n\log n\right)-\log\frac{1}{\delta}-2\log\frac{1}{1-\zeta},\end{aligned} \qquad (48)$$

[6] The version of Ref. [14, Lemma 8.8] can be obtained from Eq. (45) from Lemma 8, normalizing the Gaussian measure μ_G, and noting that $\mathbb{E}_{X\sim\mu_G(1)}[|\operatorname{Perm}(X)|^2] = n!$.

which holds with probability $1-\delta-2n^2/(m\zeta)$ over the choice of $U \sim \mu_H$.

We note that $2^{-H_\infty(P^*_{\mathrm{bs},U})} \in o(1)$; hence this term can be neglected when applying Eq. (8) in Lemma 3. Applying Lemmas 3 and 4, and the min-entropy bound (48) we obtain that the sample complexity for ϵ-certifying boson-sampling scales as

$$s_{\min} \in \Omega\left(n^{cn(\nu-1)/4}\delta^{1/4}\left(1-\zeta-2\epsilon\right)^{3/2}/\epsilon^2\right) \qquad (49)$$

with probability at least $1-\delta-2n^2/(\zeta m)$, where c is the implicit constant in (48). This completes the proof of Theorem 9a.

□

Note that the bound (45) is essential for the hardness argument of Aaronson and Arkhipov [14]. Therefore a central ingredient to the hardness argument of Aaronson and Arkhipov [14] also prohibits sample-efficient certification of boson sampling.

It is important to stress that the boson sampling hardness proof [14] covers only those instances U_S of boson sampling for which one can efficiently post-select on the collision-free outcomes. This is the case for those $U \sim \mu_H$ for which the probability weight of $\Phi^*_{m,n}$ is not smaller than polynomially small in n, i.e., $P_{\mathrm{bs},U}(\Phi^*_{m,n}) \in \Omega(1/\mathsf{poly}(n))$. Our proof method for Theorem 9a thus permits sample-efficient certification for a small fraction of the instances, in particular, those instances of $U \sim \mu_H$ for which $2\epsilon \geq P_{\mathrm{bs},U}(\Phi^*_{m,n}) > 1/\mathsf{poly}(n)$.

In part b of the theorem we can close this gap by extending the bound (48) on the min-entropy of the post-selected distribution $P^*_{\mathrm{bs},U}$ to the full output distribution $P_{\mathrm{bs},U}$, however, at the cost of restricting to $\nu > 3$. This removes the need to use Lemma 4 and hence the dependence on the probability weight of the collision-free subspace. In the remaining case with $2 < \nu \leq 3$ hardness results have not been obtained, but it is conceivable that a hardness argument can be made.

Proof of Theorem 9b. Gogolin et al. [10] have proven the following strong lower bound on the min-entropy of the boson sampling distribution (see Theorem 11 in Section C for a restatement)

$$\Pr_{U\sim\mu_H}\left[H_\infty(P_{\mathrm{bs},U}) < 2\,n\right] \in \exp\left(-\Omega(n^{\nu-2-1/n})\right), \qquad (50)$$

which holds whenever the condition of the theorem are fulfilled and in addition $\nu > 3$. In the proof, the probability measure induced on the matrices U_S is related to a certain Gaussian measure $\mu_{G_S(\sigma)}$. Then, the min-entropy bound is proven using a trivial upper bound to the permanent as well as measure concentration for $\mu_{G_S(\sigma)}$. A simple application of Theorem 2 and Lemma 3 concludes the proof. □

VI. CONCLUSION

We have shown that probability distributions with a high min-entropy cannot be certified from polynomially many samples, even when granting the certifier unlimited computational power and a full description of the target distribution. Our result applies to the problem of certifying quantum sampling problems as proposed to demonstrate a quantum speedup in a non-interactive device-independent fashion. We discuss the ironic situation that the very property that crucially contributes to the proof of approximate sampling hardness via Stockmeyer's algorithm and the Paley-Zygmund inequality — the second moments of the sampled distribution — forbids sample-efficient classical verification. Our results highlight the importance of devising more elaborate certification schemes that allow for interaction between certifier and prover, invoke further complexity-theoretic assumptions or such on the sampling device, and/or grant the certifier some small amount of quantum capacities.

ACKNOWLEDGEMENTS

We are grateful to Adam Bouland who pointed us to the literature on property testing and thus provided the missing clue for finishing this project. We would like to thank Ashley Montanaro, Bill Fefferman, Tomoyuki Morimae, Martin Schwarz, and Juan Bermejo-Vega for fruitful discussions and Aram Harrow and Anand Natarajan for sharing an early version of their related work. Finally, we would like to thank two anonymous referees for their thorough proof-checking and interesting questions which helped improve the manuscript.

D. H. and J. E. acknowledge support from the ERC (TAQ), the Templeton Foundation, the DFG (EI 519/14-1, EI 519/9-1, EI 519/7-1, CRC 183), and the European Union's Horizon 2020 research and innovation programme under grant agreement No. 817482 (PASQUANS). C. G. acknowledges support by the European Union's Marie Skłodowska-Curie Individual Fellowships (IF-EF) programme under GA: 700140 as well as financial support from ARO under contract W911NF-14-1-0098 (Quantum Characterization, Verification, and Validation), the Spanish Ministry MINECO (National Plan 15 Grant: FISICATEAMO No. FIS2016-79508-P, SEVERO OCHOA No. SEV-2015-0522), Fundació Cellex, Generalitat de Catalunya (Grants No. SGR 874 and No. 875, CERCA Programme, AGAUR Grant No. 2017 SGR 1341 and CERCA/Program), ERC (CoG QITBOX and AdG OSYRIS), EU FETPRO QUIC, EU STREP program EQuaM (FP7/2007–2017, Grant No. 323714), and the National Science Centre, Poland-Symfonia Grant No. 2016/20/W/ST4/00314.

APPENDIX

In the following, we (re)state some facts and earlier results in order to make this work self-contained.

Appendix A: Proofs for bounding the min-entropy

Here, we provide some details and proofs to statements made in Section IV A. First, we show the equivalence of the Rényi entropies (25) proceeding analogously to Ref. [46]: we simply use that for $\alpha > 1$ and $p_0 = \|P\|_\infty$ we have $p_0^\alpha \leq \sum_i p_i^\alpha$. Hence,

$$\frac{\alpha}{\alpha-1}\log(p_0) \leq \frac{1}{\alpha-1}\log\sum_{i=0}^{|\mathcal{E}_n|-1} p_i^\alpha \tag{A1}$$

$$\Leftrightarrow -\frac{\alpha}{\alpha-1}H_\infty(P) \leq -H_\alpha(P) \tag{A2}$$

$$\Leftrightarrow H_\infty(P) \geq \frac{\alpha-1}{\alpha}H_\alpha(P)\,. \tag{A3}$$

We also provide an alternative proof of Lemma 5 based on the proof of Ref. [10, Theorem 13].

Alternative proof of Lemma 5. We begin the proof by noting that

$$\Pr_{U\sim\mu_n}\left[H_\infty(P_U) \leq \log\frac{1}{\delta}\right] = \Pr_{U\sim\mu_n}\left[\exists S \in \mathcal{E}_n : P_U(S) \geq \delta\right]. \tag{A4}$$

Using the union bound (also known as Boole's inequality) we obtain that for every $\delta > 0$

$$\Pr_{U\sim\mu_n}\left[\exists S \in \mathcal{E}_n : P_U(S) \geq \delta\right] \leq \sum_{S\in\mathcal{E}_n}\Pr_{U\sim\mu_n}\left[P_U(S) \geq \delta\right]. \tag{A5}$$

Next, using Markov's inequality we can bound

$$\Pr_{U\sim\mu_n}\left[P_U(S) \geq \delta\right] \leq \frac{1}{\delta^2}\mathbb{E}_{U\sim\mu_n}\left[P_U(S)^2\right], \tag{A6}$$

which concludes the proof. □

Appendix B: Probability weight of the collision-free subspace

We recapitulate a bound of Aaronson and Arkhipov [14] on the probability weight of the collision-free subspace.

Lemma 10 ([14, Theorem 13.4])**.** *Let μ_H be the Haar measure on $U(m)$ and $m \geq n$. Then*

$$\mathbb{E}_{U\sim\mu_H}\left[P_{\mathrm{bs},U}(\Phi_{m,n}\setminus\Phi^*_{m,n})\right] \leq \frac{2n^2}{m}\,. \tag{B1}$$

Appendix C: The min-entropy bound for boson sampling

Here, we provide a slightly improved proof of the following min-entropy bound for boson sampling from [10, Theorem 12].

Theorem 11 (Min-entropy bound for boson sampling [10, Theorem 12])**.** *Let $\nu > 3$ and assume that the assertion (40) of Lemma 8 holds. Then, the boson sampling output distribution $P_{\mathrm{bs},U}$ satisfies for n bosons in $m \in \Theta(n^\nu)$ modes*

$$\Pr_{U\sim\mu_H}\left[H_\infty(P_{\mathrm{bs},U}) < 2\,n\right] \in \exp\left(-\Omega(n^{\nu-2-1/n})\right). \tag{C1}$$

The proof crucially uses the closeness of the Gaussian measure to the post-selected Haar measure as expressed by Lemma 8. Lemma 8, however, is not quite strong enough for proving Theorem 9, as we must be able to control all of $\Phi_{m,n}$ and not only the collision-free subspace $\Phi^*_{m,n}$. Fortunately, the above lemma extends naturally to all $S \in \Phi_{m,n}$ for the same scaling of m with n for which a version of Lemma 8 holds.

To state this extension we need some notation first: For every sequence S, let $\tilde{S}$ be the sequence obtained from S by removing all the zeros, i.e,

$$\tilde{S} = (\tilde{s}_1, \dots, \tilde{s}_{|\tilde{S}|}) \coloneqq (s \in S : s > 0). \tag{C2}$$

Further, let $\mu_{G_S(\sigma)}$ be the probability measure on $\mathbb{C}^{n\times n}$ obtained by drawing the real and imaginary part of every entry of a $|\tilde{S}| \times n$ matrix independently from a Gaussian distribution with mean zero and standard deviation σ and then for all $j \in [|\tilde{S}|]$ taking $\tilde{s}_j$ copies of the j^{th} row of this matrix. We can prove the following multiplicative error bound on the closeness of this measure and the Haar measure μ_H for all $S \in \Phi_{m,n}$:

Lemma 12 (Multiplicative error bound). *Let $f : \mathbb{C}^{n\times n} \to [0,1]$ be measurable, then for any m, n such that*

$$\forall S \in \Phi^*_{m,n} : \quad \mathbb{E}_{U\sim\mu_H} f(U_S) \leq (1+C)\; \mathbb{E}_{X\sim\mu_{G(1/\sqrt{m})}} f(X), \tag{C3}$$

is true for some constant $C > 0$, it holds that

$$\forall S \in \Phi_{m,n} : \quad \mathbb{E}_{U\sim\mu_H} f(U_S) \leq (1+C)\; \mathbb{E}_{X\sim\mu_{G_S(1/\sqrt{m})}} f(X). \tag{C4}$$

Proof. Let $S \in \Phi_{m,n}$, define $\tilde{S}$ as in Eq. (C2) and $m' \coloneqq |\tilde{S}|$. Define v to be the sequence containing $\tilde{s}_j$ times the integer j for every $j \in [m']$ in increasing order and w the sequence containing the positions of each of the first of the repeated rows in U_S, i.e.,

$$v \coloneqq (\underbrace{1, \dots, 1}_{\tilde{s}_1}, \underbrace{2, \dots, 2}_{\tilde{s}_2}, \dots, \underbrace{m', \dots, m'}_{\tilde{s}_{m'}}) \in (\mathbb{Z}^+)^n, \tag{C5}$$

$$w \coloneqq (1, 1+\tilde{s}_1, 1+\tilde{s}_1+\tilde{s}_2, \dots, 1+\sum_{j=1}^{m'-1} \tilde{s}_j) \in (\mathbb{Z}^+)^{m'}. \tag{C6}$$

The sequence v defines a linear embedding $\eta : \mathbb{C}^{m'\times n} \to \mathbb{C}^{n\times n}$ component wise by

$$\eta(Y)_{i,j} \coloneqq Y_{v_i,j} \quad \forall i,j \in [n], \tag{C7}$$

i.e., $\eta(Y)$ has s_j copies of the j-th row of Y. The sequence w, in turn, defines a linear projection $\pi : \mathbb{C}^{n\times n} \to \mathbb{C}^{m'\times n}$ by

$$\pi(X)_{i,j} \coloneqq X_{w_i,j} \quad \forall i \in [m'],\ j \in [n], \tag{C8}$$

in particular, $\pi(U_S)$ contains only the first out of each series of the repeated rows in U_S. Note that $\eta \circ \pi : \mathbb{C}^{n\times n} \to \mathbb{C}^{n\times n}$ is a projection onto the subspace of matrices that have the same repetition structure as U_S. Let

$$f_S \coloneqq f \circ \eta \circ \pi, \tag{C9}$$

then $f_S(U_S) = f(U_S)$ only depends on the first of the repeated rows in U_S and is independent of all the other rows. Since the Haar measure is permutation-invariant we have

$$\mathbb{E}_{U\sim\mu_H} f_S(U_S) = \mathbb{E}_{U\sim\mu_H} f_S(U_{1_n}). \tag{C10}$$

Hence, using Lemma 8 in the second step, we obtain

$$\begin{aligned}
\mathbb{E}_{U\sim\mu_H} f(U_S) &= \mathbb{E}_{U\sim\mu_H} f_S(U_{1_n}) && \text{(C11)}\\
&\leq (1+C)\, \mathbb{E}_{X\sim\mu_{G(1/\sqrt{m})}} f_S(X) && \text{(C12)}\\
&= (1+C)\, \mathbb{E}_{X\sim\mu_{G_S(1/\sqrt{m})}} f(X), && \text{(C13)}
\end{aligned}$$

which finishes the proof. □

In addition to the multiplicative error bound we need the following concentration result for the Gaussian measure $\mu_{G_S(\sigma)}$, which implies that even the largest entry of a matrix drawn from $\mu_{G_S(\sigma)}$ is unlikely to be much larger than σ.

Lemma 13 (Concentration of the Gaussian measure $\mu_{G_S(\sigma)}$)**.** *For all $n, m \in \mathbb{Z}^+$, all $S \in \Phi_{m,n}$ and all $\xi > 0$ it holds that*

$$\Pr_{X\sim\mu_{G_S(\sigma)}}\left[\max_{j,k\in[n]} |x_{j,k}| \geq \xi\right] \leq 1 - \left(1 - \operatorname{Erfc}\left(\frac{\xi}{\sqrt{2}\,\sigma}\right)\right)^{n^2}, \tag{C14}$$

where

$$\operatorname{Erfc}\left(\frac{\xi}{\sqrt{2}\,\sigma}\right) := 2\int_\xi^\infty \frac{\mathrm{e}^{-\frac{x^2}{2\,\sigma^2}}}{\sqrt{2\,\pi\,\sigma^2}}\,\mathrm{d}x \tag{C15}$$

is the complementary error function.

Proof. For Gaussian random variables we have

$$\forall \xi > 0,\ j,k \in [n] : \quad \Pr_{X\sim\mu_{G(\sigma)}}[|x_{j,k}| \geq \xi] = \operatorname{Erfc}\left(\frac{\xi}{\sqrt{2}\,\sigma}\right). \tag{C16}$$

This implies that

$$\forall \xi > 0 : \quad \Pr_{X\sim\mu_{G(\sigma)}}[\forall j,k \in [n] : |x_{j,k}| \leq \xi] = \left(1 - \operatorname{Erfc}\left(\frac{\xi}{\sqrt{2}\,\sigma}\right)\right)^{n^2}. \tag{C17}$$

At the same time, for all $S \in \Phi_{m,n}$ and $\xi > 0$ it holds that

$$\Pr_{X\sim\mu_{G_S(\sigma)}}[\forall j,k \in [n] : |x_{j,k}| \leq \xi] \geq \Pr_{X\sim\mu_{G(\sigma)}}[\forall j,k \in [n] : |x_{j,k}| \leq \xi], \tag{C18}$$

because the repetition of entries in $X \sim \mu_{G_S(\sigma)}$ only increases the chance of not having an exceptionally large entry. □

As a last ingredient we need to bound the size

$$|\Phi_{m,n}| = \binom{m+n-1}{n} \tag{C19}$$

of the sample space $\Phi_{m,n}$ of boson sampling (recall Eq. (35)). It grows faster than than exponentially with n, but if for some $\nu \geq 1$ and $c \geq 0$ it holds that $m \leq c\,n^\nu$, then

$$|\Phi_{m,n}| \leq \frac{(m+n-1)^n}{n!} \leq \left(\frac{(m+n-1)\,\mathrm{e}}{n}\right)^n \tag{C20}$$

$$\leq \mathrm{e}^n\,(c\,n^{\nu-1} + 1 - 1/n)^n \leq (2\,(c+1)\,\mathrm{e})^n\,n^{(\nu-1)\,n}. \tag{C21}$$

We now have all the ingredients rederive the desired min-entropy bound in Theorem 11.

Proof of Theorem 11. Using the union bound (also known as Boole's inequality) in the first step we obtain that for every $\epsilon > 0$

$$\Pr_{U\sim\mu_H}[\exists S \in \Phi_{m,n} : P_{\mathrm{bs},U}(S) \geq \epsilon] \tag{C22}$$

$$\leq \sum_{S\in\Phi_{m,n}} \Pr_{U\sim\mu_H}[P_{\mathrm{bs},U}(S) \geq \epsilon] \tag{C23}$$

$$\leq |\Phi_{m,n}| \max_{S\in\Phi_{m,n}} \Pr_{U\sim\mu_H}[P_{\mathrm{bs},U}(S) \geq \epsilon] \tag{C24}$$

$$= |\Phi_{m,n}| \max_{S\in\Phi_{m,n}} \Pr_{U\sim\mu_H}\left[\frac{|\operatorname{Perm}(U_S)|^2}{\prod_{j=1}^m (s_j!)} \geq \epsilon\right]. \tag{C25}$$

We now apply Lemma 12 to the indicator function

$$f(U_S) = \begin{cases} 1 & \text{if } \frac{|\operatorname{Perm}(U_S)|^2}{\prod_{j=1}^m (s_j!)} \geq \epsilon \\ 0 & \text{otherwise} \end{cases}, \tag{C26}$$

and the S for which the maximum in Eq. (C25) is attained, to obtain

$$\Pr_{U\sim\mu_H}\left[\exists S\in\Phi_{m,n} : P_{\mathrm{bs},U}(S)\geq\epsilon\right] \leq (1+C)\,|\Phi_{m,n}| \max_{S\in\Phi_{m,n}} \Pr_{X\sim\mu_{G_S(1/\sqrt{m})}}\left[\frac{|\operatorname{Perm}(X)|^2}{\prod_{j=1}^m (s_j!)}\geq\epsilon\right]. \tag{C27}$$

The definition of the permanent (recall Eq. (38)) implies that

$$\frac{|\operatorname{Perm}(X)|^2}{\prod_{j=1}^m (s_j!)} \leq |\operatorname{Perm}(X)|^2 \leq (n!)^2 \left(\max_{j,k\in[n]}|x_{j,k}|\right)^{2n}. \tag{C28}$$

Hence, for every $S\in\Phi_{m,n}$ and every $\epsilon>0$

$$\Pr_{X\sim\mu_{G_S(1/\sqrt{m})}}\left[\frac{|\operatorname{Perm}(X)|^2}{\prod_{j=1}^m (s_j!)}\geq\epsilon\right] \leq \Pr_{X\sim\mu_{G_S(1/\sqrt{m})}}\left[\max_{j,k\in[n]}|x_{j,k}| \geq \left(\frac{\sqrt{\epsilon}}{n!}\right)^{1/n}\right]. \tag{C29}$$

Plugging this into Eq. (C27), using Lemma 13 with $\xi = (\sqrt{\epsilon}/n!)^{1/n}$ and the bound on $|\Phi_{m,n}|$ from Eq. (C21) we arrive at

$$\begin{aligned}&\Pr_{U\sim\mu_H}\left[\exists S\in\Phi_{m,n} : P_{\mathrm{bs},U}(S)\geq\epsilon\right]\\ &\leq (1+C)\,(2\,(c+1)\,\mathrm{e})^n\, n^{(\nu-1)n}\left(1-\left(1-\operatorname{Erfc}\sqrt{\frac{c}{2}\frac{\epsilon^{1/n}\,n^{\nu}}{(n!)^{2/n}}}\right)^{n^2}\right).\end{aligned} \tag{C30}$$

Bounding the complementary error function by [54]

$$\operatorname{Erfc}(x) \leq \mathrm{e}^{-x^2}, \tag{C31}$$

we obtain

$$1-(1-\operatorname{Erfc}(x))^{n^2} \leq 1-\left(1-\mathrm{e}^{-x^2}\right)^{n^2} = 1-\sum_{k=0}^{n^2}\binom{n^2}{k}(-\mathrm{e}^{-x^2})^k \tag{C32}$$

$$= \sum_{k=1}^{n^2}\binom{n^2}{k}\mathrm{e}^{-x^2 k}\,(-1)^{k-1} \leq \sum_{k=1}^{n^2}(n^2\mathrm{e}/k)^k\,\mathrm{e}^{-x^2 k} \tag{C33}$$

$$\leq \sum_{k=1}^{n^2}(n^2\,\mathrm{e}^{-x^2+1})^k. \tag{C34}$$

If x is large enough such that

$$n^2\,\mathrm{e}^{-x^2+1} \leq \frac{1}{2} < 1, \tag{C35}$$

the geometric series in Eq. (C34) converges and we get the simple bound

$$1-(1-\operatorname{Erfc}(x))^{n^2} \leq \sum_{k=1}^{n^2}(n^2\mathrm{e}^{-x^2+1})^k \leq \frac{n^2\mathrm{e}^{-x^2+1}}{1-n^2\mathrm{e}^{-x^2+1}} \leq 2\,n^2\,\mathrm{e}^{-x^2+1}. \tag{C36}$$

To satisfy Eq. (C35) for large n, it is sufficient that x grows slightly faster than $\sqrt{\log(n^2)}$ and we hence need to demand a growth slightly faster than $\log(n^2)$ from the argument of the square root in the error function in Eq. (C30). Because of the bound $n! \leq \mathrm{e}^{1-n}\,n^{n+1/2}$ (a variant of Stirling's approximation) we have for the argument of that square root in Eq. (C30)

$$\frac{c}{2}\frac{\epsilon^{1/n}n^{\nu}}{(n!)^{2/n}} \geq \frac{c}{2}\frac{\epsilon^{1/n}n^{\nu}}{\mathrm{e}^{2/n-2}n^{2+1/n}} = \frac{c}{2}\frac{\epsilon^{1/n}}{\mathrm{e}^{2/n-2}}n^{\nu-2-1/n}, \tag{C37}$$

Demanding $\nu>2$ is hence all we need to be able to use the bound (C36) for large n. With the convenient choice $\epsilon=2^{-2n}$ it hence follows that for all $\nu>2$

$$\begin{aligned}&\Pr_{U\sim\mu_H}\left[\exists S\in\Phi_{m,n} : P_{\mathrm{bs},U}(S)\geq 2^{-2n}\right]\\ &\in O\left(n^2\,(2\,(c+1)\,\mathrm{e})^n\,n^{(\nu-1)n}\exp(-c\,\mathrm{e}^{-2/n+2}\,n^{\nu-2-1/n}/8)\right).\end{aligned} \tag{C38}$$

The argument of the $O(\cdot)$ is dominated by the product $n^{(\nu-1)n}\exp(-c\,\mathrm{e}^{-2/n+2}\,n^{\nu-2-1/n}/8)$, which decays for large increasing n only for $\nu > 3$. More precisely, there are constants $n_0 \in \mathbb{N}$ and $C_1, C_2, C_3 > 0$ such that for $n \geq n_0$

$$n^2\,(2\,(c+1)\,\mathrm{e})^n\,n^{(\nu-1)n}\exp(-c\,\mathrm{e}^{-2/n+2}\,n^{\nu-2-1/n}/8) \tag{C39}$$

$$= \exp\Big(2\ln(n) + n\ln(2\,(c+1)\,\mathrm{e}) + n(\nu-1)\ln n - c\,\mathrm{e}^{-2/n+2}\,n^{\nu-2-1/n}/8\Big) \tag{C40}$$

$$\leq \exp\Big(C_1 n(\nu-1)\ln n - c\,\mathrm{e}^{-2/n+2}\,n^{\nu-2-1/n}/8)\Big) \tag{C41}$$

$$\leq \exp\Big(C_1 n(\nu-1)\ln n - C_2 n^{\nu-2-1/n}\Big) \tag{C42}$$

$$\overset{\nu>3}{\leq} \exp\Big(-C_3 n^{\nu-2-1/n}\Big) \in \exp\Big(-\Omega(n^{\nu-2-1/n})\Big)\,. \tag{C43}$$

where the last inequality holds only for $\nu > 3$ since the logarithm grows slower than any power law with positive exponent. This completes the proof.

□

[1] J. Preskill, *Quantum computing and the entanglement frontier*, Bull. Am. Phys. Soc. **58** (2013), arXiv:1203.5813.

[2] 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 **339**, 798 (2013).

[3] M. Tillmann, B. Dakić, R. Heilmann, S. Nolte, A. Szameit, and P. Walther, *Experimental boson sampling*, Nature Photonics **7**, 540 (2013).

[4] 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 **339**, 794 (2013).

[5] A. Crespi, R. Osellame, R. Ramponi, D. J. Brod, E. F. Galvão, N. Spagnolo, C. Vitelli, E. Maiorino, P. Mataloni, and F. Sciarrino, *Integrated multimode interferometers with arbitrary designs for photonic boson sampling*, Nature Photonics **7**, 545 (2013).

[6] J. Carolan, J. D. A. Meinecke, P. J. Shadbolt, N. J. Russell, N. Ismail, K. Wörhoff, T. Rudolph, M. G. Thompson, J. L. O'Brien, J. C. F. Matthews, and A. Laing, *On the experimental verification of quantum complexity in linear optics*, Nature Photonics **8**, 621 (2014), arXiv:1311.2913.

[7] N. Spagnolo, C. Vitelli, M. Bentivegna, D. J. Brod, A. Crespi, F. Flamini, S. Giacomini, G. Milani, R. Ramponi, P. Mataloni, R. Osellame, E. F. Galvão, and F. Sciarrino, *Experimental validation of photonic boson sampling*, Nature Photonics **8**, 615 (2014), arXiv:1311.1622.

[8] P. Clifford and R. Clifford, The Classical Complexity of Boson Sampling, in *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA '18*, SODA '18 (Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2018) pp. 146–155, arXiv:1706.01260.

[9] A. Neville, C. Sparrow, R. Clifford, E. Johnston, P. M. Birchall, A. Montanaro, and A. Laing, *Classical boson sampling algorithms with superior performance to near-term experiments*, Nature Physics **13**, 1153 (2017), arXiv:1705.00686.

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

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

[12] M. Oszmaniec and D. J. Brod, *Classical simulation of photonic linear optics with lost particles*, New J. Phys. **20**, 092002 (2018), arXiv:1801.06166.

[13] J. Renema, V. Shchesnovich, and R. Garcia-Patron, *Quantum-to-classical transition in many-body bosonic interference*, (2018), arXiv:1809.01953.

[14] S. Aaronson and A. Arkhipov, *The computational complexity of linear optics*, Th. Comp. **9**, 143 (2013), arXiv:1011.3245.

[15] S. Boixo, S. V. Isakov, V. N. Smelyanskiy, R. Babbush, N. Ding, Z. Jiang, M. J. Bremner, J. M. Martinis, and H. Neven, *Characterizing quantum supremacy in near-term devices*, Nature Physics **14**, 595 (2018), arXiv:1608.00263.

[16] M. J. Bremner, A. Montanaro, and D. J. Shepherd, *Average-case complexity versus approximate simulation of commuting quantum computations*, Phys. Rev. Lett. **117**, 080501 (2016), arXiv:1504.07999.

[17] A. Bouland, B. Fefferman, C. Nirkhe, and U. Vazirani, *On the complexity and verification of quantum random circuit sampling*, Nature Physics **15**, 159 (2019), arXiv:1803.04402.

[18] X. Gao, S.-T. Wang, and L.-M. Duan, *Quantum supremacy for simulating a translation-invariant Ising spin model*, Phys. Rev. Lett. **118**, 040502 (2017), arXiv:1607.04947.

[19] J. Bermejo-Vega, D. Hangleiter, M. Schwarz, R. Raussendorf, and J. Eisert, *Architectures for quantum simulation showing a quantum speedup*, Phys. Rev. X **8**, 021010 (2018), arXiv:1703.00466.

[20] D. Hangleiter, J. Bermejo-Vega, M. Schwarz, and J. Eisert, *Anticoncentration theorems for schemes showing a quantum speedup*, Quantum **2**, 65 (2018), arXiv:1706.03786.

[21] A. Bouland, J. F. Fitzsimons, and D. E. Koh, Complexity classification of conjugated Clifford circuits, in *33rd Computational Complexity Conference (CCC 2018)*, Leibniz International Proceedings in Informatics (LIPIcs), Vol. 102, edited by R. A. Servedio (Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, Dagstuhl, Germany, 2018) pp. 21:1–21:25, arXiv:1709.01805.

[22] M. Yoganathan, R. Jozsa, and S. Strelchuk, *Quantum advantage of unitary Clifford circuits with magic state inputs*, arXiv:1806.03200.

[23] L. Stockmeyer, *On approximation algorithms for $\#P$*, SIAM J. Comput. **14**, 849 (1985).

[24] D. Shepherd and M. J. Bremner, *Temporally unstructured quantum computation*, Proc. Roy. Soc. A **465**, 1413 (2009), arXiv:0809.0847.

[25] S. Aaronson and A. Arkhipov, *BosonSampling is far from uniform*, arXiv:1309.7460.

[26] L. Aolita, C. Gogolin, M. Kliesch, and J. Eisert, *Reliable quantum certification of photonic state preparations*, Nature Communications **6**, 8498 (2015), arXiv:1407.4817.

[27] D. Hangleiter, M. Kliesch, M. Schwarz, and J. Eisert, *Direct certification of a class of quantum simulations*, Quantum Sci. Technol. **2**, 015004 (2017), arXiv:1602.00703.

[28] Y. Nakata, M. Koashi, and M. Murao, *Generating a state t-design by diagonal quantum circuits*, New J. Phys. **16**, 053043 (2014), arXiv:1311.1128.

[29] Y. Nakata, C. Hirche, C. Morgan, and A. Winter, *Unitary 2-designs from random X- and Z-diagonal unitaries*, J. Math. Phys. **58**, 052203 (2017), arXiv:1502.07514.

[30] L. Trevisan, M. Tulsiani, and S. Vadhan, Regularity, boosting, and efficiently simulating every high-entropy distribution, in *2009 24th Annual IEEE Conference on Computational Complexity* (2009) pp. 126–136.

[31] O. Goldreich, *Introduction to Property Testing* (Cambridge University Press, Cambridge, 2017).

[32] G. Valiant and P. Valiant, *A CLT and tight lower bounds for estimating entropy*, Tech. Rep. 10-179 (2010) eCCC.

[33] M. Walschaers, J. Kuipers, J.-D. Urbina, K. Mayer, M. C. Tichy, K. Richter, and A. Buchleitner, *Statistical benchmark for BosonSampling*, New J. Phys. **18**, 032001 (2016).

[34] J. Miller, S. Sanders, and A. Miyake, *Quantum supremacy in constant-time measurement-based computation: A unified architecture for sampling and verification*, Phys. Rev. A **96**, 062320 (2017), arXiv:1703.11002.

[35] M. Schwarz and M. Van den Nest, *Simulating quantum circuits with sparse output distributions*, arXiv:1310.6749.

[36] G. Valiant and P. Valiant, *An automatic inequality prover and instance optimal identity testing*, SIAM J. Comput. **46**, 429 (2017), eCCC, TR13-111.

[37] S. Aaronson and L. Chen, Complexity-Theoretic Foundations of Quantum Supremacy Experiments, in *32nd Computational Complexity Conference (CCC 2017)*, Leibniz International Proceedings in Informatics (LIPIcs), Vol. 79, edited by R. O'Donnell (Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, Dagstuhl, Germany, 2017) pp. 22:1–22:67, arXiv:1612.05903.

[38] S.-T. Wang and L.-M. Duan, *Certification of Boson Sampling Devices with Coarse-Grained Measurements*, (2016), arXiv:1601.02627.

[39] N. Wiebe, C. Granade, C. Ferrie, and D. G. Cory, *Hamiltonian learning and certification using quantum resources*, Phys. Rev. Lett. **112**, 190501 (2014), arXiv:1309.0876.

[40] D. Mills, A. Pappa, T. Kapourniotis, and E. Kashefi, *Information theoretically secure hypothesis test for temporally unstructured quantum computation (extended abstract)*, Electron. Proc. Theor. Comput. Sci. **266**, 209 (2018), arXiv:1704.01998.

[41] C. Bădescu, R. O'Donnell, and J. Wright, *Quantum state certification*, arXiv:1708.06002.

[42] Y. Takeuchi and T. Morimae, *Verification of Many-Qubit States*, Phys. Rev. X **8**, 021060 (2018), arXiv:1709.07575.

[43] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing*, Applied and Numerical Harmonic Analysis (Springer New York, New York, NY, 2013).

[44] T. Morimae, *Hardness of classically sampling the one-clean-qubit model with constant total variation distance error*, Phys. Rev. A **96**, 040302 (2017), arXiv:1704.03640.

[45] B. M. Terhal and D. P. DiVincenzo, *Adaptive quantum computation, constant depth quantum circuits and Arthur-Merlin games*, Quant. Inf. Comp. **4**, 134 (2004), arXiv:quant-ph/0205133.

[46] H. Wilming, M. Goihl, I. Roth, and J. Eisert, *Entanglement-ergodic quantum systems equilibrate exponentially well*, arXiv:1802.02052.

[47] B. Fefferman and C. Umans, *The Power of Quantum Fourier Sampling*, (2015), arXiv:1507.05592 [quant-ph].

[48] F. G. S. L. Brandão, A. W. Harrow, and M. Horodecki, *Local random quantum circuits are approximate polynomial-designs*, Commun. Math. Phys. **346**, 397 (2016), arXiv:1208.0692.

[49] A. W. Harrow and R. A. Low, *Random quantum circuits are approximate 2-designs*, Commun. Math. Phys. **291**, 257 (2009), arXiv:0802.1919.

[50] S. Scheel, *Permanents in linear optical networks*, arXiv:quant-ph/0406127.

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

[52] R. Lipton, *New directions in testing*, in *Distributed Computing and Cryprography*, Vol. 2 (AMS, 1991) pp. 191–202.

[53] T. Jiang, *How many entries of a typical orthogonal matrix can be approximated by independent normals?* Ann. Probab. **34**, 1497 (2006), arXiv:math/0601457.

[54] N. Ermolova and S. G. Haggman, Simplified bounds for the complementary error function; application to the performance evaluation of signal-processing systems, in *2004 12th European Signal Processing Conference* (2004) pp. 1087–1090.