

Computational advantage of quantum random sampling

Dominik Hangleiter^{*}

*Joint Center for Quantum Information and Computer Science (QIICS), University of Maryland and NIST, College Park, Maryland 20742, USA
and Joint Quantum Institute (JQI), University of Maryland and NIST,
College Park, Maryland 20742, USA*

Jens Eisert[†]

*Dahlem Center for Complex Quantum Systems, Freie Universität Berlin,
14195 Berlin, Germany,
Helmholtz-Zentrum Berlin für Materialien und Energie, 14109 Berlin, Germany,
and Fraunhofer Heinrich Hertz Institute, 10587 Berlin, Germany*

 (published 20 July 2023)

Quantum random sampling is the leading proposal for demonstrating a computational advantage of quantum computers over classical computers. Recently the first large-scale implementations of quantum random sampling have arguably surpassed the boundary of what can be simulated on existing classical hardware. Here the theoretical underpinning of quantum random sampling is comprehensively reviewed in terms of computational complexity and verifiability, as are the practical aspects of its experimental implementation using superconducting and photonic devices and its classical simulation. Open questions in the field are discussed, and perspectives for the road ahead, including potential applications of quantum random sampling, are provided.

DOI: [10.1103/RevModPhys.95.035001](https://doi.org/10.1103/RevModPhys.95.035001)

CONTENTS

I. Introduction	2	b. Showing TVD robustness	19
II. Quantum Random Sampling Schemes	5	4. Additive-error sampling hardness	20
A. Universal circuit sampling	6	a. Hiding problem instances	21
B. IQP circuit sampling	6	b. Approximate average-case hardness	21
C. Boson sampling	7	c. Hiding in boson sampling	22
D. Gaussian boson sampling	7	D. Approximate average-case hardness	23
E. Further schemes	8	1. Reduction to additive or multiplicative average-case hardness	23
III. Computational Complexity of Simulating Quantum Devices	8	2. Anticoncentration	24
A. Basics of computational complexity theory	8	a. Anticoncentration via spherical designs	25
B. Where to look for a quantum-classical separation?	9	b. Anticoncentration via computing the collision probability	26
C. Computing acceptance probabilities of randomized algorithms	10	c. Further proofs of anticoncentration	27
1. Classical acceptance probabilities	10	3. Average-case hardness: An overview	27
2. Quantum acceptance probabilities	10	4. Random self-reducibility of the permanent	28
D. Approximating GapP	11	a. Improving the success probability	28
E. Approximating #P: Stockmeyer's algorithm	13	b. Distributions over infinite fields: The case of $\mathbb{F} = \mathbb{C}$	29
1. The polynomial hierarchy	13	c. Robustness to finite-precision errors	30
2. Stockmeyer's approximate counting algorithm	13	5. Average-case hardness of quantum output probabilities	30
IV. Computational Complexity of Quantum Random Sampling	14	a. Local Taylor-series truncation	30
A. Sampling versus approximating outcome probabilities	14	b. Rational-function interpolation	31
B. Strongly simulating quantum computations	14	c. Global Taylor-series truncation	32
1. IQP circuits	15	d. From continuous to discrete subgroups	32
2. Fock boson sampling	15	6. Discussion	33
3. Gaussian boson sampling	16	E. Fine-grained results	34
C. Hardness argument	17	F. Complexity of sampling in the presence of noise	35
1. Exact sampling and worst-case hardness	17	1. Noisy output distributions	35
2. Multiplicative-error sampling hardness	17	2. Fault-tolerant random sampling	35
3. From multiplicative to additive errors	18	V. Verification	36
a. Why the total-variation distance?	18	A. Hardness of verification from classical samples	36
		B. Sample-efficient classical verification via cross-entropy benchmarking	37

^{*}mail@dhangleiter.eu

1. Heavy-outcome generation	38	C. Toward applications of quantum random sampling	69
a. Computational hardness of HOG	38	1. Exploiting randomness	69
b. Fine-graining HOG:		2. Exploiting structure	69
Binned outcome generation	39	a. Using samples to solve graph problems	70
2. Cross-entropy difference	39	b. Estimating physical quantities using Gaussian boson samplers	70
3. Linear cross-entropy benchmarking fidelity	40	D. Conclusions	71
a. Sample complexity of estimating the XEB fidelity	41	Acknowledgments	71
b. Benchmarking via XEB fidelity	41	References	72
c. Single-instance verification	42		
d. Difficulty of achieving a nontrivial XEB fidelity	42		
e. Spoofing the linear XEB fidelity	42		
C. Efficient quantum verification	44		
1. Fidelity witnessing	44		
a. Fidelity witnessing via parent Hamiltonians	44		
b. Fidelity witnesses for weighted graph states	45		
c. Fidelity witnesses for quantum optical states	45		
2. Fidelity estimation	46		
D. Efficient classical verification	47		
1. State discrimination	47		
2. Cryptographic tests	48		
E. Further approaches to the verification of quantum samplers	49		
VI. Experimental Implementations	50		
A. Universal circuit sampling with superconducting circuits	50		
1. Design of the experiment	50		
2. Benchmarking of the components	51		
a. Benchmarking of single-qubit gates	51		
b. Benchmarking of two-qubit gates	51		
c. Characterization of single-qubit measurements	51		
3. Verifying the sampling task	51		
4. Follow-up work	53		
B. Photonic implementations	53		
1. Fock boson sampling	53		
2. Gaussian boson sampling	54		
C. Further implementations of quantum random sampling	55		
VII. Classically Simulating Quantum Random Sampling Schemes	55		
A. Sampling versus computing output probabilities	56		
B. Simulating universal circuit sampling	57		
1. Using tensor networks to simulate quantum circuits	57		
2. Simulation of random quantum circuits	58		
a. State-vector simulation	58		
b. Hybrid algorithms	58		
c. Simulating the experiment of Arute <i>et al.</i>	59		
d. Efficient algorithms	61		
e. Alternative simulation schemes	61		
3. Analysis of noise	61		
C. Simulating boson-sampling protocols	62		
1. Computing probabilities: Permanents and Hafnians	62		
2. Simulating the sampling task	64		
3. Analysis of noise	65		
VIII. Perspectives	66		
A. Open questions on quantum random sampling	66		
1. Understanding random quantum circuits better	66		
2. Verification beyond XEB	67		
B. Developing novel schemes	67		
1. Improving error resilience	67		
2. Relation to analog quantum simulation	68		

I. INTRODUCTION

Dating back as far as the 1980s, researchers have been thinking about what the computational power would be of computers whose constituents follow not the laws of classical physics but rather those of quantum physics (Benioff, 1980; Feynman, 1982, 1985; Deutsch, 1985). Given that quantum mechanical systems allow for superpositions and entanglement, this might give rise to a distinct model of computation compared to the paradigmatic Turing machine model that captures classical computations.

Within the model of quantum computation (Deutsch, 1985; Bernstein and Vazirani, 1997), certain computational tasks can indeed be achieved much more efficiently than is possible using classical computing devices. While for some problems such as database search (Grover, 1996) quantum computation offers polynomial speedups over classical algorithms, for others such as factoring integer numbers (Shor, 1994, 1997) and simulating quantum systems (Lloyd, 1996) it even offers presumably exponential speedups.

Within the framework of computational complexity theory, quantum computation has also been exponentially separated from classical computation via so-called oracle separations (Bernstein and Vazirani, 1993, 1997; Simon, 1994, 1997; Raz and Tal, 2019; Yamakawa and Zhandry, 2022). The advent of quantum error correction (Shor, 1996) and the threshold theorem (Aharonov and Ben-Or, 2008) brought the notion of quantum computation closer to reality, showing that (at least in principle) errors can be corrected faster than they are generated, provided that their rate is low enough.

Since these discoveries, the search for applications of quantum computation has flourished (Montanaro, 2016; Martyn *et al.*, 2021). Quantum algorithms have been discovered for solving “classical problems,” for instance, solving structured linear equations (Harrow, Hassidim, and Lloyd, 2009), solving systems of nonlinear differential equations (J.-P. Liu *et al.*, 2021), and performing optimization tasks (Farhi *et al.*, 2000; Farhi, Goldstone, and Gutmann, 2014; Brandão and Svore, 2017). More sophisticated methods for quantum simulation have been devised, such as higher-order Trotter formulas (Childs *et al.*, 2021), qubitization (Low and Chuang, 2019), and linear combination of unitaries approaches (Childs and Wiebe, 2012), and we have a much better understanding of computational primitives possible in quantum computing in terms of the quantum singular-value transform (Gilyén *et al.*, 2019) as a general way to process quantum signals (Low and Chuang, 2017, 2019).

There already is strong evidence that the dream of a universal quantum computer may become a reality in the not-too-distant future. Quantum devices have been developed

in a plethora of experimental platforms, ranging from ultra-cold atoms trapped in an optical-lattice potential (Bloch, Dalibard, and Zwerger, 2008), Rydberg atoms in optical tweezers (Bernien *et al.*, 2017) and trapped ions (Blatt and Roos, 2012) to superconducting qubits (Clarke and Wilhelm, 2008), photonic platforms (Kok *et al.*, 2007; Bartolucci *et al.*, 2021), and silicon quantum dots (Zwanenburg *et al.*, 2013). For over a decade, special-purpose analog quantum simulators have been able to qualitatively simulate variants of the Hubbard model (Jaksch *et al.*, 1998), the Heisenberg model (Friis *et al.*, 2018), and other classically intractable Hamiltonians with high precision and tunability of parameters at scales of up to tens of thousands of atoms (Trotzky *et al.*, 2010). While much smaller still, universal quantum devices are advancing at a rapid pace. Moving beyond the proof-of-principle demonstrations of quantum algorithms on small scales (Vandersypen *et al.*, 2000, 2001), first steps toward error-corrected quantum devices are currently being made (Ofek *et al.*, 2016; Egan *et al.*, 2021; Acharya *et al.*, 2022; Krinner *et al.*, 2022; Ryan-Anderson *et al.*, 2022). The quest to actually build a universal, fault-tolerant quantum computer has now also reached industry (Reagor *et al.*, 2018; Arute *et al.*, 2019; Bartolucci *et al.*, 2021; Jurcevic *et al.*, 2021). Quantum computing has thus expanded from an area of primarily academic interest to the consistent subject of news headlines around the world.

However, the devices available right now remain far from the error-correctable regime in terms of both error rates and the sheer number of qubits and quantum operations required for quantum error correction (Häner, Roetteler, and Svore, 2017; O’Gorman and Campbell, 2017; Gheorghiu and Mosca, 2019; Gidney and Eker, 2019). Available today are noisy universal quantum devices with up to roughly 50 to 100 physical qubits (Arute *et al.*, 2019; Zhu *et al.*, 2022), as well as special-purpose quantum simulators that allow for larger system sizes but lack universal programmability. When engineering those devices, one is faced with the challenge of controlling individual quantum systems with a high degree of accuracy over long times, making their improvement and scaling a monumental challenge.

Given this profound challenge associated with building a universal, fault-tolerant quantum computer, one may—and should—ask whether we should even believe that quantum computations that outperform classical computation are physically possible. This is the question at the heart of this review. The so-called extended Church-Turing thesis states that any physically implementable model of computation can be efficiently simulated using a classical computer (Vergis, Steiglitz, and Dickinson, 1986; Bernstein and Vazirani, 1997). In particular, this thesis implies that quantum computers that exponentially outperform classical computers should not be possible. And indeed, in the entire history of computation, and despite the significant evolution of computing devices, no counterexample—other than quantum computing—has been found, lending significant credibility to the thesis. Conversely, the physical possibility of quantum computers challenges the extended Church-Turing thesis.

We can think of the extended Church-Turing thesis as a computational analog of the thesis that nature must have a description in terms of a local and realistic theory (Einstein,

Podolsky, and Rosen, 1935). Bell’s inequalities (Bell, 1964) quantitatively capture how quantum theory violates this thesis and provide a concise experimental setting to test local realism. The experimental violation of a Bell inequality (Freedman and Clauser, 1972; Aspect, Dalibard, and Roger, 1982; Aspect, Grangier, and Roger, 1982) has once and for all falsified this belief and fundamentally changed the way that we think about the interactions between the local constituents of our world. Reasonable skeptics will have been convinced of this since the last closable loopholes have been closed (Giustina *et al.*, 2015; Hensen *et al.*, 2015; Shalm *et al.*, 2015).

An experimental violation of the extended Church-Turing thesis, called quantum advantage or *quantum supremacy* (Preskill, 2012), would mark a similar milestone for the field of computing. From the perspective of computer science, it would demonstrate the physical possibility of computations that are not efficiently simulable in a classical Turing machine model. From the perspective of physics, it would demonstrate that quantum theory is applicable even in regimes that are not accessible by means of the computation that we currently have.

This gives rise to the question as to what a computational analog of a Bell inequality as a means to test local realism is. In other words, what is (i) a simple task that can be performed on noisy and intermediate-scale quantum devices that is at the same time computationally difficult to simulate for classical computers both (ii) asymptotically and (iii) in practice using available computing hardware? And what could be (iv) a simple test that this task has been successfully and unambiguously achieved so that a reasonable skeptic can be convinced?

All of these requirements are extremely challenging at different levels. The central complexity-theoretic challenge is to prove an asymptotic speedup of quantum computers over classical computers, a challenge that has remained elusive for several decades now. Next, given the intrinsic complexity of the task by the first requirement, a direct verification using only classical computing resources seems impossible at first sight. The final challenge is to actually build an intermediate-scale quantum computer that is able to outperform the classical supercomputers available today. At the same time, it is a conceptual challenge to identify ways to fairly compare near-term quantum and large-scale classical computations solving the same task since their limitations are significantly different in nature. Roughly speaking, near-term quantum devices are limited by noise, while large-scale classical devices are limited by the size of the available computers.

A conceptually simple way to achieve these theoretical requirements is to make use of the quantum algorithm for integer factoring. This is because factoring is believed to be a problem for which no efficient classical algorithm exists. In fact, a large part of the presently applied public-key cryptography is based on the hardness of factoring. Factoring is particularly suited to public-key cryptography because it is believed to define a so-called one-way function, that is, a function that can be computed easily (the product of two large prime numbers) but that is extremely difficult to invert (finding those numbers given their product). Conversely, this means that verifying a successful implementation of Shor’s algorithm is straightforward: One simply has to multiply the

output and compare it to the input. While proof-of-principle demonstrations of Shor’s algorithm have been achieved (Vandersypen *et al.*, 2001), factoring a large 2048 bit number as is used for public-key encryption via the Rivest-Shamir-Adleman (RSA) cryptosystem is estimated to require a large-scale, error-corrected universal quantum computer using roughly 2×10^7 physical qubits (Häner, Roetteler, and Svore, 2017; O’Gorman and Campbell, 2017; Gheorghiu and Mosca, 2019; Gidney and Eker, 2019), thus placing this algorithm outside the realm of what could realistically be achieved in the near future. Hence, while impressive progress is being made along these lines of thought (Barends *et al.*, 2014; Acharya *et al.*, 2022; Ryan-Anderson *et al.*, 2022), factoring cannot serve as a simple and near-term test of the computational advantage offered by quantum devices.

A particularly natural class of problems for quantum computers are *sampling problems*. Indeed, any quantum mechanical experiment can be seen as simply being a sampling experiment: given an experimental prescription, a repeated measurement will provide intrinsically random measurement outcomes according to a probability distribution determined by the Born rule. Almost 20 years ago, it was first observed that the patterns of measurement outcomes resulting from certain quantum computations could in fact be so complicated that classical computers would not be able to reproduce them (Terhal and DiVincenzo, 2004).

A simple class of computations to consider as a test of quantum devices are *random quantum computations*. Such computations are presumably not computations that solve a relevant computational problem, but they may be useful in themselves, serving at the same time as a benchmark of a given computing device and as a test of quantum computational advantage. The task of sampling from the output distribution of a random quantum computation is called quantum random sampling.

In the past 20 years, significant evidence has accumulated that for a large variety of computations, and particularly for nonuniversal computations, this task is computationally intractable for classical computers (Blais *et al.*, 2004; Bremner, Jozsa, and Shepherd, 2010; Aaronson and Arkhipov, 2013; Bremner, Montanaro, and Shepherd, 2016; Fujii and Morimae, 2017; Boixo *et al.*, 2018; Björklund, Gupt, and Quesada, 2019; Bouland *et al.*, 2022; Kondo, Mori, and Movassagh, 2022; Krovi, 2022). At the same time, there is significant evidence that current-day supercomputers have a difficult time simulating this task even for small systems comprising roughly 50–100 subsystems (Neville *et al.*, 2017; Markov *et al.*, 2018; Huang *et al.*, 2020; Bulmer *et al.*, 2022; Pan, Chen, and Zhang, 2022). Recently quantum random sampling in a classically intractable regime has been claimed to be achieved experimentally on a universal quantum processor comprising 53 qubits (Arute *et al.*, 2019), or as many as 60 qubits (Wu *et al.*, 2021; Zhu *et al.*, 2022), as well as using photonic systems (Zhong *et al.*, 2020, 2021; Madsen *et al.*, 2022).

In this review, we provide a detailed overview of quantum random sampling as a test of the presumed exponential computational advantage of quantum computers over classical ones. We show in what precise way quantum random

sampling can be seen as a computation. We explain what that computation solves, in what way it outperforms classical computations, what methods of verification are available, and what challenges arise in this context.

In Secs. II–V, we focus on the theoretical aspects of quantum random sampling: the question of how to prove an asymptotic quantum speedup, and the questions as to whether and how quantum random sampling can be verified. Here we explain how the key idea of Terhal and DiVincenzo (2004) to relate the hardness of sampling to the hardness of computing probabilities has been further developed in recent years. Building on the idea to show a collapse of the so-called polynomial hierarchy (Bremner, Jozsa, and Shepherd, 2010; Aaronson and Arkhipov, 2013) based on the classical hardness of computing quantum probabilities (Valiant, 1979; Fujii and Morimae, 2017) and the assumed availability of an efficient classical sampler, this idea has been further developed to allow for certain errors in implementation (Aaronson and Arkhipov, 2013; Bremner, Montanaro, and Shepherd, 2016) and brought closer to experimental implementation (Lund *et al.*, 2014; Hamilton *et al.*, 2017; Bermejo-Vega *et al.*, 2018; Boixo *et al.*, 2018). The question of how to verify quantum random sampling was first addressed by Shepherd and Bremner (2009), and it has been pointed out that, in its most restrictive forms, classical verification is unviable (Gogolin *et al.*, 2013; Aaronson and Arkhipov, 2014; Hangleiter *et al.*, 2019). This notwithstanding, weaker forms of classical verification indeed turn out to be possible (Aaronson and Arkhipov, 2014; Boixo *et al.*, 2018; Arute *et al.*, 2019), albeit at a potentially prohibitive computational cost (Arute *et al.*, 2019).

In Secs. VI and VII, we discuss the practical aspects of quantum random sampling, particularly experimental implementations and concrete classical simulation algorithms for quantum random sampling. In the context of experimental implementation, it is essential to fully understand and analyze the noise that remains present on the device in order to devise as-robust-as-possible schemes (Boixo *et al.*, 2018; Arute *et al.*, 2019). Likewise, from the perspective of classical simulation a central question is what features of a scheme obstruct classical algorithms (Aaronson and Chen, 2017; Markov *et al.*, 2018) and, conversely, how best to exploit “weaknesses” of a scheme or a verification method in order to devise faster simulation algorithms (Clifford and Clifford, 2020; Gao *et al.*, 2021; Bulmer *et al.*, 2022; Pan and Zhang, 2022).

We stress that the topic at hand is highly conceptual in nature, so a precise understanding of the underlying premises and an appreciation of the fine print that comes along are essential. For this reason, we have made the deliberate choice of keeping the exposition precise and accurate in most places, sometimes using formal language, while at the same time pedagogically introducing all required concepts.

What we do not discuss in this review are ways to demonstrate a quantum advantage by other means. Particularly prominent examples involve the discovery of verifiable proofs of quantumness (Brakerski *et al.*, 2018, 2020; Kahanamoku-Meyer *et al.*, 2022), for which there are recent proof-of-principle demonstrations (Zhu *et al.*, 2021). These schemes demonstrate access and control over a single

qubit via a cryptographic encoding. Yamakawa and Zhandry (2022) recently made great progress along these lines by devising a verifiable proof of computational quantum advantage based on certain random computations. In this sense, it is at the interface of quantum random sampling and cryptographic proofs of quantumness. Presumably, none of these methods can be implemented at a scale required for a quantum advantage in the intermediate term, however (Hirahara and Le Gall, 2021; Zhu *et al.*, 2021; Liu and Gheorghiu, 2022).

Before we start, we point the interested reader to more concise reviews of quantum advantage (Harrow and Montanaro, 2017), quantum random sampling (Lund, Bremner, and Ralph, 2017), and implementations of boson sampling (Brod *et al.*, 2019) that may serve as starting points in the literature. In addition, Nielsen and Chuang (2010) covered the basics of quantum computing, which we do not address here.

We begin this review by setting the stage and stating what a quantum random sampling scheme is in the first place in Sec. II. There we define universal circuit sampling, instantaneous quantum polynomial-time (IQP) circuit sampling, boson sampling, and Gaussian boson sampling; we also hint at other schemes. Section III explains the basics of computational complexity to the extent that they are needed in Sec. IV to show the computational hardness of quantum random sampling on classical computers. This discussion constitutes the heart of the review: It is precisely this fine print that is needed to appreciate the significance of experimental implementations of quantum random sampling. Section V is concerned with the question of how to verify the correctness of the implementation of a quantum random sampling scheme. In Sec. VI, we detail the experimental implementations of quantum random sampling to date. Section VII provides an overview of methods of simulation run on classical supercomputers that aim to challenge quantum implementations in their computational power. Finally, in Sec. VIII we put the findings into perspective and discuss various open questions as means of taking further steps, particularly toward explore potential applications of quantum random sampling.

II. QUANTUM RANDOM SAMPLING SCHEMES

Every experiment in quantum physics can be viewed as a sampling experiment: Measurement outcomes are intrinsically random, sampled from a probability distribution determined by the Born rule. Sampling problems are therefore natural candidates exhibiting specifically quantum features. The most prominent example of a quantum-classical divide is for a specific quantum sampling problem that cannot be reproduced classically under locality constraints: the violation of a Bell inequality (Bell, 1964). Similarly, in terms of computational complexity we expect it to be difficult to reproduce the experimental outcomes of generic quantum computations. Indeed, we can think of the corresponding experiments as violating a computational equivalent of the Bell inequality. The reasons why we expect generic computations to be hard to simulate are manifold and not precisely understood; the exponentially growing Hilbert space dimension, quantum interference leading to nonpositive amplitudes, and entanglement are some examples of distinctly quantum features

obstructing classical simulation algorithms. Roughly speaking, generic quantum computations explore the entire state space available, providing no structure that can be exploited by a classical simulation algorithm. Consequently, by this reasoning the run-time of such an algorithm must be determined by the exponential Hilbert space dimension.

To make the intuition rigorous that generic quantum computations give rise to sampling problems that are classically intractable, the idea of quantum random sampling has been introduced. In quantum random sampling problems, a quantum computation is drawn at random according to some specification. The task is then to sample from the Born rule distribution generated by this random quantum computation. There are now two notions of randomness at play: The first notion is the randomness of the computation itself, which is classical randomness used to draw the computation at random. The second notion is the intrinsically quantum randomness of individual outcomes sampled from the output distribution of that computation. Not only are such quantum random sampling schemes difficult to simulate using the known classical simulation algorithms that are already at comparably small scales, but we can also give complexity-theoretic evidence for asymptotic intractability. This evidence is independent of specific algorithms and regards the intrinsic complexity of the problem by reducing it to a paradigmatic computational problem that can be independently studied and is therefore much stronger than merely the failure of our known simulation algorithms. Quantum random sampling schemes are particularly appealing for demonstrations of quantum advantage because, as we later see, the complexity-theoretic argument applies even to certain nonuniversal computations that may be comparably easy to experimentally implement.

A quantum random sampling scheme is defined by the random choice of a quantum computation realized by a quantum circuit. A *quantum circuit* describes an arrangement of quantum gates from a certain *gate set* in some spatial and temporal order, acting on a specific set of individual quantum systems, here often taken to be qubits. In a random quantum circuit individual quantum logic gates are chosen at random from a given gate set and applied to input registers according to a certain rule. For a fixed input size n , for instance, the number of qubits in a random quantum circuit, this gives rise to a family of computations, realized as a *circuit family*, that is denoted by \mathcal{C}_n . The classical sample space Ω comprises the possible measurement outcomes.

Task 1 (Quantum random sampling).—Given as input a problem size n and a circuit C chosen at random from a family \mathcal{C}_n , sample from the output distribution $p(C)$ of the circuit applied to a reference state¹ $|0\rangle$, with the probability of an outcome $S \in \Omega$ given by

$$p_S(C) = |\langle S|C|0\rangle|^2. \quad (1)$$

Depending on whether the emphasis lies on the probability distribution over the circuits C or on the outcomes S of a fixed

¹Throughout this review, we use the term “state” both for density operators ρ and for state vectors $|\psi\rangle$ in the underlying Hilbert space.

circuit, we use $p_S(C)$ at times and use $p_C(S)$ at other times for the outcome probabilities.

In the remainder of this section, we formally introduce the most important schemes: universal circuit sampling, IQP circuit sampling, and boson sampling. These schemes recurrently appear over the course of this review, in which we discuss their and similar schemes' properties. This includes not only their complexity-theoretic analysis (Sec. IV) and the question in how far classical samples from their output distributions can be verified (Sec. V) but also their experimental implementations (Sec. VI) and specific classical simulation schemes (Sec. VII).

A. Universal circuit sampling

The most prominent example of a quantum random sampling scheme, or rather family of random sampling schemes, is *universal circuit sampling*. The rationale behind universal circuit sampling is to explore the entire Hilbert space available in small- or intermediate-scale experiments as quickly as possible. This is why it is also a universal circuit sampling scheme that was implemented to experimentally demonstrate a computational quantum advantage for the first time (Arute *et al.*, 2019).

In universal circuit sampling, quantum gates are drawn from a gate set that is universal for quantum computation: that is, any quantum computation could be implemented with gates drawn from this set. The gates are placed at certain positions in a quantum circuit architecture, which might be fixed or random. The circuit might also contain other nonrandom gates.

For example, in the experiment of Arute *et al.* (2019) a specific type of random circuit is applied: in every layer of the circuit random single-qubit gates are applied to every qubit, and a specific two-qubit entangling gate is applied to each edge of a square lattice in a particular sequence; see Fig. 1(a). The single-qubit gates are drawn from the set $\{\sqrt{X}, \sqrt{Y}, \sqrt{W}\}$ in such a way that the same single-qubit gate is not allowed to sequentially repeat. Here

$$X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad (2)$$

denote the Pauli matrices and $W = (X + Y)/\sqrt{2}$. The entangling gates are given by the *i*SWAP-like gate

$$i\text{SWAP}^* = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & -i & 0 \\ 0 & -i & 0 & 0 \\ 0 & 0 & 0 & e^{-i\pi/6} \end{pmatrix}. \quad (3)$$

As a theoretically appealing toy model of random universal circuits, consider a continuous gate set $\mathcal{G} = U(4)$ comprising all two-qubit gates. In this model, a depth- N random circuit C acting on n qubits is constructed by choosing a uniformly random gate in $G \in \mathcal{G}$ according to the Haar measure, and the pair of qubits it is applied at random (Brandão, Harrow, and Horodecki, 2016). Alternatively, we can apply the gates in a parallel architecture in which each layer of the circuit comprises random gates from \mathcal{G} applied in parallel to all qubits.

B. IQP circuit sampling

A prominent family of random quantum sampling schemes that uses restricted gate sets is given by so-called IQP circuits (Shepherd and Bremner, 2009). An IQP circuit is a commuting quantum circuit that is diagonal in the Hadamard basis. Such a circuit can always be written as $C = H^{\otimes n} D H^{\otimes n}$, where D is diagonal in the computational basis and

$$H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \quad (4)$$

denotes the Hadamard gate. IQP circuits appear naturally in the context of measurement-based quantum computation (Raussendorf and Briegel, 2001). Instances of IQP circuit families are defined by diagonal circuits comprising diagonal two-qubit gates with arbitrary phases on the diagonal (Nakata, Koashi, and Murao, 2014) and circuits of Z , controlled- Z (CZ), and controlled-controlled- Z (CCZ) gates, which flip the phase of the target qubit if and only if the control qubit (CZ) or qubits (CCZ) are in the $|1\rangle$ state (Bremner, Montanaro, and Shepherd, 2016). But one can also phrase IQP circuits in the language of Hamiltonian time evolution. In this language, an IQP circuit is given by the constant-time evolution under an Ising Hamiltonian with edge weights chosen in a specific way

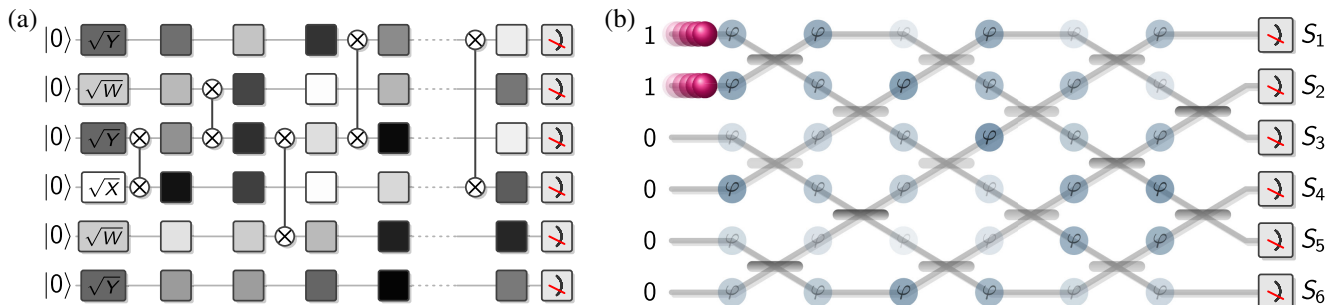


FIG. 1. Circuit diagrams for (a) random universal circuits as performed in the experiment by Arute *et al.* (2019) with random single-qubit gates from the gate set comprising \sqrt{X} , \sqrt{Y} , \sqrt{W} and fixed two-qubit entangling gates *i*SWAP* at fixed positions in the circuit, and (b) boson sampling, where passive linear optics comprising beam splitters and phase shifters are applied to a Fock input state $|1^n\rangle$ and then measured in the Fock basis with outcomes S_i .

(Bremner, Montanaro, and Shepherd, 2016). In this formulation, one can generalize IQP circuits to arbitrary multiqubit interactions: so-called X programs (Shepherd and Bremner, 2009). Another natural family of random computations in this model of computation is given by preparing a so-called cluster state (Raussendorf and Briegel, 2001; Raussendorf, Browne, and Briegel, 2003) on a square lattice and performing random local rotations around the Z axis (Haferkamp, Hangleiter, Bouland *et al.*, 2020). This model bridges a gap to quantum simulation, as it can be implemented using translation-invariant Hamiltonians (Gao, Wang, and Duan, 2017; Bermejo-Vega *et al.*, 2018).

Two specific examples of IQP circuit families that are theoretically clean and help us to illustrate important concepts in Secs. III–VIII were introduced by Bremner, Montanaro, and Shepherd (2016). An instance C_f of the first family is defined by a degree-3 Boolean polynomial $f: \{0, 1\}^n \rightarrow \{0, 1\}$ over the field $\mathbb{F}_2 = (\{0, 1\}, \oplus, \cdot)$ as

$$f(x) = \sum_{i,j,k} \alpha_{i,j,k} x_i x_j x_k + \sum_{i,j} \beta_{i,j} x_i x_j + \sum_i \gamma_i x_i, \quad (5)$$

with Boolean coefficients $\alpha_{i,j,k}, \beta_{i,j}, \gamma_i \in \{0, 1\}$ denoting whether or not CCZ, CZ, and Z gates are applied to qubits (i, j, k) , (i, j) , and i , respectively.

An instance of the second family is defined by an adjacency matrix w with entries chosen from a set of angles $A = \{0, \pi/4, \dots, 7\pi/4\}$ as

$$C_w = \exp \left[i \left(\sum_{i < j} w_{i,j} X_i X_j + \sum_i w_{i,i} X_i \right) \right], \quad (6)$$

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 $\exp(iw_{i,j} X_i X_j)$ with edge weight $w_{i,j}$ and on every vertex i , a gate $\exp(iw_{i,i} X_i)$ with vertex weight $w_{i,i}$ is performed.

C. Boson sampling

The boson-sampling scheme of Aaronson and Arkhipov (2013) is one of the most prominent and historically earliest quantum random sampling schemes. The conception of this scheme has its origins in the computational difficulty of computing the permanent of a matrix. The permanent describes the output distributions of interfering free bosons, such as single photons interfering on a beam splitter. The complexity of computing the permanent has its correspondence in a surprising physical effect: photon bunching. The experimental observation of photon bunching in the famous Hong-Ou-Mandel experiment (Hong, Ou, and Mandel, 1987) is one of the landmark experiments of quantum optics, as it was among the first experiments to experimentally confirm quantum entanglement. In this experiment, two photons interfere on a beam splitter and are measured in the photon-number basis. However, for indistinguishable photons one only ever observes zero or two photons in one of the modes, never one photon in each mode.

The boson-sampling problem generalizes this experiment. Next we increase the number of photons and let them interfere

in a complex network of beam splitters: n photons are injected into the first n of $m \in \text{poly}(n)$ modes. Those photons interfere in a linear-optical network comprising beam splitters and phase shifters that is chosen in such a way that it gives rise to a Haar-random unitary transformation of the input modes, given by $U \in U(m)$. Finally, the m output modes of the network are measured in the photon-number basis; see Fig. 1(b). As unitary mode transformations conserve the total photon number, the sample space of boson sampling is given by

$$\Phi_{m,n} = \left\{ (s_1, \dots, s_m) : \sum_{j=1}^m s_j = n \right\}, \quad (7)$$

i.e., the set of all sequences of non-negative integers of length m that sum to n . Its output distribution is

$$p_U(S) \equiv P_{\text{bs},U}(S) = |\langle S | \varphi(U) | 1_n \rangle|^2. \quad (8)$$

In Eq. (8) the state $|S\rangle$ is the Fock state corresponding to a measurement outcome $S \in \Phi_{m,n}$, $|1_n\rangle$ is the initial state with $1_n = (1, \dots, 1, 0, \dots, 0)$, and $\varphi(U)$ is the Fock space representation of the mode transformation U . To clearly distinguish the boson-sampling protocol of Aaronson and Arkhipov (2013) with output probabilities given by Eq. (8) from its variants (discussed later) we henceforth refer to it as Fock boson sampling.

D. Gaussian boson sampling

Variants of the boson-sampling protocol play with the input state and measurement basis. Most importantly, so-called Gaussian boson-sampling protocols start with a Gaussian quantum state, where the input modes are prepared in single-mode or two-mode squeezed states (Lund *et al.*, 2014; Rahimi-Keshari, Lund, and Ralph, 2015; Hamilton *et al.*, 2017; Kruse *et al.*, 2019; Grier *et al.*, 2022), or displaced squeezed states (Huh *et al.*, 2015; Quesada, 2019). The distribution of outcomes $S \in \Phi_m$ is given analogously to Eq. (8) by

$$P_{\text{GBS},U}(S) = |\langle S | \varphi(U) | g \rangle|^2, \quad (9)$$

where $|g\rangle$ is the initial Gaussian quantum state. Here the sample space

$$\Phi_m = \{(s_1, \dots, s_m) \in \mathbb{N}_0^m\} \quad (10)$$

reflects an unbounded photon number, as Gaussian states do not feature a fixed photon number. We can also think of the reverse, where a photon-number state is prepared in the input and Gaussian measurements are performed (Chabaud *et al.*, 2017; Chakhmakhchyan and Cerf, 2017; Lund, Rahimi-Keshari, and Ralph, 2017).

Gaussian boson-sampling protocols are appealing in comparison to the original proposal, as Gaussian states and measurements are experimentally much easier to implement than photon-number states and measurements. Indeed, it is in those protocols that large-scale experiments have recently been performed (Zhong *et al.*, 2020, 2021; Madsen *et al.*, 2022).

E. Further schemes

Since the first quantum random sampling schemes [IQP sampling (Bremner, Jozsa, and Shepherd, 2010) and boson sampling (Aaronson and Arkhipov, 2013)] were conceived, many more proposals for quantum random sampling schemes have been put forward. A theoretically particularly clear proposal is so-called Fourier sampling (Fefferman and Umans, 2015), which is a qubit analog of boson sampling. Another analog of boson sampling is *fermion sampling* (Oszmaniec *et al.*, 2022), for which so-called magic states are required in the input, and the closely related matchgates with magic-state inputs (Hebenstreit *et al.*, 2019). The fermionic schemes that make use of resource states as an input find their qubit analog in Clifford circuits with magic-state inputs (Hangleiter *et al.*, 2018; Yoganathan, Jozsa, and Strelchuk, 2019). The so-called one clean qubit (DQC1) model is a model in which all but one qubit is initialized in the maximally mixed state (Morimae, Fujii, and Fitzsimons, 2014; Morimae, 2017; Fujii *et al.*, 2018). This model is motivated by mixed-state quantum computations, which is a suitable framework to capture, for instance, nuclear magnetic resonance quantum processors (Negrevergne *et al.*, 2005). Other proposals include Clifford circuits that are conjugated by arbitrary product unitaries (Bouland, Fitzsimons, and Koh, 2018) and permutations of distinguishable particles in specific conditions (Aaronson *et al.*, 2016). Finally, certain models have also been proposed with the goal of closing loopholes such as the necessity to certify the correct implementation of a quantum supremacy experiment (Hangleiter *et al.*, 2017; Miller, Sanders, and Miyake, 2017), or to make such an experiment more error tolerant (Fujii, 2016; Kapourniotis and Datta, 2019).

In what follows, we discuss the properties of these schemes with respect to the possibility of using them to demonstrate a computational advantage over classical computations. Before we commence with the main focus of this review, the complexity-theoretic argument for the classical intractability of Task 1, we review some basics of computational complexity theory in Sec. III.

III. COMPUTATIONAL COMPLEXITY OF SIMULATING QUANTUM DEVICES

The previously introduced quantum random sampling schemes were devised to show computational quantum advantages of quantum devices over classical supercomputers. There are two ways for us to understand this goal: First, we can understand it in terms of the actual time required to simulate an actual experiment performing quantum random sampling. This is the realm of concrete algorithm development, and a quantum advantage in this sense is reached as soon as the available supercomputers running state-of-the-art algorithms are no longer capable of providing samples from the desired distribution. Second, we can understand it in terms of the asymptotic scaling of the best possible classical simulation algorithm. This is the realm of computational complexity theory. Computational complexity theory addresses classes of problems in terms of their intrinsic complexity in an algorithm-agnostic way. We can therefore

supplement evidence toward the first type of quantum advantage using computational complexity theory. This can help us to hedge against a “lack of imagination” in classical algorithm development.

Consider the related context of cryptography: for us to be confident in the security of a certain cryptographic scheme, it is essential that this scheme is not simply based on some problem on which known algorithms do not perform well. Rather, we collect additional evidence and, ideally, underlying reasons that in fact no algorithm can efficiently solve the problem on which the scheme is based. It is such additional, independent evidence that computational complexity theory can contribute to quantum random sampling.

Here we precisely explicate the available evidence for the classical intractability of quantum random sampling, making the intuition that quantum devices are more powerful than classical ones more rigorous. We later see which ingredients come together in a strategy to provide complexity-theoretic evidence for the hardness of sampling from, or weakly simulating, the previously defined sampling schemes. These results will constitute the complexity-theoretic underpinning of experimental prescriptions designed to demonstrate quantum computational supremacy, that is, to experimentally violate the extended Church-Turing thesis.

The argument is intricate, however, and builds on some basic results about the computational complexity of approximately computing the output probabilities of, or strongly simulating, quantum circuits, and algorithms for this task. We review those results in this section, then leverage them to weak simulation in Sec. IV.

A. Basics of computational complexity theory

To provide theoretical evidence for quantum advantage, we have to enter the realm of theoretical computer science. There classes of problems, so-called complexity classes, are studied with respect to their *computational complexity*, that is, the resources that an algorithm designed to solve problems from such a class would require in the worst case. In computational complexity theory, we can discern distinct problem classes defined by certain resource restrictions, most importantly the run-time and the memory requirement of algorithms. Understanding the relations between different complexity classes, that is, separations and inclusions between them, is the main subject of study in the theory of computational complexity. For convenience, most often *decision problems* are considered, where the task is to decide whether a given string² $x \in \{0, 1\}^*$ is in a so-called language $L \subset \{0, 1\}^*$, which is a set of bit strings. A machine that computes the Boolean function $f_L: \{0, 1\}^* \rightarrow \{0, 1\}$, which satisfies $f_L(x) = 1 \Leftrightarrow x \in L$, decides L . For example, a language L could be given by the set of all graphs for which there is a path that visits each vertex once, in binary encoding, and a string $x \in L$ is the binary encoding of a particular graph instance.

²We write the set of all finite-length bit strings as $\{0, 1\}^* = \bigcup_{n \in \mathbb{N}} \{0, 1\}^n$.

The central concept of computational complexity theory is that of an algorithm. In a simplified picture, we can think of an algorithm as computing a Boolean function $f: \{0, 1\}^* \rightarrow \{0, 1\}$ for arbitrary-length inputs. Abstractly speaking, an algorithm is a set of rules according to which a machine acts on any given input. In the case of classical algorithms, formalized as a Turing machine, those rules may involve reading bits of the input or a scratch pad and writing bits to that scratch pad, choosing a new rule according to which to continue, or stopping and outputting either 0 or 1 (Arora and Barak, 2009). We say that an algorithm is efficient if its run-time scales polynomially in the input size, given by the length $|x|$ of x .

On an actual silicon-chip computer, those rules can be implemented using certain elementary logic operations that are applied sequentially (in parallel) to some of the input registers (bits) at a time. The elementary logical operations might act on a single register or bit such as the NOT operation, on two such as OR and AND or even more registers. A set of such operations is said to be universal if an arbitrary Boolean function $f: \{0, 1\}^n \rightarrow \{0, 1\}$ can be expressed as a classical circuit using $\text{poly}(n)$ many input registers. A *classical circuit* is a mathematical model of an arrangement of classical gates implementing a logical operation that is chosen from a certain set in some spatial and temporal order computing a Boolean function. Examples of such universal sets of logical operations are {AND, NOT} and the singleton {NAND}. Using a sequence of universal logical operations, one can therefore express any other elementary logical operation. A classical circuit C_n effectively computes a function of the values of its n input registers, potentially using additional auxiliary registers. On input $x \in \{0, 1\}^n$, its outcome $C_n(x) \in \{0, 1\}$ is given by its value on a single—say, the first—output register. The size of a circuit $|C_n|$ is given by the number of gates in it. We call the model of computation in which we can execute classical circuits the circuit model.

Notice that any given circuit takes inputs of a fixed size n , while we demand that an algorithm work for any input size. We can turn a family of circuits $\{C_n\}_{n \in \mathbb{N}}$ into a meaningful algorithm³ by supplementing it with an efficient instance-generating procedure that, given the input size n , efficiently produces a description of C_n , which is then run on the input $x \in \{0, 1\}^n$. We call circuit families for which such a procedure is possible *uniform circuit families*. Uniform circuit families are therefore a realization of an algorithm in the circuit model.

The fundamental class of problems in computational complexity theory is the class \mathbf{P} , the class of problems that can be solved efficiently on a deterministic classical computer.

Definition 2 (P).—A language $L \subset \{0, 1\}^*$ is in the class \mathbf{P} if there is a classical algorithm \mathcal{A} that, given $x \in \{0, 1\}^*$ as an input, decides whether $x \in L$ in polynomial run-time in $|x|$:

$$x \in L \Leftrightarrow \mathcal{A}(x) = 1. \quad (11)$$

³Indeed, if we merely ask for the existence of a circuit family as opposed to an efficient algorithm, then this allows us to solve undecidable problems using polynomial-size circuits.

Relations between complexity classes are typically studied with respect to polynomial reductions—so-called Cook reductions—where access to a machine in \mathbf{P} is granted. A key problem in the theory of computational complexity is that the relation between different complexity classes defined with significantly different resource restrictions in mind is inherently difficult to determine. For this reason, basic relations between complexity classes are often conjectured merely based on the available evidence. The most basic and at the same time most fundamental separation in complexity theory is the belief that $\mathbf{P} \neq \mathbf{NP}$. While \mathbf{P} is the class of problems that can be efficiently computed on a classical computer, \mathbf{NP} is the class of problems that can be efficiently verified.

Definition 3 (NP).—A language $L \subset \{0, 1\}^*$ is in the class \mathbf{NP} if there is a polynomial $p: \mathbb{N} \rightarrow \mathbb{N}$ and a polynomial-time classical algorithm \mathcal{V} (called the verifier for L) such that, for every $x \in \{0, 1\}^*$,

$$x \in L \Leftrightarrow \exists y \in \{0, 1\}^{p(|x|)}: \mathcal{V}(x, y) = 1. \quad (12)$$

We call y the proof of x .

When gathering evidence for a separation between quantum and classical computation and quantum and classical sampling, in particular, we try to remain as close to problems that have been well studied, such as the conjecture $\mathbf{P} \neq \mathbf{NP}$. The main challenge is that, at the same time, the computational task must be such that it can realistically be realized on near-term quantum devices in as easy and error resilient a way as possible.

B. Where to look for a quantum-classical separation?

To better understand the complexity theory of quantum computing, we compare it to its closest cousin, randomized classical computation.⁴ We formalize randomized classical and quantum computations in terms of decision problems as complexity classes \mathbf{BPP} and \mathbf{BQP} .

Definition 4 (Classical and quantum computation).— \mathbf{BPP} (\mathbf{BQP}) is the class of all languages $L \subset \{0, 1\}^*$ for which there is a polynomial-time randomized classical (quantum) algorithm with a uniform circuit family $\{C_n\}_{n \in \mathbb{N}}$ such that, for all $n \in \mathbb{N}$ and all inputs $x \in \{0, 1\}^n$,

$$x \in L \Rightarrow \Pr[C_n(x) = 1] \geq 2/3, \quad (13)$$

$$x \notin L \Rightarrow \Pr[C_n(x) = 1] \leq 1/3, \quad (14)$$

where the probability is taken over the internal randomness of the algorithm.

Classical computations are modeled as intrinsically deterministic; only by artificially introducing randomness into the circuit do we construct a randomized classical algorithm using elementary logic gates. A randomized algorithm for a Boolean function $f: \{0, 1\}^n \times \{0, 1\}^\ell \rightarrow \{0, 1\}$ acts on both the problem input $x \in \{0, 1\}^n$ and a uniformly random bit

⁴In this section, we follow a line of thought that to our knowledge is from Scott Aaronson; see <https://www.scottaaronson.com/blog/?p=3427>.

string $r \in \{0, 1\}^\ell$ with $\ell \in \text{poly}(n)$. Randomized algorithms are at least as powerful as deterministic ones; as such, a function can simply disregard the random inputs, giving rise to a deterministic algorithm. In many practical situations, randomized algorithms turn out to be much more efficient than deterministic algorithms, however.

While classical logical gates are not generally *reversible* in that the mapping from input to output is injective, one can implement any classical computation in a circuit that uses only reversible operations (Toffoli, 1980; Fredkin and Toffoli, 1982). In other words, there are sets of reversible operations such as the three-bit Toffoli (or controlled-controlled-NOT) gate TOF (Toffoli, 1980) such that an arbitrary Boolean function can be expressed using the outcome of a single register in a computation involving only those operations.

By taking the leap to reversible classical computation, we have already made it halfway to quantum computation. Indeed, the question about the possibility of reversible classical computation was originally motivated by the observation that the laws of physics are reversible (Fredkin and Toffoli, 1982). Hence, the thinking is that a physical model of computation should be too.

Quantum circuits are a generalization of reversible classical circuits. A quantum circuit acts on qubits, the state space of which is given by \mathbb{C}^2 . The elementary operations or quantum gates are unitary matrices acting on a k -qubit input space $(\mathbb{C}^2)^{\otimes k}$, where k is a small number; typically $k = 2$. A quantum circuit acting on $m \in \text{poly}(n)$ qubit registers produces not a single bit string as an output but rather a quantum state in $(\mathbb{C}^2)^{\otimes m}$ that only upon a quantum measurement in some basis—typically the standard basis—produces a bit string as an output. Indeed, we notice that classical computation is a special case of quantum computation: If we restrict to state preparations and measurements in the standard basis and permutation matrices in that basis (which are, in particular, unitary), then we recover classical computation.

A quantum gate set \mathcal{G} is said to be *computationally universal* if an arbitrary quantum circuit acting on n qubits and using t gates can be simulated by a circuit composed of gates from \mathcal{G} up to error ϵ with overhead $\text{polylog}(n, t, 1/\epsilon)$ in terms of the numbers of both registers and gates (Aharonov, 2003). With polynomial overhead in n and t , computational universality therefore tolerates errors of the order of $2^{-\text{poly}(n,t)}$. A computationally universal gate set that serves us well in the review is the set $\{H, \text{TOF}\}$ consisting of the Hadamard and Toffoli gates. This gate set is universal for n -qubit computations when $n + 1$ many qubits are acted upon (Aharonov, 2003).

In contrast to classical computations, quantum computations are intrinsically randomized: the probability that an n -qubit quantum circuit C_n applied to an input state $|x\rangle \in \mathbb{C}^n$ results in a particular outcome y after a measurement is given by the Born rule as $|\langle y|C_n|x\rangle|^2$. We also call these probabilities the output probabilities of C_n . Indeed, it is presumably not possible to separate out the randomness from the computation, which is the case for classical computations.

An important but subtle difference between quantum and randomized classical computations presents itself in the

guise of the probability that such computations accept. This difference is a lever that allows us to separate the two types of algorithms in terms of their computational power.

C. Computing acceptance probabilities of randomized algorithms

1. Classical acceptance probabilities

We start by discussing acceptance probabilities of classical randomized algorithms before turning to quantum algorithms. The acceptance probability

$$\Pr[C_n(x) = 1] = \frac{1}{2^{p(|x|)}} \sum_{r \in \{0,1\}^{p(|x|)}} f_x(r) \quad (15)$$

of a classical randomized circuit $C_n(x)$ computing a Boolean function f_x is given by the fraction of accepting random inputs $r \in \{0, 1\}^{p(|x|)}$, where $p: \mathbb{N} \rightarrow \mathbb{N}$ is a polynomial. Computing the unnormalized acceptance probability of classical circuits is therefore clearly a #P-complete problem.⁵

Definition 5 (#P) (Arora and Barak, 2009).—The function class #P is the class of all functions $f: \{0, 1\}^* \rightarrow \mathbb{N}$ for which there is a polynomial-time classical algorithm C and a polynomial $p: \mathbb{N} \rightarrow \mathbb{N}$ such that

$$f(x) = |\{y \in \{0, 1\}^{p(|x|)} : C(x, y) = 1\}|. \quad (16)$$

In other words, #P functions, by definition, count the number of accepting inputs to a polynomial-time computation C . In contrast to BPP and BQP, which are classes of decision problems, #P is therefore a class of counting problems. In turn, we can view the decision class NP (Definition 3) as asking to decide whether there exists any input such that a computation C is accepted.

2. Quantum acceptance probabilities

We say that a quantum computation with circuit C_n accepts an input x if a measurement on $C_n|x\rangle$ results in one of a set of accepting outcomes Γ_{acc} . The acceptance probability of the computation is then given by

$$\Pr[C_n(x) = 1] = \sum_{y \in \Gamma_{\text{acc}}} |\langle y|C_n|x\rangle|^2. \quad (17)$$

For the following argument, it is sufficient to consider the set of accepting outcomes to be $\Gamma_{\text{acc}} = \{0\}$, where $0 \equiv 0^n$ denotes the all-zero outcome string; see Fenner *et al.* (1999). The acceptance probability of C_n is then given simply by a single output probability $\Pr[C_n(x) = 1] = |\langle 0|C_n|x\rangle|^2$.

We can express the acceptance probabilities of a polynomial-size quantum circuit C_n on input $x \in \{0, 1\}^\ell$ via a

⁵Given a complexity class X, we say that a problem is X hard if it is at least as hard as any problem in X in the sense that all problems in the class are polynomial-time reducible to it. We say that it is X *complete* if it is in X and X hard.

function $g_x: \{0, 1\}^{p(\ell)} \rightarrow \{+1, -1\}$ for some polynomial p as (Dawson *et al.*, 2005; Montanaro, 2017)⁶

$$\Pr[C_n(x) = 1] = \frac{1}{2^{p(\ell)}} \sum_{y \in \{0, 1\}^{p(\ell)}} g_x(y). \quad (18)$$

This is easily seen using the fact that the gate set comprising the Hadamard and the Toffoli gate is universal for quantum computing.⁷ In this gate set, we can express the all-zero amplitude of an n -qubit computation $C_n = C^{(t)} \dots C^{(1)}$ using t quantum gates $C^{(1)}, \dots, C^{(t)}$ as (Dawson *et al.*, 2005)

$$\langle 0 | C_n | x \rangle = \sum_{\lambda_1, \dots, \lambda_t} \langle 0 | C^{(t)} | \lambda_t \rangle \dots \langle \lambda_1 | C^{(1)} | x \rangle \quad (19)$$

$$= \frac{1}{\sqrt{2^h}} \sum_y s_x(y) \quad (20)$$

in terms of the number h of Hadamard gates and a signed function s_x , with the input space size given by the number of paths leading from x to 0, which is bounded by 4^t for a circuit consisting of two-qubit gates, and hence for polynomial-size circuits as $2^{\text{poly}(n)}$. This is because the matrix elements of the Toffoli gate are binary and those of the Hadamard gate are $\pm 1/\sqrt{2}$; therefore, each entry of the matrix product $C^{(t)} \dots C^{(1)}$ is a sum of numbers $(\pm 1) \times 2^{-h/2}$. We thus obtain

$$|\langle 0 | C_n | x \rangle|^2 = \frac{1}{2^h} \left| \sum_y s_x(y) \right|^2 = \frac{1}{2^h} \sum_{y, z} g_x(y, z), \quad (21)$$

where $g_x(y, z) = s_x(y)s_x(z)$.

Notice the subtle difference in the range of the function g_x versus the range of the function f_x arising in classical computation: while f_x is Boolean, g_x takes values in $\{+1, -1\}$. We can view this difference between Boolean and signed functions as a signature of *quantum interference*, as it allows for the possibility of canceling paths that was demonstrated in the Hong-Ou-Mandel experiment and discussed in Sec. I.

But we can easily translate back and forth between signed and Boolean functions via the map $g'_x(y) = [g_x(y) + 1]/2$ and reexpress

$$\sum_{y \in \{0, 1\}^{p(|x|)}} g_x(y) = |\{y: g'_x(y) = 1\}| - |\{y: g'_x(y) = 0\}|. \quad (22)$$

Notice that g'_x is again a Boolean #P function. The sum (18) can be viewed as the difference between the accepting paths of the function g'_x and its rejecting paths or, in other words, the

gap of that function. For a Boolean function $f: \{0, 1\}^n \rightarrow \{0, 1\}$ the gap is defined as

$$\text{gap}(f) = |\{y: f(y) = 1\}| - |\{y: f(y) = 0\}|, \quad (23)$$

which we normalize to

$$n\text{gap}(f) = \frac{1}{2^n} \text{gap}(f). \quad (24)$$

This is why computing functions whose values can be written as the gaps of #P functions is complete for a class called GapP.

Definition 6 (GapP) (Fenner, Fortnow, and Kurtz, 1994).—Define the function class GapP as the class of all functions $f: \{0, 1\}^* \rightarrow \mathbb{Z}$ for which there are $g, h \in \#P$ such that $f = g - h$.

Conversely, given a GapP function $g: \{0, 1\}^\ell \rightarrow \{-2^{p(\ell)}, \dots, 2^{p(\ell)}\}$ for a polynomial p , we can find an n -qubit quantum circuit $Q_g(x)$ with $n \in \text{poly}(\ell)$ that has an acceptance amplitude $\langle 0^n | Q_g(x) | 0^n \rangle = g(x)/2^n$ (Fenner *et al.*, 1999; Kondo, Mori, and Movassagh, 2022). To see this, we observe that for every GapP function g there is a polynomial-time computable function $G(x, y)$ such that $g(x) = |\{y \in \{0, 1\}^{p(|x|)}: G(x, y) = 1\}| - |\{y \in \{0, 1\}^{p(|x|)}: G(x, y) = 0\}|$. With the diagonal polynomial-size circuit $D_x = \sum_{y \in \{0, 1\}^n} (-1)^{G(x, y)} |y\rangle\langle y|$, we then find that $Q_g(x) = H^{\otimes n} D_x H^{\otimes n}$ has an acceptance amplitude $g(x)/2^n$.

Altogether we have found that acceptance probabilities of a classical circuit are given by the fraction of accepting paths of #P functions, while the acceptance probabilities of a quantum circuit C_n can be expressed as the absolute value of the normalized gap of a #P function f_0 as

$$|\langle 0^n | C_n | 0^n \rangle|^2 = |n\text{gap}(f_0)|^2. \quad (25)$$

How are GapP and #P related in terms of their computational complexity? We have already seen a simple mapping between the two, which implies that computations of GapP and #P functions are equivalent under Cook reductions⁸ that we write as

$$\text{P}^{\text{GapP}} = \text{P}^{\#P}. \quad (26)$$

Therefore, in this sense the two classes are similar. But they actually turn out to be distinct once we turn to the hardness of approximating the respective sums (15) and (18) up to a multiplicative error c .

D. Approximating GapP

Hereafter we distinguish the following notions of approximation: We say that for $c \in (0, 1]$ an estimator s provides a c -multiplicative approximation of the value S if

$$cS \leq s \leq S/c. \quad (27)$$

⁸We write a complexity class X in the exponent of another class Y to mean that a machine in Y can call an *oracle* with access to a machine solving arbitrary problems in the class X at unit time cost.

⁶We recommend the introduction to Boolean functions and their relation to quantum output probabilities given by Montanaro (2017).

⁷As discussed, since the Hadamard and Toffoli gates are computationally universal (Shi, 2002; Aharonov, 2003), the acceptance probability of an arbitrary polynomial-size computation can be expressed as the acceptance probability of such a circuit up to an error ϵ with an overhead of $\text{polylog}(1/\epsilon)$. This means that we can obtain an $O(2^{-\text{poly}(n)})$ approximation of this acceptance probability. We soon provide a more detailed discussion of such approximations and the question as to how hard it is to compute them.

We say that for $r > 0$ it is a r -relative approximation if

$$(1 - r)S \leq s \leq (1 + r)S, \quad (28)$$

and it is an ϵ -additive approximation for $\epsilon > 0$ if

$$|S - s| \leq \epsilon. \quad (29)$$

To intuitively see why there might be a difference in approximability, notice that a $\#\mathbf{P}$ sum over m -bit strings takes on values between 0 and 2^m . Typically, the values are therefore on the order of 2^m , so a constant relative error is also of that order. Conversely, \mathbf{GapP} sums take on values between -2^m and $+2^m$, but as the corresponding $\#\mathbf{P}$ function takes on an exponentially large value, the value of the \mathbf{GapP} function is the difference between two such exponentially large numbers. This difference will in general be much smaller than each individual value so a relative error is too.

A relative-error approximation of a quantity is guaranteed to have the correct sign. In contrast to relative-error approximations of $\#\mathbf{P}$ functions, which always have a positive sign, relative-error approximations of \mathbf{GapP} functions therefore teach us nontrivial sign information. In fact, this information is already sufficient to learn the exact value of any \mathbf{GapP} function up to arbitrary relative error.

Lemma 7 (Approximating \mathbf{GapP}).—Let f be a $\#\mathbf{P}$ function. Approximating $\text{gap}(f)$ up to any constant multiplicative error is then \mathbf{GapP} hard.

A detailed proof of Lemma 7 is given, for instance, in Chap. 2.2 of Hangleiter (2021). The basic idea is to use a \mathbf{GapP} oracle to iteratively compute the gap of a function f_s , which is shifted compared to the gap of f by $s2^n$. We can then compare the signs of the two gaps and vary the value of s to perform a binary search.

For a function class \mathbf{X} we define $\mathbf{Apx}_c\mathbf{X}$ as the class of problems that can be solved by approximating $\sum_x f(x)$ up to a multiplicative error c for $f \in \mathbf{X}$. We have now found that, for any $c \in (0, 1)$,

$$\mathbf{P}_{\mathbf{Apx}_c\mathbf{GapP}} = \mathbf{P}_{\mathbf{GapP}}. \quad (30)$$

Notice that in our discussion of the hardness of approximating \mathbf{GapP} using the sign information we have glossed over the fact that acceptance probabilities of quantum circuits are non-negative. And indeed it seems unlikely that those acceptance probabilities are difficult to approximate up to any constant multiplicative error.

Nevertheless, using a similar proof strategy one can prove the \mathbf{GapP} hardness of approximations for the square of the output amplitudes of quantum circuits (Terhal and DiVincenzo, 2004; Aaronson and Arkhipov, 2013; Fujii and Morimae, 2017; Goldberg and Guo, 2017). This strategy shows that multiplicative-error approximations not only get the sign correct but also the instances in which the true value is exactly zero. Additionally, there is a trivial additive-error robustness given by the spacing of the values of a normalized $\#\mathbf{P}$ function.

Lemma 8 (Approximating the absolute value of ngap).—Let $f: \{0, 1\}^\ell$ be a $\#\mathbf{P}$ function. Approximating $\text{ngap}(f)^2$ up

to (a) any relative error $\epsilon < 1/2$ or (b) additive error $1/2^{2n}$ with $n \in \text{poly}(\ell)$ is \mathbf{GapP} hard.

Proof sketch.—For part (b) of the proof we note that the additive-error robustness $1/2^{2n}$ is trivial since the spacing of the function $\text{ngap}(f)$ is given by $2/2^n$, i.e., twice the normalization of $\text{gap}(f)$ in the definition of $\text{ngap}(f)$.

For part (a) of the proof we proceed as in the proof of Lemma 7, following Proposition 8 of Bremner, Montanaro, and Shepherd (2016). The idea of the proof is to estimate $\text{ngap}(f)$ using the fact that, given a guess c , an algorithm that outputs a relative-error approximation to $|\text{ngap}(f) - c|$ can certify the correctness of c .

In the first step, we show that there is a polynomial-size classical circuit C acting on $p(\ell) + \ell + 1$ registers for some polynomial $p: \mathbb{N} \rightarrow \mathbb{N}$ that computes a shifted function f_c such that $\text{ngap}(f_c) = [\text{ngap}(f) - c]/2$ for some $c \in [-1, 1]$ such that $c = 2k/2^{p(\ell)}$, with $k \in \mathbb{N}$. To this end we make use of the following: for any polynomial $p: \mathbb{N} \rightarrow \mathbb{N}$ there is a polynomial-size circuit D_c acting on $p(\ell)$ registers computing a function g such that $\text{ngap}(g) = -c$. Now consider the polynomial-size circuit Q_c acting on $p(\ell) + \ell + 1$ registers that executes either C or D_c depending on the control register. This circuit computes a function f_c as desired.

Assume that we have an efficient algorithm \mathcal{A} that given a circuit C approximates $\text{ngap}(f_c)$ up to a relative error $\epsilon < 1$. On input Q_c this machine can certify whether $\text{ngap}(f) = c$. We now employ \mathcal{A} to estimate $\text{ngap}(f)$ using a sequence of guesses c_0, c_1, \dots for its value until we find its exact value. At each step, we have a guess c_i for c , starting with $c_0 = 0$. We use \mathcal{A} to output an estimate d_i to $|\text{ngap}(f) - c_i|$ and then apply it again to output an estimate d_i^\pm of $|\text{ngap}(f) - (c_i \pm d_i)|$. Define $c_{i+1} = c_i + d_i$ if $d_i^+ \leq d_i^-$ and as $c_i - d_i$ otherwise.

The algorithm acts contractively: Assuming that $c < \text{ngap}(f)$, we find that an estimate $d = (1 + \gamma)|c - \text{ngap}(f)|$ for some $|\gamma| < \epsilon$ satisfies

$$|c + d - \text{ngap}(f)| = |\gamma[\text{ngap}(f) - c]| \leq \epsilon|c - \text{ngap}(f)|, \quad (31)$$

and a similar inequality holds for $c - d$ if $c > \text{ngap}(f)$. Consequently, since $\text{ngap}(f)$ is an integer multiple of $2/2^n$, if the correct choice of $c \pm d$ is made in each step, the algorithm halts after $O(n)$ many steps.

It remains to be shown that the algorithm indeed halts after $O(n)$ steps. This can be seen from the equivalence

$$(1 + \epsilon)|c + d - \text{ngap}(f)| < (1 - \epsilon)|c - d - \text{ngap}(f)| \\ \Leftrightarrow (1 + \epsilon)|\gamma| < (1 - \epsilon)|2 + \gamma|, \quad (32)$$

which holds for $|\gamma| \leq \epsilon < 1/2$. The same argument immediately holds for $|\text{ngap}(f)|$, as we have not used the sign of $\text{ngap}(f)$. ■

Given the mapping of gaps to the previously described output amplitudes of quantum circuits, it therefore follows directly from Lemma 8 that approximating the output probabilities of quantum circuits is \mathbf{GapP} .

Corollary 9 (Approximating output probabilities of quantum circuits).—Approximating the output probabilities

$|\langle 0^n | C | 0^n \rangle|^2$ of an n -qubit quantum circuit comprising m gates is **GapP** complete up to (a) any relative error $\epsilon < 1/2$ or (b) the exponentially small additive error $1/2^{2^n}$.

E. Approximating #P: Stockmeyer's algorithm

For many #P-complete problems, such as computing the value of the permanent of a matrix taking values in $\{0, 1\}$, there are efficient randomized approximation schemes, including the so-called fully polynomial randomized approximation scheme (FPRAS) (Jerrum, Sinclair, and Vigoda, 2004). Many such algorithms for approximate counting are based on Markov-chain Monte Carlo methods (Jerrum, Valiant, and Vazirani, 1986; Jerrum and Sinclair, 1993). The property that those algorithms exploit is the fact that each element of the sum (15) is non-negative. Thus, the sum can be estimated by *importance sampling*, that is, sampling its elements according to their normalized weight in the sum. The intricate sign structure of **GapP** functions is what makes their relative-error approximation via such sampling algorithms difficult.

Going beyond specific algorithms, in this section we introduce a powerful general result on the approximability of such functions using a computationally restricted algorithm with access to an NP oracle from Stockmeyer (1983). Stockmeyer's algorithm can approximately count the number of accepting paths of #P functions up to small multiplicative errors, even though it is not able to exactly compute this number. It thus provides a rigorous foundation for the distinction between the approximability of **GapP** and #P. In Sec. IV, we leverage the power of this algorithm to derive rigorous separations between classical and quantum sampling algorithms.

Before we can make those statements precise, however, we need to dive a little further into the depths of computational complexity theory and define what is called the polynomial hierarchy. Stockmeyer's algorithm lies in the third level of the polynomial hierarchy. This class is much more powerful than NP, but much less powerful than #P.

1. The polynomial hierarchy

We have already seen the most important classes in the theory of computational complexity, namely, P and NP. It is no exaggeration to say that the conjecture that $P \subsetneq NP$ is indeed one of the most tested and studied unproven statement that scientists across a range of disciplines are confident about. Among other things, this intuition rests on the presumed existence of problems whose solutions are difficult to find but easy to verify. In particular, the possibility of public-key cryptography is based on the existence of such problems. It is a generalization of this statement that forms the complexity-theoretic grounding of claims to quantum supremacy. This generalization posits that the levels of an infinite hierarchy of complexity classes (the so-called polynomial hierarchy) are strict subsets of one another. Considering hypothetical algorithms within and outside of this hierarchy also allows us to understand the computational complexity of approximating #P functions.

Definition 10 (The polynomial hierarchy) (Arora and Barak, 2009).—For $i \in \mathbb{N}_0$ a language $L \subset \{0, 1\}^*$ is in Σ_i if

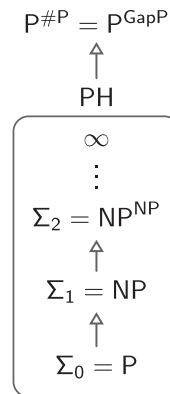


FIG. 2. The polynomial hierarchy is a hierarchy of complexity classes defined by adding consecutive NP oracles, where any layer is presumed to strictly contain all lower-lying layers. Toda's theorem (Theorem 12) states that the polynomial hierarchy is contained in $P^{\#P}$.

there is a polynomial q and a uniform polynomial-time circuit family $\{C_n\}_{n \geq 1}$ such that $x \in L$ if and only if

$$\exists u_1 \in \{0, 1\}^k \quad \forall u_2 \in \{0, 1\}^k \cdots Q_i u_i \in \{0, 1\}^k: \\ C_{|x|}(x, u_1, \dots, u_i) = 1, \quad (33)$$

where $k = q(|x|)$ and Q_i denotes a \forall or \exists quantifier depending on whether i is even or odd, respectively. The *polynomial hierarchy* PH is the set $\cup_i \Sigma_i$.

Clearly $\Sigma_i \subset \Sigma_{i+1}$. Notice that $NP = \Sigma_1$ since in Definition 3 there is only a single \exists quantifier. We can then equally characterize Σ_i as Σ_{i-1}^{NP} , so in each level an additional NP oracle is added; see Fig. 2. Intuitively, as we add alternating \exists and \forall quantifiers, the complexity of the problems solved by the circuit family $\{C_n\}$ strictly increases. Conversely, if two levels of the hierarchy coincide, then so will all other levels above them. Indeed, it is a central conjecture that the polynomial hierarchy is infinite, i.e., that every level strictly contains the previous levels. In other words, the conjecture is that “the polynomial hierarchy does not collapse.”

2. Stockmeyer's approximate counting algorithm

Indeed, it is no surprise that, given access to NP oracles, one can solve a rich class of computational problems. Nevertheless, it is surprising that one can efficiently approximate exponentially large sums up to any inverse polynomial *multiplicative error*. Stockmeyer's approximate counting algorithm (Stockmeyer, 1983) achieves this task in a low level of the polynomial hierarchy, the third level. We are now ready to state this result.

Theorem 11 (Stockmeyer, 1983; Aaronson and Arkhipov, 2013).—Given a Boolean function $f: \{0, 1\}^n \rightarrow \{0, 1\}$, let

$$p = \Pr_{x \in \{0, 1\}^n} [f(x) = 1] = \frac{1}{2^n} \sum_{x \in \{0, 1\}^n} f(x). \quad (34)$$

Thus, for all $c \geq 1 + 1/\text{poly}(n)$, there is an FBPP^{NP} machine⁹ that approximates p to within a multiplicative factor of c .

See Trevisan (2008) and Hangleiter (2021) for a sketch of the proof. Theorem 11 characterizes the complexity of approximately counting up to an inverse polynomially small multiplicative error: Since $\text{BPP} \subset \Sigma_2$ (Lautemann, 1983), and therefore $\text{BPP}^{\text{NP}} \subset \Sigma_3$, this task lies within the third level of the polynomial hierarchy. But where does this complexity class lie in relation to exactly computing a $\#\text{P}$ sum? For the answer, we refer to a final fact in complexity theory, namely, that exactly computing $\#\text{P}$ functions lets one solve any task in PH .

Theorem 12 (Toda's theorem) (Toda, 1991).—

$$\text{PH} \subset \text{P}^{\#\text{P}}. \quad (35)$$

The complexity of counting $\#\text{P}$ sums is therefore significantly easier when considering multiplicative approximations, as opposed to exact computation. Conversely, we saw in Eq. (26) and Lemma 7 that GapP does not change its complexity under multiplicative approximations. Therefore, the inclusions

$$\text{P}^{\text{Apx.}\#\text{P}} \subset \Sigma_3 \subsetneq \text{PH} \subset \text{P}^{\text{GapP}} = \text{P}^{\text{Apx.}\text{GapP}} \quad (36)$$

hold for any constant $c > 0$ since $\text{P}^{\text{GapP}} = \text{P}^{\#\text{P}} \supset \text{PH}$. The separation $\Sigma_3 \subsetneq \text{PH}$ marks the conjectured noncollapse of the polynomial hierarchy to any finite level. The same inclusions hold true when restricted to GapP functions with non-negative gaps for values of $c < 1/2$.

We have now carved out a substantial difference in complexity between quantum and randomized classical algorithms in terms of the computational complexity of approximating the respective acceptance probability to high precision. To describe quantum acceptance probabilities, negative signs are required, and hence they are GapP hard to approximate up to relative error. Conversely, classical acceptance probabilities can be expressed as sums over non-negative numbers, and hence approximating them up to relative error is in the class Σ_3 . We again emphasize that neither the quantum nor the classical algorithm should be able to multiplicatively approximate the respective acceptance probabilities, because the classes Σ_3 and GapP are not expected to be contained in BPP and BQP , respectively. Nevertheless, this difference in complexity serves as an important tool with which we can amplify harder-to-pin-down differences in the run-time of actual classical and quantum algorithms. Following this route, we eventually arrive at a conditional exponential separation for sampling tasks.

IV. COMPUTATIONAL COMPLEXITY OF QUANTUM RANDOM SAMPLING

A. Sampling versus approximating outcome probabilities

Our goal in this section is to prove not only that there is an exponential quantum versus classical divide in approximating

⁹ FBPP is the function-class equivalent of the decision class BPP , that is, the class of functions computable in probabilistic polynomial time with bounded failure probability.

output probabilities of computations but also that this divide reappears when it actually comes to performing such computations, that is, performing the corresponding sampling. Randomized algorithms indeed seem to be the perfect playground where we might see a quantum advantage, since any quantum computation naturally produces random samples from the distribution determined by the Born rule, while classical randomized algorithms require external randomness.

To make a rigorous statement about randomized computations, we consider the task of sampling from a given distribution, not caring about specific outcomes of the computation. To be able to apply the machinery of complexity theory and Stockmeyer's algorithm, in particular, it also proves useful to consider the task of sampling from randomly chosen quantum computations.

The key idea that we use to make a rigorous statement about the complexity of classical and quantum sampling is to relate the task of sampling from a distribution to computing its output probabilities. In doing so, we leverage the complexity-theoretic difference between computing classical and quantum output probabilities to classical and quantum sampling. The key technical ingredient when doing so is Stockmeyer's algorithm. We observe that Stockmeyer's counting theorem (Theorem 11) can be directly applied to estimating the acceptance probability, and in fact all output probabilities, of so-called derandomizable sampling algorithms, which are deterministic algorithms with random inputs, as previously discussed; see Definition 3.11 and the proof of Theorem 1.1 given by Aaronson and Arkhipov (2013).

*Definition 13 (Derandomizable sampling).—*A derandomizable sampling algorithm is an algorithm \mathcal{A} that takes as an input a particular instance $y \in \{0, 1\}^n$ of a problem, as well as a uniformly random string $r \in \{0, 1\}^{\text{poly}(|y|)}$ and outputs a random bit string $x = \mathcal{A}(y, r)$ that is distributed according to

$$p_y(x) = \Pr_r[\mathcal{A}(y, r) = x]. \quad (37)$$

If \mathcal{A} is such a derandomizable algorithm, we can use Stockmeyer's algorithm to estimate its output probabilities (37). To do so, we define its input function as

$$f_y: \{0, 1\}^{\text{poly}(|y|)} \rightarrow \{0, 1\},$$

$$r \mapsto \begin{cases} 1 & \text{if } \mathcal{A}(y, r) = x, \\ 0 & \text{otherwise.} \end{cases} \quad (38)$$

The output of Stockmeyer's approximation algorithm will then be a $[1 - 1/\text{poly}(|y|)]$ -multiplicative approximation to the probability $p_y(x)$. This provides the sought-after connection between sampling and approximation of probabilities that form the basis of the forthcoming proofs of sampling hardness.

B. Strongly simulating quantum computations

For the specific schemes presented in Sec. II, approximating the output probabilities is in fact a GapP -hard task and

thus just as hard as for arbitrary quantum computations. Generally, and this is particularly true for universal random circuits, the output probabilities of a circuit family \mathcal{C} are GapP hard to approximate if the circuit family generates the entirety of BQP after so-called postselection (Fujii and Morimae, 2017). In a postselection argument we compare two probabilistic complexity classes by granting ourselves the ability to restrict attention to a certain subset of desired outcomes even if that subset has an exponentially small probability. A post-selected class postA is defined as a class of decision problems that we can solve using a computation within \mathbf{A} and post-selecting on certain outcomes with a bounded error (Fujii and Morimae, 2017).

Definition 14 (Postselected class) (Fujii and Morimae, 2017).—A language L is in the class postA if there is a uniform family of circuits $\{C_x\}$ associated with \mathbf{A} for which there are a single output register O_x and a $\text{poly}(|x|)$ -size postselection register P_x such that (i) if $x \in L$, then $\Pr(O_x = 1 | P_x = 00 \cdots 0) \geq 2/3$, and (ii) if $x \notin L$, then $\Pr(O_x = 1 | P_x = 00 \cdots 0) \leq 1/3$.

Aaronson (2005) showed that $\text{postBQP} = \text{PP}$, where PP is the decision-problem equivalent of $\#\text{P}$ that asks whether at least half of the inputs are accepted. This implies that $\text{P}^{\text{postBQP}} = \text{P}^{\text{PP}} = \text{P}^{\#\text{P}} \supset \text{PH}$. Building on this result, Fujii and Morimae (2017) demonstrated that if $\text{postA} = \text{postBQP}$, then a machine that approximates the output probabilities of circuits associated with \mathbf{A} up to a multiplicative error $1/\sqrt{2} < c < 1$ can be used to decide any problem in PP , and hence also any problem in GapP . This is because the $\text{postA} = \text{postBQP}$ condition ensures that \mathbf{A} is rich enough to encode the output probabilities of arbitrary quantum computations and hence gaps of $\#\text{P}$ functions.

Taking a different perspective, one can show that the output probability of a universal quantum circuit can encode hard instances of the *Jones polynomial* (Kuperberg, 2015; Goldberg and Guo, 2017; Mann and Bremner, 2017) as well as *Tutte polynomials* (Kuperberg, 2015; Goldberg and Guo, 2017) and certain Ising model partition functions (Bremner, Montanaro, and Shepherd, 2016; Boixo *et al.*, 2018). In particular, estimating those quantities up to a relative error $1/4 + o(1)$ is $\#\text{P}$ hard.¹⁰ Expressing the output probabilities in terms of such quantities, which have been studied in detail in the literature, also proves to be extremely useful once we get to approximate sampling hardness.

Similarly, the output probabilities of several restricted quantum computational models, including the previously discussed ones, can be expressed in terms of universal quantities that are GapP hard to approximate (Fefferman and Umans, 2015; Gao, Wang, and Duan, 2017; Miller, Sanders, and Miyake, 2017; Morimae, 2017; Bermejo-Vega *et al.*, 2018; Bouland, Fitzsimons, and Koh, 2018; Fujii *et al.*, 2018). In the following, we illustrate how this is achieved using the paradigmatic schemes introduced in Sec. II.

¹⁰Notice that achieving a relative error $1/4 + o(1)$ is slightly more demanding than a multiplicative error $1/\sqrt{2}$.

1. IQP circuits

As a particularly illustrative example of such reasoning, for IQP circuits one finds¹¹ that $\text{postIQP} = \text{postBQP}$. In addition, for IQP circuits defined by a weighted adjacency matrix W [see Eq. (6)] the output amplitude

$$\langle 0 | C_W | 0 \rangle = \frac{1}{2^n} Z_W \quad (39)$$

can be expressed as an imaginary-temperature partition function of an Ising model (Bremner, Montanaro, and Shepherd, 2016; Fujii and Morimae, 2017):

$$Z_W = \sum_{z \in \{\pm 1\}^n} \exp \left[i \left(\sum_{i < j} w_{i,j} z_i z_j + \sum_i w_{i,i} z_i \right) \right]. \quad (40)$$

An analogous reduction can be made for the universal circuits of Boixo *et al.* (2018) with CZ gates. The modulus square $|Z_W|^2$ of such partition functions has been shown to be GapP hard to approximate up to a relative error $1/4 + o(1)$ (Fujii and Morimae, 2017; Goldberg and Guo, 2017).

For an IQP circuit C_f defined by a Boolean degree-3 polynomial f with coefficient vectors α, β , and γ [see Eq. (5)], one finds that the all-zero amplitude is given by the gap of f as¹²

$$\begin{aligned} \langle 0 | H^{\otimes n} C_f H^{\otimes n} | 0 \rangle &= \frac{1}{2^n} \sum_{x,y} \langle y | C_f | x \rangle \\ &= \frac{1}{2^n} \sum_x (-1)^{f(x)} = \text{ngap}(f). \end{aligned} \quad (44)$$

We have seen that approximating the gaps of arbitrary $\#\text{P}$ functions f up to multiplicative errors $1/\sqrt{2}$ is GapP complete. This remains true when the function f is restricted to a degree-3 Boolean polynomial over the field \mathbb{F}_2 since IQP circuits are universal with postselection (Bremner, Montanaro, and Shepherd, 2016).

2. Fock boson sampling

The output distribution $P_{\text{bs},U}$ of a Fock boson-sampling experiment [see Eq. (8)] can be expressed as (Scheel, 2008)

¹¹This can be shown using a gadget to implement the Hadamard gate via teleportation, the idea being that what IQP circuits lack in universality is the possibility to switch between X and Z bases. By measuring a single output line, one can teleport a Hadamard gate to an arbitrary position in the circuit using gate teleportation (Bremner, Montanaro, and Shepherd, 2016; Montanaro, 2017).

¹²To see this, note that

$$Z_i |x\rangle = (-1)^{x_i}, \quad (41)$$

$$\text{CZ}_{i,j} |x\rangle = (-1)^{x_i x_j}, \quad (42)$$

$$\text{CCZ}_{i,j,k} |x\rangle = (-1)^{x_i x_j x_k}. \quad (43)$$

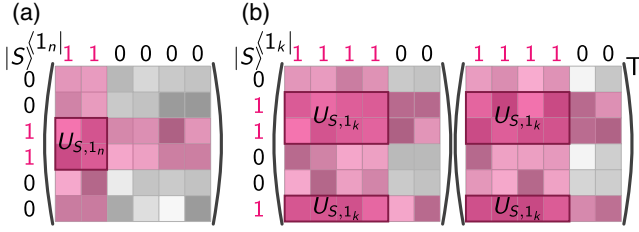


FIG. 3. (a) The output probabilities of Fock boson sampling [Eq. (45)] can be expressed as the modulus squared of the permanent of a submatrix $U_{S,1_n}$ of the Haar-random unitary U constructed by discarding rows and columns according to the outcome and input registers $|S\rangle$ and $|1\rangle_k$. (b) Analogously, the output probabilities of Gaussian boson sampling [Eq. (54)] with squeezed state inputs on k modes are proportional to the modulus squared of the Hafnian of $U_{S,1_k} U_{S,1_k}^T$.

$$P_{\text{bs},U}(S) = \frac{|\text{Perm}(U_{S,1_n})|^2}{\prod_{j=1}^m (s_j!)} \quad (45)$$

in terms of the permanent of the matrix $U_{S,1_n} \in \mathbb{C}^{n \times n}$, which can be obtained from U according to the following prescription. Define the submatrix $U_{S,S'}$ with $S, S' \in \mathbb{N}^m$ as follows: for all $j, k \in [m] = \{1, 2, \dots, m\}$, keep a matrix comprising S_j copies of the j th row of U and now write S'_j copies of the k th column of that matrix into $U_{S,S'}$; see Fig. 3(a). For so-called collision-free outcomes $S \in \Phi_{m,n}$, that is, outcomes with entries given only by 0 or 1, $U_{S,1_n}$ is therefore a certain submatrix of U . The permanent of a matrix $X = (x_{j,k}) \in \mathbb{C}^{n \times n}$ is defined analogously to the determinant, but without the negative signs, as

$$\text{Perm}(X) = \sum_{\tau \in \text{Sym}_n} \prod_{j=1}^n x_{j,\tau(j)}, \quad (46)$$

where Sym_n labels all permutations of the set $[n] = \{1, 2, \dots, n\}$.

It is a well-known fact that computing the permanent of a matrix is a problem that is #P hard even when one restricts to binary matrices (Valiant, 1979). At the same time, its close cousin the determinant can be exactly computed in polynomial time. Theorem 4.3 of Aaronson and Arkhipov (2013) extended the notable result of Valiant (1979) to approximations of the modulus squared of the permanent up to multiplicative errors. More precisely, they showed that for any $c \in [1/\text{poly}(n), 1]$ approximating $\text{Perm}(X)^2$ up to multiplicative error c for $X \in \mathbb{R}^{n \times n}$ remains GapP hard by a reduction similar to that used to prove Lemma 7 on multiplicative-error GapP hardness of computing the modulus of the gap of a #P function.

3. Gaussian boson sampling

Similarly, the output distribution of Gaussian boson sampling [see Eq. (9)] can be expressed as (Hamilton et al., 2017; Kruse et al., 2019)

$$P_{\text{GBS},U}(S) = \det(\sigma_Q)^{-1/2} \frac{\text{Haf}(M_S)}{\prod_{j=1}^m (s_j!)}, \quad (47)$$

in terms of the so-called Hafnian of a matrix M_S that is constructed as follows. Let $\sigma \in \mathbb{C}^{2m \times 2m}$ be the covariance matrix¹³ of the Gaussian state $\varphi(U)|g\rangle$ prior to the measurement and $\sigma_Q = \sigma + \mathbb{1}_{2m}/2$. Set

$$M = \begin{pmatrix} 0_m & \mathbb{1}_m \\ \mathbb{1}_m & 0_m \end{pmatrix} (\mathbb{1}_{2m} - \sigma_Q^{-1}), \quad (48)$$

where $\mathbb{1}_m$ denotes the $m \times m$ identity matrix. Analogously to how we construct a submatrix $U_{S,S'}$ from U , we obtain the submatrix M_S of M as follows: for every $j \in [m]$, M_S comprises S_j copies of the j th and $(m+j)$ th row and column of M , respectively; see Fig. 3(b). Hence, if $n = \sum_j S_j$ many photons are detected, then M_S is a symmetric $2n \times 2n$ complex matrix. Like the permanent, the Hafnian of a matrix is a certain polynomial in its matrix entries and is defined for a matrix $A \in \mathbb{C}^{2n \times 2n}$ as

$$\text{Haf}(A) = \sum_{\sigma \in \text{PMP}(2n)} \prod_{j=1}^n A_{\sigma(2j-1), \sigma(2j)}, \quad (49)$$

where $\text{PMP}(2n)$ is the set of all perfect matching permutations of $2n$ elements, that is, permutations $\sigma: [2n] \rightarrow [2n]$ that for every i satisfy $\sigma(2i-1) < \sigma(2i)$ and $\sigma(2i-1) < \sigma(2i+1)$ (Barvinok, 2016a). In particular, the permanent of A can be written as a special case of the Hafnian as

$$\text{Perm}(A) = \text{Haf} \begin{pmatrix} 0 & A \\ A^T & 0 \end{pmatrix}, \quad (50)$$

and hence approximating the Hafnian is at least as difficult as approximating the permanent, namely, GapP hard, in the worst case.

The output probabilities of Gaussian boson sampling take a particularly simple form if the input state $|g\rangle$ is a product of single-mode squeezed states with squeezing parameters r_i , which is the setting that has been studied in experiments (Zhong et al., 2020, 2021). In this case, the covariance matrix σ of the Gaussian state before detection can easily be derived to be given as

$$\sigma = \frac{1}{2} \begin{pmatrix} U & 0 \\ 0 & U^* \end{pmatrix} \Sigma \Sigma^\dagger \begin{pmatrix} U^\dagger & 0 \\ 0 & U^T \end{pmatrix}, \quad (51)$$

with $U \in U(m)$ the Haar-random unitary transformation of the input modes, and

$$\Sigma = \begin{pmatrix} \bigoplus_{i=1}^m \cosh(r_i) & \bigoplus_{i=1}^m \sinh(r_i) \\ \bigoplus_{i=1}^m \sinh(r_i) & \bigoplus_{i=1}^m \cosh(r_i) \end{pmatrix}. \quad (52)$$

The output probabilities can then be written in terms of the matrix $A = U[\bigoplus_{i=1}^m \tanh(r_i)]U^T$ as

$$P_{\text{GBS},U}(S) = \frac{1}{\prod_{j=1}^m \cosh(r_j)} |\text{Haf}(A_{S,S})|^2, \quad (53)$$

¹³See Kok and Lovett (2010) for an introduction to continuous-variable quantum information processing.

recalling the definition of $A_{S,S'}$ from Sec. IV.B.2; see also Fig. 3(b). These probabilities take a particularly simple form whenever k out of the m modes are prepared in single-mode squeezed states with a uniform squeezing parameter r and the other $m - k$ modes are prepared in the vacuum state. In this case

$$P_{\text{GBS},U}(S) = \frac{\tanh^k(r)}{\cosh^k(r)} |\text{Haf}(U_{S,1_k} U_{S,1_k}^T)|^2. \quad (54)$$

C. Hardness argument

We are now in a position to prove that under certain conditions on the quantum circuit family \mathcal{C} sampling from the output distribution of a random instance $C \in \mathcal{C}$ cannot be done in classical polynomial time in the size of C , i.e., polynomial in the number of qubits. The idea of the proof is to exploit the fact that approximating output probabilities of unitaries in \mathcal{C} is **GapP** hard. In contrast, if there was an efficient (derandomizable) sampling algorithm for a random $C \in \mathcal{C}$, then we could approximate its output probability using Stockmeyer’s algorithm. But because Stockmeyer’s algorithm lies in the third level of the polynomial hierarchy, the existence of such an algorithm implies that $\Sigma_3 \supset \text{P}^{\text{GapP}} \supset \text{PH}$: the polynomial hierarchy collapses to its third level. Assuming the generalized $\text{P} \neq \text{NP}$ conjecture that the polynomial hierarchy is infinite, this rules out the existence of an efficient sampling algorithm for circuits from \mathcal{C} . In the following we present this argument, which was given in detail by Bremner, Jozsa, and Shepherd (2010), Bremner, Montanaro, and Shepherd (2016), and Aaronson and Arkhipov (2013).

1. Exact sampling and worst-case hardness

We formalize the previously sketched idea in the following theorem.

Theorem 15 (Exact sampling hardness).—Let \mathcal{C} be a family of quantum circuits such that there is a constant $c \in (0, 1]$ for which approximating the output probabilities up to multiplicative error c is **GapP** hard. If there is an exact derandomizable sampling algorithm for circuits in \mathcal{C} , then the polynomial hierarchy collapses to its third level Σ_3 .

Proof.—Suppose that there is a derandomizable sampling algorithm \mathcal{A} that, given as an input a description of a circuit $C \in \mathcal{C}$, could efficiently sample from its output distribution $p(C)$ as defined in Eq. (1). We can then apply Stockmeyer’s algorithm (Theorem 11) to the function f_C defined in Eq. (38). In time $\text{poly}(1/c)$ and within the third level Σ_3 of the polynomial hierarchy, the output of this procedure will produce a multiplicative-error estimate $q_0(C)$ of the output probability $p_0(C)$ that satisfies

$$p_0(C)c \leq q_0(C) \leq p_0(C)/c. \quad (55)$$

But since approximating $p_0(C)$ is a **GapP**-hard task by assumption, this implies that the polynomial hierarchy collapses to Σ_3 . ■

Notice two important subtleties of the argument: To prove exact sampling hardness, it is crucial that the output

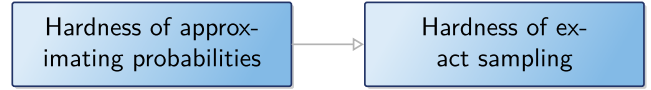


FIG. 4. In the proof of Theorem 15, the idea is to relate the hardness of approximating the probabilities in a distribution to the hardness of exactly sampling from that distribution.

probabilities are not only **GapP** hard to compute exactly but even to approximate up to some constant relative error; see Fig. 4. Meanwhile, it is sufficient for exact sampling hardness that there is no algorithm that efficiently computes all instances of the output probabilities. In other words, the argument relies on *worst-case hardness* of approximating the output probabilities since a single “hard instance” is sufficient for it.

What happens, though, once the sampling algorithm is allowed to make some error as compared to the ideal target distribution? Indeed, while an ideal quantum device samples from the ideal distribution, no such device can exist. Every physical realization of the ideal model, be it in terms of a classical simulation algorithm or a quantum implementation, will inevitably lead to errors so that it is able to only approximately sample from the target distribution. Such errors may be due to finite-precision issues intrinsic to computation or noise in the physical implementation of quantum random sampling using near-term quantum devices.

Does hardness of sampling still hold in the presence of errors on the sampled distribution? And if so, what types and magnitudes of errors are tolerated?

2. Multiplicative-error sampling hardness

As a first step, the proof of sampling hardness can be extended from exact sampling to sampling from a probability distribution p that is multiplicatively close to the target distribution $p(C)$ in the sense that for some constant $d \in (0, 1]$ each probability p_x satisfies

$$dp_x(C) \leq p_x \leq p_x(C)/d. \quad (56)$$

We can then easily amend the proof of Theorem 15 for this case to prove multiplicative-error robustness.

Multiplicative-error robustness of Theorem 15.—Assume that there is an efficient classical sampling algorithm \mathcal{A} that achieves the following task: Given as an input a description of a circuit $C \in \mathcal{C}$, produce a sample from a probability distribution p that approximates the distribution $p(C)$ defined in Eq. (1) up to a multiplicative error d as in Eq. (56). We can then use Stockmeyer’s algorithm to generate an approximation q_0 of the output probability p_0 that is correct up to any constant multiplicative error c :

$$cp_0 \leq q_0 \leq p_0/c. \quad (57)$$

But the probability p was multiplicatively close to the ideal output probability $p_0(C)$ to begin with, so we obtain

$$cdp_0(C) \leq cp_0 \leq q_0 \leq p_0/c \leq p_0(C)/(cd), \quad (58)$$

that is, an overall multiplicative-error approximation to the probability $p_0(C)$ with constant multiplicative error cd . If c and d are chosen such that the probability $p_0(C)$ is **GapP** hard to approximate up to multiplicative error cd , the existence of an efficient sampling algorithm with multiplicative-error guarantee cd implies the collapse of the polynomial hierarchy. ■

3. From multiplicative to additive errors

We saw in our discussion about the approximability of **GapP** how extraordinarily demanding multiplicative errors are in the guise of Lemma 7. There we used that such approximations always preserve the sign of a quantity and, moreover, attain 100% accuracy if the quantity is 0. Similarly, for the sampling task, there is no difference in complexity when one allows for constant multiplicative errors compared to the exact case. And indeed, to satisfy such a notion of approximation, an algorithm would need to account for the size of all of the exponentially many probabilities, some of which may be computer-precision close to zero to begin with. While this notion of approximation may be achievable using a fault-tolerant quantum device and a computation using ultra-high precision that scales with the system size, this state of affairs seems implausible in practice.

What is a more plausible notion of approximation then? In the following, we consider approximations q to a target distribution p in terms of the total-variation distance (TVD)

$$\|p - q\|_{\text{TV}} = \frac{1}{2} \sum_x |p(x) - q(x)| \quad (59)$$

between p and q . The TVD measures the maximal distinguishability of two probability distributions in terms of the optimal distinguishing strategy (Watrous, 2018) and is therefore a natural measure of statistical distance. But why is the TVD a sensible measure to consider when quantum advantage via quantum random sampling is considered? While the answer to this question is not entirely clear, there are several arguments that one might make.

a. Why the total-variation distance?

The first argument comes from the perspective of classical simulation algorithms. Indeed, a fundamental notion of imprecision of a randomized algorithm such as a sampling algorithm is given by additive errors. To see why, observe that a classical computer makes use of a constant precision representation of numbers. This gives rise to an additive error on all computations that is exponentially small in the number of digits of the representation. Going a step further, imperfections in an algorithm often give rise to additive errors on the result. One may therefore argue that the precision that is achievable by classical algorithms is fundamentally—and often in practice—simply an additive error, and the TVD is a natural way of capturing this error. At the same time, this line of reasoning implies that the precision of computing individual probabilities in the process of sampling needs to scale with the size of the system; see Sec. IV.A for details.

The second argument comes from the perspective of the noisy quantum device. This argument observes that any device error is reflected in an additive error on the distribution. To see this, we observe that for both coherent and incoherent errors we can write an erroneously prepared quantum state ρ_ϵ as a convex mixture of the target state $\rho = U|0\rangle\langle 0|U^\dagger$ and some other quantum state σ orthogonal to it as

$$\rho_\epsilon = (1 - \epsilon)U|0\rangle\langle 0|U^\dagger + \epsilon\sigma. \quad (60)$$

Consequently, the output distributions of the noisy and ideal states when measured in the standard basis $p(\rho)$ and $p(\rho_\epsilon)$ satisfy

$$\|p(\rho) - p(\rho_\epsilon)\|_{\text{TV}} = \frac{1}{2} \sum_x |p_x(\rho) - p_x(\rho_\epsilon)| \quad (61)$$

$$\leq \frac{1}{2} \max_{\{M_x\}_x} |\text{Tr}[(\rho - \rho_\epsilon)M_x]| \quad (62)$$

$$= \|\rho - \rho_\epsilon\|_{\text{Tr}} = \epsilon, \quad (63)$$

where the maximization runs over arbitrary positive operator-valued measures (POVMs) $\{M_x\}_x$. Here we have defined the trace distance $\|\cdot\|_{\text{Tr}}$, which is identical to the TVD for diagonal quantum states. The trace distance, analogously to the TVD, measures the maximal distinguishability of two quantum states in terms of the optimal quantum distinguishing strategy (Watrous, 2018). Since the trace distance maximizes over all possible measurement strategies, it upper bounds the TVD between the outcome distributions, which is given by fixing a measurement basis.

However, it is important to note that trace or total-variation distance are not good models of physically realistic errors occurring on a noisy quantum device. These errors (like the imprecision of classical computers) are independent of the size of the system. Hence, as constant local gate errors occur in a quantum circuit, the trace distance of its output state scales linearly in the number of gates, which quickly increases to a trivial value. To make the TVD meaningful from the perspective of a noisy quantum device, we thus need to scale down the local errors as we scale up the circuit size.

Finally, as we later see, the TVD arises naturally when one considers exact sampling algorithms that work only in the average case. Since average-case algorithms are natural for random quantum circuits, this provides further justification for the TVD. Compared to other statistical distances such as the Kullback-Leibler (KL) divergence, the TVD also turns out to be the measure that is amenable to the proof technique that we present in the following.

To summarize this discussion, the TVD is a notion of robustness for both classical and near-term quantum algorithms solving the sampling task. The smallest meaningful and nontrivial notion of approximation may be to consider the task of sampling up to constant TVD. This requires only relatively mild error or precision scalings of the individual components of the respective algorithms on the order of $1/m$. Such scaling of local algorithmic errors is already extremely demanding, however.

b. Showing TVD robustness

In the following, we consider the task of sampling from a distribution q that is ϵ close to a target output distribution $p(C)$ of a quantum circuit C in the sense that

$$\|p(C) - q\|_{\text{TV}} \leq \epsilon. \quad (64)$$

Our goal is to show that this task is hard for classical computers. Compared to exact and multiplicative-error sampling hardness, this endeavor is faced with the challenge that as ϵ increases, so does the legroom for classical simulation: to show hardness we have to prove that sampling from any distribution within an ϵ TVD neighborhood of the target distribution is classically intractable. We are faced with a dramatically increased burden in the proof as hardness needs to be shown for an entire volume of probability distributions rather than a single point. As $\epsilon \rightarrow 1$, the output state of the computation becomes classically simulable as the uniform distribution, in particular, is always within this error bound of the target distribution. But the uniform distribution is easy to sample from even an exponentially large sample space by repeated unbiased coin tosses.

Given what we have seen so far, there is a fundamental discrepancy between how the proof of exact sampling hardness can naturally be made robust to noise and the errors that inherently occur in realistic settings. The discrepancy is one between the utterly unrealistic notion of multiplicative errors on all probabilities and the more realistic notion of additive errors on the global outcome distribution. The question we now focus on is whether we can overcome this hurdle.

In technical terms, what we now prove is that no efficient classical algorithm \mathcal{A} taking as an input an efficient description of C exists that samples from any distribution q such that $\|q - p(C)\|_{\text{TV}} \leq \epsilon$ for a constant $\epsilon > 0$. Again we make use of Stockmeyer's approximate counting algorithm with a derandomizable sampling algorithm as an input in order to take the step from hardness of approximating probabilities. How can we take the leap from proving robust hardness-of-sampling results for multiplicative errors to those for additive errors?

To approach an answer to this question, we conceive of the sampling algorithm \mathcal{A} as an adversarial party that, given U as an input, tries to adversarially obstruct the approximate counting algorithm in its goal of approximating specific probabilities. The adversarially acting sampling algorithm is, however, constrained to sample from a distribution satisfying the respective error bounds. The following observations regarding the nature of additive errors in contrast to multiplicative ones are instructive.

- (1) When the sampling algorithm is constrained to multiplicative errors on individual probabilities, the total *additive error* it can make depends on the shape of the distribution. In particular, every individual probability will be correct up to an error that depends on its size. In contrast, the additive-error constraint allows the adversarial party much more flexibility. An additive error can be viewed as a total *error budget* that may be distributed across the individual probabilities at will. In particular, a few probabilities can come with large

relative errors supposing that the other ones are correct up to a small additive error.

- (2) When proving multiplicative-error robustness, the shape and volume of the region in the space of probability distributions on a sample space Ω of which hardness is proven depend heavily on the specific shape of the distribution. In contrast, for additive-error robustness the volume and shape of this region are sensitive only to boundaries of Ω .
- (3) Approximating output probabilities of quantum computations up to an inverse polynomial additive error does not remain hard for GapP but only for BQP; see Theorem 3 of De las Cuevas *et al.* (2011). Only for inverse, exponentially small additive errors $\pm 1/2^n$ do those approximations again become GapP hard. This is easily seen using the fact that normalized gaps of Boolean functions acting on $\{0, 1\}^n$ take on only values that are integer multiples of $2/2^n$. Approximating those gaps up to an additive error $< 1/2^n$ is therefore just as difficult as exactly computing them.¹⁴

What can we take away from these observations? Point (3) implies that to prove a polynomial-hierarchy collapse via Stockmeyer's algorithm we must still rely on the hardness of approximating output probabilities of circuit families up to relative errors or exponentially small additive errors.

Points (1) and (2) shine light on two sides of the same coin. In contrast to the case of multiplicative robustness, we cannot rely on the hardness of estimating individual probabilities that might be small. In particular, it cannot be the case that only one of the circuits within \mathcal{C} has a single output probability on which all classical algorithms fail. Instead, we must rely on circuit families for which not only are single outcome probabilities of some members of the family difficult to compute but also a large—constant—fraction of all output probabilities of the circuit family must be difficult to compute. This idea is formalized within the notion of *average-case complexity*: Approximating the outcome probabilities of quantum circuits must be difficult for a large fraction of the instances, where an instance is defined by a specific quantum circuit.

In particular, average-case complexity therefore requires that not all but instead few of those hard probabilities can be small, i.e., smaller than, say, doubly exponentially small while few large ones are easy to approximate. Indeed, if this were the case, since small quantities have small relative errors, the adversarially acting sampling algorithm could easily distribute the better part of its constant error budget on the few large probabilities while at the same time still passing the relative-error threshold on the small probabilities. In this way they would meet the constraint imposed by the global additive error but not achieve a provably difficult task, as the error on the computationally intractable probabilities would be too large. Rather, there must be a large fraction of difficult instances that are reasonably large, say, at least as large as uniform probabilities $\sim 1/|\Omega|$ on the sample space Ω . This idea is formalized within the notion of *anticoncentration*, which is a

¹⁴See also the Supplementary Material of Bremner, Montanaro, and Shepherd (2016).

condition on the probability that a randomly drawn problem instance (again specified by a circuit and an outcome string) is reasonably large. Anticoncentration constrains how the adversarial player can distribute their error budget: they can choose between getting many probabilities right with small errors, but making larger errors on a few outcomes, say, inverse polynomial errors on polynomially many probabilities, or getting all probabilities right with reasonably small inverse-exponential errors. These observations were made by Aaronson and Arkhipov (2013), who observed that the natural problem in boson sampling, namely, computing a permanent, is an average-case hard problem.

In the previous discussion, we touched on a point that we had glossed over in our discussion of exact sampling hardness: it is key to random circuit sampling schemes that there are two notions of probability at play. First, there is the random choice of a circuit from the family \mathcal{C} . Second, there is the random choice of an outcome string S that is distributed according to $p(U)$. Equally, there are two probability distributions: the distribution according to which random circuits are drawn and the outcome distribution of each such random circuit. These notions are crucially distinct.

As we later see, the choice of random circuit instances is essential to providing evidence for the additive-error robust hardness of simulating quantum circuits. The second notion of probability is intrinsic to our choice of problem. In the end, we aim to prove the hardness of a sampling task. This is a task requiring randomness: we want to obtain a random sample from a distribution that we in turn chose at random from another ensemble.

4. Additive-error sampling hardness

Given average-case hardness of approximating the output probabilities, we can prove a hardness-of-sampling result that is robust to constant additive errors. We proceed analogously to the proof of multiplicative robustness, following the sketch in Fig. 5.

Additive-error robustness of Theorem 15.—Assume that there is an efficient, derandomizable classical algorithm that takes as an input a description of a circuit instance C from a family \mathcal{C} and outputs samples distributed according to a probability distribution p that satisfies

$$\|p - p(C)\|_{\text{TV}} \leq \epsilon. \quad (65)$$

In Eq. (65) $p(C)$ is the ideal target distribution defined in Eq. (1). We use this sampling algorithm in order to approximate a random problem instance as given by the output probability $p_0(C) = |\langle 0|C|0\rangle|^2$ of C .

According to Task 1, random sampling, we generate an instance by drawing $C \in \mathcal{C}$ at random. To estimate the value of this instance, we use Stockmeyer's approximate counting algorithm with input given by the algorithm \mathcal{A} , the circuit instance C , and the outcome string 0^n . Using access to its NP oracle, Stockmeyer's algorithm will output a multiplicative-error approximation q_0 of the noisy output probability p_0 satisfying

$$|q_0 - p_0| \leq c p_0 \quad (66)$$

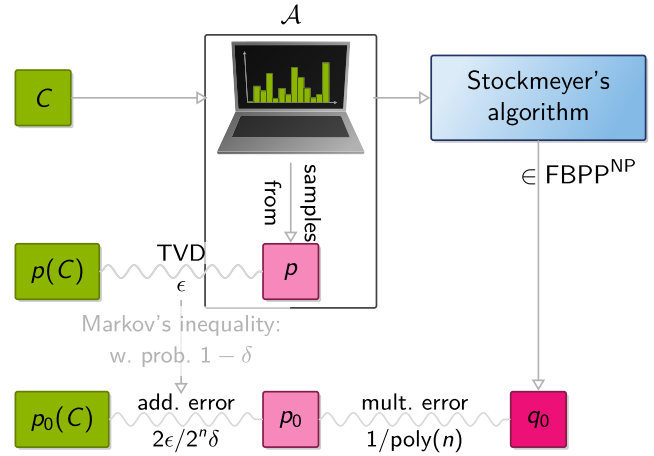


FIG. 5. Outline of the proof strategy for additive-error sampling hardness. A derandomizable sampling algorithm \mathcal{A} , given C as an input, samples from a distribution p that is ϵ close in total-variation distance (TVD) to the target distribution $p(C)$. Using Markov's inequality and the hiding property, this implies that the output probability p_0 of p is within additive error $2\epsilon/2^n\delta$ of the ideal output probability $p_0(C)$, with a probability of at least $1 - \delta$. Given \mathcal{A} as an input, Stockmeyer's algorithm can infer a $[1/\text{poly}(n)]$ -multiplicative approximation q_0 of the approximate output probability p_0 in the third level of the polynomial hierarchy.

in time $\text{poly}(n, 1/c)$ within the third level Σ_3 of the polynomial hierarchy.

Our goal is to bound the error

$$|q_0 - p_0(C)| \leq |q_0 - p_0| + |p_0 - p_0(C)|. \quad (67)$$

Equation (66) already provides the first half of this bound. For the second bound we need to leverage the total-variation-distance bound (65) on the global distributions p and $p(C)$ to an error bound on the individual probabilities p_0 and $p_0(C)$.

To obtain such a bound, consider again the sampling algorithm \mathcal{A} . Remember that *qua*, as a derandomizable algorithm on input U , r with a uniformly random $r \in \{0, 1\}^{\text{poly}(n)}$, will output a random sample from p such that

$$p_x(C) = \Pr[C \text{ outputs } x], \quad (68)$$

$$p_x = \Pr_r[\mathcal{A} \text{ outputs } x \text{ on input } C]. \quad (69)$$

Acting adversarially, the algorithm \mathcal{A} wants to maximize the error $|p_0 - p_0(C)|$. To do so, it needs to have some prior information about which of the outcome strings are more likely to be queried in Stockmeyer's algorithm given a certain input C so that it can distribute more of its constant error budget on those outcomes. This information would manifest itself in a distribution of outcomes x that is nonuniform (and in fact concentrated on the single all-zero outcome) from the perspective of \mathcal{A} given C (Aaronson and Arkhipov, 2013). This is because the all-zero outcome is always the one that we are interested in. But if it were able to distribute all of its constant error budget on this single outcome, then it would not

be able to achieve a difficult task, which is what we are trying to show.

a. Hiding problem instances

To see how we can achieve the result that this distribution over outcomes is not biased toward a few outcomes but rather uniform over all outcomes, consider the distribution over circuits C_y obtained by drawing $C \in \mathcal{C}$ at random and then appending X gates $X_1^{y_1} \cdot X_2^{y_2} \cdots X_n^{y_n}$ for uniformly random $y \in \{0, 1\}^n$ to the end of the circuit (Bremner, Montanaro, and Shepherd, 2016). We can then reexpress the outcome probabilities of C_y as

$$p_x(C_y) = |\langle x | C_y | 0 \rangle|^2 = |\langle 0 | C_{x \oplus y} | 0 \rangle|^2 = p_0(C_{x \oplus y}). \quad (70)$$

Consequently, the same problem instance C can equivalently be obtained when one provides the adversary \mathcal{A} with an instance C_y for uniformly random y and then queries Stockmeyer's algorithm on the outcome y . When aiming to estimate the problem instance $p_0(C)$, we can therefore hide the instance C in the circuit C_y by randomly appending X gates according to a uniformly random y and then querying Stockmeyer's algorithm on outcome y . But since y is hidden from \mathcal{A} , the distribution over outcomes on which we are going to query Stockmeyer's algorithm to obtain the output probability is uniformly random, and it cannot bias its error toward any given outcome.

For this to work, it is crucial that \mathcal{A} cannot distinguish whether we have directly generated a random problem instance C for which we are directly interested in the all-zero outcome, or whether we have first drawn a random $C \in \mathcal{C}$ and then hidden this instance by constructing the unitary C_y with uniformly random y and query on the outcome y (Aaronson and Arkhipov, 2013). Hence, the probability of directly drawing C_y must be the same as that of drawing C and then appending uniformly random X gates according to y .

Generally, we therefore say that a circuit family \mathcal{C} has the *hiding property* if (a) there is an efficient instance-generating procedure that converts a given problem instance $C \in \mathcal{C}$ and a uniformly random outcome y into another problem instance C_y , and (b) the distribution over circuits is invariant under this procedure, i.e.,

$$\Pr_{C_y \sim \mathcal{C}} [C_y] = \Pr_{C \sim \mathcal{C}, y \sim \{0,1\}^n} [C_y]. \quad (71)$$

The hiding property holds naturally for most random circuit families, and, in particular, also for universal random circuits where each gate is drawn from the Haar measure. This is because the Haar measure is left and right invariant under arbitrary unitaries and the Pauli- X gate is one such unitary.

If the hiding property holds, without loss of generality we can therefore always query Stockmeyer's algorithm on the all-zero outcome of C , making use of the fact that this outcome is indistinguishable from a uniformly random one from the perspective of \mathcal{A} . Conversely, we can conceive of the outcomes of the circuits that we query Stockmeyer's algorithm on as being uniformly distributed from the perspective of \mathcal{A} . In this case, we can apply Markov's inequality to obtain a bound

on the error for individual probabilities. For uniformly random x we obtain that

$$\begin{aligned} \Pr_{x \in \{0,1\}^n} \left[|p_x - p_x(C)| \geq \frac{1}{\delta} \mathbb{E}_{x \in \{0,1\}^n} [|p_x - p_x(C)|] \right] \\ = \Pr_{x \in \{0,1\}^n} \left[|p_x - p_x(C)| \leq \frac{2\epsilon}{\delta 2^n} \right] \leq \delta \end{aligned} \quad (72)$$

since

$$\begin{aligned} \mathbb{E}_{x \in \{0,1\}^n} [|p_x - p_x(C)|] &= \frac{1}{2^n} \sum_{x \in \{0,1\}^n} |p_x - p_x(C)| \\ &= \frac{2}{2^n} \|P - P(C)\|_{\text{TV}} = \frac{2\epsilon}{2^n}. \end{aligned} \quad (73)$$

Combining Eqs. (67) and (72), we now find that, with a probability of at least $1 - \delta$ over the inputs, the error of the estimate q_0 output by Stockmeyer's approximate counting algorithm satisfies

$$|q_0 - p_0(C)| \leq \frac{1}{\text{poly}(n)} p_0 + \frac{2\epsilon}{2^n \delta} \quad (74)$$

$$\leq \frac{1}{\text{poly}(n)} p_0(C) + \frac{2\epsilon}{2^n \delta} \left(1 + \frac{1}{\text{poly}(n)} \right). \quad (75)$$

The bound in Eqs. (74) and (75) is a mixture of an exponentially small additive and inverse polynomially small multiplicative error. However, the error bound does not hold for all possible inputs to Stockmeyer's algorithm; it holds only for a $1 - \delta$ fraction of the inputs. By the hiding property this corresponds to a $1 - \delta$ fraction of the problem instances.

b. Approximate average-case hardness

To show the hardness of the sampling task, we need to show that achieving this error on an arbitrary $1 - \delta$ fraction of the outputs is sufficient for a collapse of the polynomial hierarchy.¹⁵ Indeed, our procedure involving Stockmeyer's algorithm is precisely such an algorithm (in the third level of the polynomial hierarchy). A sufficient condition to show such a polynomial-hierarchy collapse is then the following: The problem of estimating the probabilities remains **GapP** hard even when using a polynomial-time algorithm that succeeds on only a constant fraction of the instances. In other words, an algorithm solving the estimation problem for $p_0(C)$ with error (75) and the success probability given by the respective fraction of the instances (i.e., $1 - \delta$) is as powerful as an arbitrary **GapP** algorithm. This contrasts with the proof of exact sampling, where it was merely required that the estimation problem is **GapP** hard in the worst case, that is, for a machine that is required to succeed on all instances.

Making this intuition rigorous is the idea of average-case hardness.

Definition 16 (Approximate average-case hardness).—Let $\Gamma \in (0, 1)$ and $\epsilon > 0$. A function class \mathbf{F} is average-case

¹⁵This is because Markov's inequality does not control the instances on which the bound fails.

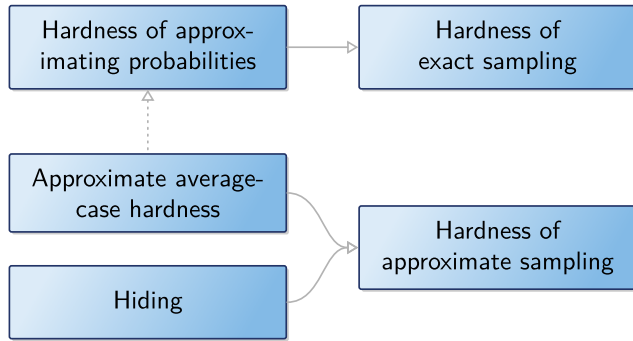


FIG. 6. While in the proof of exact sampling hardness it was sufficient to build on the hardness of approximating the output probabilities of quantum circuits, in order to prove hardness of approximate sampling further properties of the circuit family \mathcal{C} are required: approximate average-case hardness of computing the output probabilities and the hiding property.

hard with constant Γ and error ε if approximating any Γ fraction of the instances in F up to error ε is GapP hard.

If approximate average-case hardness holds with respect to the error (75), the existence of an efficient sampling algorithm \mathcal{A} for the output distribution of a random instance $C \in \mathcal{C}$ implies that we can approximate GapP -hard probabilities in the third level of the polynomial hierarchy using Stockmeyer’s algorithm. The polynomial hierarchy collapses.

We have proven approximate sampling hardness; see Fig. 6.

Theorem 17 (Additively robust sampling hardness).— Consider a circuit family \mathcal{C} that satisfies (1) the hiding property and (2) approximate average-case hardness up to error (75) on any $1 - \delta$ fraction of the instances. Suppose that there is an efficient classical sampling algorithm \mathcal{A} that, given $C \in \mathcal{C}$ drawn at random as an input with success probability at least $1 - \delta$ over \mathcal{C} , outputs samples from an additive approximation p to the outcome distribution $p(C)$ satisfying $\|p - p(C)\|_{\text{TV}} \leq \varepsilon$. The polynomial hierarchy then collapses.

We have taken a long route from the complexity-theoretic foundations of quantum speedups to rigorous and approximate hardness-of-sampling arguments relevant to near-term quantum technology. The complexity-theoretic foundations of quantum speedups manifested themselves in the GapP vs $\#\text{P}$ dichotomy: while multiplicatively approximating the acceptance probabilities of classical circuits can be done on the third level of the polynomial hierarchy, this task remains GapP complete for certain quantum circuit families. We then saw how the at-first-sight different tasks of sampling from a probability distribution (weakly simulating it) and approximating its outcome probabilities (strongly simulating it) are related on a rigorous level: Stockmeyer’s approximate counting algorithm and the concept of the polynomial hierarchy proved essential to this question. Building on those methods, we could show that the task of sampling from the output distribution of certain random quantum computations cannot be achieved using an efficient classical algorithm. In a last step, we aimed to make this result robust to realistic errors, that is, additive errors in total-variation distance on the level of the output distributions. Making this leap involved stronger properties of the output distribution, however: approximate

average-case hardness and the hiding property. The way that we have formulated Theorem 17 provides a general framework for providing a hardness argument for approximately sampling from the output distributions of quantum circuit families. But in order to complete the proof the two properties (hiding and approximate average-case hardness) need to be shown for specific circuit families.

We hinted that the hiding property trivially holds for most circuit families: to show this, we merely need to show that X gates at the end of the circuit do not alter the circuit family. The only instances of circuit families for which hiding is nontrivial are boson-sampling protocols. We sketch the argument here.

c. Hiding in boson sampling

We saw in Sec. II that the output probabilities of Fock boson sampling are given by permanents (45) of submatrices of Haar-random unitaries. Conceivably, though, there is some structure in such submatrices. To see this, consider the case in which we obtain all bosons in a single mode as the outcome, i.e., $S = (n, 0, 0, \dots)$. In this case, all columns of the submatrix $U_{S,1_n}$ are equal, and this can plausibly be exploited to approximate $|\text{Perm}(U_{S,1_n})|^2$. In other words, because of the structure in the matrix, the specific outcome cannot be hidden. However, Aaronson and Arkhipov (2013) showed that under certain conditions hiding holds in Fock boson sampling in virtue of the fact that the output probabilities of a random boson-sampling instance are determined by permanents of approximately Gauss-random, and therefore highly unstructured, matrices.

To achieve this, Aaronson and Arkhipov (2013) considered the collision-free boson-sampling distribution $P_{\text{bs},U}^*$. The distribution $P_{\text{bs},U}^*$ is obtained from $P_{\text{bs},U}$ by discarding all output sequences S with more than one boson per mode, i.e., all S that are not in the set of collision-free sequences

$$\Phi_{m,n}^* = \{S \in \Phi_{m,n} : \forall s \in S : s \in \{0, 1\}\}. \quad (76)$$

Why are collision-free outcomes advantageous when proving hardness? Intuitively, this is because for collision-free outcomes the submatrix $U_{S,1_n}$ has much less structure than for outcomes with collisions because there are no repeated rows or columns. If, moreover, the size of $U_{S,1_n}$ becomes sufficiently small compared to the full size of U , neither does there remain any of the structure in U stemming from the orthogonality of its columns.

The hiding property then follows from two facts. First, we need to justify that restricting our attention to collision-free outcomes is valid. This is true if postselecting onto the collision-free subspace can be done efficiently, in the sense that its probability weight is at least a constant, and Aaronson and Arkhipov proved that this is the case if m grows sufficiently fast with n and at least as $m \in \Omega(n^2)$; see Jiang (2006), Arkhipov and Kuperberg (2012), and Theorem 13.4 of Aaronson and Arkhipov (2013). Second, Aaronson and Arkhipov proved that if m grows even faster, namely, as $m \in \Omega[n^5 \log(n)^2]$, the measure induced on $U \sim \mu_H$ 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 a standard deviation $\sigma = 1/\sqrt{m}$ on $n \times n$ matrices. Consequently, regardless of which submatrix we choose, i.e., which collision-free outcome we obtain, the distribution of the submatrices is approximately Gaussian.

Conversely, Aaronson and Arkhipov (2013) proved in Lemma 5.8 that, given a Gauss-random instance $X \sim \mu_G(\sigma)$ as input, there is a BPP^{NP} algorithm¹⁶ that, given X , hides this matrix in a large unitary matrix in the sense that it generates a Haar-random $U \in U(m)$ such that there is a uniformly random $S \in \Phi_{m,n}^*$ such that $X = U_{S,1_n}$. This provides the instance-generating algorithm. Hiding a Gauss-random instance X is therefore possible when constructing a larger unitary matrix of which X is a uniformly random submatrix, similar to how we hid a qubit circuit C by appending uniformly random X gates to it.

A similar reasoning can be applied to Gaussian boson sampling, albeit with a slightly different distribution (Deshpande *et al.*, 2022). Recall that the matrices of which the Hafnian is computed in Gaussian boson sampling with k single-mode squeezed inputs and n detected photons in m modes are of the form $U_{S,1_k} U_{S,1_k}^T$, which for collision-free outcomes are outer products of random $n \times k$ submatrices of the linear-optical unitary U . For those matrices, hiding plausibly holds with respect to symmetric Gaussian matrices XX^T , where $X \sim \mathcal{G}_{n,k}(0, 1/m)$ is an $n \times k$ matrix drawn from the Gaussian distribution on $n \times k$ complex matrices. Indeed, this is provably true in two regimes (Deshpande *et al.*, 2022): First, for $m \in O(k^5 \log^2 k)$ and $k = n$ the submatrices are individually Gaussian distributed by the result of Aaronson and Arkhipov (2013), and hence we can also bound the distance to the distribution of XX^T . Second, for $k = m$ Jiang (2009) showed that whenever $n \in o(\sqrt{m}/\log m)$ the distribution of $n \times n$ submatrices of UU^T for unitary U converges asymptotically to the distribution of XX^T , where $X \sim \mathcal{G}_{n,m}(0, 1/m)$ is an $n \times m$ complex Gaussian matrix. For the intermediate regime $m^{1/5} < k < m$, there is numerical evidence that the hiding property remains true (Deshpande *et al.*, 2022). The instance-generating algorithm of Aaronson and Arkhipov (2013) in their Lemma 5.8 will also work for this setting provided that the distributions of $U_{S,1_k} U_{S,1_k}^T$ for unitary U and XX^T for Gaussian $X \sim \mathcal{G}_{n,k}(0, 1/m)$ are close not only in TVD but also in a slightly stronger multiplicative sense. This is because the instance-generating algorithm simply postselects on the matrix XX^T , appearing as a submatrix of $U1_k U^T$ by making use of the NP oracle.

An efficient way of constructing a Gaussian boson-sampling scheme that comes without needing to scale $m \in \Omega[\text{poly}(n)]$ was discovered by Grier *et al.* (2022). They observed that by programming a Gaussian boson-sampling device in a bespoke way, it is possible to encode the permanent of an arbitrary matrix in the output probabilities. Specifically, they considered a bipartite system of $2m$ modes. The input state is given by a product of two-mode squeezed states on

modes i and $i + m$ for $i = 1, \dots, m$ with squeezing parameters r_1, \dots, r_m . In other words, the two halves of a two-mode squeezed state are associated with the two partitions. A bipartite unitary mode transformation $U \otimes V$ is then applied to the system, and all modes are measured in the Fock basis. This gives rise to output probabilities that are proportional to a function of a submatrix of a matrix $C = U \text{diag}(r) V^\dagger$, where $r = (r_1, \dots, r_m)$ is the vector of squeezing values. Since this is simply a singular-value decomposition, by choosing r , U , and V bespoke, C can be programmed to be an arbitrary matrix, and, in particular, a Gaussian one that satisfies the hiding property by definition.

Proving approximate average-case hardness is an entirely different story, however, and remains the central open theory problem in the context of quantum random sampling. However, much work has been put into gathering evidence for the truth of approximate average-case hardness. In Sec. IV.D, we discuss this evidence.

D. Approximate average-case hardness

To prove approximate average-case hardness, it is helpful to simplify the rather baroque error mixture (75) on any $1 - \delta$ fraction to something more familiar: an exponentially small additive or a constant multiplicative error. Indeed, for those errors we already know the worst-case hardness of approximating the output probabilities, and hence a necessary condition is true.

1. Reduction to additive or multiplicative average-case hardness

To achieve our goal, we begin by observing that, depending on which one of the two terms in Eq. (75) is larger, the error will be a relatively or exponentially small additive, respectively. Hence, if we are able to determine the comparative size of the two terms, we can reduce the error to a simpler form. Specifically, if in the error bound (75) the probability $p_0(C)$ is smaller than $\alpha/2^n$ for some constant $\alpha > 0$, then Eq. (75) can be upper bounded in terms of an additive error $[2\epsilon/\delta + \alpha + o(1)]/2^n$ if it is larger than $\alpha/2^n$. Equation (75) can therefore be upper bounded in terms of a relative error $2\epsilon/(\alpha\delta) + o(1)$.

To reduce the error (75) to an exponentially small additive error, we can make use of the concentration of the probabilities around their mean (given by $1/2^n$) using Markov's inequality

$$\Pr_{C \sim \mathcal{C}} \left[p_0(C) \geq \frac{1}{2^n \alpha} \right] \leq \alpha, \quad (77)$$

where the probability is taken over the choice of problem instances. Since the probability in Eq. (77) runs over the choice of random circuit, while in Eq. (72) it runs only over the uniformly random choice of outcome, the failure probabilities are independent of one another. Hence, both bounds are satisfied with probability $(1 - \delta)(1 - \alpha)$, in which case Eq. (75) is upper bounded by an exponentially small additive error

$$|q_0 - p_0(C)| \leq \left[\frac{2\epsilon}{\delta} + \frac{1}{\text{poly}(n)} \left(1 + \frac{1}{\alpha} \right) \right] \frac{1}{2^n}. \quad (78)$$

¹⁶Like Stockmeyer's algorithm, this algorithm is therefore on the third level of PH.

To reduce the error (75) to the arguably more “natural” case (Aaronson and Arkhipov, 2013) of a constant relative-error approximation, we invoke the so-called anticoncentration property introduced by Aaronson and Arkhipov (2013).

Definition 18 (Anticoncentration).—We say that a circuit family \mathcal{C} anticoncentrates if for a constant $\alpha > 0$ there is a $\gamma(\alpha) > 0$ independent of n such that

$$\Pr_{C \sim \mathcal{C}} \left[p_0(C) \geq \frac{\alpha}{2^n} \right] \geq \gamma(\alpha). \quad (79)$$

Since the failure probabilities δ and $\gamma(\alpha)$ are independent, bounds (72) and (79) are both satisfied with a probability of at least $\gamma(\alpha)(1 - \delta)$, in which case we obtain the relative-error bound

$$|q_0 - p_0(C)| \leq \left(\frac{2\epsilon}{\delta\alpha} + \frac{1}{\text{poly}(n)} \right) p_0(C). \quad (80)$$

For the relative-error case, we can set $\alpha = 1/c$, $\epsilon = \gamma(\alpha)/4$, and $\delta = \gamma(\alpha)/2$ to obtain a $c/2 + o(1)$ relative-error approximation of $p_0(C)$ with a probability at least $\gamma(1 - \gamma/2)$ over the choice of instances. For the additive-error case, we can set $2\epsilon/\delta = \kappa/2$ and α constant to obtain a $[\kappa/2 + o(1)]/2^n$ additive approximation of $p_0(C)$ with a probability at least $4\epsilon\alpha/\kappa$ over the choice of instances.

We have reduced approximate average-case hardness (condition 2 of Theorem 17) to either (2a) additive approximate average-case hardness up to an exponentially small additive error $O(2^{-n})$ on any γ fraction or (2b) relative approximate average-case hardness up to a relative error $1/4$ on any $\gamma(1 - \gamma/2)$ fraction and (2c) anticoncentration for $\alpha = 1$ with constant $\gamma = \gamma(\alpha)$.

To date no proof of additive or relative approximate average-case hardness exists. But to see why a multiplicative-error average-case hardness conjecture is plausibly true for GapP functions, again consider the previous argument. For typical #P functions the number of accepting paths is exponentially large, and hence a multiplicative error is also of the same order of magnitude. In contrast, for typical GapP functions, as differences of #P functions, the number of accepting paths is a difference between two exponentially large numbers, which is often orders of magnitude smaller than each such number. This is why for #P functions we often do not expect approximate average-case hardness, while for GapP functions this conjecture seems reasonable.

Another argument in favor of approximate average-case hardness makes use of a universal quantity such as the Ising partition function (40) (Bremner, Jozsa, and Shepherd, 2010; Bremner, Montanaro, and Shepherd, 2016; Goldberg and Guo, 2017; Boixo *et al.*, 2018), the Tutte partition function (Goldberg and Guo, 2017), or the Jones polynomial (Kuperberg, 2015; Mann and Bremner, 2017). This argument observes that as we draw random instances of an Ising partition function Z_W no additional structure is present, unlike a worst-case instance, which a hypothetical approximation algorithm might be able to exploit.

While one might argue that these arguments are relatively weak, there have also not been counterexamples to

approximate average-case hardness in the standard settings. In the following, we see further and more substantial technical evidence of the additive average-case hardness conjecture.

2. Anticoncentration

We begin with the anticoncentration property (Definition 18). The anticoncentration property allows us to reduce the baroque error (75) to a relative error, arguably the most natural error if we want to prove the hardness of approximating the probabilities because GapP naturally allows one to reduce relative errors to an exact computation. But anticoncentration can also serve as evidence for the additive approximate average-case hardness property to hold. By ruling out that almost all outcome probabilities are less than inverse exponentially small, anticoncentration rules out that an inverse-exponential additive-error approximation is trivial: we cannot simply guess 0 for all probabilities and be almost always correct if anticoncentration holds.

In this sense, a certain degree of anticoncentration is required to hold for approximate average-case hardness to be true. Note, however, that anticoncentration is not a necessary property for hardness of sampling to hold, and neither is approximate average-case hardness. Both properties are merely used in the proof strategy that we describe in this section. But while approximate average-case hardness is sufficient for approximate hardness of sampling, anticoncentration is not.

Notice that to prove anticoncentration we merely need to derive statistical properties of the respective random circuit families. To see this, we make use of the Paley-Zygmund inequality (Bremner, Montanaro, and Shepherd, 2016), a lower-bound analog to Markov’s inequality, which states that, for a random variable Z with $0 \leq Z \leq 1$,

$$\Pr [Z > \alpha \mathbb{E}[Z]] \geq (1 - \alpha)^2 \frac{\mathbb{E}[Z]^2}{\mathbb{E}[Z^2]}. \quad (81)$$

Using the Paley-Zygmund inequality we can therefore reduce the anticoncentration property to the value of the second moments of the random circuit ensemble as

$$\Pr \left[p_0(C) > \frac{\alpha}{2^n} \right] \geq (1 - \alpha)^2 \frac{2^{-2n}}{\mathbb{E}[p_0(C)^2]}. \quad (82)$$

The normalized second moment $2^n \mathbb{E}[p_0(C)^2]$ is also often referred to as the average collision probability.¹⁷ To prove anticoncentration for quantum random sampling, it is therefore sufficient to bound this average collision probability as $O(2^{-n})$.

This scaling of the average collision probability as $O(d^{-1})$ for quantum states in dimension d is precisely the scaling that one obtains when drawing a quantum state $|\psi\rangle$ uniformly at random on the complex unit sphere $S(\mathbb{C}^d)$ and measuring it in the computational basis. Equivalently, we can draw a unitary $U \sim U(d)$ uniformly at random and apply it to a reference state as $U|0\rangle$, giving rise to a uniformly distributed quantum state. The corresponding uniform measure $\mathcal{U}_{S(\mathbb{C}^d)}$ on the unit

¹⁷The collision probability of a distribution p is given by $\sum_x p(x)^2$.

sphere is therefore invariant under the action of unitaries $U(d)$. For this measure, we can compute the k th moment projector as

$$M^k = \int_{S(\mathbb{C}^d)} (|\psi\rangle\langle\psi|^{\otimes k}) d\mathcal{U}_{S(\mathbb{C}^d)}(\psi) = \frac{P_{[k]}}{D_{[k]}}, \quad (83)$$

where $P_{[k]}$ is the projector on the symmetric subspace of k tensor copies and $D_{[k]} = \binom{d+k-1}{k}$ is the dimension of that subspace. See [Kliesch and Roth \(2021\)](#) for a pedagogical introduction to random unitaries and states.

For uniformly random quantum states we can now compute the second moments of the output probabilities $|\langle x|\psi\rangle|^2$ as

$$\mathbb{E}[|\langle x|\psi\rangle|^4] = \mathbb{E}[\langle x|^{\otimes 2}(|\psi\rangle\langle\psi|)^{\otimes 2}|x\rangle^{\otimes 2}] \quad (84)$$

$$= \langle x|^{\otimes 2} \mathbb{E}[(|\psi\rangle\langle\psi|)^{\otimes 2}] |x\rangle^{\otimes 2} \quad (85)$$

$$= D_{[2]}^{-1} \langle x|^{\otimes 2} P_{[2]} |x\rangle^{\otimes 2} \quad (86)$$

$$= D_{[2]}^{-1} = \frac{2}{(d+1)d}, \quad (87)$$

where we specify that $|x\rangle^{\otimes 2}$ is in the symmetric subspace such that the projector $P_{[2]} = (\mathbb{1} + \mathbb{S})/2$ with swap operator

$$\mathbb{S} = \sum_{i,j=1}^d |i\rangle\langle j| \langle j| \langle i| \quad (88)$$

acts trivially on it.

For uniformly random quantum states, we therefore obtain from Eq. (82) that the anticoncentration property holds with a success probability of at least $(1 - \alpha^2)/2$. Proving anticoncentration of quantum circuit families can therefore be viewed as proving that the output probabilities of these families behave up to constant factors just like the output probabilities of uniformly random quantum states in terms of their average collision probability, their second moment. To prove bounds on the average collision probability, one can now proceed in various different ways. One can directly bound the average collision probability, or one can show that the output states of circuits drawn from the family already behave sufficiently similarly to uniformly random states. We now sketch the two most important ways via which anticoncentration can be proven for random quantum circuits: the so-called design property and statistical-mechanics mappings.

a. Anticoncentration via spherical designs

While the circuit families proposed for quantum random sampling do not generate a uniformly random quantum state $C|0\rangle$, several families have the strong property that they mimic uniform randomness at the level of the second moment. A family of vectors $\Psi = \{|\psi_i\rangle\}_i$ that mimics uniform randomness for the k th moments in the sense that

$$M_{\Psi}^k = \frac{1}{|\Psi|} \sum_i (|\psi_i\rangle\langle\psi_i|)^{\otimes k} = \frac{P_{[k]}}{D_{[k]}} \quad (89)$$

forms a so-called complex (spherical) k -design. We can slightly relax the notion of a k -design to approximations thereof and say that a family Ψ is a relative ϵ -approximate k -design if

$$(1 - \epsilon)M_{\Psi}^k \leq \frac{P_{[k]}}{D_{[k]}} \leq (1 + \epsilon)M_{\Psi}^k. \quad (90)$$

The proof of the following theorem then directly follows [Hangleiter et al. \(2018\)](#).

Lemma 19 (Anticoncentration of 2-designs).—Let Ψ be a relative ϵ -approximate 2-design on $S(\mathbb{C}^d)$. The output probabilities $|\langle 0|\psi\rangle|^2$ of a randomly chosen $|\psi\rangle \in \Psi$ then anticoncentrate in the sense that, for $0 \leq \alpha \leq 1$,

$$\Pr_{|\psi\rangle \sim \Psi} \left(|\langle 0|\psi\rangle|^2 > \frac{\alpha(1 - \epsilon)}{d} \right) \geq \frac{(1 - \alpha)^2(1 - \epsilon)^2}{2(1 + \epsilon)}. \quad (91)$$

Several circuit families considered for quantum random sampling approximately exhibit the 2-design property when applied to a reference state. This holds, in particular, for universal random circuits in various settings. For random circuits, one can even prove a stronger property, namely, that they are *unitary designs*, mimicking uniform randomness on the unitary group as opposed to the complex sphere. Unitary designs by definition have the property that their columns form spherical designs; hence, Lemma 19 applies to them. Historically, the first proof of the 2-design property for random circuits was from [Harrow and Low \(2009\)](#), albeit for a weaker (additive) notion of approximation than is required for the proof of anticoncentration.

[Brandão, Harrow, and Horodecki \(2016\)](#) proved the stronger result that random circuits on n qubits arranged in a linear chain form an ϵ -approximate unitary k -design if they contain $O\{\text{poly}(k)n[n + \log(1/\epsilon)]\}$ many gates. The circuits that they considered are composed of two-qubit gates that are applied either to random neighboring qubits or in an alternating parallel “brickwork” configuration. The individual gates may be drawn either from a universal gate set containing its own inverses or uniformly (Haar) randomly. The key idea of the proof of [Brandão, Harrow, and Horodecki \(2016\)](#) was to map the design property to the gap of a local, frustration-free Hamiltonian, the local terms of which correspond to the individual two-qubit gates of the circuit and act on $4k$ many qubits, using the so-called detectability lemma ([Aharonov et al., 2009](#); [Anshu, Arad, and Vidick, 2016](#)). The gap of this Hamiltonian can then be bounded using the seminal result of [Nachtergaele \(1996\)](#). [Haferkamp \(2022\)](#) recently improved this result by showing a milder polynomial dependence in k , thereby providing an improved bound on the spectral gap. The same technique can also be applied to show the design property for other circuit families that encode universal quantum circuits, such as random measurement-based quantum computations ([Haferkamp, Hangleiter, Bouland et al., 2020](#)).

Further examples of postselected-universal circuit families that exhibit the 2-design property, and therefore anticoncentration, are conjugated Clifford circuits ([Bouland, Fitzsimons, and Koh, 2018](#)), Clifford circuits with magic-state inputs ([Hangleiter et al., 2018](#); [Yoganathan, Jozsa, and Strelchuk,](#)

2019), and diagonal quantum circuits applied to the state $|+\rangle^{\otimes n}$ (Nakata, Koashi, and Murao, 2014; Hangleiter *et al.*, 2018).

Improving the result of Brandão, Harrow, and Horodecki (2016) to lattices of arbitrary dimension, Harrow and Mehraban (2023) proved that random universal circuits arranged on a lattice of dimension D generate an approximate k -design using $\text{poly}(k)n^{1+1/D}$ many gates. This result reflects the intuition that, due to the fact that correlations in a parallel brickwork circuit spread ballistically, sufficiently random quantum states can arise only in a depth that scales linearly with the diameter of the system, and hence as $n^{1/D}$.

b. Anticoncentration via computing the collision probability

While this intuition is presumably true for the design property of random circuits, it was recently proven that anticoncentration already arises in logarithmic depth for nearest-neighbor random circuits in one dimension with uniformly random two-qubit gates (Barak, Chou, and Gao, 2021). To prove this result, Barak, Chou, and Gao (2021) directly bounded the average collision probability, that is, the second moment $2^n \mathbb{E}[p_0(C)^2]$, using a mapping to a statistical-mechanics model of Zhou and Nahum (2019). Dalzell, Hunter-Jones, and Brandão (2022) showed that this result is tight by complementing it with an $O(n \log(n))$ lower bound on the circuit size that holds for arbitrary geometries. For architectures with arbitrary connectivity, they further showed that $5n \log(n)/6$ many gates are necessary and sufficient (up to subleading corrections) for exponentially small collision probabilities. This in fact also holds directly for the anticoncentration property (Deshpande, Niroula *et al.*, 2022).

We now sketch the idea of these proofs, following Hunter-Jones (2019). The idea is to again exploit the properties of the moment operator, albeit now at the level of the individual quantum gates in the random circuit. For uniformly Haar-random unitaries, we can, analogously to Eq. (83), define a moment operator M_H^k on $U(d)$. This moment operator is characterized by so-called Weingarten functions (Wg) as (Brouwer and Beenakker, 1996; Hunter-Jones, 2019)

$$M_H^k = \mathbb{E}_{U \sim \mu_H} [U^{\otimes k} \otimes \bar{U}^{\otimes k}] = \sum_{\sigma, \pi \in S_k} \text{Wg}(\sigma^{-1}\pi, d) |\sigma\rangle\langle\pi|. \quad (92)$$

In Eq. (92) $|\sigma\rangle = [1 \otimes r(\sigma)]|\Omega\rangle$, where r is the representation of the symmetric group S_k on $(\mathbb{C}^d)^{\otimes k}$ that permutes the vectors in the tensor product and $|\Omega\rangle = \sum_{j=1}^d |j\rangle|j\rangle$ is the maximally entangled state up to normalization. To evaluate formulas involving the moment operator (92), it is useful to develop a graphical language for the moment operator. In this language, we can express the identity and the swap operator on two tensor copies, as well as the corresponding maximally entangled state, as rewirings of single-copy identities as follows¹⁸:

$$\mathbb{1} = \text{---}, \quad \mathbb{S} = \text{---}, \quad |\Omega\rangle = \text{---}, \quad \langle\Omega| = \text{---}. \quad (93)$$

¹⁸See Bridgeman and Chubb (2017) for an introduction to the graphical representation.

Hence, we can write

$$|\mathbb{S}\rangle\langle\mathbb{1}| = \text{---} \text{---} \text{---}. \quad (94)$$

For quantum circuits composed of Haar-random two-qubit unitaries, we can now evaluate the expectation value locally, and the global moment operator is given by

$$\mathbb{E}_{U_1, \dots, U_m \sim U(4)} [U^{\otimes 2} \otimes \bar{U}^{\otimes 2}] = \prod_{i=1}^m \left(\mathbb{E}_{U_i \sim U(4)} [U_i^{\otimes 2} \otimes \bar{U}_i^{\otimes 2}] \right), \quad (95)$$

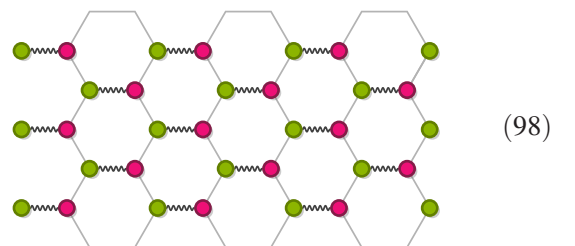
where $U = \prod_i U_i$. In an abuse of notation, we take the expectation over the individual quantum gates at their respective locations in the quantum circuit. Using Weingarten calculus, we can now evaluate the Weingarten formula for $k = 2$, obtaining the result in graphical representation as

$$\mathbb{E}_{U \sim U(d)} \left[\text{---} \right] = \frac{1}{d^2 - 1} \left[\text{---} \right]. \quad (96)$$

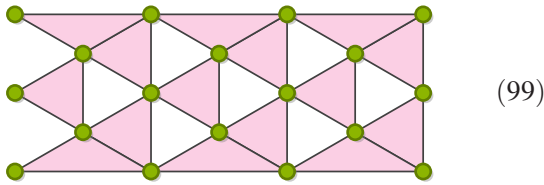
We can view the expectation value of a single two-qubit gate as an effective vertex

$$\mathbb{E}_{U \sim U(d)} \left[\text{---} \right] \rightarrow \text{---}, \quad (97)$$

where the vertices can take one of two values $\mathbb{1}$ or \mathbb{S} (corresponding to a spin up or down) that tell us how to contract each of the incoming or outgoing edges, and the curly edge between the vertices corresponds to a weight that is given by $-1/d/(d^2 - 1)$ for the configurations $\langle\mathbb{S}|\mathbb{1}\rangle$ and $\langle\mathbb{1}|\mathbb{S}\rangle$, and by $1/(d^2 - 1)$ otherwise. The contractions themselves will pick up different values; for example, for a single contraction we obtain $\langle\mathbb{S}|\mathbb{1}\rangle = \langle\mathbb{1}|\mathbb{S}\rangle = d$ and $\langle\mathbb{1}|\mathbb{1}\rangle = \langle\mathbb{S}|\mathbb{S}\rangle = d^2$. Computing the second moment $\mathbb{E}_U |\langle x|U|0\rangle|^4$ now corresponds to computing a partition function over all local ‘‘spin’’ (also known as permutation) configurations, with the corresponding weights and boundary conditions determined by $|x\rangle$ and $|0\rangle$:



One can now sum over the pink vertices, giving rise to a new statistical mechanical model. This model is defined by terms acting on the plaquettes of a triangular lattice:



The plaquette terms are now just functions of permutations of the local spins with dimension k , which are nonzero only if the product of the permutations on a plaquette is the identity. For $k = 2$, this allows one to perform simple domain-wall counting arguments in order to bound the value of the average collision probability.

c. Further proofs of anticoncentration

An example of computing the second moments that makes use of the expression of the circuit amplitudes as a partition function is given by IQP circuits. For those circuits it is possible to directly compute the average collision probability, making use of the simple structure of the output probabilities as an Ising partition function (Bremner, Montanaro, and Shepherd, 2016); see Eq. (40). There is also a direct proof of anticoncentration that does not rely on bounding second moments for the DQC1 model (Morimae, 2017).

The most important schemes for which anticoncentration has remained elusive are boson-sampling protocols. For Fock boson sampling, one can also compute the second moment of the output probabilities by making use of the hiding property such that the well-studied properties of Gaussian matrices can be exploited to compute $\mathbb{E}_{X \sim \mathcal{G}}[|\text{Perm}(X)|^2] = n!$ and $\mathbb{E}_{X \sim \mathcal{G}}[|\text{Perm}(X)|^4]/(n!)^2 = n + 1$ (Aaronson and Arkhipov, 2013). The value of the second moments translates to a bound on the anticoncentration probability γ in Eq. (79) given by $1/(n + 1)$ (Aaronson and Arkhipov, 2013). While numerical evidence suggests that anticoncentration is true for Fock boson sampling (Aaronson and Arkhipov, 2013), second moments are therefore not sufficient to prove this. Improving this bound, Tao and Vu (2009) proved that the permanent of $n \times n$ Bernoulli matrices is of the order of $n^{n(1/2-\epsilon)}$ with probability $1 - n^{-0.1}$, while a bound of the order of $n^{n\{1/2-O(\log(n))\}}$ with inverse polynomial failure probability would be required for anticoncentration (Aaronson and Arkhipov, 2013). While this result may be extended to Gaussian distributions over \mathbb{C} , it is unclear how to further improve it (Tao and Vu, 2009). As a way around this, one might try to use higher moments of the Fock boson-sampling distribution in order to obtain tighter bounds than are provided by the Paley-Zygmund inequality. First steps in this direction were taken by Nezami (2021), who characterized all moments of the distribution of Gaussian permanents and computed the lower ones but concluded that a closed formula for all moments may be sufficient to prove anticoncentration. For Gaussian boson sampling the situation remains even more elusive, as here the distribution over which moments of the Hafnian (9) need to be computed is the so-called circular

orthogonal ensemble (COE), which is given by symmetric Gaussian matrices of the form XX^T with $X \sim \mathcal{G}$; see Sec. IV.C.4.c.

Remember that anticoncentration is merely a necessary condition for additively approximate average-case hardness, and a means to reduce this to relative-error approximations. It remains to prove the approximate average-case hardness conjecture in either its additive or its relative-error version. This is the focus of Sec. IV.D.3.

3. Average-case hardness: An overview

Generally speaking average-case complexity is a crucial question in cryptography and comes with a number of interesting peculiarities. However, we have few handles on average-case complexity and proofs of average-case hardness are possible for only a few complexity classes. The question of average-case hardness was first posed by Levin (1986) as a means to narrow down problem classes in which one can hope for simulation algorithms that work on average. What is the complexity of an instance drawn at random from some distribution μ over all possible problems? An important question in the context of average-case complexity is one that was posed by Levin (1986): How does the average-case complexity of a problem class depend on the distribution? If one defines a probability measure to be supported on hard problem instances only, average-case complexity equals worst-case complexity. There even exists a single so-called universal distribution for which the average-case complexity of any algorithm equals its worst-case complexity (Li and Vitányi, 1992). The strong dependence on the distribution is part of the reason why average-case complexity under natural measures such as the uniform measure has remained largely elusive.

Results that characterize average-case complexity of certain problems are known only for counting problems. The key conceptual idea underlying proofs of average-case hardness for such problems is the notion of *random self-reducibility*. We say that a computational problem is randomly self-reducible if we can polynomially reduce the problem of evaluating any fixed instance x to evaluating random instances y_1, \dots, y_k with a bounded probability that is independent of the input. Random self-reducibility is therefore a particular type of worst-to-average-case reduction: We assume that there is a machine that solves random instances with probability bounded away from 1 over a given distribution and then use this machine to try to efficiently solve an arbitrary fixed instance. If this is possible, then such a machine allows us to solve any instance in a time that is polynomially equivalent to the time it takes to solve a random instance. Hence, the problem must be as hard on average over this distribution as in the worst case.

A first step toward proving approximate average-case hardness of quantum output probabilities (that also constitutes a necessary condition) is to prove average-case hardness of near-exactly computing those output probabilities for the respective circuit family. Average-case complexity for near-exact computation was pioneered by Lipton (1991) for the permanent as it prominently features in boson sampling (Aaronson and Arkhipov, 2013). The key idea of Lipton's method is to use polynomial interpolation in order to

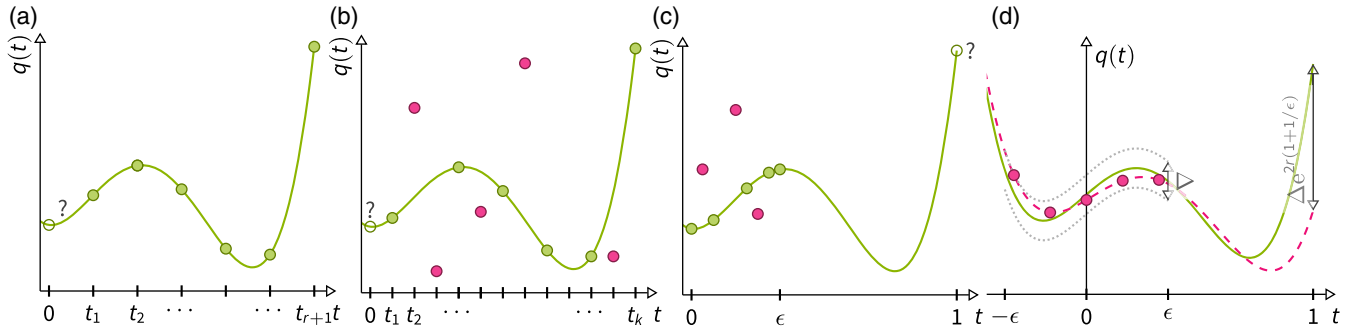


FIG. 7. (a) From at least $r + 1$ interpolation points $(t_i, q(t_i))$ one can efficiently interpolate a polynomial $q(t)$ of degree r . (b) Using the Berlekamp-Welch decoding algorithm (Welch and Berlekamp, 1986) for the Solomon-Reed code, one can reconstruct a degree- r polynomial from k points (t_i, y_i) if at least $(k + r)/2$ of those points are correct. (c) When drawing instances from a distribution on the infinite field \mathbb{C} as opposed to the uniform measure over a finite field, the interpolation points are chosen from the interval $[0, \epsilon]$ for $\epsilon = 1/\text{poly}(n)$ so that the distribution of $G(t)$ in Eq. (108) does not deviate too far from the original distribution. (d) Using the result of Theorem 23 by Rakhmanov (2007), one can bound the interpolation error of a degree- r polynomial in the interval $(-\epsilon, \epsilon)$ when given evaluation points that are correct up to an error Δ (with inverse polynomial failure probability). Using the Lemma 22 of Paturi (1992), one can then bound the extrapolation error when extrapolating to the hard problem instance at $t = 1$.

interpolate from certain judiciously chosen random instances to an arbitrary, fixed instance. This method is possible if the quantity in question can be written as a polynomial in the input parameters. While random quantum circuits lack this structure, the polynomial interpolation method of Lipton's can in fact be adapted to a broad class of quantum random sampling schemes (Movassagh, 2018, 2020; Bouland *et al.*, 2019, 2022; Kondo, Mori, and Movassagh, 2022; Krovi, 2022). In the following, we introduce and discuss these methods, which eventually come close to proving approximate average-case hardness in that they tolerate an additive error of $O(2^{-O(m)})$ for random universal circuits, where m is the number of gates in the circuit (Krovi, 2022). However, the step to inverse exponential $O(2^{-n})$ or relative-error average-case complexity remains wide open, and indeed remains the central open question in the field of quantum supremacy from a complexity-theoretic viewpoint.

4. Random self-reducibility of the permanent

We start with the simplest and historically original proof of average-case hardness for $\#\text{P}$ -random self-reducibility of the permanent over a finite field \mathbb{F} with respect to the uniform distribution over that field. Recall the following definition of the permanent of an $n \times n$ matrix X over \mathbb{F} [Eq. (46)]:

$$\text{Perm}(X) = \sum_{\sigma \in \mathcal{S}_n} \prod_{j=1}^n x_{j, \sigma(j)}. \quad (100)$$

The underlying structure in which the proof of random self-reducibility for the permanent is rooted is the algebraic fact that it is a degree- n polynomial in the matrix entries of X [and a degree- $2n$ polynomial in the case of $|\text{Perm}(X)|^2$]. Concretely, the idea is the following: Given an arbitrary instance $A \in \mathbb{F}^{n \times n}$, draw a uniformly random matrix B and for $t \in \mathbb{F}$ define the matrix

$$E(t) = A + tB \quad (101)$$

for $t \in \mathbb{F}$. We think of A as a hard instance. Notice that, for any fixed value of $t \neq 0$, $E(t)$ is distributed uniformly over \mathbb{F} . This is in spite of the fact that $E(t)$ and $E(t')$ are correlated for values $t, t' \in \mathbb{F}$. As the permanent is a degree- n polynomial in the matrix entries of an $n \times n$ matrix, the permanent of the matrix $E(t)$ is a degree- n polynomial $q(t) = \text{Perm}[E(t)]$ in t .

We now assume that there is an efficient machine \mathcal{O} that computes $\text{Perm}(X)$ for uniformly random instances X with failure probability δ . This algorithm, while it may fail to evaluate $q(0) \equiv \text{Perm}(A)$, will by assumption likely evaluate $q(t_i)$ correctly for some choice of evaluation points t_i . The idea is to infer $q(0)$ from the values of q at the points $\{t_i\}_i$ using polynomial interpolation; see Fig. 7(a).

We can now query \mathcal{O} on $n + 1$ distinct points $t_1, \dots, t_{n+1} \neq 0$, obtaining the values $q(t_i)$.¹⁹ Applying a union bound, the probability that all of those values are correct is lower bounded by $1 - (n + 1)\delta$. Setting $\delta = 1/3n$, we thus obtain $n + 1$ correct pairs $\{(t_i, q(t_i)), i \in [n + 1]\}$ with a probability of at least $2/3 - 1/3n$. But q is a degree- n polynomial, and hence those points uniquely determine q . We can now solve a linear system of equations to interpolate the polynomial q and compute $q(0) = \text{Perm}(A)$. Hence, an algorithm that solves random instances with a probability of at least $1 - 1/3n$ is able to solve arbitrary instances and computing the permanent over finite fields is average-case hard on any $1 - 1/3n$ fraction of the instances.

a. Improving the success probability

Being correct on any $1 - 1/(3n)$ fraction of the instances is a strong requirement on the evaluation algorithm, however, and by contraposition requires only that at most a $1/3n$ fraction of the instances indeed need to be $\#\text{P}$ hard to compute. It is desirable to lower this requirement as far as

¹⁹Notice that this requires the size of \mathbb{F} to be at least $n + 2$, and hence Lipton's proof does not work for the field \mathbb{F}_2 . Indeed, in this case there are also known approximation schemes for the permanent (Jerrum, Sinclair, and Vigoda, 2004).

possible to make stronger statements and assess average-case hardness as well as possible.

Indeed, we can bring down the requirement on \mathcal{O} to work correctly only for a constant $1/2 + 1/\text{poly}(n)$ fraction of the instances (Gemmell *et al.*, 1991; Gemmell and Sudan, 1992); see also Arora and Barak (2009). The idea is to use error-correction techniques for polynomial codes such as the Reed-Solomon one (Reed and Solomon, 1960), where a string of n symbols is identified with the coefficients of a degree- $(n-1)$ polynomial. Decoding algorithms for such codes output the correct polynomial even in the presence of some amount of errors.

An error-correction algorithm for Reed-Solomon codes that will be extremely useful for our purposes is the algorithm by Welch and Berlekamp (1986), as it works over arbitrary fields and can even be extended to rational-function interpolation (Movassagh, 2018, 2020).

Theorem 20 (Unique decoding for Reed-Solomon codes) (Welch and Berlekamp, 1986).—Let q be a degree- r polynomial over any field \mathbb{F} . Suppose that we are given k pairs of elements $\{(t_i, y_i)\}_{i \in [k]}$ with all t_i distinct, with the condition that $y_i = q(t_i)$ for at least $\max(r+1, (k+r)/2)$ points. One can then uniquely recover q exactly in $\text{poly}(k, r)$ deterministic time.

We illustrate decoding with errors in Fig. 7(b). Notice that for polynomially large k the Berlekamp-Welch decoding algorithm tolerates an error rate that is arbitrarily close to half. The Berlekamp-Welch algorithm is thus optimal in that, as soon as less than half of the points are correct, no unique solution is guaranteed to exist.

This issue is addressed by so-called list-decoding algorithms, which output a list of compatible solutions, given the observation that there cannot be too many such solutions (Arora and Barak, 2009). These algorithms were developed (Beaver and Feigenbaum, 1990; Lipton, 1991) for so-called Reed-Muller codes (Muller, 1954; Reed, 1954) over finite fields of which the Reed-Solomon one is a special case (Sudan, 1997). Using list-decoding algorithms, average-case hardness of the permanent over sufficiently large finite fields has even been shown for any inverse polynomial fraction of correct points (Cai, Pavan, and Sivakumar, 1999); see Guruswami (2006) for an overview of such approaches.

We now illustrate the use of the Berlekamp-Welch algorithm to prove average-case hardness given by Gemmell *et al.* (1991): Using the Berlekamp-Welch algorithm, we can query the oracle \mathcal{O} a number of times given by $k > 2(n+1)$ at distinct points t_i , obtaining pairs $(t_i, \mathcal{O}(t_i))$. We can then upper bound the probability that less than $(k+n)/2$ of the obtained data points are correct, as

$$\Pr \left[|\{i, \mathcal{O}(t_i) \neq q(t_i)\}| > k - \frac{k+n}{2} \right] < \frac{2\delta k}{k-n} \quad (102)$$

using Markov's inequality. This probability is at most $1/2$ if the failure probability of \mathcal{O} satisfies

$$\delta < \frac{1}{4} \left(1 - \frac{k}{n} \right). \quad (103)$$

Hence, the decoding procedure succeeds using k samples as long as \mathcal{O} works on a $3/4 + k/(4n) = 3/4 + 1/\text{poly}(n)$ fraction of the instances. Using an interpolation path to A , which is a polynomial in k , Gemmell and Sudan (1992) showed that this can be further improved to a $1/2 + 1/\text{poly}(n)$ fraction.

b. Distributions over infinite fields: The case of $\mathbb{F} = \mathbb{C}$

When considering the output probabilities of Fock boson sampling [Eq. (8)] and Gaussian boson sampling [Eq. (9)], and also looking ahead of quantum circuits, the matrices in question have entries in not a finite but rather an infinite field, the complex numbers $\mathbb{F} = \mathbb{C}$. In this case, we are faced with two additional technical difficulties: First, there is no uniform or translation-invariant measure over the complex numbers. This means that when we construct the random matrix $E(t)$ as in Eq. (101) by drawing a random matrix B from some distribution μ , then $E(t)$ will be distributed according to some distribution μ' depending on the value of t and the hard instance A . Second, assuming that we have found a solution to this problem, the previously used polynomial interpolation and error-correction techniques for the case of finite fields fail if we only have a finite approximation of the values of $q(t_i)$. Numerically dealing with real numbers will, however, inevitably lead to finite-precision errors on the order of $2^{-\text{poly}(n)}$.

We can circumvent the first problem by choosing values of t that are small such that the difference between μ' and μ in total-variation distance is small. As the total-variation distance upper bounds the difference in probability that the two distributions assign to a specific event, this difference translates to an additional contribution to the failure probability of \mathcal{O} .

The natural distribution over \mathbb{C} that also appears in the Fock boson-sampling problem is the complex normal distribution $\mathcal{N}_{\mathbb{C}}(\mu, \sigma)$ with mean μ and variance σ^2 . The following lemma, a variation of Lemma 7.4 of Aaronson and Arkhipov (2013), bounds the total-variation distance between slightly shifted and squashed product Gaussian distributions with products of the standard distribution.

Lemma 21 (Autocorrelation of Gaussian distributions).—For the distributions

$$\mathcal{D}_1 = \mathcal{N}_{\mathbb{C}}(0, (1 - \epsilon^2)\sigma)^M, \quad (104)$$

$$\mathcal{D}_2 = \prod_{i=1}^M \mathcal{N}_{\mathbb{C}}(v_i, \sigma), \quad (105)$$

with $v = (v_1, v_2, \dots, v_M) \in \mathbb{C}^M$ and $\epsilon, \sigma > 0$, it holds that

$$\|\mathcal{D}_1 - \mathcal{N}_{\mathbb{C}}(0, \sigma)^M\|_{\text{TV}} \leq 2M\epsilon, \quad (106)$$

$$\|\mathcal{D}_2 - \mathcal{N}_{\mathbb{C}}(0, \sigma)^M\|_{\text{TV}} \leq \frac{1}{\sigma} \|v\|_{\ell_1}. \quad (107)$$

The same result holds for the uniform distribution $\mathcal{U}_{\mathbb{C}}(\mu, \sigma)$ centered around μ with cutoff σ .

For an arbitrary matrix $A = (a_{i,j})_{i,j}$ we now define the family of matrices

$$G(t) = tA + (1-t)B \quad (108)$$

similarly by drawing standard normal distributed instances $B \in \mathbb{C}^{n \times n}$. The matrix $E(t)$ is then distributed according to the new distribution

$$\mathcal{D} = \prod_{i,j=1}^n \mathcal{N}_{\mathbb{C}}(ta_{i,j}, (1-t)^2). \quad (109)$$

Choosing equidistant values of t_i in the interval $(0, \epsilon]$ for some cutoff $\epsilon > 0$ will then result in a success probability of the algorithm \mathcal{O} that has a failure probability δ that is given by

$$\Pr[\mathcal{O}(t_i) = q(t_i)] \geq 1 - \delta - \|\mathcal{D} - \mathcal{G}_{\mathbb{C}}(0, 1)^{n^2}\|_{\text{TV}} \quad (110)$$

$$\geq 1 - \delta - 6n^2\epsilon. \quad (111)$$

The remainder of the argument follows analogously by choosing $\epsilon = \delta/6n^2$. We illustrate the procedure in Fig. 7(c).

c. Robustness to finite-precision errors

The finite-precision problem requires somewhat more powerful machinery: using bounds on the stable extrapolation and interpolation of polynomials, we can recover the original proof using polynomial interpolation. This comes at a cost, however: we cannot make use of the error-correction techniques of Berlekamp and Welch anymore, because those techniques require that some of the points are evaluated exactly.

The two results that have been identified as being helpful to this effort by Aaronson and Arkhipov (2013) are a lemma by Paturi (1992) and a theorem by Rakhmanov (2007).

Lemma 22 (Stable extrapolation) (Paturi, 1992).—Let $p: \mathbb{R} \rightarrow \mathbb{R}$ be a polynomial of degree r , and suppose that $|p(x)| \leq \Delta$ for all x such that $|x| \leq \epsilon$. Thus, $|p(1)| \leq \Delta e^{2r(1+1/\epsilon)}$.

Theorem 23 (Stable interpolation) (Rakhmanov, 2007).—Let E_k denote the set of k equidistant points in $(-1, 1)$. For a polynomial $p: \mathbb{R} \rightarrow \mathbb{R}$ of degree r such that $|p(x)| \leq 1$ for all $x \in E_k$, it then holds that

$$|p(x)| \leq C \log \left(\frac{\pi}{\arctan[(k/r)\sqrt{\mathcal{R}^2 - x^2}]} \right) \quad (112)$$

for $|x| \leq \mathcal{R} = \sqrt{1 - r^2/k^2}$.

We can now apply these results to the polynomial $p(t) = q(t) - q'(t)$, where $q'(t)$ is the polynomial defined by the slightly erroneous values $q'(t_i)$ of $q(t_i)$ satisfying $|q'(t_i) - q(t_i)| \leq 2^{-O(n^c)}$ for a sufficiently large c . Using the result of Rakhmanov (2007), we can bound the error between q and q' between the evaluation points. Using Paturi's lemma (Paturi, 1992), we can then bound the error tolerance when extrapolating to $q(1)$; see Fig. 7(d).

Note that exactly the same arguments apply to the output probabilities of Gaussian boson sampling, which are given by the squared Hafnian $|\text{Haf}(XX^T)|^2$ for Gaussian $X \in \mathbb{C}^{2n \times 2k}$;

recall Eq. (47). The squared Hafnian is a degree- $2n$ polynomial in its matrix entries [recall its relation Eq. (50) to the permanent], and hence a degree- $4n$ polynomial in the matrix entries of the Gaussian-distributed matrix X .

5. Average-case hardness of quantum output probabilities

We now turn to average-case hardness of the output probabilities of quantum circuits. We first observe that there is a natural polynomial structure on the success probabilities of quantum circuits. For a quantum circuit $C = C_m \cdots C_2 C_1$ comprising m gates C_i acting on n qubits, the output amplitudes can be expressed in terms of the following path integral:

$$\langle 0|C|0 \rangle = \sum_{\lambda_1, \dots, \lambda_{m-1} \in \{0,1\}^n} \langle 0|C_m|\lambda_{m-1} \rangle \cdots \langle \lambda_2|C_2|\lambda_1 \rangle \langle C_1|0 \rangle. \quad (113)$$

Consider that C is drawn from some measure μ_C that defines a circuit family \mathcal{C} . Some of the gates in C might be randomly drawn from a gate set \mathcal{G} , while others might be fixed across all $C \in \mathcal{C}$.

Now we are faced with a severe issue when trying to instantiate the idea of Lipton (1991), however: when trying to construct an equivalent of $E(t)$ by choosing random instances B for a fixed worst-case circuit A , the matrix given by $A + tB$ will not be unitary for $t \neq 0$, and therefore does not define a valid problem instance. This is because the unitary matrices form a group with respect not to addition but to multiplication. How then can we perform a worst-to-average-case reduction? A natural idea is to make use of the group structure by multiplying A and B in a gatewise fashion in a way that is polynomial in an interpolation parameter and then showing that the distribution of the resulting instances does not deviate too much from the distribution of B . We can do so in different ways.

a. Local Taylor-series truncation

On a high level, the first approach saves the polynomial structure of Eq. (101) by making use of a Taylor expansion. We interpolate between a hard and a random instance as follows. For a hard instance of a circuit C with random gates C_1, \dots, C_m drawn uniformly from a continuous subgroup \mathcal{G} of the corresponding unitary group $U(d)$, we define a new circuit by setting each gate

$$C_i(t) = C_i H_i e^{-ith_i}, \quad (114)$$

where H_i is Haar random in \mathcal{G} and $h_i = -i \log H_i$ is its generator. If the resulting circuit is denoted as $C \circ H(t)$, $C_i(0)$ is Haar random in \mathcal{G} , while for $t = 1$ we recover the original gate C_i . As with average-case hardness of Gauss-random permanents, for small t the gate $H_i e^{-ith_i}$ looks almost Haar random. One therefore wants to follow the same previously given procedure to extrapolate to $t = 1$, given values of $|\langle 0|C \circ H_t|0 \rangle|^2$.

However, the gates $C_i(t)$, and hence the output probability $|\langle 0|C \circ H_t|0 \rangle|^2$, are not low-degree polynomials in t ; thus,

polynomial interpolation cannot be applied. An easy way to circumvent this problem is to consider Taylor approximations of the deformed gates $C_i(t)$. We define the (t, K) -truncated and perturbed Haar measure on the circuit family \mathcal{C} by replacing each Haar-random gate H_i in a circuit C with

$$G_i = H_i \left(\sum_{k=0}^K \frac{(-ih_i t)^k}{k!} \right). \quad (115)$$

We can now use the standard Suzuki bound on Taylor truncations

$$|\langle \psi | C_i G_i - C_i H_i e^{-ith_i} | \psi \rangle| \leq \frac{\kappa}{K!} \quad (116)$$

for a constant $\kappa > 0$, set $K \in \text{poly}(n)$, use an analog of Lemma 21 to complete a worst-to-average-case reduction for exactly computing the probabilities on any $3/4 + 1/\text{poly}(n)$ fraction of the instances. Alternatively, as discussed, we can apply the stability results by [Paturi \(1992\)](#) and [Rakhmanov \(2007\)](#) to achieve robustness to additive errors $2^{-\text{poly}(n)}$ on a $1 - 1/\text{poly}(n)$ fraction of the instances; see also [Bouland et al. \(2019\)](#).

A notable caveat of this approach is that in the reduction the unitary group remains since the Taylor truncation of e^{-ith_i} is nonunitary. This means that average-case hardness is achieved not for exactly evaluating the circuit success probabilities but only for exactly evaluating numbers $p_0(C)'$, which are $2^{-\text{poly}(n)}$ -additive approximations thereof and which do not correspond to success probabilities of valid quantum circuits. Nevertheless, average-case hardness of those numbers is a necessary requirement for the additive approximate average-case hardness property and hence serves as evidence for the conjecture. In addition, the additive approximate average-case hardness conjecture of the truncated distribution is equivalent to the additive approximate average-case hardness conjecture of the nontruncated distribution ([Bouland et al., 2019](#)). For a more detailed discussion of this caveat, see [Movassagh \(2020\)](#) and [Napp et al. \(2022\)](#).

b. Rational-function interpolation

A more natural interpolation that remains within the unitary group and is much more error robust makes use of the *Cayley function*

$$f(x) = \frac{1 + ix}{1 - ix} \quad (117)$$

for $x \in \mathbb{R}$, defining $f(-\infty) = -1$. The Cayley function is a bijection between $\mathbb{R} \cup \{-\infty\}$ and the complex unit circle. Observing that unitary matrices have eigenvalues on the complex unit circle, a Haar random unitary matrix $H \in U(d)$ can therefore be uniquely represented as

$$H = f(h), \quad h = h^\dagger, \quad (118)$$

and $H^\dagger = f(-h)$. For each quantum gate $C_i \in U(d)$ we can then construct the path

$$C_i(t) = C_i f(th_i), \quad (119)$$

with $h_i = f^{-1}(C_i^\dagger H_i)$ for Haar-random H_i such that $C_i(0) = C_i$ and $C_i(1) = H_i$. The interpolated gate (119) can be expressed as a fraction of two degree- d polynomials using the spectral decomposition of $h = \sum_{\alpha=1}^d h_{i,\alpha} |\psi_{i,\alpha}\rangle \langle \psi_{i,\alpha}|$ as

$$C_i(t) = \frac{1}{q_k(t)} \sum_{\alpha=1}^d p_{i,\alpha}(t) C_i |\psi_{i,\alpha}\rangle \langle \psi_{i,\alpha}|, \quad (120)$$

with

$$q_i(t) = \prod_{\alpha=1}^d (1 + ith_{i,\alpha}), \quad (121)$$

$$p_{i,\alpha}(t) = f(h_{i,\alpha}) (1 - th_{i,\alpha}) \prod_{\beta \in [d] \setminus \alpha} (1 + ith_{i,\beta}). \quad (122)$$

Denote the circuit resulting from this interpolation as $C \star H(t)$. Now one can bound the total-variation distance for the distribution \mathcal{D}_ϵ on the circuit obtained when choosing $t = 1 - \epsilon$ as $O(m\epsilon)$ ([Movassagh, 2020](#)).

However, while the techniques we have used thus far were useful for polynomial interpolation, we now need to extrapolate a rational function. As a first step, one can generalize the Berlekamp-Welch algorithm to rational functions with degrees k_1 and k_2 in the numerator and denominator, respectively ([Gemmell and Sudan, 1992](#); [Movassagh, 2018](#)). This algorithm requires the number of evaluation points t_i to be at least $k_1 + k_2 + 2e$, where e is the number of errors made by the evaluation algorithm \mathcal{O} .

A barrier to making this result robust lies in the fact that the results on stable interpolation ([Rakhmanov, 2007](#)) and extrapolation ([Paturi, 1992](#)) of low-degree polynomials do not apply to rational functions. [Movassagh \(2020\)](#) observed, however, that the output probabilities of the interpolated circuit can be reduced to a polynomial. To see this, observe that the output probabilities can be written as the following fraction of two polynomials $Q(t) = \prod_{i=1}^m q_i(t)$ and $P(t) = \prod_{i=1}^m \sum_{\alpha} p_{i,\alpha}(t) C_i |\psi_{i,\alpha}\rangle \langle \psi_{i,\alpha}|$:

$$|\langle 0 | C \star H(t) | 0 \rangle|^2 = \frac{|\langle 0 | P(t) | 0 \rangle|^2}{|Q(t)|^2}. \quad (123)$$

But as we can compute $Q(t)$ exactly in time $\Theta(m)$, we can reduce the rational function to a polynomial function by multiplying $|\langle 0 | C \star H(t) | 0 \rangle|^2$ by $|Q(t)|^2$. Now one can show that $|Q(t)|^2 \leq 1 + O(m\epsilon)$, so when choosing $\epsilon = 1/m$ the additional error incurred due to this multiplication is a multiplicative $O(1)$ error. The scaling of the extrapolation error in $|\langle 0 | C \star H(t) | 0 \rangle|^2$ is therefore not disturbed when interpolating $|Q(t)|^2 |\langle 0 | C \star H(t) | 0 \rangle|^2$ instead.

Now we can again resort to Lemma 22 and Theorem 23 in order to compute the robustness as $2^{-O(m/\epsilon)} = 2^{-O(m^2)}$ on any $1 - 1/\text{poly}(n)$ fraction of the instances ([Movassagh, 2020](#)) where $(0, \epsilon]$ defines the interval on which the success probabilities of $C(t)$ are evaluated. [Kondo, Mori, and Movassagh \(2022\)](#) observed that this can be further improved

TABLE I. Comparison of the average-case hardness results for random quantum circuits on n qubits with m gates.

Reference	Path	Interpolation method	Robustness	Instance fraction
Bouland <i>et al.</i> (2019) ^a	Truncated local Taylor series	Berlekamp-Welch (BW)	Exact	$3/4 + 1/\text{poly}(n)$
	Truncated local Taylor series	Paturi + Rakhmanov	$2^{-\text{poly}(n)}$	$1 - 1/\text{poly}(n)$
Movassagh (2018)	Cayley paths	Rational BW	Exact	$3/4 + 1/\text{poly}(n)$
Movassagh (2020)	Cayley paths	Paturi + Rakhmanov	$2^{-O(m^3)}$	$1 - 1/\text{poly}(n)$
Bouland <i>et al.</i> (2022)	Cayley paths	Robust BW in BPP ^{NP}	$2^{-O(m \log m)}$	$3/4 + 1/\text{poly}(n)$
Kondo, Mori, and Movassagh (2022)	Cayley paths	Lagrange interpolation + error bounds	$2^{-O(m \log m)}$	$1 - 1/O(m)$
Krovi (2022)	Truncated global Taylor series	Robust BW in BPP	$2^{-O(m)}$	$> 3/4$

^aNote that Bouland *et al.* (2019) proved average-case hardness for a nonunitary circuit whose output probabilities were $2^{-\text{poly}(n)}$ close to the ideal output probabilities. The robustness that we reference is with respect to this nonunitary circuit; see the main text for a discussion of this point.

using the same strategy if Lagrange polynomials are used for the interpolation. For those polynomials, they found results analogous to Lemma 22 of Paturi (1992), and Theorem 23 of Rakhmanov (2007) to obtain a robustness of $2^{-O(m \log m)}$ on any $1 - 1/O(m)$ fraction of the instances.

The limitation of this approach, however, is that there is no error-correction procedure such that all results of the oracle need to be correct, giving rise to a small tolerated failure probability because a union bound needs to be applied. Aiming to circumvent this issue, Bouland *et al.* (2022) observed that the failure probability can further be improved to $3/4 + 1/\text{poly}(n)$ while retaining the same error scaling $2^{-O(m \log m)}$ by making use of an NP oracle. They achieve this by constructing a more robust Berlekamp-Welch algorithm for polynomial interpolation over the real numbers. This algorithm makes use of the NP oracle in addition to randomness, and is therefore in the third level of the polynomial hierarchy.

c. Global Taylor-series truncation

Krovi (2022) recently observed that, rather than performing a Taylor-series truncation on the level of individual gates, one can perform such a truncation on the level of the global output distribution. The key observation of Krovi (2022) is that the output probabilities of circuits interpolated via Eq. (114) can be expressed as a path integral

$$p(t) = \sum_r e^{-i(t/m)\Delta\phi_r} A_r \quad (124)$$

in terms of the 4^{2m} paths r , $|A_r| \leq 1$ and $|\Delta\phi_r|/m \in O(1)$. Here the coefficients A_r can be thought of as the path weights and $\Delta\phi_r$ can be considered their phases. Performing an appropriately chosen Taylor-series truncation of $p(t)$, one finds that a degree- $O(m/\log m)$ polynomial is sufficient to achieve error robustness $2^{-O(m)}$ for circuits with Haar-random two-qubit gates and in fact robustness $2^{-O(n)}$ for IQP circuits. This result thus reduces the gap to the required robustness of 2^{-n} to constants in the exponent. By making use of recent results in polynomial interpolation that use specifically chosen points (Kane, Karmalkar, and Price, 2017), one can further improve the success probability of the interpolation to a constant without the need for an NP oracle, as shown by Bouland *et al.* (2022).

We summarize the various average-case hardness results²⁰ just discussed in Table I.

d. From continuous to discrete subgroups

A key issue to note in the worst-to-average-case reductions on the unitary group is that the random gates in the circuit families need to be drawn from *continuous subgroups* of the unitary group. Only if this is the case can one choose values of the interpolation parameter t that are small enough that the measure on the gate set is not perturbed too much in the interpolation step. In particular, this implies that the reduction does not apply to discrete gate sets, and for some architectures the choice of random gates must be modified for the reduction to apply. For instance, to apply the average-case hardness results to the IQP circuit family defined in Eq. (6), we need to choose the edge weights $w_{i,j}$ uniformly from the unit circle S^1 rather than from a discrete set of angles; see also Haferkamp, Hangleiter, Eisert, and Gluza (2020).

One step in the direction of achieving an exact average-case hardness reduction for a discrete gate set was taken by Dalzell *et al.* (2020) with their Theorem 6. They considered the discrete family of IQP circuits whose output amplitudes are given by gaps of degree-3 Boolean polynomials; see Eqs. (5) and (44). Specifically, they showed a recursive reduction from the gap of a degree-3 polynomial with random degree-1 terms (but fixed degree-2 and degree-3 terms) to the gap of a worst-case polynomial (with the same degree-2 and degree-3 terms). This translates to an exact average-case hardness result over a certain discrete family of IQP circuits. There are two problems with this approach, however. First, the family is specific since it depends explicitly on the degree-2 and degree-3 terms of a worst-case instance. Second, it does not work for the output probabilities, since these no longer contain sign information about the gap, which is crucial for the reduction; compare also the proof of approximate worst-case hardness of GapP discussed in Sec. III.D. This strategy is still worth noting, however, since it is intrinsically distinct from the previously discussed polynomial interpolation approaches and might yield another path to proving approximate average-case hardness.

²⁰We also note that a formulation of the aforementioned proof strategy using the language of representations of Lie groups is provided by Oszmaniec *et al.* (2022).

6. Discussion

Using the previously discussed techniques, we are currently able to prove approximate average-case hardness for universal random circuits with robustness $2^{-O(m)}$, where m is the number of gates in the circuit. This is further improved for IQP circuits to $2^{-O(n)}$, where n is the number of qubits. To prove the approximate average-case hardness conjecture, we would need to improve this to $O(2^{-n})$, however. Can we hope to prove such a result? The key technical obstacle on the way to addressing this question is the instability of polynomials with respect to variations in the interpolation points. Indeed, we saw in Paturi's lemma (Lemma 22) that the extrapolation error of a bounded error polynomial scales exponentially in the degree r and size of the interval ϵ on which the bound holds, and the version used by [Kondo, Mori, and Movassagh \(2022\)](#) scales as an order- d Chebyshev polynomial in ϵ . As we have to make this interval inverse polynomially small to maintain closeness of the probability distributions, it results in a strong increase of the Paturi bound that can be counter-weighted only by an inversely scaling error bound on the interval $(-\epsilon, \epsilon)$. Small variations of a polynomial at a few points can thus lead to large variations far from those points.

Random self-reducibility thus seems doomed when it comes to additive robustness of success probabilities on the order of 2^{-n} , as would be necessary for the quantum supremacy conjecture. Indeed, [Aaronson and Arkhipov \(2013\)](#) argued that polynomial interpolation faces a significant barrier. They claimed that—in the presence of anticoncentration—the fact that polynomial interpolation is linear in the coefficients and hence linear with respect to additive errors prohibits it from allowing approximate average-case hardness to be proven. Roughly speaking, this is because even if two polynomials agree up to exponentially small error in an interval, they may exponentially disagree outside of that interval, while at the same time the target value of the polynomial might not be exponentially larger. Hence, constant relative-error approximations in the evaluation interval could translate to exponentially larger relative-error approximations at the target point. The suggestion of [Aaronson and Arkhipov \(2013\)](#) was then to make use of a restricted class of polynomials that are not closed under addition, but that are at the same time able to capture the quantity of interest.

Making this argument somewhat quantitative, [Boulund et al. \(2022\)](#) investigated the applicability of random self-reducibility in the context of noisy circuits with error detection; see also [Aaronson and Arkhipov \(2013\)](#). They showed that even noisy, error-detectable probabilities that are conjectured to be $2^{-O(m)}$ close to uniform ([Boixo et al., 2018](#)) remain $\#\text{P}$ hard to compute up to error $2^{-16m \log m - O(m)}$ in the average case via random self-reducibility. But they argue that this implies that the average-case robustness of $2^{-O(m \log m)}$ is essentially optimal for this technique up to log factors in the exponent. The result of [Krovi \(2022\)](#) has further removed the log factors in the exponent, providing a matching $2^{-O(m)}$ scaling of the robustness for universal random circuits and $2^{-O(n)}$ for IQP circuits. Similarly, for the case of Fock boson sampling on $m = O(n^c)$ many modes, they were able to show an even tighter error bound of $e^{-(c+4)n \log n - O(n)}$, which is only

constant factors in the exponent away from the $e^{-n \log n}$ robustness required to prove the approximate average-case hardness conjecture.

Note that, while the scaling conjecture of [Boixo et al. \(2018\)](#) was recently shown in the low-noise limit ([Dalzell, Hunter-Jones, and Brandão, 2021](#)), at high-noise strengths [Deshpande, Niroula et al. \(2022\)](#) proved the expected convergence of the probabilities to uniform as $2^{-n-O(d)}$ ([Deshpande, Niroula et al., 2022](#)); see Sec. IV.F.1 for a discussion of these results. The latter result may significantly lower the barrier for random circuits.

As another piece of evidence complicating a proof of approximate average-case hardness, for a somewhat baroque constant-depth universal circuit architecture in two dimensions, [Napp et al. \(2022\)](#) proved that no approximate worst-to-average-case reduction that enables the Stockmeyer argument is possible. Rather, this architecture admits algorithms for both strong and weak simulation that are efficient in large fractions of instances. But for the same architecture strong simulation is classically intractable in the worst case unless GapP admits a polynomial-time algorithm and the polynomial hierarchy collapses to its third level, respectively. Moreover, they provided numerical evidence that random constant-depth universal circuits in two dimensions are efficiently simulable on average in practice. Strengthening this point, [Deshpande, Niroula et al. \(2022\)](#) showed that at sublogarithmic depth almost all probabilities are subexponentially small for random universal circuits, so anticoncentration does not hold. This implies that the trivial algorithm that always outputs 0 is a good additive approximate average-case strong simulator for this case. Note, however, that this does not imply an average-case approximate sampling algorithm. Technically speaking, the upshot of these results is that any technique to prove approximate average-case hardness must be sensitive to the depth of the circuit since we do not expect any technique to work at low depth. Moreover, while hardness of approximate sampling might hold for certain sublogarithmic depths, we are barred from proving it via the Stockmeyer argument.

For an approximate worst-to-average-case reduction we would require, it seems, quantum success probabilities that are extremely robust to noise in generic instances. Techniques such as quantum error correction ([Raussendorf, Harrington, and Goyal, 2006](#)) might at first sight seem ideally suited for this task, but in such approaches errors need to be actively corrected. While in the framework of quantum sampling active correction can be bypassed using postselection ([Fujii, 2016](#); [Kapourniotis and Datta, 2019](#)), this means that only those probabilities corresponding to specific measurement outcomes on subsystems will be protected against errors. Since the postselection registers comprise at least a constant fraction of all registers, the protected probabilities comprise only a $2^{-\Omega(n)}$ fraction of the instances. But by the hiding property every outcome probability is in one-to-one correspondence with the acceptance probability of a circuit from the family. Thus, postselected fault tolerance seems to be in conflict with average-case hardness.

To summarize, as it stands we have strong complexity-theoretic evidence of the hardness of *exact sampling* from the output distributions of quantum random sampling schemes.

This evidence is provided by the conjectured noncollapse of the polynomial hierarchy, which is a direct generalization of the unanimously believed $P \neq NP$ conjecture, whose failure would have extreme consequences on our widely tested view of the computational complexity of many different problems. Conversely, the evidence for the hardness of *approximate sampling* is substantially weaker since it is based only on the approximate average-case hardness conjecture. The failure of this conjecture, while presumably unlikely, would not result in any meaningful consequences in complexity theory. But while, as sketched in this section, there remain significant hurdles, proving this conjecture might still be possible in the not-too-distant future.

E. Fine-grained results

The previously discussed complexity-theoretic arguments rule out an efficient classical simulation algorithm under the assumption of the noncollapse of the polynomial hierarchy and approximate average-case hardness of computing the respective output probabilities. However, they do not, and cannot, make any quantitative statements about lower bounds on the run-time of any classical simulation algorithm. But a convincing demonstration of quantum advantage requires relying not only on asymptotic complexity-theoretic statements but also on evidence that, for the given finite size of the experiment, there is no classical algorithm that can solve the problem using a reasonable amount of resources.

This is the point at which so-called fine-grained complexity results continue to try to provide lower bounds on the run-time of classical simulation algorithms. The key idea in such results is to leverage versions of the so-called strong exponential-time hypothesis (SETH), which states that certain NP-complete problems cannot be solved in time faster than 2^{an} in the input size n for some constant a , depending on the type of problem. These conjectures may then be leveraged to conjecture a fine-grained version of the collapse of the polynomial hierarchy.

We now discuss this idea more concretely using the example of IQP circuits with output probabilities given by the squared gap of degree-3 polynomials; see Eqs. (5) and (44). Dalzell *et al.* (2020) provided a fine-grained hardness argument for this circuit family via a closely related problem that they call poly3-NONBALANCED . The input to this problem is a degree-3 Boolean polynomial f , and the task is to decide whether $\text{gap}(f) \neq 0$, i.e., whether the function f has a different number of 0 and 1 outputs. Since computing the gap of degree-3 Boolean polynomials is $\#P$ complete (Montanaro, 2017), this problem is complete for a complexity class called coC=P . A language L is contained in coC=P if there exists a polynomial-time algorithm M such that, for all $x \in \{0, 1\}^*$,

$$x \in L \Leftrightarrow \text{gap}(M(x, \cdot)) \neq 0, \quad (125)$$

and is therefore closely related to the class PP where the condition is

$$x \in L \Leftrightarrow \text{gap}(M(x, \cdot)) < 0. \quad (126)$$

coC=P is analogous to PP and $\#P$ in that an oracle to coC=P is sufficient to solve any problem in the polynomial hierarchy

(Toda and Ogiwara, 1992) and, conversely, an efficient algorithm for coC=P within the polynomial hierarchy would imply a collapse of PH .

The idea of fine-grained supremacy results is now analogous to the Stockmeyer argument to assume the existence of an efficient classical derandomizable sampling algorithm for the output distribution of an n -qubit IQP circuit C_f up to a multiplicative error, using $g(n)$ gates and $t(n)$ time steps. This algorithm gives rise to a nondeterministic algorithm for poly3-NONBALANCED running in $s(n)$ steps, in the sense that it accepts if and only if there is at least one computational path (i.e., input of randomness) giving rise to the all-zero sample. The fine-grained advantage result now relies on the following conjecture (Dalzell *et al.*, 2020).

Conjecture 24 [$\text{poly3-NSETH}(a)$].—Any nondeterministic classical algorithm that solves poly3-NONBALANCED requires in the worst case 2^{an-1} time steps, where n is the number of variables in the poly3-NONBALANCED instance.

This conjecture directly yields a lower bound on the time complexity of the assumed classical sampling algorithm as $t(n) \geq 2^{an-1}$. Omitting some fine print about the computational model in which this conjecture is phrased here,²¹ the best known limit on a is given by $a < 0.9965$ (Lokshtanov *et al.*, 2017).

Analogously to the proof of additive-error sampling hardness via Stockmeyer's algorithm, fine-grained statements can be made for additive errors assuming an average-case lower bound on the run-time of a classical algorithm. From this it is possible to estimate the number of qubits required to show a quantum advantage such that no classical computer will be able to reproduce the task. Dalzell *et al.* (2020) estimated that IQP circuit sampling on roughly 200 qubits and 10^6 gates would require at least a century using a classical simulation algorithm running on state-of-the-art supercomputers.

Furthermore, statements can be made for different models by relating their simulation to well-studied problems such as poly3-NONBALANCED . In particular, this has been done for boson sampling (Dalzell *et al.*, 2020), as well as in the DQC1 model and Clifford + T universal circuit sampling (Morimae and Tamaki, 2019).

Huang, Newman, and Szegedy (2020) pursued a complementary approach on fine-grained results by considering strong simulation of quantum circuits via certain simulation algorithms. Specifically, they considered a subclass of classical simulation algorithms, which they called monotone simulators. Roughly speaking, a monotone simulator is one that does not explicitly make use of the specific values of the nonzero matrix entries of the gates. A counterexample to a monotone method is therefore the simulator of Bravyi and Gosset (2016), which explicitly uses the number of T gates in the circuit. Note, however, that a T gate does not differ from a simulable Z or S gate in terms of the locations of the nonzero matrix entries. Nonetheless, most tensor-network-based

²¹Since fine-grained complexity is about the concrete run-time, one has to fix the computational model. Typically, fine-grained complexity results are stated in terms of the so-called word RAM model (Williams, 2015).

methods (see Sec. VII for details) are well captured by the monotone framework. They show an explicit lower bound of $\tilde{O}(2^{n-3})$ on the run-time of such monotone simulators. Furthermore, invoking the exponential-time hypothesis they provide a $2^{n-o(n)}$ lower bound on strong simulations of quantum circuits.

F. Complexity of sampling in the presence of noise

The complexity-theoretic analysis we have seen thus far pertains to constant total-variation-distance errors. While this is a meaningful notion of robustness, it is extremely challenging to achieve such errors in a scalable way: doing so requires local gate errors to scale at most inversely with the circuit size. Since local gate errors are the experimental bottleneck in any implementation of quantum random sampling, it is therefore natural to ask whether the sampling task remains hard in the presence of constant local gate errors. Constant gate errors tend to give rise to a TVD between the experimental output distribution and the target distribution that deviates from unity only by an inverse exponential. But it might still be the case that the sampling task remains difficult for a classical computer.

There are at least two ways of approaching this question. First, we can ask the following: Given certain local errors in a quantum random sampling scheme, what is the complexity of sampling from the output distribution? Second, we can ask whether it is possible to design a quantum random sampling scheme that is robust to constant local errors. While the first question requires an analysis of the noisy output distribution from the perspective of computational complexity, the second question might be solved by encoding a random sampling scheme in a fault-tolerant way. We sketch some results in these areas in the following.

1. Noisy output distributions

A natural noise model in the context of universal random circuits is given by single-qubit noise channels after each two-qubit gate in the circuit, since the fidelity of two-qubit gates is typically much worse than the single-qubit fidelity (Arute *et al.*, 2019). Assume for simplicity that the noise channel is gate independent, or that all two-qubit gates and the associated noise channel are the same, and that its average gate fidelity is given by $1 - \epsilon$. This model was analyzed by Dalzell, Hunter-Jones, and Brandão (2021) and Deshpande, Niroula *et al.* (2022) in different regimes of the parameter ϵ .

Dalzell, Hunter-Jones, and Brandão (2021) compared the output distribution of the noisy circuit p_{noisy} to the “white-noise distribution” with respect to an ideal distribution p_{ideal} . Given a fidelity F , the white-noise distribution is defined as

$$p_{\text{wn}} = Fp_{\text{ideal}} + (1 - F)p_{\text{unif}}, \quad (127)$$

where p_{unif} is the uniform distribution. Approximately sampling from the white-noise distribution p_{wn} with inverse polynomial fidelity F within TVD error ϵF is just as difficult as approximately sampling from the ideal distribution p_{ideal} within TVD error ϵ , given that p_{ideal} anticoncentrates in the sense that it has exponentially small second moments; see

Theorem 4 of Dalzell, Hunter-Jones, and Brandão (2021). Notice that achieving an inverse polynomial fidelity would still require a local error rate of $\Theta(1/n)$ for circuits of a size $O(n \log(n))$, which is the minimal size required for anti-concentration to hold; see Sec. IV.D.2.

Dalzell, Hunter-Jones, and Brandão (2021) showed that the distance of the noisy distribution approaches the uniform distribution as $e^{-2m\epsilon + O(m\epsilon^2)}$, i.e., exponentially in the circuit size. At the same time, the distance to the white-noise distribution with fidelity parameter $F = e^{-2m\epsilon + O(m\epsilon^2)}$ scales as $O(F\epsilon\sqrt{m})$ in the regime in which the noise parameter is small in the sense that $\epsilon n \log(n) \ll 1$ and the circuit family satisfies the anticoncentration property, requiring $m \in \Omega[n \log(n)]$. Since the distance to the white-noise distribution scales as a square root in the circuit size, their result showed a quadratic improvement in the required noise level for random quantum circuits as compared to the worst case for which the error would grow as $O(\epsilon m)$. To summarize, the average fidelity decay is exponential in m , and the typical distance to the corresponding white-noise distribution grows slower than the worst case. Consequently, there is now an optimal scaling of m with n that achieves the minimal error to an appropriate white-noise distribution in terms of the circuit size. It is in this regime that the cross-entropy benchmarking (XEB) fidelity translates to a TVD bound and provides the best measure of quantum advantage; see also the discussion in Sec. V.B.3.

Meanwhile, Deshpande, Niroula *et al.* (2022) showed that in the regime of large noise $\epsilon \in O(1)$ the expected total-variation distance to the uniform distribution is lower bounded by $\exp[-O(d)]$, where d is the depth of the circuit. In certain regimes this result also holds for typical instances. In light of the result of Dalzell, Hunter-Jones, and Brandão (2021) that showed a fidelity decay in the circuit size $m = nd$, this is a surprisingly slow decay.

Notice that the respective bounds translate to a concentration bound on the distance of the individual probabilities to uniform as $2^{-O(m)-n}$ and $2^{-O(d)-n}$, respectively, by a Markov bound on the TVD. We also stress that the two results consider complementary regimes and that their respective proof techniques fail beyond the considered regime. It remains an interesting open problem to analyze the entire distribution of the TVD between the noisy distribution and the uniform distribution as well as its noise dependence in more detail.

2. Fault-tolerant random sampling

As an alternative approach, one can consider the possibility of embedding quantum random sampling in a fault-tolerant encoding wherein error syndrome measurements are part of the sampling scheme. Fujii (2016) observed that sampling from the entire distribution of such an encoding remains worst-case hard in the presence of noise. This is because one may postselect on the syndrome measurements returning the no-error outcomes. In this case the conditional distribution on the sampling measurements is given simply by the ideal distribution, provided that the corresponding postselection probability is nonzero. Consequently, exact simulation of the noisy distribution remains computationally intractable in the worst case, provided that the local error rates are below the threshold for the encoding used.

Kapourniotis and Datta (2019) provided an explicit example of such an encoding in the measurement-based model of quantum computing, which also allows for an efficient verification scheme. However, it is unclear to what extent the approximate average-case hardness conjecture required for this scheme is plausible, since it is based on the postselected success of magic-state distillation. Building on the ideas of Bravyi *et al.* (2020), Mezher *et al.* (2020) developed high-dimensional and interactive measurement-based protocols in which this is achieved for every instance by appropriate classical postselection.

V. VERIFICATION

In Sec. IV, we discussed the complexity-theoretic evidence for the classical intractability of quantum random sampling. But in order to demonstrate a quantum advantage via quantum random sampling the quantum implementation must be sufficiently accurate. It is therefore essential to verify that a claimed implementation of quantum random sampling in fact achieves the purported task.

The verification task is extremely challenging, however. This is due to the difficulty of verifying sampling tasks in general, as well as the impossibility of efficiently simulating a sufficiently accurate implementation of quantum random sampling or computing the corresponding output probabilities. In this section, we review different approaches to the verification problem, both inefficient and efficient ones.

Clarifying the verification problem somewhat more formally is a first nontrivial task since there are various distinct settings in which we can conceive of verification—we might allow for interaction between a skeptic and a quantum device that is claimed to produce samples from the correct distribution, or merely claimed to perform a task that is classically not efficiently solvable. We might ask to verify the device just from the samples it produces, or we might allow access to the quantum state of the device, i.e., by performing measurements in different bases.

We begin by reviewing the reason why naive verification from samples alone is impossible in the absence of assumptions on the device simply because too many samples from the device would be required in Sec. V.A. We then move on to sample-efficient but computationally inefficient protocols for different verification settings that simply use samples from the device in Sec. V.B. Given the previous result, such protocols require assumptions on the device, or verify a weaker statement than the correctness of the samples. In Sec. V.C, we then consider the setting in which we have direct access to the output state $C|0\rangle$ of the computation. This allows fully efficient and yet rigorous certification protocols for quantum sampling schemes that assume accurate quantum measurements in certain restricted bases. Finally, we discuss verification schemes that involve several rounds of interaction between a skeptic verifier and the quantum device under investigation in Sec. V.D.

A. Hardness of verification from classical samples

In this section, we discuss a simple argument for why verifying the samples from quantum random sampling

schemes typically requires exponentially many samples and is therefore infeasible—the quantum device would need to be run exponentially many times. To this end, one can invoke the result by Valiant and Valiant (2017) on optimal identity testing and properties of the output probability distribution of quantum random sampling (Hangleiter *et al.*, 2019).

Theorem 25 (Optimal identity testing) (Valiant and Valiant, 2017).—There are constants $c_1, c_2 > 0$ such that for any $\epsilon > 0$ and any target distribution P there is a test that, given samples from a distribution Q , distinguishes whether $P = Q$ or $\|P - Q\|_{\text{TV}} > \epsilon$, when promised that one is the case, given

$$c_1 \max \left\{ \frac{1}{\epsilon}, \frac{1}{\epsilon^2} \|P_{-\epsilon/\sqrt{16}}^{-\max}\|_{\ell_{2/3}} \right\} \quad (128)$$

many samples. On the other hand, there is no such test from fewer than

$$c_2 \max \left\{ \frac{1}{\epsilon}, \frac{1}{\epsilon^2} \|P_{-2\epsilon}^{-\max}\|_{\ell_{2/3}} \right\} \quad (129)$$

samples.

For a vector of non-negative numbers P we define $P^{-\max}$ to be the vector obtained from P by setting the largest entry to zero, and $P_{-\epsilon}$ to be the vector obtained from P by setting all of the smallest entries to zero such that the sum of the removed entries is less than ϵ . Moreover, $\|P\|_{\ell_{2/3}} = (\sum_x p_x^{2/3})^{3/2}$. The $\ell_{2/3}$ norm of $P_{-\epsilon}^{-\max}$ therefore completely characterizes the asymptotic complexity of identity testing up to constant factors in ϵ . The intuition behind the result of Valiant and Valiant (2017) is that the largest probability and the tail of the distribution are easily detected in an identity test because a constant deviation in these parts of the distribution will be visible in the samples obtained with high probability. An important corollary of their result, which was known prior to it [see Goldreich (2017)], is that the complexity of testing against the uniform distribution on a sample space Ω requires $O(\sqrt{|\Omega|})$ samples, while verification requires fewer samples for more peaked distributions.

Lower bounds on the certifiability of quantum random sampling, intuitively speaking, follow from the fact that the output distributions of the schemes are extremely flat with high probability. Technically speaking, we obtain the lower bounds from bounding the $\ell_{2/3}$ norm. The second moments that were used to prove anticentration are sufficient for that. To see this, following Hangleiter *et al.* (2019), we first observe that the $\ell_{2/3}$ norm can be lower bounded in terms of the largest probability p_0 of a distribution P as

$$\|P_{-\epsilon}^{-\max}\|_{\ell_{2/3}} \geq p_0^{-1/2} (1 - \epsilon - p_0)^{3/2}. \quad (130)$$

We then observe that the Rényi-2 entropy $H_2(p) = -\log \sum_x p_x^2$ upper bounds the largest probability as

$$\log p_0 \leq -\frac{1}{2} H_2(p). \quad (131)$$

But now we can use the fact that most quantum random sampling schemes given by a circuit family \mathcal{C} have bounded

average collision probabilities (see Sec. IV.D.2) and that they concentrate around the mean by Markov's inequality as

$$\sum_x \mathbb{E}_{C \sim \mathcal{C}} [p_x(C)^2] \leq O(2^{-n}/\delta), \quad (132)$$

with a probability of at least $1 - \delta$. This implies that the Rényi-2 entropy is bounded as $H_2(P_C) \geq n + \log[O(\delta)]$. Consequently, the largest probability is exponentially small with high probability, i.e., $\log p_0 \leq -\{n + \log[O(\delta)]\}/2$. The sample complexity of certifying quantum random sampling from samples scales at least as $\Omega(2^{n/4+O(\delta)})$ with a probability of $1 - \delta$ over the choice of circuit instance. Note that even though the second moments of the boson-sampling probabilities are not sufficiently small to prove anticoncentration, they are small enough to prohibit sample-efficient verification (Gogolin *et al.*, 2013; Hangleiter *et al.*, 2019).

While one might think that this result is actually not too bad in that few enough samples may be required for intermediate-scale instances of quantum random sampling, the optimal identity test of Valiant and Valiant (2017), which employs a variant of the χ^2 test, is highly impractical in that the constants involved are much too large. In addition, the problem becomes more challenging when the test is also required to accept distributions that are not too far away from the ideal distribution. This is because this requirement poses an additional constraint on the testing protocol.

B. Sample-efficient classical verification via cross-entropy benchmarking

To overcome the obstacle of exponential sample complexity, one may consider a weaker requirement than verifying the full total-variation distance. The most prominent approach that achieves this is a family of tests, which we label *cross-entropy benchmarking*. These tests were introduced in a series of works (Aaronson and Chen, 2017; Boixo *et al.*, 2018; Neill *et al.*, 2018; Arute *et al.*, 2019). The central idea is to use multiplicative measures of similarity between the implemented “noisy” distribution Q and the ideal target distribution P_C that measure the correlation between the two distributions. We can express those measures as follows.

Definition 26 (Cross-entropy measures).—Let $f: [0, 1] \rightarrow \mathbb{R}$ be a monotonically increasing function. Define

$$F_f(Q, P_C) = \sum_{x \in \{0,1\}^n} Q(x) f(P_C(x)) \quad (133)$$

as the cross-entropy measure corresponding to f .

The first observation that we can make is that by a Chernoff bound the cross-entropy measures F_f can be sample efficiently estimated from a number of samples that depends on the variance of $f(P_C(x))$ over x and scales as $1/\epsilon^2$ in the estimation error. For exponentially small values of $F_f(Q, P_C)$ the error ϵ needs to scale inverse exponentially too. Hence, sample efficiency is lost in that case.

The second observation is that estimating cross-entropy measures is computationally inefficient for quantum advantage schemes since the probabilities $P_C(x)$ of the ideal distribution (or a function thereof) need to be computed for

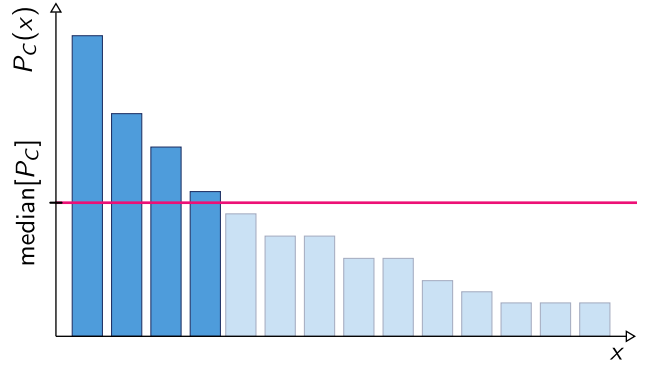


FIG. 8. In the task heavy-outcome generation (HOG) one is asked to output a list of strings $\{x_1, \dots, x_k\}$ for which $P_C(x_i) \geq \text{med}(P_C)$.

the observed outcomes. As we later see, this constitutes an important obstacle to their practical usage in verifying quantum random sampling in the quantum advantage regime.

While different variants of the measure are interpreted differently, the intuition underlying all such measures is the following: those distributions that get the heavy outcomes of a quantum computation correct will score well on cross-entropy measures because these outcomes dominate the measure (Aaronson and Chen, 2017). One can characterize heavy outcomes as those bit strings $x \in \{0, 1\}^n$ for which the probability $P_C(x)$ of obtaining x is large, for example, larger than the median of P_C ; see Fig. 8.

Before we introduce the most important measures (heavy-outcome generation, cross-entropy difference, and cross-entropy benchmarking fidelity), we discuss in more detail the shape of the outcome distribution of random quantum circuits. Consider the success probability $p_U(0) = |\langle 0|U|0\rangle|^2$ of a Haar-random unitary $U \in U(d)$. The distribution of $p = p_U(0)$ over the choice of U is given by the so-called Porter-Thomas distribution (Porter and Thomas, 1956), which is asymptotically exponentially distributed as²²

$$P_{\text{PT}}(p) = (d-1)(1-p)^d \xrightarrow{d \gg 1} d \exp(-dp). \quad (134)$$

For $d \gg 1$ one can now invoke Levy's lemma²³ (Ledoux, 2005) to see that the finite distribution of outcome

²²See Chap. 4.9 of Haake (2010) for the derivation.

²³Levy's lemma (Ledoux, 2005) can be stated as follows. Given a function $f: S^D \rightarrow \mathbb{R}$ defined on the D -dimensional hypersphere S^D with zero mean and an $x \in S^D$ chosen uniformly at random,

$$\Pr[|f(x)| \geq \epsilon] \leq 2 \exp\left(-\frac{2C(D+1)\epsilon^2}{\eta^2}\right), \quad (135)$$

where $\eta > 0$ is the Lipschitz constant of f and $C > 0$ is a constant. For normalized quantum state vectors of a complex vector space of dimension d , $D = 2d - 1$. The heuristic intuition developed here is that for random processes with an approximately constant Lipschitz constant, one would expect the fluctuation to scale approximately as the inverse square root of the dimension $d = 2^n$ of the underlying Hilbert space.

probabilities of a fixed, Haar-randomly drawn unitary is expected to be $O(1/\sqrt{d})$ close to the Porter-Thomas distribution. While exactly implementing Haar-random unitaries via a quantum circuit requires exponentially many gates, it was numerically shown by Boixo *et al.* (2018) that the output distribution of universal random circuits quickly tends toward the Porter-Thomas distribution in terms of the lower moments of the distribution. This evidence serves as justification for the use of properties of the Porter-Thomas distribution, as opposed to merely the second moments of the distribution, in the analysis of cross-entropy measures.

1. Heavy-outcome generation

The most basic cross-entropy measure that serves as intuition for the more involved measures that we later discuss is based on the so-called heavy-outcome generation (HOG) task, which was introduced by Aaronson and Chen (2017).

Problem 27 (HOG) (Aaronson and Chen, 2017).—Given as input a random quantum circuit $C \in \mathcal{C}$ from a family \mathcal{C} , generate distinct output strings x_1, \dots, x_k , at least a $2/3$ fraction of which have a probability greater than the median of P_C $\text{med}[P_C]$.

HOG is equivalent to achieving a nonzero score in the HOG fidelity

$$F_{\text{HOG}}(Q, P_C) = \frac{2}{\ln 2} \sum_{x \in \{0,1\}^n} Q(x) [\theta(P_C(x) - \text{med}[P_C]) - \frac{1}{2}], \quad (136)$$

defined in terms of the step function $\theta: \mathbb{R} \rightarrow \{0, 1\}$, which is 0 for $x < 0$ and 1 otherwise.

Because it is defined in terms of the bias of the target distribution, F_{HOG} can be sample efficiently estimated. The median can be estimated efficiently up to a small error from few samples. Given k samples $\{x_0, \dots, x_k\}$ from a noisy distribution Q , we then need to compute the probabilities $P_C(x_i)$ and compare them to the median. By Hoeffding's inequality this can be achieved with error $O(1/\sqrt{k})$ with exponentially small failure probability.

We now discuss the properties of F_{HOG} . If Q is maximally noisy (that is, the uniform distribution), then

$$F_{\text{HOG}}(Q, P_C) = \frac{2}{\ln 2} \left(\frac{1}{2^n} |\{x: P_C(x) \geq \text{med}[P_C]\}| - \frac{1}{2} \right) = 0, \quad (137)$$

as the median is defined as the largest number such that the sum of the output probabilities of C exceeding that number is at least $1/2$. On the other hand, in an ideal implementation for which $Q = P_C$, $F_{\text{HOG}}(Q, P_C) > 0$ so long as P_C is nonuniform. This is because, by definition, the probabilities above the median are larger than those below the median, and hence the probability weight above the median is at least $1/2$. More specifically, if the outcome probabilities $P_C(x)$ are Porter-Thomas distributed, then $F_{\text{HOG}}(P_C, P_C) = 1$. To see this, observe that the median of the exponential distribution is

given by $\ln 2/2^n$ and the total probability weight of P_C above the median is then given by²⁴

$$\sum_{x \in \{0,1\}^n} P_C(x) \theta(P_C(x) - \ln 2/2^n) \approx \int_{\ln 2/2^n}^{\infty} 2^n e^{-2^n p} dp = \frac{1 + \ln 2}{2}. \quad (138)$$

More generally, a distribution that scores well in terms of F_{HOG} will therefore tend to be closer to an ideal implementation of P_C in terms of total-variation distance. This is rigorously true if the noisy distribution is a convex mixture

$$Q_\lambda(x) = (1 - \lambda)P_C(x) + \lambda \frac{1}{2^n} \quad (139)$$

of the ideal target distribution and the uniform distribution with $\lambda \in [0, 1]$.

There are also distributions, however, which score well on the HOG fidelity but are far away from P_C . To see this, take the distribution that is supported on $\{x: P_C(x) \geq \text{med}[P_C]\}$. This distribution will have a HOG fidelity of $1/\ln 2 > 1$ even though its total-variation distance to P_C is at least $(1 - \ln 2)/2$.

a. Computational hardness of HOG

It is presumably difficult to find a distribution that has high support on the heavy outcomes of the target distribution though. Scoring well on F_{HOG} may thus be computationally hard even though it is a strictly easier task than approximately sampling from the target distribution. To see this, observe that the ability to sample from the correct distribution implies the ability to score well on F_{HOG} , but not vice versa since F_{HOG} does not quantify the TVD. Aaronson and Chen (2017) conjectured precisely that HOG is computationally intractable for random quantum circuits. To support this conjecture, they reduced it to the hardness of deciding whether $|\langle 0|C|0\rangle|^2$ is larger than $\text{med}[P_C]$, with a probability of at least $1/2 + \Omega(2^{-n})$ over the choice of C . The *quantum threshold assumption* (QUATH) states that this task is computationally intractable for classical computers. To reduce QUATH to HOG, we simply assume that there is an efficient routine solving HOG. Given a quantum circuit C , we can run that routine on the circuit $C' = \prod_i X_i^{z_i}$, where z is a uniformly random string. If z is contained in the k output samples, then we output YES; otherwise, we output YES with probability $1/2$ and NO otherwise. This procedure decides whether z is a heavy string for C' or, equivalently, whether 0^n is heavy for C , with success a probability of at least $1/2 + \Omega(1/2^n)$ since z is uniformly random.

Conversely, HOG can be solved by a quantum algorithm for circuits with a probability weight above the median greater than $2/3$ with a probability of at least $1 - \exp[-\Omega(k)]$. Aaronson and Chen (2017) provided a proof that this is indeed the case with high probability by showing in their

²⁴See also Aaronson and Chen (2017), footnote 3.

Lemma 8 that in expectation the probability weight above the median is lower bounded by $5/8$.

The HOG test and the HOG fidelity F_{HOG} can therefore be considered benchmarks for quantum random sampling based on evidence independent of the argument presented in Sec. IV. While HOG and QUATH may be plausible conjectures; however, the level of complexity-theoretic evidence for both QUATH and the intractability of HOG is extremely weak. This occurs because we have no independent underpinning of those conjectures such as the noncollapse of the polynomial hierarchy, which is independently grounded in significant evidence.

b. Fine-graining HOG: Binned outcome generation

A natural way to connect the properties of the HOG fidelity with the TVD is to bin probabilities in a more fine-grained fashion (Bouland *et al.*, 2019). This retains the complexity-theoretic intuition behind HOG that producing outcomes that are correlated with the ideal distribution is hard, and is also more directly supported by evidence for the intractability of simulating quantum random sampling within a constant TVD. A natural starting point for such a more fine-grained measure is to observe that HOG effectively divides the probabilities into two bins: those that are larger than the median and those that are smaller. The HOG benchmark is then obtained from testing whether the empirically obtained samples satisfy certain properties expected from ideally distributed samples on the respective bins. The sample efficiency of computing this benchmark can be retained even when generalizing it to polynomially many bins and comparing the number of observed outcomes per bin with the number of expected outcomes.

Given that the distribution of outcome probabilities is expected to be an exponential distribution, the natural way to bin is to choose a larger number of bins. Concretely, we can choose m equifilled bins $[p_i, p_{i+1})$ satisfying

$$\int_{p_i}^{p_{i+1}} 2^n e^{-2^n p} dp = \frac{1}{m} \quad (140)$$

for $i = 1, \dots, m$, $p_0 = 0$, and $p_m = 1$. Define $\Omega = \{[p_i, p_{i+1})\}_{i \in [m]}$. The task of binned outcome generation (BOG) (Bouland *et al.*, 2019) is to obtain a good, i.e., low, value of the binned distance

$$d_{\text{BOG}}(Q, P_C) = \sum_{x \in \Omega} \left| \frac{1}{2^n} \sum_{x \in \{0,1\}^n} [Q(x) - P_C(x)] \delta(P_C(x) \in X) \right| \quad (141)$$

$$= \sum_{x \in \Omega} \left| \frac{1}{2^n} \sum_{x \in \{0,1\}^n} Q(x) \delta(P_C(x) \in X) - \frac{1}{m} \right|, \quad (142)$$

where Eq. (142) is true if P_C is Porter-Thomas distributed. This is a discretized estimator of the total-variation distance of the outcome distribution and can be estimated from polynomially many samples; see Fig. 9. Indeed, for $Q = P_C$ this measure is 0, while for any $Q \neq P_C$ it converges to

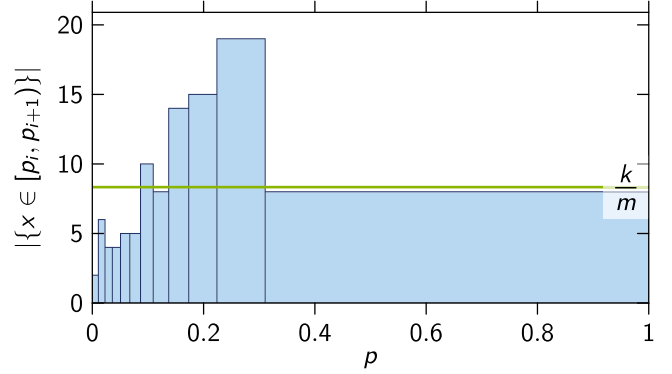


FIG. 9. The fine-grained generalization of heavy-outcome generation is to bin the samples x_1, \dots, x_k from the noisy distribution Q according to the probabilities $P_C(x_i)$. This constitutes a coarse-grained estimator of the total-variation distance between Q and P_C . Since P_C is nearly exponentially distributed for random circuits, a suitable choice of m bins $[p_i, p_{i+1})$ is such that they are equifilled with a $1/m$ fraction of the ideal samples. This is shown for a noisy exponential distribution on an ($n = 3$)-qubit sample space, with $m = 12$ bins and $k = 100$ samples.

$\|Q - P_C\|_{\text{TV}}$ as $m \rightarrow \infty$. Canonne and Wimmer (2020) proved that such *binned identity testing* with k bins up to error ϵ is possible using $O(k/\epsilon^2)$ many samples, and moreover that this is asymptotically optimal.

2. Cross-entropy difference

While HOG and its variants are conceptually intuitive, in practice we want to capture as much about the distribution as possible given the available samples. To capture correlations between the distribution Q and P_C as well as possible, an appealing measure is the cross entropy (Boixo *et al.*, 2018)

$$\text{CE}(Q, P_C) = - \sum_x Q(x) \log P_C(x). \quad (143)$$

The cross entropy is a well-known statistical measure of similarity between two distributions and measures correlations between the two distributions (Murphy, 2012). It also gives rise to a distance measure between Q and P_C , known as the cross-entropy difference²⁵

$$d_{\text{CE}}(Q, P_C) = \text{CE}(Q, P_C) - H(P_C) \quad (145)$$

$$= \sum_{x \in \{0,1\}^n} [Q(x) - P_C(x)] \log \frac{1}{P_C(x)}, \quad (146)$$

where H denotes the Shannon entropy.

²⁵Note that Boixo *et al.* (2018) defined the cross-entropy difference in terms of Eq. (133) as the deviation of cross entropy between Q and P_C from the cross entropy between the uniform distribution and P_C ,

$$F_{\text{XE}}(Q, P_C) = \text{CE}(1/2^n, P_C) - \text{CE}(Q, P_C). \quad (144)$$

But how does the cross-entropy difference fare when applied to the task of verifying quantum supremacy distributions? Using the assumption that the ideal probabilities are exponentially distributed, we observe that it constitutes a good measure for distributions of the form Q_λ in Eq. (139) (Boixo *et al.*, 2018):

$$d_{\text{CE}}(Q_\lambda, P_C) = (1-\lambda)d_{\text{CE}}(P_C, P_C) + \lambda d_{\text{CE}}(1/2^n, P_C) \quad (147)$$

$$\approx (1-\lambda)0 + \lambda 1 = \lambda. \quad (148)$$

To see why this is the case, we can compute the expectation value of $H(P_C)$ over the random choice of C as (Boixo *et al.*, 2018)

$$\mathbb{E}_C[H(P_C)] = -\sum_x \mathbb{E}_C[P_C(x) \log P_C(x)] \quad (149)$$

$$= -2^n \int_0^\infty 2^n e^{-2^n p} p \log p dp \quad (150)$$

$$= n - 1 + \gamma, \quad (151)$$

where $\gamma \approx 0.5774$ is the Euler constant. Likewise, the cross entropy between P_C and the uniform distribution is in expectation given by

$$\mathbb{E}_C[\text{CE}(1/2^n, P_C)] = -\frac{1}{2^n} \sum_{x \in \{0,1\}^n} \mathbb{E}_C[\log P_C(x)] \quad (152)$$

$$= -\int_0^\infty 2^n e^{-2^n p} \log p dp \quad (153)$$

$$= n + \gamma. \quad (154)$$

From this we obtain $\mathbb{E}_C[d_{\text{CE}}(1/2^n, P_C)] = 1$.

By the previous argument that the probabilities $P_C(x)$ for a given Haar-random and large enough unitary C are pairwise independently identically distributed according to the Porter-Thomas distribution, with high probability over the choice of C , $d_{\text{CE}}(1/2^n, P_C) = 1$ for a fixed circuit. Conversely, as the cross entropy reduces to the Shannon entropy for $Q = P_C$ we trivially have $d_{\text{CE}}(P_C, P_C) = 0$. To summarize, the cross-entropy difference attains the value 1 for the uniform distribution and vanishes for the ideal distribution, giving rise to linear interpolation (147) for states of the form Q_λ . Notice that this is equally true for any noisy distribution

$$Q'_\lambda = (1-\lambda)P_C + \lambda Q', \quad (155)$$

in which the uniform distribution is replaced by a distribution Q' that is uncorrelated with P_C , i.e., $\mathbb{E}_C[\text{CE}(Q', P_C)] = -\sum_x Q'(x) \mathbb{E}_C[\log P_C(x)]$.

Under certain conditions the cross-entropy difference in fact bounds the total-variation distance (Bouland *et al.*, 2019). To see this, notice that the definition of the cross-entropy difference is similar to that of the Kullback-Leibler divergence

$$D_{\text{KL}}(Q||P_C) = \text{CE}(Q, P_C) - H(Q), \quad (156)$$

which bounds the total-variation distance by Pinsker's inequality as

$$\|Q - P_C\|_{\text{TV}} \leq \sqrt{D_{\text{KL}}(Q||P_C)/2}. \quad (157)$$

Hence, if the cross-entropy difference satisfies $d_{\text{CE}}(Q, P_C) \leq \epsilon$ and the noise is entropy increasing such that $H(Q) \geq H(P_C)$, we have

$$\|Q - P_C\|_{\text{TV}} \leq \sqrt{D_{\text{KL}}(Q||P_C)/2} \quad (158)$$

$$\leq \sqrt{d_{\text{CE}}(Q, P_C)/2} \leq \sqrt{\epsilon/2}. \quad (159)$$

The condition $H(Q) \geq H(P_C)$ is a fairly general condition on the type of noise under which the total-variation distance bound (158) holds. But it is also a condition that cannot be checked from fewer than exponentially many samples from Q . Moreover, one can easily construct examples of distributions that violate the inequality (158) (Bouland *et al.*, 2019). Those examples fare well on the cross-entropy difference but are far from the ideal target distribution.

The cross-entropy difference can be efficiently estimated up to accuracy ϵ with failure probability α from

$$m \geq \frac{[n + O(\log n)]^2}{2\epsilon^2} \log(2/\alpha) \quad (160)$$

many independently identically distributed (iid) samples from Q . To derive Eq. (160), we apply Hoeffding's inequality and assume that the probabilities $P_C(x)$ are Porter-Thomas distributed. We obtain that, with a probability of at least $1 - 1/f(n)$ over the choice of U , the probabilities $P_C(x)$ satisfy

$$2^{-2n}/f(n) \leq P_C(x) \leq [n + \log f(n)]2^{-n}, \quad (161)$$

such that their logarithms $\log P_C(x)$ differ only by a constant factor of $\sim(2 + O[\log(f(n))])$ from $-n$.

3. Linear cross-entropy benchmarking fidelity

The most widely used cross-entropy benchmark is the XEB fidelity introduced by Arute *et al.* (2019). This measure simply chooses f to be the identity function, up to rescaling and shifting, $f_{\text{XEB}}(x) = 2^n x - 1$, such that

$$F_{\text{XEB}}(Q, P_C) = \sum_{x \in \{0,1\}^n} Q(x)[2^n P_C(x) - 1]. \quad (162)$$

The XEB fidelity has the virtue that it can be meaningfully applied in two variants: in the first variant, it is a variant of a randomized-benchmarking protocol with the goal of obtaining a fidelity measure averaged over random sequences of quantum gates. This variant is a special instance of randomized benchmarking (Helsen *et al.*, 2019, 2022; Y. Liu *et al.*, 2021) and can be applied to gates acting on a few qubits (Arute *et al.*, 2019). In its second reading, it can be used as a verification protocol for single instances of quantum random

sampling. By making use of a typicality argument based on Levy’s lemma, guarantees for the average randomized-benchmarking behavior can be transferred to the single-instance application. Therefore, the XEB fidelity unifies the idea of benchmarking a quantum processor by running random computations on it with the idea of demonstrating a quantum advantage via sampling from the output distribution of such circuits.

Even though it may serve as a measure of the fidelity of a single circuit instance for a large number of qubits, the XEB fidelity is intrinsically an average-case measure, and its ability to verify single instances is derived merely from the fact that these instances are typical. Given the choice of rescaling and shifting, the average XEB fidelity for a family of quantum circuits \mathcal{C} that gives rise to a spherical 2-design (recall Sec. IV.D.2) indeed gives rise to a meaningful measure of quantum advantage in the sense that

$$\begin{aligned} & \mathbb{E}_{\mathcal{C} \sim \mathcal{C}}[F_{\text{XEB}}(Q_C, P_C)] \\ &= \begin{cases} \sum_x 2^n \mathbb{E}_C[P_C(x)^2] - 1 \approx 1 & Q_C = P_C, \\ \sum_x P_C(x) - 1 = 0 & Q_C = 1/2^n \end{cases} \quad (163) \end{aligned}$$

in the extreme cases in which, for every C , Q_C is the ideal target distribution and the uniform distribution, respectively. In the following, we discuss in more detail these interpretations of the XEB fidelity, and the extent to which the XEB provides a meaningful measure of quantum advantage.

a. Sample complexity of estimating the XEB fidelity

For Haar-random unitaries, F_{XEB} can be estimated up to error ϵ with a probability of at least $1 - \delta$ from

$$\ell \geq \frac{e^2}{2e^2} \ln^2 \left(\frac{2}{2\delta} \right) \ln \left(\frac{2}{\delta} \right) \quad (164)$$

many samples (Hangleiter, 2021; Kliesch and Roth, 2021). Moreover, using the bounds (161) on the size of the probabilities $P_C(x)$, we can estimate the average XEB fidelity $\mathbb{E}_C[F_{\text{XEB}}(Q, P_C)]$ up to error 2ϵ with failure probability δ from

$$\ell_C \geq \frac{1}{2e^2} \log \frac{2}{\delta} \quad (165)$$

many distinct random circuits and

$$\ell \geq \frac{[n + O(\log n)]^2}{2e^2 / \ell_C^2} \log(2/\delta) \quad (166)$$

many samples per circuit (Hangleiter, 2021). In fact, an $O(1)$ bound on the variance of $\mathbb{E}[F_{\text{XEB}}]$ is true even if only the third moments of the circuit are close to the Haar-random value and the noise is gate independent (Helsen *et al.*, 2022).

b. Benchmarking via XEB fidelity

We now sketch how XEB can be used to benchmark a quantum device. For instance, Arute *et al.* (2019) analyzed how to estimate the depolarization error p_c per cycle of the

computation using the XEB fidelity. We now follow their argument. Consider the noisy quantum state

$$\rho_C = \epsilon_d C|0\rangle\langle 0|C^\dagger + (1 - \epsilon_d)\chi_C \quad (167)$$

after applying a random circuit C with d gate layers; see Eq. (75). In Eq. (167) ϵ_d describes the effect of errors on the state and in the case of $\chi_C = \mathbb{1}/2^n$ is interpreted as the depolarization fidelity. We assume now that the erroneous state χ_C is uncorrelated with C in the sense that the probabilities of a computational-basis measurement are uncorrelated as $\mathbb{E}_C[\langle x|\chi_C|x\rangle\langle x|C|0\rangle\langle 0|C^\dagger|x\rangle] = \mathbb{E}_C[\langle x|\chi_C|x\rangle]\mathbb{E}_C[\langle x|C|0\rangle\langle 0|C^\dagger|x\rangle]$.

When averaging or “twirling” over random unitaries that form a unitary design, we would then expect to obtain a fully mixed state

$$\mathbb{E}_C[C^\dagger \chi_C C] = \frac{\mathbb{1}}{2^n} \quad (168)$$

such that one might expect

$$\mathbb{E}_C[C^\dagger \rho_C C] = \bar{\epsilon}_d |0\rangle\langle 0| + (1 - \bar{\epsilon}_d) \frac{\mathbb{1}}{2^n}, \quad (169)$$

where $\bar{\epsilon}_d$ denotes the average of the individual values of ϵ_d over the random choice of unitaries. Equation (169) precisely describes the effect of a depolarizing channel acting in each cycle of the computation with depolarization fidelity p_c such that $p_c^d = \bar{\epsilon}_d$.

We obtain an expression of the circuit-averaged XEB fidelity in terms of the depolarization fidelity

$$\mathbb{E}_C[F_{\text{XEB}}(Q, P_C)] = p_c^d \left(2^n \sum_x \mathbb{E}_C[P_C(x)^2] - 1 \right), \quad (170)$$

where Q is the output distribution of the noisy state ρ_C and P_C is as usual the output distribution of $C|0\rangle$. We can now use Eq. (170) in order to estimate p_c from $F_{\text{XEB}}(Q, P_C)$. To do this, we classically estimate the quantity in brackets in Eq. (170) and obtain

$$p_c^d \approx \frac{\widehat{F_{\text{XEB}}}(Q, P_C)}{2^n \sum_x \mathbb{E}_U[P_C(x)^2] - 1}, \quad (171)$$

where $\widehat{F_{\text{XEB}}}(Q, P_C)$ denotes the empirical estimate of $F_{\text{XEB}}(Q, P_C)$ for a fixed circuit and $\overline{F_{\text{XEB}}}(Q, P_C)$ denotes the empirical average over random circuits. From an exponential fit of p_c^d for various values of d one can now estimate p_c .

Notice that in writing Eq. (169), we have used the average XEB fidelity $\overline{F_{\text{XEB}}}$ as a proxy for the average fidelity \bar{F} of the quantum state. Arguments for why the assumption that the noise is uncorrelated from the circuit should be true are essential for substantiating that connection.

Y. Liu *et al.* (2021) provided further credence to the connection between average fidelity and average XEB fidelity by performing numerical simulations. They also further substantiated the claim that the model of Arute *et al.* (2019) is valid, even in certain cases in which their

uncorrelated noise assumption does not hold. To this end, they considered “random circuit sampling benchmarking” in the spirit of randomized benchmarking. Specifically, they formalized the protocol of [Arute et al. \(2019\)](#) by estimating the average *quantum fidelity* $\overline{F_d}$ of quantum circuits of increasing depth $d = 1, \dots, D$, and finally performing an exponential fit $F = Ae^{-\lambda d}$. If (a) the average fidelity is in fact well fitted by a single exponential decay and (b) the average XEB fidelity is a good proxy of the average quantum fidelity, then this model matches XEB benchmarking as performed by [Arute et al. \(2019\)](#).

[Y. Liu et al. \(2021\)](#) made the connection (a) by proving the following: Consider random circuits that comprise layers of arbitrary non-Clifford gates (say, the two-qubit *i*SWAP gates) and single-qubit Haar-random gates.²⁶ Now suppose that every layer of non-Clifford (two-qubit) gates comes with a Pauli noise channel $\mathcal{N}(\rho) = \sum_{\alpha \in \{0,1,2,3\}^n} p_\alpha \sigma_\alpha \rho \sigma_\alpha$, where σ_α denote the n -qubit Pauli matrices and p_α are their coefficients. The average fidelity $\mathbb{E}F_d$ of depth- d circuits does in fact decay exponentially in the total error $\lambda = \sum_{\alpha \neq 0} p_\alpha$ in the sense that $e^{-\lambda d} \leq \mathbb{E}F_d \leq e^{-\lambda d}(1 + K\lambda)$ for $d \ll 2^n$ up to a first-order approximation in λ ; see also the related discussion of [Helsen et al. \(2022\)](#).

For the second connection (b), they performed numerical simulations for various noise models. To this end, they made use of a somewhat more versatile fidelity estimator that is closely related to the XEB fidelity that was introduced by [Rinott, Shoham, and Kalai \(2022\)](#).²⁷ Intuitively speaking, in this “unbiased XEB” estimator, instead of multiplying the ideal probability by $1/2^n$, one multiplies it by the inverse second moments of the ideal output distributions

$$F_{\text{XEB},u}(Q, P_C) = \frac{F_{\text{XEB}}(Q, P_C)}{\mathbb{E}_C[F_{\text{XEB}}(Q, P_C)]}. \quad (172)$$

This means that it is normalized on average to unity not only for deep quantum circuits that have designlike moments [recall Eq. (163)] but also for more shallow circuits with differing second moments. [Y. Liu et al. \(2021\)](#) found good agreement between the fidelity and their unbiased XEB fidelity for various correlated noise models and, moreover, showed that the variance of the XEB fidelity scales as $O[1/\ell + \lambda^2(\mathbb{E}F)^2]$ in the number of samples ℓ collected per circuit. The unbiased estimator (172) was recently tested as a measure of fidelity in small instances of measurement-based quantum random sampling ([Ringbauer et al., 2022](#)).

Note also that the maximum-likelihood estimator (MLE) for the fidelity was analyzed by [Rinott, Shoham, and Kalai \(2022\)](#). They found that the MLE had a smaller bias and variance than the linear XEB estimator and, like the unbiased XEB estimator, was therefore a better fidelity estimator. They also found—as noted by [Arute et al. \(2019\)](#)—that, in the regime of small depolarization fidelity $\epsilon_d \ll 1$, the XEB fidelity estimator converged to the MLE of the fidelity.

²⁶This is the setup of a cycle benchmarking protocol ([Erhard et al., 2019](#)).

²⁷Unbiased estimators for other scenarios were discussed by [Y. Liu et al. \(2021\)](#) and [Choi et al. \(2023\)](#).

c. Single-instance verification

When the number of qubits is large and the unitary C is drawn Haar randomly, Levy’s lemma implies that the fluctuations around the expectation value over C [Eq. (163)] are expected to be $O(1/\sqrt{2^n})$. Consequently, for a large number of qubits, the fidelity concentrates around its expected value over the choice of random circuits ([Arute et al., 2019](#)).

For a large number of qubits, following [Arute et al. \(2019\)](#) we again write the noisy implementation of the quantum state $C|0\rangle$ as

$$\rho_C = FC|0\rangle\langle 0|C^\dagger + (1-F)\chi_C, \quad (173)$$

where the mixed state χ_C describes the effect of noise and $F = \langle 0|C^\dagger \rho_C C|0\rangle$ is the fidelity of ρ_C and the target state $C|0\rangle$. We can now make the assumption that χ_C is uncorrelated from $C|0\rangle$ in the sense that ([Arute et al., 2019](#))

$$\sum_x \langle x|\chi_C|x\rangle f(p_C(x)) = \frac{1}{2^n} \sum_x f(p_C(x)) + \epsilon \quad (174)$$

for $\epsilon \ll F$. By the Levy’s lemma argument, [Arute et al. \(2019\)](#) expected a typical fluctuation $\epsilon \in O(1/\sqrt{2^n})$.

Large parts of the analysis of the theoretical proposal of random circuit sampling ([Boixo et al., 2018](#)) and the experimental realization thereof ([Arute et al., 2019](#)) are indeed dedicated to validating the assumption of uncorrelated noise. This can be done by numerically studying realistic error models such as random Pauli errors. To summarize, given that the previously sketched arguments hold, the XEB fidelity quantifies the fidelity $F_{\text{XEB}}(Q, P_C) = F$ up to a deviation of the order of $1/\sqrt{2^n}$.

d. Difficulty of achieving a nontrivial XEB fidelity

As with HOG, we expect that achieving an exponentially small score in the XEB fidelity $b/2^n$ for constant $b > 1$, formalized as the task XHOG, is computationally hard. This is because intuitively XHOG is a refined version of HOG in which the outcomes have to be produced according to their actual weight. Analogously to the argument reducing HOG to QUATH ([Aaronson and Chen, 2017](#)), XHOG can be reduced to an analogous conjecture XQUATH ([Aaronson and Gunn, 2019](#)). XQUATH states that given a circuit $C \sim \mathcal{C}$, there is no efficient classical algorithm that produces an estimate p of $p_C(0)$ such that

$$\mathbb{E}\{[p_C(0) - p]^2\} = \mathbb{E}\{[p_C(0) - 2^{-n}]^2\} - \Omega(2^{-3n}), \quad (175)$$

where the expectation is taken over the choice of random circuit and the algorithm’s internal randomness.

e. Spoofing the linear XEB fidelity

To summarize the previous discussion, the XEB fidelity serves two distinct functions ([Gao et al., 2021](#)). First, the argument of [Aaronson and Gunn \(2019\)](#) suggested that achieving a nontrivial XEB value is a computationally intractable task for random quantum circuits. Second, the

XEB fidelity serves as a proxy for the quantum fidelity (Arute *et al.*, 2019; Y. Liu *et al.*, 2021; Choi *et al.*, 2023).

Zhou, Stoudenmire, and Waintal (2020) and Gao *et al.* (2021) observed, however, that the XEB fidelity in fact overestimates the quantum fidelity in certain settings, leading to weaknesses that can be exploited by an adversarial classical simulator. More concretely, Gao *et al.* (2021) characterized the conditions under which the XEB fidelity serves as a good proxy of the quantum fidelity when comparing a noisy quantum device to an ideal circuit. Based on these conditions, they demonstrated that the XEB fidelity is not a reliable measure of quantum advantage in an “adversarial setting” in which these conditions can be violated.²⁸ The explicit argument of Zhou, Stoudenmire, and Waintal (2020) and Gao *et al.* (2021) was based on three properties of the XEB fidelity that make it distinct from the fidelity.

First, the fidelity and the XEB fidelity exhibit different scaling behavior as multiple quantum systems are combined into a larger one: whereas the quantum fidelity generally decreases exponentially in the number of combined systems, the XEB fidelity generally increases. To see this, consider k disjoint n -qubit quantum systems with XEB fidelity values $\chi_i = 2^n \sum_x q_i(x) p_i(x) - 1$ and fidelities F_i for $i = 1, \dots, k$, where q_i and p_i are the output probabilities corresponding to the respective noisy and ideal circuits. The fidelity then scales multiplicatively as $F = \prod_i F_i$, whereas the total XEB fidelity scales as

$$\chi = 2^{kn} \sum_{x_i} \prod_i p_i(x_i) q_i(x_i) - 1 \quad (176)$$

$$= \prod_i (\chi_i + 1) - 1 \approx \sum_i \chi_i, \quad (177)$$

assuming that $\chi_i \ll 1$. This difference in scaling behavior is fundamental to the fact that the first term of the XEB fidelity tends toward a nonzero value (namely, unity) as p and q become uncorrelated from one another, which is explicitly subtracted.

Second, their values may be distinct for highly correlated errors. To see this intuitively, consider a noisy quantum circuit with m gates and independently and homogeneously distributed random errors across the circuit at rate e . The probability that no error occurs is then given by $(1 - e)^m$. If the presence of one or more errors leads to vanishing contributions to the XEB or the fidelity, then both will be equal to $(1 - e)^m$. However, outside of some limiting cases, there are nonzero correction terms for finite-size systems. Consider a single bit-flip error at depth t in a 1D random circuit. In the Heisenberg picture, we can propagate $X(t)$ backward in time and consider its effect on the initial state $|0^n\rangle$. If the dynamics are chaotic, then $X(t)$ becomes a linear combination of $4^{|s|}$ Pauli strings, the support of which grows linearly as $|s| \approx 2ct$ with an effective “scrambling velocity” c . But out of those operators $\sim 2^{|s|}$ are products of $\mathbb{1}$ and Z , and hence they do not cause an error on the input state $|0^n\rangle$. Consequently, a single error

contributes $O(2^{-2ct})$ to the XEB fidelity and quantum fidelity alike. Conversely, we can forward propagate the error, but now the argument holds only for the XEB fidelity because measurements are performed in the Z basis, while all terms contribute to the quantum fidelity, leading to a distinct behavior. Gao *et al.* (2021) further argued that this difference can be amplified when considering specific spatial error patterns and provided a lower bound on the total correction.

In the complementary “benign setting” of errors distributed independently and homogeneously across the system, they found necessary and sufficient conditions for the XEB fidelity and the quantum fidelity to agree, namely, that $nef(c) \ll 1$, where $f(c) \in O(1)$ is a decreasing function depending on the architecture details. Via a mapping to a statistical-mechanics model analogous to the one introduced in Sec. IV.D.2, they derived a diffusion-reaction for how errors evolve in the circuit and analyzed it for different ensembles of random gates. Using this model, they explored the intuition just described quantitatively, finding that the XEB fidelity starts to deviate from the fidelity for strong noise.

Third, because the XEB fidelity quantifies the correlations between the distribution q and p , complete knowledge of p allows one to amplify those correlations by choosing q adversarially.

Building on those insights as well as a spoofing algorithm of the XEB fidelity for low-depth quantum circuits (Barak, Chou, and Gao, 2021), Gao *et al.* (2021) constructed an algorithm that achieves high scores for large quantum circuits. The key idea of this algorithm is to approximate the ideal circuit with a circuit that is given by a product over smaller subsystems, each of which can be simulated on a classical computer. To achieve this, given a number of subsystems to divide the circuit in, they removed entangling gates between those subsystems. Using the algorithm, they achieved a score of 1.85×10^{-4} in 0.6 s on a single graphics processing unit (GPU), while the experiment at Google by Arute *et al.* (2019) achieved 2.24×10^{-3} , and a similar ratio for the larger follow-up experiments was attained at USTC (Wu *et al.*, 2021; Zhu *et al.*, 2022). They found, however, that for small system sizes the ratio between the performance of their algorithm and the experimental score increases and conjectured that their algorithm will achieve an advantage over the quantum value of the XEB fidelity. Relating to the hardness argument of Aaronson and Gunn (2019), their algorithm seems to refute the XQUATH conjecture. More concretely, Gao *et al.* (2021) showed that for 1D circuits their algorithm achieves a XEB fidelity that scales inversely exponentially $e^{-O(d)}$ in the circuit depth.²⁹ On the other hand, they showed that the XEB score of a variant of their algorithm precisely reflects the statement of the XQUATH conjecture in terms of probability estimation on average. Consequently, their results refute the XQUATH conjecture for circuits of sublinear depth $d \in o(n)$.

Given this discussion, achieving a quantum advantage in terms of cross-entropy benchmarking via quantum random sampling boils down to the question as to whether the inverse-exponential scaling of the quantum score of the linear XEB fidelity can be beaten by another inverse-exponential scaling

²⁸A similar overestimation of the fidelity has also been observed in the literature on randomized benchmarking (Boone *et al.*, 2019).

²⁹A slightly weaker statement also holds for 2D circuits.

of a classical algorithm. And it seems that it can. A different way of benchmarking quantum advantage experiments from the linear XEB fidelity thus seems to be necessary to demonstrate that quantum devices are in fact able to scalably outperform classical algorithms in an adversarial setting. To this end, note that the spoofing algorithm of Zhou, Stoudenmire, and Waintal (2020) and Gao *et al.* (2021) presumably do not work for the cross-entropy difference as it intrinsically builds on the linearity of the linear XEB fidelity. In spite of the results for the linear XEB, the cross-entropy difference remains a potentially valid means of benchmarking quantum advantage.

We again stress, however, that while XEB measures may be estimated from a few samples, all variants of XEB suffer from the problem that their evaluation is computationally inefficient. This limits their practical usage to a regime just below the quantum advantage threshold in which classically computing the output probabilities is still possible, but at a high cost. Alternatively, as we see in Sec. VI, other quantities might be used in order to feasibly obtain an estimate of XEB measures. In the following, we discuss an alternative approach that does not suffer from the conceptual—in terms of quantifying quantum advantage—or computational—in terms of its efficient evaluation—disadvantages of XEB.

C. Efficient quantum verification

An approach that is both natural in an experimental setting and a direct follow-up of the previous discussion regarding the relation between XEB fidelity and quantum fidelity is to verify the sampling task directly on the level of the quantum state. This is reasonable: In an experimental setting, we know that there is a quantum state on which measurements are performed. Therefore, we can exploit access to that quantum state in order to circumvent the no-go result of Sec. V.A and potentially achieve fully efficient verification of the TVD between the experimental and the target distribution, assuming that the measurements are carried out correctly.

Verification of a quantum state is possible if we have access to an ideal state preparation via a swap test, or by verification protocols that use measurements along the direction of the target state (Pallister, Linden, and Montanaro, 2018). But, assuming that this capacity would already assume the ability to prepare the ideal target state, a reasonable quantum protocol for verifying quantum random sampling schemes should therefore make use of restricted quantum capacities only, such as the ability to implement single-qubit measurements or to prepare single-qubit states reliably. Experimentally, such assumptions are extremely well justified: in most platforms single-qubit gate fidelities are orders of magnitude better than entangling-gate fidelities. It is also entirely different in kind when compared to assumptions on the global effect of the noise on the outcome probability distribution P_C such as the assumption $H(Q) \geq H(P_C)$ that was necessary for a cross-entropy-based test to yield bounds on the total-variation distance: it is an assumption on single-qubit measurements and therefore local. This means that it can be verified to the same degree that one can characterize those measurement apparatuses. For single- or two-qubit measurements this is

possible using a tool such as gate set tomography (Blume-Kohout *et al.*, 2013, 2017; Merkel *et al.*, 2013; Greenbaum, 2015; Cerfontaine, Otten, and Bluhm, 2020; Helsen *et al.*, 2021; Brieger, Roth, and Kliesch, 2023) or the device-independent verification of quantum processes and instruments (Sekatski *et al.*, 2018).

In contrast to classical verification from samples where we were given classical samples from an *a priori* untrusted device, we can conceptualize quantum verification as the task to verify the preparation of a certain quantum state with a deep circuit using components of the device that are well characterized and known to work correctly.

In the following we see protocols that are able to verify or estimate the quantum fidelity between two quantum states σ and $|\psi\rangle\langle\psi|$,

$$F(\sigma, |\psi\rangle\langle\psi|) = \langle\psi|\sigma|\psi\rangle. \quad (178)$$

Via the Fuchs–van de Graaf inequality, the fidelity bounds the TVD via the trace distance

$$\|p_\sigma - p_\psi\|_{\text{TV}} \leq \|\sigma - |\psi\rangle\langle\psi|\|_{\text{Tr}} \leq \sqrt{1 - F(\sigma, |\psi\rangle\langle\psi|)}, \quad (179)$$

where p_σ and p_ψ are the output distributions of σ and $|\psi\rangle\langle\psi|$ in the standard basis, respectively.

Generally, we can think about such protocols in terms of their information gain versus their complexity in terms of number of measurements and distinct measurement settings as well as assumptions made in the derivation of the protocol (Eisert *et al.*, 2020). While protocols with low complexity tend to yield little information about an underlying quantum state, protocols with higher complexity can reveal more information about that state. In the following, we discuss two types of protocols to verify the output states of quantum random sampling via the fidelity: fidelity witnessing and fidelity estimation.

1. Fidelity witnessing

We call an observable W a fidelity witness for a target state ρ if (Gluza *et al.*, 2018) (i) $\text{Tr}[\sigma W] = 1$ iff $\rho = \sigma$ and (ii) $\text{Tr}[\sigma W] \leq F(\rho, \sigma)$. Conceptually speaking, fidelity witnesses are much like entanglement witnesses (Gühne and Tóth, 2009) in that they cut a hyperplane through quantum state space, which detects a property of quantum states: Those states that lie on the left of the hyperplane defined by $\text{Tr}[W\sigma] \geq F_T$ are guaranteed to have a high fidelity of at least F_T since $\text{Tr}[W\sigma]$ lower bounds $F(\rho, \sigma)$. For those states on the right of the hyperplane (satisfying $\text{Tr}[W\sigma] < F_T$) we cannot make a statement about their fidelity. Conversely, though, all states σ with low fidelity $F(\rho, \sigma) \leq F_T$ are guaranteed to lie to the right of the hyperplane as $\text{Tr}[W\sigma] \leq F(\rho, \sigma) \leq F_T$. We illustrate the idea of a fidelity witness in Fig. 10.

a. Fidelity witnessing via parent Hamiltonians

A simple fidelity witness $W_H = \mathbb{1} - H/\Delta$ can be constructed for the ground state of a Hamiltonian H with gap Δ (Cramer *et al.*, 2010; Hangleiter *et al.*, 2017). To see this, one can simply expand the Hamiltonian with ground state energy

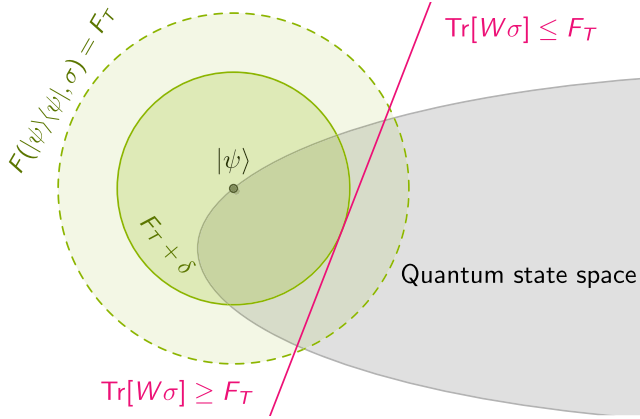


FIG. 10. Given a target state $\rho = |\psi\rangle\langle\psi|$, a fidelity witness W for ρ provides a lower bound on the fidelity $F(\rho, \sigma) \geq \text{Tr}[W\sigma]$ so that, in particular, all states σ such that $F(\rho, \sigma) \leq F_T$, it also holds that $\text{Tr}[W\sigma] \leq F_T$. Conversely, all states σ satisfying $\text{Tr}[W\sigma] \geq F_T$ will also satisfy $F(\rho, \sigma) \geq F_T$. There is a gap $\delta \geq 1 - F_T$ such that all states σ with fidelity $F(\rho, \sigma) \geq F_T + \delta$ lie on the left side of the witness.

set to 0 in its eigenbasis $|i\rangle$ with eigenvalues λ_i in order to bound the fidelity between the ground state $|0\rangle\langle 0|$ and a state preparation ρ using

$$\begin{aligned} \text{Tr}(H\sigma) &= \sum_{i=1}^d \lambda_i \text{Tr}(|i\rangle\langle i|\sigma) \geq \Delta \sum_{i=1}^d \text{Tr}(|i\rangle\langle i|\sigma) \\ &= \Delta[1 - \text{Tr}(|0\rangle\langle 0|\sigma)] = \Delta[1 - F(|0\rangle\langle 0|, \sigma)]. \end{aligned} \quad (180)$$

To apply this witness, it is required to have knowledge of both the ground state energy and the gap of the Hamiltonian in question. Applying this fidelity witness to quantum random sampling, Hangleiter *et al.* (2017) observed that arbitrary quantum computations and, in particular, those required for quantum random sampling can be embedded in the ground state of a frustration-free, local Hamiltonian via the Feynman-Kitaev history state construction. This protocol finds a particularly natural application in the measurement-based model of quantum computation, which is universal for quantum computation (Raussendorf and Briegel, 2001; Raussendorf, Browne, and Briegel, 2003). Since the prepared quantum state in measurement-based quantum computing is a stabilizer state, it is the ground state of a local, commuting Hamiltonian with gap 2 comprising the stabilizers, which are product operators. The state preparations of quantum random sampling schemes in the measurement-based model can therefore be verified via fidelity witnessing using only trusted single-qubit measurements (Gao, Wang, and Duan, 2017; Bermejo-Vega *et al.*, 2018).

We now illustrate this point and define the cluster state on N qubits on a lattice as

$$|\text{CS}\rangle = \left(\prod_{\langle i,j \rangle} \text{CZ}_{i,j} \right) H^{\otimes N} |0^N\rangle, \quad (181)$$

where the symbol $\langle i, j \rangle$ denotes nearest neighbors on a lattice. Arbitrary quantum computations can be driven by single-qubit

operations on that state: adaptive measurements at the correct angles in the x - y plane (multiples of $\pi/8$ suffice) (Mantri, Demarie, and Fitzsimons, 2017). Assuming highly accurate single-qubit operations and measurements, we can now use the fidelity witness in order to verify the premeasurement quantum state $|\text{CS}\rangle$.

To do so, we need to derive a “parent Hamiltonian” that has $|\text{CS}\rangle$ as its ground state. This can be easily done by observing that the diagonal Hamiltonian

$$H_0 = - \sum_{i=1}^N Z_i \quad (182)$$

has the all-zero state $|0^N\rangle$ as its ground state with ground state energy $E_0 = -N$ and gap $\Delta = 2$. Our strategy to derive a parent Hamiltonian H of $|\text{CS}\rangle$ is based on the observation that conjugation by unitary transformations U preserves the eigenvalues such that $U|0^N\rangle$ is a ground state of UH_0U^\dagger with ground state energy E_0 and gap Δ . Inserting $U = \left(\prod_{\langle i,j \rangle} \text{CZ}_{i,j} \right) H^{\otimes N}$ and using the relation $\text{CZ}(X \otimes \mathbb{1})\text{CZ} = X \otimes Z$, we obtain that the Hamiltonian

$$H = - \sum_{i=1}^N \left(X_i \prod_{j \in \partial i} Z_j \right) = - \sum_{i=1}^N S_i \quad (183)$$

is a parent of $|\text{CS}\rangle$ with ground state energy $E_0 = -N$ and gap $\Delta = 2$. In Eq. (183) $\partial i = \{j \in V : (i, j) \in E\}$ denotes the neighborhood of site i on a graph $G = (V, E)$. The operators $S_i = X_i \sum_{j \in \partial i} Z_j$ are often called stabilizers of $|\text{CS}\rangle$. The same applies if we rotate the cluster state locally prior to a computational-basis measurement.

The fidelity witness has also been applied to the verification of IQP circuits the diagonal part of which comprises Z , CZ , and the non-Clifford CCZ gate defined in Eq. (5) (Miller, Sanders, and Miyake, 2017). While the resulting nonlocal stabilizers h_i are not directly products of Pauli operators in the same way as we obtained $\text{CZ}(X \otimes \mathbb{1})\text{CZ} = X \otimes Z$, Miller, Sanders, and Miyake (2017) showed that single-qubit Pauli- X and Pauli- Z measurements suffice to measure those stabilizers. More precisely, a measurement of the stabilizer h_i can be achieved by measuring $X_i \prod_{j \neq i} Z_j$ with outcome $v = (v_1, \dots, v_n)$ and returning $(-1)^{\partial_i f(v) + v_i}$, where $\partial_i f(x) = f(x_1, \dots, x_i + 1, \dots, x_n) - f(x_1, \dots, x_i, \dots, x_n)$.

b. Fidelity witnesses for weighted graph states

Efficient fidelity estimation protocols for arbitrary weighted graph states as they are generated by the IQP circuit C_W with arbitrary weights $w_{i,j}$ were developed by Morimae, Takeuchi, and Hayashi (2017), Hayashi and Takeuchi (2019), and Zhu and Hayashi (2019). Those circuits have been seen to give rise to graph states in which not only vertices (as in the previous example) but also edges have arbitrary weights, so-called weighted graph states.

c. Fidelity witnesses for quantum optical states

Another approach to constructing fidelity witnesses was discovered by Chabaud, Grosshans *et al.* (2021) in the context

of linear-optical state preparations as a means to verify the output state of the boson-sampling protocol given by $\varphi(U)|1_n\rangle$ [Eq. (8)], where U is a Haar-random linear-optical unitary. They observed that if certain Gaussian measurements are performed on the state $\varphi(U)|1_n\rangle$, then one can efficiently simulate the effect of the linear-optical unitary in the post-processing. Specifically, consider a single-mode heterodyne measurement with POVM elements $|\alpha\rangle\langle\alpha|$, where $|\alpha\rangle\alpha = e^{a\hat{a}^\dagger - \bar{a}a}|0\rangle$ is a coherent state. The effect of a linear-optical unitary multimode heterodyne POVM element $\pi^{-m} \prod_i |\alpha_i\rangle\langle\alpha_i|$ is to transform it into another element $\pi^{-m} \prod_i |\beta_i\rangle\langle\beta_i|$, where $\varphi(U) \prod_i |\alpha_i\rangle = \prod_i |\beta_i\rangle$ and the values of β_i are efficiently computable. This idea can be used to verify a noisy state preparation σ of $\varphi(U)|1_n\rangle$ by performing heterodyne measurements, obtaining outcomes α_i and reinterpreting the outcomes as β_i . Now we observe that the fidelity of the quantum state σ with a pure product state $\psi = \prod_i |\psi_i\rangle\langle\psi_i|$ can be bounded as

$$F(\psi, \sigma) \geq 1 - \sum_{i=1}^m [1 - F(\psi_i, \sigma_i)] \geq 1 - m[1 - F(\psi, \sigma)], \quad (184)$$

where $\rho_i = \text{Tr}_{1, \dots, m \setminus \{i\}} \rho$ is the reduced state of ρ on the i th mode. This reduces the verification problem to estimating the single-mode fidelities $F(\psi_i, \sigma_i)$. Chabaud, Grosshans *et al.* (2021) showed that this is possible using only heterodyne measurements on the state σ_i if ψ_i has bounded support in the Fock basis. The second inequality in Eq. (184), moreover, shows that the witness has a certain robustness to noise.

A similar protocol for Gaussian states, and hence Gaussian boson sampling, was developed by Aolita *et al.* (2015). In this protocol, a witness is constructed directly on the level of the m -mode quantum state preparation σ_p , again observing that the time evolution can be inverted classically for Gaussian measurements. More precisely, observe that

$$W = 1 - \sum_{i=1}^m n_i \quad (185)$$

witnesses the vacuum state $|0^m\rangle$, and hence $\tilde{W} = 1 - \sum_i \tilde{n}_i$, with $\tilde{n}_i = U n_i U^\dagger$, witnesses the state $U|0^m\rangle$. Since the number operator can be measured using homodyne (x and p) measurements that can be seen through the equality $n_i = x_i^2 + p_i^2 - 1/2$, and since the action of a Gaussian unitary U on those operators can be computed efficiently, defining $r_{2i-1} = x_i$ and $r_{2i} = p_i$, the vector r is transformed as

$$U^\dagger r U = S r + d = \tilde{r}, \quad (186)$$

where S is a symplectic matrix corresponding to U and $d \in \mathbb{R}^{2m}$. Measuring all elements of \tilde{r}^2 , i.e., certain linear combinations of $x_i p_j, x_i x_j$ and $p_i p_j$, thus allows one to estimate $\sum_i \tilde{n}_i$ and hence the witness of $U|0^m\rangle$ for any Gaussian state.

All of the fidelity witnesses in this section can be written in the form $W = 1 - \sum_{i=1}^k w_i$ with operators w_i that we need to

measure in an experiment. The sample complexity to achieve an overall estimation error ϵ thus scales as $O(k(\epsilon/k)^{-2}) = O(k^3/\epsilon^2)$ since the error of every individual term needs to scale as ϵ/k .

A downside of fidelity witnesses is that while they provide a bound on the fidelity and are therefore well suited to verify state preparations that are close to the ideal target state, the bound provided by the witness typically becomes loose rather quickly, and hence the value of the witness becomes trivial even while the fidelity is still reasonably high. This motivates one to directly estimate the fidelity, which, while potentially more difficult, yields much more detailed information regarding the state preparation.

2. Fidelity estimation

In certain settings, fidelity estimation is possible with a constant number of samples via the so-called direct fidelity estimation protocol of Flammia and Liu (2011), and similar to the protocols proposed by Bourennane *et al.* (2004), Kiesel *et al.* (2005), Tóth and Gühne (2005), and Pallister, Linden, and Montanaro (2018). Using direct fidelity estimation, we can estimate the fidelity of imperfect state preparations σ with pure target states of the form

$$\rho = \sum_{\lambda \in \Lambda} p_\lambda A_\lambda \quad (187)$$

in terms of normal operators $\{A_\lambda\}_{\lambda \in \Lambda}$ weighted by probabilities p_λ .

The idea is the following: Decompose $A_\lambda = \sum_{a \in \text{spec}(A_\lambda)} a \pi_\lambda^a$ in terms of its eigenprojectors π_λ^a . The fidelity can then be written as

$$F(\rho, \sigma) = \sum_{\lambda} \sum_{a \in \text{spec}(A_\lambda)} p_\lambda \text{Tr}[\pi_\lambda^a \sigma] a, \quad (188)$$

and hence it can be estimated by sampling $\lambda \leftarrow p_\lambda$ and measuring A_λ on the state preparation σ , obtaining outcome a with probability $\text{Tr}[\sigma \pi_\lambda^a]$. Given k samples a_i obtained in this way, the fidelity can then be estimated as $\hat{F}(\rho, \sigma) = (1/m) \sum_{i=1}^m a_i$ with error ϵ using $O(1/\epsilon^2)$ many samples.

For the protocol to be efficiently possible in practice, the following requirements are necessary.

- (i) For each $\lambda \in \Lambda$, A_λ can be efficiently measured. In particular, this is the case if $A_\lambda = A_{\lambda_1} \otimes \dots \otimes A_{\lambda_n}$, with $\lambda = (\lambda_1, \dots, \lambda_n)$, is a product of single-qubit operators A_{λ_i} .
- (ii) For each $\lambda \in \Lambda$, $\text{spec}(A_\lambda) \subset [a_\lambda, b_\lambda]$ for constants $a_\lambda, b_\lambda \in \mathbb{R}$.
- (iii) The probability distribution $p = (p_\lambda)_{\lambda \in \Lambda}$ can be efficiently classically sampled.

A particularly simple application of the protocol is its application to stabilizer states such as the locally rotated cluster state $|\text{CS}\rangle$, since such a state is in the joint $+1$ eigenspace of the stabilizer operators (Flammia and Liu, 2011). A state $|\psi\rangle$ stabilized by n operators S_i with eigenvalues ± 1 can therefore be expressed as $|\psi\rangle\langle\psi| = \prod_i (\mathbb{1} - S_i)/2 = 2^{-n} \sum_{\lambda \in \mathcal{S}} s_\lambda$, where \mathcal{S} denotes the stabilizer

group of $|\psi\rangle$ that is generated by the n operators S_i . Thus, it can be efficiently applied to quantum random sampling architectures that are based on state preparations that are locally equivalent to stabilizer states, particularly ones based on measurement-based computations (Hangleiter, 2021; Ringbauer *et al.*, 2022). Notice, though, that universal random circuits are not of this type.

A potential drawback of the direct fidelity estimation protocol as opposed to fidelity witnesses is that in principle it requires a different *measurement setting* in each run of the experiment. In contrast, to evaluate the fidelity witness only two distinct measurement settings are repeated many times. Thus, while the overall quantum sample complexity is dramatically reduced from $O(n^3)$ to $O(1)$ in the number of qubits, the measurement setting complexity is increased from $O(1)$ to $O(1/\epsilon^2)$ in the estimation error. Depending on the experimental setting at hand there may well be a trade-off between the time required to switch between settings and the time required for many repetitions of the same measurement setting; see Ringbauer *et al.* (2022). It has also been noted that, when restricting the operators A_j to Pauli operators, the sample complexity of verification scales exponentially in the number of non-Clifford gates in the circuit (Leone, Oliviero, and Hamma, 2023).

A closely related fidelity estimation protocol is the so-called shadow fidelity estimation (Huang, Kueng, and Preskill, 2020). In this protocol, measurements are performed in a random Clifford basis; see Kliesch and Roth (2021) for an explanation. The sample complexity of shadow fidelity is also constant, but it is computationally inefficient for non-Clifford states since overlaps between the target state and an arbitrary stabilizer state need to be computed. Another fidelity estimation protocol that can be applied to quantum random sampling schemes is the adaptive protocol of Bennink (2021), which requires two auxiliary qubits and entangling gates between the unknown state preparation and those auxiliary qubits and on-the-fly classical computation. This scheme is sample efficient precisely for anticoncentrating distributions with exponentially small collision probability. To even further reduce the experimental effort of verification as compared to direct fidelity estimation, one would need to improve the scaling in the tolerated estimation error ϵ . For stabilizer states this was studied by Kaley, Kyriallidis, and Linke (2019).

D. Efficient classical verification

Thus far we have seen, on the one hand, classical verification methods that are sample efficient in that they require only a few (polynomially many) samples from the quantum device but require exponential computational runtime. On the other hand, we have seen quantum verification tools that are fully efficient but require trust in an experimental quantum measurement and are experimentally more demanding since they require measurements in different local bases. We conclude our discussion of verification protocols with classical verification protocols that are fully efficient but make other types of assumptions than experimental ones, or yield less information about the implemented distribution.

1. State discrimination

Rather than trying to certify the full target distribution in the TVD, we can alternatively discriminate the experimentally implemented distribution from our best guess of what a noisy distribution or a nearby classically simulable distribution could be. One can see the full verification task in this mindset as distinguishing the imperfect preparation against all possible distributions that are at least ϵ far from the target distribution.

The discrimination task was considered by Gogolin *et al.* (2013) in a setting of a highly restricted client aiming to verify a boson sampler just from the histogram of outcomes without using the information about which outcome has been obtained. They showed that in this setting a boson-sampling distribution cannot be distinguished from a uniform one and prompted the development of a fully efficient and simple state discrimination test that makes use of the actual outcomes (Aaronson and Arkhipov, 2014). To date state discrimination remains the most convincing way to validate boson-sampling experiments, as it is unclear whether the XEB fidelity yields a meaningful benchmark of boson-sampling experiments.

We illustrate the idea by means of the test of Aaronson and Arkhipov (2014) for discriminating the Fock boson-sampling distribution from the uniform distribution. The idea is to use the so-called row-norm estimator for a matrix $X \in \mathbb{C}(n \times n)$,

$$R^*(X) = \frac{1}{n^n} \prod_{i=1}^n R_i(X), \quad (189)$$

where $R_i(X) = \|x_i\|_2^2 = |x_{i,1}|^2 + \dots + |x_{i,n}|^2$ is the norm squared of the i th row of X . Indeed, for a Gaussian normal matrix $X \sim \mathcal{N} \equiv \mathcal{N}_{\mathbb{C}}(0, 1)^{n \times n}$ one expects $\mathbb{E}_{X \sim \mathcal{N}}[R^*(X)] = 1$. The fluctuations around this value depend on whether experimental samples are chosen from the boson-sampling distribution or a uniform distribution and can be exploited to discriminate a device from uniform. To discriminate a distribution from uniform, we compute $R^*(U_{S,1_n})$ for a few samples S and compare the outcome to one's expectation. To see why this achieves the task, we let \mathcal{H} be the distribution \mathcal{N} with distribution function $p_{\mathcal{N}}(X)$ scaled by the probability of obtaining the corresponding outcome, i.e., $p_{\mathcal{H}}(X) = p_{\mathcal{N}}(X)P(X)$. When specializing to boson sampling, the matrix X will be an approximately Gaussian-distributed submatrix $U_{S,1_n}$ of the linear-optical unitary U . Remember that the probability of obtaining this matrix, which corresponds to the outcome S , is given by $P_U(S) = |\text{Perm}(U_{S,1_n})|^2/n!$; see Eq. (45). One finds that (Aaronson and Arkhipov, 2014)

$$\begin{aligned} & \Pr_{\mathcal{H}}[R^* \geq 1] - \Pr_{\mathcal{N}}[R^* \geq 1] \\ &= \frac{1}{2} \mathbb{E}_{\mathcal{N}}[|R^* - 1|] \geq 0.146 - O\left(\frac{1}{\sqrt{n}}\right). \end{aligned} \quad (190)$$

In other words, the row-norm estimator $R^*(X)$ is slightly correlated with $\text{Perm}(X)$. An intuitive reason for this is that multiplying every row of X by the same scalar c also multiplies $\text{Perm}(X)$ by c (Aaronson and Arkhipov, 2014). At the same time, it can be computed in time $O(n^2)$.

To discriminate a boson sampler from the uniform distribution, one therefore needs to simply collect k samples S_1, \dots, S_k from a device claimed to realize a boson sampler and compute $(1/k) \sum_{i=1}^k |R^*(U_{S_i}) - 1|$ up to sufficiently high precision so as to confidently distinguish the resulting value from 0.³⁰

In the same framework, one can distinguish a boson sampler against other, somewhat more informed distributions such as a distribution of distinguishable particles that are sent through the linear-optical network (Carolan *et al.*, 2014; Spagnolo *et al.*, 2014). In the experiments of Zhong *et al.* (2020, 2021), the output distribution was additionally distinguished from a thermal distribution. To distinguish from any classically efficient distribution, they used the Bayesian likelihood ratio estimator

$$c = \frac{\Pr(\{x_1, \dots, x_S\} | P_0)}{\Pr(\{x_1, \dots, x_S\} | P_0) + \Pr(\{x_1, \dots, x_S\} | Q)}, \quad (191)$$

where the likelihood of obtaining the experimental samples x_1, \dots, x_S is evaluated with respect to both the ideal target distribution P_0 and a distribution Q that we want to distinguish from P_0 .

An additional experimentally motivated test ruling out spoofing distributions that makes use of low-order marginal probabilities (Villalonga *et al.*, 2021) performed by Zhong *et al.* (2021) is to measure these marginals. One can then compare them to the theoretical predictions, thereby ruling out the possibility that a distribution that agrees only on the first two or three marginals is a good spoofing distribution.

The efficient state discrimination tests for boson sampling highlight a key difference between the output distributions of variants of boson sampling and universal circuit sampling: for universal circuit sampling we expect the output distribution to not even to be efficiently distinguishable from the uniform distribution. This expectation can be understood in various readings. It can first be viewed from the perspective of HOG-like tests since high performance on a HOG-like test serves as a discriminator against the uniform distribution. Conversely, if HOG is indeed a computationally difficult task, then this provides evidence that discriminating against uniform is also a difficult task. Indeed, it is difficult to imagine a way of discriminating against uniform that does not make use of a HOG-like estimator. Stilck França and García-Patrón (2022) made this intuition more rigorous. They showed that if there are functions defining a cross-entropy measure (133) that gives rise to a sample-efficient state discrimination test, then full verification of the total-variation distance will be sample-efficiently possible in a multiround scheme. Since we do not believe the latter to be possible, the result of Stilck França and García-Patrón (2022) serves as more formal evidence against the possibility of efficient state discrimination for random quantum circuits.

³⁰This may be done in a Bayesian framework (Carolan *et al.*, 2014).

2. Cryptographic tests

A completely orthogonal but promising avenue of verifying sampling schemes was pioneered by Shepherd and Bremner (2009): By allowing the certifier to choose the classical input to the sampling device rather than drawing it fully at random, it may be possible to efficiently certify that a quantum device has performed a task that no classical device could have solved under cryptographic assumptions on the hardness of certain tasks. This could be facilitated by checking a previously hidden bias in the obtained samples for a certain family of IQP circuits (Shepherd and Bremner, 2009).

It is instructive to understand the idea behind such a test of computational quantumness. The protocol of Shepherd and Bremner (2009) was formulated for a certain family of IQP circuits called X programs. An X program acting on n qubits is defined by a list of pairs $(\theta_p, p) \in [0, 2\pi] \times \{0, 1\}^n$ and acts as

$$|0\rangle \mapsto \exp\left(i \sum_p \theta_p \prod_{j=1}^n X_j^{p_j}\right) |0\rangle. \quad (192)$$

For the purposes of the quantumness test it is sufficient to choose a constant value of θ that is the same for every nonvanishing term in the Hamiltonian. In this case, an X program with k nonvanishing Hamiltonian terms acting on n qubits can be represented by a 0/1 matrix $P \in \{0, 1\}^{k \times n}$. Each row of this matrix specifies a Hamiltonian term, and it is easy to see that the output distribution of such an X program is given by

$$P_P(x) = \left| \sum_{a \in \{0,1\}^k: P^T a = x} \cos(\theta)^{k - \text{wt}(a)} \sin(\theta)^{\text{wt}(a)} \right|^2, \quad (193)$$

where $\text{wt}(a) = |\{l \in [k]: a_l = 1\}|$ is the Hamming weight of the binary string $a \in \{0, 1\}^k$.

For a random variable X taking values in $\{0, 1\}^n$ and $s \in \{0, 1\}^n$, the *bias* of X in the direction of s is simply the probability that a sample $x \sim P_P$ is orthogonal to s , i.e., that $x^T s = 0$. The key idea of the test of computational quantumness is to hide a string s , the output probability distribution of an X program, in such a way that this string s cannot be determined efficiently. At the same time, however, the bias of the output distribution of the X program in direction s is significantly larger than the bias of any cheating distribution that can be efficiently obtained using classical computing resources. In particular, the bias of the output distribution P_P of the X program defined by a matrix $P \in \{0, 1\}^{k \times n}$ and angle θ is given by

$$\Pr_{x \sim P_P} [x^T s = 0] = \sum_{x: x^T s = 0} P_P(x). \quad (194)$$

To achieve this, Shepherd and Bremner (2009) noticed that the matrix P can be viewed as the generator matrix of a linear code. That is, the columns of P span the code space $\mathcal{C} = \{Pd: d \in \{0, 1\}^n\}$. If we let P_s be the $n_s \times n$ submatrix

of P obtained by deleting all rows p for which $p^T s = 0^{31}$ and we let \mathcal{C}_s be the code generated by P_s , then we can rewrite the bias (194) of P_p as (Shepherd and Bremner, 2009)

$$\Pr_{x \sim P_p}[x^T s = 0] = \mathbb{E}_{c \sim \mathcal{C}_s} \{\cos^2[\theta(n_s - 2\text{wt}(c))]\}. \quad (195)$$

We can now set a quantum challenge that is intrinsically verifiable in the following way. We choose a code \mathcal{C}_s and a value of θ in such a way that both the bias (195) is strictly larger than $1/2$ and any classical strategy can achieve only a bias that is significantly lower, say, by a constant. We then choose a generating matrix P_s for \mathcal{C}_s such that s is not orthogonal to any of the rows of P_s . Finally, we obfuscate this matrix by adding rows that are orthogonal to s , permuting all rows and potentially performing reversible column operations, giving rise to a matrix P . Given samples from the distribution P_p , we can now distinguish the hypothesis that the sampling device has quantum capacities from the hypothesis that it is cheating to compare the frequencies of outcomes that are orthogonal to the hidden string s . Notice that this protocol does not certify that the samples are distributed according to the correct distribution. Therefore, it does not constitute a work-around to the no-go theorem of Sec. V.A based on cryptographic assumptions. Like the HOG test (Problem 27), this cryptographic test of quantumness merely certifies that the device has the capacity to do something that presumably (under assumptions) no classical computing device could have achieved.

The suggestion of Shepherd and Bremner (2009) is to use *quadratic residue codes* and a obfuscation procedure that exploits specific properties of these codes (such as that the full-weight vector is always a code word). They conjectured that recovering the matrix P_s from the obfuscated matrix P is NP complete. Choosing $\theta = \pi/8$, this construction gives rise to a bias that serendipitously matches that of the Bell inequality: $\cos^2(\pi/8) \approx 0.854$ for the quantum value, and $3/4$ for the best classical strategy discussed by Bremner, Jozsa, and Shepherd (2010).³²

Note also that besides the security assumption on the obfuscation procedure, additional conjectures need to be made (Shepherd and Bremner, 2009) for such a test to achieve its goal: First, the distribution P_p of a randomly selected X program with a constant $\theta = \pi/8$ should be hard to sample from, so only a quantum device can perform this task. Second, the output distribution should be sufficiently flat in the sense that its Rényi 2-entropy or *collision entropy* is close to maximal, i.e., $H_2(P_p) = \Omega(n)$, so cheating becomes more difficult.

Iterating the importance of extensively testing cryptographic assumptions for their security, Kahanamoku-Meyer (2019) developed a classical cheating strategy for the protocol proposed by Shepherd and Bremner (2009). Given a description of an X program in the form of the matrix P , one finds that the cheating strategy extracts the secret vector s with probability arbitrarily close to unity in an empirically observed average run-time of $O(n^3)$.

In a similar mindset, albeit without restricting to sampling tasks for which there is strong complexity-theoretic evidence

for hardness, cryptographic tests of quantumness were devised by Brakerski *et al.* (2018, 2020). They made use of so-called trapdoor claw-free functions to delegate a simple task that no classical device can efficiently solve, but a quantum device succeeds with a higher probability. A trapdoor claw-free function is a two-to-one efficiently computable function f such that it is difficult to find a *claw* x, x' for which $f(x) = f(x')$, but it becomes easy when given access to the trapdoor. Thus, while a classical algorithm can only ever hold $y = f(x)$ and x , but not at the same time x' , a quantum algorithm can compute f in superposition and therefore hold y as well as a superposition $|x\rangle + |x'\rangle$. The idea of the proof is to exploit this superposition: we can ask the device to perform a measurement in the computational basis, obtaining x or x' , or in the Hadamard basis obtaining d for which $d(x \oplus x') = 0$. This reveals some information about x and x' that is not accessible to a classical device. Such protocols were recently improved into much simpler functions (Kahanamoku-Meyer *et al.*, 2022) and low-depth implementations (Hirahara and Le Gall, 2021; Liu and Gheorghiu, 2022), bringing their experimental demonstration within closer reach (Zhu *et al.*, 2021).

Using such trapdoor claw-free functions, it is also possible to classically delegate a BQP computation to a fully untrusted quantum server (Mahadev, 2018), and even to verify sampling problems (Chung *et al.*, 2020). A drawback of the protocol of Chung *et al.* (2020), however, is that it has only inverse polynomially large soundness, so it cannot be used as a subroutine in secure computation problems. More severely, for an application to verifying quantum random sampling, the overhead is unfeasibly large.

E. Further approaches to the verification of quantum samplers

Another approach to verification of quantum states from measurements, blind verified quantum computation, was developed by Broadbent, Fitzsimons, and Kashefi (2009) and Fitzsimons and Kashefi (2017). While the protocols discussed in Sec. V.C.1 make use of the ability of the experimenter to measure single qubits with high fidelity, blind verified quantum computing presupposes the ability to accurately prepare single qubits. And indeed, blind verified quantum computing also applies measurement-based computation using cluster states, thereby exploiting the property that single-qubit phase gates commute through the state preparation. While in our approaches the imperfect state preparation is directly verified, in verified blind quantum computing so-called trap qubits are employed. The outcome of measurements on those qubits is deterministic and can thus be checked to build confidence in the correct functioning of an untrusted quantum server. By turning blind quantum computing upside down, a “*post hoc* verification protocol” for quantum computations was developed by Fitzsimons, Hajdusek, and Morimae (2018).

To build trust in the correct functioning of a sampling device, one can also resort to weaker types of verification than direct verification of the quantum state or output distribution. For instance, instead of directly running a randomly chosen unitary circuit, one can run specific computations on the device, the output distribution of which is highly structured, such as the quantum Fourier transform (Tichy *et al.*, 2014).

³¹This leaves only rows for which $p^T s = 1$.

³²There is no proof that this $3/4$ is the optimal classical value.

Finally, one can build trust in the device from certain efficiently computable benchmarks such as two-point correlation functions (Phillips *et al.*, 2019), higher correlation functions (Zhong *et al.*, 2021), the click-number distribution in boson sampling with threshold detection (Drummond *et al.*, 2022), and comparisons to a coarse-grained distribution (Wang and Duan, 2016).

Note also that, with the exception of the classical verification protocol due to Mahadev (2018) and Chung *et al.* (2020), most of the verification protocols considered here require iid state preparations, which is an additional assumption (albeit a realistic one). To relax this assumption to the non-iid case, one can make use of de Finetti arguments (de Finetti, 1937; Hudson and Moody, 1976; Caves, Fuchs, and Schack, 2002; König and Renner, 2005). This was done by Takeuchi and Morimae (2018), optimized to graph states by Takeuchi *et al.* (2019) and Markham and Krause (2020), and optimized to bosonic states by Chabaud *et al.* (2020) and Chabaud, Grosshans *et al.* (2021).

VI. EXPERIMENTAL IMPLEMENTATIONS

It is the comparative simplicity of quantum random sampling schemes that renders them particularly compelling for an implementation on current-day devices. In contrast to other proposals for quantum advantage, they do not precisely require interactive or multiround feedback. Moreover, comparably small circuit sizes are required such that it might be possible to implement the circuits with non-negligible fidelity without full-fledged quantum error correction. This makes quantum random sampling schemes attractive as proofs of quantum advantage from an experimental point of view. Experimental implementations of quantum random sampling start with the first proof-of-principle demonstrations of boson sampling (Broome *et al.*, 2013; Crespi *et al.*, 2013; Spring *et al.*, 2013; Tillmann *et al.*, 2013) and universal circuit sampling (Neill *et al.*, 2018) and culminate in recent large-scale implementations of universal circuit sampling (Arute *et al.*, 2019; Wu *et al.*, 2021; Zhu *et al.*, 2022) and Gaussian boson sampling (Zhong *et al.*, 2020, 2021; Madsen *et al.*, 2022), which are arguably in the classically intractable regime. In this section, we summarize important technological developments and experimental subtleties of quantum random sampling implementations, with a focus on universal circuit sampling.

A. Universal circuit sampling with superconducting circuits

At the current state of the art, universal circuit sampling is most feasibly implemented using superconducting transmon devices. The first large-scale experiment aimed at reaching a quantum advantage was performed in such an architecture (Arute *et al.*, 2019). This experiment is a landmark experiment that arguably first reached the regime of a quantum advantage over the capabilities of classical supercomputers, and hence the “quantum supremacy” regime. We therefore provide more detail to the discussion of this experiment, as an exemplary discussion *pars pro toto*. The experiment implemented a random circuit consisting of up to 20 layers

of the universal random circuits introduced in Sec. II.A acting on 53 qubits.

1. Design of the experiment

The experiment of Arute *et al.* (2019) was performed on a *transmon* superconducting chip referred to as the *Sycamore* chip. Transmons are superconducting charge qubits that have been designed to be less sensitive to charge noise than is common in other settings, a feature that renders them particularly attractive for use in quantum computational schemes. Generally speaking, in a superconducting circuit currents and voltages behave quantum mechanically, as conduction electrons condense into a macroscopic quantum state. For this to be possible and to ensure that the ambient thermal energy is reduced to well below the native energy scales of the qubits, cryogenic temperatures are required. The extremely low temperatures of ~ 20 mK required for the experiments are currently accessible only in dilution refrigerators. Each of the qubits can be seen as a nonlinear superconducting resonator operating at 5–7 GHz. These qubits can be tuned by resorting to 2 degrees of freedom. On the one hand, there is a microwave drive that allows one to drive Rabi oscillations of the qubit. On the other hand, there is a magnetic flux control that allows one to tune the frequency.

During a quantum circuit, the qubits are tuned to three different frequencies: First, there is the qubit idle frequency at which single-qubit gates are performed. Second, there is an interaction frequency to which neighboring qubits are tuned in order to interact. The idle frequency is chosen such that there is as little crosstalk as possible during single-qubit gates, while at the same time the frequency distance required for interaction with its neighbors is minimized. Finally, the qubits are tuned to a readout frequency. When one selects those frequencies, there are trade-offs to be accounted for between energy relaxation, dephasing, leakage, and control imperfections (Arute *et al.*, 2019). At the idle frequency, single-qubit gates are implemented by driving the qubits with 25 ns microwave pulses.

In the Sycamore superconducting-qubit architecture, two-qubit gates are implemented using adjustable couplers. Since the qubits are arranged in a planar two-dimensional architecture, these couplers are naturally placed between nearest neighbors on a lattice.³³ The couplers allow one to quickly switch on and off a coupling of up to 40 MHz by tuning the frequency of the coupler qubits. Specifically, the coupling is achieved by tuning neighboring qubits’ frequencies on resonance and turning on a 20 MHz coupling for 12 ns. The coupling in the system natively gives rise to the two-qubit gate

$$\text{fSim}(\theta, \phi) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos(\theta) & -i \sin(\theta) & 0 \\ 0 & -i \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 0 & e^{-i\phi} \end{pmatrix}, \quad (196)$$

³³This architecture has also been chosen to be forward compatible with the realization of a surface code for quantum error correction.

with tunable angles θ and ϕ . In Eq. (196) the angle θ is interpreted as the swap angle and the angle ϕ is a conditional phase. The fSim gate captures a wide range of entangling gates, including the *i*SWAP gate with $\theta = \pi/2$ and $\phi = 0$, as well as the CZ gate with $\theta = 0$ and $\phi = \pi$.

In the experiment of Arute *et al.* (2019), *i*SWAP-like fSim gates close to *i*SWAP* (3) with $\theta \approx \pi/2$ and a conditional phase $\phi \approx \pi/6$ were performed. The specific phases of each two-qubit gate corresponding to a particular physical coupler between two qubits varied around their ideal values. Arute *et al.* (2019) were able to measure the precise angle, thus ensuring a higher accuracy of the resulting computational task. The uncertainty in the actual angles implemented in the circuit can be viewed as limited programmability of the device: some parameters of the circuit are determined only contingently on the specific physical implementation. More recently progress has been made toward achieving full programmability of the angles in the fSim gate (Foxen *et al.*, 2020).

Every qubit can be read out by means of a linear resonator. To this end, the qubit frequency is tuned to its readout value and coupled to a far-detuned resonator via a neighboring coupler (Blais *et al.*, 2004; Gambetta *et al.*, 2006; Bultink *et al.*, 2018). As the qubit state changes from $|0\rangle$ to $|1\rangle$, there is a frequency shift in the resonator that can be read out via the phase shift incurred by a microwave probe signal applied to the resonator (Arute *et al.*, 2019). On the chip, the qubits are divided into groups of six qubits that are each coupled to their own resonator, but resonators within a group are simultaneously read out via frequency multiplexing. Overall, the architecture consists of 53 such transmon qubits, each of which is connected to a readout device, with 86 couplers connecting nearest-neighbor qubits.

2. Benchmarking of the components

Substantial efforts have been made to carefully benchmark the experiment.

a. Benchmarking of single-qubit gates

At the lowest level, benchmarking of the experiment was performed on the level of the individual components of the device. For the individual components, the single-qubit operations, the entangling gates, and the readout were benchmarked individually. For both single-qubit gates and two-qubit gates as well as the benchmarking of the entire device, Arute *et al.* (2019) made use of linear XEB. Here we return to the line of thought developed in Sec. V.B.3 and put it into the context of experimental findings. As developed there, XEB provides a unified picture for average-case benchmarking of small-scale operations in the sense of randomized benchmarking, on the one hand (Arute *et al.*, 2019; Y. Liu *et al.*, 2021), and single-instance benchmarking of typical large-scale quantum states, on the other hand (Arute *et al.*, 2019). The use of linear XEB is attractive in this context, as this procedure does not require the classical computation of all possible events, but classical simulations need only to compute the likelihood of the set of bit strings obtained in an experiment.

For the benchmarking of single-qubit gates, linear XEB benchmarking has been used to estimate the probability of an

error occurring on the single-qubit level. For each qubit, a sequence of a variable number of randomly selected gates is applied and $F_{\text{XEB}}(Q, P_C)$, as defined in Eq. (162) and discussed in Sec. V.B.3, is estimated. The resulting scheme can be seen as a randomized-benchmarking protocol (Helsen *et al.*, 2021; Y. Liu *et al.*, 2021). One finds a decay of the signal in the length ℓ of the sequence that is well described by an exponential dependence of the form $(1 - 3e_1/4)^\ell$, where $e_1 \in [0, 1]$ is the single-qubit Pauli error probability. The single-qubit error e_1 over the various qubits follows a distribution that is estimated by suitable histograms. From these histograms, one can then estimate an average of about $e_1 = 0.16\%$ in simultaneous operation of the qubits on the chip.

b. Benchmarking of two-qubit gates

For the linear XEB benchmarking of the two-qubit gates, as in the single-qubit case, sequences of cycles are employed. Now, each cycle consists of randomly chosen single-qubit gates followed by the *i*SWAP* two-qubit gate. This gives rise to an interleaved randomized-benchmarking scheme (Arute *et al.*, 2019) in which the same logic as for single-qubit gate benchmarking is applied: an exponential curve is fitted to the decay, and one can estimate the two-qubit error rate e_2 by subtracting the single-qubit error rate e_1 . After appropriate corrections for dispersive shifts and crosstalk, an average of about $e_2 = 0.62\%$ is found when gates are operated simultaneously on the chip. Finally, the combined single- and two-qubit error rate e_{2C} , which characterizes a single layer of a gate cycle, is measured to be $e_{2C} = 0.93\%$ on average.

c. Characterization of single-qubit measurements

Measurement errors of single-qubit readout are obtained by preparing $|0\rangle$ and $|1\rangle$ and performing a measurement of the state. The identification error is taken to be the probability that the qubit was read out in a state other than that intended, giving rise to a median identification error of 0.97% for the $|0\rangle$ state and 4.5% for the $|1\rangle$ state (Arute *et al.*, 2019). The fact that the state-preparation fidelity is much higher than the measurement fidelity justifies this procedure.

In a second step, multiqubit readout is characterized by preparing and measuring 150 random classical bit strings with 53 qubits and repeating each measurement 3000 times, resulting in a 13.6% probability of correctly identifying the state. This can be decomposed to a median error for the simultaneous single-qubit readout of 1.8% for $|0\rangle$ and 5.1% for $|1\rangle$, giving an overall simultaneous readout error of about 3.8%.

3. Verifying the sampling task

The entire setup of the experiment has been tailored to achieve a quantum computational advantage. Benchmarking the individual components builds trust in the functioning of the 53-qubit device as a whole, but does not yet constitute a test of quantum advantage, as outlined in Sec. I. In light of the hardness of rigorous verification of the sampling task using the samples as explained in Sec. V.A, the entire scheme has been benchmarked via linear XEB but is now applied to

typical instances of high-dimensional quantum states, as discussed in Sec. V.B.3. As noted there, while the linear XEB fidelity does not yield a rigorous certificate for the sampling task, achieving a nontrivial XEB value might be a computationally difficult task in itself. Having said that, the claim of *Arute et al. (2019)* is indeed to have performed the *sampling task* to nontrivial precision.

To estimate the XEB fidelity, the probability of each bit string obtained in the experiment needs to be computed. As further detailed in Sec. VII, for the full random quantum circuit this is beyond the reach of classical computers. This is why proxy methods need to be used in order to reduce the complexity of computing the output probabilities of the implemented quantum circuits. Specifically, *Arute et al. (2019)* made use of three different simulation strategies.

In a full circuit simulation, the exact output probabilities of a given quantum circuit are computed. In “patch circuits” one removes all two-qubit gates along a slice through the 2D qubit array such that the circuit is split into two unconnected parts and the overall fidelity is simply the product of two fidelities. In “elided circuits” one removes a fraction of two-qubit gates between the two partitions of the qubits such that the parts are coupled, but less entanglement is being generated.

To benchmark the patch circuit and elided circuit method against the full circuit method as a means to estimate the XEB fidelity, *Arute et al. (2019)* performed what they called verification circuits. The circuits were chosen in such a way that a full circuit simulation was still possible. Specifically,

two-qubit gates were arranged in a simplifiable tiling so that circuits with exactly the same gate count as in the full experiment were easier to classically simulate. For circuits with 14 cycles on up to 53 qubits, this allowed for the comparison of the three different methods of estimating the XEB fidelity [see Fig. 11(a)], showing that all methods yield roughly the same value for the XEB fidelity.

For full circuit simulation of up to 43 qubits, a “Schrödinger-type” simulation algorithm is run for the simulation of the full quantum state, making use of 100 000 cores and a 250 terabyte memory. For larger qubit sizes, a hybrid “Schrödinger-Heisenberg-type” simulation algorithm is run.

Offering further justification, *Arute et al. (2019)* provided a model for how the XEB fidelity $F_{\text{XEB}}(Q, P_C)$ scales given the errors obtained for the individual circuit components, yielding good agreement with the predictions obtained via the various simulation methods. Altogether these tests constitute justification for the use of the “elided” and “patch” methods as a substitute of full circuit simulation when computing the XEB benchmark.

In the supremacy regime of 53 qubits and a depth of 20, elided and patch circuit methods remain close to the error model (Fig. 11), yielding a value of $F_{\text{XEB}}(Q, P_C) \approx (2.24 \pm 0.21) \times 10^{-3}$ averaged over ten circuit instances. Here the error bar is a σ interval, where σ combines statistical errors of the finite-sample XEB benchmark and systematic errors due to the elided simulation method. This shows that the XEB value is larger than 10^{-3} with 5σ significance. Since the XEB fidelity scales inverse exponentially, the number of

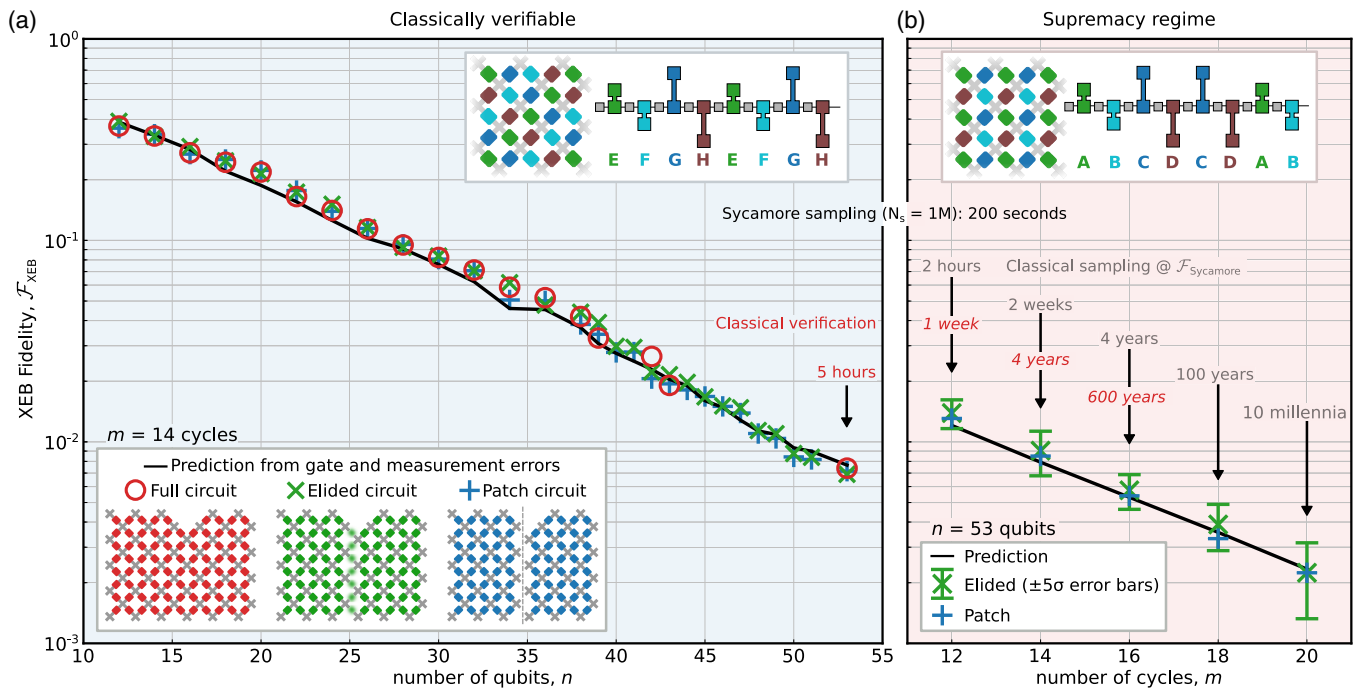


FIG. 11. (a) To build trust in the proxy methods (elided and patch circuits) for full circuit simulation used to estimate the XEB fidelity, sampling on the quantum processor is performed using circuits with the same gate count as the “supremacy circuits,” but in a simplifiable pattern and with a depth of 14. Each data point is an average of the XEB fidelity of ten circuit instances with $(0.5\text{--}2.5) \times 10^6$ many samples per instance. The solid line represents the predicted value of the XEB fidelity given the error model. (b) In the “supremacy regime” of 53 qubits and a depth of up to 20, the elided and patch methods are used to estimate the XEB fidelity and the classical simulation time for verification and sampling is extrapolated. From *Arute et al., 2019*.

samples required to obtain the required significance scales exponentially.

To further substantiate the claim that, on the quantum device, the sampling task has in fact been achieved to nontrivial accuracy, [Arute *et al.* \(2019\)](#) performed further tests. First, they compared the values of the linear XEB with the logarithmic XEB or cross-entropy difference (145); see Sec. V.B.2 for a discussion. This measure is expected to have a larger variance than the linear XEB fidelity, as it puts more weight on the tail of the distribution but at the same time relates more closely to the actual total-variation distance ([Bouland *et al.*, 2019](#)). [Arute *et al.* \(2019\)](#) argued that both measures can serve as a proxy for the quantum fidelity, as discussed in Sec. V.B.3. Second, they analyzed in more detail the distribution of bit string probabilities obtained in the experiment. They found an excellent fit with the expected Porter-Thomas distribution of the outcome probabilities and performed hypothesis tests to reject the hypothesis that the samples stemmed from a uniform distribution.

The claim of having achieved quantum computational advantage in a practical sense is substantiated by extrapolating the computational effort to estimate the computational cost of the quantum advantage circuits to larger system sizes. [Arute *et al.* \(2019\)](#) estimated that for $n = 53$ and $d = 14$, sampling of 3×10^6 bit strings with 0.01 fidelity would take about a year. By extrapolation, they then argued that for the full $n = 53$ and $d = 20$ obtaining 10^6 samples on the quantum processor takes about 200 s, while sampling to a comparable fidelity classically would take 10 000 yr on 10^6 cores, and the verification of the fidelity would require millions of years. These claims have naturally been challenged by the new, improved classical simulation methods explained in Sec. VII.

4. Follow-up work

[Wu *et al.* \(2021\)](#) and [Zhu *et al.* \(2022\)](#) followed up on the landmark experiment of [Arute *et al.* \(2019\)](#) by presenting comprehensive and qualitatively similar data from a superconducting platform, but with a larger number of qubits and larger circuit sizes. The superconducting processor of [Wu *et al.* \(2021\)](#) and [Zhu *et al.* \(2022\)](#) of $n = 66$ transmon qubits, which are coupled by 110 tunable nearest-neighbor couplers.

However, quantitatively the experiment improved in several ways on the experiment of [Arute *et al.* \(2019\)](#). [Wu *et al.* \(2021\)](#) benchmarked the device using 56-qubit, depth-20 random Sycamore circuits [i.e., in the same scheme as [Arute *et al.*, 2019](#)] and achieved comparable error rates. They found a XEB fidelity of 0.0662% for roughly 10^7 bit strings observed in the experiment. [Zhu *et al.* \(2022\)](#) improved upon this and measured a XEB fidelity of 0.0758% for 60-qubit, 22-cycle circuits, and $(3.66 \pm 0.345) \times 10^{-4}$ for 60-qubit, 24-cycle circuits. Their experiment improved on that of [Arute *et al.* \(2019\)](#), especially when it came to readout fidelity, for which they achieved an average fidelity of 2.26%. [Zhu *et al.* \(2022\)](#) estimated that the sampling task would require about 4 orders of magnitude more resources than the sampling task considered by [Arute *et al.* \(2019\)](#).

To summarize this discussion, [Arute *et al.* \(2019\)](#), [Wu *et al.* \(2021\)](#), and [Zhu *et al.* \(2022\)](#) all claimed a significant

advantage for their respective quantum devices over all possible classical algorithms applied to the same task. In a nutshell, the advantage claim of those experiments is based on a placeholder for the linear XEB fidelity that can be computed in the advantage regime, as well as empirical and numerical evidence for the validity of this estimator. In Sec. VII, we discuss how and to what extent this quantum advantage claim is challenged by tailored classical simulation algorithms, as well as how the particular choice of benchmark affects the claim.

B. Photonic implementations

Historically preceding implementations using superconducting quantum circuits, photonic implementations of variants of boson sampling have developed significantly over the past decade. These fall under implementations of the original proposal of [Aaronson and Arkhipov \(2013\)](#) to make use of initial Fock state preparations, as well as implementations of the Gaussian boson-sampling protocol initially proposed by [Lund *et al.* \(2014\)](#) and refined by [Hamilton *et al.* \(2017\)](#) and [Kruse *et al.* \(2019\)](#).

1. Fock boson sampling

Soon after the proposal of boson sampling became available ([Aaronson and Arkhipov, 2013](#)), the first experiments with photonic systems were conducted, all at around the same time ([Broome *et al.*, 2013](#); [Crespi *et al.*, 2013](#); [Spring *et al.*, 2013](#); [Tillmann *et al.*, 2013](#)). These first implementations involved a comparably small number of modes and photons, even though these early experiments were often performed on photonic chips in integrated optics. [Spring *et al.* \(2013\)](#) presented data from an experiment involving $m = 6$ modes and $n = 3$ and 4 photons, resorting to silica-on-silicon integrated waveguide circuits. In such waveguide circuits fabricated by ultraviolet writing, evanescent waves overlap, giving rise to effective beam-splitter arrays. In this experiment, two parametric down-conversion pair sources are used to inject up to four photons into a photonic circuit. In other words, the sources are not used in a heralded mode, where one port provides a classical signal for the presence of a photon in the other port, but both output ports of the sources are fed into the device. The dominant sources of inaccuracy in this type of sampling are consequently multiphoton emission, as well as the partial distinguishability of our photon sources.

In fact, limitations of single-photon sources to date still constitute a key limitation in the way of large-scale implementations of Fock boson-sampling experiments. Postselection is used to ensure that higher photon numbers that are intrinsically also produced in the process do not substantially contribute. To build trust in the functioning of the device, the measured relative frequencies of outcomes in which the photons are detected in distinct modes are compared with the expected numbers. This is possible since, up to these system sizes, the relevant probabilities can still be classically computed.

The experiment of [Crespi *et al.* \(2013\)](#) also showed three-photon interference in an integrated interferometer involving $m = 5$ optical modes. Similarly, [Tillmann *et al.* \(2013\)](#)

presented data from three photons in a $m = 5$ mode integrated optical interferometer. In each case, single photons were created using parametric down-conversion. Broome *et al.* (2013) performed boson sampling in a tunable architecture on $m = 6$ modes with $n = 2$ and 3 photons. Here polarization controllers at the inputs and outputs can be used to perform different unitary evolutions.

The next step in implementation sized up the instances slightly to $n = 3$ photons in $m = 9$ modes (Carolan *et al.*, 2014; Spagnolo *et al.*, 2014). More significantly, both Carolan *et al.* (2014) and Spagnolo *et al.* (2014) performed the efficient state discrimination test proposed by Aaronson and Arkhipov (2014) in order to distinguish the experimental samples from a uniform distribution. Carolan *et al.* (2014) furthermore distinguished the samples from a distribution obtained if the bosons were distinguishable, making use of a technique called bosonic clouding. More recently Giordani *et al.* (2018) experimentally demonstrated a way to efficiently witness multiphoton interference in an $n = 3$ photon experiment, following a proposal of Walschaers *et al.* (2016).

These small-scale experimental results were more recently brought to a new level in terms of large-scale photonic implementations. This advance was made possible by substantial technological development (Loredo *et al.*, 2017; Wang *et al.*, 2017). On the one hand, solid-state sources of highly efficient, pure, and indistinguishable single photons have been developed. Such quantum dot-micropillar systems allow for the deterministic generation of indistinguishable single photons with a high sample rate. On the other hand, the transmissivity of linear-optical circuits has been dramatically improved with the development of ultralow-loss optical circuits. These developments allowed Wang *et al.* (2017) to implement a $n = 5$ -photon, $m = 9$ -mode boson sampler with a high sample rate. Improving those components even further by integrating the optical circuit in a three-dimensional architecture, Wang *et al.* (2019) performed a boson-sampling experiment with $n = 20$ photons and $m = 60$ modes: the largest implementation of Fock boson sampling to date. For a detailed discussion of the early photonic implementations of boson sampling, see the review by Brod *et al.* (2019).

2. Gaussian boson sampling

Gaussian boson sampling allows for even larger system sizes, given the comparably easy availability of suitable sources. Recall that in Gaussian boson sampling single-mode squeezed states are prepared at the input, whereas in Fock boson sampling single-qubit Fock states need to be prepared, a much more challenging task.

After early demonstrations of the so-called scattershot boson-sampling variant of Gaussian boson sampling (GBS) (Bentivegna *et al.*, 2015; Zhong *et al.*, 2018; Paesani *et al.*, 2019), Zhong *et al.* (2020) performed a large-scale GBS experiment that involved 50 input single-mode squeezed states featuring high indistinguishability and squeezing parameters. The resource states are fed into a large-scale bulk optical (and hence not integrated) interferometer with full connectivity among $m = 100$ modes that implements a random transformation with low loss. Some randomness in this interferometer is physical: the interferometer is fabricated

to implement a certain unitary transformation, but imperfections of the process alter the targeted unitary. To obtain an accurate description of the unitary, the interferometer is characterized *post hoc* via tomography. Strictly speaking, the boson-sampling device used in this and all previous experiments is therefore not a programmable device. Rather, it is designed to implement a specific transformation that is slightly altered in the fabrication process. The output of the interferometer is then sampled from making use of high-efficiency single-photon detectors. In this experiment, up to $n = 76$ output photon clicks have been detected.

This scheme was improved by Zhong *et al.* (2021) in two ways. First, a restricted programmability of the boson-sampling device was achieved by making use of the capacity to vary the phase of the input squeezed states. This can also be viewed as introducing programmable phases in the random unitary transformation. Second, the experiment was pushed further to detecting $n = 113$ photon events at the output of a photonic circuit comprising $m = 144$ optical modes. Key to the latter improvement is the availability of a high-brightness and scalable quantum light source that was developed for this purpose. This source builds on methods of the stimulated emission of squeezed photons, which are improved to achieve near-unity purity and high efficiency.

In principle, these experiments can be efficiently verified in their functioning using quantum measurements (Chabaud, Grosshans *et al.*, 2021); see Sec. V.C.1. While such tests were performed in this experiment, subsystem properties (namely, low-order mode marginals) were used by Zhong *et al.* (2021) to efficiently distinguish them from classically simulable distributions such as distinguishable photons and thermal states; see Sec. V.D.1. To this end, they used a variant of Bayesian likelihood ratio estimators, which can be recast as a ratio of cross-entropy scores. In a similar vein, Drummond *et al.* (2022) found good agreement between the distribution of the total number of clicks of the threshold detectors observed by Zhong *et al.* (2021) with the theoretical click-number distribution, including some decoherence effects.

Recently Madsen *et al.* (2022) performed Gaussian boson sampling using time multiplexing in order to implement low-depth but high-dimensional unitary mode transformations, as proposed by Deshpande *et al.* (2022). The lower depth of the unitary transformation allows larger system sizes to be reached since the loss does not contribute as much. At the same time, classical simulation may become easier, but Deshpande *et al.* (2022) provided numerical evidence that low-depth, high-dimensional transformations remain computationally intractable in practice. The experiment used $m = 216$ single-mode squeezed input states, a linear-optical transformation with three-dimensional connectivity, and photon-number-resolving detectors. The average number of detected photons is 125. To benchmark the experiment, Madsen *et al.* (2022) applied a number of tests. For the events with a low photon number of $n \leq 6$ and $m = 16$, they computed the TVD between the experimental and the target distribution. In the intermediate regime of photon numbers $n \leq 26$ and $m = 216$ modes, they estimated the cross-entropy difference as well as the Bayesian estimator of Zhong *et al.* (2021) in order to compare them to the potential classical spoofing algorithms discussed in Sec. V.D.1. Finally, in the

classically intractable regime, they computed first- and second-order cumulants of the experimental distribution.

We close this section by mentioning that a variant of the original scheme of Gaussian boson sampling was implemented by [Thekkadath *et al.* \(2022\)](#) on an interferometer comprising $m = 15$ modes. This scheme allows for shifts of the input squeezed states in phase space. Such displacements are useful when anticipating applications of Gaussian boson sampling, as sketched in Sec. VIII. A direct implementation of a scheme of approximating vibronic spectroscopy with imperfect quantum optics as a variant of boson sampling was reported by [Clements *et al.* \(2018\)](#).

C. Further implementations of quantum random sampling

The previously discussed schemes of quantum random sampling are by far the most common schemes that have been implemented experimentally. That said, platforms aside from superconducting and photonic architectures have also been considered, sometimes even leading to an actual experimental realization. [Wang *et al.* \(2020\)](#) suggested overcoming the challenge of preparing and detecting bosonic quantum states in photonic implementations and implemented a boson-sampling protocol in a two-mode superconducting device, thereby deviating from the common implementations of boson sampling on photonic platforms. This is used for simulating molecular vibronic spectra, as suggested by [Huh *et al.* \(2015\)](#).

Quantum random sampling in the measurement-based model of quantum computing was recently demonstrated on small scales by [Ringbauer *et al.* \(2022\)](#). The advantage of this approach over gate-based circuits is that, in principle, significantly less device control is required. This is because all entangling gates are fixed and can be applied in a single layer, and only a single layer of random Z-type rotations are required ([Bermejo-Vega *et al.*, 2018](#); [Haferkamp, Hangleiter, Bouland *et al.*, 2020](#)). The trade-off of this approach compared to a gate-based one is therefore one between depth of the circuit and space: in order to achieve a hard-to-simulate circuit comparable to that of [Arute *et al.* \(2019\)](#), 2500–10 000 qubits are presumably required. [Ringbauer *et al.* \(2022\)](#) made this trade-off explicit: By “recycling”—i.e., measuring and re-preparing—certain qubits during the computation while keeping the remaining qubits coherent, depth of the physically implemented circuit can be traded with the number of qubits available in the device. A major advantage of the measurement-based approach to quantum random sampling is that it is possible to efficiently witness and measure the quantum fidelity using single-qubit measurements, as discussed in Secs. V.C.1 and V.C.2 ([Hangleiter *et al.*, 2017](#); [Bermejo-Vega *et al.*, 2018](#); [Hangleiter, 2021](#)). This allows one to perform the benchmarking and verification methods discussed in Sec. V.B.3 using the quantum fidelity, and thereby circumvent important caveats of the XEB fidelity.

Along similar lines, to lessen the burden of actually explicitly implementing random circuits in a gate-based approach, a number of schemes have been suggested that would in effect give rise to such circuits, but they are based on physical interaction mechanisms. For example, [Muraleedharan, Miyake, and Deutsch \(2019\)](#) considered the complexity of a probability distribution associated with an

ensemble of noninteracting massive bosons undergoing a quantum random walk on a one-dimensional lattice. These settings are potentially more feasible to implement in cold atomic systems. In fact, the coherent cold collisions that have already been experimentally implemented ([Mandel *et al.*, 2003](#)) in systems of neutral ultracold atoms in optical lattices gives rise to precisely the interaction required for the implementation of the schemes of [Bermejo-Vega *et al.* \(2018\)](#) and [Haferkamp, Hangleiter, Bouland *et al.* \(2020\)](#), which allows for efficient quantum verification.

VII. CLASSICALLY SIMULATING QUANTUM RANDOM SAMPLING SCHEMES

Random quantum sampling schemes are set up to showcase the computational power of quantum devices to demonstrate that there are computational advantages of paradigmatic quantum computers over classical computers. The rigorous statements discussed in Sec. IV always involve a separation in the scaling of classical versus quantum computations. Such statements show that as systems are scaled up the speed of the respective quantum computations will at some point certainly surpass that of every classical algorithm. But how large does one actually have to make a quantum sampler such that it cannot be simulated classically? In other words, what is the finite-size behavior of the complexity of simulating quantum random sampling?

This question be explicitly answered only for specific classical algorithms at a time.³⁴ The effort to devise such specific algorithms constitutes a crucial part in the quest of demonstrating a quantum advantage and thereby violating the extended Church-Turing thesis: One has to demonstrate not only that the scaling is possible in principle but also that the frontier determined by the best available classical algorithm run on the fastest available supercomputers can be surpassed using actual quantum devices.

We can conceive of this situation as a competition between classical algorithms with an unfavorable scaling of the complexity, but run on extremely large supercomputers, and small but extremely noisy quantum devices. In the absence of quantum error correction, both competitors will hit a ceiling sooner or later, and the competition between classical and quantum devices is determined by which ceiling is more favorable: Roughly speaking, the quantum device, which is constrained by the noise present in current-day experiments, will hit the simulation barrier as the circuit size reaches the tolerated error divided by the local gate error. Conversely, the classical algorithm, which is constrained by the scaling of the simulation task, will hit a barrier once the time or space complexity reaches the tolerable limit determined by the speed and memory size of current-day supercomputers.

What is and what is not possible in this situation depends heavily on the precise setting considered: Is the goal to exactly simulate the sampling task, to sample from a distribution close in the TVD, or to simulate a quantum experiment while including realistic amounts and sources of noise? Or is it to

³⁴Alternatively, one can invoke fine-grained complexity assumptions, as discussed in Sec. IV.E.

score at least as high as the quantum device on a given benchmark, potentially via a means other than simulating the sampling task? Depending on the task at hand, a classical simulation algorithm may be able to exploit weaknesses in the benchmark, or optimally exploit the available time and space resources to beat the performance of a noisy quantum device.³⁵

In this section, we provide an overview of classical simulation algorithms for different tasks related to quantum random sampling. We categorize those tasks into two categories: first, computing the output probabilities, a task that inevitably requires exponential precision for random instances, as almost all probabilities are exponentially small (recall Sec. V.A), and second, simulating the sampling task. Computing the output probabilities is first and foremost required as a subroutine of most sampling algorithms, and also for the estimation of the XEB fidelity of an experimental system. In contrast, the goals of simulating the sampling task are manifold: the goal can be to sample from the ideal output distribution or a distribution close to it, it can be to simulate a noisy quantum experiment as well as possible, or it can be to achieve high scores on a given quantum advantage benchmark such as the XEB fidelity.

A. Sampling versus computing output probabilities

Computing the output probabilities of a random quantum computation, or strongly simulating it, involves computing the output amplitudes of a quantum circuit. On a high level, most classical algorithms for computing probabilities can be broadly categorized into Feynman-type algorithms and Schrödinger-type algorithms (Aaronson and Chen, 2017). Consider a quantum circuit with m gates acting on n qubits. A Schrödinger algorithm stores and consecutively updates the entire state using $\sim 2^n$ space and $\sim m2^n$ time. A Feynman algorithm, in contrast, makes use of a path-integral formulation of the output amplitudes [recall Eq. (19)] that expresses them as a sum of $\sim 4^m$ many products of m matrix entries of the quantum gates in the circuit. Such an algorithm computes each term and sums all terms up consecutively, therefore requiring merely $\sim m + n$ space. In fact, Aaronson and Chen (2017) showed that Feynman algorithms can have a much reduced run-time for local circuits that can be decomposed into d layers of m/d gates by recursively computing sums over paths over portions of the circuit. This gives rise to a run-time scaling of $O(n(2d)^{n+1})$ for general circuits and $2^{O(d\sqrt{n})}$ for circuits on a two-dimensional grid, while the space consumption scales as $n \log n$. Typically $m \gg n$ and hence, depending on the setting at hand, space or time may be the limiting factor and determine the choice of simulation algorithm.

In practice, more intricate algorithms are used, but the basic idea often remains the same. For qubit-based architectures, most importantly universal random circuits, hybrid Schrödinger-Feynman-type algorithms turn out to be the most efficient in practice. The most important tool here is so-called

tensor-network algorithms [see Bridgeman and Chubb (2017) for an introduction], which allow the exploitation of locality structure in quantum circuits. For boson-sampling schemes, Feynman algorithms are natural since the output probabilities are expressed in terms of matrix polynomials in the entries of an $n \times n$ linear-optical unitary, albeit with exponentially many terms. Locality cannot in most instances be meaningfully exploited for those systems.

We now turn to the task of sampling from the output distribution of a quantum circuit, or weakly simulating it. Computing the probabilities is not sufficient for sampling from a given distribution, and in fact is not even necessary, however. Having said that, computing the output probabilities is often the key subroutine of sampling algorithms, and all methods of sampling that we are aware of make use of that subroutine. We sketch the most important ideas for how to sample from a given distribution that are used in simulations of quantum random sampling.

First, there are *ancestral sampling* techniques. Here the idea is that in order to sample from a multivariate distribution [say, a distribution p over length- n bit strings with probabilities $p(x_1, \dots, x_n)$], we can iteratively sample from the marginal distributions of larger and larger portions of the bit string. In the first step of such an algorithm, we sample a bit y_1 from the marginal distribution $p_1 = \sum_{x_2, \dots, x_n} p(\cdot, x_2, \dots, x_n)$, in the second step we sample from the conditional distribution $p(\cdot | y_1)$, etc. The key obstacle to notice in this approach is that it requires an algorithm not only for individual probabilities but also for all marginals of the distribution, a potentially considerably more difficult task, as it naively requires summing over exponentially many probabilities.

Second, there are *rejection-sampling* techniques. The idea of rejection sampling is to generate a sample y from a distribution q , as well as a uniformly random number $u \in [0, 1]$ in the first step. The distribution q should be such that we can efficiently sample from it and it must satisfy $p(x) \leq cq(x)$ for some number c and all x . In the second step, the sample x is accepted if $ucq(x) \leq p(x)$ and rejected otherwise. If it is rejected, the procedure is repeated. The expected number of probabilities that need to be computed per sample is given by c . Rejection sampling has a natural geometrical intuition: Suppose that q is the uniform distribution over length- n bit strings and $c = 2^n$. We then sample uniformly random points in the rectangle $\{0, 1\}^n \times [0, 1]$ and accept a sample if it lies within the histogram of the distribution p .

There are also so-called Markov-chain Monte Carlo techniques. Here the idea is to set up a Markov chain of bit strings $x_1 \rightarrow x_2 \rightarrow \dots \rightarrow x_m$ that converges to the target distribution p as its stationary distribution. This Markov chain is specified by the probability $P_t(x)$ of being in state x at time step t and rates $W_{x \rightarrow x'}$ for the transition $x \rightarrow x'$ that determine the probability of moving from state x to state x' . The overall idea is to construct the Markov chain based on a proposal distribution q . The proposal distribution determines the probability $q(x'|x)$ of moving to state x' given that the Markov chain is in state x . The transition probabilities are then given by

$$W_{x \rightarrow x'} = \Pr[\text{accept} | (x'|x)]q(x'|x). \quad (197)$$

³⁵A quantitative analysis of the competing scalings for the task of sampling from the exact or noisy distribution as measured by the linear XEB fidelity was made by Zlokapa, Boixo, and Lidar (2023).

A simple choice of the acceptance probability is the Metropolis choice,

$$\Pr[\text{accept}|(x'|x)] = \min \left\{ \frac{p(x')q(x|x')}{p(x)q(x'|x)}, 1 \right\}. \quad (198)$$

Equation (198) has a favorable property in that it depends only on the ratio $p(x')/p(x)$. This means that one need not be able to compute those probabilities directly, but rather only a function $f \propto p$.

B. Simulating universal circuit sampling

The best studied family of quantum circuits is universal random circuits, particularly the circuits implemented by Arute *et al.* (2019) and, subsequently, Wu *et al.* (2021) and Zhu *et al.* (2022). Recall that these circuits comprise single-qubit gates \sqrt{X} , \sqrt{Y} , \sqrt{W} and the two-qubit gate $i\text{SWAP}^* = \text{fSim}(\pi/2, \pi/6)$. The goal of large-scale simulations performed for this task has been to compare the performance of inefficient classical algorithms, potentially including approximations and the noisy quantum devices in the lab. The methods devised for this task have similar complexity for the task of computing the probabilities and simulating the experiment since amplitude estimation dominates the computational cost. Nonetheless, the number of amplitudes required for sampling typically scales linearly in the number of samples, and hence producing millions of samples can be prohibitively costly, while computing a single amplitude is achievable.

The figure of merit in terms of which the success of these simulations is measured is either the fidelity of the classical representation of an approximate quantum state in cases in which such a representation exists or the XEB fidelity of the produced classical samples as a classical benchmark that acts as a placeholder for the circuit fidelity.

1. Using tensor networks to simulate quantum circuits

The most important tool for the simulation of universal random circuits is *tensor networks* (Markov and Shi, 2008; Boixo *et al.*, 2017). The basic idea of a tensor network is to express a quantity of interest in terms of a network of multi-index tensors in which the edges correspond to a prescription to sum over the corresponding index. Amplitudes of quantum circuits are therefore naturally tensor networks since two-qubit gates are rank-4 tensors, single-qubit gates are rank-2 tensors, and a product state is simply a product of vectors (rank-1 tensors). The circuit description is merely a rule specifying how to connect those tensors. To compute the quantity of interest, one then needs to contract the tensors across their edges, i.e., perform tensor multiplication by summing over the corresponding index; see Fig. 12. The contraction complexity is determined by the largest dimension of an index that appears in a particular contraction scheme, which is roughly determined by the tree width of the underlying graph (Markov and Shi, 2008).

While the properties of one-dimensional efficient tensor networks can be computed efficiently in the dimensions and size of the tensor network, this no longer generally holds true

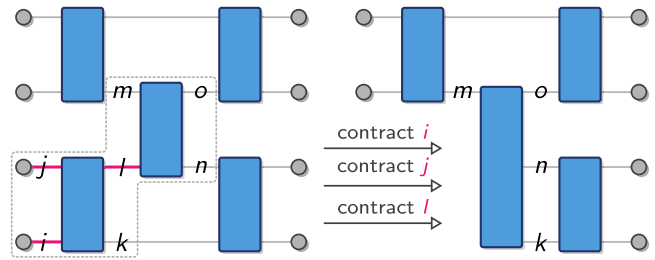


FIG. 12. In a tensor network every edge corresponds to a rule to sum over the corresponding index of the neighboring tensors. In a quantum circuit two-qubit gates are represented as four-index tensors (boxes) and single-qubit computational-basis states (vertices) are single index tensors or vectors. Contracting an edge with a neighboring computational-basis state (indices i and j) corresponds to selecting a slice of the neighboring tensor. Contracting an edge between two arbitrary tensors (index l) corresponds to summing over the entries of the neighboring tensors over that index, resulting in a new, larger tensor.

for higher-dimensional geometries that do not admit a linear contraction scheme (Schuch *et al.*, 2007; Haferkamp, Hangleiter, Eisert, and Gluza, 2020). Nonetheless, it often remains possible to find contraction schemes that scale much better than the worst-case run-time in practice.

Tensor networks admit various sampling algorithms. One can make use of ancestral sampling because the data structure of a tensor network naturally admits the computation of marginals at a cost that is similar to the cost of computing an individual output amplitude (Ferris and Vidal, 2012). Still, this method is costly since every sample requires n different contractions of the circuit tensor network.

For the output distributions of random universal circuits, variants of rejection sampling are much more efficient, however. This is because the output distribution of random universal circuits is exponentially (or Porter-Thomas) distributed, which implies that the largest probability is exponentially small with inverse polynomial failure probability over the choice of the random circuit [recall Eq. (161)]. Choosing the uniform proposal distribution and the bound $c = \log(2^n/\epsilon)$ in the rejection-sampling algorithm (see Sec. VII.A), one can therefore simulate Porter-Thomas-distributed probability distributions for $n = 49$ up to error $\epsilon = 10^{-3}$ using 41 probabilities per bit string on average (Markov *et al.*, 2018). Further improving on this, Markov *et al.* (2018) introduced a “frugal” sampling scheme that reduces the fraction of rejected strings. To do so, frugal rejection sampling chooses c such that the upper tail of the distribution with probabilities $> c/2^n$ has a fixed weight ϵ and accepts all proposed strings x_j with unit probability if their probability is larger than $2^n p(x_j)/c$; see Fig. 13. This effectively reduces the probability of such outcomes to $c/2^n$ while improving the average number of probabilities required per sample and making them independent of n . At the same time, it introduces an error of the sampled distribution compared to the target distribution. Quantitatively, this error is given by $2 \exp[-c/(1 - e^{-c})]$ as measured by the TVD of the sampled distribution to the ideal one assuming exponentially distributed probabilities. For instance, for $c = 10$ it is given by $\sim 10^{-4}$.

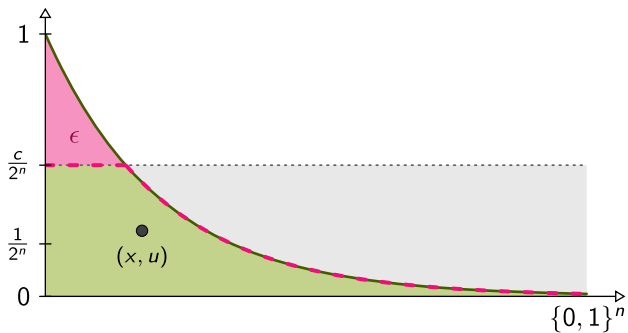


FIG. 13. In frugal rejection sampling, we sample a point (x, u) uniformly at random over the area $\{0, 1\}^n \times [0, c/2^n]$. A sample is accepted if $u \leq p(x)$ (green area) and rejected otherwise. For Porter-Thomas exponentially distributed outcome probabilities (solid green line), this will result in a TVD error $\epsilon = 2 \exp[-c/(1 - e^{-c})]$ of the actually sampled distribution (dashed pink line) compared to the target distribution.

A further important advantage of rejection sampling over ancestral sampling is that all probabilities can be precomputed. This is crucial because it allows a more efficient use of the contractions of a tensor network. Instead of contracting a new tensor network for each amplitude, the desired output strings are stored in a large tensor such that only a single (albeit slightly more complex) contraction is required for the entire batch. For instance, in the algorithm of Markov *et al.* (2018) saving 10^7 amplitudes instead of a single one leads to only a 2.76-fold slowdown of the simulation.

2. Simulation of random quantum circuits

The question of how to approximately simulate random quantum circuits thus boils down to the question of how to best contract the tensor network representing the quantum circuit. In the following, we further discuss various techniques for doing so.

a. State-vector simulation

The simplest algorithms in some sense for the simulation of quantum circuits merely store the entire quantum state and time evolve that state (De Raedt *et al.*, 2007, 2019; Smelyanskiy, Sawaya, and Aspuru-Guzik, 2016; Häner and Steiger, 2017; Pednault *et al.*, 2019). Here a key challenge is to exploit all the available storage on a large computer. For this, the state-vector simulation needs to be distributed among the different parts of the storage. To our knowledge, the largest such simulation runs on 49 (7×7) qubits (Pednault *et al.*, 2017; Li *et al.*, 2018).

An alternative and arguably more natural way of storing multipartite quantum states is given by a tensor network, as previously discussed. To compute all amplitudes of the output state of the quantum circuit, one can contract the tensor network along the time dimension, giving rise to a tensor-network representation of the output state. The description complexity of a tensor-network state is bounded by $\sim n2^n$, while the simulation time scales as $\sim m2^n$ in the worst case. This approach has been pursued in a number of works (McCaskey *et al.*, 2018; Guo *et al.*, 2019; Pan *et al.*, 2020;

Zhou, Stoudenmire, and Waintal, 2020), building on work in the simulation of quantum many-body systems (Schollwöck, 2005; Verstraete, Cirac, and Murg, 2008). While tensor networks can efficiently approximate states with low entanglement (Schollwöck, 2005), this is not the case for random quantum circuits that have high entanglement by construction.

Indeed, an important feature of tensor-network algorithms is that they allow for a natural way of relaxing the precision of the simulation. When a two-qubit gate is contracted into a two-site tensor network, the dimension of the tensors are multiplied. To keep the storage effort constant, the usual approach is to perform a singular-value decomposition of the new, larger tensor and to then truncate the smallest singular values. Thus, the tensor size is kept fixed. For quantum states with low entanglement the singular-value distribution will be nontrivial, allowing for an efficient approximation scheme. For random quantum circuits, however, the singular-value distribution tends to be flat, so a reduction of the bond dimension results in large errors (Markov and Shi, 2008; Guo *et al.*, 2019).

The introduced error rate due to such truncation can be viewed as analogous to a finite gate fidelity in a real quantum circuit. Such a sequential compression was pursued for one- and two-dimensional quantum circuits by Zhou, Stoudenmire, and Waintal (2020). Using this approach, fidelities of the output state on the same order of magnitude as seen in the experiment by Arute *et al.* (2019) can be reached for a two-dimensional circuit with CZ entangling gates acting on 54 qubits. These simulations could be carried out on a laptop computer in a few hours. For 20 qubits, the linear XEB fidelity of the resulting state classically can be computed using the exact probabilities that are obtained from an untruncated tensor-network contraction. Note that this approach does not yet achieve the advantage regime of $F_{\text{XEB}} \approx 0.002$ for the *i*SWAP* entangling gate, which is considerably more difficult to simulate. Zhou, Stoudenmire, and Waintal (2020) estimated that this would require a bond dimension of roughly 10^4 , which is an order of magnitude above what is needed for the CZ gate.

Even for algorithms that do not involve approximations, a clever choice of contraction order can yield better run times, however. For instance, Guo *et al.* (2019) provided a simulator for quantum circuits acting on a two-dimensional lattice based on specific contraction strategies of the tensor-network representation of the quantum state after the circuit has been applied. For a lattice of side length L and a circuit of depth d , their most generic contraction scheme achieved space and time complexities of the resulting algorithm that scale as $2^{d(L+1)/8}$ and $L^2 2^{d(L+1)/8}$, respectively. This allowed for the computation of a single output probability of a random quantum circuit with CZ entangling gates of depth 26 on 10×10 qubits on a supercomputer in 9 min and a circuit of depth 40 on 7×7 qubits in 31 min and 92.51 terabytes memory usage.

b. Hybrid algorithms

Implementing the idea of Aaronson and Chen (2017) to balance memory consumption and computation time in a Schrödinger-Feynman hybrid algorithm, Chen, Zhang *et al.*

(2018), Chen, Zhou *et al.* (2018), and Li *et al.* (2018) introduced “slicing algorithms” in which the system is sliced into smaller subcircuits that are independently simulated. Every time an entangling gate occurs between those subcircuits, the number of independent circuits to be simulated is multiplied by the Schmidt product rank of the entangling gate. By judiciously choosing the slices, one can thus optimally balance the memory consumption and computation time.

All of the previously mentioned simulations used CZ entangling gates. In the universal circuit sampling experiments (Arute *et al.*, 2019; Wu *et al.*, 2021; Zhu *et al.*, 2022), the entangling gates are ones that are close to the i SWAP* gate, however. This gate is significantly more challenging to simulate. This is because while the CZ gate can be decomposed into a sum of two equally weighted product operators as $CZ = |0\rangle\langle 0| \otimes \mathbb{1} + |1\rangle\langle 1| \otimes Z$, the i SWAP* gate saturates the decomposition rank of 4 with equal magnitude weights as

$$i\text{SWAP}^* = |0\rangle\langle 0| \otimes |0\rangle\langle 0| + e^{-i\pi/6}|1\rangle\langle 1| \otimes |1\rangle\langle 1| - i|0\rangle\langle 1| \otimes |1\rangle\langle 0| - i|1\rangle\langle 0| \otimes |0\rangle\langle 1|. \quad (199)$$

Roughly speaking, the effort of simulating circuits including i SWAP-like gates will therefore be quadratically larger in the number of gates across partitions of the circuit compared to circuits with CZ entangling operations.

Markov *et al.* (2018) exploited such a decomposition to match a given fidelity in a classical simulation. They made use of the observation that all two-qubit gate paths have equal weight in absolute value, while the remainder of the circuit is chaotic, meaning that different paths contribute roughly equally to the final amplitude (Villalonga *et al.*, 2020). This implies that one may estimate an output probability to a given fidelity by simply summing over a fraction of the paths given by the fidelity. This allows the simulator to produce a (correlated) sample of M bit strings with target fidelity f at the same cost as computing fM noiseless amplitudes.

Villalonga *et al.* (2019) further showed that even faster sampling can be achieved by recycling an initial tensor contraction to obtain contractions for nearby bit strings. The resulting simulation algorithm has been executed on one of the fastest supercomputers available to simulate with fidelity 0.5% depth-40 7×7 random circuits with CZ entangling gates in 2.44 h, and depth-24 11×11 circuits in 0.28 h (Villalonga *et al.*, 2020).

A number of works have aimed at finding the optimal way of contracting the corresponding tensor networks by finding good contraction paths that keep the tensors relatively small (Chen, Zhang *et al.*, 2018; Chen, Zhou *et al.*, 2018; Guo *et al.*, 2019; Huang *et al.*, 2020; Pan *et al.*, 2020; Schutski *et al.*, 2020; Gray and Kourtis, 2021; Guo, Zhao, and Huang, 2021). An approach that is closely related to tensor-network contraction was introduced by Boixo *et al.* (2017). This approach makes use of undirected graphical models that are probabilistic models for which a graph expresses the conditional dependence structure between random variables (Barber, 2012).³⁶ The key idea of this approach is the following:

³⁶As such, they are closely related to tensor networks (Glasser *et al.*, 2019).

When representing the quantum circuit by a product of unitary matrices acting at different clock cycles, expressions for probabilities can be viewed as a path integral with individual paths formed by a sequence of the computational-basis states. The dependencies can then be cast into the form of a probabilistic graphical model, except that in contrast to actual probabilistic models, the factors in general take complex values. To evaluate the resulting expressions, a new variant of a variable elimination algorithm (Murphy, 2012) has been suggested. This algorithm allows one to sample from the output distribution of circuits featuring a sufficiently small tree width, as well as to estimate the XEB benchmark. Chen, Zhang *et al.* (2018) and C. Huang *et al.* (2021) further improved upon this approach and combined it with tensor-network contraction techniques for an application in a parallelized architecture.

Other interesting variants of circuit contraction schemes were proposed by Chen *et al.* (2020), inspired by quantum teleportation, to swap space and time in order to take advantage of low-depth quantum circuits. Finally, Kalachev, Pantelev, and Yung (2021) devised a “multitensor-contraction scheme” in which the tensor-network contraction is performed by assigning a so-called contraction tree with a recursive relation. In this relation, certain precomputed sub-expressions are reused as often as possible to speed up the overall computation. In this way, they are able to compute individual probabilities of Sycamore circuits with a depth of up to 16.

c. Simulating the experiment of Arute *et al.*

The previously mentioned methods have been used to approximately simulate different random circuits or different sizes than those performed by Arute *et al.* (2019), Wu *et al.* (2021), and Zhu *et al.* (2022). To fairly compare the noisy experiment with a classical algorithm, it is necessary to perform the same task (or at least fairly comparable tasks) in the first place, however. It is not fully clear what exactly that task should be. Ideally, the task on which we compare quantum and classical algorithms is to produce samples from the correct probability distribution. However, this point of view has the issue that it is not possible to verify the distribution. Arute *et al.* (2019) seemed to have precisely this in mind when they argued that the linear XEB fidelity is a placeholder for the quantum fidelity and performed further tests to corroborate this, such as computing the logarithmic cross entropy and estimating the entropy of the sampled distribution. Alternatively, we could think that the task on which the experiment has to be beaten is merely to score high on the XEB benchmark. This interpretation has the advantage that there is a clear-cut benchmark, which, while not efficiently computable, can at least be sample-efficiently estimated (see Sec. V.B.3) and is well defined.

The latter approach was taken by Pan and Zhang (2022). They devised a tensor-contraction method that allowed them to exactly compute a certain subset of the probabilities. The basic idea of their “big head” algorithm is to identify a bottleneck in the contraction of the tensor network and split the tensor network into two parts across that edge, which is close to the output. This gives rise to a large “head part” of the

network and a “tail part” of the network. The output qubits in the head part of the network are projected onto a fixed bit string $s_1 = (0, \dots, 0)$. The tail part of the network is much smaller than the head part and contains a subset of the output qubits. Thus, the head part of the network need only be contracted once, while the output amplitude of a bit string (s_1, s_2) can be computed as the inner product of the vectors corresponding to the contractions of the head and tail parts. In this way, Pan and Zhang (2022) were able to obtain 2^{21} correlated bit strings, and postselecting onto the 10^6 largest ones gave a XEB score of 0.736. The probabilities computed by the algorithm are exact. Y. A. Liu *et al.* (2021) reduced the run-time of the algorithm to a few minutes by making use of a large supercomputer.

Now this method arguably does not produce uncorrelated or independent samples from the correct distribution and does therefore not achieve the sampling task associated with the universal circuit. X. Liu *et al.* (2021) made use of the algorithm of Pan and Zhang (2022) in order to produce perfect samples from the target distribution. To this end, they left fewer legs of the tensor network open (6 instead of 21) and used the outcomes to produce a single perfect sample of depth-20 circuits in 276 s on a supercomputer. To produce many samples with smaller XEB fidelity, these perfect samples can be diluted by uniform samples, as proposed by Huang *et al.* (2020). Producing 10^6 samples with a XEB fidelity of 0.2% is therefore equivalent to producing 2000 perfect samples.

Pan, Chen, and Zhang (2022) pursued a different strategy and achieved the approximate sampling task by artificially introducing approximations, and using a sparse representation of the output state. First, they “drilled holes” into the tensor network by judiciously removing a few of its edges at various positions in the circuit. To achieve this, they removed k pairs of edges of the i SWAP* gate, which allows them to significantly reduce the complexity while decreasing the fidelity of the state by roughly a factor of 2^{-2k} . Second, they computed the output probabilities associated with L uniformly random groups of l correlated bit strings. By removing $2k = 8$ edges from the tensor network they were able to compute 2^{26} uncorrelated batches of correlated probabilities. Using those they obtained 2^{20} independent samples from a state with a fidelity or, equivalently, XEB fidelity of $\approx 0.37\%$. The simulation has a cost of about 15 h on a small cluster of 512 GPUs. The samples produced in this way passed the same tests that were performed by Arute *et al.* (2019) to validate the experimental samples.

An analogous approach was pursued by Kalachev *et al.* (2021), who devised a slicing procedure based on maximizing the norm of partially summed slices to match a targeted fidelity. As with the approach of Pan, Chen, and Zhang (2022), this gives rise to batches of correlated probabilities. Kalachev *et al.* (2021) further provided an optimized sampling procedure that minimized the sampling overhead in terms of how many probabilities needed to be computed to a factor of 2. They estimated that this would allow them to sample from a distribution with a fidelity of 0.2% using 15 months time on a single GPU. By postselecting on the largest amplitudes in a number of batches, they were able to spoof the linear XEB benchmark with a value of 0.47% in 4 h on a single GPU.

In a similar spirit, and as previously discussed in more detail, Zhou, Stoudenmire, and Waintal (2020), Barak, Chou, and Gao (2021), and Gao *et al.* (2021) provided evidence that weaknesses in the linear XEB fidelity can be exploited in order to devise classical algorithms with a score comparable to that of noisy quantum devices. Specifically, the algorithm of Gao *et al.* (2021) scored only 1 order of magnitude below the experiment of Arute *et al.* (2019) using a laptop computer. It is projected to keep roughly constant score for larger circuits, while the experimental score is expected to further decrease exponentially. At the same time, the quantum fidelity of the quantum state from which those samples were produced is presumably exponentially small. It may therefore remain difficult to sample from the output distribution of a state with comparably high fidelity.

In Table II we compare the most advanced algorithms for approximately computing the output probabilities of Sycamore circuits. These algorithms are used as crucial subroutines in algorithms that approximately sample from the output distribution of those circuits, and algorithms that perform the weaker task of outputting samples with a high XEB score.

Summarizing the previous discussion, it is fair to say that the experiment of Arute *et al.* (2019) has been simulated on conventional computers, probably most convincingly by Pan, Chen, and Zhang (2022). However, all of the simulation methods mentioned here fail already for the slightly larger implementation of universal circuit sampling by Zhu *et al.* (2022). We stress that our discussion once again highlights how difficult it is to fairly compare different spoofing strategies to experimental samples, either of which can be validated only by incomplete methods such as cross-entropy benchmarking. For example, one can argue that the bit strings produced by Pan and Zhang (2022) have already outperformed the experimental samples in terms of the relevant

TABLE II. Comparison of the time and space complexity of computing bit string probabilities of Sycamore circuits with 53 qubits for selected simulation schemes in terms of the total number of floating point operations (FLOPs). Data from Kalachev, Panteleev, and Yung (2021), Kalachev *et al.* (2021), Pan, Chen, and Zhang (2022), and Pan and Zhang (2022).

Reference	Depth	Bit strings		Time complexity (FLOPs)	Space complexity
Kalachev, Panteleev, and Yung (2021)	16	2×10^6	Uncorrelated	1.1×10^{19}	?
Gray and Kourtis (2021)	20	1	Uncorrelated	3.1×10^{22}	2^{27}
Huang <i>et al.</i> (2020)	20	64	Uncorrelated	6.7×10^{18}	2^{29}
Pan and Zhang (2022)	20	2×10^6	Correlated	4.5×10^{18}	2^{30}
Pan, Chen, and Zhang (2022)	20	2^{26}	Uncorrelated and correlated	3.5×10^{18}	2^{30}
Kalachev <i>et al.</i> (2021)	20	2^{25}	Correlated	6.9×10^{18}	?

benchmark, the XEB fidelity, since this is the benchmark that [Arute et al. \(2019\)](#) have decided on as the central quantity characterizing the quality of their experiment (granting that they did perform further benchmarks). But one could equally well argue that the samples of [Arute et al. \(2019\)](#) are actually approximately sampled from the targeted distribution. In this reading, a high XEB fidelity is one of many features that those samples should have. In addition, they should be independent samples, should have a high entropy, and could even be sampled from the output distribution of a quantum state that has high fidelity with the ideal target state. Evidence for all of those features was collected in the experiment of [Arute et al. \(2019\)](#) by means of various tests. In this reading only the samples of [Pan, Chen, and Zhang \(2022\)](#) can actually be said to “reproduce” the experiment, as viewed through those tests. This discussion highlights the importance of clearly identifying and stating reproducible criteria under which we will consider quantum random sampling to be successfully achieved, on a classical or a quantum computer and in the absence of an unambiguous and efficient means of verifying samples.

d. Efficient algorithms

While the previous simulation algorithms have exponential run times or result in large errors, [Napp et al. \(2022\)](#) provided both numerical and analytical evidence that shallow (depth-3) universal circuits in a two-dimensional brickwork architecture can be strongly simulated as well as efficiently weakly simulated within a constant total-variation distance error. They did so in a twofold approach: They first numerically demonstrated approximate simulation of random universal circuits in a 400×400 brickwork architecture using a tensor-network algorithm (which is worst-case hard to simulate strongly). They then provided analytical evidence for easiness using a mapping to a recently developed model consisting of alternating rounds of random unitaries and weak measurements ([Bao, Choi, and Altman, 2020](#); [Jian et al., 2020](#)); see also Sec. IV.D.6.

e. Alternative simulation schemes

Yet another approach, of an entirely different type, is to make use of the so-called stabilizer decomposition of quantum states. This method is based on the observation that stabilizer states, that is, states generated by Clifford circuits can be efficiently simulated both weakly and strongly ([Gottesman, 1997](#)). Circuits that comprise additional non-Clifford gates can then be expressed as linear combinations of Clifford circuits ([Aaronson and Gottesman, 2004](#); [Bravyi and Gosset, 2016](#); [Bennink et al., 2017](#); [Qassim, Pashayan, and Gosset, 2021](#)). The complexity of this scheme grows exponentially in the stabilizer rank χ of a quantum state, that is, the number of stabilizer states in this decomposition. Since the number of non-Clifford gates typically grows much faster than the number of qubits, this approach is currently not practically useful, however.

To summarize the previously mentioned efforts, we conclude that, using sophisticated classical algorithms, modern supercomputers can keep track of existing experimental schemes of universal circuit sampling.

3. Analysis of noise

Intuitively speaking, noise should render the simulation of quantum random sampling schemes less computationally demanding. In an idealized scenario, if local depolarizing noise with constant strength is applied at the end of a quantum circuit, the output distribution will be close to the uniform distribution. However, often it is *a priori* unclear how to exploit specific types of noise in a particular simulation algorithm and, more specifically, which noise levels will be classically simulable. It has therefore been a subject of some research to delineate regions—determined by the type and strength of noise—in which quantum random sampling schemes are efficiently simulable via classical algorithms. Conversely, one can ask whether it is possible to mitigate certain forms of noise without resorting to quantum error-correction techniques.

Early on [Aharonov and Ben-Or \(1996\)](#) had already considered the effect of depolarizing noise on the complexity of quantum circuit simulation. They found a polynomial-time algorithm for noisy circuits whenever the depolarizing fidelity is higher than some threshold. More specifically to the case of quantum random sampling, [Bremner, Montanaro, and Shepherd \(2017\)](#) showed that IQP circuits subject to local depolarizing noise at the end of the circuit are classically simulable for any constant noise strength, provided the ideal distribution in question is sufficiently anticoncentrated. The key idea of their simulation scheme is to make use of a simulation algorithm based on a sparse Fourier representation of the output distribution for sufficiently anticoncentrated IQP distributions ([Kushilevitz and Mansour, 1993](#); [Schwarz and van den Nest, 2013](#)). For measurement depolarizing noise with strength ϵ and distributions with collision probability $\leq \alpha/2^n$, their algorithm runs in time $O(n^{\log(\alpha/\delta)/\epsilon})$ to sample from the target distribution up to TVD δ . This simulation scheme can be further extended to universal circuits ([Yung and Gao, 2017](#)) using a measurement-based embedding ([Gao, Wang, and Duan, 2017](#)) and then exploiting the algorithm of [Bremner, Montanaro, and Shepherd \(2017\)](#) on individual branches of that embedding. It may not be clear, however, to what extent the considered type of noise channel (local depolarizing noise at the end of the circuit) is actually realistic and reflective of common physical sources of quantum noise ([Boixo, Smelyanskiy, and Neven, 2017](#); [Boixo et al., 2018](#)). It has also been noted that, asymptotically, such Fourier-based simulation algorithms are no more efficient than trivial algorithms ([Boixo, Smelyanskiy, and Neven, 2017](#)). Nonetheless, there may well be an intermediate regime in which an advantage can be gained by exploiting the specific structure of the Fourier coefficients.

Following up on this, [Gao and Duan \(2018\)](#) proved convergence to the uniform distribution for local Pauli noise associated with single-qubit gates. This convergence result was recently refined by [Dalzell, Hunter-Jones, and Brandão \(2021\)](#) and [Deshpande, Niroula et al. \(2022\)](#), who further delineated the regime in which we expect classical simulation algorithms to be feasible; recall Sec. IV.F.1. From the result of [Deshpande, Niroula et al. \(2022\)](#), it follows that random circuits with a constant amount of noise are efficiently simulatable up to inverse polynomial TVD (by the trivial

algorithm that simply outputs uniform samples) whenever the depth grows as $\omega(\log n)$, since in this case the TVD between the noisy output distribution and the uniform distribution is smaller than any inverse polynomial.

Gao and Duan (2018) made a significant step forward from this and gave an average-case simulation algorithm for the output distributions of universal quantum circuits with a noiseless Clifford part and Pauli noise on non-Clifford gates with noise strength η . The output distribution of these noisy circuits is nontrivial in that it is far from uniform yet simulable up to TVD error ϵ with a run-time of $n^{O(\log 1/\epsilon)/\eta}$, and hence efficient for constant $\epsilon, \eta > 0$. This algorithm runs in quasipolynomial time only if the goal is to simulate the noisy circuit up to inverse polynomial TVD. This regime is significant since an algorithm that can simulate a noisy experimental circuit only up to constant TVD can be efficiently distinguished from the actual noisy experiment at polynomial overhead. Building on this algorithm of Gao and Duan (2018), Aharonov *et al.* (2022) closed this gap and found an algorithm that can efficiently simulate a noisy universal random circuit with constant local depolarizing noise after every gate up to any inverse polynomial TVD whenever the output distribution anticoncentrates. Since anticoncentration requires at least logarithmic depth (Dalzell, Hunter-Jones, and Brandão, 2022), the algorithm of Aharonov *et al.* (2022) is therefore nontrivial precisely in the regime of logarithmic depth. Given previous results, this is the regime in which one might hope for an asymptotic quantum advantage even for quantum circuits with a constant amount of noise (Deshpande, Niroula *et al.*, 2022); see our discussion of these results in Sec. IV.F.1. These results hence show that random quantum circuits of logarithmic depth do not offer a “sweet spot” at which anticoncentration already sets in, and yet constant noise levels are not yet overwhelming.

We now sketch the idea of the algorithm of Gao and Duan (2018) and Aharonov *et al.* (2022), which draws its key ideas from the work of Bremner, Montanaro, and Shepherd (2017). The starting observation of the algorithm is that the ideal output distribution of a quantum circuit $C = U_d U_{d-1} \cdots U_1$ can be expressed as a Pauli path integral

$$P_C(x) = \sum_{s_0, \dots, s_d \in \mathbb{P}_n} \text{Tr}[|x\rangle\langle x|_{s_d}] \text{Tr}[s_d U_d s_{d-1} U_d^\dagger] \cdots \text{Tr}[s_1 U_1 s_0 U_1^\dagger] \text{Tr}[s_0 |0^n\rangle\langle 0^n|] \quad (200)$$

$$=: \sum_{s \in \mathbb{P}_n^{d+1}} f(C, s, x), \quad (201)$$

where \mathbb{P}_n is the n -qubit Pauli group. This can be easily seen from the fact that the Pauli matrices form a complete operator basis, and therefore $\text{Tr}[U\rho U^\dagger s] = \sum_{t \in \mathbb{P}_n} \text{Tr}[U t U^\dagger s] \text{Tr}[\rho t]$. We can also think of the Pauli path integral as a Fourier decomposition of the output probabilities.

In the Fourier representation, the effect of local depolarizing noise can be easily analyzed since it acts simply as $\mathcal{E}(\rho) = (1 - \epsilon)\rho + \epsilon \text{Tr}[\rho] \mathbb{1}/2^n$. The contribution of a Pauli path of a noisy quantum circuit to the total output probability

thus decays with the number of nonidentity Pauli operators in it (the Hamming weight of s) as³⁷

$$\tilde{p}_C(x) = \sum_{s \in \mathbb{P}_n^{d+1}} (1 - \epsilon)^{|s|} f(C, s, x). \quad (202)$$

Aharonov *et al.* (2022) showed that the sum can be approximated by including only path weights $f(C, s, x)$ with Hamming weight $|s| \leq \ell$ incurring a TVD error on the order of $2^{-\Omega(\ell)}$ on average. They then showed that the truncated sum can be calculated efficiently using the knowledge that the low-weight Pauli paths are sparse in that most of them actually have weight 0, using ideas similar to those of Kushilevitz and Mansour (1993) and Bremner, Montanaro, and Shepherd (2017) on computing quantities with a sparse Fourier spectrum. This completes an algorithm for an approximate strong simulation. The algorithm for approximating the probabilities can be straightforwardly extended to an algorithm that also approximates all marginals (over bits of the measurement outcome x) of the truncated noisy distribution. Consequently, the marginal sampling algorithm can be used to sample from the output distribution up to TVD $2^{-\Omega(\ell)}$ in time $2^{O(\ell)}$.

For IQP circuits it has also been shown that it is possible to classically protect against noise (Bremner, Montanaro, and Shepherd, 2017): Using classical coding techniques, one can encode a smaller IQP circuit \mathcal{C} redundantly in a larger one \mathcal{C}' such that, even if local depolarizing noise is applied to the output of \mathcal{C}' , one can sample efficiently from a distribution arbitrarily close to the ideal output distribution of \mathcal{C} . It is not at all clear, however, how these coding techniques (which are similar to the concepts employed in the idea to use cryptography to verify IQP circuits discussed in Sec. V.D.2) can be extended beyond IQP circuits.

C. Simulating boson-sampling protocols

Classical simulation methods for boson-sampling naturally exploit the expression of the output probabilities in terms of the permanent or related matrix polynomials. The individual terms in those polynomials can be viewed as the weights of a Feynman path-integral expansion of the polynomial, and hence Feynman-type algorithms are natural candidates for the simulation of those schemes.

1. Computing probabilities: Permanents and Hafnians

Computing the output probabilities of boson sampling amounts to computing the permanent (46) for Fock input states and the Hafnian (49) for Gaussian input states. The naive run-time of computing the permanent of an $n \times n$ matrix scales linearly in the number of all permutations of n elements, given by $n!$, multiplied by the complexity of computing the product of n numbers, given by n^2 , while the space complexity is given by $O(n)$. Similarly, we can express the Hafnian as a sum over all perfect matching permutations of $2n$ elements, and hence the worst-case run-time is given by n^2 times the number of perfect matching

³⁷This reflects an analogous expression derived by Bremner, Montanaro, and Shepherd (2017) for noisy IQP circuits.

permutations $|\text{PMP}(2n)| = (2n - 1)!! = 1 \times 3 \times 5 \times \dots \times (2n - 1)$ (Gupt, Izaac, and Quesada, 2019).

These worst-case estimates can be significantly improved, however, via reexpressions of the permanent and the Hafnian, respectively. Indeed, Ryser (1963) found a way to reexpress the permanent via the principle of inclusion and exclusion as a sum of 2^n terms, and hence the complexity of computing the permanent is reduced to $O(n^2 2^n)$ and further to $O(n 2^n)$ using Gray codes. Alternative expressions for the permanent with the same number of terms, and hence the same complexity, were found by Glynn (2010) using the polarization identity for symmetric tensors and making use of partial derivatives.³⁸ Similarly, the Hafnian of a $n \times n$ matrix can also be computed in time $O(n^3 2^{n/2})$ (Björklund, 2012). The Ryser formula for the permanent can be further reduced to incorporate collision events, thereby reducing the number of terms from $O(n 2^n)$ to $\prod_i (n_i + 1)$, where n_i is the number of photons observed in the mode i (Shchesnovich, 2013; Tichy, 2014; Chin and Huh, 2018). Algorithms based on these reexpressions remain fastest for permanents and Hafnians (Wu *et al.*, 2018; Björklund, Gupta, and Quesada, 2019; Gupta, Izaac, and Quesada, 2019), allowing for the computation of matrix permanents of sizes up to 54×54 (Lundow and Markström, 2022). Their run times can also be further improved by exploiting specific structures such as the sparsity or the matrix bandwidth (Lundow and Markström, 2022).

A natural way to exploit path-integral expressions for an approximate computation of permanents and Hafnians is to randomly sample out paths and sum up their weights to construct a randomized estimator of the permanent or Hafnian. Gurvits (2003) did precisely that, making use of Ryser's or Glynn's formula to obtain an algorithm that takes time $O(n^2/e^2)$ to achieve an additive error $\pm \epsilon \|A\|^n$ estimate of the permanent of A . Aaronson and Hance (2012) generalized the algorithm, obtaining an improved run-time for permanents with repeated rows and columns, corresponding to bunching events, and derandomizing the algorithm for non-negative matrices. Furthermore, it can be extended to arbitrary input states (Yung, Gao, and Huh, 2019).

In specific instances one can also obtain multiplicative-error approximations in subexponential or even polynomial time. Such results delineate regimes in which the permanent is in fact not $\#\text{P}$ hard to approximate, and hence the sampling task will not be intractable either. Specifically, for non-negative matrices Jerrum, Sinclair, and Vigoda (2004) gave a Markov-chain Monte Carlo-based randomized algorithm that was able to approximate the permanent up to multiplicative error ϵ in time $\text{poly}(n, 1/\epsilon)$ while, deterministically, only an approximation factor of 2^n is currently achievable (Linial, Samorodnitsky, and Wigderson, 1998; Barvinok, 1999; Gurvits and Samorodnitsky, 2002). Using a method based on a Taylor-series approximation of the complex polynomial $f(z) = \ln\{\text{Perm}[J + z(A - J)]\}$, where $z \in \mathbb{C}$ and J is the matrix filled with ones, Barvinok identified certain regimes for which quasipolynomial relative-error approximations of the permanent and the Hafnian are possible. This is the case if the

function $f(z)$ is holomorphic on the unit disk for matrices with entries $a_{i,j}$ satisfying $|a_{i,j} - 1| \leq 0.19$ (Barvinok, 2016b), matrices with entries satisfying $\delta < a_{i,j} \leq 1$ (Barvinok, 2017), and diagonally dominant matrices (Barvinok, 2019). An interesting case is that of positive semidefinite matrices, as it has been shown that exactly computing the permanent of such matrices remains $\#\text{P}$ hard (Grier and Schaeffer, 2018), but multiplicative-error approximation algorithms in BPP^{NP} (Rahimi-Keshari, Lund, and Ralph, 2015) and with quasipolynomial run times (Anari *et al.*, 2017; Barvinok, 2020) exist in some circumstances. Building on the approach of Barvinok, Eldar and Mehraban (2018) showed that for random Gaussian matrices with nonzero but vanishing mean there is a quasipolynomial-time algorithm that approximates the permanent to within a multiplicative error.

Physically interesting cases include the case of low-rank matrices since such matrices determine the probabilities of outcomes with collisions: for constant rank, the corresponding permanents can be computed efficiently in the matrix dimension (Barvinok, 1996). Quesada (2019) and Quesada *et al.* (2019) analyzed the complexity of computing the output probabilities of Gaussian states with finite displacement. In this case, the probabilities correspond to so-called loop Hafnians (Björklund, Gupta, and Quesada, 2019), which can be viewed as counting the perfect matching of a graph with self-loops. Using similar techniques, Chabaud, Ferrini *et al.* (2021) and Chabaud and Walschaers (2022) found efficient algorithms for states with polynomial stellar rank and polynomial support over the Fock basis. Another physically relevant simplifying modification is to analyze the complexity of computing the outcome probabilities if the detectors can only distinguish between 0 and at least 1 photon, so-called threshold detectors. In this case, the output probabilities can be expressed in terms of what Quesada, Arrazola, and Killoran (2018) called the Torontonians. The complexity of directly computing the Torontonian is given by $O(n^3 2^n)$, which is equivalent to the complexity of directly computing the Hafnian. It remains an open question whether threshold detectors significantly reduce the complexity of simulating Gaussian boson-sampling experiments. Furthermore, the output probabilities of Gaussian boson sampling with local and shallow linear-optical circuits can be efficiently computed by making use of the banded structure of the adjacency matrix (Qi, Cifuentes *et al.*, 2020).

Exploiting the fact that for realistic experiments the number of modes is not much larger than the number of observed photons as required by the proofs of hardness (see Sec. IV.C.4.c) and the fact that threshold detectors are used that distinguish only between 0 and ≥ 1 photons, Popova and Rubtsov (2021) introduced an iterative series of approximations to the ideal outcome probabilities. To this end, they exploited the finding that low-order moments $\sum_k k^j p_n(k)$ can be efficiently computed with the complexity scaling exponential in j . Here $p_n(k)$ is the probability that photons have been detected in k of m detectors, conditioned on a total number n of photons. They then solved an inverse-moment problem to estimate $p_n(k)$, projecting that, up to $j = 4$ th order, probabilities of a $m = 100$ mode device can be estimated with 50% relative accuracy.

³⁸See Huh (2022) for a use of the Glynn formula in a quantum algorithm for permanent estimation.

2. Simulating the sampling task

Given the entirely different structure of the circuits in variants of boson-sampling schemes, the sampling algorithms used also differ in type from simulations of universal circuit sampling. An important further distinction in the quantitative comparison of classical simulation algorithms to actual experiments is the lack of a simple benchmark analogous to the XEB fidelity. As discussed in Sec. V.D, the most important way to verify boson-sampling experiments is state discrimination schemes, as well as certain efficiently computable quantities such as low-order correlations of the respective distributions. Consequently, in boson-sampling experiments it is also much less clear at which point quantum advantage has been reached experimentally.

The first competitive classical simulation algorithm for Fock state boson sampling used the previously described Markov-chain Monte Carlo method, giving rise to a much better run-time than the naive worst-case complexity (Neville *et al.*, 2017). This algorithm takes into account noise in actual devices, in particular, photon loss, which is the dominant source of errors. This approximate sampling algorithm was vastly improved by Clifford and Clifford (2018), who provided an exact boson-sampling algorithm with the same improved run-time of $O(n2^n + \text{poly}(m, n))$ as compared to the worst-case run-time of $O(\binom{m+n-1}{n}n2^n)$, where n corresponds to the number of photons and m is the number of output modes. In follow-up work by Clifford and Clifford (2020), this algorithm was further improved, achieving an average-case time complexity that is much lower when m is proportional to n . When $m = n$ specifically, the algorithm runs in time approximately $O(n1.69^n)$ on average. The sampling algorithms by Clifford and Clifford are based on ancestral or marginal sampling. The key insight of their algorithms is an expression of the low-order photon marginals in terms of permanents of smaller and smaller matrices, so the run-time of the algorithm is dominated by the final marginal, where a single permanent of the full $n \times n$ matrix needs to be computed that in the worst case is given by $O(n2^n)$. Altogether, these results indicate that Fock boson samplers require at least ~ 40 photons before one can hope to surpass the capabilities of currently available classical computers.

An exact algorithm for Gaussian boson sampling with threshold detectors (Quesada, Arrazola, and Killoran, 2018) that was implemented by Gupta *et al.* (2020) requires exponential space since the entire probability distribution needs to be saved. Quesada and Arrazola (2020) improved on this and devised an exponential-time exact sampling algorithm that uses only polynomial space and has a run-time $O(n^32^n)$ for generating a single sample with n photons. To achieve a run-time scaling proportional to the run-time required for a single Hafnian computation, Bulmer *et al.* (2022) and Quesada *et al.* (2022) gave algorithms with a further quadratic improvement, achieving a run-time $O(n^32^{n/2})$. The key idea of Quesada *et al.* (2022) was to first perform a virtual heterodyne measurement in all modes. Such a measurement can be efficiently simulated. One can iteratively replace the heterodyne outcomes with photon-number measurements and sample from the photon-number distribution conditioned on the heterodyne outcomes in the remaining modes. These

probabilities are described by loop Hafnians of matrices with increasing size, similar to how the algorithm of Clifford and Clifford (2020) expressed probabilities in terms of smaller permanents for standard boson sampling. The idea of an algorithm for Gaussian boson sampling with threshold detectors by Bulmer *et al.* (2022) is to simulate a photon-number-resolving measurement and then set all nonzero photon numbers in a sample to 1. In the dilute regime, this reduces the computation to a loop Hafnian of size $n \times n$ containing $2^{n/2}$ terms. Bulmer *et al.* (2022) then provided a construction that reduces sampling in the nondilute regime to sampling in the dilute regime by artificially introducing “submodes” for each detector.

Bulmer *et al.* (2022) also presented the most advanced implementation of near-exact sampling algorithms for Gaussian boson sampling with photon-number-resolving detectors. To this end, they implemented the ancestral sampling algorithm of Quesada *et al.* (2022) with a variety of improvements. Specifically, they reduced the run-time of computing loop Hafnians by making use of an inclusion-exclusion principle on pairs of photons and using a so-called finite difference sieve analogous to Glynn’s formula. Furthermore, they exploited threshold detectors explicitly in their sampling algorithm. For low photon density, simulating photon-number-resolving detectors and reducing collisions subsequently are advantageous for computing the Torontonian that exactly describe the output distribution with threshold detectors. Running on an $\sim 100\,000$ core supercomputer, they were able to simulate $m = 60$ modes with up to 80 photons observed by photon-number-resolving detectors with a mean time per sample of 3 s, and $m = 100$ modes with up to 60 click events with a mean time per sample of 8.4 s. Finally, they generated a single 92-photon event in $m = 100$ modes and photon-number-resolving detectors in 82 min.

A complementary approach was pursued by Villalonga *et al.* (2021). The idea is to sample from a distribution that reproduces the low-order mode marginals of the ideal target distribution. These low-order marginals can indeed be classically efficiently calculated since they are determined simply by a submatrix of the covariance matrix. The practical challenge is to efficiently sample from a distribution with the correct marginals. Villalonga *et al.* (2021) presented two heuristic approaches that were able to achieve this. The first heuristic employs a maximum-entropy principle that corresponds to a Boltzmann machine, i.e., a distribution of the form $p(\mathbf{z}) = (1/Z) \exp(\sum_i \lambda_i z_i + \sum_{i < j} \lambda_{i,j} z_i z_j + \dots)$, where Z is the partition function that normalizes the distribution. To find the correct parameters $\lambda_i, \lambda_{i,j}, \dots$, a mean-field approximation is used for the second order and costly log-likelihood minimization is employed for higher orders. Another method makes use of a greedy algorithm to generate samples with the correct low-order marginals with a cost exponential in the order of the marginal. Villalonga *et al.* (2021) implemented their sampler using the ideal second- and third-order marginals and compared it to the experiment of Zhong *et al.* (2021) using $m = 144$ modes and squeezing values that give rise to an average photon number of up to 66.9. The total-variation distance of the low-order marginal distributions of up to 14 modes compared to the corresponding ideal distribution is

lower than that of the experimental distribution. In a similar vein, first steps toward an approach analogous to that of Clifford and Clifford were taken by [Renema \(2020a\)](#), who computed low-order marginals in terms of photons rather than modes, potentially offering a better approximation of the ideal distribution.

Another variant of a spoofing algorithm for Gaussian boson sampling was recently proposed by [Martínez-Cifuentes, Fonseca-Romero, and Quesada \(2022\)](#) and exploited the fact that the quantum device is noisy. The idea is to replace the input squeezed states with so-called squashed states, that is, coherent states with vacuum fluctuations in one quadrature and larger fluctuations in the other. Linearly transformed squashed states are classical Gaussian states in that a photon-number measurement can be efficiently simulated classically. The definition of squashed states is motivated by the fact that loss in the network can be incorporated by replacing the initial squeezed states with squeezed thermal states. Squashed states are indeed those Gaussian states that best approximate squeezed thermal states and are at the same time classically simulable in the photon-number basis. [Martínez-Cifuentes, Fonseca-Romero, and Quesada \(2022\)](#) found that while the experiment of [Zhong *et al.* \(2020\)](#) can be spoofed by squashed-state Gaussian boson sampling in the sense that the correlations in the distribution match the ideal correlations better than the experiment, the more recent experiment of [Zhong *et al.* \(2021\)](#) cannot. However, these approaches cannot be applied to universal circuit sampling, because the approximation that reproduces marginals and correlations up to a constant order would be exponentially close to the uniform distribution due to the highly entangled nature of the output distribution.

Note also that the complexity of Fock boson sampling has been considered under locality constraints ([Deshpande *et al.*, 2018](#); [Maskara *et al.*, 2022](#)). In a certain setting, this structure renders the classical simulation of boson sampling efficient. [Oh, Lim, Fefferman, and Jiang \(2022\)](#) followed up on these results and derived general algorithms for Fock boson sampling and Gaussian boson sampling that exploit the graph structure of a linear-optical circuit. For a sufficiently small tree width of the interaction graph, i.e., in particular, for low-depth, geometrically local linear-optical circuits, this exact sampling is efficient.

3. Analysis of noise

In boson-sampling experiments with photons, the dominant sources of noise are losses of photons due to finite transmittivity of waveguides and other optical elements, finite distinguishability of the photons due to imperfect time or frequency synchronization between the single-photon sources, so-called mode mismatch. The asymptotic effects of these noise types has been studied extensively for photon loss and detector noise (dark counts) ([Rahimi-Keshari, Ralph, and Caves, 2016](#); [Oszmaniec and Brod, 2018](#); [García-Patrón, Renema, and Shchesnovich, 2019](#); [Moylett *et al.*, 2019](#); [Renema, Shchesnovich, and García-Patrón, 2019](#); [Qi, Brod *et al.*, 2020](#); [Oh *et al.*, 2021](#)) and partial distinguishability of the photons ([Shchesnovich, 2014](#); [Tichy, 2015](#); [Rahimi-Keshari, Ralph, and Caves, 2016](#); [Renema *et al.*, 2018](#); [Moylett *et al.*, 2019](#); [Renema, 2020b](#)). The overall observation

of these studies is that already comparably low noise levels drive the output probability distribution closer to distributions that are simulable with less effort.

An interesting “toy” noise model for boson sampling was considered by [Kalai and Kindler \(2014\)](#). In this model, additive Gaussian noise is applied to the random Gaussian submatrix of which the permanent is taken to compute the outcome probabilities of boson sampling; see Eq. (45). [Kalai and Kindler \(2014\)](#) showed that the collision-free output probabilities of boson sampling with a constant amount of such Gaussian noise can be approximated by sparse low-degree polynomials. This gives an efficient approximation algorithm for the noisy output probabilities with constant precision. This noise model turns out to be appealing: on the one hand, it “preserves the mathematical connection to random Gaussian matrices, used to establish hardness of boson sampling” ([Shchesnovich, 2019](#)). As [Shchesnovich \(2019\)](#) showed, on the other hand, this noise model is closely related to experimentally more relevant noise sources: it is equivalent to photon loss at the input of the interferometer and dark counts in the measurement that exactly compensate for the lost photons, as well as partial distinguishability of bosons.

While [Kalai and Kindler \(2014\)](#) did not provide a total-variation-distance bound on the approximate noisy distributions (with approximations given by the low-degree polynomial), [Shchesnovich \(2019\)](#) provided such a bound and showed that it can be made inverse polynomially small at a polynomial cost in the time it takes to compute the corresponding probabilities. Analogously to [Renema *et al.* \(2018\)](#) and [Renema \(2020b\)](#), they then argued that a Metropolis Markov-chain Monte Carlo algorithm can be used to efficiently sample from this distribution.³⁹ The result of [Shchesnovich \(2019\)](#) can thus be viewed as unifying several more specific previous results ([Arkhipov, 2015](#); [Leverrier and García-Patrón, 2015](#); [Aaronson and Brod, 2016](#); [Oszmaniec and Brod, 2018](#); [Renema *et al.*, 2018](#); [García-Patrón, Renema, and Shchesnovich, 2019](#); [Renema, Shchesnovich, and García-Patrón, 2019](#)) on the easiness and hardness of noisy boson sampling in certain noise regimes [see Table I of [Shchesnovich \(2019\)](#) for an overview] and gives rise to the following heuristic picture: Boson sampling with noise strength on the order of $\Omega(1)$ can be simulated classically to total-variation-distance error ϵ with polynomial effort in n and $1/\epsilon$. Conversely, for a noise strength scaling as $O(1/n)$ the simulation complexity remains the same as that of ideal boson sampling. In other words, constant local noise renders boson sampling classically simulatable, while local noise scaling inversely with the number of photons presumably remains classically intractable, with the intermediate regime remaining open.

With the Fourier picture of [Bremner, Montanaro, and Shepherd \(2017\)](#), [Gao and Duan \(2018\)](#), and [Aharonov *et al.* \(2022\)](#) in mind, [Oh, Jiang, and Fefferman \(2023\)](#) built upon those prior works and provided a fully provable algorithm for sampling from the output distribution of Fock boson sampling with Gaussian noise according to the model of [Kalai and](#)

³⁹Notice, though, that this falls short of an efficiency proof since none of those works actually bounds the mixing time of the corresponding Markov chain.

Kindler (2014). Their sampling algorithm is based on the low-degree polynomial decomposition of Kalai and Kindler (2014), which they showed also works for the marginals of Fock boson sampling written in first quantization. This allowed them to compute all marginals of the low-degree approximation to the noisy probabilities and hence provides an approximate sampling algorithm for the noisy distribution. At a constant noise rate, the total run-time of the algorithm is quasipolynomial and given by $n^{O(\log n, \log(1/\epsilon), \log(1/\delta))}$ per sample to within a total-variation distance $\epsilon > 0$ for a proportion $1 - \delta$ of Haar-random unitaries, assuming the hiding condition $m \in \omega(n^5)$.⁴⁰ It is unclear whether this algorithm extends to natural noise sources such as photon loss and distinguishability, however.

A compelling physical picture for why noise renders classical simulations tractable was developed by Renema, Shchesnovich, and García-Patrón (2019) and Renema (2020b), who conceived of the boson-sampling distribution as arising from interference processes with increasing order. The effect of physical noise, including, in particular, photon loss and photon distinguishability, in a precise way results in higher-order interference terms to contribute exponentially less. This gives rise to a distribution that just arises from low-order interference. The resulting distribution can therefore be classically simulated efficiently. A similar intuition, albeit on the level of individual modes, is followed using the simulation algorithm of Villalonga *et al.* (2021). Shchesnovich (2021, 2022) showed, however, that the output data from classical simulation methods based on lower-order multiboson interferences can be efficiently distinguished from a noisy boson-sampling distribution since the higher-order correlations remain sufficiently significant. This matches the observation of Zhong *et al.* (2021), who found that higher-order correlations remain present in experimental data.

To conclude, the fact that noise renders the classical simulation of imperfect devices less computationally demanding adds a challenge to the experimental realization of quantum random sampling schemes that show an unambiguous quantum advantage.

VIII. PERSPECTIVES

The field of quantum random sampling has now reached a state in which the theoretical foundations are thoroughly explored, and important but extremely difficult open questions have been identified. It has reached a state in which we have seen first demonstrations on the verge of classical intractability and first pushbacks from classical algorithms. In this review, we have discussed these theoretical and practical aspects of quantum random sampling.

But what is the road ahead? Some features of this road are clear; important technical questions such as approximate

⁴⁰To achieve a provable polynomial-time algorithm as with universal circuits, the total noise rate has to scale like $1 - x^\gamma$, with $\gamma = \Omega(\log n)$ and a constant $x \in [0, 1)$. Kalai and Kindler (2014) argued that this is also the fair comparison since a constant noise rate per gate results in an overall noise that scales with the number of gates.

average-case hardness remain to be tackled, and quantum devices and classical algorithms alike are going to be further improved. At the same time, the leap from demonstrations of quantum computational advantage via quantum random sampling or other means of achieving a practically useful task with a quantum advantage seems large.

This section is more than an outlook. While we summarize the key open question, we also provide summaries of ideas to highlight interesting future directions of the field. We start by summarizing the key open questions in the field of quantum random sampling, most of which appeared in previous sections. We then take a broader perspective on the field of quantum advantages in general and quantum random sampling schemes in particular to see what questions have already been comprehensively settled, what is ahead, and what are reasonable next steps. In particular, we begin our outlook by drawing connections between quantum random sampling and other fields, such as quantum simulation. Finally, we sketch some ideas that have been developed with the goal of practical applications of quantum random sampling in mind. These applications make direct use of either the randomness of quantum random sampling or the programmability of a quantum random sampler in order to solve a specific task.

A. Open questions on quantum random sampling

Throughout this review, we have highlighted important open technical questions regarding our understanding of quantum random sampling. We summarize some of the most important ones here.

1. Understanding random quantum circuits better

From the perspective of the computational complexity of quantum random sampling, the key open question is to prove approximate average-case hardness, as discussed in Secs. IV.D.5 and IV.D.6. Currently approximate average-case hardness is a conjecture that is based solely on the lack of efficient classical simulation algorithms and the observation that random instances do not offer any additional structure that a classical simulation algorithm might be exploited to perform better than in the worst case. While we have progressively moved forward on this question by making polynomial interpolation techniques more robust (Bouland, Fitzsimons, and Koh, 2018; Movassagh, 2020; Bouland *et al.*, 2022; Kondo, Mori, and Movassagh, 2022; Krovi, 2022), there remain fundamental barriers to improving this result to the required robustness $O(2^{-n})$, as discussed in Sec. IV.D.6. It seems that, from this point onward, polynomial interpolation alone will not be able to help us solve the question of approximate average-case hardness, and new proof ideas will be required. The development of new methods, while most pressing, is also elusive and constitutes a major challenge that reaches beyond the field of quantum random sampling schemes all the way into the midst of computational complexity theory. For example, Aaronson and Arkhipov (2013) suggested making use of a restricted class of polynomials not closed under addition that are at the same time able to capture the quantity of interest.

We now zoom out from the details of the proof of robust sampling hardness and consider the task of sampling from the output distribution of a random quantum circuit up to a

constant total-variation-distance error. For such an overall constant additive error on the global distribution to be achieved, the gate errors need to scale inversely as $1/(m + 2n)$ with the total number m of gate applications and n single-qubit state preparations and measurements. But in experiments the gate application, state-preparation, and measurement errors typically do not scale in the size of the system, or circuit, but rather are fixed using physical details. This raises the question of what the optimal trade-off for achieving a quantum advantage is in terms of circuit depth and system size. While random short depth circuits might be easy to sample from with a global error budget (Napp *et al.*, 2022), too large of a circuit will incur a large amount of errors that renders classical simulation trivial. This is why the detailed study of noise in random circuits and its effect on the output distribution is paramount to better understanding the computationally most difficult regimes. First steps toward this were recently taken by the complementing approaches of Dalzell, Hunter-Jones, and Brandão (2021) and Deshpande, Niroula *et al.* (2022). These two works study the convergence to the white-noise or uniform distribution in the regimes of low gate noise $\epsilon \in \tilde{O}(1/n)$ or constant gate noise, respectively. A better understanding of how different types of experimentally relevant noise affect the output distribution of typical random quantum circuits is thus paramount to optimizing the parameters in a demonstration of quantum advantage.

2. Verification beyond XEB

The issue of noise in random quantum circuits directly leads to the next open question. In Sec. V, we discussed in what sense samples from random quantum circuits can be verified. The standard measure of quantum advantage as of today is the linear XEB fidelity (162). On the one hand, this is because it offers the best available compromise between being practically viable and providing a meaningful benchmark for achieving a nontrivial task on the quantum device. On the other hand, it is because it provides a unified view for the use of quantum random sampling as a benchmark of a quantum device and as a means to demonstrate a computational advantage. However, neither interpretation of XEB fidelity is fully understood.

Coming from the perspective of quantum advantage demonstrations, there is the question under which circumstances cross-entropy type measures (and particularly the XEB fidelity) can yield certificates for the global distribution. The logarithmic XEB fidelity, for instance, provides rigorous bounds on the total-variation distance only if the noise in the device is such that it increases the entropy of the ideal distribution (Bouland *et al.*, 2019). But estimating the entropy of the noisy distribution is an infeasible task in itself.

Coming from a practical perspective of device development and characterization, the question remains to identify in which settings random quantum circuits can be used to benchmark noise in the quantum circuit. As discussed in Sec. V.B.3, Y. Liu *et al.* (2021c) made some first steps toward understanding how a noise parameter can be extracted from the XEB fidelity for the case of global Pauli noise via a perturbative analysis in the noise parameter. Going beyond perturbative methods, further steps in this direction might

make use of the framework of Fourier analysis for randomized benchmarking (Helsen *et al.*, 2022). Ultimately, one wants to analyze gate-dependent noise channels in a local quantum circuit.

Finally, we mention that the most important problem with the use of XEB to verify quantum random sampling is the fact that evaluating XEB-like measures, while sample efficient, incurs the exponential computational cost of estimating some of the target probabilities. Going beyond XEB, an interesting open problem is whether the no-go result prohibiting sample-efficient verification of flat distributions discussed in Sec. V.A can be circumvented in random sampling schemes with larger second moments. Is there any “room in the middle” between exponentially flat distributions that are hard to verify but anticoncentrate and polynomially concentrated distributions that do not anticoncentrate but are sample-efficiently verifiable? If there is, then distributions could exist that we can sample-efficiently verify from classical samples and that we can sample from efficiently on a quantum computer but cannot efficiently sample from on a classical computer; see also Hangleiter *et al.* (2019). Works such as the one by Morimae (2017) proving anticoncentration of the DQC1 model without resorting to a second-moment bound might yield some leeway in this direction.

But one may also ask whether there are fully efficient ways of verifying that a classically intractable task has been achieved via quantum random sampling without resorting to directly verifying the total-variation distance. We have mentioned ideas for the verification of quantum samplers that make use of cryptographic secret hiding (Shepherd and Bremner, 2009) in a delegated scheme. But such ideas remain prone to classical attacks (Kahanamoku-Meyer, 2019) or remain orthogonal to the spirit of quantum random sampling as a specifically simple, unstructured task that is executed on a given quantum device. That said, for simple tasks, proofs of quantumness that might not be too far from the realm of practical feasibility can be devised using such ideas (Hirahara and Le Gall, 2021; Zhu *et al.*, 2021; Kahanamoku-Meyer *et al.*, 2022; Liu and Gheorghiu, 2022). Interesting progress in the direction of merging these worlds with public verifiability of NP problems such as factoring was recently made by Yamakawa and Zhandry (2022). It remains an interesting question to further explore the possibility of verifying quantum random sampling efficiently.

B. Developing novel schemes

Moving beyond better understanding the current schemes, there is the overarching question of how quantum random sampling schemes can be extended beyond their current realm of applicability. This both regards the extension from digital quantum devices to analog ones and leads to a larger error resilience.

1. Improving error resilience

Given all the strengths of the various approaches to quantum random sampling, it is also limited in its capacity to demonstrate quantum speedups. This is because these schemes do not allow for any type of error correction, making

a region that is nontrivially accessible with finite errors limited. Going from relative to additive errors has been a tremendous technical achievement, matching complexity-theoretic arguments more closely with experimental desiderata, but it still falls short of capturing fully realistic errors. The central challenge one has to overcome when realizing quantum supremacy is thus to bring this barrier down as far as possible such that the computing capabilities of classical computers can be surpassed before the barrier is hit.

Ultimately, one wants to make the hardness of quantum sampling robust to constant local errors. This can indeed be achieved for universal computations using quantum error-correction codes. However, quantum error correction is intrinsically based on the continuous measurement of error syndromes, giving information about which errors have occurred during one cycle of the computation. Those errors then need to be actively corrected, requiring an elaborate machinery that is again well outside the realm of what we envision the context of quantum random sampling to be. In addition, from a conceptual point of view quantum random sampling is intrinsically based on a global property of the outcome state, namely, the full probability distribution. To make this global property robust to constant local errors will therefore likely require one to invoke a global error-detection machinery such as that of [Bremner, Montanaro, and Shepherd \(2017\)](#).

One might think that coherent errors do not constitute a specifically grave problem for quantum random sampling schemes since, say, Pauli errors can often simply be absorbed in the random ensemble, giving rise to a different computation distributed according to the same ensemble. We stress, however, that to maintain hardness of sampling we actually need to know how the circuit has changed due to the errors. In other words, the errors need to be “heralded.” But continuous measurements of syndromes complicate the computation significantly. Conversely, if the ongoing computation is not continuously measured in every gate cycle, it is not clear under which circumstances such a “heralded noise model” is actually realistic. Finding a way around this obstacle, possibly using error detection and *post hoc* corrections, is the major challenge in making quantum random sampling schemes robust to physical noise and thus scalable.

Further intuition on the resilience of quantum circuits to local errors is also provided by the analysis of constant-depth quantum circuits: [Bravyi, Gosset, and König \(2018\)](#) showed that constant-depth quantum circuits are more powerful than their classical counterparts. Any classical probabilistic circuit composed of bounded fan-in gates that solves what [Bravyi, Gosset, and König \(2018\)](#) called the two-dimensional hidden linear function problem with high probability must have a depth that is at least logarithmic in the system size. In contrast, the same problem can be solved with certainty by a constant-depth quantum circuit that is composed of one- and two-qubit quantum gates that act on a two-dimensional lattice. This scheme is robust to noise in that the aforementioned separation in computational power persists even when the shallow quantum circuits are restricted to three dimensions and are corrupted by noise ([Bravyi et al., 2020](#)). Technically, the argument supporting this conclusion is rooted in ideas on the generation of a

long-range entanglement in noisy three-dimensional cluster states ([Raussendorf, Bravyi, and Harrington, 2005](#)).

In a similar spirit, the first ideas for using nonadaptive error correction by embedding a computation in an error-correction code have been made ([Fujii, 2016](#); [Kapourniotis and Datta, 2019](#)) and, indeed, if experimental errors remain within the specific error model considered, sampling hardness remains. However, robustness may be lost since the distribution for which an approximate average-case hardness conjecture for the outcome probabilities holds has significantly changed compared to the non-error-corrected distribution.

2. Relation to analog quantum simulation

Aside from the largely technical open questions discussed thus far, another avenue for making progress en route to larger-scale implementations of quantum random sampling is to connect it to the setting of analog quantum simulators. Such devices offer a limited amount of control, but often a large number of coherently and highly accurately controlled quantum degrees of freedom, which in several instances cannot be simulated by even the best classical algorithms ([Trotzky et al., 2012](#); [Braun et al., 2015](#); [Choi et al., 2016](#); [Debnath et al., 2016](#); [Ebadi et al., 2021](#)).

Along these lines a reasonable goal would be to prove a rigorous complexity-theoretic separation for a task that is natural in a physics mindset in general, and quantum simulations in particular. Quantities that first come to mind here are measurements of k -point correlation functions of the type $\langle b_i^\dagger b_j \rangle$. First steps toward this were taken by [Novo, Bermejo-Vega, and García-Patrón \(2021\)](#), who showed that one can run a Stockmeyer argument for the task of reproducing the statistics of an energy measurement of a local Hamiltonian. Deviating from the mindset of quantum random sampling, [Baez et al. \(2020\)](#) showed a quantum advantage for the estimation of dynamical structure factors, providing the insight that performing measurements on quantum states arising from time evolution under local Hamiltonians is BQP complete ([Nagaj and Wocjan, 2008](#); [Vollbrecht and Cirac, 2008](#); [Nagaj, 2012](#)) closer to experimental reality.

These works are based simply on the assumption that quantum computers are more powerful than classical computers and therefore do not offer independent evidence for this separation. In technical terms, they show a much weaker complexity-theoretic consequence than a collapse of the polynomial hierarchy, namely, that $\text{BPP} = \text{BQP}$. Coming from a complexity-theoretical perspective, they are thus begging the question as, from this perspective, one would like to precisely collect evidence that $\text{BPP} \neq \text{BQP}$. Accepting this, it is still not obvious whether one would expect average-case hardness of the respective tasks for problems in BQP. When one comes from a more practically minded perspective, accepting $\text{BQP} \not\subseteq \text{BPP}$ is a fair assumption. Such ideas may thus help to demonstrate quantum advantages for tasks that are more useful than sampling alone. From a technological perspective, it is interesting to see whether one can reach the regime in which quantum advantages in this sense are conceivable.

Another interesting perspective that has been considered in this context is the relation of sampling hardness to physical

phenomena. For instance, phase transitions in sampling complexity of two-dimensional bosonic lattice systems were considered by [Deshpande *et al.* \(2018\)](#) and [Maskara *et al.* \(2022\)](#). Here the idea is to vary a physical parameter of the system, in this case, the spacing between bosons in the initial state, and to consider the complexity as a function of time when evolving the system. In a similar vein, [Ehrenberg *et al.* \(2022\)](#) studied transitions in the complexity of sampling from the output distribution of many-body-localizing time evolution. In such approaches, the hope is to narrow down and better understand the physical mechanisms underlying sampling complexity.

C. Toward applications of quantum random sampling

What is next? On the road toward practically useful quantum computers, quantum random sampling schemes are an important stepping stone. But quantum random sampling has been conceived as a proof-of-principle task to show that quantum devices have the capability to computationally outperform classical computers and nothing more. It is therefore not set up to realize practically interesting applications in their own right. However, a natural next question is whether one can exploit the provable speedup over classical sampling algorithms on the specific random sampling task for relevant practically motivated applications. Here we discuss some of these first steps at identifying applications of quantum random sampling.

Roughly speaking, these applications of quantum random sampling fall into two categories. On the one hand, there are applications that exploit the intrinsic quantum randomness of typical quantum circuits. Such applications make use of the fact that the output distributions of random quantum circuits are highly unstructured or, in technical terms, have a high min-entropy bound, as explained in Sec. V.A. On the other hand, there are applications that take programmable quantum random sampling devices as their starting point and ask the following question: What applications can those devices be used for? In such applications, the structure of the output distributions is explicitly exploited to solve a computational task or to serve as a subroutine in an algorithm solving such a task. In the following we explain some of the ideas in this mindset with the goal of giving the interested reader a concrete idea about potentially engaging directions of study.

1. Exploiting randomness

One of the most promising near-term applications of quantum devices is the generation of certified random numbers. In the classical world, bits that are perfectly random in that they are unpredictable not only to the user of the device but to any observer cannot be realized in principle, because the laws of classical physics are deterministic. In practice one has to therefore rely on (albeit possibly extremely weak and plausible) hypotheses to design pseudorandom-number generators. Going beyond this, so-called true random-number generators exploit physical processes from the realm of classical physics that are hard to predict. Quantum random-number generators make use of the intrinsic randomness offered by quantum mechanics. The possibility of harnessing

this randomness makes quantum technologies attractive as a means of generating certified random numbers ([Acín and Masanes, 2016](#)) that cannot be predicted by any adversary.

Given that the output distributions of random quantum circuits have a high min-entropy, statistically verified quantum random sampling would naturally give rise to a large number of intrinsically random bits. In the absence of such statistical tests, [Aaronson \(2018, 2019\)](#) proposed protocols for certified randomness that use universal circuit sampling and the XEB benchmark. The proof of security of the proposed protocols is based on a strong and highly nonstandard complexity-theoretic conjecture on the hardness of what [Aaronson \(2019\)](#) called the long list quantum sample verification (LLQSV) problem. This problem asks one to distinguish exponentially many output bit strings from a quantum random sampler, given by an oracle, from uniformly random numbers. More specifically, the conjecture is that LLQSV is not in a complexity class called QCAM which contains AM, BQP, etc.

[Bassirian *et al.* \(2021\)](#) provided complexity-theoretic evidence in support of the classical intractability of this problem. This support holds to the same standard as the evidence for computational hardness of achieving a high XEB score via XHOG and HOG. They prove two statements regarding the hardness of LLQSV or, in other words, the hardness of distinguishing the high min-entropy samples from the quantum device from uniformly random samples. They do so in the black-box model in which query access to a random Boolean function is granted, instead of a random circuit. First, [Bassirian *et al.* \(2021\)](#) proved an average-case linear min-entropy bound for quantum algorithms that pass a XEB-like test. Second, they showed that no BQP or PH algorithm can solve the LLQSV problem, thereby individually showing separations from major classes contained in QCAM. To do so, they reduced it to a variant of the so-called correlation problem introduced by [Aaronson \(2010\)](#). These results imply that if one believes that quantum circuits viewed as random functions are sufficiently unstructured, then quantum random sampling can generate random samples that are certified by a XEB-like test.

In a different vein, the fact that quantum states prepared by random quantum circuits are highly entangled might be useful in *quantum metrology*. In this context, it is not the flatness of the classical output distribution that is exploited, but rather the full quantum state. Along these lines, [Oszmaniec *et al.* \(2016\)](#) studied how useful random bosonic states are for quantum metrology. They indeed found that a close to optimal Heisenberg scaling is typically achieved. [Valido and García-Ripoll \(2021\)](#) explored the phase sensitivity of generic linear interferometric schemes using Gaussian resources and measurements in what could be called boson-sampling-inspired strategies. Multimode metrology via a variant of Gaussian boson sampling was studied by [Guanzon, Lund, and Ralph \(2021\)](#). Finally, it has been suggested that the high min-entropy bound of the output distributions can be exploited to devise cryptographic schemes ([Nikolopoulos, 2019](#); [Huang, Kok, and Luo, 2021](#); [Z. Huang *et al.*, 2021](#)).

2. Exploiting structure

Rather than exploiting the randomness of quantum random samplers, one may alternatively program such devices in a

bespoke way in order to solve computational problems. Such applications make use of the structure in the output probability distributions, an idea which has been most developed to date for variants of Gaussian boson sampling. We now give two examples that argue along these lines. While the first example makes specific use of samples, the second example uses samples in order to estimate probabilities.

a. Using samples to solve graph problems

A natural class of problems that can be studied in the context of Gaussian boson sampling is graph problems. This is because the Hafnian (49) of an adjacency matrix of a graph equals the number of *perfect matchings* of that graph, that is, the number of disjoint sets of edges in which every vertex of the graph is connected to exactly one edge.

As an example, consider the so-called densest k -subgraph problem (Arrazola and Bromley, 2018). This problem asks one, given a graph G with n vertices, to find the subgraph with $k < n$ vertices that has the largest number of edges. Recall that the probability $P_{\text{GBS},U}(S)$ [Eq. (47)] of obtaining a collision-free output pattern S in Gaussian boson sampling is determined by the Hafnian of a submatrix M_S of a certain matrix M [Eq. (48)] that depends on the covariance matrix of the input state. Given the adjacency matrix $A \in \{0, 1\}^{m \times m}$ of G , we can now choose the squeezing parameters and linear-optical unitary in order to “program” that matrix to be

$$M = c(A \oplus A), \quad (203)$$

where $c < \lambda^{-1}$ and λ is the largest eigenvalue of A . The corresponding Gaussian state is pure and hence is a valid state that can be prepared in Gaussian boson sampling. The output probabilities postselected on the collision-free subspace will then be proportional to $|\text{Haf}(A_S)|^2$, where A_S is a submatrix of A determined by the outcome S or, equivalently, the adjacency matrix of a subgraph of G with vertices selected by S . Since the Hafnian of an adjacency matrix equals the number of perfect matchings of the corresponding graph, the larger the number of perfect matchings in a subgraph, the more likely its corresponding sample is obtained as an output in Gaussian boson sampling.

The next step is to establish a connection between the number of perfect matchings in a graph and its density. On an intuitive level, the number of perfect matchings corresponds to the density of a graph since a graph with many perfect matchings will have many edges. Indeed, the number of perfect matchings provides a lower bound to the number of edges in the graph (Aghabali *et al.*, 2015). Consequently, by programming the quantum device in an appropriate way, one can sample from a distribution that has a bias in favor of dense subgraphs. For this reason, stochastic algorithms (Lee *et al.*, 2010) for the densest k -subgraph problem that make use of uniform randomness can be enhanced by having access to samples drawn from the output distribution of Gaussian boson sampling. This was recently demonstrated in a proof-of-principle experiment using time-bin-encoded GBS (Sempere-Llagostera *et al.*, 2022).

Arrazola, Bromley, and Rebentrost (2018) followed a similar line of thought upon introducing an NP-hard problem referred to as Max-Haf. They showed that access to samples

from the Gaussian boson-sampling distribution defined by the probabilities $P_{\text{GBS},U}(S)$ of obtaining the output pattern S can enhance classical stochastic algorithms for this problem. They not only presented the idea and compared the performance of this algorithm with classical algorithms based on uniform randomness but also reviewed numerical data from use cases. Brádler *et al.* (2018) discussed the problem of actually finding perfect matchings of arbitrary graphs enhanced by having access to samples from Gaussian boson sampling.

Coming from a perspective of quantum machine learning, Jahangiri *et al.* (2020) proposed an application of quantum random sampling to statistical modeling. Havlíček *et al.* (2019) showed how minimally enhanced IQP circuits might be used to enhance the feature space of machine-learning algorithms for supervised learning. More concretely, samples from Gaussian boson samplers can be utilized to construct feature vectors of graphs that give rise a natural measure of similarity between graphs (Schuld *et al.*, 2020). The connection to quantum-enhanced machine learning was made even more explicit by Banchi, Quesada, and Arrazola (2020), who showed how Gaussian boson-sampling devices can be trained in the following sense: Analytical gradient formulas for the GBS distribution can be exploited when training devices using gradient-descent-based methods. Finally, Chabaud, Markham, and Sohbi (2021) studied supervised learning using minimal extensions of Fock boson sampling.

b. Estimating physical quantities using Gaussian boson samplers

Using the samples from a quantum device in order to estimate outcome probabilities is the basis of a line of thought initiated by Huh *et al.* (2015). When preparing *displaced squeezed states* at the input of a linear-optical device, the output probabilities of a Gaussian boson sampler can be used to estimate so-called Franck-Condon factors, which represent the transition frequencies of molecular vibronic spectra. This is a problem for which no efficient classical algorithm is currently known. In this way, Franck-Condon factors can be estimated from Gaussian boson-sampling data. Following up on this, Jnane *et al.* (2021) suggested an analog quantum simulation of molecular vibronic spectra based on boson-sampling-like schemes, incorporating the non-Condon scattering operation with a quadratically small truncation error. Pursuing a similar aim, *molecular docking* was studied by Banchi *et al.* (2020), who suggested that Gaussian boson samplers provide insights into molecular docking configurations, which are spatial orientations that molecules assume when they bind to larger proteins. Connecting these ideas to the loop Hafnian picture of the output probabilities, Quesada (2019) suggested estimating Franck-Condon factors by counting perfect matchings of graphs with loops. To this end, he showed that the Franck-Condon factor associated with a transition between initial and final vibrational states in two different potential energy surfaces can be reduced to the number of perfect matchings of a suitable weighted graph with loops. Clements *et al.* (2018) explored the impact of experimental imperfections on the performance of the protocol of Huh *et al.* (2015) for performing quantum simulations of vibronic spectroscopy, providing stringent benchmarks that have to be met by experiments. This work also discussed

practically meaningful examples such as Franck-Condon factors for vibronic transitions in molecules such as tropolone. Departing from the previously mentioned prescriptions in a different way, Wang *et al.* (2020) implemented a small-scale instance of the protocol of Huh *et al.* (2015) in a two-mode superconducting device.

Known classical simulation methods for boson sampling with sparse outputs, as presented by Roga and Takeoka (2020) and Oh, Lim, Fefferman, and Jiang (2022), have challenged these results in that it was argued that the instances considered when sampling from Franck-Condon factors is often sparse in the appropriate sense. Technically, this work demonstrated that the computationally costly support detection step, i.e., the localization of the largest element from a long list, can be reduced to solving an Ising model that can be solved in polynomial time under suitable conditions. Oh, Lim, Wong *et al.* (2022) followed up on this line of thought by presenting a quantum-inspired classical algorithm for molecular vibronic spectra. Technically, they found an exact solution of the Fourier components of molecular vibronic spectra at zero temperature using a positive P -representation method. The resulting algorithm resembles that of Baiardi, Bloino, and Barone (2013).

Both of these lines of work are contributions that show the potential of achieving computational advantages in practically motivated problems using Gaussian boson-sampling devices. At the same time, as the classical algorithms by Roga and Takeoka (2020), Oh, Lim, Fefferman, and Jiang (2022), and Oh, Lim, Wong *et al.* (2022) showed, it may be possible to find classical algorithms that are efficient for those instances of Gaussian boson sampling that are used to solve a specific computational problem. For these instances there is no complexity-theoretic reason analogous to the polynomial-hierarchy collapse to believe in a quantum speedup. Rather, we are now moving into the realm of comparing quantum algorithms with the best classical algorithm for specific problems, as one would also expect when considering practically relevant problems.

D. Conclusions

In this review, we have provided a comprehensive overview of the efforts aimed at understanding in theory and demonstrating in practice the computational advantage of quantum random sampling over classical computation. Quantum random sampling schemes are particularly attractive, as they are simple conceptually and have comparably small experimental desiderata. On the highest level, there seem to be two main lessons that can be drawn from the research efforts that are the focus of this review.

One of those lessons is of a *foundational nature*. Ultimately, the questions asked in endeavors to show quantum advantages with quantum random sampling schemes follow up on the thoughts of Turing about the intertwining of the complexity of processes in nature and about what can be computed using the mechanisms allowed by natural laws. Boldly stated, the question on the desk is as follows: What is, after all, the computational nature of nature? In more elaborate words, can all naturally feasible computations be efficiently described within a classical Turing machine model? The

extended Church-Turing thesis asserts that this is indeed the case, but it is challenged by the onset of physical quantum computers. We have walked a long route along this path, starting with theoretical arguments against the validity of the extended Church-Turing thesis and progressing to the question of how to verify those claims experimentally. Further efforts in realizing sampling schemes will shine light onto this matter.

The other lesson relates to *technological issues*. Present efforts toward realizing quantum advantage schemes cannot be underestimated in their importance of providing guidance for the next steps to be taken in the development of quantum technologies. The experimental demonstration of quantum random sampling schemes provides an impetus for achieving unprecedented control in experiments and for pursuing large-scale quantum computations. The next steps are to pursue practically motivated quantum algorithms on such quantum devices, a process that is well under way. Some schemes can be seen as variations of quantum random sampling schemes that address pragmatically motivated questions. This applies to photonic experiments that explore vibronic spectra (Clements *et al.*, 2018; Wang *et al.*, 2020), implement variational schemes (Peruzzo *et al.*, 2014), or examine quantum simulations of processes in statistical physics (Somhorst *et al.*, 2023). The layout of the superconducting quantum advantage experiment of Arute *et al.* (2019) has been made forward compatible with a realization of the surface code (Satzinger *et al.*, 2021). Indeed, arguably the most substantial next step will be to achieve fault tolerance in quantum computing, a step that may still be relatively far off. The efforts on quantum random sampling schemes can be seen as a first milestone in this direction.

In a similar way, the questions of “what next” apply to theoretical research. Steps have been taken toward developing protocols that show a more practically minded quantum advantage. Quantum approximate optimization algorithms in their various variants suggest addressing questions of combinatoric optimization (Farhi, Goldstone, and Gutmann, 2014; Zhou *et al.*, 2020), and variational quantum eigensolvers may solve variational principles beyond the capabilities of classical efficient variational methods (McClean *et al.*, 2016). These applications are thought to be pursued without quantum error correction, but the key question remains open as to what noise levels quantum devices may ultimately tolerate while maintaining a quantum advantage (Stilck França and García-Patrón, 2021). The efforts toward achieving quantum advantages can be seen as a first stepping stone en route to building useful quantum computers and an invitation to master the next hurdle along that route.

ACKNOWLEDGMENTS

We thank Scott Aaronson, Dorit Aharonov, Juani Bermejo-Vega, Adam Bouland, Ulysse Chabaud, Abhinav Deshpande, Ish Dhand, Adam Ehrenberg, Bill Fefferman, Raúl García-Patrón, Christian Gogolin, Alexey Gorshkov, Jonas Haferkamp, Aram Harrow, Marcel Hinsche, Marios Ioannou, Martin Kliesch, Austin Lund, Arthur Mehta, Ashley Montanaro, Tomoyuki Morimae, Hakop Pashayan, Jelmer Renema, Nicolás Quesada, Robert Raussendorf, Ingo Roth, Martin

Schwarz, and Barbara Terhal for numerous exciting and enlightening discussions of quantum random sampling in recent years. We also thank Sergio Boixo, Bill Fefferman, Jonas Haferkamp, Chao-Yang Lu, Ingo Roth, Pedram Roushan, and especially Ulysse Chabaud and Abhinav Deshpande for the helpful and extensive comments on drafts of this work. D. H. is grateful specifically to Marcel Hinsche for insights into Gaussian boson sampling, to Scott Aaronson and Marcel Hinsche for discussions of the hiding problem, to Abhinav Deshpande for discussions surrounding the comparison of near-term quantum and large-scale classical algorithms, to Jonas Helsen and Ingo Roth for discussions about XEB fidelity, and to Bill Fefferman for discussions regarding anticoncentration and the relevance of the total-variation distance. D. H. acknowledges financial support from the U.S. Department of Defense through a QuICS Hartree Fellowship. This work was completed while D. H. was visiting the Simons Institute for the Theory of Computing. J. E. acknowledges funding from the DFG (CRC 183 and EI 519/21-1), the BMBF (PhoQuant, QPIC-1, DAQC, HYBRID, MUNIQ-ATOMS, FermiQP, and QSolid), the BMWK (EniQmA and PlanQK), QuantERA (HQCC), Munich Quantum Valley (MQV-K8), and the Einstein Foundation (Einstein Research Unit on Quantum Devices). He has also received funding from the EU's Horizon 2020 research and innovation program under Grant Agreement No. 817482 (PASQuanS2, Millennium).

Note added.— Sections II–V build on previously unpublished parts of the work of Hangleiter (2021).

REFERENCES

- Aaghabali, M., S. Akbari, S. Friedland, K. Markström, and Z. Tajfirouz, 2015, “Upper bounds on the number of perfect matchings and directed 2-factors in graphs with given number of vertices and edges,” *Eur. J. Comb.* **45**, 132–144.
- Aaronson, S., 2005, “Quantum computing, post-selection, and probabilistic polynomial-time,” *Proc. R. Soc. A* **461**, 3473–3482.
- Aaronson, S., 2010, “BQP and the polynomial hierarchy,” in *Proceedings of the 42nd ACM Symposium on Theory of Computing (STOC '10)*, Cambridge, MA, 2010 (Association for Computing Machinery, New York), pp. 141–150, [10.1145/1806689.1806711](https://doi.org/10.1145/1806689.1806711).
- Aaronson, S., 2018, “Certified randomness from quantum supremacy,” PowerPoint presentation (accessed on May 14, 2022), <https://www.youtube.com/watch?v=hf7-Elx1Y4w>.
- Aaronson, S., 2019, “Aspects of certified randomness from quantum supremacy,” PowerPoint presentation (accessed on September 4, 2020).
- Aaronson, S., and A. Arkhipov, 2013, “The computational complexity of linear optics,” *Theory Comput.* **9**, 143–252.
- Aaronson, S., and A. Arkhipov, 2014, “BosonSampling is far from uniform,” *Quantum Inf. Comput.* **14**, 1383–1423.
- Aaronson, S., A. Bouland, G. Kuperberg, and S. Mehraban, 2016, “The computational complexity of ball permutations,” [arXiv:1610.06646](https://arxiv.org/abs/1610.06646).
- Aaronson, S., and D. J. Brod, 2016, “BosonSampling with lost photons,” *Phys. Rev. A* **93**, 012335.
- Aaronson, S., and L. Chen, 2017, “Complexity-theoretic foundations of quantum supremacy experiments,” in *Proceeding of the 32nd Computational Complexity Conference (CCC 2017)*, Riga, 2017, Leibniz International Proceedings in Informatics (LIPIcs) Vol. 79, edited by Ryan O’Donnell (Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, Dagstuhl, Germany), pp. 22:1–22:67.
- Aaronson, S., and D. Gottesman, 2004, “Improved simulation of stabilizer circuits,” *Phys. Rev. A* **70**, 052328.
- Aaronson, S., and S. Gunn, 2019, “On the classical hardness of spoofing linear cross-entropy benchmarking,” [arXiv:1910.12085](https://arxiv.org/abs/1910.12085).
- Aaronson, S., and T. Hance, 2012, “Generalizing and derandomizing Gurvits’s approximation algorithm for the permanent,” [arXiv:1212.0025](https://arxiv.org/abs/1212.0025).
- Acharya, R., *et al.*, 2022, “Suppressing quantum errors by scaling a surface code logical qubit,” [arXiv:2207.06431](https://arxiv.org/abs/2207.06431).
- Acín, A., and L. Masanes, 2016, “Certified randomness in quantum physics,” *Nature (London)* **540**, 213–219.
- Aharonov, D., 2003, “A simple proof that Toffoli and Hadamard are quantum universal,” [arXiv:quant-ph/0301040](https://arxiv.org/abs/quant-ph/0301040).
- Aharonov, D., I. Arad, Z. Landau, and U. Vazirani, 2009, “The detectability lemma and quantum gap amplification,” in *Proceedings of the 41st Annual Symposium on Theory of Computing (STOC '09)*, Bethesda, MD, 2009 (Association for Computing Machinery, New York), pp. 417–426, [10.1145/1536414.1536472](https://doi.org/10.1145/1536414.1536472).
- Aharonov, D., and M. Ben-Or, 1996, “Polynomial simulations of decohered quantum computers,” in *Proceedings of the 37th Conference on Foundations of Computer Science, Burlington, VT, 1996* (IEEE, New York), pp. 46–55, [10.1109/SFCS.1996.548463](https://doi.org/10.1109/SFCS.1996.548463).
- Aharonov, D., and M. Ben-Or, 2008, “Fault-tolerant quantum computation with constant error rate,” *SIAM J. Comput.* **38**, 1207–1282.
- Aharonov, D., X. Gao, Z. Landau, Y. Liu, and U. Vazirani, 2022, “A polynomial-time classical algorithm for noisy random circuit sampling,” [arXiv:2211.03999](https://arxiv.org/abs/2211.03999).
- Anari, N., L. Gurvits, S. O. Gharan, and A. Saberi, 2017, “Simply exponential approximation of the permanent of positive semi-definite matrices,” [arXiv:1704.03486](https://arxiv.org/abs/1704.03486).
- Anshu, A., I. Arad, and T. Vidick, 2016, “Simple proof of the detectability lemma and spectral gap amplification,” *Phys. Rev. B* **93**, 205142.
- Aolita, L., C. Gogolin, M. Kliesch, and J. Eisert, 2015, “Reliable quantum certification of photonic state preparations,” *Nat. Commun.* **6**, 8498.
- Arkhipov, A., 2015, “Boson sampling is robust against small errors in the network matrix,” *Phys. Rev. A* **92**, 062326.
- Arkhipov, A., and G. Kuperberg, 2012, “The bosonic birthday paradox,” *Geom. Topol. Monogr.* **18**, 1–7.
- Arora, S., and B. Barak, 2009, *Computational Complexity: A Modern Approach* (Cambridge University Press, Cambridge, England).
- Arrazola, J. M., and T. R. Bromley, 2018, “Using Gaussian Boson Sampling to Find Dense Subgraphs,” *Phys. Rev. Lett.* **121**, 030503.
- Arrazola, J. M., T. R. Bromley, and P. Rebentrost, 2018, “Quantum approximate optimization with Gaussian boson sampling,” *Phys. Rev. A* **98**, 012322.
- Arute, F., *et al.*, 2019, “Quantum supremacy using a programmable superconducting processor,” *Nature (London)* **574**, 505–510.
- Aspect, A., J. Dalibard, and G. Roger, 1982, “Experimental Test of Bell’s Inequalities Using Time-Varying Analyzers,” *Phys. Rev. Lett.* **49**, 1804–1807.
- Aspect, A., P. Grangier, and G. Roger, 1982, “Experimental Realization of Einstein-Podolsky-Rosen-Bohm Gedanken Experiment: A New Violation of Bell’s Inequalities,” *Phys. Rev. Lett.* **49**, 91–94.
- Baez, M. L., M. Goihl, J. Haferkamp, J. Bermejo-Vega, M. Gluza, and J. Eisert, 2020, “Dynamical structure factors of dynamical

- quantum simulators,” *Proc. Natl. Acad. Sci. U.S.A.* **117**, 26123–26134.
- Baiardi, A., J. Bloino, and V. Barone, 2013, “General time dependent approach to vibronic spectroscopy including Franck-Condon, Herzberg-Teller, and Duschinsky effects,” *J. Chem. Theory Comput.* **9**, 4097–4115.
- Banchi, L., M. Fingerhuth, T. Babej, C. Ing, and J. M. Arrazola, 2020, “Molecular docking with Gaussian boson sampling,” *Sci. Adv.* **6**, eaax1950.
- Banchi, L., N. Quesada, and J. M. Arrazola, 2020, “Training Gaussian boson sampling distributions,” *Phys. Rev. A* **102**, 012417.
- Bao, Y., S. Choi, and E. Altman, 2020, “Theory of the phase transition in random unitary circuits with measurements,” *Phys. Rev. B* **101**, 104301.
- Barak, B., C.-N. Chou, and X. Gao, 2021, “Spoofing linear cross-entropy benchmarking in shallow quantum circuits,” in *Proceedings of the 12th Innovations in Theoretical Computer Science Conference (ITCS 2021)*, 2021, Leibniz International Proceedings in Informatics (LIPIcs) Vol. 185, edited by James R. Lee (Schloss Dagstuhl–Leibniz-Zentrum für Informatik, Dagstuhl, Germany), pp. 30:1–30:20.
- Barber, D., 2012, *Bayesian Reasoning and Machine Learning* (Cambridge University Press, Cambridge, England).
- Barends, R., *et al.*, 2014, “Superconducting quantum circuits at the surface code threshold for fault tolerance,” *Nature (London)* **508**, 500–503.
- Bartolucci, S., *et al.*, 2021, “Fusion-based quantum computation,” [arXiv:2101.09310](https://arxiv.org/abs/2101.09310).
- Barvinok, A., 1999, “Polynomial time algorithms to approximate permanents and mixed discriminants within a simply exponential factor,” *Random Struct. Algorithms* **14**, 29–61.
- Barvinok, A., 2016a, *Combinatorics and Complexity of Partition Functions, Algorithms and Combinatorics* (Springer International Publishing, Cham, Switzerland).
- Barvinok, A., 2016b, “Computing the permanent of (some) complex matrices,” *Found. Comput. Math.* **16**, 329–342.
- Barvinok, A., 2017, “Approximating permanents and Hafnians,” *Discrete Anal.* **2**, 34.
- Barvinok, A., 2019, “Computing permanents of complex diagonally dominant matrices and tensors,” *Isr. J. Math.* **232**, 931–945.
- Barvinok, A., 2020, “A remark on approximating permanents of positive definite matrices,” [arXiv:2005.06344](https://arxiv.org/abs/2005.06344).
- Barvinok, A. I., 1996, “Two algorithmic results for the traveling salesman problem,” *Math. Oper. Res.* **21**, 65–84.
- Bassirian, R., A. Bouland, B. Fefferman, S. Gunn, and A. Tal, 2021, “On certified randomness from quantum advantage experiments,” [arXiv:2111.14846](https://arxiv.org/abs/2111.14846).
- Beaver, D., and J. Feigenbaum, 1990, “Hiding instances in multi-oracle queries,” in *Proceedings of the 7th Annual Symposium on Theoretical Aspects of Computer Science (STACS '90)*, Rouen, France, Lecture Notes in Computer Science, edited by C. Choffrut and T. Lengauer (Springer, New York), pp. 37–48, [10.1007/3-540-52282-4_30](https://doi.org/10.1007/3-540-52282-4_30).
- Bell, J. S., 1964, “On the Einstein Podolsky Rosen paradox,” *Physics* **1**, 195–200.
- Benioff, P., 1980, “The computer as a physical system: A microscopic quantum mechanical Hamiltonian model of computers as represented by Turing machines,” *J. Stat. Phys.* **22**, 563–591.
- Bennink, R. S., 2021, “Efficient verification of anticoncentrated quantum states,” *npj Quantum Inf.* **7**, 127.
- Bennink, R. S., E. M. Ferragut, T. S. Humble, J. A. Laska, J. J. Nutaro, M. G. Pleszkoch, and R. C. Pooser, 2017, “Unbiased simulation of near-Clifford quantum circuits,” *Phys. Rev. A* **95**, 062337.
- Bentivegna, M., *et al.*, 2015, “Experimental scattershot boson sampling,” *Sci. Adv.* **1**, e1400255.
- Bermejo-Vega, J., D. Hangleiter, M. Schwarz, R. Raussendorf, and J. Eisert, 2018, “Architectures for Quantum Simulation Showing a Quantum Speedup,” *Phys. Rev. X* **8**, 021010.
- Bernien, H., *et al.*, 2017, “Probing many-body dynamics on a 51-atom quantum simulator,” *Nature (London)* **551**, 579–584.
- Bernstein, E., and U. Vazirani, 1993, “Quantum complexity theory,” in *Proceedings of the 25th Annual ACM Symposium on the Theory of Computing (STOC '93)*, San Diego, 1993 (Association for Computing Machinery, New York), pp. 11–20.
- Bernstein, E., and U. Vazirani, 1997, “Quantum complexity theory,” *SIAM J. Comput.* **26**, 1411–1473.
- Björklund, A., 2012, “Counting perfect matchings as fast as Ryser,” in *Proceedings of the 2012 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA '12)*, Kyoto, 2012 (Association for Computing Machinery, New York), pp. 914–921, [10.1137/1.9781611973099.73](https://doi.org/10.1137/1.9781611973099.73).
- Björklund, A., B. Gupt, and N. Quesada, 2019, “A faster Hafnian formula for complex matrices and its benchmarking on a super-computer,” *J. Exp. Algorithmics* **24**, 1.11:1.
- Blais, A., R.-S. Huang, A. Wallraff, S. M. Girvin, and R. J. Schoelkopf, 2004, “Cavity quantum electrodynamics for superconducting electrical circuits: An architecture for quantum computation,” *Phys. Rev. A* **69**, 062320.
- Blatt, R., and C. F. Roos, 2012, “Quantum simulations with trapped ions,” *Nat. Phys.* **8**, 277–284.
- Bloch, I., J. Dalibard, and W. Zwerger, 2008, “Many-body physics with ultracold gases,” *Rev. Mod. Phys.* **80**, 885–964.
- Blume-Kohout, R., J. K. Gamble, E. Nielsen, K. Rudinger, J. Mizrahi, K. Fortier, and P. Maunz, 2017, “Demonstration of qubit operations below a rigorous fault tolerance threshold with gate set tomography,” *Nat. Commun.* **8**, 14485.
- Blume-Kohout, R., J. King Gamble, E. Nielsen, J. Mizrahi, J. D. Sterk, and P. Maunz, 2013, “Robust, self-consistent, closed-form tomography of quantum logic gates on a trapped ion qubit,” [arXiv:1310.4492](https://arxiv.org/abs/1310.4492).
- Boixo, S., S. V. Isakov, V. N. Smelyanskiy, R. Babbush, N. Ding, Z. Jiang, M. J. Bremner, J. M. Martinis, and H. Neven, 2018, “Characterizing quantum supremacy in near-term devices,” *Nat. Phys.* **14**, 595–600.
- Boixo, S., S. V. Isakov, V. N. Smelyanskiy, and H. Neven, 2017, “Simulation of low-depth quantum circuits as complex undirected graphical models,” [arXiv:1712.05384](https://arxiv.org/abs/1712.05384).
- Boixo, S., V. N. Smelyanskiy, and H. Neven, 2017, “Fourier analysis of sampling from noisy chaotic quantum circuits,” [arXiv:1708.01875](https://arxiv.org/abs/1708.01875).
- Boone, K., A. Carignan-Dugas, J. J. Wallman, and J. Emerson, 2019, “Randomized benchmarking under different gatesets,” *Phys. Rev. A* **99**, 032329.
- Bouland, A., B. Fefferman, Z. Landau, and Y. Liu, 2022, “Noise and the frontier of quantum supremacy,” *Proceedings of the 62nd Annual IEEE Symposium on Foundations of Computer Science (FOCS 2021)*, 2021 (IEEE, New York), pp. 1308–1317, [10.1109/FOCS52979.2021.00127](https://doi.org/10.1109/FOCS52979.2021.00127).
- Bouland, A., B. Fefferman, C. Nirkhe, and U. Vazirani, 2019, “On the complexity and verification of quantum random circuit sampling,” *Nat. Phys.* **15**, 159–163.
- Bouland, A., J. F. Fitzsimons, and D. E. Koh, 2018, “Complexity classification of conjugated Clifford circuits,” in *Proceedings of the 33rd Computational Complexity Conference (CCC 2018)*,

- San Diego, 2018*, Leibniz International Proceedings in Informatics (LIPIcs) Vol. 102, edited by Rocco A. Servedio (Schloss Dagstuhl, Leibniz-Zentrum für Informatik, Dagstuhl, Germany), pp. 21:1–21:25.
- Bourennane, M., M. Eibl, C. Kurtsiefer, S. Gaertner, H. Weinfurter, O. Gühne, P. Hyllus, D. Bruß, M. Lewenstein, and A. Sanpera, 2004, “Experimental Detection of Multipartite Entanglement Using Witness Operators,” *Phys. Rev. Lett.* **92**, 087902.
- Brádler, K., P.-L. Dallaire-Demers, P. Reberntrost, D. Su, and C. Weedbrook, 2018, “Gaussian boson sampling for perfect matchings of arbitrary graphs,” *Phys. Rev. A* **98**, 032310.
- Brakerski, Z., P. Christiano, U. Mahadev, U. Vazirani, and T. Vidick, 2018, “A cryptographic test of quantumness and certifiable randomness from a single quantum device,” in *Proceedings of the 59th Annual IEEE Symposium on Foundations of Computer Science (FOCS), Paris, 2018* (IEEE, New York), pp. 320–331, [10.1109/FOCS.2018.00038](https://doi.org/10.1109/FOCS.2018.00038).
- Brakerski, Z., V. Koppula, U. Vazirani, and T. Vidick, 2020, “Simpler proofs of quantumness,” [arXiv:2005.04826](https://arxiv.org/abs/2005.04826).
- Brandão, F. G. S. L., A. W. Harrow, and M. Horodecki, 2016, “Local random quantum circuits are approximate polynomial-designs,” *Commun. Math. Phys.* **346**, 397–434.
- Brandão, F. G. S. L., and K. Svore, 2017, “Quantum speed-ups for semidefinite programming,” in *Proceeding of the 58th Annual IEEE Symposium on Foundations of Computer Science (FOCS), Berkeley, CA, 2017* (IEEE, New York), pp. 415–426, [10.1109/FOCS.2017.45](https://doi.org/10.1109/FOCS.2017.45).
- Braun, S., M. Friesdorf, S. S. Hodgman, M. Schreiber, J. P. Ronzheimer, A. Riera, M. del Rey, I. Bloch, J. Eisert, and U. Schneider, 2015, “Emergence of coherence and the dynamics of quantum phase transitions,” *Proc. Natl. Acad. Sci. U.S.A.* **112**, 3641.
- Bravyi, S., and D. Gosset, 2016, “Improved Classical Simulation of Quantum Circuits Dominated by Clifford Gates,” *Phys. Rev. Lett.* **116**, 250501.
- Bravyi, S., D. Gosset, and R. König, 2018, “Quantum advantage with shallow circuits,” *Science* **362**, 308.
- Bravyi, S., D. Gosset, R. König, and M. Tomamichel, 2020, “Quantum advantage with noisy shallow circuits,” *Nat. Phys.* **16**, 1040–1045.
- Bremner, M. J., R. Jozsa, and D. J. Shepherd, 2010, “Classical simulation of commuting quantum computations implies collapse of the polynomial hierarchy,” *Proc. R. Soc. A* **467**, 459–472.
- Bremner, M. J., A. Montanaro, and D. J. Shepherd, 2016, “Average-Case Complexity versus Approximate Simulation of Commuting Quantum Computations,” *Phys. Rev. Lett.* **117**, 080501.
- Bremner, M. J., A. Montanaro, and D. J. Shepherd, 2017, “Achieving quantum supremacy with sparse and noisy commuting quantum computations,” *Quantum* **1**, 8.
- Bridgeman, J. C., and C. T. Chubb, 2017, “Hand-waving and interpretive dance: An introductory course on tensor networks,” *J. Phys. A* **50**, 223001.
- Brieger, R., I. Roth, and M. Kliesch, 2023, “Compressive gate set tomography,” *PRX Quantum* **4**, 010325.
- Broadbent, A., J. Fitzsimons, and E. Kashefi, 2009, “Universal blind quantum computation,” in *Proceedings of the 50th Annual IEEE Symposium on Foundations of Computer Science (FOCS), Atlanta, 2009* (IEEE, New York), pp. 517–526, [10.1109/FOCS.2009.36](https://doi.org/10.1109/FOCS.2009.36).
- Brod, D. J., E. F. Galvão, A. Crespi, R. Osellame, N. Spagnolo, and F. Sciarrino, 2019, “Photonic implementation of boson sampling: A review,” *Adv. Photonics* **1**, 034001.
- Broome, M. A., A. Fedrizzi, S. Rahimi-Keshari, J. Dove, S. Aaronson, T. C. Ralph, and A. G. White, 2013, “Photonic boson sampling in a tunable circuit,” *Science* **339**, 794–798.
- Brouwer, P. W., and C. W. J. Beenakker, 1996, “Diagrammatic method of integration over the unitary group, with applications to quantum transport in mesoscopic systems,” *J. Math. Phys. (N.Y.)* **37**, 4904–4934.
- Bulmer, J. F. F., *et al.*, 2022, “The boundary for quantum advantage in Gaussian boson sampling,” *Sci. Adv.* **8**, eabl9236.
- Bultink, C. C., B. Tarasinski, N. Haandbæk, S. Poletto, N. Haider, D. J. Michalak, A. Bruno, and L. DiCarlo, 2018, “General method for extracting the quantum efficiency of dispersive qubit readout in circuit QED,” *Appl. Phys. Lett.* **112**, 092601.
- Cai, J.-Y., A. Pavan, and D. Sivakumar, 1999, “On the hardness of permanent,” in *Proceedings of the 16th Annual Symposium on Theoretical Aspects of Computer Science (STACS '99), Trier, Germany, 1999*, Lecture Notes in Computer Science, edited by C. Meinel and S. Tison (Springer, Berlin), pp. 90–99, [10.1007/3-540-49116-3_8](https://doi.org/10.1007/3-540-49116-3_8).
- Canonne, C. L., and K. Wimmer, 2020, “Testing data binnings,” [arXiv:2004.12893](https://arxiv.org/abs/2004.12893).
- Carolan, J., *et al.*, 2014, “On the experimental verification of quantum complexity in linear optics,” *Nat. Photonics* **8**, 621–626.
- Caves, C. M., C. A. Fuchs, and R. Schack, 2002, “Unknown quantum states: The quantum de Finetti representation,” *J. Math. Phys. (N.Y.)* **43**, 4537–4559.
- Cerfontaine, P., R. Otten, and H. Bluhm, 2020, “Self-Consistent Calibration of Quantum-Gate Sets,” *Phys. Rev. Appl.* **13**, 044071.
- Chabaud, U., T. Douce, F. Grosshans, E. Kashefi, and D. Markham, 2020, “Building trust for continuous variable quantum states,” [arXiv:1905.12700](https://arxiv.org/abs/1905.12700).
- Chabaud, U., T. Douce, D. Markham, P. van Loock, E. Kashefi, and G. Ferrini, 2017, “Continuous-variable sampling from photon-added or photon-subtracted squeezed states,” *Phys. Rev. A* **96**, 062307.
- Chabaud, U., G. Ferrini, F. Grosshans, and D. Markham, 2021, “Classical simulation of Gaussian quantum circuits with non-Gaussian input states,” *Phys. Rev. Res.* **3**, 033018.
- Chabaud, U., F. Grosshans, E. Kashefi, and D. Markham, 2021, “Efficient verification of boson sampling,” *Quantum* **5**, 578.
- Chabaud, U., D. Markham, and A. Sohbi, 2021, “Quantum machine learning with adaptive linear optics,” *Quantum* **5**, 496.
- Chabaud, U., and M. Walschaers, 2022, “Resources for bosonic quantum computational advantage,” [arXiv:2207.11781](https://arxiv.org/abs/2207.11781).
- Chakhmakhchyan, L., and N. J. Cerf, 2017, “Boson sampling with Gaussian measurements,” *Phys. Rev. A* **96**, 032326.
- Chen, J., F. Zhang, C. Huang, M. Newman, and Y. Shi, 2018, “Classical simulation of intermediate-size quantum circuits,” [arXiv:1805.01450](https://arxiv.org/abs/1805.01450).
- Chen, M.-C., R. Li, L. Gan, X. Zhu, G. Yang, C.-Y. Lu, and J.-W. Pan, 2020, “Quantum-Teleportation-Inspired Algorithm for Sampling Large Random Quantum Circuits,” *Phys. Rev. Lett.* **124**, 080502.
- Chen, Z.-Y., Q. Zhou, C. Xue, X. Yang, G.-C. Guo, and G.-P. Guo, 2018, “64-qubit quantum circuit simulation,” *Sci. Bull.* **63**, 964–971.
- Childs, A. M., Y. Su, M. C. Tran, N. Wiebe, and S. Zhu, 2021, “Theory of Trotter Error with Commutator Scaling,” *Phys. Rev. X* **11**, 011020.
- Childs, A. M., and N. Wiebe, 2012, “Hamiltonian simulation using linear combinations of unitary operations,” *Quantum Inf. Comput.* **12**, 901–924.

- Chin, S., and J. Huh, 2018, “Generalized concurrence in boson sampling,” *Sci. Rep.* **8**, 6101.
- Choi, J., *et al.*, 2023, “Preparing random states and benchmarking with many-body quantum chaos,” *Nature (London)* **613**, 468.
- Choi, J.-y., S. Hild, J. Zeiher, P. Schauf, A. Rubio-Abadal, T. Yefsah, V. Khemani, D. A. Huse, I. Bloch, and C. Gross, 2016, “Exploring the many-body localization transition in two dimensions,” *Science* **352**, 1547–1552.
- Chung, K.-M., Y. Lee, H.-H. Lin, and X. Wu, 2020, “Constant-round blind classical verification of quantum sampling,” *arXiv:2012.04848*.
- Clarke, J., and F. K. Wilhelm, 2008, “Superconducting quantum bits,” *Nature (London)* **453**, 1031–1042.
- Clements, W. R., J. J. Renema, A. Eckstein, A. A. Valido, A. Lita, T. Gerrits, S. W. Nam, W. S. Kolthammer, J. Huh, and I. A. Walmsley, 2018, “Approximating vibronic spectroscopy with imperfect quantum optics,” *J. Phys. B* **51**, 245503.
- Clifford, P., and R. Clifford, 2018, “The classical complexity of Boson sampling,” in *Proceedings of the 29th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA '18), New Orleans, 2018* (Society for Industrial and Applied Mathematics, Philadelphia), pp. 146–155, [10.1137/1.9781611975031.10](https://doi.org/10.1137/1.9781611975031.10).
- Clifford, P., and R. Clifford, 2020, “Faster classical boson sampling,” *arXiv:2005.04214*.
- Cramer, M., M. B. Plenio, S. T. Flammia, R. Somma, D. Gross, S. D. Bartlett, O. Landon-Cardinal, D. Poulin, and Y.-K. Liu, 2010, “Efficient quantum state tomography,” *Nat. Commun.* **1**, 149.
- Crespi, A., R. Osellame, R. Ramponi, D. J. Brod, E. F. Galvão, N. Spagnolo, C. Vitelli, E. Maiorino, P. Mataloni, and F. Sciarrino, 2013, “Integrated multimode interferometers with arbitrary designs for photonic boson sampling,” *Nat. Photonics* **7**, 545–549.
- Dalzell, A. M., A. W. Harrow, D. E. Koh, and R. L. La Placa, 2020, “How many qubits are needed for quantum computational supremacy?,” *Quantum* **4**, 264.
- Dalzell, A. M., N. Hunter-Jones, and F. G. S. L. Brandão, 2021, “Random quantum circuits transform local noise into global white noise,” *arXiv:2111.14907*.
- Dalzell, A. M., N. Hunter-Jones, and F. G. S. L. Brandão, 2022, “Random quantum circuits anticoncentrate in log depth,” *PRX Quantum* **3**, 010333.
- Dawson, C. M., H. L. Haselgrove, A. P. Hines, D. Mortimer, M. A. Nielsen, and T. J. Osborne, 2005, “Quantum computing and polynomial equations over the finite field \mathbb{Z}_2 ,” *Quantum Inf. Comput.* **5**, 102–112.
- Debnath, S., N. M. Linke, C. Figgatt, K. A. Landsman, K. Wright, and C. Monroe, 2016, “Demonstration of a small programmable quantum computer with atomic qubits,” *Nature (London)* **536**, 63–66.
- de Finetti, B., 1937, “La prévision: Ses lois logiques, ses sources subjectives,” *Ann. Inst. Henri Poincaré* **7**, 1–68, http://www.numdam.org/item/AIHP_1937__7_1_1_0.pdf.
- De las Cuevas, G, W. Dür, M. Van den Nest, and M. A. Martin-Delgado, 2011, “Quantum algorithms for classical lattice models,” *New J. Phys.* **13**, 093021.
- De Raedt, H., F. Jin, D. Willsch, M. Willsch, N. Yoshioka, N. Ito, S. Yuan, and K. Michielsen, 2019, “Massively parallel quantum computer simulator, eleven years later,” *Comput. Phys. Commun.* **237**, 47–61.
- De Raedt, K., K. Michielsen, H. De Raedt, B. Trieu, G. Arnold, M. Richter, Th. Lippert, H. Watanabe, and N. Ito, 2007, “Massively parallel quantum computer simulator,” *Comput. Phys. Commun.* **176**, 121–136.
- Deshpande, A., B. Fefferman, M. C. Tran, M. Foss-Feig, and A. V. Gorshkov, 2018, “Dynamical Phase Transitions in Sampling Complexity,” *Phys. Rev. Lett.* **121**, 030501.
- Deshpande, A., P. Niroula, O. Shtanko, A. V. Gorshkov, B. Fefferman, and M. J. Gullans, 2022, “Tight bounds on the convergence of noisy random circuits to the uniform distribution,” *PRX Quantum* **3**, 040329.
- Deshpande, A., *et al.*, 2022, “Quantum computational advantage via high-dimensional Gaussian boson sampling,” *Sci. Adv.* **8**, eabi7894.
- Deutsch, D., 1985, “Quantum theory, the Church-Turing principle and the universal quantum computer,” *Proc. R. Soc. A* **400**, 97–117.
- Drummond, P. D., B. Opanchuk, A. Delliios, and M. D. Reid, 2022, “Simulating complex networks in phase space: Gaussian boson sampling,” *Phys. Rev. A* **105**, 012427.
- Ebadi, S., *et al.*, 2021, “Quantum phases of matter on a 256-atom programmable quantum simulator,” *Nature (London)* **595**, 227–232.
- Egan, L., *et al.*, 2021, “Fault-tolerant control of an error-corrected qubit,” *Nature (London)* **598**, 281–286.
- Ehrenberg, A., A. Deshpande, C. L. Baldwin, D. A. Abanin, and A. V. Gorshkov, 2022, “Simulation complexity of many-body localized systems,” *arXiv:2205.12967*.
- Einstein, A., B. Podolsky, and N. Rosen, 1935, “Can quantum-mechanical description of physical reality be considered complete?,” *Phys. Rev.* **47**, 777–780.
- Eisert, J., D. Hangleiter, N. Walk, I. Roth, D. Markham, R. Parekh, U. Chabaud, and E. Kashefi, 2020, “Quantum certification and benchmarking,” *Nat. Rev. Phys.* **2**, 382–390.
- Eldar, L., and S. Mehraban, 2018, “Approximating the permanent of a random matrix with vanishing mean,” in *Proceedings of the 59th IEEE Annual Symposium on Foundations of Computer Science (FOCS 2018), Paris, 2018* (IEEE, New York), pp. 23–34, [10.1109/FOCS.2018.00012](https://doi.org/10.1109/FOCS.2018.00012).
- Erhard, A., J. J. Wallman, L. Postler, M. Meth, R. Stricker, E. A. Martinez, P. Schindler, T. Monz, J. Emerson, and R. Blatt, 2019, “Characterizing large-scale quantum computers via cycle benchmarking,” *Nat. Commun.* **10**, 5347.
- Farhi, E., J. Goldstone, and S. Gutmann, 2014, “A quantum approximate optimization algorithm,” *arXiv:1411.4028*.
- Farhi, E., J. Goldstone, S. Gutmann, and M. Sipser, 2000, “Quantum computation by adiabatic evolution,” *arXiv:quant-ph/0001106*.
- Fefferman, B., and C. Umans, 2015, “The power of quantum Fourier sampling,” *arXiv:1507.05592*.
- Fenner, S., F. Green, S. Homer, and R. Pruim, 1999, “Determining acceptance possibility for a quantum computation is hard for the polynomial hierarchy,” *Proc. R. Soc. A* **455**, 3953–3966.
- Fenner, S. A., L. J. Fortnow, and S. A. Kurtz, 1994, “Gap-definable counting classes,” *J. Comput. Syst. Sci.* **48**, 116–148.
- Ferris, A. J., and G. Vidal, 2012, “Perfect sampling with unitary tensor networks,” *Phys. Rev. B* **85**, 165146.
- Feynman, R. P., 1982, “Simulating physics with computers,” *Int. J. Theor. Phys.* **21**, 467–488.
- Feynman, R. P., 1985, “Quantum mechanical computers,” *Opt. News* **11**, 11–20.
- Fitzsimons, J. F., M. Hajdusek, and T. Morimae, 2018, “Post Hoc Verification of Quantum Computation,” *Phys. Rev. Lett.* **120**, 040501.
- Fitzsimons, J. F., and E. Kashefi, 2017, “Unconditionally verifiable blind quantum computation,” *Phys. Rev. A* **96**, 012303.
- Flammia, S. T., and Y.-K. Liu, 2011, “Direct Fidelity Estimation from Few Pauli Measurements,” *Phys. Rev. Lett.* **106**, 230501.
- Foxen, B., *et al.*, 2020, “Demonstrating a Continuous Set of Two-Qubit Gates for Near-Term Quantum Algorithms,” *Phys. Rev. Lett.* **125**, 120504.

- Fredkin, E., and T. Toffoli, 1982, “Conservative logic,” *Int. J. Theor. Phys.* **21**, 219–253.
- Freedman, S. J., and J. F. Clauser, 1972, “Experimental Test of Local Hidden-Variable Theories,” *Phys. Rev. Lett.* **28**, 938–941.
- Friis, N., *et al.*, 2018, “Observation of Entangled States of a Fully Controlled 20-Qubit System,” *Phys. Rev. X* **8**, 021012.
- Fujii, K., 2016, “Noise threshold of quantum supremacy,” *arXiv:1610.03632*.
- Fujii, K., H. Kobayashi, T. Morimae, H. Nishimura, S. Tamate, and S. Tani, 2018, “Impossibility of Classically Simulating One-Clean-Qubit Model with Multiplicative Error,” *Phys. Rev. Lett.* **120**, 200502.
- Fujii, K., and T. Morimae, 2017, “Commuting quantum circuits and complexity of Ising partition functions,” *New J. Phys.* **19**, 033003.
- Gambetta, J., A. Blais, D. I. Schuster, A. Wallraff, L. Frunzio, J. Majer, M. H. Devoret, S. M. Girvin, and R. J. Schoelkopf, 2006, “Qubit-photon interactions in a cavity: Measurement-induced dephasing and number splitting,” *Phys. Rev. A* **74**, 042318.
- Gao, X., and L. Duan, 2018, “Efficient classical simulation of noisy quantum computation,” *arXiv:1810.03176*.
- Gao, X., M. Kalinowski, C.-N. Chou, M. D. Lukin, B. Barak, and S. Choi, 2021, “Limitations of linear cross-entropy as a measure for quantum advantage,” *arXiv:2112.01657*.
- Gao, X., S.-T. Wang, and L.-M. Duan, 2017, “Quantum Supremacy for Simulating a Translation-Invariant Ising Spin Model,” *Phys. Rev. Lett.* **118**, 040502.
- García-Patrón, R., J. J. Renema, and V. S. Shchesnovich, 2019, “Simulating boson sampling in lossy architectures,” *Quantum* **3**, 169.
- Gemmel, P., R. Lipton, R. Rubinfeld, M. Sudan, and A. Wigderson, 1991, “Self-testing/correcting for polynomials and for approximate functions,” in *Proceedings of the 23rd Annual ACM Symposium on Theory in Computing (STOC '91), New Orleans, 1991* (Association for Computing Machinery, New York), pp. 33–42, [10.1145/103418.103429](https://doi.org/10.1145/103418.103429).
- Gemmel, P., and M. Sudan, 1992, “Highly resilient correctors for polynomials,” *Inf. Process. Lett.* **43**, 169–174.
- Gheorghiu, V., and M. Mosca, 2019, “Benchmarking the quantum cryptanalysis of symmetric, public-key and hash-based cryptographic schemes,” *arXiv:1902.02332*.
- Gidney, C., and M. Eker, 2019, “How to factor 2048 bit RSA integers in 8 hours using 20 million noisy qubits,” *arXiv:1905.09749*.
- Gilyén, A., Y. Su, G. H. Low, and N. Wiebe, 2019, “Quantum singular value transformation and beyond: Exponential improvements for quantum matrix arithmetics,” in *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing (STOC '19), Phoenix, 2019* (Association for Computing Machinery, New York), pp. 193–204, [10.1145/3313276.3316366](https://doi.org/10.1145/3313276.3316366).
- Giordani, T., *et al.*, 2018, “Experimental statistical signature of many-body quantum interference,” *Nat. Photonics* **12**, 173–178.
- Giustina, M., *et al.*, 2015, “Significant-Loophole-Free Test of Bell’s Theorem with Entangled Photons,” *Phys. Rev. Lett.* **115**, 250401.
- Glasser, I., R. Sweke, N. Pancotti, J. Eisert, and J. I. Cirac, 2019, “Expressive power of tensor-network factorizations for probabilistic modelling, with applications from hidden Markov models to quantum machine learning,” *arXiv:1907.03741*.
- Gluz, M., M. Kliesch, J. Eisert, and L. Aolita, 2018, “Fidelity Witnesses for Fermionic Quantum Simulations,” *Phys. Rev. Lett.* **120**, 190501.
- Glynn, D. G., 2010, “The permanent of a square matrix,” *Eur. J. Combinatorics* **31**, 1887–1891.
- Gogolin, C., M. Kliesch, L. Aolita, and J. Eisert, 2013, “Boson-sampling in the light of sample complexity,” *arXiv:1306.3995*.
- Goldberg, L. A., and H. Guo, 2017, “The complexity of approximating complex-valued Ising and Tutte partition functions,” *Comput. Complex.* **26**, 765–833.
- Goldreich, O., 2017, *Introduction to Property Testing* (Cambridge University Press, Cambridge, England).
- Gottesman, D., 1997, “Stabilizer codes and quantum error correction,” Ph.D. thesis (California Institute of Technology).
- Gray, J., and S. Kourtis, 2021, “Hyper-optimized tensor network contraction,” *Quantum* **5**, 410.
- Greenbaum, D., 2015, “Introduction to quantum gate set tomography,” *arXiv:1509.02921*.
- Grier, D., D. J. Brod, J. M. Arrazola, M. B. A. Alonso, and N. Quesada, 2022, “The complexity of bipartite Gaussian boson sampling,” *Quantum* **6**, 863.
- Grier, D., and L. Schaeffer, 2018, “New hardness results for the permanent using linear optics,” *arXiv:1610.04670*.
- Grover, L. K., 1996, “A fast quantum mechanical algorithm for database search,” in *Proceedings of the 28th Annual ACM Symposium on Theory of Computing (STOC '96), Philadelphia, 1996* (Association for Computing Machinery, New York), pp. 212–219, [10.1145/237814.237866](https://doi.org/10.1145/237814.237866).
- Guanzon, J. J., A. P. Lund, and T. C. Ralph, 2021, “Multimode metrology via scattershot sampling,” *Phys. Rev. A* **104**, 032607.
- Gühne, O., and G. Tóth, 2009, “Entanglement detection,” *Phys. Rep.* **474**, 1.
- Guo, C., Y. Zhao, and H.-L. Huang, 2021, “Verifying Random Quantum Circuits with Arbitrary Geometry Using Tensor Network States Algorithm,” *Phys. Rev. Lett.* **126**, 070502.
- Guo, C., *et al.*, 2019, “General-Purpose Quantum Circuit Simulator with Projected Entangled-Pair States and the Quantum Supremacy Frontier,” *Phys. Rev. Lett.* **123**, 190501.
- Gupt, B., J. M. Arrazola, N. Quesada, and T. R. Bromley, 2020, “Classical benchmarking of Gaussian boson sampling on the Titan supercomputer,” *Quantum Inf. Process.* **19**, 249.
- Gupt, B., J. Izaac, and N. Quesada, 2019, “The Walrus: A library for the calculation of Hafnians, Hermite polynomials and Gaussian boson sampling,” *J. Open Source Software* **4**, 1705.
- Guruswami, V., 2006, “List decoding in average-case complexity and pseudorandomness,” in *Proceedings of the IEEE Information Theory Workshop (ITW '06), Chengdu, China, 2006* (IEEE, New York), pp. 32–36, [10.1109/ITW.2006.1633776](https://doi.org/10.1109/ITW.2006.1633776).
- Gurvits, L., 2003, “Classical deterministic complexity of Edmonds’ problem and quantum entanglement,” *arXiv:quant-ph/0303055*.
- Gurvits, L., and A. Samorodnitsky, 2002, “A deterministic algorithm for approximating the mixed discriminant and mixed volume, and a combinatorial corollary,” *Discrete Comput. Geom.* **27**, 531–550.
- Haake, F., 2010, *Quantum Signatures of Chaos*, Springer Series in Synergetics Vol. 54 (Springer, Berlin).
- Haferkamp, J., 2022, “Random quantum circuits are approximate unitary t -designs in depth $o(nt^{5+o(1)})$,” *Quantum* **6**, 795.
- Haferkamp, J., D. Hangleiter, A. Bouland, B. Fefferman, J. Eisert, and J. Bermejo-Vega, 2020, “Closing Gaps of a Quantum Advantage with Short-Time Hamiltonian Dynamics,” *Phys. Rev. Lett.* **125**, 250501.
- Haferkamp, J., D. Hangleiter, J. Eisert, and M. Gluz, 2020, “Contracting projected entangled pair states is average-case hard,” *Phys. Rev. Res.* **2**, 013010.
- Hamilton, C. S., R. Kruse, L. Sansoni, S. Barkhofen, C. Silberhorn, and I. Jex, 2017, “Gaussian Boson Sampling,” *Phys. Rev. Lett.* **119**, 170501.

- Häner, T., M. Roetteler, and K. M. Svore, 2017, “Factoring using $2n + 2$ qubits with Toffoli based modular multiplication,” *Quantum Inf. Comput.* **17**, 673–684.
- Häner, T., and D. S. Steiger, 2017, “0.5 petabyte simulation of a 45-qubit quantum circuit,” in *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis (SC '17), Denver, 2017* (Association for Computing Machinery, New York), pp. 1–10, [10.1145/3126908.3126947](https://doi.org/10.1145/3126908.3126947).
- Hangleiter, D., 2021, “Sampling and the complexity of nature,” Ph.D. thesis (Freie Universität Berlin).
- Hangleiter, D., J. Bermejo-Vega, M. Schwarz, and J. Eisert, 2018, “Anticoncentration theorems for schemes showing a quantum speedup,” *Quantum* **2**, 65.
- Hangleiter, D., M. Kliesch, J. Eisert, and C. Gogolin, 2019, “Sample Complexity of Device-Independently Certified Quantum Supremacy,” *Phys. Rev. Lett.* **122**, 210502.
- Hangleiter, D., M. Kliesch, M. Schwarz, and J. Eisert, 2017, “Direct certification of a class of quantum simulations,” *Quantum Sci. Technol.* **2**, 015004.
- Harrow, A. W., A. Hassidim, and S. Lloyd, 2009, “Quantum Algorithm for Linear Systems of Equations,” *Phys. Rev. Lett.* **103**, 150502.
- Harrow, A. W., and R. A. Low, 2009, “Random quantum circuits are approximate 2-designs,” *Commun. Math. Phys.* **291**, 257–302.
- Harrow, A. W., and S. Mehraban, 2023, “Approximate unitary t -designs by short random quantum circuits using nearest-neighbor and long-range gates,” *Commun. Math. Phys.* **401**, 1531–1626.
- Harrow, A. W., and A. Montanaro, 2017, “Quantum computational supremacy,” *Nature (London)* **549**, 203–209.
- Havlíček, Vojtech, Antonio D. Córcoles, Kristan Temme, Aram W. Harrow, Abhinav Kandala, Jerry M. Chow, and Jay M. Gambetta, 2019, “Supervised learning with quantum-enhanced feature spaces,” *Nature (London)* **567**, 209–212.
- Hayashi, M., and Y. Takeuchi, 2019, “Verifying commuting quantum computations via fidelity estimation of weighted graph states,” *New J. Phys.* **21**, 093060.
- Hebenstreit, M., R. Jozsa, B. Kraus, S. Strelchuk, and M. Yoganathan, 2019, “All Pure Fermionic Non-Gaussian States Are Magic States for Matchgate Computations,” *Phys. Rev. Lett.* **123**, 080503.
- Helsen, J., M. Ioannou, I. Roth, J. Kitzinger, E. Onorati, A. H. Werner, and J. Eisert, 2021, “Estimating gate-set properties from random sequences,” [arXiv:2110.13178](https://arxiv.org/abs/2110.13178).
- Helsen, J., I. Roth, E. Onorati, A. H. Werner, and J. Eisert, 2022, “A general framework for randomized benchmarking,” *PRX Quantum* **3**, 020357.
- Helsen, J., X. Xue, L. M. K. Vandersypen, and S. Wehner, 2019, “A new class of efficient randomized benchmarking protocols,” *npj Quantum Inf.* **5**, 1–9.
- Hensen, B., *et al.*, 2015, “Loophole-free Bell inequality violation using electron spins separated by 1.3 kilometres,” *Nature (London)* **526**, 682–686.
- Hirahara, S., and F. Le Gall, 2021, “Test of quantumness with small-depth quantum circuits,” *Proceedings of the 46th International Symposium on Mathematical Foundations of Computer Science (MFCS 2021), Tallinn, Estonia, 2021*, Leibniz International Proceedings in Informatics (LIPIcs) Vol. 202, pp. 59:1–59:15, [10.4230/LIPIcs.MFCS.2021.59](https://doi.org/10.4230/LIPIcs.MFCS.2021.59).
- Hong, C. K., Z. Y. Ou, and L. Mandel, 1987, “Measurement of Subpicosecond Time Intervals between Two Photons by Interference,” *Phys. Rev. Lett.* **59**, 2044–2046.
- Huang, C., M. Newman, and M. Szegedy, 2020, “Explicit lower bounds on strong quantum simulation,” *IEEE Trans. Inf. Theory* **66**, 5585–5600.
- Huang, C., *et al.*, 2020, “Classical simulation of quantum supremacy circuits,” [arXiv:2005.06787](https://arxiv.org/abs/2005.06787).
- Huang, C., *et al.*, 2021, “Efficient parallelization of tensor network contraction for simulating quantum computation,” *Nat. Comput. Sci.* **1**, 578–587.
- Huang, H.-Y., R. Kueng, and J. Preskill, 2020, “Predicting many properties of a quantum system from very few measurements,” *Nat. Phys.* **16**, 1050–1057.
- Huang, Z., P. Kok, and C. Lupo, 2021, “Fault-tolerant quantum data locking,” *Phys. Rev. A* **103**, 052611.
- Huang, Z., P. P. Rohde, D. W. Berry, P. Kok, J. P. Dowling, and C. Lupo, 2021, “Photonic quantum data locking,” *Quantum* **5**, 447.
- Hudson, R. L., and G. R. Moody, 1976, “Locally normal symmetric states and an analogue of de Finetti’s theorem,” *Z. Wahrsch. Verw. Geb.* **33**, 343–351.
- Huh, J., 2022, “A fast quantum algorithm for computing matrix permanent,” [arXiv:2205.01328v2](https://arxiv.org/abs/2205.01328v2).
- Huh, J., G. G. Guerreschi, B. Peropadre, J. R. McClean, and A. Aspuru-Guzik, 2015, “Boson sampling for molecular vibronic spectra,” *Nat. Photonics* **9**, 615–620.
- Hunter-Jones, N., 2019, “Unitary designs from statistical mechanics in random quantum circuits,” [arXiv:1905.12053](https://arxiv.org/abs/1905.12053).
- Jahangiri, S., J. M. Arrazola, N. Quesada, and N. Killoran, 2020, “Point processes with Gaussian boson sampling,” *Phys. Rev. E* **101**, 022134.
- Jaksch, D., C. Bruder, J. I. Cirac, C. W. Gardiner, and P. Zoller, 1998, “Cold Bosonic Atoms in Optical Lattices,” *Phys. Rev. Lett.* **81**, 3108.
- Jerrum, M. R., and A. Sinclair, 1993, “Polynomial-time approximation algorithms for the Ising model,” *SIAM J. Comput.* **22**, 1087–1116.
- Jerrum, M. R., A. Sinclair, and E. Vigoda, 2004, “A polynomial-time approximation algorithm for the permanent of a matrix with non-negative entries,” *J. Assoc. Comput. Mach.* **51**, 671–697.
- Jerrum, M. R., L. G. Valiant, and V. V. Vazirani, 1986, “Random generation of combinatorial structures from a uniform distribution,” *Theor. Comput. Sci.* **43**, 169–188.
- Jian, C.-M., Y.-Z. You, R. Vasseur, and A. W. W. Ludwig, 2020, “Measurement-induced criticality in random quantum circuits,” *Phys. Rev. B* **101**, 104302.
- Jiang, T., 2006, “How many entries of a typical orthogonal matrix can be approximated by independent normals?,” *Ann. Probab.* **34**, 1497–1529.
- Jiang, T., 2009, “The entries of circular orthogonal ensembles,” *J. Math. Phys. (N.Y.)* **50**, 063302.
- Jnane, H., N. P. D. Sawaya, B. Peropadre, A. Aspuru-Guzik, R. Garcia-Patron, and J. Huh, 2021, “Analog quantum simulation of non-Condon effects in molecular spectroscopy,” *ACS Photonics* **8**, 2007–2016.
- Jurcevic, P., *et al.*, 2021, “Demonstration of quantum volume 64 on a superconducting quantum computing system,” *Quantum Sci. Technol.* **6**, 025020.
- Kahanamoku-Meyer, G. D., 2019, “Forging quantum data: Classically defeating an IQP-based quantum test,” [arXiv:1912.05547](https://arxiv.org/abs/1912.05547).
- Kahanamoku-Meyer, G. D., S. Choi, U. V. Vazirani, and N. Y. Yao, 2022, “Classically verifiable quantum advantage from a computational Bell test,” *Nat. Phys.* **18**, 918–924.
- Kalachev, G., P. Pantelev, and M.-H. Yung, 2021, “Recursive multi-tensor contraction for XEB verification of quantum circuits,” [arXiv:2108.05665](https://arxiv.org/abs/2108.05665).

- Kalachev, G., P. Panteleev, P.-F. Zhou, and M.-H. Yung, 2021, “Classical sampling of random quantum circuits with bounded fidelity,” *arXiv:2112.15083*.
- Kalai, G., and G. Kindler, 2014, “Gaussian noise sensitivity and BosonSampling,” *arXiv:1409.3093*.
- Kalev, A., A. Kyrillidis, and N. M. Linke, 2019, “Validating and certifying stabilizer states,” *Phys. Rev. A* **99**, 042337.
- Kane, D., S. Karmalkar, and E. Price, 2017, “Robust polynomial regression up to the information theoretic limit,” in *Proceedings of the 58th IEEE Annual Symposium on Foundations of Computer Science (FOCS), Berkeley, 2017* (IEEE New York), pp. 391–402, 10.1109/FOCS.2017.43.
- Kapourniotis, T., and A. Datta, 2019, “Nonadaptive fault-tolerant verification of quantum supremacy with noise,” *Quantum* **3**, 164.
- Kiesel, N., C. Schmid, U. Weber, G. Tóth, O. Gühne, R. Ursin, and H. Weinfurter, 2005, “Experimental Analysis of a Four-Qubit Photon Cluster State,” *Phys. Rev. Lett.* **95**, 210502.
- Kliesch, M., and I. Roth, 2021, “Theory of quantum system certification,” *PRX Quantum* **2**, 010201.
- Kok, P., and B. W. Lovett, 2010, *Introduction to Optical Quantum Information Processing* (Cambridge University Press, Cambridge, England).
- Kok, P., W. J. Munro, K. Nemoto, T. C. Ralph, J. P. Dowling, and G. J. Milburn, 2007, “Linear optical quantum computing with photonic qubits,” *Rev. Mod. Phys.* **79**, 135–174.
- Kondo, Y., R. Mori, and R. Movassagh, 2022, “Quantum supremacy and hardness of estimating output probabilities of quantum circuits,” in *Proceedings of the 62nd IEEE Annual Symposium on Foundations of Computer Science (FOCS), Denver, 2021* (IEEE, New York), pp. 1296–1307, 10.1109/FOCS52979.2021.00126.
- König, R., and R. Renner, 2005, “A de Finetti representation for finite symmetric quantum states,” *J. Math. Phys. (N.Y.)* **46**, 122108.
- Krinner, S., *et al.*, 2022, “Realizing repeated quantum error correction in a distance-three surface code,” *Nature (London)* **605**, 669–674.
- Krovi, H., 2022, “Average-case hardness of estimating probabilities of random quantum circuits with a linear scaling in the error exponent,” *arXiv:2206.05642*.
- Kruse, R., C. S. Hamilton, L. Sansoni, S. Barkhofen, C. Silberhorn, and I. Jex, 2019, “A detailed study of Gaussian boson sampling,” *Phys. Rev. A* **100**, 032326.
- Kuperberg, G., 2015, “How hard is it to approximate the Jones polynomial?,” *Theory Comput.* **11**, 183–219.
- Kushilevitz, E., and Y. Mansour, 1993, “Learning decision trees using the Fourier spectrum,” *SIAM J. Comput.* **22**, 1331–1348.
- Lautemann, C., 1983, “BPP and the polynomial hierarchy,” *Inf. Process. Lett.* **17**, 215–217.
- Ledoux, M., 2005, *The Concentration of Measure Phenomenon*, Mathematical Surveys and Monographs Vol. 89 (American Mathematical Society, Providence).
- Lee, V. E., N. Ruan, R. Jin, and C. Aggarwal, 2010, “A survey of algorithms for dense subgraph discovery,” in *Managing and Mining Graph Data*, edited by C. C. Aggarwal and H. Wang (Springer, Berlin), pp. 303–336, 10.1007/978-1-4419-6045-0_10.
- Leone, L., S. F. E. Oliviero, and A. Hamma, 2023, “Nonstabilizer-ness determining the hardness of direct fidelity estimation,” *Phys. Rev. A* **107**, 022429.
- Leverrier, A., and R. García-Patrón, 2015, “Analysis of circuit imperfections in BosonSampling,” *Quantum Inf. Comput.* **15**, 0489–0512.
- Levin, L. A., 1986, “Average case complete problems,” *SIAM J. Comput.* **15**, 285–286.
- Li, M., and P. M. B. Vitányi, 1992, “Average case complexity under the universal distribution equals worst-case complexity,” *Inf. Process. Lett.* **42**, 145–149.
- Li, R., B. Wu, M. Ying, X. Sun, and G. Yang, 2018, “Quantum supremacy circuit simulation on Sunway TaihuLight,” *arXiv:1804.04797*.
- Linial, N., A. Samorodnitsky, and A. Wigderson, 1998, “A deterministic strongly polynomial algorithm for matrix scaling and approximate permanents,” in *Proceedings of the 30th Annual ACM Symposium on Theory of Computing (STOC '98), Dallas, 1998* (Association for Computing Machinery, New York), pp. 644–652, 10.1145/276698.276880.
- Lipton, R., 1991, “New directions in testing,” in *Distributed Computing and Cryptography*, DIMACS Series in Discrete Mathematics and Theoretical Computer Science Vol. 2, edited by J. Feigenbaum and M. Merritt (Association for Computing Machinery, New York), pp. 191–202.
- Liu, J.-P., H. Ø. Kolden, H. K. Krovi, N. F. Loureiro, K. Trivisa, and A. M. Childs, 2021, “Efficient quantum algorithm for dissipative nonlinear differential equations,” *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2026805118.
- Liu, X., *et al.*, 2021, “Redefining the quantum supremacy baseline with a new generation Sunway supercomputer,” *arXiv:2111.01066*.
- Liu, Y., M. Otten, R. Bassirianjahromi, L. Jiang, and B. Fefferman, 2021, “Benchmarking near-term quantum computers via random circuit sampling,” *arXiv:2105.05232*.
- Liu, Y. A., *et al.*, 2021, “Closing the ‘quantum supremacy’ gap: Achieving real-time simulation of a random quantum circuit using a new Sunway supercomputer,” in *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis (SC '21), St. Louis, 2021* (Association for Computing Machinery, New York), pp. 1–12, 10.1145/3458817.3487399.
- Liu, Z., and A. Gheorghiu, 2022, “Depth-efficient proofs of quantumness,” *Quantum* **6**, 807.
- Lloyd, S., 1996, “Universal quantum simulators,” *Science* **273**, 1073–1078.
- Lokshantov, D., R. Paturi, S. Tamaki, R. Williams, and H. Yu, 2017, “Beating brute force for systems of polynomial equations over finite fields,” in *Proceedings of the 28th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA), Barcelona, 2017*, edited by P. N. Klein (Society for Industrial and Applied Mathematics, Philadelphia), pp. 2190–2202, 10.1137/1.9781611974782.143.
- Loredo, J. C., M. A. Broome, P. Hilaire, O. Gazzano, I. Sagnes, A. Lemaitre, M. P. Almeida, P. Senellart, and A. G. White, 2017, “Boson Sampling with Single-Photon Fock States from a Bright Solid-State Source,” *Phys. Rev. Lett.* **118**, 130503.
- Low, G. H., and I. L. Chuang, 2017, “Optimal Hamiltonian Simulation by Quantum Signal Processing,” *Phys. Rev. Lett.* **118**, 010501.
- Low, G. H., and I. L. Chuang, 2019, “Hamiltonian simulation by qubitization,” *Quantum* **3**, 163.
- Lund, A. P., M. J. Bremner, and T. C. Ralph, 2017, “Quantum sampling problems, BosonSampling and quantum supremacy,” *npj Quantum Inf.* **3**, 15.
- Lund, A. P., A. Laing, S. Rahimi-Keshari, T. Rudolph, J. L. O’Brien, and T. C. Ralph, 2014, “Boson Sampling from a Gaussian State,” *Phys. Rev. Lett.* **113**, 100502.
- Lund, A. P., S. Rahimi-Keshari, and T. C. Ralph, 2017, “Exact boson sampling using Gaussian continuous variable measurements,” *Phys. Rev. A* **96**, 022301.
- Lundow, P. H., and K. Markström, 2022, “Efficient computation of permanents, with applications to boson sampling and random matrices,” *J. Comput. Phys.* **455**, 110990.

- Madsen, L. S., *et al.*, 2022, “Quantum computational advantage with a programmable photonic processor,” *Nature (London)* **606**, 75–81.
- Mahadev, U., 2018, “Classical verification of quantum computations,” [arXiv:1804.01082](https://arxiv.org/abs/1804.01082).
- Mandel, O., M. Greiner, A. Widera, T. Rom, T. W. Hänsch, and I. Bloch, 2003, “Controlled collisions for multi-particle entanglement of optically trapped atoms,” *Nature (London)* **425**, 937–940.
- Mann, R. L., and M. J. Bremner, 2017, “On the complexity of random quantum computations and the Jones polynomial,” [arXiv:1711.00686](https://arxiv.org/abs/1711.00686).
- Mantri, A., T. F. Demarie, and J. F. Fitzsimons, 2017, “Universality of quantum computation with cluster states and (X, Y) -plane measurements,” *Sci. Rep.* **7**, 1–7.
- Markham, D., and A. Krause, 2020, “A simple protocol for certifying graph states and applications in quantum networks,” *Cryptography* **4**, 3.
- Markov, I. L., A. Fatima, S. V. Isakov, and S. Boixo, 2018, “Quantum supremacy is both closer and farther than it appears,” [arXiv:1807.10749](https://arxiv.org/abs/1807.10749).
- Markov, I. L., and Y. Shi, 2008, “Simulating quantum computation by contracting tensor networks,” *SIAM J. Comput.* **38**, 963–981.
- Martínez-Cifuentes, J., K. M. Fonseca-Romero, and N. Quesada, 2022, “Classical models are a better explanation of the Jiuzhang Gaussian boson samplers than their targeted squeezed light models,” [arXiv:2207.10058](https://arxiv.org/abs/2207.10058).
- Martyn, J. M., Z. M. Rossi, A. K. Tan, and I. L. Chuang, 2021, “Grand unification of quantum algorithms,” *PRX Quantum* **2**, 040203.
- Maskara, N., A. Deshpande, M. C. Tran, A. Ehrenberg, B. Fefferman, and A. V. Gorshkov, 2022, “Complexity Phase Diagram for Interacting and Long-Range Bosonic Hamiltonians,” *Phys. Rev. Lett.* **129**, 150604.
- McCaskey, A., E. Dumitrescu, M. Chen, D. Lyakh, and T. Humble, 2018, “Validating quantum-classical programming models with tensor network simulations,” *PLoS One* **13**, e0206704.
- McClean, J. R., J. Romero, R. Babbush, and A. Aspuru-Guzik, 2016, “The theory of variational hybrid quantum-classical algorithms,” *New J. Phys.* **18**, 023023.
- Merkel, S. T., J. M. Gambetta, J. A. Smolin, S. Poletto, A. D. Córcoles, B. R. Johnson, C. A. Ryan, and M. Steffen, 2013, “Self-consistent quantum process tomography,” *Phys. Rev. A* **87**, 062119.
- Mezher, R., J. Ghalbouni, J. Dgheim, and D. Markham, 2020, “Fault-tolerant quantum speedup from constant depth quantum circuits,” *Phys. Rev. Res.* **2**, 033444.
- Miller, J., S. Sanders, and A. Miyake, 2017, “Quantum supremacy in constant-time measurement-based computation: A unified architecture for sampling and verification,” *Phys. Rev. A* **96**, 062320.
- Montanaro, A., 2016, “Quantum algorithms: An overview,” *npj Quantum Inf.* **2**, 15023.
- Montanaro, A., 2017, “Quantum circuits and low-degree polynomials over \mathbb{F}_2 ,” *J. Phys. A* **50**, 084002.
- Morimae, T., 2017, “Hardness of classically sampling the one-clean-qubit model with constant total variation distance error,” *Phys. Rev. A* **96**, 040302.
- Morimae, T., K. Fujii, and J. F. Fitzsimons, 2014, “Hardness of Classically Simulating the One-Clean-Qubit Model,” *Phys. Rev. Lett.* **112**, 130502.
- Morimae, T., Y. Takeuchi, and M. Hayashi, 2017, “Verification of hypergraph states,” *Phys. Rev. A* **96**, 062321.
- Morimae, T., and S. Tamaki, 2019, “Fine-grained quantum computational supremacy,” *Quantum Inf. Comput.* **19**, 1089–1115.
- Movassagh, R., 2018, “Efficient unitary paths and quantum computational supremacy: A proof of average-case hardness of random circuit sampling,” [arXiv:1810.04681](https://arxiv.org/abs/1810.04681).
- Movassagh, R., 2020, “Quantum supremacy and random circuits,” [arXiv:1909.06210](https://arxiv.org/abs/1909.06210).
- Moylett, A. E., R. García-Patrón, J. J. Renema, and P. S. Turner, 2019, “Classically simulating near-term partially-distinguishable and lossy boson sampling,” *Quantum Sci. Technol.* **5**, 015001.
- Muller, D. E., 1954, “Application of Boolean algebra to switching circuit design and to error detection,” *Trans. IRE Prof. Group Ind. Electron. EC-3*, 6–12.
- Muraleedharan, G., A. Miyake, and I. H. Deutsch, 2019, “Quantum computational supremacy in the sampling of bosonic random walkers on a one-dimensional lattice,” *New J. Phys.* **21**, 055003.
- Murphy, K. P., 2012, *Machine Learning: A Probabilistic Perspective*, Adaptive Computation and Machine Learning Series (MIT Press, Cambridge, MA).
- Nachtergaele, B., 1996, “The spectral gap for some spin chains with discrete symmetry breaking,” *Commun. Math. Phys.* **175**, 565–606.
- Nagaj, D., 2012, “Universal two-body-Hamiltonian quantum computing,” *Phys. Rev. A* **85**, 032330.
- Nagaj, D., and P. Wocjan, 2008, “Hamiltonian quantum cellular automata in one dimension,” *Phys. Rev. A* **78**, 032311.
- Nakata, Y., M. Koashi, and M. Murao, 2014, “Generating a state t -design by diagonal quantum circuits,” *New J. Phys.* **16**, 053043.
- Napp, J. C., R. L. La Placa, A. M. Dalzell, F. G. S. L. Brandão, and A. W. Harrow, 2022, “Efficient Classical Simulation of Random Shallow 2D Quantum Circuits,” *Phys. Rev. X* **12**, 021021.
- Negrevergne, C., R. Somma, G. Ortiz, E. Knill, and R. Laflamme, 2005, “Liquid-state NMR simulations of quantum many-body problems,” *Phys. Rev. A* **71**, 032344.
- Neill, C., *et al.*, 2018, “A blueprint for demonstrating quantum supremacy with superconducting qubits,” *Science* **360**, 195–199.
- Neville, A., C. Sparrow, R. Clifford, E. Johnston, P. M. Birchall, A. Montanaro, and A. Laing, 2017, “Classical boson sampling algorithms with superior performance to near-term experiments,” *Nat. Phys.* **13**, 1153–1157.
- Nezami, S., 2021, “Permanent of random matrices from representation theory: Moments, numerics, concentration, and comments on hardness of boson-sampling,” [arXiv:2104.06423](https://arxiv.org/abs/2104.06423).
- Nielsen, M. A., and I. L. Chuang, 2010, *Quantum Computation and Quantum Information*, 10th ed. (Cambridge University Press, Cambridge, England).
- Nikolopoulos, G. M., 2019, “Cryptographic one-way function based on boson sampling,” *Quantum Inf. Process.* **18**, 259.
- Novo, L., J. Bermejo-Vega, and R. García-Patrón, 2021, “Quantum advantage from energy measurements of many-body quantum systems,” *Quantum* **5**, 465.
- Ofek, N., *et al.*, 2016, “Extending the lifetime of a quantum bit with error correction in superconducting circuits,” *Nature (London)* **536**, 441–445.
- O’Gorman, J., and E. T. Campbell, 2017, “Quantum computation with realistic magic-state factories,” *Phys. Rev. A* **95**, 032338.
- Oh, C., L. Jiang, and B. Fefferman, 2023, “On classical simulation algorithms for noisy boson sampling,” [arXiv:2301.11532](https://arxiv.org/abs/2301.11532).
- Oh, C., Y. Lim, B. Fefferman, and L. Jiang, 2022, “Classical Simulation of Boson Sampling Based on Graph Structure,” *Phys. Rev. Lett.* **128**, 190501.
- Oh, C., Y. Lim, Y. Wong, B. Fefferman, and L. Jiang, 2022, “Quantum-inspired classical algorithm for molecular vibronic spectra,” [arXiv:2202.01861](https://arxiv.org/abs/2202.01861).

- Oh, C., K. Noh, B. Fefferman, and L. Jiang, 2021, “Classical simulation of lossy boson sampling using matrix product operators,” *Phys. Rev. A* **104**, 022407.
- Oszmaniec, M., R. Augusiak, C. Gogolin, J. Kolodnyski, A. Acín, and M. Lewenstein, 2016, “Random Bosonic States for Robust Quantum Metrology,” *Phys. Rev. X* **6**, 041044.
- Oszmaniec, M., and D.J. Brod, 2018, “Classical simulation of photonic linear optics with lost particles,” *New J. Phys.* **20**, 092002.
- Oszmaniec, M., N. Dangniam, M. E. S. Morales, and Z. Zimborás, 2022, “Fermion sampling: A robust quantum computational advantage scheme using fermionic linear optics and magic input states,” *PRX Quantum* **3**, 020328.
- Paesani, S., Y. Ding, R. Santagati, L. Chakhmakhchyan, C. Vigliar, K. Rottwitz, L. K. Oxenløwe, J. Wang, M. G. Thompson, and A. Laing, 2019, “Generation and sampling of quantum states of light in a silicon chip,” *Nat. Phys.* **15**, 925–929.
- Pallister, S., N. Linden, and A. Montanaro, 2018, “Optimal Verification of Entangled States with Local Measurements,” *Phys. Rev. Lett.* **120**, 170502.
- Pan, F., K. Chen, and P. Zhang, 2022, “Solving the Sampling Problem of the Sycamore Quantum Circuits,” *Phys. Rev. Lett.* **129**, 090502.
- Pan, F., and P. Zhang, 2022, “Simulation of Quantum Circuits Using the Big-Batch Tensor Network Method,” *Phys. Rev. Lett.* **128**, 030501.
- Pan, F., P. Zhou, S. Li, and P. Zhang, 2020, “Contracting Arbitrary Tensor Networks: General Approximate Algorithm and Applications in Graphical Models and Quantum Circuit Simulations,” *Phys. Rev. Lett.* **125**, 060503.
- Paturi, R., 1992, “On the degree of polynomials that approximate symmetric Boolean functions (preliminary version),” in *Proceedings of the 24th Annual ACM Symposium on Theory of Computing (STOC '92), Victoria, British Columbia, Canada, 1992* (Association for Computing Machinery, New York), pp. 468–474, 10.1145/129712.129758.
- Pednault, E., J. A. Gunnels, G. Nannicini, L. Horesh, T. Magerlein, E. Solomonik, E. W. Draeger, E. T. Holland, and R. Wisnieff, 2017, “Pareto-efficient quantum circuit simulation using tensor contraction deferral,” [arXiv:1710.05867](https://arxiv.org/abs/1710.05867).
- Pednault, E., J. A. Gunnels, G. Nannicini, L. Horesh, and R. Wisnieff, 2019, “Leveraging secondary storage to simulate deep 54-qubit Sycamore circuits,” [arXiv:1910.09534](https://arxiv.org/abs/1910.09534).
- Peruzzo, A., J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. O’Brien, 2014, “A variational eigenvalue solver on a photonic quantum processor,” *Nat. Commun.* **5**, 4213.
- Phillips, D. S., M. Walschaers, J. J. Renema, I. A. Walmsley, N. Treps, and J. Sperling, 2019, “Benchmarking of Gaussian boson sampling using two-point correlators,” *Phys. Rev. A* **99**, 023836.
- Popova, A. S., and A. N. Rubtsov, 2021, “Cracking the quantum advantage threshold for Gaussian boson sampling,” [arXiv:2106.01445](https://arxiv.org/abs/2106.01445).
- Porter, C. E., and R. G. Thomas, 1956, “Fluctuations of nuclear reaction widths,” *Phys. Rev.* **104**, 483–491.
- Preskill, J., 2012, “Quantum computing and the entanglement frontier,” [arXiv:1203.5813](https://arxiv.org/abs/1203.5813).
- Qassim, H., H. Pashayan, and D. Gosset, 2021, “Improved upper bounds on the stabilizer rank of magic states,” *Quantum* **5**, 606.
- Qi, H., D. J. Brod, N. Quesada, and R. García-Patrón, 2020, “Regimes of Classical Simulability for Noisy Gaussian Boson Sampling,” *Phys. Rev. Lett.* **124**, 100502.
- Qi, H., D. Cifuentes, K. Brádler, R. Israel, T. Kalajdziewski, and N. Quesada, 2020, “Efficient sampling from shallow Gaussian quantum-optical circuits with local interactions,” [arXiv:2009.11824](https://arxiv.org/abs/2009.11824).
- Quesada, N., 2019, “Franck-Condon factors by counting perfect matchings of graphs with loops,” *J. Chem. Phys.* **150**, 164113.
- Quesada, N., and J. M. Arrazola, 2020, “Exact simulation of Gaussian boson sampling in polynomial space and exponential time,” *Phys. Rev. Res.* **2**, 023005.
- Quesada, N., J. M. Arrazola, and N. Killoran, 2018, “Gaussian boson sampling using threshold detectors,” *Phys. Rev. A* **98**, 062322.
- Quesada, N., R. S. Chadwick, B. A. Bell, J. M. Arrazola, T. Vincent, H. Qi, and R. García-Patrón, 2022, “Quadratic speed-up for simulating Gaussian boson sampling,” *PRX Quantum* **3**, 010306.
- Quesada, N., L. G. Helt, J. Izaac, J. M. Arrazola, R. Shahrokhshahi, C. R. Myers, and K. K. Sabapathy, 2019, “Simulating realistic non-Gaussian state preparation,” *Phys. Rev. A* **100**, 022341.
- Rahimi-Keshari, S., A. P. Lund, and T. C. Ralph, 2015, “What Can Quantum Optics Say about Computational Complexity Theory?,” *Phys. Rev. Lett.* **114**, 060501.
- Rahimi-Keshari, S., T. C. Ralph, and C. M. Caves, 2016, “Sufficient Conditions for Efficient Classical Simulation of Quantum Optics,” *Phys. Rev. X* **6**, 021039.
- Rakhmanov, E. A., 2007, “Bounds for polynomials with a unit discrete norm,” *Ann. Math.* **165**, 55–88.
- Raussendorf, R., S. Bravyi, and J. Harrington, 2005, “Long-range quantum entanglement in noisy cluster states,” *Phys. Rev. A* **71**, 062313.
- Raussendorf, R., and H. J. Briegel, 2001, “A One-Way Quantum Computer,” *Phys. Rev. Lett.* **86**, 5188–5191.
- Raussendorf, R., D. E. Browne, and H. J. Briegel, 2003, “Measurement-based quantum computation on cluster states,” *Phys. Rev. A* **68**, 022312.
- Raussendorf, R., J. Harrington, and K. Goyal, 2006, “A fault-tolerant one-way quantum computer,” *Ann. Phys. (Amsterdam)* **321**, 2242–2270.
- Raz, R., and A. Tal, 2019, “Oracle separation of BQP and PH,” in *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing (STOC 2019), Phoenix, 2019* (Association for Computing Machinery, New York), pp. 13–23, 10.1145/3313276.3316315.
- Reagor, M., *et al.*, 2018, “Demonstration of universal parametric entangling gates on a multi-qubit lattice,” *Sci. Adv.* **4**, eaao3603.
- Reed, I. S., 1954, “A class of multiple-error-correcting codes and the decoding scheme,” *Trans. IRE Prof. Group Inf. Theory* **4**, 38–49.
- Reed, I. S., and G. Solomon, 1960, “Polynomial codes over certain finite fields,” *J. Soc. Ind. Appl. Math.* **8**, 300–304.
- Renema, J. J., 2020a, “Marginal probabilities in boson samplers with arbitrary input states,” [arXiv:2012.14917](https://arxiv.org/abs/2012.14917).
- Renema, J. J., 2020b, “Simulability of partially distinguishable superposition and Gaussian boson sampling,” *Phys. Rev. A* **101**, 063840.
- Renema, J. J., A. Menssen, W. R. Clements, G. Triginer, W. S. Kolthammer, and I. A. Walmsley, 2018, “Efficient Classical Algorithm for Boson Sampling with Partially Distinguishable Photons,” *Phys. Rev. Lett.* **120**, 220502.
- Renema, J. J., V. S. Shchesnovich, and R. García-Patrón, 2019, “Classical simulability of noisy boson sampling,” [arXiv:1809.01953v2](https://arxiv.org/abs/1809.01953v2).
- Ringbauer, M., *et al.*, 2022, “Verifiable measurement-based quantum random sampling with trapped ions” (to be published).
- Rinott, Y., T. Shoham, and G. Kalai, 2022, “Statistical aspects of the quantum supremacy demonstration,” *Stat. Sci.* **37**, 322–347.
- Roga, W., and M. Takeoka, 2020, “Classical simulation of boson sampling with sparse output,” *Sci. Rep.* **10**, 14739.

- Ryan-Anderson, C., *et al.*, 2022, “Implementing fault-tolerant entangling gates on the five-qubit code and the color code,” *arXiv:2208.01863*.
- Ryser, H. J., 1963, *Combinatorial Mathematics* (American Mathematical Society, Providence).
- Satzinger, K. J., *et al.*, 2021, “Realizing topologically ordered states on a quantum processor,” *Science* **374**, 1237–1241.
- Scheel, S., 2008, “Permanents in linear optical networks,” *Acta Phys. Slovaca* **58**, 675.
- Schollwöck, U., 2005, “The density-matrix renormalization group,” *Rev. Mod. Phys.* **77**, 259–315.
- Schuch, N., M. M. Wolf, F. Verstraete, and J. I. Cirac, 2007, “Computational Complexity of Projected Entangled Pair States,” *Phys. Rev. Lett.* **98**, 140506.
- Schuld, M., K. Brádler, R. Israel, D. Su, and B. Gupta, 2020, “Measuring the similarity of graphs with a Gaussian boson sampler,” *Phys. Rev. A* **101**, 032314.
- Schutski, R., T. Khakhulin, I. Oseledets, and D. Kolmakov, 2020, “Simple heuristics for efficient parallel tensor contraction and quantum circuit simulation,” *Phys. Rev. A* **102**, 062614.
- Schwarz, M., and M. van den Nest, 2013, “Simulating quantum circuits with sparse output distributions,” *arXiv:1310.6749*.
- Sekatski, P., J.-D. Bancal, S. Wagner, and N. Sangouard, 2018, “Certifying the Building Blocks of Quantum Computers from Bell’s Theorem,” *Phys. Rev. Lett.* **121**, 180505.
- Sempere-Llagostera, S., R. B. Patel, I. A. Walmsley, and W. S. Kolthammer, 2022, “Experimentally finding dense subgraphs using a time-bin encoded Gaussian boson sampling device,” *arXiv:2204.05254*.
- Shalm, L. K., *et al.*, 2015, “Strong Loophole-Free Test of Local Realism,” *Phys. Rev. Lett.* **115**, 250402.
- Shchesnovich, V., 2022, “Boson sampling cannot be faithfully simulated by only the lower-order multi-boson interferences,” *arXiv:2204.07792*.
- Shchesnovich, V. S., 2013, “Asymptotic evaluation of bosonic probability amplitudes in linear unitary networks in the case of large number of bosons,” *Int. J. Quantum. Inf.* **11**, 1350045.
- Shchesnovich, V. S., 2014, “Sufficient condition for the mode mismatch of single photons for scalability of the boson-sampling computer,” *Phys. Rev. A* **89**, 022333.
- Shchesnovich, V. S., 2019, “Noise in boson sampling and the threshold of efficient classical simulatability,” *Phys. Rev. A* **100**, 012340.
- Shchesnovich, V. S., 2021, “Distinguishing noisy boson sampling from classical simulations,” *Quantum* **5**, 423.
- Shepherd, D., and M. J. Bremner, 2009, “Temporally unstructured quantum computation,” *Proc. R. Soc. A* **465**, 1413–1439.
- Shi, Y., 2002, “Both Toffoli and controlled-NOT need little help to do universal quantum computation,” *arXiv:quant-ph/0205115*.
- Shor, P. W., 1994, “Algorithms for quantum computation: Discrete logarithms and factoring,” in *Proceedings of the 35th Annual Symposium on Foundations of Computer Science, Santa Fe, 1994* (IEEE, New York), pp. 124–134, [10.1109/SFCS.1994.365700](https://doi.org/10.1109/SFCS.1994.365700).
- Shor, P. W., 1996, “Fault-tolerant quantum computation,” in *Proceedings of the 37th Annual Symposium on Foundations of Computer Science, Burlington, VT, 1996* (IEEE, New York), pp. 56–65, [10.1109/SFCS.1996.548464](https://doi.org/10.1109/SFCS.1996.548464).
- Shor, P. W., 1997, “Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer,” *SIAM J. Comput.* **26**, 1484–1509.
- Simon, D. R., 1994, “On the power of quantum computation,” in *Proceedings of the 35th Annual Symposium on Foundations of Computer Science (SFCS ’94), Santa Fe, 1994* (IEEE, New York), pp. 116–123, [10.1109/SFCS.1994.365701](https://doi.org/10.1109/SFCS.1994.365701).
- Simon, D. R., 1997, “On the power of quantum computation,” *SIAM J. Comput.* **26**, 1474–1483.
- Smelyanskiy, M., N. P. D. Sawaya, and A. Aspuru-Guzik, 2016, “qHiPSTER: The quantum high performance software testing environment,” *arXiv:1601.07195*.
- Somhorst, F. H. B., *et al.*, 2023, “Quantum simulation of thermodynamics in an integrated quantum photonic processor,” *Nat. Commun.* **14**, 3895.
- Spagnolo, N., *et al.*, 2014, “Efficient experimental validation of photonic boson sampling against the uniform distribution,” *Nat. Photonics* **8**, 615–620.
- Spring, J. B., *et al.*, 2013, “Boson sampling on a photonic chip,” *Science* **339**, 798–801.
- Stilck França, D., and R. García-Patrón, 2021, “Limitations of optimization algorithms on noisy quantum devices,” *Nat. Phys.* **17**, 1221–1227.
- Stilck França, D., and R. García-Patrón, 2022, “A game of quantum advantage: Linking verification and simulation,” *Quantum* **6**, 753.
- Stockmeyer, L., 1983, “The complexity of approximate counting,” *Proceedings of the 15th Annual ACM Symposium on Theory of Computing (STOC ’83), Boston, 1983* (Association for Computing Machinery, New York), 118–126, [10.1145/800061.808740](https://doi.org/10.1145/800061.808740).
- Sudan, M., 1997, “Decoding of Reed Solomon codes beyond the error-correction bound,” *J. Complexity* **13**, 180–193.
- Takeuchi, Y., A. Mantri, T. Morimae, A. Mizutani, and J. F. Fitzsimons, 2019, “Resource-efficient verification of quantum computing using Serfling’s bound,” *npj Quantum Inf.* **5**, 1–8.
- Takeuchi, Y., and T. Morimae, 2018, “Verification of Many-Qubit States,” *Phys. Rev. X* **8**, 021060.
- Tao, T., and V. Vu, 2009, “On the permanent of random Bernoulli matrices,” *Adv. Math.* **220**, 657–669.
- Terhal, B. M., and D. P. DiVincenzo, 2004, “Adaptive quantum computation, constant depth quantum circuits and Arthur-Merlin games,” *Quantum Inf. Comput.* **4**, 134–145.
- Thekkadath, G. S., S. Sempere-Llagostera, B. A. Bell, R. B. Patel, M. S. Kim, and I. A. Walmsley, 2022, “Experimental demonstration of Gaussian boson sampling with displacement,” *PRX Quantum* **3**, 020336.
- Tichy, M. C., 2014, “Interference of identical particles from entanglement to boson-sampling,” *J. Phys. B* **47**, 103001.
- Tichy, M. C., 2015, “Sampling of partially distinguishable bosons and the relation to the multidimensional permanent,” *Phys. Rev. A* **91**, 022316.
- Tichy, M. C., K. Mayer, A. Buchleitner, and K. Mølmer, 2014, “Stringent and Efficient Assessment of Boson-Sampling Devices,” *Phys. Rev. Lett.* **113**, 020502.
- Tillmann, M., B. Dakić, R. Heilmann, S. Nolte, A. Szameit, and P. Walther, 2013, “Experimental boson sampling,” *Nat. Photonics* **7**, 540–544.
- Toda, S., 1991, “PP is as hard as the polynomial-time hierarchy,” *SIAM J. Comput.* **20**, 865–877.
- Toda, S., and M. Ogiwara, 1992, “Counting classes are at least as hard as the polynomial-time hierarchy,” *SIAM J. Comput.* **21**, 316–328.
- Toffoli, T., 1980, “Reversible computing,” in *International Colloquium on Automata, Languages and Programming (ICALP ’80), Noordwijkerhout, Netherlands, 1980*, Lecture Notes in Computer Science Vol. 85, edited by Jaco de Bakker and Jan van Leeuwen (Springer, New York), pp. 632–644, [10.1007/3-540-10003-2_104](https://doi.org/10.1007/3-540-10003-2_104).

- Tóth, G., and O. Gühne, 2005, “Entanglement detection in the stabilizer formalism,” *Phys. Rev. A* **72**, 022340.
- Trevisan, L., 2008, “Lecture 6: Approximate counting,” in Lecture Notes on Computational Complexity (accessed April 2, 2022), <https://lucatrevisan.github.io/cs278-08/lecture06.pdf>.
- Trotzky, S., Y.-A. Chen, A. Flesch, I. P. McCulloch, U. Schollwöck, J. Eisert, and I. Bloch, 2012, “Probing the relaxation towards equilibrium in an isolated strongly correlated one-dimensional Bose gas,” *Nat. Phys.* **8**, 325–330.
- Trotzky, S., L. Pollet, F. Gerbier, U. Schnorrberger, I. Bloch, N. V. Prokof'ev, B. Svistunov, and M. Troyer, 2010, “Suppression of the critical temperature for superfluidity near the Mott transition: Validating a quantum simulator,” *Nat. Phys.* **6**, 998.
- Valiant, G., and P. Valiant, 2017, “An automatic inequality prover and instance optimal identity testing,” *SIAM J. Comput.* **46**, 429–455.
- Valiant, L. G., 1979, “The complexity of computing the permanent,” *Theor. Comput. Sci.* **8**, 189–201.
- Valido, A. A., and J. J. García-Ripoll, 2021, “Gaussian phase sensitivity of boson-sampling-inspired strategies,” *Phys. Rev. A* **103**, 032613.
- Vandersypen, L. M. K., M. Steffen, G. Breyta, C. S. Yannoni, R. Cleve, and I. L. Chuang, 2000, “Experimental Realization of an Order-Finding Algorithm with an NMR Quantum Computer,” *Phys. Rev. Lett.* **85**, 5452–5455.
- Vandersypen, L. M. K., M. Steffen, G. Breyta, C. S. Yannoni, M. H. Sherwood, and I. L. Chuang, 2001, “Experimental realization of Shor’s quantum factoring algorithm using nuclear magnetic resonance,” *Nature (London)* **414**, 883–887.
- Vergis, A., K. Steiglitz, and B. Dickinson, 1986, “The complexity of analog computation,” *Math. Comput. Simul.* **28**, 91–113.
- Verstraete, F., J. I. Cirac, and V. Murg, 2008, “Matrix product states, projected entangled pair states, and variational renormalization group methods for quantum spin systems,” *Adv. Phys.* **57**, 143.
- Villalonga, B., S. Boixo, B. Nelson, C. Henze, E. Rieffel, R. Biswas, and S. Mandrà, 2019, “A flexible high-performance simulator for verifying and benchmarking quantum circuits implemented on real hardware,” *npj Quantum Inf.* **5**, 1–16.
- Villalonga, B., D. Lyakh, S. Boixo, H. Neven, T. S. Humble, R. Biswas, E. G. Rieffel, A. Ho, and S. Mandrà, 2020, “Establishing the quantum supremacy frontier with a 281 Pfl/s simulation,” *Quantum Sci. Technol.* **5**, 034003.
- Villalonga, B., M. Y. Niu, L. Li, H. Neven, J. C. Platt, V. N. Smelyanskiy, and S. Boixo, 2021, “Efficient approximation of experimental Gaussian boson sampling,” [arXiv:2109.11525](https://arxiv.org/abs/2109.11525).
- Vollbrecht, K. G. H., and J. I. Cirac, 2008, “Quantum Simulators, Continuous-Time Automata, and Translationally Invariant Systems,” *Phys. Rev. Lett.* **100**, 010501.
- Walschaers, M., J. Kuipers, J.-D. Urbina, K. Mayer, M. C. Tichy, K. Richter, and A. Buchleitner, 2016, “Statistical benchmark for BosonSampling,” *New J. Phys.* **18**, 032001.
- Wang, C. S., *et al.*, 2020, “Efficient Multiphoton Sampling of Molecular Vibronic Spectra on a Superconducting Bosonic Processor,” *Phys. Rev. X* **10**, 021060.
- Wang, H., *et al.*, 2017, “High-efficiency multiphoton boson sampling,” *Nat. Photonics* **11**, 361–365.
- Wang, H., *et al.*, 2019, “Boson Sampling with 20 Input Photons and a 60-Mode Interferometer in a 10^{14} -Dimensional Hilbert Space,” *Phys. Rev. Lett.* **123**, 250503.
- Wang, S.-T., and L.-M. Duan, 2016, “Certification of boson sampling devices with coarse-grained measurements,” [arXiv:1601.02627](https://arxiv.org/abs/1601.02627).
- Watrous, J., 2018, *The Theory of Quantum Information*, 1st ed. (Cambridge University Press, Cambridge, England).
- Welch, L. R., and E. R. Berlekamp, 1986, “Error correction for algebraic block codes,” U.S. Patent No. US4633470A, <https://patents.google.com/patent/US4633470A/en>.
- Williams, V. V., 2015, “Hardness of easy problems: Basing hardness on popular conjectures such as the strong exponential time hypothesis,” in *Proceedings of the 10th International Symposium on Parameterized and Exact Computation (IPEC 2015)*, Patras, Greece, 2015, Leibniz International Proceedings in Informatics (LIPIcs) Vol. 43, edited by T. Husfeldt and I. Kanj (Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, Dagstuhl, Germany), pp. 17–29, [10.4230/LIPIcs.IPEC.2015.17](https://doi.org/10.4230/LIPIcs.IPEC.2015.17).
- Wu, J., Y. Liu, B. Zhang, X. Jin, Y. Wang, H. Wang, and X. Yang, 2018, “A benchmark test of boson sampling on Tianhe-2 super-computer,” *Natl. Sci. Rev.* **5**, 715–720.
- Wu, Y., *et al.*, 2021, “Strong Quantum Computational Advantage Using a Superconducting Quantum Processor,” *Phys. Rev. Lett.* **127**, 180501.
- Yamakawa, T., and M. Zhandry, 2022, “Verifiable quantum advantage without structure,” [arXiv:2204.02063](https://arxiv.org/abs/2204.02063).
- Yoganathan, M., R. Jozsa, and S. Strelchuk, 2019, “Quantum advantage of unitary Clifford circuits with magic state inputs,” *Proc. R. Soc. A* **475**, 20180427.
- Yung, M.-H., and X. Gao, 2017, “Can chaotic quantum circuits maintain quantum supremacy under noise?,” [arXiv:1706.08913](https://arxiv.org/abs/1706.08913).
- Yung, M.-H., X. Gao, and J. Huh, 2019, “Universal bound on sampling bosons in linear optics and its computational implications,” *Natl. Sci. Rev.* **6**, 719–729.
- Zhong, H.-S., *et al.*, 2018, “12-Photon Entanglement and Scalable Scattershot Boson Sampling with Optimal Entangled-Photon Pairs from Parametric Down-Conversion,” *Phys. Rev. Lett.* **121**, 250505.
- Zhong, H.-S., *et al.*, 2020, “Quantum computational advantage using photons,” *Science* **370**, 1460–1463.
- Zhong, H.-S., *et al.*, 2021, “Phase-Programmable Gaussian Boson Sampling Using Stimulated Squeezed Light,” *Phys. Rev. Lett.* **127**, 180502.
- Zhou, L., S.-T. Wang, S. Choi, H. Pichler, and M. D. Lukin, 2020, “Quantum Approximate Optimization Algorithm: Performance, Mechanism, and Implementation on Near-Term Devices,” *Phys. Rev. X* **10**, 021067.
- Zhou, T., and A. Nahum, 2019, “Emergent statistical mechanics of entanglement in random unitary circuits,” *Phys. Rev. B* **99**, 174205.
- Zhou, Y., E. M. Stoudenmire, and X. Waintal, 2020, “What Limits the Simulation of Quantum Computers?,” *Phys. Rev. X* **10**, 041038.
- Zhu, D., *et al.*, 2021, “Interactive protocols for classically-verifiable quantum advantage,” [arXiv:2112.05156](https://arxiv.org/abs/2112.05156).
- Zhu, H., and M. Hayashi, 2019, “Efficient Verification of Hypergraph States,” *Phys. Rev. Appl.* **12**, 054047.
- Zhu, Q., *et al.*, 2022, “Quantum computational advantage via 60-qubit 24-cycle random circuit sampling,” *Sci. Bull.* **67**, 240–245.
- Zlokapa, A., S. Boixo, and D. Lidar, 2023, “Boundaries of quantum supremacy via random circuit sampling,” *npj Quantum Inf.* **9**, 36.
- Zwanenburg, F. A., A. S. Dzurak, A. Morello, M. Y. Simmons, L. C. L. Hollenberg, G. Klimeck, S. Rogge, S. N. Coppersmith, and M. A. Eriksson, 2013, “Silicon quantum electronics,” *Rev. Mod. Phys.* **85**, 961–1019.