Google's 2019 "Quantum Supremacy" Claims: Data, Documentation, and Discussion

Gil Kalai, Yosef Rinott & Tomer Shoham

October 11, 2022

Abstract

In October 2019, *Nature* published a paper [3] describing an experimental work that took place at Google. The paper claims to demonstrate quantum (computational) supremacy on a 53-qubit quantum computer. Since September 2019 the authors have been involved in a long-term project to study various statistical aspects of the Google experiment. In particular, we have been trying to gather the relevant data and information, to reconstruct and verify those parts of the Google 2019 supremacy experiments that are based on classical computations (unless they require too heavy computation), and to put the data under statistical analysis. We have now (August 2022) concluded the part relating to the gathering of data and information needed for our study of the 2019 Google experiment, and this document describes the available data and information for the Google 2019 experiment and some of our results and plans.

1 Introduction

The 2019 paper "Quantum supremacy using a programmable superconducting processor" [3] claimed that Google's Sycamore processor performed a certain computation in about 200 seconds, while a state-of-the-art classical supercomputer would take, according to Google's team estimates, approximately 10,000 years to perform the same computation. Google's Sycamore quantum computer performed a *sampling task*; that is, it generates random bitstrings, of length 53, with considerable noise, from a certain discrete probability distribution supported on all such 2^{53} bitstrings. The specific Google's sampling task is referred to as random circuit sampling (RCS, for short). Google's announcement of quantum supremacy was compared by various writers (See, e.g. [1]) to landmark technological achievements such as the Wright brothers' invention of a motor-operated airplane, the Sputnik, and the landing on the moon, as well as to landmark scientific achievement such as Fermi's demonstration of a nuclear chain reaction, the discovery of the Higgs boson, and the LIGO detection of gravitational waves.

In 2020 a team from the University of Science and Technology of China (USTC) claimed [28] that the sampling task computed by their photonic *Jiuzhang* quantum computer, would take 2.5 billion years to perform on a classical supercomputer. USTC's quantum computers also required about 200 seconds for their tasks. This task is referred to as Gaussian boson sampling (GBS, for short). In 2021, another team from USTC repeated the Google 2019 RCS experiment with a superconducting processor *Zuchongzhi* of 60 qubits and depth 24 [27, 30], and claimed to achieve an even stronger form of quantum advantage compared to the Google 2019 experiment.

The Google experiment represented a very large leap in various aspects of the human ability to control noisy quantum systems. For example the Google AI team previously reported an experiment with nine qubits [15]. This leap is especially impressive in terms of the dimensions of the Hilbert space representing a state of the computer. (From dimension 100–500 in [15] to dimension 10^{16} in [3].)

As for Google's "quantum supremacy" claims, progress by several groups [16, 17, 29, 11, 12, 18, 8] (and more) exhibits classical algorithms which are ten orders of magnitude faster than those used in the Google paper. This was achieved, for example, by Pan, Chen, and Zhang [17] in 2021. Our study concentrates on other aspects of the Google experiment.

As we already mentioned, the Google's announcement was regarded as a major scientific and technological event. On its own it gave some evidence that the "strong Church–Turing thesis" had been violated and it was described as an ironclad refutation of claims by some scientists (including Kalai) that quantum computation is not possible. Of no less importance is that quantum supremacy was considered as a major intermediate step towards exhibiting experimentally quantum error-correction codes needed for building larger quantum computers. The announcement of quantum supremacy stirred a great deal of enthusiasm among scientists and in the general public, and garnered significant media attention. It had substantial

Figure 1: *The price of bitcoin in a period of 4 weeks around 23.9.2019. Source: CoinDesk*

impact; for example, following the media attention surrounding the leaking of the supremacy claims around September 22, 2019, the value of bitcoin (and other digital currencies) sharply dropped by more than 10% and it is a reasonable possibility that given the potential effects of quantum computers on the safety of digital currencies, the quantum supremacy claims caused this drop.

Putting the Google supremacy claims under scrutiny

The Google paper appeared in October 2019 and a month earlier it was briefly posted on a NASA server and became publicly available. Since the announcement of the quantum supremacy claim several researchers, including Kalai, raised various concerns. Since December 2019, the authors have been involved in a long-term project to study various statistical aspects of the Google experiment. In particular, we have been trying to gather the relevant data and information and to reconstruct and verify those parts of the Google 2019 supremacy experiments that are based on classical computations. (Unless they require too heavy computation; we carried some heavy computation on the cloud for which we put a cap of 2000 dollars on our spending.) We also performed several "sanity tests" of the experiment.

2 Google's 2019 quantum supremacy claim

2.1 A brief background

In this draft we will assume knowledge of the Google 2019 experiment, Google's noise model, Google's $\mathcal{F}_{XEB}$ linear cross entropy fidelity estimator, and Google's formula (77) in [4] for predicting the fidelity of a circuit from the fidelity of its components. We will give here a brief summary of these topics.

The Google 2019 experiment is based on the building of a quantum computer (circuit) with n superconducting qubits, that perform m rounds of computation. The computation is carried out by a 1-qubit and 2-qubit gates. At the end of the computation the qubits are measured, leading to a string of zeroes and ones of length n. The ultimate experiment was for $n = 53$ and $m = 20$. It involved 1113 1-qubit gates and 430 2-qubit gates. For that experiment the Google team produced a sample of three million 0-1 vectors of length 53.

Every circuit C with n qubits describes a probability distribution $\mathbf{P}_C(x)$ for 0-1 vectors of length n. (In fact, it describes a 2^n-dimensional vector of complex amplitudes; to every 0-1 vector x, there is an associated amplitude $z(x)$ and $\mathbf{P}_C(x) = |z(x)|^2$.) The quantum computer enables to sample according to the probability distribution $\mathbf{P}_C(x)$ with a considerable amount of noise. When n and m are not too large classical simulations enable to compute the amplitudes themselves (and hence the probabilities $\mathbf{P}_C(x)$). Google's supremacy claims are based on the fact that these classical simulations quickly become infeasible as n and m grow.

The Google basic noise model for the noisy samples their Sycamore device actually produces is

$$\mathbf{N}_C(x) = \phi \mathbf{P}_C + (1-\phi)2^{-n}, \tag{1}$$

where ϕ is the *fidelity*, a parameter that roughly describes the quality of the sample. (The fidelity has a precise meaning in terms of the actual noisy quantum process carried out by the Google Sycamore device.)

Based on their noise model (and the fact that the distribution $\mathbf{P}_C$ is an instance of a Porter–Thomas distribution) the Google paper describes a statistic called the linear cross entropy estimator (and denoted by $\mathcal{F}_{XEB}$.) Once the quantum computer produces a sequence $\tilde{\mathbf{x}}$ of N samples $\tilde{\mathbf{x}} = (\tilde{\mathbf{x}}^{(1)}, \tilde{\mathbf{x}}^{(2)}, \ldots, \tilde{\mathbf{x}}^{(N)})$, the following "linear cross entropy" estimator $\mathcal{F}_{XEB}$ for the fidelity is computed

$$\mathcal{F}_{XEB}(\tilde{\mathbf{x}}) = \frac{1}{N}\sum_{i=1}^{N} 2^n \mathbf{P}_C(\tilde{\mathbf{x}}^{(i)}) - 1. \tag{2}$$

Computing $\mathcal{F}_{XEB}$ requires knowledge of $\mathbf{P}_C(x)$ for sampled bitstrings.

The Google supremacy claim is based also on the following a priori prediction for the fidelity of a circuit based on the probabilities of error for the individual components.

$$\hat{\phi} \ = \ \prod_{g\in\mathcal{G}_1}(1-e_g)\prod_{g\in\mathcal{G}_2}(1-e_g)\prod_{q\in\mathcal{Q}}(1-e_q). \tag{3}$$

Here $\mathcal{G}_1$ is the set of 1-gates (gates operating on a single qubit), $\mathcal{G}_2$ is the set of 2-gates (gates operating on two qubits), and $\mathcal{Q}$ is the set of qubits. For a gate g, the term e_g in the formula refers to the probability of an error (1 minus the fidelity) of the individual gate g. For a qubit q, e_q is the probability of a read-out error when we measure the qubit q.

The Google supremacy paper [3] made two crucial claims regarding the ultimate 53-qubit samples.

A) The fidelity ϕ of their sample is above 1/1000.

B) Producing a sample with similar fidelity would require 10,000 years on a supercomputer.

For claim A) regarding the value of ϕ, the argument relies on an extrapolation argument that has two ingredients. One ingredient is a few hundred experiments in the classically tractable regime: the regime where the probability distribution $\mathbf{P}_C$ can be computed by a classical computer and the performance of the quantum computer can be tested directly. The other ingredient is the theoretical formula (3) for predicting the fidelity. According to the paper, the fidelity of entire circuits closely agrees with the prediction of formula (3) (Formula (77) in [4])with a deviation below 10–20 percent. There are around 200 reported experiments in the classically tractable regime including ones carried out on simplified circuits (which are easier to simulate on classical computers). These experiments support the claim that the prediction given by Formula (77) for the fidelity is indeed very robust and applies to the 53-qubit circuit in the supremacy regime.

For claim B) regarding the classical difficulty, the Google team mainly relies on extrapolation from the running time of a specific algorithm they used. They also rely on the computational complexity support for the assertion that the task at hand is asymptotically difficult.

2.2 The types of circuits of the Google experiment

The circuits used in the 2019 Google supremacy experiment had the following structure. The qubits were arranged on a planar grid, so a single qubit was identified via two coordinates, like qubit $(3, 3)$. The circuits had two types of layers: one type of layer consists of 2-gates acting on pairs of (neighboring) qubits. After each such layer of two gates there was another layer of randomly chosen 1-gates acting on every qubit. The layers of 1-gates consist of the "programmable" ingredient in the experiment and they change from circuit to circuit. The layers of 2-gates are fixed throughout the experiment according to a certain pattern. The pattern **EFGH** was used in all experiments conducted between February to May 2019 and a new pattern **ABCDCDAB** was used in June 2019 to produce the samples of the "supremacy" circuit, namely circuits which would require a huge classical computation. Each letter like **E** corresponds to a fixed set of 2-gates acting in parallel on the qubits (with the convention that for circuits with a smaller number of qubits we regard only 2-gates that involve the qubits in the circuit). So Pattern **EFGH** means that we first apply a layer of random 1-gates on all qubits, then apply 2-gates according to **E**, next apply another layer of random 1-gates on all qubits, then apply 2-gates according to **F** and continue (in a periodic manner). The new pattern **ABCDCDAB** is based on new types of layers for the 2-gates and on a period of length eight.

The depth m of a circuit refers to the number of layers of 2-qubits. So, for example, the 2-gates layers of a circuit with pattern **EFGH** of depth $m = 14$ are **E, F, G, H, E, F, G, H, E, F, G, H, E, F**.

The six types of circuits

The Google 2019 experiment relies on six types of circuits. The first three types are:

a) The full circuits of pattern **EFGH**,

b) The elided circuits of pattern **EFGH**,

c) The patch circuits of pattern **EFGH**.

Types b) and c) are simplified forms of type a) for which there are quick algorithms to compute the amplitudes for all values of n up to 53. In particular, patch circuits consist of two separate circuits on two disjoint sets of qubits. The Google paper [3] is based on running the quantum computer on a few hundred circuits of type a) b) and c) and computing the fidelity estimator based on the amplitudes.

There was also a preliminary stage of calibration that, based on experiments on 1-qubit and 2-qubit circuits, determined the precise adjustments for the experimental 2-gates. Those adjustments are the same for all circuits of type a), b) and c). See Section 3.

As far as we understand, the initial plan of the Google team to demonstrate quantum supremacy was to compute the empirical fidelities for circuits of type b) and c) for a number of qubits between 12 and 53 and for circuits of type a) for a number of qubits up to $n = 43$ and to use this information to estimate the fidelity for circuits of type a) with 53 qubits and depths (m) between 12 and 20.

However, in May 2019, the Google team discovered an efficient classical algorithm for the circuits of type a). (Because of this discovery, the full circuits with pattern **EFGH** are also referred to as "verifiable full circuits.") The new algorithm discovered by the Google team is related to tensor networks methods which later led to important discoveries in this direction.[1] Using this algorithm, the Google team computed the amplitudes for the experimental bitstrings for type a) circuits with $n = 53$ $m = 14$.

Subsequently, for the purpose of demonstrating quantum supremacy, the Google team moved in May 2019 to a different architecture based on a new pattern **ABCDCDAB** which is harder to simulate by classical computers. This led to the following three new types of circuits.

[1] The Google team pointed out that they also implemented a related tensor network method in 2017 [6].

d) The full circuits with pattern **ABCDCDAB**. (These circuits are also referred to as "supremacy full circuits.")

e) The elided circuits of pattern **ABCDCDAB**,

f) The patch circuits of pattern **ABCDCDAB**.

For this architecture the Google team produced samples only with $n = 53$ and depths $12, 14, 16, 18$ and 20. The simulators available to the Google team were not powerful enough for computing the amplitudes and for checking the fidelity for circuits of type (d).

The circuits with pattern **ABCDCDAB** required to run the calibration process again on 1-qubit and 2-qubit circuits based on the new architecture and this has led to new adjustments for the 2-gates. Circuits of the new pattern **ABCDCDAB** (and the calibration process) were not tested for less than 53 qubits.

2.3 A chronology of the Google Sycamore experiments

Prior to the development of the Sycamore quantum computer, in 2018, the Google team had attempted to develop a 72-qubit chip called "Bristlecone," but due to difficulties the team later proceeded to the development of the Sycamore 54-qubit computer (with 53 effective qubits).

Here is a list of experiments for the Sycamore quantum computer including early experiments based on earlier calibration methods. For every experiment fresh new random circuits were generated. (Namely, the improvements in the calibration procedures in a certain experiment were not based on improving the fidelity for circuits that were used in an earlier experiment.)

Each experiment took 1-2 days to perform. Each of the first six experiments is based on an improved method for calibration. There were six experiments for circuits of pattern **EFGH** and one (the last) experiment is based on the new architecture with pattern **ABCDCDAB**.

2.3.1 Earlier experiments

1. Date: 02/08/2019

   Full circuits with pattern **EFGH**; numbers of qubits: $2, 3, 4, \ldots, 49$

2. Date: 02/13/2019

   Full circuits with pattern **EFGH**; numbers of qubits: $2, 3, 4, \ldots, 49$

3. Date: 03/14/2019

   Full circuits with pattern **EFGH**; numbers of qubits: $2, 3, 4, \ldots, 51$

4. Date: 03/22/2019

   Full circuits with pattern **EFGH**; numbers of qubits: $33, 34, \ldots, 51$

5. Date: 03/27/2019

   Full circuits with pattern **EFGH**; numbers of qubits: $10, 20, 33, 34, 35, 36, 37, 38, 51$

2.3.2 The experiments on which the Google paper is based

6. Date: 04/22/2019

   Patch, elided, and full circuits with pattern **EFGH**; Numbers of qubits: $12, 14, 16, \ldots, 36, 38, 39, 40, \ldots, 50, 51$, depth $m = 14$

   A few days later the Sycamore quantum computer produced the 53-qubit samples for patch, elided, and full circuits with the **EFGH** pattern also with depth $m = 14$.

7. Date: 06/13/2019

   Numbers of qubits: 53, depth $m = 12, 14, 16, 18, 20$

   Type of circuits: patch, elided, full with pattern **ABCDCDAB**.

3 The nature of the calibration process

The Sycamore quantum computer has systematic deviations from the ideal circuit it describes, and the calibration process accounts for small systematic errors in the experimental circuits compared to the random circuits they represent. The calibration process is crucial since the combined effect of such systematic errors can slash the fidelity to zero and it is important to account for them.

By the calibration process we refer in this paper to a method which, based on multiple runs of 1-qubit and 2-qubit quantum circuits, adjusted the definition of the 2-gates of the experimental circuits so that certain systematic forms of noise will be greatly reduced. This calibration adjustment was carried out simultaneously for all the experiments for patch, elided, and verifiable full circuits.

It is easy to identify from the Python program how the random circuits were modified by the calibration. (We could also use the QSIM files for that purpose.) For example, consider a single 2-gate acting on two qubits $A = (3, 3)$ and $B = (3, 4)$.

In the ideal description of the circuit, this 2-gate is described in the Python program for the circuit as follows:

```
cirq.Moment(
   cirq.FSimGate(theta=1.57079632679, phi=0.52359877559).
   on( cirq.GridQubit(3, 3), cirq.GridQubit(3, 4)),
```

In this 2-gate (that is sometimes referred to as the "standard 2-gate") $\theta = \pi/2$, and $\phi = \pi/6$. We will now show the part of the program that describes 2-gates modifications for this specific gate. The calibration consists of adding to the description of the circuits, two fixed rotations on each of the 2 qubits involved in this particular 2-gate.

The two added single qubit rotations are

```
cirq.Rz (np.pi * 2.5333591271878086).  on (cirq.GridQubit (3, 3)),
cirq.Rz (np.pi * -2.4748096263683066).  on (cirq.GridQubit (3, 4)),
```

and these two added single qubit rotations are the same for every appearance of the 2-gate involving A and B. The values that appear here in the single qubit rotations, as well as the adjustments for the definition of 2-gates, were computed based on many experiments for 1-qubit and 2-qubit circuits for these two qubits (3,3) and (3,4). (We neither have the precise algorithm for this computation nor the data with the outcomes of the 1-qubit and 2-qubit experiments.)

In addition, the 2-gate acting on qubits A and B is modified to

```
cirq.Moment(
   cirq.FSimGate(theta=1.2947043217999283, phi=0.4859467238431821).
   on( cirq.GridQubit(3, 3), cirq.GridQubit(3, 4)),
```

The calibrated values $\theta = 1.2947043217999283$ and $\phi = 0.4859467238431821$ are the same for all 2-gates acting on these two qubits in all circuits. (Sometimes the calibrated 2-gate is referred to as "native" 2-gate.) These values for θ and ϕ replace the values for the "standard" gate in the original random circuit $\theta = \pi/2$, and $\phi = \pi/6$.

Remarks:

1. We note that the adjustments for 2-gate involving qubits A and B depends (only) on multiple runs of 1-qubit and 2-qubit quantum circuits involving these two qubits.

2. We also note that there might have been additional actions to account for the fact that certain components of the Sycamore chip degrade over time.

3. In the December 2020 Internet discussion [22] Sergio Boixo from the Google team referred to Figure S30 in [4] and asserted that in the specific case of the 2019 experiment using "Sycamore" standard gates instead of native gates halves the fidelity. However, this claim referred only to the two-gate adjustments and not to the one-gate adjustments, and moving to the original circuit with standard gates slashes the fidelity to zero. (The effect of the two-gate adjustments represents the worse-case situation for $n = 53, m = 20$, and it is smaller for smaller circuits.)

Sergio Boixo later clarified that since 2020 (for later experiments but not for the 2019 supremacy experiment) the Google team developed a method to carry out the 1-qubit calibrations using physical control of the device rather than using a modification of the definition of the circuit. (This later development is not relevant to the 2019 experiment and we did not study it.)

4 Data, documentations, and discussions

4.1 The supplementary data

The Google paper was leaked around September 23, 2019 and the paper was published by "Nature" a month later on October 23, 2019. The data for the Google 2019 experiment can be found in [5]. The Google team uploaded data to the server in five dates as follows:

October 22/23, 2019, publication date

1. Files with 500,000 bitstrings for many of the experimental circuits with $n < 53$ were uploaded and for $n = 53$ samples consists of a few million bitstrings were uploaded. The bitstrings for several numbers of qubits n were missing on this date (e.g. missing for n=16, 32, 34, 38, 39, 42, ..., 51). In addition, the bitstrings for the patch circuits were missing.

2. QSIM files with descriptions for all the experimental circuits except the patch circuits were uploaded, as well as Python files for computing their amplitudes.

January 23, 2020

Missing bitstrings for full and elided circuits were uploaded. Amplitude files for all circuits (not including patch circuits) except (roughly) 200 circuits were uploaded. Some additional data including information on noise for individual components of the Sycamore quantum computer was uploaded.

January 22, 2021 & May 18, 2021

Amplitude files for all circuits (not including patch circuits) that were computed by the Google team were uploaded except six circuits for which the amplitudes were computed in Jülich (Jülich Research Center, Germany).

June 13, 2022

The Jülich amplitudes were uploaded and also bitstring files (100,000 bitstrings per circuit) and QSIM and Python files for the patch circuits.

4.2 Requests for data

4.2.1 Early requests from October 2019

Here are Kalai's requests from the Google team from early October 2019.

(a) Larger samples from the quantum circuits, comparison of the empirical distribution with the model.

(b) The bitstrings produced by the quantum computer and a description of the circuits

(c) The amplitudes that were computed for the samples for each circuit

(d) The full list of amplitudes. (Here we refer to the 2^n amplitides for all bitstrings and not only those in the sample.)

(e) Timetable for the experiments and calibrations

(f) (November 11, 2019.) The values of the individual fidelities e_q and e_g for every qubit q and gate g used in Formula 3.

4.2.2 Subsequent requests

(g) Readout errors information. We discussed this matter in January 2021 and we received useful data regarding the readout errors shortly afterwards. In this case, we received the required data quickly, in a very satisfactory way.

(h) Amplitudes for the verifiable experiments. (September 2021) As we already mentioned, the Google team developed useful algorithms and programs to compute the amplitudes for the experiments of pattern **EFGH**. (This was the reason for moving to a different pattern for the "supremacy experiment.") They used these programs for 10 53-qubit depth-14 circuits (this required several hours for each circuit on the Google supercomputer). For a useful reliability-control of the experiment we proposed that the amplitudes be computed for the remaining (over 100) circuits, and at least for 2 specific ones.

(i) The optimization programs to move from data on 1-qubit and 2-qubit circuits to 1-gate and 2-gate corrections in the definition of the circuits and the raw data gathered from 1-qubit and 2-qubit circuits. We requested it in January 2022.

4.3 Data provided by the Google team

1. Google's researchers were under press embargo and did not respond to most issues before the paper appeared.

2. As we already mentioned, on October 23, 2019 the paper was published in Nature, the supplementary raw data contains (b) (bit strings, description of the circuits in a form of a Python program to compute the amplitudes) for many of the circuits used in the experiments and on January 2020 bitstrings for other circuits were added. We requested the data for the "patch circuits" in June 2021 and again in December 2021. The data for the patch circuits was uploaded on the server in June 2022. (As we mentioned, the data has only 100K samples per every patch circuit.)

3. The Google team promised to supply the amplitudes (that they computed as part of the experiment) for their samples (c) . They uploaded amplitudes for many circuits in January 2020, and for additional two hundred circuits in May 2021. Uploading the amplitudes that were computed for 6 circuits in an external facility (Jülich) required approval of the Jülich team and they were uploaded in June 2022.

4. Regarding the request (d) for the full list of amplitudes, the Google team informed us that the full lists of amplitudes were discarded. (Indeed, for large values of n this is a huge amount of data.) Moreover, for circuits with pattern **EFGH** the Google team developed (around May 2019) algorithms for computing the amplitudes for the sample that did not require computing all the 2^n amplitudes and these algorithms were used in some of these cases.

5. Regarding larger samples (a). Until August 2022, the Google team did not supply any further samples produced by the quantum computer. As for comparison between the empirical distribution and the noise model, the Google team referred (already in September 2019) for such a comparison to an earlier experiment (on 9 qubits) [15].

6. Regarding timetable and some details for the experiments and calibrations (request (e)), in May 2022 we received a brief timetable and some useful details regarding the final experiments and five earlier ones. See Section 2.3.

7. Regarding item (h), the Google team told us that they would not perform these further computations as they require a lot of human and computational efforts.

   However, using new algorithms, amplitudes for all the verifiable experiments were computed by Kalachev, Panteleev, and Yung [11] and the results give a strong support to Google's 2019 claims and predictions regarding the linear cross entropy of their samples.

8. Regarding item (i), the Google team informed us that they could not share the full proprietary calibration system (that required many years of development) needed to move from data on 1-qubit and 2-qubit to 1-gate and 2-gate corrections in the definition of the circuits. They

pointed out that the main innovation of the calibration program was made public.

9. In July 2022, Adam Zalcman and Sergio Boixo gave us a useful Python program for splitting the patch circuits into the two patches.

10. Regarding item (f) the individual gate- and qubit- fidelities. We asked for it several times in 2019 and early 2020 and again in September 2022.

4.4 The discussion with the Google team

Since October 2019 through August 2022, we had good discussions (initiated by Scott Aaronson), mainly by email, with the Google team and especially with John Martinis and Sergio Boixo, on various aspects of their experiment and on some of our findings. Overall, the Google team welcomed us (and others) in putting their experiment under careful scrutiny, even though they had been aware since 2019 of the first author's concerns [10] about the reliability of the Google 2019 Sycamore paper. The discussions were easy going and in good atmosphere. A video discussion between Boixo, Kalai, Rinott and Shoham in October 2021 was especially fruitful. The basic methodology of trying to pass the data from the Google experiment through various "sanity tests" was overall agreed upon by the Google team although at times we had different interpretations of specific findings. There was a single issue regarding our concerns about the calibration process in connection with the 2019 Google video [9], discussed in late 2021, that led to a somewhat more tense exchange, and in particular where John Martinis criticized attempts to pass judgment on the Google 2019 experiment based on a short video meant for general audience, rather than on the paper itself. Still, also in this case, we had a useful discussion which ultimately shed some light on aspects of the Google 2019 experiment.

Overall, we were not shy to ask for data and information, and in a few cases the Google team was not shy to decline our requests. Most of our requests for data and information were met, although arguably, most of our requests should have been part of the supplementary material of the Google paper to start with.

Since December 2020 until November 2021, we conducted in parallel a useful discussion with Chao-Yang Lu and members of the Gaussian boson sampling team from USTC along with members of the Google team and other researchers regarding the issue of spoofing and k-point correlations of the Gaussian boson sampling experiment. (This discussion was also initiated by Scott Aaronson.) We also had a brief email correspondence with the USTC team that replicated the Google 2019 experiment.

4.5 The discrepancy with the outcomes of Jülich Research Center team

One little mystery that was settled following our discussions with the Google team involved the data from the Jülich team. The amplitudes for 6 large circuits with $39, 42$ and 43 qubits, were computed by researchers from Jülich's research center using their own powerful simulators and (later, after the publication of the paper) when the Google team checked the computation with their own simulators, in some cases the amplitudes were different, and in one case also the estimated fidelity was lower than expected. Initially, the thought was that the (small) differences between the outcomes were caused by numerical differences between the simulators. However, this was not the reason. It turned out that the problem was not a numerical difference between the simulators but rather using a (slightly) different calibration method. (See Section 3.) In the simulations coming from Jülich the researchers used circuits and measurements corresponding to the correct experiment, but the Google team slightly improved the parameters of the circuits a few days after running the experiment, and the Jülich team used an older version. This explains the discrepancy in the amplitudes and why the fidelity estimated with the Jülich simulation was lower for one of the circuits. In other words, the amplitudes were computed for precisely the same experiment, but the calibration, namely the formalization of the experiment as a quantum circuit (the description of the "native" gates) was slightly different.

4.6 Data from other sources

It could have been valuable to perform some of our statistical analysis on data from other quantum computers and we tried without success to obtain similar data (description of calibrated circuits, bitstrings and amplitude files) from other quantum computers especially those from IBM.

Also by now, Google, NASA, and various other groups have powerful simulators that also allow to introduce noise. This provides further opportunity for our statistical analysis, which, so far, we have not used.

4.7 The USTC replication

In June 2021 scientists from USTC [27] described a close replication of the Google 2019 experiment with 56-qubit depth 20 quantum circuits on their *Zuchongzhi* quantum computer. Later in September 2021, they described an improved experiment with 60-qubit depth-24 circuit [30]. The team from USTC shared the description of the circuits (in Matlab) for each $n = 15, 18, 21, 24, ..., 54, 56$, and it seems that they sampled (or shared) less bitstrings than Google did; for example, for $n = 15$ they sampled 200K bit strings, whereas the Google team sampled 500K. Carefully studying the data from these experiments (and perhaps asking for additional data) would be an interesting direction for further research.

4.8 Discussions over scientific blogs

There were useful discussions over a few scientific blogs regarding the Google 2019 experiment, and especially over Aaronson's blog "Shtetl Optimized" (SO) and also over Kalai's blog "Combinatorics and More." For example, in a discussion over SO (December 2020) a commentator "Till" asserted [25] that the calibration process actually adjusted the definition of the circuit to the device. Namely, that the random circuits were generated with "standard gates" but the definition of 2-gates was modified in the calibration process (in the same way for all circuits). (See Section 3.) This was new for us and for several others who thought that the calibration process is a physical process performed on the Sycamore device. Aaronson's own conclusion of the discussion was: "So, my summary would be that yes, there's a calibration phase, and the 2-qubit gates used depend on the outcome of that phase, but there's still a clear separation enforced between the calibration phase and the actual running of the QC."

5 Two proposals for future experiments

5.1 A proposal for blind experiments

In [20] we proposed the following protocol for an evaluation of Google's experiment (here we describe a small variant). First, Google will share the parameters of their calibration. Next, independent scientists will prepare several programs (circuits) for Sycamore to be run with about n qubits, for which computing the sampling probabilities should be a task that takes several months (on a classical computer). These programs will be sent to Google for implementation. Google will send back the implemented programs and large samples that they produce in a short time, which is assumed to preclude computation of the relevant amplitudes of the calibrated circuits. Using classical computers the scientists will take their time and compute the set of amplitudes for each calibrated circuit. They will then evaluate the relation between those amplitudes and the samples they received. Such a protocol is likely to be relevant to other quantum supremacy demonstrations which are being pursued very actively these days. Overall the Google team welcomed the idea of conducting blind experiments in the future and agreed to our specific proposed protocol.

Remark: Over an Internet discussion (on Aaronson's blog, Feb. 2020) Craig Gidney (a member of the Google team) [23] asserted that various experiments for the early pattern that were initially considered computationally hard, turned out to be easier than expected and they served and could further serve as sort of blind tests for the 2019 experiment. The Google team made a similar comment in our email correspondence in January 2021: "We did publish a lot of data that no-one has been able to analyze yet. I hope that eventually some of this data will be analyzed, which will be an interesting confirmation. Analyzing 39 and 40 qubit has actually become easy in the meantime." These remarks motivated our request (h) in Section 4.2.2. As we mentioned, the vast progress in simulation techniques has made it possible to examine the fidelity estimators and they were verified by Kalachev, Panteleev, and Yung, in [11].

5.2 Testing calibration strategies by other groups on Google's Sycamore data

As a follow up to our request for the calibration programs (item (i) in Section 4.2.2) that could not be met, we raised the possibility of sharing raw data gathered from 1-qubit and 2-qubit circuits which formed the input for the calibration programs needed to describe the "native gates". This will give an opportunity to other groups to test their own calibration methods (without revealing Google's full proprietary calibration system). The Google team asserted that this might be possible in principle, but it will require some considerable efforts, and it is not clear if the 2019 data was kept. It might be better and easier to implement this proposal on more recent or even future Sycamore experiments. Given the central place of the calibration stage in NISQ experiments this direction could be valuable in its own right.

6 On the evaluation of the Google 2019 experiment

6.1 Five central questions

As we already mentioned, several ingredients of the Google Sycamore experiment represent major progress in human ability to control quantum systems. When we bring the Sycamore 2019 to test, there are some central problems that come up:

a) Were the sampling tasks achieved as claimed?

In our opinion, the findings of our paper [20] show that the answer is negative. A "sampling task" in the traditional sense, namely reaching (approximately) a sample from some known-in-advanced distribution (or even known-after-the-fact distribution) was not achieved in the 2019 experiment. The empirical distribution is quite different from the Google basic noise model (1) and moving to our more detailed noise model gives only a small improvement. The Google team disagrees with our opinion and notes that they write explicitely in the supplement [4] (around Equation (24)) that they *do not* necessarily assume the basic noise model (1).

b) Are the statistical tools for estimating the fidelity satisfactory?

Our study in [20] shows that given the Google noise model, the $\mathcal{F}_{XEB}$-estimator for the fidelity is quite good. (We offer some improvements mainly when the number of qubits is small.) It is an interesting question if this estimator measures fidelity when we are far away from the Google noise model, see [8].

We note that it might be legitimate to base quantum supremacy claims on the hardness to sample with a high value of $\mathcal{F}_{XEB}$, without referring to the question of whether a specific sampling task was achieved or if $\mathcal{F}_{XEB}$ genuinely estimates the fidelity, see [2, 8].

c) Are the claims regarding the estimated fidelity of the samples valid?

Here, for example, the Google team claimed to achieve for random circuit sampling (full circuits) with $m = 14$, for $n = 12$ with estimated ($\mathcal{F}_{XEB}$) fidelity 0.3694, for $n = 22$ with estimated fidelity 0.165, $n = 32$ with estimated fidelity 0.071. Even leaving aside larger depth and larger numbers of qubits, the question is to what extent we can regard this achievement as solid. We note that the specific computations, regarding the ($\mathcal{F}_{XEB}$) value of the samples given the description of the (calibrated) circuits were verified, and the remaining question is about the quality of the overall claim.

d) Are the claims regarding the predictive power of the a-priori fidelity estimations valid? (Here we refer to Formula (3).)

Here, for example, the a-priori prediction based on Formula (3) (Formula (77) in [4]) for $m = 14$ and $n = 12, 22, 32$ are $0.386, 0.1554, 0.062$ respectively. Question d) is whether we have solid evidence for the claim that the simple formula (77) (and the statistical independence assumptions allowing it) provides accurate prediction of the $\mathcal{F}_{XEB}$ estimator. Here, the specific computations were not checked (the individual fidelities were not shared, see Section 4 item (f)), but they are supported by our approximate computations based on averaged values of the fidelities. As in the previous item, the remaining question is about the quality of the overall claim and here the excellent prediction power of the a priori fidelity estimation raised concerns about it (item 10 in Section 6.2, below).

e) Are the claims regarding the classical difficulty of the sampling task correct?

Regarding Google's "quantum supremacy" claims, a very short summary is that by now classical algorithms are ten orders of magnitude faster than those used in the Google paper and hence the speed-up is ten orders of magnitude lower than Google's fantastic claims. See, e.g., [16, 17, 29, 11, 12, 18]. The Google paper claims that their ultimate task that required 200 seconds for the quantum computer would require 10,000 years on a powerful supercomputer. With the new algorithms the task can be done in a matter of seconds.

A few days after the publication of the Google supremacy paper researchers from IBM [18] exhibited a theoretical improvement of six orders of magnitudes (albeit on a more powerful supercomputer). In response, the Google team pointed out that their paper [3] did anticipate some progress in classical algorithms: "We expect that lower simulation costs than reported here will eventually be achieved, but we also expect that they will be consistently outpaced by hardware improvements on larger quantum processors." As a matter of fact, the Google team welcomed better algorithms and wrote in [3] that the bitstring samples from all circuits have been archived [5] "to encourage development and testing of more advanced verification algorithms."

While weakening Google's supremacy claims, the progress in classical algorithms has also led to major support for Google's claims on items c) and d). In the recent work by Gleb Kalachev, Pavel Panteleev, and Man-Hong Yung, [11] the authors were able to compute the fidelity for samples for which 2019 Google algorithms were too slow to handle. The fidelity values that Gleb, Pavel, and Man-Hong computed agree perfectly with Google's (77) prediction. This gives a very strong support to an affirmative answer to items c) and d). Another major support for both claims c) and d) are the 2021 replications by a group in USTC of the Google 2019 experiments [27, 30]. Yet more support for claims c) and d) came from our own study. In [20] we offered (as a "sanity test") an alternative way to estimate the fidelity based on a "secondary" component of the theoretical distributions that arises from readout errors. In a subsequent work [21] we checked this alternative estimators for the fidelity and found a good match with the $\mathcal{F}_{XEB}$ estimators for $n \leq 30$. This work nicely implements Fourier tools.

We remark that on many items mentioned in this section we carried out statistical analysis that is not presented here, and on most items there is room for more detailed discussion. This is especially the case for the comparison of the empirical distribution with the noise model (6.1a) and the prediction power of equation (77) (6.1d).

Three very concrete questions for circuits with 22 qubits

To be completely concrete let us ask three questions about (verifiable full) random circuit sampling of a circuit C of the kind discussed in the Google paper with $n = 22$ qubits and depth $m = 14$.

1. Can humanity produce at present samples which are good approximation of the Google noise model or any other specific noise model?

2. Did humanity reach the ability to produce samples for quantum circuit C with $\mathcal{F}_{XEB}$ fidelity estimated above 0.15?

3. Did humanity reach the ability to predict, for a quantum circuit C, with good accuracy, the $\mathcal{F}_{XEB}$ fidelity estimator based on the fidelity of the individual components of this circuit?

The findings of our paper [20] indicate that the answer to the first question is negative. The Google supremacy paper and subsequent confirmations present a strong case for a positive answer to the other two questions. But there are remaining doubts and concerns that need to be carefully checked, and not enough replications. (We are aware only of the Google experiment itself and the two USTC replications.)[2]

6.2 Confirmations, refutations, concerns, and weaknesses

Here is a list, without much commentary of weaknesses, concerns, confirmations, and refutations of the Google 2019 quantum supremacy experiment.

A) Confirmations

1. The computations of the fidelity estimators and related computations reported in the Google paper were confirmed by us up to our own computational limits (and several other groups confirmed it well beyond what we did).

2. Major confirmation by Kalachev, Panteleev, and Yung [11] for the Google 2019 paper's experimental claims in the "supremacy region." This gives a strong support to the reliability of the 2019 Google experiment.

3. Our readout errors/Fourier study largely supports (for $n \leq 30$) the experimental claims and the experiment's reliability [21], see Figure 2.

[2]For comparison, after the first Wright brothers flight on December 17, 1903 (120 feet) there were subsequent flights: 852 feet on January 1, 1904; 1,297 feet on February 23 1904; 1,810 feet on May 14, 1904; 4,781 feet on September 23, 1904; 5,360 feet on December 17, 1904; 12,540 feet on January 1, 1905; 24 miles on May 22, 1905; 34 miles on July 4, 1905; 39 miles on October 5, 1905; 56 miles on November 23, 1905, and 63 miles on December 31, 1905. (Source: GPT-3)

4. A successful replication was announced and published by a group from USTC [27, 30].

5. The Google fidelity estimators and statistical methodology are overall good [20].

B) Refutations

6. Google's fantastic supremacy claims were largely (but not fully) refuted [16, 17, 29, 11, 12, 18]. (See Section 6.1 above.)

7. The data does not fit the Google noise model [20] or any other specific model. (See Section 6.1 above.); the Google team disagrees on this point.)

AB) Mixed confirmation/refutation

8. The boson sampling quantum supremacy experiments give a mixed signal. They were regarded as independent confirmation of quantum supremacy using a very different quantum device, however the computational hardness claims are in tension with old (Kalai and Kindler [14]) and new [19, 26] works.

C) Concerns

9. The Google experiment represents amazing leaps in human ability to control quantum devices. The non-gradual advances are surprising and there isn't a clear technological reason behind them.[3]

10. The prediction power for the a priori fidelity estimation (Formula (77) in [4]) raised concerns about reliability. This matter is raised and briefly discussed in [13] and is also mentioned in [20].

There are two important remarks regarding this concern: First, the Google team and other researchers regard this prediction power as one of the main

[3]The Google supremacy experiment was often compared to IBM's "Deep Blue" victory against the chess world champion Kasparov; in that case, however, progress was gradual and earlier chess programs could already win against chess grand-masters.

achievements of the Google experiment. Second, [11] confirmed these predictions for many circuits for which the Google team could not even compute the amplitudes! (This is a sort of "blind test" verification.)

11. When you move away from Google's noise model, the fidelity estimators may not capture the actual fidelity of the quantum system. (See e.g. [8].)

12. Some aspects of the calibration process may seem overly surprising (this represents some preliminary analysis that we presented in our email discussions in November 2021).

13. There are no sufficient replications of the supremacy experiment and verifications of the experimental findings even in the 12-40 qubit range. (Neither by other groups nor by the Google team itself.)

D) Weaknesses

14. While the experiment was characterized as performing a sampling computational task, the empirical distribution was not compared to the noise model, and the experimental samples are too small to allow such comparison (directly) when $n > 14$.

15. The supremacy claims were based on a single algorithm. The Google team itself developed better algorithms to related computational tasks.

    When the Google team found an improved type of algorithms (around May 2019) they switched to a different class of circuits for which the new algorithm did not apply without sufficient evidence that similar improvements could not be made for them as well. (Later, as we mentioned, improvements were also achieved to the new class of circuits.)

16. The results were not presented to the academic community before publication (e.g., as an arXiv publication), and careful review of the experiment prior to publication may have been insufficient.

17. The main fidelity estimator is not unbiased and can be improved [20].

18. Moving to a new pattern **ABCDCDAB** weakened the power of the extrapolation argument. For the new pattern no experiments for circuits with less than 53 qubits were conducted. (This was a weakness in

planning the experiment although some of the fidelity predictions are now confirmed [11].)

19. Improvements of the calibration process were interlaced (namely occurred at the same time) with the experiment. The description of the calibration process in Google's popular video [9] may be in tension with the claim that there was a clear separation between the calibration stage and the experiment itself.

20. The calibration process itself somewhat weakens the claim for a "programmable quantum computer". (This concern was raised by a commentator "Till" in an Internet discussion [25].)

21. The effect of the calibration process is very large (it is not a "fine tuning" of some sort).

22. The precise programs for the calibration process are commercial secrets. (However, the main innovation in calibration was the use of XEB, and that code is open sourced.)

23. The levels of transparency and of documentation of the Google 2019 experiment seem insufficient.

Of course, some of the items above are controversial and also many of these items are related. Items (15), (16) are a-priori weaknesses that turned out to be crucial (refutation (6) above). The recent confirmations (2) indicate that weakness (18) may have no baring on the assertions for large circuits, but we still regard it as a weakness in the planning the experiment. (The Google team disagrees with us.)

We also note that expectations expressed in the Google 2019 paper [3] seem overly optimistic: "Quantum processors have thus reached the regime of quantum supremacy. We expect that their computational power will continue to grow at a double-exponential rate: the classical cost of simulating a quantum circuit increases exponentially with computational volume, and hardware improvements will probably follow a quantum-processor equivalent of Moore's law, doubling this computational volume every few years." (Compare with later analysis in [31, 8].)

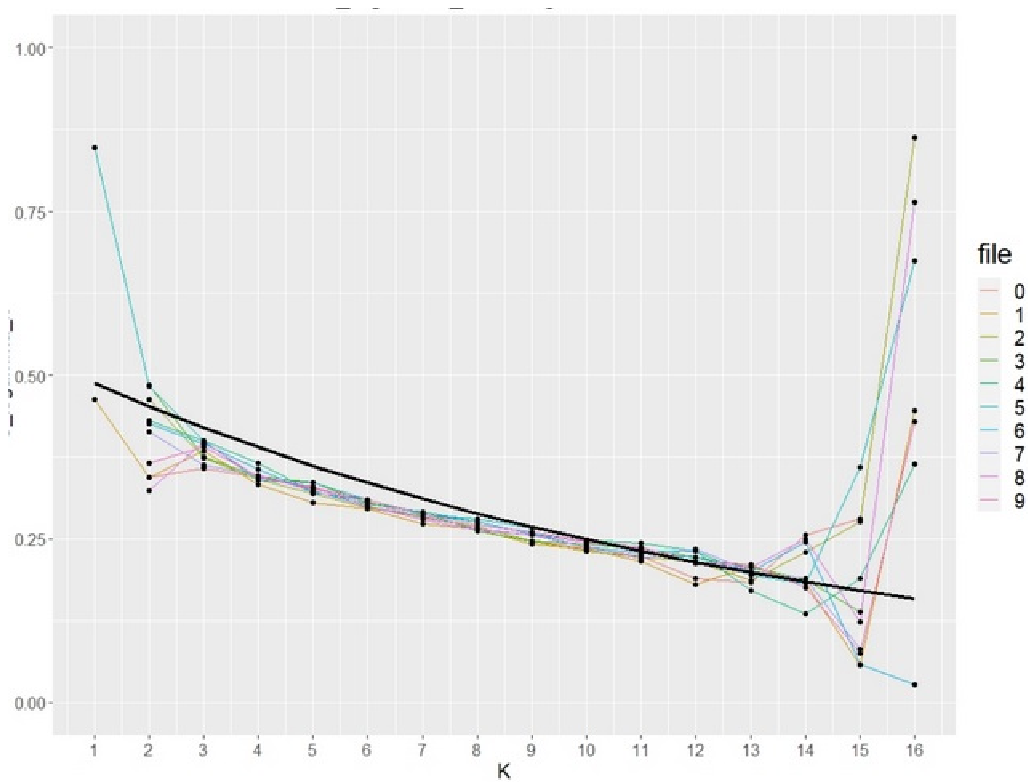

Figure 2: *The decay of Fourier–Walsh contributions as predicted by the theory (bold black curve)([7, 21]) and as demonstrated by samples of a 16-qubit Sycamore computer (ten random circuits).*

7 Where we are (September 2022)

Now that the Google 2019 experiment data-gathering stage needed for our study has largely come to an end, here is a brief summary of where we are.

1. We compared the empirical distribution with Google's noise model and with other noise models. Some of this analysis is discussed in [20], and a few other findings were discussed with the Google team.

2. In [20] we offered an alternative way (that could serve as a "sanity test") to estimate the fidelity based on a linear cross entropy test for a "secondary" component of the theoretical distributions that arises from readout errors. In a subsequent work [21] we used Fourier methods to check this alternative estimator for the fidelity, and found a good match with the primary linear cross entropy fidelity estimators for $n \leq 30$.

3. We carried out some statistical analysis of the effect of the calibration process. Continuing in this direction, we plan, among other things, to study a proposal of John Martinis (made on Aaronson's blog in December 2020 [24], as well as in our discussions) for improving the calibration.

4. We plan to extend some of our statistical studies to the new data from June 2022, and especially to the data for the patch circuits. There is also some remaining analysis to be carried out for the earlier data. We also plan to extend our statistical study to data coming from other NISQ computers and from simulators that incorporate the effect of noise.

5. There is one important data item that is still not publicly available, namely, the fidelities of individual components that are used in Formula 3 (Formula (77) in [4]). This information consists only of a few hundred real numbers and we hope that the Google team will share it in the near future.

References

[1] S. Aaronson, Why Google's quantum supremacy milestone matters, Opinion, *New York Times* Oct. 30, 2019.

[2] S. Aaronson and S. Gunn, On the classical hardness of spoofing linear cross-entropy benchmarking (2019), arXiv:1910.12085.

[3] F. Arute et al., Quantum supremacy using a programmable superconducting processor, *Nature,* 574 (2019), 505–510.

[4] F. Arute et al. (2019), Supplementary information for "Quantum supremacy using a programmable superconducting processor," (2019) arXiv:1910.11333.

[5] F. Arute et al. (2019), Supplementary data for "Quantum supremacy using a programmable superconducting processor," https://datadryad.org/stash/dataset/doi:10.5061/dryad.k6t1rj8.

[6] S. Boixo, S. V. Isakov, V. N. Smelyanskiy, H. Neven, Simulation of low-depth quantum circuits as complex undirected graphical models, arXiv:1712.05384, 2017.

[7] S. Boixo, V. N. Smelyanskiy, and H. Neven, Fourier analysis of sampling from noisy chaotic quantum circuits. arXiv:1708.01875 (2017).

[8] X. Gao, M. Kalinowski, C.-N. Chou, M. D. Lukin, B. Barak, and S. Choi, Limitations of Linear Cross-Entropy as a Measure for Quantum Advantage, 2021, arXiv:2112.01657.

[9] Google, Demonstrating Quantum Supremacy, 4:42 minutes video for a large audience, Oct. 23, 2019. URL: https://youtu.be/-ZNEzzDcllU .

[10] S. Irani (Moderator), Supremacy Panel, Hebrew University of Jerusalem, Dec. 2019. Participants: D. Aharonov, B. Barak, A. Bouland, G. Kalai, S. Aaronson, S. Boixo, and U. Vazirani. (Video.) URL: https://youtu.be/_Yb7uIGBynU .

[11] G. Kalachev, P. Panteleev, and M.-H. Yung, Multi-tensor contraction for XEB verification of quantum circuits, arXiv:2108.05665, 2021.

[12] G. Kalachev, P. Panteleev, P. F. Zhou, and M.-H. Yung, Classical sampling of random quantum circuits with bounded fidelity, arXiv:2112.15083, 2021.

[13] G. Kalai, The argument against Quantum Computers, the quantum laws of nature, and Google's supremacy claims, in: *The Intercontinental Academia Laws: 'Rigidity and Dynamics,'* (M. J. Hannon and E. Z. Rabinovici (ed.)) Proceedings of the ICA Workshops 2018&2019, Singapore and Birmingham, World Scientific, to appear. arXiv:2008.05188.

[14] G. Kalai and G. Kindler, Gaussian noise sensitivity and BosonSampling (2014), arXiv:1409.3093.

[15] C. Neill et al., A blueprint for demonstrating quantum supremacy with superconducting qubits. *Science* 360 (2018), 195-199

[16] F. Pan and P. Zhang, Simulating the Sycamore quantum supremacy circuits. arXiv:2103.03074

[17] F. Pan, K. Chen, and P. Zhang, Solving the sampling problem of the Sycamore quantum supremacy circuits, *Phys. Rev. Lett.* 129 (2022), 090502

[18] E. Pednault, J. A. Gunnels, G. Nannicini, L. Horesh, and R. Wisnieff, Leveraging secondary storage to simulate deep 54-qubit Sycamore circuits (2019), arXiv:1910.09534.

[19] A. S. Popova and A. N. Rubtsov, Cracking the quantum advantage threshold for Gaussian boson sampling, 2021, arXiv:2106.01445.

[20] Y. Rinott, T. Shoham, and G. Kalai, Statistical aspects of the quantum supremacy demonstration, *Statistical Science* 37 (2022), 322–347.

[21] Y. Rinott, T. Shoham, and G. Kalai, Quantum Advantage Demonstrations via Random Circuit Sampling: Fourier Expansion and Statistics, *manuscript in progress.*

[22] *Shtetl Optimized*, A comment by S. Boixo , Dec. 2020. URL: https://scottaaronson.blog/?p=5159#comment-1870651 .

[23] *Shtetl Optimized*, A comment by C. Gidney, , Feb. 2020. URL: https://scottaaronson.blog/?p=4608#comment-1830545 .

[24] *Shtetl Optimized*, A comment by J. Martinis, Dec. 2020. URL: https://scottaaronson.blog/?p=5159#comment-1870734 .

[25] *Shtetl Optimized*, A comment by “Till”, Dec. 2020. URL: https://scottaaronson.blog/?p=5159#comment-1869118 .

[26] B. Villalonga, M. Yuezhen Niu, L. Li, H. Neven, J. C. Platt, V. N. Smelyanskiy, and S. Boixo, Efficient approximation of experimental Gaussian boson sampling, 2021, arXiv:2109.11525.

[27] Y. Wu et al., Strong quantum computational advantage using a superconducting quantum processor. *Phys. Rev. Lett.* 127, 180501.

[28] H.-S. Zhong et al., Quantum computational advantage using photons, *Science* 370 (2020), 1230 1460–1463.

[29] Y. Zhou, E. M. Stoudenmire, and X. Waintal, What limits the simulation of quantum computers? arXiv preprint arXiv:2002.07730 (2020).

[30] Q. Zhu et al., Quantum Computational Advantage via 60-Qubit 24-Cycle Random Circuit Sampling, *Science Bulletin*, Volume 67 (2022), 240-245. arXiv:2109.03494.

[31] A. Zlokapa, S. Boixo, and D. Lidar, Boundaries of quantum supremacy via random circuit sampling, arXiv:2005.02464, 2020.