Quantum supremacy and hardness of estimating output probabilities of quantum circuits

Yasuhiro Kondo, Ryuhei Mori
School of Computing
Tokyo Institute of Technology
Tokyo, Japan
mori@c.titech.ac.jp

Ramis Movassagh
IBM Quantum
MIT-IBM Watson AI Lab
Cambridge MA, U.S.A.
ramis@us.ibm.com

***Abstract*—Motivated by the recent experimental demonstrations of quantum supremacy, proving the hardness of the output of random quantum circuits is an imperative near term goal. We prove under the complexity theoretical assumption of the non-collapse of the polynomial hierarchy that approximating the output probabilities of random quantum circuits to within $\exp(-\Omega(m \log m))$ additive error is hard for any classical computer, where m is the number of gates in the quantum computation. More precisely, we show that the above problem is $\#\mathsf{P}$-hard under $\mathsf{BPP}^{\mathsf{NP}}$ reduction. In the recent experiments, the quantum circuit has n-qubits and the architecture is a two-dimensional grid of size $\sqrt{n} \times \sqrt{n}$ [1]. Indeed for constant depth circuits approximating the output probabilities to within $2^{-\Omega(n \log n)}$ is hard. For circuits of depth $\log n$ or $\sqrt{n}$ for which the anti-concentration property holds, approximating the output probabilities to within $2^{-\Omega(n \log^2 n)}$ and $2^{-\Omega(n^{3/2} \log n)}$ is hard respectively. We then show that the hardness results extend to any open neighborhood of an arbitrary (fixed) circuit including the trivial circuit with identity gates. We made an effort to find the best proofs and proved these results from first principles, which do not use the standard techniques such as the Berlekamp–Welch algorithm, the usual Paturi's lemma, and Rakhmanov's result.**

***Index Terms*—Quantum supremacy; quantum complexity; quantum circuits; random circuit sampling; extended Church-Turing thesis; average-case hardness.**

I. Introduction and related work

Moore's law for classical (super-)computers is reaching a saturation point because if the computation were done with smaller components confined to smaller spaces, then the quantum effects would become relevant. Consequently alternative models and architectures are being investigated to empower the future of computation. Among the many proposals, quantum computing is currently the only model of computation that could potentially exponentially outperform any classical computer. Proving this in affirmative has been a main driving force in the field of quantum computation.

For quantum computers to have the awesome computational power just described the so called Extended Church-Turing Thesis (ECTT) would need to be refuted. ECTT states that a probabilistic Turing machine can efficiently simulate any model of computation that can be realized in Nature (i.e., a realistic computation). A single computational task that would provably refute ECTT would be sufficient. Therefore it is an imperative near-term goal to show that for a given computational task (whatever it may be) a quantum computer can provably outperform any classical computer by running in a time that is exponentially faster. It is then necessary that an actual experiment is performed to demonstrate the separation. Hence to refute the ECTT one needs a solid complexity theoretical foundation and an experimental demonstration. This event (i.e., refutation of ECTT) would be a watershed moment in the history of computation, which would usher the era of quantum supremacy.

It is noteworthy that quantum computation has already demonstrated classical ascendancy for search problems, where Grover's algorithm provably gives a quadratic speed-up over the best possible classical search algorithms [2]. More striking is the Simons problem which proves that on a quantum computer a hidden sub-string can be found exponentially faster than on a classical computer [3]. Its generalization, discovered by Shor, showed that factorization of large composite integers can be done exponentially faster than the best known classical algorithms [4], with significant implications for cybersecurity and cryptography. The inception of quantum computation harks back to Feynman's 1981 paper, which argued that simulation of quantum matter would be exponentially hastened by a quantum computer [5]. Grover's algorithm and Simon's algorithm have provable quadratic and exponential speed up, respectively, against classical algorithms in terms of the query complexity. However, the exponential separation of the power of quantum computers over the classical ones in terms of computational complexity has not been proved to date. In fact, the refutation of ECTT remains a major open problem. What would be a good task that would establish the exponential separation in the near-term?

Modern quantum supremacy proposals are based on the hardness of sampling. The hardness of sampling relies on complexity theoretical assumptions that predate quantum computing. It is known that there is no classical efficient algorithm sampling outputs of the worst-case quantum circuit unless the polynomial hierarchy collapses [6], [7]. Related to the sampling problem, the computation of output probability for the worst-case quantum circuit is classically hard [7], [8]. The first proposal for demonstrating sampling-based quantum supremacy by a near-term quantum computer was the original BosonSampling paper of Aaronson and Arkhipov [9] in which they showed that producing samples from a distribution that mimics the distribution of a linear optical system is classically hard. Later, Bremner et al showed that a class of circuits known as IQP circuits are also classically hard to sample from [10]. The foremost candidate for demonstrating quantum supremacy has been the so-called Random Circuit Sampling (RCS) problem [11], which states that for any classical computer it is hard to produce samples from a distribution that is close to the distribution of a local quantum circuit whose local gates are randomly and independently are drawn uniformly from the space of all possible gates.

Demonstration of quantum supremacy is ultimately given by an experiment for which there is solid complexity theoretical evidence of hardness of the task at hand. Indeed Google did an experiment that involved a random circuit with 53 qubits to demonstrate the hardness of RCS [1]. Soon after new classical algorithms emerged that challenged the claim [12]–[14]. It remains an open problem to mathematically prove the hardness of sampling. A fruitful approach is to prove the hardness of sampling by proving the hardness of approximating probability amplitudes of the quantum circuit. In particular, if the probability amplitudes obey an anti-concentration property [15], [16], then one can use Stockmeyer's algorithm [17] to prove that it is sufficient to prove that the amplitudes are hard to approximate to within $2^{-n}/\mathrm{poly}(n)$ additive error.

The first theoretical evidence for the hardness of computing the output probabilities was given by Bouland et al [18], who showed that the computation of the amplitudes of a *non-unitary* approximation of the actual quantum circuit is hard unless the polynomial hierarchy collapses. In [19] the Cayley path was introduced, which is a unitary matrix-valued path. It was shown that the exact probability amplitudes of the (i.e., unitary) random quantum circuit is #P-hard, and that even approximating the amplitudes to within 2^{-m^c}, where c is a quantified constant and m is the number of gates, remains #P-hard. The validity of the hardness with respect to additive error approximation is referred to as *robustness*.

A. *Summary of this work*

In this work we substantially (super-polynomially) improve the robustness to $2^{-\Omega(m\log m)}$, where m is the number of gates. Therefore, our result proves that approximating the probability amplitudes to within $2^{-\Omega(n\log n)}$ is hard for constant depth circuits. In order to use Stockmeyer's algorithm to prove hardness of sampling from the hardness of approximating probability amplitudes with respect to additive errors, one needs a further property of anti-concentration. This property has been proved for circuits of depths $\log n$ [16] and $\sqrt{n}$ [15]. Our robustness bound for circuits of depth $\log n$ and $\sqrt{n}$ is $2^{-\Omega\left(n\log^2 n\right)}$ and $2^{-\Omega(n^{3/2}\log n)}$ respectively.

In proving this result, we took an entirely a new approach that does not follow the standard techniques of the past [9], [18], [20]. Instead of using standard paths such as $\theta X + (1-\theta)Y$ in BosonSampling [9], or the truncation of the Taylor series used in [18], we rely on the Cayley path introduced in [19]. Furthermore, instead of Paturi's lemma [21], which has become standard in the field for bounding the polynomial extrapolation errors [9], [18], and Rakhmanov's result [22], which is usually used to extend error bounds on a disrete set of points to a uniform error bound in a region [18], we use the Lagrange polynomials for estimating the error of polynomial extrapolation. In fact we do not use the well-known Berlekamp–Welch algorithm [23], which has instabilities in the presence of uniform noise.

We prove two theorems for the hardness of RCS that complement one another by using oracles of different strengths yet requiring different success probabilities (Theorems 1 and 2). We prove our hardness results from first principles and hope that the new approach helps to overcome the insurmountable difficulties that the standard techniques meet. Our main results are the following two theorems:

Theorem 1 (Simplified)**.** *It is* #P*-hard under* BPP*-reduction to approximate* $|\langle 0|C|0\rangle|^2$ *to within the additive error* $2^{-\Omega(m\log m)}$ *for* $1-O(1/m)$ *fraction of quantum circuits* C*.*

Theorem 2 (Simplified)**.** *It is* #P*-hard under* $\mathsf{BPP}^{\mathsf{NP}}$*-reduction to approximate* $|\langle 0|C|0\rangle|^2$ *to within the additive error* $2^{-\Omega(m\log m)}$ *for* $\frac{3}{4}+\frac{1}{\mathrm{poly}(n)}$ *fraction of quantum circuits* C*.*

In other word, unless the polynomial hierarchy collapses to finite level, the above tasks are outside the polynomial hierarchy. We summarize the proof structures that culminate in Theorems 1 and 2 in Figures 1 and 2 respectively.

We then show that the Cayley path can be utilized in much the same way to prove that any circuit with

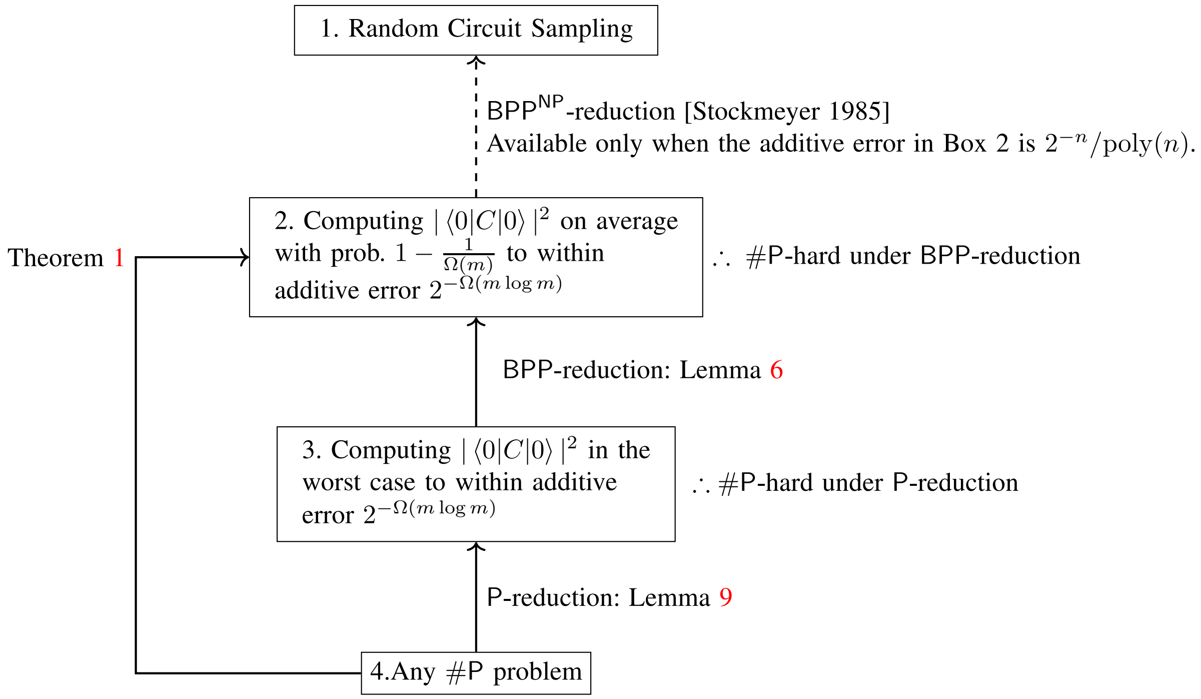
1. Random Circuit Sampling
BPP$^{\mathsf{NP}}$-reduction [Stockmeyer 1985]
Available only when the additive error in Box 2 is $2^{-n}/\mathrm{poly}(n)$.
Theorem 1
2. Computing $|\langle 0|C|0\rangle|^2$ on average with prob. $1-\frac{1}{\Omega(m)}$ to within additive error $2^{-\Omega(m\log m)}$
$\therefore$ #P-hard under BPP-reduction
BPP-reduction: Lemma 6
3. Computing $|\langle 0|C|0\rangle|^2$ in the worst case to within additive error $2^{-\Omega(m\log m)}$
$\therefore$ #P-hard under P-reduction
P-reduction: Lemma 9
4.Any #P problem

Fig. 1: The proof structure and reductions for Theorem 1

the same architecture as the worst-case circuit also has the same hardness properties as shown above. A perhaps surprising corollary is that sampling from circuits close to identity as $\#P$-hard.

Theorem 3 (Simplified)**.** *Theorems 1 and 2 hold in the case that the circuit C is within any open neighborhood of a fixed circuit. This also applies to the trivial circuit $C = I$ with identity gates (see Corollary 2.)*

We remark that Bouland et al [24] claim the same robustness as our Theorem 2 and arrived at it independently. Our respective papers are different in details. We use Lagrange interpolation for bounding the errors induced by polynomial extrapolation, which gives simple and direct proofs. More complicated techniques were needed in [24] for bounding the errors due to polynomial extrapolation, which they call Lagrang*ian* interpolation.

B. *Open problems and future work*

Our results can be applied to BosonSampling. We hope to see extension and application of these techniques to the complexity of BosonSampling [25] for which there was a recent experimental breakthrough [26].

Applications of the Cayley path for randomizing quantum gates have found use in other contexts [27] and it would be interesting to see fresh new applications.

The main open problem is to prove the hardness of sampling for random quantum circuits. In particular, improving our additive error robustness bounds to $2^{-n}/\mathrm{poly}(n)$ for random circuits that have the anti-concentration property would be sufficient. The overarching goal of proving the quantum supremacy conjecture is achieved (i.e., Corollary 2 below is proved) if the following conjecture is proved in the affirmative:

Conjecture 1. *Approximating $|\langle 0|C|0\rangle|^2$ with probability of $\frac{3}{4}+\frac{1}{\mathrm{poly}(n)}$ over the choice of quantum circuits C to within the additive error $2^{-n}/\mathrm{poly}(n)$ implies the collapse of the polynomial hierarchy to a finite level.*

Assuming Conjecture 1 and anti-concentration of output probabilities, we obtain the hardness of RCS via Stockmeyer's theorem.

Theorem 4 (Stockmeyer [17])**.** *Given a Boolean function $f:\{0,1\}^n \to \{0,1\}$, let*

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

Then there exists an $\mathsf{FBPP}^{\mathsf{NP}^f}$ *machine that approximates p to within any multiplicative factor of* $1+1/\mathrm{poly}(n)$.

In conclusion, Theorem 4 along with the anti-concentration property of the output probabilities [15], and Conjecture 1 prove the following major open problem:

Conjecture 2. *Classically sampling from any distribution with a total variation distance of* $1/\mathrm{poly}(n)$ *from*

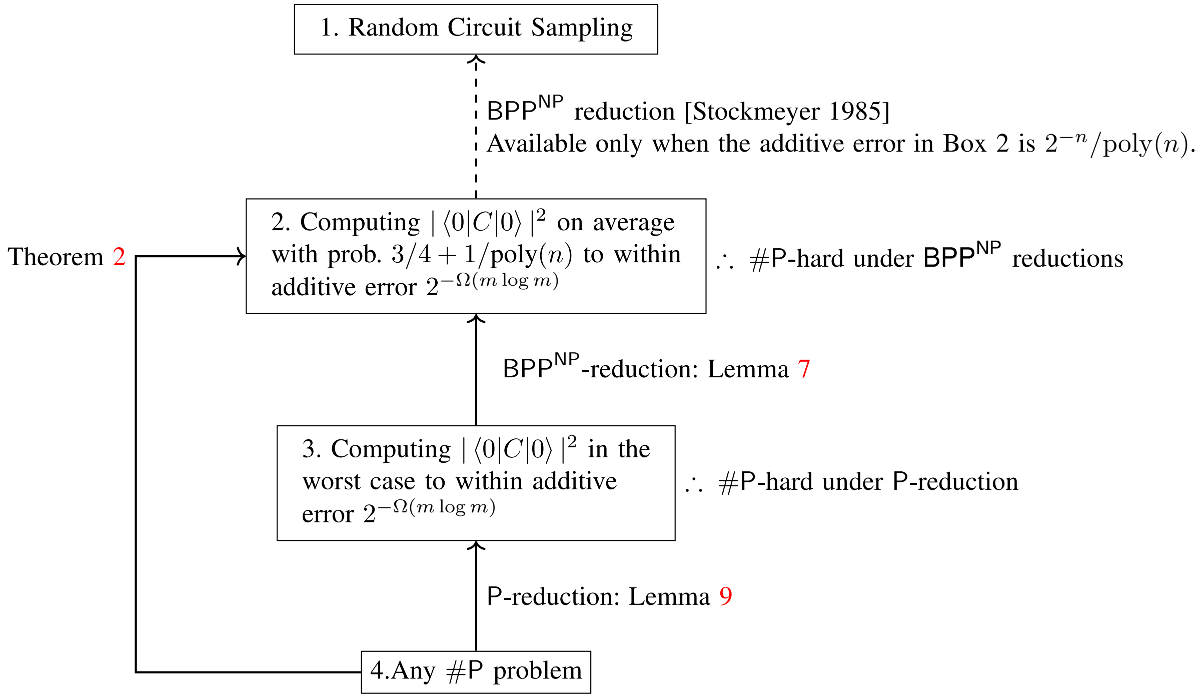
1. Random Circuit Sampling
BPP^NP reduction [Stockmeyer 1985]
Available only when the additive error in Box 2 is $2^{-n}/\mathrm{poly}(n)$.
Theorem 2
2. Computing $|\langle 0|C|0\rangle|^2$ on average with prob. $3/4+1/\mathrm{poly}(n)$ to within additive error $2^{-\Omega(m\log m)}$
∴ #P-hard under BPP^NP reductions
BPP^NP-reduction: Lemma 7
3. Computing $|\langle 0|C|0\rangle|^2$ in the worst case to within additive error $2^{-\Omega(m\log m)}$
∴ #P-hard under P-reduction
P-reduction: Lemma 9
4.Any #P problem

Fig. 2: The proof structure and reductions for Theorem 2

the output distribution of the random quantum circuit is hard unless the polynomial hierarchy collapses to finite level.

II. Cayley Path

We now define the interpolation between any two gates of the quantum computation (i.e., any two unitaries) based on the Cayley path, which was first introduced in [19]. Suppose $U_0, U_1 \in \mathbb{U}(N)$ are unitary matrices and we wish to interpolate between U_0 and U_1 via a path with nice algebraic properties that can be utilized in our reductions below. Let $\theta \in \mathbb{R}$ and $f(\theta)$ be the *Cayley function*

$$f(\theta) = \frac{1+i\theta}{1-i\theta}\,, \tag{1}$$

where one defines $f(-\infty) = -1$. The Cayley function as just defined is a bijection between $\mathbb{R}\cup\{-\infty\}$ and the unit circle in the complex plane.

The proposed path is

$$U(\theta) = U_0\, f(\theta h) = \sum_{\alpha=1}^{N} f(\theta h_\alpha)\, U_0 |\psi_\alpha\rangle\langle\psi_\alpha|\,, \tag{2}$$

where h is a hermitian matrix defined by $h = f^{-1}\left(U_0^\dagger U_1\right)$, which is guaranteed to exist by the bijection property. $U(\theta)$ is a unitary matrix as it is a product of two unitary matrices. Note that $U(0) = U_0 f(0) = U_0$ and $U(1) = U_0 U_0^\dagger U_1 = U_1$ as desired. We now derive the algebraic dependence of the entries of $U(\theta)$ on θ.

Let the spectral decomposition of h be $h = \sum_{\alpha=1}^{N} h_\alpha |\psi_\alpha\rangle\langle\psi_\alpha|$. Using the definition of the Cayley function and the foregoing equations we write

$$U(\theta) = \frac{1}{q(\theta)} \sum_{\alpha=1}^{N} p_\alpha(\theta)\; U_0\; |\psi_\alpha\rangle\langle\psi_\alpha|\,, \tag{3}$$

where $q(\theta)$ and $p_\alpha(\theta)$ are univariate scalar complex polynomials of degree N in θ:

$$\begin{aligned} q(\theta) &= \prod_{\alpha=1}^{N}(1 - i\theta h_\alpha) \\ p_\alpha(\theta) &= (1 + i\theta h_\alpha) \prod_{\beta\in[N]\setminus\alpha} (1 - i\theta h_\beta). \end{aligned} \tag{4}$$

In what follows we wish to apply the Cayley path to each one of the gates in the quantum circuit. For this we first need to formally define the architecture of the circuit.

Definition 1 (Architecture)**.** The architecture $\mathcal{A}$ is a poly-time uniform family $\{A_\ell\}_{\ell=1,2,\ldots}$ of quantum circuit where all quantum gates are "blank" and not specified. The quantum circuit A_ℓ has $n(\ell)$ qubits and $m(\ell)$ quantum gates. If all quantum gates of A_ℓ are specified, then the quantum circuit is specified. We say an architecture is *local* if each gate acts on at most two qubits.

In this paper all architectures are assumed to be local.

Definition 2. Let $\mathcal{H}_{\mathcal{A}}$ be the distribution over circuits with architecture $\mathcal{A}$ with an implicitly fixed ℓ (see Definition 1), whose local gates are unitary matrices drawn independently and at random from the Haar measure.

Given a fixed architecture and a quantum circuit C whose k-th gate is C_k, we consider the randomized quantum circuit by replacing all quantum gates of C with quantum gates drawn from the Haar measure. Then, we consider a Cayley path interpolation between each gate of the fixed circuit and the randomized one.

Let $C(\theta)$ denote the quantum circuit with m gates:

$$C(\theta) = \mathcal{C}_m(\theta) \cdots \mathcal{C}_2(\theta)\, \mathcal{C}_1(\theta)\,, \tag{5}$$

where $\mathcal{C}_k(\theta) = \mathbb{I} \otimes C_k(\theta)$ is a unitary matrix that only acts non-trivially on the qubits that $C_k(\theta)$ acts on. Here each local unitary gate is a unitary-valued Cayley path $C_k(\theta)$:

$$C_k(\theta) = C_k\; f(\theta h_k)\ , \tag{6}$$

where $f(\theta h_k)$ is a unitary matrix and h_k is hermitian $h_k^\dagger = h_k$. Suppose C_k is a fixed gate of a quantum computation and $H_k \equiv f(h_k)$ is a Haar unitary matrix then $C_k(0) = C_k$. Moreover, by the translation invariance of the Haar measure $C_k(1) = C_k H_k$ is a Haar random gate. Hence we have an interpolation scheme between any fixed gate and a Haar random gate.

Definition 3. Let us denote by $\mathcal{H}_{\mathcal{A},\Delta}$ the distribution over circuits whose local gates are drawn from the distribution induced by the Cayley path for $\theta = 1 - \Delta$ for $\Delta \in [0, 1]$.

The randomness of the quantum circuit under the Cayley path is quantified in this lemma:

Lemma 1 (Total Variation Distance [19])**.** *For a circuit with m gates and an architecture $\mathcal{A}$, the total variation distance between $\mathcal{H}_{\mathcal{A}}$ and $\mathcal{H}_{\mathcal{A},\Delta}$ is $O(m\Delta)$.*

We make the dependence on k explicit in Eqs. (3) and (4) by denoting $p_\alpha(\theta) \mapsto p_{k,\alpha}(\theta)$ and $q(\theta) \mapsto q_k(\theta)$. We can now express Eq. (6) as

$$C_k(\theta) = \frac{1}{q_k(\theta)} \sum_{\alpha=1}^{N} p_{k,\alpha}(\theta)\; C_k |\psi_{k,\alpha}\rangle\langle\psi_{k,\alpha}| \tag{7}$$

where $q_k(\theta) = \prod_{\alpha=1}^{N}(1 - i\theta h_{k,\alpha})$

$$p_{k,\alpha}(\theta) = \sum_{\alpha=1}^{N}(1 + i\theta h_{k,\alpha}) \prod_{\beta\in[N]\setminus\alpha} (1 - i\theta h_{k,\beta}). \tag{8}$$

It will be useful to make a change of variables to $\theta = 1 - x$ such that generic instances correspond to $x = 0$ and #P-hard point to $x = +1$.

The probability amplitude of starting the quantum computation in the state $|0^n\rangle$ and measuring the string $|0^n\rangle$ is $p_0(x) \equiv |\langle 0^n|C(x)|0^n\rangle|^2$. Note that at $x = 1$ we recover the worst case #P-hard instance probability amplitude and $x = 0$ corresponds to the probability amplitude of the generic random circuit. Using $\theta = 1-x$ in Eqs. (7)-(8), the circuit has the algebraic form [19]

$$\begin{aligned} |\langle 0^n|C(x)|0^n\rangle|^2 &= \left|\langle 0^n| \prod_{k=1}^{m} \mathcal{C}_k(x)|0^n\rangle\right|^2 \\ &\equiv \frac{|\langle 0^n|P(x)|0^n\rangle|^2}{|Q(x)|^2}, \end{aligned} \tag{9}$$

where

$$P(x) \equiv \sum_{\alpha_1,\ldots,\alpha_m=1}^{N} \prod_{k=1}^{m} g_{k,\alpha_k}(x)\; \mathcal{C}_k\,(x=1)\, |\psi_{k,\alpha_k}\rangle\langle\psi_{k,\alpha_k}| \tag{10}$$

and

$$|Q(x)|^2 \equiv \prod_{k=1}^{m} \prod_{\alpha_k=1}^{N} \left| 1 + ix\frac{h_{k,\alpha_k}}{r_{k,\alpha_k}} e^{iu_{k,\alpha_k}} \right|^2 \tag{11}$$

$$\begin{aligned} g_{k,\alpha_k}(x) \equiv & \left[e^{iu_{k,\alpha_k}} - ix\frac{h_{k,\alpha_k}}{r_{k,\alpha_k}} \right] \\ & \times \prod_{\beta_k\in[N]\setminus\alpha_k} \left[e^{-iu_{k,\beta_k}} + ix\frac{h_{k,\beta_k}}{r_{k,\beta_k}} \right]. \end{aligned}$$

Here, we let $1 \pm ih_{k,\alpha_k} = r_{k,\alpha_k} e^{\pm iu_{k,\alpha_k}}$ with r_{k,α_k} and u_{k,α_k} defined as $r_{k,\alpha_k} = \sqrt{1 + h_{k,\alpha_k}^2}$ and $u_{k,\alpha_k} = \arctan(h_{k,\alpha_k})$.

The quantity $|Q(x)|^2$ can be pre-computed in time $\Theta(m)$ as it only depends on the eigenvalues of the local terms which are matrices of size at most $N = 4$. Since $\frac{h_{k,\alpha_k}}{r_{k,\alpha_k}} < 1$ and for generic circuits $|x| \le \Delta = O(m^{-1})$, it is easily seen that $|Q(x)|^2$ is very near one:

$$\begin{aligned} |Q(x)|^2 &\le \prod_{k=1}^{m} \prod_{\alpha_k=1}^{N} \left| 1 + ix\frac{h_{k,\alpha_k}}{r_{k,\alpha_k}} e^{iu_{k,\alpha_k}} \right|^2 \\ &\le 1 + O(m\Delta)\ . \end{aligned} \tag{12}$$

III. Proof of average-case robustness

Our goal here is to prove that *approximating* $p_0(x) \equiv |\langle 0^n|C(x)|0^n\rangle|^2$ to within ϵ additive error is *hard* for as large an ϵ as possible. That is given an x_i and a classical algorithm that promises to give us $p_0(x_i) + \epsilon_i$ efficiently (polynomial classical time), where $|\epsilon_i| \le \epsilon \ll 1$, we wish to construct a low degree algebraic function $\tilde{p}(x)$ whose extrapolation to $x = 1$ is guaranteed to be *hard*.

Since for any x, $|Q(x)|^2$ can be computed in time $\Theta(m)$ we can reduce the rational functional form of $p_0(x)$ to a polynomial by multiplying through by $|Q(x)|^2$, which for any given x can be treated as simply a constant. Let us denote by the "exact" polynomial

$$p_e(x) = |\langle 0^n|P(x)|0^n\rangle|^2 = |Q(x)|^2 p_0(x) \tag{13}$$

of degree $8m$ where we treated $Q(x)$ as a known constant. Therefore, we have at our disposal a set of tuples $(x_i, p_e(x_i) + \epsilon_i |Q(x_i)|^2)$. In Eq. (12) we showed that $|Q(x)|^2 \leq 1 + O(m\Delta)$, and by taking $\Delta = O(m^{-1})$ we are guaranteed to have $|Q(x)|^2 \approx 1$. This shows that the additive error $|Q(x_i)|^2 \epsilon_i \approx \epsilon_i$.

Let the difference of the exact polynomial $p_e(x)$ from the one that results from the extrapolation of the erroneous polynomial $\tilde{p}(x)$ be defined by

$$p(x) \equiv \tilde{p}(x) - p_e(x).$$

We are promised that for all $|x_i| \leq \Delta$, $|p(x_i)| \leq \epsilon$ and wish to show that $|p(1)|$ is sufficiently small such that it falls within a region whose hardness is guaranteed. We will return to the quantification of this region in Section IV. For now let us bound the polynomial extrapolation error $|p(1)|$.

The following Lemmas 2 and 3 and Corollary 1 on this page are presented to show that one may obtain robustness beyond what the traditional Paturi's lemma allows. Nevertheless this approach will face other difficulties. Hence, starting in Lemma 4 we prove the results independent of this approach.

Traditionally this bound is obtained using Paturi's lemma [21], which we recall:

Lemma 2. *[Paturi's lemma [21]] Let $p(x)$ be a polynomial of degree d, and suppose $|p(x)| \leq \epsilon$ for $|x| \leq \Delta$ where $\Delta \in (0,1)$. Then $p(1) \leq \epsilon \exp[2d(1 + \Delta^{-1})]$*

For $k \geq 0$, let us denote by $T_k(x)$ the k^{th} Chebyshev polynomial, which is a degree k algebraic polynomial defined by

$$T_k(x) = \frac{1}{2}\left[(x + \sqrt{x^2 - 1})^k + (x - \sqrt{x^2 - 1})^k\right]$$

for $x > 1$. Paturi has another result in the same paper (Corollary 2 in [21]), which says

Corollary 1. *[Paturi's Corollary [21]] Let $p(x)$ be a polynomial of degree at most d. Assume $|p(x)| \leq \epsilon$ in the interval $[-\Delta, \Delta]$ for some $0 < \Delta \leq 1$. We then have $|p(x)| \leq \epsilon \left|T_d(1 + \frac{|x| - \Delta}{\Delta})\right|$ for all $|x| \geq \Delta$ where T_d denotes the Chebyshev polynomial of degree d.*

We shall use the latter and prove the following lemma

Lemma 3. *Let $p(x)$ be a polynomial of degree d, and suppose $|p(x)| \leq \epsilon$ for $|x| \leq \Delta$ where $\Delta \in (0,1)$. Then*

$$|p(1)| < \epsilon\ \exp(d \log|2\Delta^{-1}|).$$

Proof. From Paturi's corollary we have $|p(1)| \leq \epsilon |T_d(\Delta^{-1})|$. Moreover,

$$\begin{aligned}|T_d(x)| &\leq \frac{1}{2}\left[\left|(x + \sqrt{x^2 - 1})^d\right| + \left|(x - \sqrt{x^2 - 1})^d\right|\right] \\ &< \left(|x| + \sqrt{x^2 - 1}\right)^d < |2x|^d = e^{d\log(2|x|)}.\end{aligned}$$

We conclude that $|p(1)| < \epsilon \exp\left[d\log(2\Delta^{-1})\right] = \epsilon \exp\left[-(d\log\Delta)\left(1 - \frac{1}{\log_2\Delta}\right)\right]$. □

An issue one faces in making robustness claims is that the discrete bound of $|p(x_i)| \leq \epsilon$ for $|x_i| \leq \epsilon$, does not readily imply a uniform bound $|p(x)| \leq \epsilon$ for all $|x| \leq \Delta$. This is traditionally remedied by Rakhmanov's result [22] (see [18], [19]). Here we will do without Rakhmanov's result. To do so, take $d + 1$ points in the interval $[-\Delta, \Delta]$ and estimate the function $p(x)$ using the Lagrange interpolation technique. We prove

Lemma 4. *Let $p(x)$ be a polynomial of degree at most d. Let $\Delta \in (0,1)$. Assume that $|p(x_j)| \leq \epsilon$ for all of the $d+1$ equally-spaced points $x_j = -\Delta + \frac{2j}{d}\Delta$ for $j = 0, 1, \ldots, d$. Then*

$$|p(1)| < \epsilon\,\frac{\exp\left[d(1 + \log\Delta^{-1})\right]}{\sqrt{2\pi d}}\,. \tag{14}$$

Proof. Let $p_j = p(x_j)$ for all $j = \{0, 1, 2, \ldots, d\}$, where by assumption $|p_j| \leq \epsilon$. The Lagrange representation of the function $p(x)$ writes

$$p(x) = \sum_{j=0}^{d} p_j\ \delta_j(x), \qquad \delta_j(x) \equiv \frac{\prod_{\ell \neq j} x - x_\ell}{\prod_{\ell \neq j} x_j - x_\ell}\,.$$

By triangular inequality we have $|p(1)| \leq \epsilon \sum_{j=0}^{d} |\delta_j(1)|$. Moreover, using $x_j = -x_{d-j}$ and the fact that $|x_j| < 1$ for all j we have

$$\begin{aligned}|\delta_j(1)| &= \frac{\prod_{\ell \neq j} |1 - x_\ell|}{\prod_{\ell \neq j} |x_j - x_\ell|} \\ &= \frac{(1 + x_j)\prod_{\ell \in \{0,1,\ldots,\lfloor (d-1)/2 \rfloor\}\setminus\{j,\, d-j\}}(1 - x_\ell^2)}{\prod_{\ell \neq j} |x_j - x_\ell|} \\ &< \frac{(1 + x_j)}{\prod_{\ell \neq j} |x_j - x_\ell|}.\end{aligned}$$

Since $x_j - x_\ell = \frac{2\Delta}{d}(j - \ell)$, we have $\prod_{\ell \neq j} |x_j - x_\ell| = (\frac{2\Delta}{d})^d \prod_{\ell \neq j} |j - \ell|$. Moreover $\prod_{\ell \neq j} |j - \ell| = \prod_{\ell \in \{0,1,2,\ldots,,j-1,j+1,\ldots,d\}} |j - \ell| = j!(d-j)!$ and we obtain

$$|\delta_j(1)| = \left(\frac{d}{2\Delta}\right)^d \frac{(1 + x_j)}{\prod_{\ell \neq j} |j - \ell|} = \left(\frac{d}{2\Delta}\right)^d \frac{(1 + x_j)}{j!\,(d-j)!}\,.$$

We express the bound $|p(1)| \leq \epsilon \sum_{j=0}^{d} |\delta_j(1)|$ as

$$\begin{aligned}|p(1)| &< \epsilon \left(\frac{d}{2\Delta}\right)^d \sum_{j=0}^{d} \frac{(1 + x_j)}{j!\,(d-j)!} \\ &= \epsilon \left(\frac{d}{2\Delta}\right)^d \frac{1}{d!} \sum_{j=0}^{d} \binom{d}{j} (1 + x_j).\end{aligned}$$

Using the symmetry of $x_j = -x_{d-j}$ we have $(1+x_j) + (1+x_{d-j}) = 2$ and it is easy to see that (irrespective of the parity of d)

$$\sum_{j=0}^{d} \binom{d}{j} (1+x_j) = 2^d.$$

By Stirling's inequality $n! \geq \sqrt{2\pi n}\frac{n^n}{e^n}$, we conclude that

$$|p(1)| < \epsilon \left(\frac{d}{2\Delta}\right)^d \frac{2^d}{d!} \leq \epsilon \frac{(e\Delta^{-1})^d}{\sqrt{2\pi d}}$$
$$= \epsilon \frac{\exp\left[d(1+\log\Delta^{-1})\right]}{\sqrt{2\pi d}}.$$

□

Remark 1. Note that the choice of the equally-spaced $d+1$ points is not exactly optimal. We choose the $d+1$ points $\{\Delta\cos(0), \Delta\cos(\pi/d), \Delta\cos(2\pi/d), \ldots, \Delta\cos((d-1)\pi/d), \Delta\cos(\pi)\}$, which are the extrema of the Chebyshev polynomials $T_d(x/\Delta)$. Then the worst-case polynomial is the Chebyshev polynomial $\epsilon T_d(x/\Delta)$, and the same bound as in Lemma 3 is obtained. The issue with Lemma 3 remains to be that it either assumes a uniform bound or requires using Chebyshev extrema as just described. To prove Theorem 2 we need Lemma 5. This lemma does not give the freedom to choose the points exactly. Therefore, we cannot use Chebyshev extrema at will that saturate the Paturi's Lemma (Lemma 3)

Similarly, we obtain the following lemma where $d+1$ points are chosen from L equally-spaced points in $[-\Delta, \Delta]$.

Lemma 5. *Let $p(x)$ be a polynomial of degree at most d, and L an integer at least $d+1$. Let $a_0, a_1, \ldots, a_d$ be integers satisfying $0 \leq a_0 < a_1 < \cdots < a_d \leq L-1$. Let $\Delta \in (0,1)$. Assume that $|p(x_j)| \leq \epsilon$ for all of the $d+1$ points $x_j = -\Delta + \frac{2a_j}{L-1}\Delta$ for $j = 0, 1, 2, \ldots, d$. Then*

$$|p(1)| \leq \epsilon \frac{\exp[d(1+\log((1+\Delta^{-1})\frac{L-1}{d}))]}{\sqrt{2\pi d}}.$$

The proof of Lemma 5 is similar to the proof of Lemma 4, and is presented in Appendix A.

Armed with the new extrapolation error bound, we proceed to prove the hardness of evaluating the probability amplitudes of general circuits with an additive error. In the following, we show the reductions from the worst-case to the average-case computation of the output probability of the quantum circuit. While real numbers appear in the reduction algorithms, they should be represented by $\mathrm{poly}(m)$ bits. The rounding only causes additional errors of size $2^{-\mathrm{poly}(m)}$, which will not affect our bounds and conclusions. As in [9] and for simplicity, we ignore rounding issues in the proofs, and work with real numbers.

Lemma 6 (Strong oracle)**.** *Let $\delta > 0$ be a constant and $\mathcal{O}$ a classical oracle that takes as input the classical description of the quantum circuit C in the architecture $\mathcal{A}$ and outputs $\mathcal{O}(C)$ that satisfies*

$$\Pr_{C\sim\mathcal{H}_\mathcal{A}}\left[\,|\mathcal{O}(C) - |\langle 0^n|C|0^n\rangle|^2\,| \leq \epsilon\,\right] \geq 1 - \frac{\delta}{8m+1}.$$

Then there exists a classical probabilistic polynomial-time algorithm $\mathcal{R}$ with access to $\mathcal{O}$ that outputs $\mathcal{R}^{\mathcal{O}}(C)$ satisfying:

$$\Pr_{\mathcal{R}}\left[\,|\mathcal{R}^{\mathcal{O}}(C) - |\langle 0^n|C|0^n\rangle|^2\,| < \epsilon\exp\left[O(m\log m)\right]\,\right]$$
$$\geq 1 - \delta - \frac{1}{\mathrm{poly}(m)}.$$

Proof. Previously we proved that the total variation distance between $\mathcal{H}_\mathcal{A}$ and $\mathcal{H}_{\mathcal{A},\Delta}$ is $O(m\Delta)$ (see Lemma 1). Hence invoking the oracle $\mathcal{O}$ it holds that

$$\Pr_{C\sim\mathcal{H}_{\mathcal{A},\Delta}}\left[\,|\mathcal{O}(C) - |\langle 0^n|C|0^n\rangle|^2\,| > \epsilon\,\right]$$
$$\leq \frac{\delta}{8m+1} + O(m\Delta).$$

By the union bound, the probability that at least one of the $8m+1$ points evaluated by $\mathcal{O}$ has error larger than ϵ is at most $\delta + (8m+1)O(m\Delta)$. By choosing $\Delta = \Theta(m^{-k})$ for some constant $k > 2$, from Lemma 1, the error probability is at most $\delta + 1/\mathrm{poly}(m)$. From Lemma 4 we know that if all $8m+1$ evaluation points have an error at most ϵ, then the extrapolation error from $x \in [-\Delta, \Delta]$ to $x = 1$ in the Lagrange extrapolation is given (via Eqs. (12) and (14))

$$\epsilon\,|Q(1+\Delta)|^2 \frac{\exp\left[8m(1+\log\Delta^{-1})\right]}{\sqrt{16\pi m}}$$
$$\leq \epsilon\,(1+O(m\Delta))\frac{\exp\left[8m(1+\log\Delta^{-1})\right]}{\sqrt{16\pi m}}.$$

Since $\Delta = \Theta(m^{-k})$ for some constant $k > 2$, we obtain the Lemma. □

It is desirable to make the oracle $\mathcal{O}$ as weak as possible. In the following Lemma we show that this can be done at the expense of introducing an NP-machine.

The high level idea for the following lemma is that if we know the points x_i for which the additive error committed by the classical oracle O is sufficiently small then we can apply Lemma 5 to do Lagrange extrapolation. However, to find those points we need to call an NP-oracle. Then we call the classical oracle O that succeeds in approximating $|\langle 0^n|C|0^n\rangle|^2$ with probability at least $\frac{3}{4} + \delta$ over the random choice of quantum circuits.

Lemma 7 (Weak oracle)**.** *Let $\delta > 0$ be a constant and $\mathcal{O}$ a classical oracle that takes as input the classical*

Algorithm 1 The reduction algorithm $\mathcal{R}$ where $d := 8m$ and $L := \lceil (d+1)/\delta \rceil$

```
function $\mathcal{R}(C)$
    Draw a fixed quantum circuit $H$ according to $\mathcal{H}_\mathcal{A}$
    for $i \in \{0, \ldots, L-1\}$ do
        $x_i \leftarrow \Delta(2i - L + 1)/(L-1)$
        $y_i \leftarrow \mathcal{O}\left(C(x_i)\right) |Q(x_i)|^2$        ▷ $C(x_i) \in \mathcal{H}_{\mathcal{A},\Delta}$ with $H$ kept fixed for all $x_i$.
                                                                            See Eqs. (5)–(9) and Def. 3.

    $l \leftarrow 0$
    $r \leftarrow 2$
    loop   poly$(m)$ times
        $c \leftarrow (l+r)/2$
        if $W\left(1_d, (x_i, y_i)_{i=0}^{L-1}, l, c\right)$ then        ▷ $W$ is an NP-oracle
            $r \leftarrow c$
        else
            $l \leftarrow c$
    return $l$                                                              ▷ $l$ is the approximation for $\tilde{p}(1)$
```

description of the quantum circuit C in the architecture $\mathcal{A}$ and outputs $\mathcal{O}(C)$ that satisfies

$$\Pr_{C\sim\mathcal{H}_\mathcal{A}}\left[\,|\,\mathcal{O}(C) - |\langle 0^n|C|0^n\rangle|^2\,| \le \epsilon\,\right] \ge \frac{3}{4} + \delta.$$

Then there exists a classical probabilistic algorithm $\mathcal{R}$ with the oracle access to $\mathcal{O}$ and an NP*-machine which outputs $\mathcal{R}^{\mathcal{O},\mathsf{NP}}(C)$ in time* poly(n,m) *satisfying:*

$$\Pr_{\mathcal{R}}\left[\left|\mathcal{R}^{\mathcal{O},\mathsf{NP}}(C) - |\langle 0^n|C|0^n\rangle|^2\right| < \epsilon \exp\left[O(m\log\frac{m}{\delta})\right]\right] \ge \frac{1}{2} + \delta - \frac{1}{\text{poly}(m)}\ . \tag{15}$$

Proof. We first describe the probabilistic algorithm $\mathcal{R}$ for the reduction. Then, we show that the algorithm $\mathcal{R}$ satisfies the conditions in the lemma. We define the NP-oracle W as the oracle that solves the following NP problem:

(Input) A positive integer d in the unary representation, L pairs $\{(x_i, y_i) \in \mathbb{R}^2\}_{i\in\{0,1,\ldots,L-1\}}$, and $l, r \in \mathbb{R}$ such that $l < r$.

(Output) *True*: if there exists a polynomial $\tilde{p}(x) = \sum_{j=0}^{d} a_j x^j$ such that

$$\left|\{i \in \{0,\ldots,L-1\} : |\tilde{p}(x_i) - y_i| \le |Q(x_i)|^2\epsilon\}\right| \ge (1+\delta)L/2,$$

and $\tilde{p}(1) \in [l, r)$. *False*: otherwise.

This problem is in NP since for a given certificate $\tilde{p}(x)$, the conditions above can be verified in polynomial time. We now describe the reduction algorithm $\mathcal{R}$ shown in Algorithm 1. Then, we will show that algorithm $\mathcal{R}$ satisfies the conditions of the lemma.

$\mathcal{R}$ is a probabilistic polynomial-time algorithm accessing the oracle $\mathcal{O}$ and the NP-oracle W. We set $\Delta = \Theta(m^{-k})$ for some constant $k \ge 2$ so that the total variation distance between $\mathcal{H}_\mathcal{A}$ and $\mathcal{H}_{\mathcal{A},\Delta}$ is $O(m\Delta) = 1/\text{poly}(m)$ by Lemma 1. By L calls to the oracle $\mathcal{O}$, the probability that at least $(1+\delta)L/2$ points are computed with error of at most ϵ is at least $1/2 + \delta - 1/\text{poly}(m)$; the latter follows from Markov's inequality (see Appendix B for details). So we can assume that at least $(1+\delta)L/2$ points are computed with error at most ϵ. In other words,

$$\left|\{i \in \{0,\ldots,L-1\} : |p_e(x_i) - y_i| \le |Q(x_i)|^2\epsilon\}\right| \ge (1+\delta)L/2$$

for the degree-d polynomial $p_e(x) := |\langle 0^n| P(x) |0^n\rangle|^2$ in Eq. (13). In this case, $W\left(1_d, (x_i, y_i)_{i=0}^{L-1}, 0, 2\right)$ is true where $[0, 2)$ is the initial region for the binary search since $p_e(x)$ can be seen as a certificate that satisfies the conditions in W, and by Eq. (12) we have $p_e(x) \le |Q(x)|^2 < 2$. From the binary search in Algorithm 1, we obtain l and r such that $0 \le r - l \le 2^{-\text{poly}(m)}$ and $W(1_d, (x_i, y_i)_{i=0}^{L-1}, l, r)$ is true. Let $\tilde{p}(x) = \sum_{j=0}^{d} a_j x^j$ be a certificate for $W(1_d, (x_i, y_i)_{i=0}^{L-1}, l, r)$. Then, $\tilde{p}(x)$ satisfies

$$\left|\{i \in \{0,\ldots,L-1\} : |\tilde{p}(x_i) - y_i| \le |Q(x_i)|^2\epsilon\}\right| \ge (1+\delta)L/2\ ,$$

and $|\tilde{p}(1) - \mathcal{R}(C)| \le 2^{-\text{poly}(m)}$ because $\mathcal{R}(C) = l$ and $\tilde{p}(1) \in [l, r)$.

In the following, we show that $|\tilde{p}(1) - p_e(1)| \le \epsilon\exp\{O(m\log m)\}$. Define the two sets S_{p_e} and $S_{\tilde{p}}$ by

$$S_{p_e} = \left\{i \in \{0,\ldots,L-1\} : |p_e(x_i) - y_i| \le |Q(x_i)|^2\epsilon\right\},$$
$$S_{\tilde{p}} = \left\{i \in \{0,\ldots,L-1\} : |\tilde{p}(x_i) - y_i| \le |Q(x_i)|^2\epsilon\right\}.$$

Since S_{p_e} and $S_{\tilde{p}}$ have sizes that are at least $(1+\delta)L/2$ there exists a non-empty intersection. We have $|S_{p_e} \cap S_{\tilde{p}}| = |S_{p_e}| + |S_{\tilde{p}}| - |S_{p_e} \cup S_{\tilde{p}}| \ge (1+\delta)L - L \ge \delta L \ge d+1$.

Since $p_e(x)$ and $\tilde{p}(x)$ are degree d polynomials by assumption, and $|p_e(x_i) - \tilde{p}(x_i)| \leq |p_e(x_i) - y_i| + |y_i - \tilde{p}(x_i)| \leq 2|Q(x_i)|^2\epsilon$ for at least $d+1$ points in $x_0, \ldots, x_{L-1} \in [-\Delta, \Delta]$. Then by Lemma 5 we obtain the desired result

$$|p_e(1) - \tilde{p}(1)| \leq \epsilon(2 + O(m\Delta)) \\ \times \exp\left[d\left(1 + \log\left((1+\Delta^{-1})\frac{L-1}{d}\right)\right)\right].$$

Hence, $|p_e(1) - \mathcal{R}(C)| \leq \epsilon(2 + O(m\Delta)) \exp\left[d(1 + \log((1+\Delta^{-1})(L-1)/d))\right] + 2^{-\mathrm{poly}(m)}$. Since $\Delta = \Theta(m^{-k})$ for some constant $k \geq 2$, we obtain Eq. (15) in the Lemma. □

If we take $\delta = 1/\mathrm{poly}(n)$, the success probability of $1/2 + \delta$ can be boosted to a constant greater than $1/2$ by $O(\delta^{-2})$ calls to the oracle.

IV. #P-HARDNESS OF THE COMPUTATION OF PROBABILITY AMPLITUDE FOR THE WORST-CASE QUANTUM CIRCUIT

Definition 4 (#P [28]). A function $f\colon \{0,1\}^* \to \mathbb{N}$ is in #P if there exists a polynomial $p\colon \mathbb{N} \to \mathbb{N}$ and a polynomial-time deterministic Turing machine M such that for every $x \in \{0,1\}^*$,

$$f(x) = \left|\left\{y \in \{0,1\}^{p(|x|)} : M(x,y) = 1\right\}\right|.$$

Lemma 8 (Equivalent to Thm 3.2 in Fenner et al [8]). *For any $f \in$ #P there is a polynomial-time uniform family of quantum circuits $\{C_\ell(x)\}$ and a polynomial p such that for all x of length ℓ,*

$$|\langle 0^{p(\ell)}|C_\ell(x)|0^{p(\ell)}\rangle|^2 = \left(1 - \frac{f(x)}{2^{p(\ell)-1}}\right)^2. \qquad (16)$$

Proof. For any $f \in$ #P, there is a poly-time deterministic Turing machine $M(x,y)$ such that $f(x) = |\{y \in \{0,1\}^{p(\ell)} : M(x,y) = 1\}|$ from Definition 4 where $p(\ell)$ is a polynomial in $\ell = |x|$. Let the poly-time uniform family of quantum circuits be $C_\ell(x) = H^{\otimes p(\ell)} V_x H^{\otimes p(\ell)}$ where H is the Hadamard gate and

$$V_x = \sum_{y \in \{0,1\}^{p(\ell)}} (-1)^{M(x,y)} |y\rangle\langle y|.$$

Then

$$\langle 0^{p(\ell)}|C_\ell(x)|0^{p(\ell)}\rangle = \frac{1}{2^{p(\ell)}} \sum_{y \in \{0,1\}^{p(\ell)}} (-1)^{M(x,y)}$$
$$= \frac{1}{2^{p(\ell)}} \left(|\{y : M(x,y) = 0\}| - |\{y : M(x,y) = 1\}|\right)$$
$$= 1 - \frac{f(x)}{2^{p(\ell)-1}}.$$

□

Lemma 9. *Let $\mathcal{O}$ be an oracle that for an arbitrary given quantum circuit C with m gates computes $|\langle 0^n|C|0^n\rangle|^2$ with the additive error less than 2^{-m^μ} for some constant $\mu > 0$. Then there exists an $\mathsf{FP}^{\mathcal{O}}$ algorithm that solves any #P problem.*

Proof. Let f be an arbitrary #P function. Let $M(x,y)$ be defined as before such that $f(x) = |\{y \in \{0,1\}^{p(|x|)} : M(x,y) = 1\}|$.

Since in the two cases $f(x) = 2^{p(|x|)-1} \pm c$ one obtains the same probability (Eq. (16)), we introduce a function g that is sign unambiguous with respect to $1 - \frac{g(x)}{2^{p(\ell)-1}}$ and is in one-to-one correspondence with the the probability amplitude. Let $g\colon \{0,1\}^* \to \mathbb{N}$ be a function defined by $g(x) = f(x) + 2^{p(|x|)}$. Here, $G(x,y,z) \equiv z \vee M(x,y)$ satisfies $g(x) = |\{y \in \{0,1\}^{p(|x|)}, z \in \{0,1\} : G(x,y,z) = 1\}|$. Since $G(x,y,z)$ is computable in polynomial time, g is also a #P function. We will show an $\mathsf{FP}^{\mathcal{O}}$ algorithm that computes $g(x)$, and therefore it also computes $f(x)$.

As in Lemma 8, there exists poly-time uniform family of quantum circuits $C_\ell(x)$ such that $|\langle 0^n|C_\ell(x)|0^n\rangle|^2 = (1 - g(x)/2^{n-1})^2$ where $n \equiv p(|x|) + 1$. The value of $g(x)$ is exactly determined from the approximation of $(1 - g(x)/2^{n-1})^2$ to within the additive error less than $1/2^{2n-1}$. Therefore, to compute $f(x)$, it is sufficient to compute $(1 - g(x)/2^{n-1})^2$ with additive error less than $1/2^{2n-1}$.

For computing $|\langle 0^n|C_\ell(x)|0^n\rangle|^2$ with small additive error with the aid of the oracle $\mathcal{O}$, we use an enlarged quantum circuit $C_\ell(x) \cdot \mathbb{I}_1^k$, where $\mathbb{I}_1^k$ is k identity gates acting on the first qubit (the number of qubits remains the same). Since $|\langle 0^n|C_\ell(x) \cdot \mathbb{I}_1^k|0^n\rangle|^2 = |\langle 0^n|C_\ell(x)|0^n\rangle|^2$, by assumption of the lemma $\mathcal{O}(C_\ell(x) \cdot \mathbb{I}_1^k)$ outputs $(1 - g(x)/2^{n-1})^2$ with additive error less than $2^{-(m+k)^\mu}$ where m is the number of gates in $C_\ell(x)$. By choosing $k = \lceil (2n)^{1/\mu} \rceil$, the additive error is upper bounded by $2^{-k^\mu} \leq 2^{-2n}$. □

These prove our main theorems which we restate along with their proofs:

Theorem 1. *There is an architecture $\mathcal{A}$ such that it is #P-hard under* BPP*-reduction to approximate $|\langle 0|C|0\rangle|^2$ with probability $1 - O(1/m)$ over the choice of $C \sim \mathcal{H}_{\mathcal{A}}$ to within the additive error $2^{-\Omega(m \log m)}$.*

Proof. This is immediate from Lemmas 6 and 9. □

Theorem 2. *There is an architecture $\mathcal{A}$ such that it is #P-hard under* $\mathsf{BPP}^{\mathsf{NP}}$*-reduction to approximate $|\langle 0|C|0\rangle|^2$ with probability $\frac{3}{4} + \frac{1}{\mathrm{poly}(n)}$ over the choice of $C \sim \mathcal{H}_{\mathcal{A}}$ to within the additive error $2^{-\Omega(m \log m)}$.*

Proof. This is immediate from Lemmas 7 and 9. □

Remark 2. Since #P-hardness implies $\mathsf{C_{=}P}$-hardness with respect to Turing reduction, in the statements of the main theorems above one can simply replace #P-hard with $\mathsf{C}_=\mathsf{P}$-hard with respect to Turing reduction.

V. Hardness of fixed circuits

In this section we show that any circuit with the same architecture as the worst-case circuit also has the same hardness properties as shown above. A perhaps surprising corollary is that sampling from circuits close to identity as $\#P$-hard.

Theorem 3. *Let $U = U_m \cdots U_1$ be a fixed quantum circuit with an architecture $\mathcal{A}$. Let $\mathcal{O}$ be an oracle approximating the output probability to within the additive error $\exp(-\Omega(m \log m))$ of a circuit with the architecture $\mathcal{A}$ whose local gates are Δ-close to U_k in spectral norm, where $\Delta = o(m^{-1})$. Then, there is a $\mathsf{BPP}^{\mathsf{NP},\mathcal{O}}$ algorithm solving #P-problem.*

Proof. Let $U_k(\theta) = U_k f(\theta h_k)$, where $f(h_k) = U_k^\dagger C_k$ and C_k is the corresponding gate of the worst-case circuit. We have $U(0) = U_k$ and $U_k(1) = C_k$. Let us quantify the distance between the distribution over $U_k(\theta)$ and U_k for $|\theta| \leq \Delta$.

Let $\{\lambda_j\}_j$ be eigenvalues of the Hermitian matrix h_k. Using the spectral norm and its invariance under unitary multiplication we have

$$\begin{aligned} &\|U_k(0) - U_k(\theta)\|_\infty = \|f_k(0) - f_k(\theta h_k)\|_\infty \\ &= \max_j \left|1 - \frac{1+i\theta\lambda_j}{1-i\theta\lambda_j}\right| = 2|\theta| \max_j \left|\frac{\lambda_j}{1-i\theta\lambda_j}\right| \\ &\leq 2\Delta \max_j |\lambda_j| = 2\Delta \|h_k\|_\infty = O(\Delta)\ , \end{aligned}$$

where we used $\|h_k\|_\infty = O(1)$ because of the following argument. Now $\|h\|_\infty$ is large if the unitary matrix $U_k^\dagger C_k$ has an eigenvalue near -1. We have the freedom to multiply each C_k by a global phase $C_k \to e^{i\varphi_k} C_k$, as this will not change the observables. In particular, all probabilities and $\#P$-hardness of the worst-case circuit C are invariant under this transformation. The multiplication of the global phase can be used to avoid the eigenvalue near -1. Here, this unitary matrix has size at most four, and we can always multiply C_k by an appropriate $e^{i\varphi_k}$ to rotate any eigenvalues close to -1 such that all eigenvalues of $U_k^\dagger C_k$ are a constant distance away from -1 rendering $\max_j |\lambda_j| = \|h_k\|_\infty \leq f^{-1}(\exp(i3\pi/4)) = O(1)$. □

This theorem leads to the following result:

Corollary 2. *In Theorem 3 the local gates can be identity $U_k = I$.*

Gottesman-Knill theorem states that circuits made up of Clifford gates are classically efficient to simulate. The Sum-over-Cliffords method of [29][section 2.3] gives classical algorithms for sampling the output of near-Clifford circuits with non-trivial improvement on the exponential scaling of the run-time. However the above corollary shows that approximating the probability amplitudes of *near* Clifford circuits to within the stated additive error cannot be performed efficiently on a classical computer unless the polynomial hierarchy collapses to finite level.

Lastly, we comment on a class of quantum circuits known as IQP circuit [6], [10], [30]. These circuits have the following form $C = H^{\otimes n} C_Z H^{\otimes n}$– that is, the first and last layers are n Hadamard gates and all intermediate gates are diagonal in the Z-basis. We believe our techniques along with the proof of the total variation distance in [19] can be generalized to random IQP circuits in a straightforward manner. This would imply the hardness of approximating the output probabilities of average-case IQP circuits.

Acknowledgement: R. Movassagh acknowledges funding from the MIT-IBM Watson AI Lab under the project *Machine Learning in Hilbert space*. The research was supported by the IBM Research Frontiers Institute. R. Mori was supported in part by JST PRESTO Grant Number JPMJPR1867 and JSPS KAKENHI Grant Numbers JP17K17711, JP18H04090, JP20H04138, and JP20H05966.

References

[1] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. Brandao, D. A. Buell *et al.*, “Quantum supremacy using a programmable superconducting processor,” *Nature*, vol. 574, no. 7779, pp. 505–510, 2019.

[2] L. K. Grover, “A fast quantum mechanical algorithm for database search,” in *Proceedings of the twenty-eighth annual ACM symposium on Theory of computing*, 1996, pp. 212–219.

[3] D. R. Simon, “On the power of quantum computation,” *SIAM journal on computing*, vol. 26, no. 5, pp. 1474–1483, 1997.

[4] P. W. Shor, “Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer,” *SIAM review*, vol. 41, no. 2, pp. 303–332, 1999.

[5] R. P. Feynman, “Quantum mechanical computers,” *Foundations of physics*, vol. 16, no. 6, pp. 507–531, 1986.

[6] M. J. Bremner, R. Jozsa, and D. J. Shepherd, “Classical simulation of commuting quantum computations implies collapse of the polynomial hierarchy,” in *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 467, no. 2126. The Royal Society, 2011, pp. 459–472.

[7] B. M. Terhal and D. P. DiVincenzo, “Adaptive quantum computation, constant depth quantum circuits and Arthur–Merlin games,” *Quant. Inf. Comp.*, vol. 4, no. 2, pp. 134–145, 2004.

[8] S. Fenner, F. Green, S. Homer, and R. Pruim, “Determining acceptance possibility for a quantum computation is hard for the polynomial hierarchy,” *arXiv preprint quant-ph/9812056*, 1998.

[9] S. Aaronson and A. Arkhipov, “The computational complexity of linear optics,” in *Proceedings of the forty-third annual ACM symposium on Theory of computing*. ACM, 2011, pp. 333–342.

[10] M. J. Bremner, A. Montanaro, and D. J. Shepherd, “Average-case complexity versus approximate simulation of commuting quantum computations,” *Physical review letters*, vol. 117, no. 8, p. 080501, 2016.

[11] S. Boixo, S. V. Isakov, V. N. Smelyanskiy, R. Babbush, N. Ding, Z. Jiang, M. J. Bremner, J. M. Martinis, and H. Neven, “Characterizing quantum supremacy in near-term devices,” *Nature Physics*, vol. 14, no. 6, p. 595, 2018.

[12] J. Napp, R. L. La Placa, A. M. Dalzell, F. G. Brandao, and A. W. Harrow, "Efficient classical simulation of random shallow 2D quantum circuits," *arXiv preprint arXiv:2001.00021*, 2020.

[13] C. Huang, F. Zhang, M. Newman, J. Cai, X. Gao, Z. Tian, J. Wu, H. Xu, H. Yu, B. Yuan *et al.*, "Classical simulation of quantum supremacy circuits," *arXiv preprint arXiv:2005.06787*, 2020.

[14] E. Pednault, J. A. Gunnels, G. Nannicini, L. Horesh, and R. Wisnieff, "Leveraging secondary storage to simulate deep 54-qubit sycamore circuits," *arXiv preprint arXiv:1910.09534*, 2019.

[15] A. Harrow and S. Mehraban, "Approximate unitary t-designs by short random quantum circuits using nearest-neighbor and long-range gates," *arXiv preprint arXiv:1809.06957*, 2018.

[16] A. M. Dalzell, N. Hunter-Jones, and F. G. Brandão, "Random quantum circuits anti-concentrate in log depth," *arXiv preprint arXiv:2011.12277*, 2020.

[17] L. Stockmeyer, "On approximation algorithms for #P," *SIAM Journal on Computing*, vol. 14, no. 4, pp. 849–861, 1985.

[18] A. Bouland, B. Fefferman, C. Nirkhe, and U. Vazirani, "On the complexity and verification of quantum random circuit sampling," *Nature Physics*, vol. 15, no. 2, p. 159, 2019.

[19] R. Movassagh, "Quantum supremacy and random circuits," *arXiv preprint arXiv:1909.06210*, 2020.

[20] ——, "Efficient unitary paths and quantum computational supremacy: A proof of average-case hardness of random circuit sampling," *arXiv preprint arXiv:1810.04681*, 2018.

[21] R. Paturi, "On the degree of polynomials that approximate symmetric boolean functions (preliminary version)," in *Proceedings of the twenty-fourth annual ACM symposium on Theory of computing*. ACM, 1992, pp. 468–474.

[22] E. A. Rakhmanov, "Bounds for polynomials with a unit discrete norm," *Annals of mathematics*, pp. 55–88, 2007.

[23] L. R. Welch and E. R. Berlekamp, "Error correction for algebraic block codes," Dec. 30 1986, uS Patent 4 633 470.

[24] A. Bouland, B. Fefferman, Z. Landau, and Y. Liu, "Noise and the frontier of quantum supremacy," *arXiv preprint arXiv:2102.01738*, 2021.

[25] Y. Kondo, R. Mori, and R. Movassagh, "Fine-grained analysis and improved robustness of quantum supremacy for bosonsampling," *in preparation*, 2021.

[26] H.-S. Zhong, H. Wang, Y.-H. Deng, M.-C. Chen, L.-C. Peng, Y.-H. Luo, J. Qin, D. Wu, X. Ding, Y. Hu *et al.*, "Quantum computational advantage using photons," *Science*, vol. 370, no. 6523, pp. 1460–1463, 2020.

[27] M. Oszmaniec, N. Dangniam, M. E. Morales, and Z. Zimborás, "Fermion sampling: a robust quantum computational advantage scheme using fermionic linear optics and magic input states," *arXiv preprint arXiv:2012.15825*, 2020.

[28] S. Arora and B. Barak, *Computational Complexity: A Modern Approach*, 1st ed. USA: Cambridge University Press, 2009.

[29] S. Bravyi, D. Browne, P. Calpin, E. Campbell, D. Gosset, and M. Howard, "Simulation of quantum circuits by low-rank stabilizer decompositions," *Quantum*, vol. 3, p. 181, 2019.

[30] M. J. Bremner, A. Montanaro, and D. J. Shepherd, "Achieving quantum supremacy with sparse and noisy commuting quantum computations," *arXiv preprint arXiv:1610.01808*, 2016.

Appendix A
Proof of Lemma 5

Let $p_j = p(x_j)$ for all $j = \{0, 1, 2, \ldots, d\}$, where by assumption $|p_j| \leq \epsilon$. The Lagrange representation of the function $p(x)$ writes

$$p(x) = \sum_{j=0}^{d} p_j \; \delta_j(x), \qquad \delta_j(x) \equiv \frac{\prod_{\ell \neq j} x - x_\ell}{\prod_{\ell \neq j} x_j - x_\ell} \; .$$

By triangular inequality we have $|p(1)| \leq \epsilon \sum_{j=0}^{d} |\delta_j(1)|$. Moreover, using the fact that $|x_j| \leq \Delta$ for all j we have

$$\begin{aligned} |\delta_j(1)| &= \frac{\prod_{\ell \neq j} |1 - x_\ell|}{\prod_{\ell \neq j} |x_j - x_\ell|} < \frac{\prod_{\ell \neq j} (1 + \Delta)}{\prod_{\ell \neq j} |x_j - x_\ell|} \\ &< \frac{(1 + \Delta)^d}{\prod_{\ell \neq j} |x_j - x_\ell|}. \end{aligned}$$

Since $|x_j - x_\ell| = \frac{2\Delta}{L-1}|a_j - a_\ell| \geq \frac{2\Delta}{L-1}|j - \ell|$, we have $\prod_{\ell \neq j} |x_j - x_\ell| \geq (\frac{2\Delta}{L-1})^d \prod_{\ell \neq j} |j - \ell|$. Moreover $\prod_{\ell \neq j} |j - \ell| = \prod_{\ell \in \{0,1,2,\ldots,,j-1,j+1,\ldots,d\}} |j - \ell| = j!(d-j)!$ and we obtain

$$\begin{aligned} |\delta_j(1)| &< \left(\frac{L-1}{2\Delta}\right)^d \frac{(1+\Delta)^d}{\prod_{\ell \neq j} |j - \ell|} \\ &= \left(\frac{L-1}{2\Delta}\right)^d \frac{(1+\Delta)^d}{j! \, (d-j)!} \; . \end{aligned}$$

We express the bound $|p(1)| \leq \epsilon \sum_{j=0}^{d} |\delta_j(1)|$ as

$$\begin{aligned} |p(1)| &< \epsilon \left(\frac{(L-1)(1+\Delta)}{2\Delta}\right)^d \sum_{j=0}^{d} \frac{1}{j! \, (d-j)!} \\ &= \epsilon \left(\frac{(L-1)(1+\Delta)}{2\Delta}\right)^d \frac{1}{d!} \sum_{j=0}^{d} \binom{d}{j} \\ &= \epsilon \left(\frac{(L-1)(1+\Delta)}{2\Delta}\right)^d \frac{2^d}{d!}. \end{aligned}$$

By Stirling's inequality $n! \geq \sqrt{2\pi n} \frac{n^n}{\mathrm{e}^n}$, we conclude that

$$\begin{aligned} |p(1)| &< \epsilon \left(\frac{(L-1)(1+\Delta)}{2\Delta}\right)^d \frac{2^d}{d!} \\ &\leq \epsilon \frac{\left(\mathrm{e}(L-1)(1+\Delta^{-1})/d\right)^d}{\sqrt{2\pi d}} \\ &= \epsilon \frac{\exp[d(1 + \log((1+\Delta^{-1})\frac{L-1}{d}))]}{\sqrt{2\pi d}}. \end{aligned}$$

Appendix B
The proof of the concentration

The proof was originally given in [9], see also Theorem 1 in [19]. From Lemma 1,

$$\Pr_{C \sim \mathcal{H}_{\mathcal{A},\Delta}} \left[|\mathcal{O}(C) - p(\Delta)| \leq |Q(\Delta)|^2 \epsilon \right] \geq \frac{3}{4} + \delta - O(m\Delta)$$

Let

$$\Theta := \{ i \in \{0, \ldots, L-1\} \mid |p(x_i) - y_i| \leq |Q(x_i)|^2 \epsilon \}.$$

Then

$$\begin{aligned}
&\Pr_{C\sim\mathcal{H}_{\mathcal{A},\Delta}}[|\Theta| \geq (1+\delta)L/2] \\
&= 1 - \Pr_{C\sim\mathcal{H}_{\mathcal{A},\Delta}}[|\Theta| < (1+\delta)L/2] \\
&= 1 - \Pr_{C\sim\mathcal{H}_{\mathcal{A},\Delta}}[L - |\Theta| > (1-\delta)L/2] \\
&\geq 1 - \frac{(\frac{1}{4} - \delta + O(m\Delta))L}{(1-\delta)L/2} \\
&= \frac{1}{2} + \frac{3}{2}\frac{\delta}{1-\delta} - O(m\Delta).
\end{aligned}$$

The inequality is obtained from Markov's inequality.