Power of Quantum Computation with Few Clean Qubits*

Keisuke Fujii[1], Hirotada Kobayashi[2], Tomoyuki Morimae[3], Harumichi Nishimura[4], Shuhei Tamate[5], and Seiichiro Tani[6]

1 **The Hakubi Center for Advanced Research and Quantum Optics Group, Division of Physics and Astronomy, Graduate School of Science, Kyoto University, Kyoto, Japan**
2 **Principles of Informatics Research Division, National Institute of Informatics, Tokyo, Japan**
3 **Advanced Scientific Research Leaders Development Unit, Gunma University, Kiryu, Gunma, Japan**
4 **Department of Computer Science and Mathematical Informatics, Graduate School of Information Science, Nagoya University, Nagoya, Aichi, Japan**
5 **Principles of Informatics Research Division, National Institute of Informatics, Tokyo, Japan**
6 **NTT Communication Science Laboratories, NTT Corporation, Atsugi, Kanagawa, Japan**

Abstract

This paper investigates the power of polynomial-time quantum computation in which only a very limited number of qubits are initially clean in the $|0\rangle$ state, and all the remaining qubits are initially in the totally mixed state. No initializations of qubits are allowed during the computation, nor are intermediate measurements. The main contribution of this paper is to develop unexpectedly strong error-reduction methods for such quantum computations that simultaneously reduce the number of necessary clean qubits. It is proved that any problem solvable by a polynomial-time quantum computation with one-sided bounded error that uses logarithmically many clean qubits is also solvable with exponentially small one-sided error using just two clean qubits, and with polynomially small one-sided error using just one clean qubit. It is further proved in the two-sided-error case that any problem solvable by such a computation with a constant gap between completeness and soundness using logarithmically many clean qubits is also solvable with exponentially small two-sided error using just two clean qubits. If only one clean qubit is available, the problem is again still solvable with exponentially small error in one of the completeness and soundness and with polynomially small error in the other. An immediate consequence is that the Trace Estimation problem defined with *fixed* constant threshold parameters is complete for $\mathrm{BQ}_{[1]}\mathrm{P}$ and $\mathrm{BQ}_{\log}\mathrm{P}$, the classes of problems solvable by polynomial-time quantum computations with completeness 2/3 and soundness 1/3 using just one and logarithmically many clean qubits, respectively. The techniques used for proving the error-reduction results may be of independent interest in themselves, and one of the technical tools can also be used to show the hardness of weak classical simulations of one-clean-qubit computations (i.e., DQC1 computations).

1998 ACM Subject Classification F.1.2 Modes of Computation, F.1.3 Complexity Measures and Classes

Keywords and phrases DQC1, quantum computing, complete problems, error reduction

Digital Object Identifier 10.4230/LIPIcs.ICALP.2016.13

* A full version [11] of this paper is available at arXiv.org e-Print archive, arXiv:1509.07276 [quant-ph].

43rd International Colloquium on Automata, Languages, and Programming (ICALP 2016).
Editors: Ioannis Chatzigiannakis, Michael Mitzenmacher, Yuval Rabani, and Davide Sangiorgi;
Article No. 13; pp. 13:1–13:14

Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

1 Introduction

1.1 Background

One of the most important goals in quantum information processing is to realize a quantum mechanical machine whose computational ability is superior to classical computers. The ultimate goal is, of course, to realize a large scale universal quantum computer, which still seems to be many years off despite extensive experimental efforts. Plenty of attention has thus been paid to "intermediate" (i.e., non-universal) models of quantum computation, which are somehow easier to physically implement. Such intermediate models do not offer universal quantum computation, but are believed to still be able to solve some problems that are hard for classical computers.

The *deterministic quantum computation with one quantum bit (DQC1)*, often mentioned as the *one-clean-qubit model*, is one of the most well-studied examples of such intermediate models. This model was introduced by Knill and Laflamme [15] to reflect some actual experimental setups such as nuclear magnetic resonance (NMR), where pure clean qubits are very hard to prepare and therefore are considered as very expensive resources. For example, in nuclear spin ensemble systems such as liquid state NMR systems, it is usually extremely hard, although not impossible, to polarize a spin (i.e., to initialize a qubit to state $|0\rangle$), since energy scale of a nuclear spin qubit is quite small, while it is favorable for long coherence time. A DQC1 computation over w qubits starts with the initial state of the totally mixed state except for a single clean qubit, namely, $|0\rangle\langle 0| \otimes \left(\frac{I}{2}\right)^{\otimes(w-1)}$. After applying a polynomial-size *unitary* quantum circuit to this state, only a single output qubit is measured in the computational basis at the end of the computing in order to read out the computation result. No initializations of qubits are allowed during the computation, nor are intermediate measurements. The DQC1 model is believed not to have full computational power of the standard polynomial-time quantum computation, and is indeed strictly less powerful under some reasonable assumptions [5]. At first glance the model even looks easy to classically simulate and does not seem to offer any quantum advantage, partly because its highly-mixed initial state obviously lacks "quantumness" such as entanglement, coherence, and discord, which are widely believed to be origins of the power of quantum information processing, and also because any time-evolution over a single-qubit state or a totally mixed state is trivially simulatable by a classical computation. Nevertheless, the DQC1 model is not trivial, either, in the sense that it can efficiently solve several problems for which no efficient classical algorithms are known, such as estimating the spectral density [15], testing integrability [20], calculating the fidelity decay [19], approximating the Jones and HOMFLY polynomials [23, 13], and approximating an invariant of 3-manifolds [12]. As many of these problems have physically meaningful applications, the DQC1 model is one of the most important intermediate quantum computation models.

Despite its importance explained thus far and the fact that tons of papers in physics have focused on it, very little has been studied on the genuinely complexity-theoretic aspects of the DQC1 model (to the best knowledge of the authors, no such studies exist other than Refs. [5, 21, 22]). The primal purpose of the present paper is to establish for the first time the fundamental core of detailed complexity-theoretic treatments of the DQC1 model and its generalization. To provide the very base of the study of computational complexity of such models, this paper investigates how robust these models are against computation error.

Computation error is an inherent feature of quantum computing, as the outcome of a computation is inevitably probabilistic and hence may not always be correct. Error reduction, or success-probability amplification, is thus one of the most fundamental issues in quantum

computing. Computation error can be efficiently reduced to be negligibly small in many standard computation models via a simple repetition-based method. Typical examples are polynomial-time quantum computations with bounded error, and in particular, the error can be made exponentially small in BQP both in completeness and in soundness, which provides a reasonable ground for the well-used definition of BQP that employs bounds 2/3 and 1/3 for completeness and soundness, respectively. In many other computation models, however, it is unclear whether the error can be reduced efficiently by the standard repetition-based method, and more generally, whether error reduction itself is possible. Typically, for models with very limited computational resources like space-bounded quantum computations, it is simply impossible to repeat the original computation sufficiently many times, which becomes an enormous obstacle to error reduction when initializations of qubits are disallowed after the computation starts. Indeed, it is impossible in the case of one-way quantum finite state automata to reduce computation error below a certain constant [4]. Also, the reducibility of computation error is unclear in various logarithmic-space quantum computations. For computations of one-sided bounded error performed by logarithmic-space quantum Turing machines, Watrous [28] presented a nontrivial method that reduces the error to be exponentially small. Other than this result, error-reduction techniques have not been developed much for space-bounded quantum computations.[1]

The computation models with few clean qubits, including DQC1, may be viewed as variants of space-bounded quantum computations in a sense, and thus, it is highly nontrivial to reduce computation error in these models. On the other hand, the reducibility of computation error is particularly desirable in these models, as the DQC1 computations mentioned above that solve the classically-hard problems in fact solve the decisional versions of the problems only with two-sided bounded error. Computation error can be quite large in such computations, and the gap between completeness and soundness is allowed to be polynomially small. The only method known for amplifying success probability of these computations is to sequentially repeat an attempt of the computation polynomially many times, but this requires the clean qubit to be initialized every time after finishing one attempt, and moreover, the result of each attempt must be recorded to classical work space prepared outside of the DQC1 model. It is definitely more desirable if computation error can be reduced without such initializations, the operations that are very expensive for the model. The situation is similar even when the number of clean qubits is allowed to be logarithmically many with respect to the input length. It is also known that any quantum computation of two-sided bounded error that uses logarithmically many clean qubits can be simulated by a quantum computation still of two-sided bounded error that uses just one clean qubit, but the known method for this simulation considerably increases the computational error, and the gap between completeness and soundness becomes polynomially small.

1.2 The results

This paper develops methods of reducing computation error in quantum computations with few clean qubits, including the DQC1 model. As will be presented below, the methods

[1] After the completion of this work, Fefferman, Kobayashi, Lin, Morimae, and Nishimura [10] developed methods of error reduction for space-bounded unitary quantum computations. Both of this very recent method and the one by Watrous [28] do not apply to quantum computations with few clean qubits, for these methods assume the easiness of "exact initialization check" (i.e., the easiness of checking whether the given state is *exactly* equal to the initial state of the computation), which is no longer the case for quantum computations with few clean qubits where many qubits are initially in the totally mixed state.

proposed are unexpectedly powerful and are able to simultaneously reduce both computation error and the number of necessary clean qubits, providing an almost fully satisfying solution in the cases of one-sided bounded error. In the two-sided-error case, the methods in this paper are applicable only when there is a constant gap between completeness and soundness in the original computation, but still significantly improve the situation of quantum computations with few clean qubits as to both the reducibility of computation error and the reducibility of the number of necessary clean qubits. These results are the first error-reducible properties for intermediate quantum computation models, not limited to the DQC1 model.

The results may alternatively be interpreted as that any problem solvable by a DQC1 computation with constant computation error is still solvable with constant computation error even when a bit noisy initial state is given instead of the ideal one-clean-qubit state, for the problem is also solvable with very small error when given an ideal one-clean-qubit initial state, thanks to the error-reduction results. This is perhaps very helpful in actual implementation, as the initial state prepared does not need to be very close to the ideal one-clean-qubit state, and can be away from it by, say, a constant δ in trace distance to still have success probability close to $1-\delta$. In particular, the qubit that is supposed to be clean does not need to be thoroughly purified and may be noisy to some extent.

The result for the two-sided-error case has another implication that the power of DQC1 computations with small two-sided error is characterized by the Trace Estimation problem defined with *fixed* constant threshold parameters. This may also be viewed as the first "gap amplification" result for the Trace Estimation problem. The Trace Estimation problem is ubiquitous in quantum many-body physics as observables, related to various important quantities in physics like the fidelity decay characterizing quantum chaos [9] and the Jones polynomials corresponding to the expected values of the Wilson loops in SU(2) Chern-Simons topological quantum field theory [29]. The results thus provide a useful tool to understand computational complexity of such quantum many-body systems, and establish a new bridge between computational complexity theory and quantum many-body physics.

Simultaneous reducibility of computation error and the number of clean qubits. Let $\mathrm{Q_{log}P}(c,s)$, $\mathrm{Q_{[1]}P}(c,s)$, and $\mathrm{Q_{[2]}P}(c,s)$ denote the classes of problems solvable by polynomial-time quantum computations with completeness c and soundness s that uses logarithmically many clean qubits, one clean qubit, and two clean qubits, respectively. First, in the one-sided-error case, it is proved that any problem solvable by a polynomial-time quantum computation with one-sided bounded error that uses logarithmically many clean qubits is also solvable by that with exponentially small one-sided error using just two clean qubits. If only one clean qubit is available, the problem is still solvable with polynomially small one-sided error (and thus with any small constant one-sided error).

▶ **Theorem 1.1.** *For any polynomially bounded function $p\colon \mathbb{Z}^+ \to \mathbb{N}$ and any polynomial-time computable function $s\colon \mathbb{Z}^+ \to [0,1]$ satisfying $1-s \geq \frac{1}{q}$ for some polynomially bounded function $q\colon \mathbb{Z}^+ \to \mathbb{N}$,*

$$\mathrm{Q_{log}P}(1,s) \subseteq \mathrm{Q_{[2]}P}(1,2^{-p}).$$

▶ **Theorem 1.2.** *For any polynomially bounded function $p\colon \mathbb{Z}^+ \to \mathbb{N}$ and any polynomial-time computable function $s\colon \mathbb{Z}^+ \to [0,1]$ satisfying $1-s \geq \frac{1}{q}$ for some polynomially bounded function $q\colon \mathbb{Z}^+ \to \mathbb{N}$,*

$$\mathrm{Q_{log}P}(1,s) \subseteq \mathrm{Q_{[1]}P}\Big(1, \frac{1}{p}\Big).$$

The above two theorems are for the case of perfect completeness, and similar statements hold even for the case of perfect soundness, by considering the complement of the problem.

▶ **Corollary 1.3.** *For any polynomially bounded function* $p\colon \mathbb{Z}^+ \to \mathbb{N}$ *and any polynomial-time computable function* $c\colon \mathbb{Z}^+ \to [0,1]$ *satisfying* $c \geq \frac{1}{q}$ *for some polynomially bounded function* $q\colon \mathbb{Z}^+ \to \mathbb{N}$*,*

$$\mathrm{Q_{log}P}(c,0) \subseteq \mathrm{Q_{[2]}P}(1-2^{-p},0) \quad \textit{and} \quad \mathrm{Q_{log}P}(c,0) \subseteq \mathrm{Q_{[1]}P}\Big(1-\frac{1}{p},0\Big).$$

In the two-sided-error case, it is proved that any problem solvable by a polynomial-time quantum computation that uses logarithmically many clean qubits and has a constant gap between completeness and soundness can also be solved by that with exponentially small two-sided error using just two clean qubits. If only one clean qubit is available, the problem is again still solvable with exponentially small error in one of the completeness and soundness and polynomially small error in the other.

▶ **Theorem 1.4.** *For any polynomially bounded function* $p\colon \mathbb{Z}^+ \to \mathbb{N}$ *and any constants* c *and* s *in* $(0,1)$ *satisfying* $c > s$*,*

$$\mathrm{Q_{log}P}(c,s) \subseteq \mathrm{Q_{[2]}P}(1-2^{-p},2^{-p}).$$

▶ **Theorem 1.5.** *For any polynomially bounded function* $p\colon \mathbb{Z}^+ \to \mathbb{N}$ *and any constants* c *and* s *in* $(0,1)$ *satisfying* $c > s$*,*

$$\mathrm{Q_{log}P}(c,s) \subseteq \mathrm{Q_{[1]}P}\Big(1-2^{-p},\frac{1}{p}\Big) \cap \mathrm{Q_{[1]}P}\Big(1-\frac{1}{p},2^{-p}\Big).$$

The ideas for the proofs of these statements and techniques developed therein may be of independent interest in themselves, and will be overviewed in Section 2.

Completeness results for Trace Estimation problem. Define the complexity classes $\mathrm{BQ_{log}P}$ and $\mathrm{BQ_{[1]}P}$ by $\mathrm{BQ_{log}P} = \mathrm{Q_{log}P}\big(\frac{2}{3},\frac{1}{3}\big)$ and $\mathrm{BQ_{[1]}P} = \mathrm{Q_{[1]}P}\big(\frac{2}{3},\frac{1}{3}\big)$, respectively. An immediate but important consequence of Theorem 1.5 is that the Trace Estimation problem is complete for $\mathrm{BQ_{log}P}$ and $\mathrm{BQ_{[1]}P}$ under polynomial-time many-one reduction, even when the problem is defined with *fixed* constant parameters that specify the bounds on normalized traces in the yes-instance and no-instance cases.

Given a description of a quantum circuit that specifies a unitary transformation U, the Trace Estimation problem specified with two parameters a and b satisfying $-1 \leq b < a \leq 1$ is the problem of deciding whether the real part of the normalized trace of U is at least a or it is at most b.

Trace Estimation Problem: $\mathrm{TrEst}(a,b)$

Input:	A description of a quantum circuit Q that implements a unitary transformation U over n qubits.
Yes Instances:	$\frac{1}{2^n}\Re(\mathrm{tr}\,U) \geq a$.
No Instances:	$\frac{1}{2^n}\Re(\mathrm{tr}\,U) \leq b$.

The paper by Knill and Laflamme [15] that introduced the DQC1 model already pointed out that this problem is closely related to the DQC1 computation. This point was further clarified in the succeeding literature (see Refs. [21, 22, 23], for instance). More precisely, consider a variant of the Trace Estimation problem where the two parameters a and b may depend on the input length (i.e., the length of the description of Q). It is known that this version of the Trace Estimation problem, for any a and b such that the gap $a - b$ is bounded from below by an inverse-polynomial with respect to the input length, can be solved by a DQC1 computation with *some* two-sided bounded error where the completeness and soundness parameters c and s depend on a and b. It is also known that, for any two nonnegative parameters a and b such that the gap $a - b$ is bounded from below by an inverse-polynomial with respect to the input length, the corresponding version of the Trace Estimation problem is hard for the complexity class $\mathrm{Q}_{[1]}\mathrm{P}(c, s)$ for *some* completeness and soundness parameters c and s that depend on a and b. Hence, the Trace Estimation problem essentially characterizes the power of the DQC1 computation. One subtle matter to be pointed out in the existing arguments above is that, when the parameters a and b are fixed for the Trace Estimation problem, the completeness c and soundness s with which the problem is in $\mathrm{Q}_{[1]}\mathrm{P}(c, s)$ are *different* from the completeness c' and soundness s' with which the problem is hard for $\mathrm{Q}_{[1]}\mathrm{P}(c', s')$. Namely, given two nonnegative parameters a and b of the problem, the computation solves the problem with completeness $c = (1 + a)/2$ and soundness $s = (1 + b)/2$, while the problem is hard for the class with completeness $c' = a/4$ and soundness $s' = b/4$. Therefore, the existing arguments are slightly short for proving $\mathrm{BQ}_{[1]}\mathrm{P}$-completeness of the Trace Estimation problem with fixed parameters a and b (and $\mathrm{Q}_{[1]}\mathrm{P}(c, s)$-completeness of that for fixed completeness and soundness parameters c and s, in general).

In contrast, with Theorem 1.5 in hand, it is immediate to show that the Trace Estimation problem is complete for $\mathrm{BQ}_{\log}\mathrm{P}$ and for $\mathrm{BQ}_{[1]}\mathrm{P}$ for any constants a and b satisfying $0 < b < a < 1$.

▶ **Theorem 1.6.** *For any constants a and b in $(0, 1)$ satisfying $a > b$, $\mathrm{TrEst}(a, b)$ is complete for $\mathrm{BQ}_{\log}\mathrm{P}$ and for $\mathrm{BQ}_{[1]}\mathrm{P}$ under polynomial-time many-one reduction.*

Hardness of weak classical simulations of DQC1 computation. Recently, quite a few number of studies focused on the hardness of *weak* classical simulations of restricted models of quantum computing under some reasonable assumptions [26, 7, 2, 18, 14, 17, 25, 8, 24]. Namely, a plausible assumption in complexity theory leads to the impossibility of efficient sampling by a classical computer according to an output probability distribution generatable with a quantum computing model. Among them are the IQP model [7] and the Boson sampling [2], both of which are proved hard for classical computers to simulate within multiplicative error, unless the polynomial-time hierarchy collapses to the third level (in fact, the main result of Ref. [2] is a much more meaningful hardness result on the weak simulatability of the Boson sampling within *polynomially small additive error*, but which needs a much stronger complexity assumption than the collapse of polynomial-time hierarchy).

An interesting question to ask is whether a similar result holds even for the DQC1 model. Very recently, Morimae, Fujii, and Fitzsimons [17] settled the case of the $\mathrm{DQC1}_m$-type computation, the generalization of the DQC1 model that allows m output qubits to be measured at the end of the computation, by proving that a $\mathrm{DQC1}_m$-type computation with $m \geq 3$ cannot be simulated within multiplicative error unless the polynomial-time hierarchy collapses to the third level. Their proof essentially shows that any PostBQP circuit can be simulated by a $\mathrm{DQC1}_3$-type computation, where PostBQP is the complexity class

corresponding to bounded-error quantum polynomial-time computations with postselection, which is known equivalent to PP [1]. By an argument similar to that in Ref. [7], it follows that PP is in PostBPP (the version of BPP with postselection), if the $\mathrm{DQC1}_3$-type computation is classically simulatable within multiplicative error. Together with Toda's theorem [27], this implies the collapse of the polynomial-time hierarchy to the third level.

One obvious drawback of the existing argument above is an inevitable postselection measurement inherent to the definition of PostBQP. This becomes a quite essential obstacle when trying to extend this argument to the DQC1 model, where only one qubit is allowed to be measured. To deal with the DQC1 model, this paper takes a different approach by considering the complexity class NQP introduced in Ref. [3] or the class SBQP introduced in Ref. [16]. Let $\mathrm{NQ}_{[1]}\mathrm{P}$ and $\mathrm{SBQ}_{[1]}\mathrm{P}$ be the variants of NQP and SBQP, respectively, in which the quantum computation performed is restricted to a DQC1 computation. From one of the technical tools used for proving the main results of this paper, it is immediate to show that the restriction to a DQC1 computation does not change the classes NQP and SBQP.

▶ **Theorem 1.7.** $\mathrm{NQP} = \mathrm{NQ}_{[1]}\mathrm{P}$ *and* $\mathrm{SBQP} = \mathrm{SBQ}_{[1]}\mathrm{P}$.

If any DQC1 computation were classically simulatable within multiplicative error, however, the class $\mathrm{NQ}_{[1]}\mathrm{P}$ would be included in NP and the class $\mathrm{SBQ}_{[1]}\mathrm{P}$ would be included in SBP, where SBP is a classical version of SBQP in short, introduced in Ref. [6]. Similarly, if any DQC1 computation were classically simulatable within exponentially small additive error, both $\mathrm{NQ}_{[1]}\mathrm{P}$ and $\mathrm{SBQ}_{[1]}\mathrm{P}$ would be included in SBP. Combined with Theorem 1.7, any of the inclusions $\mathrm{NQ}_{[1]}\mathrm{P} \subseteq \mathrm{NP}$, $\mathrm{SBQ}_{[1]}\mathrm{P} \subseteq \mathrm{SBP}$, and $\mathrm{NQ}_{[1]}\mathrm{P} \subseteq \mathrm{SBP}$ further implies an implausible consequence that PH = AM, which in particular implies the collapse of the polynomial-time hierarchy to the second level. Accordingly, the following theorem holds.

▶ **Theorem 1.8.** *The DQC1 model is not classically simulatable either within multiplicative error or exponentially small additive error, unless* PH = AM.

The above argument based on NQP and SBQP to prove Theorem 1.8 is very general, and can also be used to show the hardness of weak classical simulations of other quantum computing models. In particular, it can replace the existing argument based on PostBQP, which was developed in Ref. [7] and has appeared frequently in the literature [2, 14, 17, 25, 8, 24]. This also weakens the complexity assumption necessary to prove the hardness results for such models, including the IQP model [7] and the Boson sampling [2] (the polynomial-time hierarchy now collapses to the second level, rather than the third level when using PostBQP). Moreover, the hardness results for such models now hold for any constant multiplicative error $c \geq 1$, rather than only for c satisfying $1 \leq c < \sqrt{2}$ as in Refs. [7, 17].

2 Overview of error-reduction results

This section presents an overview of the proofs for the error reduction results. First, Subsection 2.1 provides high-level descriptions of the proofs of Theorems 1.1 and 1.2, the theorems for the one-sided error case of perfect completeness. Compared with the two-sided-error case, the proof construction is relatively simpler in the perfect-completeness case, but already involves most of key technical ingredients of this paper. Subsection 2.2 then explains the further idea that proves Theorems 1.4 and 1.5, the theorems for the two-sided-error case.

2.1 Proof ideas of Theorems 1.1 and 1.2

Let $A = (A_{\mathrm{yes}}, A_{\mathrm{no}})$ be any problem in $\mathrm{Q_{log}P}(1, s)$, where the function s defining the soundness is bounded away from one by an inverse-polynomial, and consider a polynomial-time

uniformly generated family of quantum circuits that puts A in $\mathrm{Q_{log}P}(1, s)$. Let Q_x denote the quantum circuit from this family when the input is x, where Q_x acts over $w(|x|)$ qubits for some polynomially bounded function w, and is supposed to be applied to the initial state $(|0\rangle\langle 0|)^{\otimes k(|x|)} \otimes \left(\frac{I}{2}\right)^{\otimes(w(|x|)-k(|x|))}$ that contains exactly $k(|x|)$ clean qubits, for some logarithmically bounded function k.

Theorems 1.1 and 1.2 are proved by constructing circuits with desirable properties from the original circuit Q_x. The construction is essentially the same for both of the two theorems and consists of three stages of transformations of circuits: The first stage reduces the number of necessary clean qubits to just one, while keeping perfect completeness and soundness still bounded away from one by an inverse-polynomial. The second stage then makes the acceptance probability of no-instances arbitrarily close to 1/2, still using just one clean qubit and keeping perfect completeness. Here, it not only makes the soundness (i.e., the upper bound of the acceptance probability of no-instances) close to 1/2, but also makes the acceptance probability of no-instances *at least* 1/2. Finally, in the case of Theorem 1.2, the third stage further reduces soundness error to be polynomially small with the use of just one clean qubit, while preserving the perfect completeness property. If one more clean qubit is available, the third stage can achieve exponentially small soundness, which leads to Theorem 1.1. The analyses of the third stage effectively use the fact that the acceptance probability of no-instances is close to 1/2 after the transformation of the second stage.

The rest of this subsection sketches the ideas that realize each of these three stages.

One-Clean-Qubit Simulation Procedure. The first stage uses a procedure called the One-Clean-Qubit Simulation Procedure. Given the quantum circuit Q_x with a specification of the number $k(|x|)$ of clean qubits, this procedure results in a quantum circuit R_x such that the input state to R_x is supposed to contain just one clean qubit, and when applied to the one-clean-qubit initial state, the acceptance probability of R_x is still one if x is in A_{yes}, while it is at most $1 - \delta(|x|)$ if x is in A_{no}, where δ is an inverse-polynomial function determined by $\delta = 2^{-k}(1 - s)$. It is stressed that the One-Clean-Qubit Simulation Procedure preserves perfect completeness, which is in stark contrast to the straightforward method of one-clean-qubit simulation.

Consider the $k(|x|)$-clean-qubit computation performed with Q_x. Let Q denote the quantum register consisting of the $k(|x|)$ initially clean qubits, and let R denote the quantum register consisting of the remaining $w(|x|) - k(|x|)$ qubits that are initially in the totally mixed state. Further let $\mathsf{Q}^{(1)}$ denote the single-qubit quantum register consisting of the first qubit of Q, which corresponds to the output qubit of Q_x. In the one-clean-qubit simulation of Q_x by R_x, the $k(|x|)$ qubits in Q are supposed to be in the totally mixed state initially and R_x tries to simulate Q_x only when Q initially contains the clean all-zero state. To do so, R_x uses another quantum register O consisting of just a single qubit, and this qubit in O is the only qubit that is supposed to be initially clean.

For ease of explanations, assume for a while that all the qubits in Q are also initially clean even in the case of R_x. The key idea in the construction of R_x is the following simulation of Q_x that makes use of the phase-flip transformation: The simulation first applies the Hadamard transformation H to the qubit in O and then flips the phase if and only if the content of O is 1 *and* the simulation of Q_x results in rejection (which is realized by performing Q_x to (Q, R) and then applying the controlled-Z transformation to $(\mathsf{O}, \mathsf{Q}^{(1)})$, where the content 1 in $\mathsf{Q}^{(1)}$ is assumed to correspond to the rejection in the original computation by Q_x). The simulation further performs the inverse of Q_x to (Q, R) and again applies H to O. At the end of the simulation, the qubit in O is measured in the computational basis, where measuring 0

corresponds to acceptance. The point is that this phase-flip-based construction provides a quite "faithful" simulation of Q_x, meaning that the rejection probability of the simulation is polynomially related to the rejection probability of the original computation of Q_x (and in particular, the simulation never rejects when the original computation never rejects, i.e., it preserves the perfect completeness property).

As mentioned before, all the qubits in Q are supposed to be in the totally mixed state initially in the one-clean-qubit simulation of Q_x by R_x, and R_x tries to simulate Q_x only when Q initially contains the clean all-zero state. To achieve this, each of the applications of the Hadamard transformation H is replaced by an application of the controlled-H transformation so that H is applied only when all the qubits in Q are in state $|0\rangle$. By considering the one-clean-qubit computations with the circuit family induced by R_x, the perfect completeness property is preserved and soundness is still bounded away from one by an inverse-polynomial (although the rejection probability becomes smaller for no-instances by a multiplicative factor of 2^{-k}, where notice that 2^{-k} is an inverse-polynomial as k is a logarithmically bounded function).

Randomness Amplification Procedure. The second stage uses the procedure called the Randomness Amplification Procedure. Given the circuit R_x constructed in the first stage, this procedure results in a quantum circuit R'_x such that the input state to R'_x is still supposed to contain just one clean qubit, and when applied to the one-clean-qubit initial state, the acceptance probability of R'_x is still one if x is in $A_{\rm yes}$, while it is in the interval $\left[\frac{1}{2}, \frac{1}{2} + \varepsilon(|x|)\right]$ if x is in $A_{\rm no}$ for some sufficiently small function ε.

Consider the one-clean-qubit computation performed with R_x. Let O denote the single-qubit register consisting of the initially clean qubit, which is also the output qubit of R_x. Let R denote the quantum register consisting of all the other qubits that are initially in the totally mixed state (by the construction of R_x, R consists of $w(|x|)$ qubits).

Suppose that the qubit in O is measured in the computational basis after R_x is applied to the one-clean-qubit initial state $|0\rangle\langle 0| \otimes \left(\frac{I}{2}\right)^{\otimes w(|x|)}$ in (O, R). Obviously from the property of R_x, the measurement results in 0 with probability exactly equal to the acceptance probability $p_{\rm acc}$ of the one-clean-qubit computation with R_x. Now suppose that R_x is applied to a slightly different initial state $|1\rangle\langle 1| \otimes \left(\frac{I}{2}\right)^{\otimes w(|x|)}$ in (O, R), where O initially contains $|1\rangle$ instead of $|0\rangle$ and all the qubits in R are again initially in the totally mixed state. The key property here to be proved is that, in this case, the measurement over the qubit in O in the computational basis results in 1 again with probability exactly $p_{\rm acc}$, the acceptance probability of the one-clean-qubit computation with R_x. This implies that, after the application of R_x to (O, R) with all the qubits in R being in the totally mixed state, the content of O remains the same with probability exactly $p_{\rm acc}$, and is flipped with probability exactly $1 - p_{\rm acc}$, the rejection probability of the original one-clean-qubit computation with R_x, regardless of the initial content of O.

The above observation leads to the following construction of the circuit R'_x. The construction of R'_x is basically a sequential repetition of the original circuit R_x. The number N of repetitions is polynomially many with respect to the input length $|x|$, and the point is that the register O is reused for each repetition, and only the qubits in R are refreshed after each repetition (by preparing N registers $\mathsf{R}_1, \ldots, \mathsf{R}_N$, each of which consists of $w(|x|)$ qubits, the same number of qubits as R, all of which are initially in the totally mixed state). After each repetition the qubit in O is measured in the computational basis (in the actual construction, this step is exactly simulated without any measurement – a single-qubit totally mixed state is prepared as a fresh ancilla qubit for each repetition so that the content of O is copied to

this ancilla qubit using the CNOT transformation, and this ancilla qubit is never touched after this CNOT application). Now, no matter which measurement result is obtained at the jth repetition for every j in $\{1,\ldots,N\}$, the register O is reused as it is, and the circuit R_x is simply applied to $(\mathsf{O},\mathsf{R}_{j+1})$ at the $(j+1)$st repetition. After the N repetitions, the qubit in O is measured in the computational basis, which is the output of R'_x (the output 0 corresponds to acceptance). The point is that at each repetition, the content of O is flipped with probability exactly equal to the rejection probability of the original one-clean-qubit computation of R_x. Taking into account that O is initially in state $|0\rangle$, the computation of R'_x results in acceptance if and only if the content of O is flipped even number of times during the N repetitions. An analysis on Bernoulli trials then shows that, when the acceptance probability of the original one-clean-qubit computation of R_x was in the interval $\left[\frac{1}{2},1\right)$, the acceptance probability of the one-clean-qubit computation of R'_x is at least $1/2$ and converges linearly to $1/2$ with respect to the repetition number. On the other hand, when the acceptance probability of the original R_x was one, the content of O is never flipped during the computation of R'_x, and thus the acceptance probability of R'_x remains one.

Stability Checking Procedures. In the case of Theorem 1.2, the third stage uses the procedure called the One-Clean-Qubit Stability Checking Procedure. Given the circuit R'_x constructed in the second stage, this procedure results in a quantum circuit R''_x such that the input state to R''_x is still supposed to contain just one clean qubit, and when applied to the one-clean-qubit initial state, the acceptance probability of R''_x is still one if x is in A_{yes}, while it is $1/p(|x|)$ if x is in A_{no} for a polynomially bounded function p predetermined arbitrarily.

Consider the one-clean-qubit computation performed with R'_x. Let Q denote the single-qubit register consisting of the initially clean qubit, which is also the output qubit of R'_x. Let R denote the quantum register consisting of all the other qubits that are initially in the totally mixed state, and let $w'(|x|)$ denote the number of qubits in R.

Again the key observation is that, after the application of R'_x to (Q,R) with all the qubits in R being in the totally mixed state (followed by the measurement over the qubit in Q in the computational basis), the content of Q is flipped with probability exactly equal to the rejection probability of the original one-clean-qubit computation with R'_x, regardless of the initial content of Q.

This leads to the following construction of the circuit R''_x. The construction of R''_x is again basically a sequential repetition of the original circuit R'_x, but this time the qubit in Q is also supposed to be initially in the totally mixed state. The circuit R'_x is repeatedly applied $2N$ times, where N is a power of two and is polynomially many with respect to the input length $|x|$, and again the register Q is reused for each repetition, and only the qubits in R are refreshed after each repetition (by preparing $2N$ registers $\mathsf{R}_1,\ldots,\mathsf{R}_{2N}$, each of which consists of $w'(|x|)$ qubits, all of which are initially in the totally mixed state). The key idea for the construction of R''_x is to use a counter that counts the number of attempts such that the measurement over the qubit in Q results in $|1\rangle$ after the application of R'_x (again each measurement is simulated by a CNOT application using an ancilla qubit of a totally mixed state). Notice that the content of Q is never flipped regardless of the initial content of Q, if the original acceptance probability is one in the one-clean-qubit computation with R'_x. Hence, in this case the counter value either stationarily remains its initial value or is increased exactly by $2N$, the number of repetitions. On the other hand, if the original acceptance probability is close to $1/2$ in the one-clean-qubit computation with R'_x, the content of Q is flipped with probability close to $1/2$ after each application of R'_x regardless of the initial

content of Q. This means that, after each application of R'_x, the measurement over the qubit in Q results in $|1\rangle$ with probability close to 1/2 regardless of the initial content of Q, and thus, the increment of the counter value must be distributed around $\frac{1}{2} \cdot 2N = N$ with very high probability. Now, if the counter value is taken modulo $2N$ and if the unique initially clean qubit is prepared for the most significant bit of the counter (which picks the initial counter value from the set $\{0, \ldots, N-1\}$ uniformly at random), the computational-basis measurement over this most significant qubit of the counter always results in $|0\rangle$ if x is in A_{yes}, while it results in $|1\rangle$ with very high probability if x is in A_{no} (which can be made at least $1 - \frac{1}{p(|x|)}$ for an arbitrarily chosen polynomially bounded function p, by taking an appropriately large number N).

One drawback of the construction of R''_x above via the One-Clean-Qubit Stability Checking Procedure is that, in the case of no-instances, there inevitably exist some "bad" initial counter values in $\{0, \ldots, N-1\}$ with which R''_x is forced to accept with unallowably high probability. For instance, if the initial counter value is 0, R''_x is forced to accept when the increment of the counter is less than N, which happens with probability at least a constant. This is the essential reason why the current approach achieves only a polynomially small soundness in the one-clean-qubit case in Theorem 1.2, as the number of possible initial counter values can be at most polynomially many (otherwise the number of repetitions must be super-polynomially many) and even just one "bad" initial value is problematic to go beyond polynomially small soundness. In contrast, if not just one but two clean qubits are available, one can remove the possibility of "bad" initial counter values, which results in the Two-Clean-Qubit Stability Checking Procedure. This time, the circuit R'_x is repeatedly applied $8N$ times, and the counter value is taken modulo $8N$. The two initially clean qubits are prepared for the most and second-most significant bits of the counter, which results in picking the initial counter value from the set $\{0, \ldots, 2N-1\}$ uniformly at random. Now the point is that the counter value can be increased by N before the repetition so that the actual initial value of the counter is in $\{N, \ldots, 3N-1\}$, which discards the tail sets $\{0, \ldots, N-1\}$ and $\{3N, \ldots, 4N-1\}$ of the set $\{0, \ldots, 4N-1\}$. As the size of the tail sets discarded is sufficiently large, there no longer exists any "bad" initial counter value, which leads to the exponentially small soundness in the two-clean-qubit case in Theorem 1.1.

2.2 Proof ideas of Theorems 1.4 and 1.5

The results for the two-sided error case need more complicated arguments and is proved in eight stages of transformations in total, which are split into three parts.

The first part consists of three stages, and proves that any problem solvable with constant completeness and soundness using logarithmically many clean qubits is also solvable with constant completeness and soundness using just one clean qubit. At the first stage of the first part, by a standard repetition with a threshold-value decision, one first reduces errors to be sufficiently small constants, say, completeness 15/16 and soundness 1/16. For this, if the starting computation has a constant gap between completeness and soundness, one requires only a constant number of repetitions, and thus, the resulting computation still requires only logarithmically many clean qubits. The second stage of the first part then reduces the number of clean qubits to just one. The procedure in this stage is exactly the One-Clean-Qubit Simulation Procedure developed in the first stage of the one-sided error case. The gap between completeness and soundness becomes only an inverse-polynomial by this transformation, but the point is that the gap is still sufficiently larger (i.e., a constant times larger) than the completeness error. Now the third stage of the first part transforms the computation resulting from the second stage to the computation that still uses only one

clean qubit and has constant completeness and soundness. The procedure in this stage is exactly the Randomness Amplification Procedure, developed in the second stage of the one-sided error case, and it makes use of the difference of the rates of convergence to 1/2 of the acceptance probability between the yes- and no-instance cases.

The second part consists of two stages, and proves that any problem solvable with constant completeness and soundness using just one clean qubit is also solvable with almost-perfect (i.e., exponentially close to one) completeness and soundness below 1/2 using just logarithmically many clean qubits. At the first stage of the second part, one reduces both of the completeness and soundness errors to be polynomially small, again by a standard repetition with a threshold-value decision. Note that the computation resulting from the first part requires only one clean qubit. Thus, even when repeated logarithmically many times, the resulting computation uses just logarithmically many clean qubits, and achieves polynomially small errors. The second stage of the second part then repeatedly attempts the computation resulting from the first stage polynomially many times, and accepts if at least one of the attempts results in acceptance (i.e., takes OR of the attempts). A straightforward repetition requires polynomially many clean qubits, and to avoid this problem, after each repetition one tries to recover the clean qubits for reuse by applying the inverse of the computation (the failure of this recovery step is counted as an "acceptance" when taking the OR). This results in a computation that still requires only logarithmically many clean qubits, and has completeness exponentially close to one, while soundness is still below 1/2.

Now the third part is essentially the same as the three-stage transformation of the one-sided error case. From the computation resulting from the second part, the first stage of the third part decreases the number of clean qubits to just one, via the One-Clean-Qubit Simulation Procedure. The completeness of the resulting computation is still exponentially close to one and its soundness is bounded away from one by an inverse-polynomial. The second stage of the third part then applies the Randomness Amplification Procedure to make the acceptance probability of no-instances arbitrarily close to 1/2, while keeping completeness exponentially close to one. Finally, the third stage of the third part proves that one can further decrease soundness error to be polynomially small using just one qubit via the One-Clean-Qubit Stability Checking Procedure, or to be exponentially small using just two qubits via the Two-Clean-Qubit Stability Checking Procedure, while keeping completeness exponentially close to one.

By considering the complement problem, the above argument can also prove the case of exponentially small soundness error in Theorem 1.5.

Acknowledgements. The authors are grateful to Richard Cleve, François Le Gall, Keiji Matsumoto, and Yasuhiro Takahashi for very useful discussions. Keisuke Fujii is supported by the Grant-in-Aid for Research Activity Start-up No. 25887034 of the Japan Society for the Promotion of Science. Hirotada Kobayashi and Harumichi Nishimura are supported by the Grant-in-Aid for Scientific Research (A) Nos. 24240001 and 16H01705 of the Japan Society for the Promotion of Science. Tomoyuki Morimae is supported by the Program to Disseminate Tenure Tracking System of the Ministry of Education, Culture, Sports, Science and Technology in Japan, the Grant-in-Aid for Scientific Research on Innovative Areas No. 15H00850 of the Ministry of Education, Culture, Sports, Science and Technology in Japan, and the Grant-in-Aid for Young Scientists (B) No. 26730003 of the Japan Society for the Promotion of Science. Harumichi Nishimura is also supported by the Grant-in-Aid for Scientific Research on Innovative Areas No. 24106009 of the Ministry of Education, Culture, Sports, Science and Technology in Japan, which Hirotada Kobayashi and Seiichiro Tani are

also grateful to. Harumichi Nishimura further acknowledges support from the Grant-in-Aid for Scientific Research (C) No. 25330012 of the Japan Society for the Promotion of Science. Part of the work of Shuhei Tamate was done while this author was at the RIKEN Center for Emergent Matter Science, Wako, Saitama, Japan.

References

1 Scott Aaronson. Quantum computing, postselection, and probabilistic polynomial-time. *Proceedings of the Royal Society A*, 461(2063):3473–3482, 2005. doi:10.1098/rspa.2005.1546.

2 Scott Aaronson and Alex Arkhipov. The computational complexity of linear optics. *Theory of Computing*, 9:143–252, 2013. doi:10.4086/toc.2013.v009a004.

3 Leonard M. Adleman, Jonathan DeMarrais, and Ming-Deh A. Huang. Quantum computability. *SIAM Journal on Computing*, 26(5):1524–1540, 1997. doi:10.1137/S0097539795293639.

4 Andris Ambainis and Rūsiņš Freivalds. 1-way quantum finite automata: strengths, weaknesses and generalizations. In *39th Annual Symposium on Foundations of Computer Science*, pages 332–341, 1998. doi:10.1109/SFCS.1998.743469.

5 Andris Ambainis, Leonard J. Schulman, and Umesh Vazirani. Computing with highly mixed states. *Journal of the ACM*, 53(3):507–531, 2006. doi:10.1145/1147954.1147962.

6 Elmar Böhler, Christian Glaßer, and Daniel Meister. Error-bounded probabilistic computations between MA and AM. *Journal of Computer and System Sciences*, 72(6):1043–1076, 2006. doi:10.1016/j.jcss.2006.05.001.

7 Michael J. Bremner, Richard Jozsa, and Dan J. Shepherd. Classical simulation of commuting quantum computations implies collapse of the polynomial hierarchy. *Proceedings of the Royal Society A*, 467(2126):459–472, 2011. doi:10.1098/rspa.2010.0301.

8 Daniel J. Brod. The complexity of simulating constant-depth BosonSampling. *Physical Review A*, 91(4):article 042316, 2015. doi:10.1103/PhysRevA.91.042316.

9 Joseph Emerson, Yaakov S. Weinstein, Seth Lloyd, and D. G. Cory. Fidelity decay as an efficient indicator of quantum chaos. *Physical Review Letters*, 89(28), 2002. doi:10.1103/PhysRevLett.89.284102.

10 Bill Fefferman, Hirotada Kobayashi, Cedric Yen-Yu Lin, Tomoyuki Morimae, and Harumichi Nishimura. Space-efficient error reduction for unitary quantum computations. In *Automata, Languages, and Programming, 43rd International Colloquium, ICALP 2016, Proceedings*, Leibniz International Proceedings in Informatics, 2016.

11 Keisuke Fujii, Hirotada Kobayashi, Tomoyuki Morimae, Harumichi Nishimura, Shuhei Tamate, and Seiichiro Tani. Power of quantum computation with few clean qubits. arXiv.org e-Print archive, arXiv:1509.07276 [quant-ph], 2015. arXiv:1507.05592.

12 Stephen P. Jordan and Gorjan Alagic. Approximating the Turaev-Viro invariant of mapping tori is complete for one clean qubit. In Dave Bacon, Miguel Martin-Delgado, and Martin Roetteler, editors, *Theory of Quantum Computation, Communication, and Cryptography, 6th Conference, TQC 2011, Madrid, Spain, May 24–26, 2011, Revised Selected Papers*, volume 6745 of *Lecture Notes in Computer Science*, pages 53–72. Springer-Verlag, 2014. doi:10.1007/978-3-642-54429-3_5.

13 Stephen P. Jordan and Pawel Wocjan. Estimating Jones and HOMFLY polynomials with one clean qubit. *Quantum Information and Computation*, 9(3–4):0264–0289, 2009.

14 Richard Jozsa and Maarten Van den Nest. Classical simulation complexity of extended Clifford circuits. *Quantum Information and Computation*, 14(7–8):0633–0648, 2014.

15 E. Knill and R. Laflamme. Power of one bit of quantum information. *Physical Review Letters*, 81(25):5672–5675, 1998. doi:10.1103/PhysRevLett.81.5672.

16 Greg Kuperberg. How hard is it to approximate the Jones polynomial? *Theory of Computing*, 11:183–219 (article 6), 2015. doi:10.4086/toc.2015.v011a006.

17 Tomoyuki Morimae, Keisuke Fujii, and Joseph F. Fitzsimons. Hardness of classically simulating the one-clean-qubit model. *Physical Review Letters*, 112(13):article 130502, 2014. doi:10.1103/PhysRevLett.112.130502.

18 Xiaotong Ni and Maarten Van den Nest. Commuting quantum circuits: Efficient classical simulations versus hardness results. *Quantum Information and Computation*, 13(1–2):0054–0072, 2013.

19 David Poulin, Robin Blume-Kohout, Raymond Laflamme, and Harold Ollivier. Exponential speedup with a single bit of quantum information: Measuring the average fidelity decay. *Physical Review Letters*, 92(17), 2004. doi:10.1103/PhysRevLett.92.177906.

20 David Poulin, Raymond Laflamme, G. J. Milburn, and Juan Pablo Paz. Testing integrability with a single bit of quantum information. *Physical Review A*, 68(2):article 022302, 2003. doi:10.1103/PhysRevA.68.022302.

21 D. J. Shepherd. Computation with unitaries and one pure qubit. arXiv.org e-Print archive, arXiv:quant-ph/0608132, 2006. arXiv:quant-ph/0608132.

22 Daniel James Shepherd. *Quantum Complexity : restrictions on algorithms and architectures*. PhD thesis, Department of Computer Science, Faculty of Engineering, University of Bristol, 2009. arXiv.org e-Print archive, arXiv:1005.1425 [cs.CC]. arXiv:1005.1425.

23 Peter W. Shor and Stephen P. Jordan. Estimating Jones polynomials is a complete problem for one clean qubit. *Quantum Information and Computation*, 8(8–9):0681–0714, 2008.

24 Yasuhiro Takahashi, Seiichiro Tani, Takeshi Yamazaki, and Kazuyuki Tanaka. Commuting quantum circuits with few outputs are unlikely to be classically simulatable. In *Computing and Combinatorics, 21st International Conference, COCOON 2015*, volume 9198 of *Lecture Notes in Computer Science*, pages 223–234, 2015. doi:10.1007/978-3-319-21398-9_18.

25 Yasuhiro Takahashi, Takeshi Yamazaki, and Kazuyuki Tanaka. Hardness of classically simulating quantum circuits with unbounded Toffoli and fan-out gates. *Quantum Information and Computation*, 14(13–14):1149–1164, 2014.

26 Barbara M. Terhal and David P. DiVincenzo. Adaptive quantum computation, constant depth quantum circuits and Arthur-Merlin games. *Quantum Information and Computation*, 4(2):134–145, 2004.

27 Seinosuke Toda. PP is as hard as the polynomial-time hierarchy. *SIAM Journal on Computing*, 20(5):865–877, 1991. doi:10.1137/0220053.

28 John Watrous. Quantum simulations of classical random walks and undirected graph connectivity. *Journal of Computer and System Sciences*, 62(2):376–391, 2001. doi:10.1006/jcss.2000.1732.

29 Edward Witten. Quantum field theory and the Jones polynomial. *Communications in Mathematical Physics*, 121(3):351–399, 1989. doi:10.1007/BF01217730.