Fault-Tolerant Quantum Solvability of Noisy Binary Linear Problem

Wooyeong Song,1,2 Youngrong Lim,3 Kabgyun Jeong,4,3 Jinho Yeon,2 Jung Jun Park,3 M. S. Kim,5,3 and Jeongho Bang6,7

1Center for Quantum Information, Korea Institute of Science and Technology, Seoul 02792, Korea
2Department of Physics, Hanyang University, Seoul 04763, Korea
3School of Computational Sciences, Korea Institute for Advanced Study, Seoul 02455, Korea
4Research Institute of Mathematics, Seoul National University, Seoul 08826, Korea
5QOLS, Blackett Laboratory, Imperial College London, London SW7 2AZ, United Kingdom
6Electronics and Telecommunications Research Institute, Daejeon 34129, Korea
(Received September 24, 2021)

The noisy binary linear problem (NBLP) is a known computationally intractable problem. Thus, NBLP offers primitives for post-quantum cryptography. An efficient quantum NBLP algorithm that exhibits a polynomial quantum sample and time complexities has recently been proposed. However, such a scenario cannot be accommodated because the preparation and use of a quantum sample would be costly and difficult when |G| is large. It could even offset the quantum speedup achieved. Although QC can make the fullest use of the quantum sample to efficiently solve NBLP, it has not been determined whether the intrinsic hardness of NBLP is reduced. Accordingly, the security level of the post-quantum cryptography has not been determined; thus, we are uncertain whether the NBLP hardness has been overcome.

In this Letter, we provide a complete analysis of the quantum solvability of the NBLP. To this end, we considered two essential and independent processes of the algorithm. One is the process of loading the samples (∈ G) into an entangled state |ψ⟩, which is denoted by P_{\psi}. The P_{\psi} is analysed to introduce a useful quantum gadget, called quantum random-access memory (QRAM). The other process involves the application of the main algorithm kernel P_A, which is an optimised set of elementary gate operations. Here, we analyse an extendable form of P_A, which can cover multiple problems, and apply the result to the binary setting. Subsequently, we analyse the number of repetitions of P_{\psi} + P_A required to determine the solution s, where the exponential reduction in the quantum-sample complexity (argued in Refs. [5][8]) is (re)derived and discussed, more intensively. Such an analysis allows us to account for the overall resource-consuming aspect, thereby facilitating a proper discussion of the quantum solvability of NBLP.

The analysis was conducted in the context of fault-tolerant QC. Here, we consider the Clifford + T library, assuming that an effective quantum error-correction code is embedded. We then minimise the overall number of gate layers, particularly those of T (or T†) gates, often called T-depth complexity. Because T and T† are
much more costly to implement than any Clifford gates, the T-depth is often used to approximate the time complexity of a quantum circuit \[11–13\]. In this context, we define a reasonable cost \( C \) in terms of the computational time as

\[
C \equiv (T\text{-depth of } P_\psi + T\text{-depth of } P_A) \times S,
\]

where \( S \) denotes the number of repetitions of \( P_\psi + P_A \) for the completion of the algorithm.

In this letter, we analyse the (I) T-depth of \( P_\psi \), (II), T-depth of \( P_A \), and (III) repetitions \( S \). We realise that (I), (II), and (III) are closely related, and thus they are not separately studied for NBLP. We also evaluated the number of logical qubits required to complete \( P_\psi \) and \( P_A \), which is termed the width. Finally, \( C \) is estimated, and the quantum solvability of the NBLP is discussed.

Algorithm overview.— We briefly outline the entire procedure of the quantum NBLP algorithm.

(A.1) A state \( |\psi\rangle \) of a quantum sample is prepared in the following form: \( |\psi\rangle = \frac{1}{\sqrt{2^n}} \sum_i |a_i b_i\rangle \), where the summation \( \sum_i \) includes only the inputs in \( \mathcal{S} \), and \( q = \lceil \log_2(\mathcal{S}) \rceil \leq n \) \[27\], which can be regarded as the factor that determines the “superposition size” of \( |\psi\rangle \).

(A.2) Given a quantum sample \( |\psi\rangle \), we run \( P_A \). Formally, \( P_A \) is given as the Bernstein–Vazirani (BV) kernel:

\[
P_A = \text{QFT}^{2^n-1},
\]

where \( \text{QFT} \) is the d-dimensional quantum Fourier transform (QFT): \( \text{QFT}_d |j\rangle = \frac{1}{\sqrt{d}} \sum_k e^{2\pi i kj/d} |k\rangle \). In NBLPs, \( P_A \) becomes \( \text{QFT}^{2^n-1} = \hat{H}^{2^n-1} \), where \( \hat{H} \) is the Hadamard transform: \( |j\rangle \rightarrow \frac{1}{\sqrt{2^n}} \sum_k e^{2\pi i jk/n} |k\rangle \) \((j, k = 0, 1)\). The output state \( \hat{H}^{2^n-1} |\psi\rangle \) is then expressed as

\[
\frac{1}{\sqrt{2^n+1}} \sum_i \sum_k (-1)^{n(k+sk^*)+r_a k^*} |k\rangle |k^*\rangle,
\]

where \( k \in \{0, 1\}^n \) and \( k^* \in \{0, 1\} \).

(A.3) We measure the qubit state \( |k^*\rangle \). Here, if we measure \( k^* = 0 \), no information on \( s \) can be retrieved from the remaining state,

\[
\frac{1}{\sqrt{2^n+q}} \sum_i \sum_k (-1)^{n(k+sk^*)+r_a k} |k\rangle,
\]

and the failure is returned. Otherwise (i.e. if \( k^* = 1 \)), we obtain the remaining state

\[
\frac{1}{\sqrt{2^n+q}} \sum_i \sum_k (-1)^{n(k+sk^*)+r_a} |k\rangle.
\]

By solving Eq. (5), we obtain the candidate \( k \). Here, the true solution \( s \) can be measured (i.e. \( k = s \)) with a certain probability, which is expressed as \( P(k = s|k^* = 1) \) \[28\].

(A.4) Repeating (A.1)–(A.3), we determine the most frequently measured \( k \) as the true solution \( s \), which is referred to as “majority-voting.” The condition of majority-voting is analysed later. Additional details are provided in the Supplementary Material.

Analysis (I): Resource counts for \( P_\psi \).—Our idea of understanding and analysing \( P_\psi \) (or step (A.1)) of NBLP refers to the RAM used in classical computing. For a given database, for example, \( \mathcal{S} \) in our case, the RAM reads a memory location specified by an address, such as \( \gamma \), and returns the data, for example, \( D_{\gamma} \) \[14\].

A quantum version of RAM, called QRAM, can access the target address and output the allocated data, such that \[9\]

\[
\frac{1}{\sqrt{|\mathcal{S}|}} \sum_{\gamma \in \mathcal{S}} |\gamma\rangle |\text{null}\rangle \rightarrow \frac{1}{\sqrt{|\mathcal{S}|}} \sum_{\gamma \in \mathcal{S}} |\gamma\rangle |D_{\gamma}\rangle,
\]

where \( |\gamma\rangle \) denotes the address, \( |\text{null}\rangle \) is the null state, and \( \mathcal{S} \) denotes the space of the addresses. Now, let us consider how such a process can be used to prepare \( |\psi\rangle \):
First, we set \( \{D_{\gamma}\} = \{(a, b_a)_{\gamma}\} \) for all samples in \( \mathcal{S} \), where \( (a, b_a)_{\gamma} \) is a sample allocated by \( \gamma \). By considering \( |\mathcal{S}| = |\mathcal{S}| = 2^n \), the address symbol \( \gamma \) can be expressed as a q-tuple of binary number: \( \gamma = \gamma_0 \gamma_1 \ldots \gamma_{q-1} \), where \( \gamma_j \in \{0, 1\} \) for all \( j = 0, 1, \ldots, q - 1 \). Then, Eq. (7) yields the address-and-data entangled state as

\[
|\Psi\rangle = \frac{1}{\sqrt{2^n}} \sum_{\gamma \in \mathcal{S}} |\gamma\rangle |(a, b_a)_{\gamma}\rangle.
\]

After decoupling the address and data, we can retrieve \( |\psi\rangle \) \[29\]. Here, we assume that decoupling (sometimes, called a “fan-in”) is possible and the resource required is at most the same as those for coupling in Eq. (7). Furthermore, the symbols of \( \gamma \) can be “incorporated into” or “synchronised with” those of \( a \). Thus, we focus on Eq. (7) in the analysis of \( P_\psi \).

An efficient implementation of Eq. (7) is applied to a scheme of the bucket-brigade QRAM because it prevents the usual overhead \[30\]. For some small-scale memories, the bucket-brigade QRAM leads to considerable resource savings, which renders the quantum speedups tangible \[7\]. However, the realisation of such savings will not be substantial in a large-scale memory because the bucket-brigade QRAM is vulnerable to errors. Moreover, the logical-qubit size and T-depth complexity are both \( O(2^n) \) \[15\]. Thus, the application of recent state-of-art techniques associated with QRAM \[31\] to implement \( P_\psi \) for NBLP is yet to be determined, and it remains unclear whether it will considerably reduce the T-depth.

Herein, our first result can be stated as

Resource Estimation (RE) 1. Resource counts for implementing \( P_\psi \) are as follows:

\[
W_{P_\psi} = O(2^{\log n+q}) \quad \text{and} \quad T_{D,P_\psi} = O(nq),
\]
where $W_{P_{\psi}}$ and $T_{D,P_{\psi}}$ represent the number of logical qubits and $T$-depth of $P_{\psi}$, respectively.

Notably, $T_{D,P_{\psi}}$ can be a polynomial in $n$. Our concrete design recipes for $P_{\psi}$ and $T$-depth optimisation for NBLP are described in the Supplementary Material.

Analysis (II): Resource counts for $P_A$.—Next, we consider the resource for $P_A$. Here, by considering the formal definition of the BV kernel (as in Eq. (7)), we start by investigating the $T$-depth of an arbitrary l-qubit QFT. Usually, the quantum circuit for an l-qubit QFT can be synthesised with the controlled-$\hat{R}_k$ gates and $\hat{H}$, where $\hat{R}_k$ denotes the single-qubit rotation: $\hat{R}_k = |0\rangle \langle 0| + e^{i\pi\theta_k} |1\rangle \langle 1|$. Typically, an ideal QFT circuit requires $\frac{l(l-1)}{2} = O(l^2)$-controlled-$\hat{R}_k$ gates with $\hat{H}^\otimes l$, with $\theta_k = 2^{-k}$ $(k = 1, 2, \ldots, l - 1)$. In practice, however, an l-qubit QFT is implemented within a small fixed error $\Delta$, with $\theta_k = 2^{-k}$ $(k = 1, 2, \ldots, l)$ satisfying $2 \leq \beta \leq l - 1$. Thus, the (so-called) approximate-QFT (AQFT) is constructed based on $\frac{(2l-\beta)(l-1)}{2} = O(l^2)$ controlled-$\hat{R}_k$ gates. However, the condition $\beta < l - 1$ implies that a specific error $\Delta$ is unavoidable because the rotation angle $\theta_k$ that is smaller than the threshold value $2^{-\beta}$ is discarded, limiting the choice of $\beta$. The lower bound of each of $\beta$ is known as $O(\log l)$ (Chap. 5 of Ref. [16] for the fundamentals).

To realise an l-qubit AQFT circuit in a fault-tolerant manner, $\beta = O(\log l)$ can be considered. Then, all controlled-$\hat{R}_k$ gates with $\theta_k \leq 2^{-O(\log l)}$ are discarded with the error bounded by $\Delta$, and the controlled-$\hat{R}_k$ gate counts are reduced from $O(l^2)$ to $O(l\log \frac{l}{\Delta})$. The remaining controlled-$\hat{R}_k$ gates are decomposed into Clifford+$T$ gates, where the fault-tolerance overhead is involved. Consequently, we can obtain an l-qubit AQFT circuit featured by $O(l\log \frac{l}{\Delta} \times \log \frac{\log l}{\Delta})$ number of $T$ (or $T'$) gates, which allows the $T$-count of $O(l \log^2 l)$. For all effective QC (specifically, for $\Delta \geq (2^{-l})$, we can neglect the dependence on $\Delta$, Consequently, by noting that $T$-depth is upper bounded by $T$-count in general, we obtain [32].

$$T_{D,AQFT} \leq T_{C,AQFT} = O \left( l \log^2 l \right),$$

where $T_{C,AQFT}$ denotes the $T$-count of l-qubit AQFT. This result is well appreciated in a fault-tolerant manner.

Based on the above analysis, we obtained the second result:

Resource Estimation (RE) 2. We can implement $P_A$ in NBLP, with

$$W_{P_A} = O(n) \text{ and } T_{D,P_A} = N/A.$$  (11)

The estimation can be validated as follows: In NBLP (i.e. a binary problem), $P_A$ is only the $(n + 1)$-fold product of the Hadamard transform, i.e. $P_A = \text{QFT}^\otimes n + 1 = \hat{H}^\otimes n + 1$. Thus, the number of logical qubits is $n + 1$. Although the circuit may run with some additional ancilla qubits, $W_{P_A}$ scales as $O(n)$. Subsequently, it implies zero $T$-depth complexity, because controlled $\hat{R}_k$ gates are not required. Hence, RE[2] holds. This result is a straightforward consequence of $P_A = \hat{H}^\otimes n + 1$. However, the analysis of AQFT would be useful, particularly when the BV kernel is applied to a general problem setting, such as a multinary problem [33]. Hence, both $W_{P_A}$ and $T_{D,P_A}$ are generally polynomially bounded in $n$.

Majority-voting conditions.—Prior to analysing (III), we derive the condition of majority-voting [performed in (A.4)]. First, we calculate the probability ($P_S$), where $k$ measured at (A.3) is equal to the true solution $s$. By substituting $k = s$ into Eq. (6), we can calculate $P_S$ as

$$P_S = P(k = s | k^* = 1)P(k^* = 1) = \left\| \frac{1}{\sqrt{2^n + 1}} \sum_a \omega_{s,a}^{(2n) + c_s} |s\rangle \right\|^2 = \frac{1}{2^{n-q+1}} \left\| \sum_a \omega_{s,a}^{(1)\gamma_s} \right\|^2,$$

where $P(k^* = 1) = \frac{1}{2}$. Here, we cast a useful concentration bound, the so-called Chernoff–Hoeffding (CH) inequality [18]: For $t \ll O(2^n)$,

$$P \left( |\overline{U} - \mathbb{E}(\overline{U}_a) | \geq t \right) \leq 2e^{-\frac{2t^2}{2^n}},$$

where $\overline{U}_a = (−1)^{s,a}, \overline{U} = \frac{1}{2^n} \sum_a \overline{U}_a$, and $\mathbb{E}(\overline{U}_a)$ denotes the expectation of $\overline{U}_a$. If we assume that the order of $q$ is larger than $O(\log_2 n)$, the right-hand side term in Eq. (13) is negligible, and $P \left( |\overline{U} - \mathbb{E}(\overline{U}_a) | \geq t \right) = 0$ for a large $n$. Note that we have used the definition [D]: If a factor is as small as $O(e^{-n})$, the factor can be negligible for a large $n$ and set to zero. Accordingly, we obtain the following expression:

$$|\overline{U} - \mathbb{E}(\overline{U}_a) | < t,$$

and using Eqs. (12) and (14), we can obtain the lower bound of $P_S$ such that

$$P_S = \frac{1}{2^{n-q+1}} |\overline{U}|^2 > P_{S,\text{inf}} = \frac{1}{2^{n-q+1}} |2\eta - t|^2,$$

where we use $\mathbb{E}(\overline{U}_a) = (\frac{1}{2} + \eta) - (\frac{1}{2} - \eta) = 2\eta$.

Next, we consider the probability $P_F = P(k \neq s)$, where the measured $k$ is not equal to the solution $s$. For convenience, $P(k = \tilde{s})$ denotes $P(k \neq s)$, where $\tilde{s} = s + \phi$. $\phi = \phi_\ell \phi_{\ell-1} \cdots \phi_{n-1}$ is an arbitrary $n$-tuple of binary numbers $\phi_j \in \{0, 1\}$, except for $\phi = 00\cdots 0$. Then, from Eq. (6), $P_F$ is calculated as

$$P_F = \frac{1}{2^{n-q+1}} \left\| \sum_a \omega_{s,a}^{(1)\gamma_s} \right\|^2,$$

where $\gamma_s = q_s(\phi + a)$.
Here, we recall the CH inequality in Eq. (13) by letting $U_a = (-1)^{a \Phi + \epsilon_s}$ and $F = \frac{1}{2} \sum_a U_a$. It should be noted that, in this case, $E(U_a) = 0$ because $a \Phi$ and $a \Phi + \epsilon_s$ are either 0 or 1 with probability $\frac{1}{2}$. Subsequently, because $O(q)$ is larger than $O(\log_2 n)$ and $e^{-\frac{1}{2} \pi n^2}$ is negligible owing to the definition [D], $P(\|F\| \geq t) = 0$. Thus, we have
\[ \|U\| < t. \]
(17)
By using Eqs. (16) and (17), the upper bound for $P_F$ is obtained as follows:
\[ P_F = \frac{1}{2^{n-q+1}} \|U\|^2 < P_{F,\text{sup}} = \frac{1}{2^{n-q+1}} |t|^2. \]
(18)
Next, we specify the conditions required for the majority-voting to work:
\[ P_{S,\text{inf}} > P_{F,\text{sup}} \iff t < \eta. \]
(19)
This condition should be satisfied; otherwise, we cannot completely rule out the possibility of identifying a “false” solution $\tilde{s}$ in (A.4).

Analysis (III): Number of repetitions $S$.—Finally, we analyse the number of repetitions $S$. Let us assume that a candidate solution $k$ is obtained, completing (A.1)–(A.3). The process is then repeated until $M$ candidates are collected, and finally the most frequently occurring $k$ is chosen among $M$ at (A.4). Here, we let $x_k = 1$ (or $x_k = 0$) when the true solution $s$ (or a false solution $\tilde{s}$) is measured after (A.1)–(A.3). Subsequently, let $X$ be the number of times that the true solution $k = s$ is measured among $M$. Thus, $X = \sum_{k=1}^{M} x_k$. Because all values of $x_k$ are independent. In such a setting, we can use a statistical inequality, i.e. the Chernoff bound [19]:

For any $\epsilon > 0$,
\[ P(\|X - \mu\| \geq \epsilon \mu) \leq 2e^{-\frac{\epsilon^2}{2} \mu}, \]
(20)
where $\mu = E(\mathbb{1}_{k=s}) = MP_S$, and $\mathbb{1}_{k=s}$ is the indicator function of $k = s$. By letting $2e^{-\frac{\epsilon^2}{2} \mu} \leq \delta$ with $\delta \in (0, 1]$, we can derive the following theorem:
\[ P(\|X - P_S\| \geq \epsilon') \leq \delta \text{ iff } M \geq \frac{3}{\epsilon'^2} \ln \frac{2}{\delta}, \]
(21)
where $X = \frac{3}{\epsilon'^2} \ln \frac{2}{\delta}$ samples, $X$ can be estimated within the interval $[P_S - \epsilon', P_S + \epsilon']$ with a probability of at least $1 - \delta$. This is sometimes referred to as the sampling theorem. Therefore, noting that the Chernoff bound gives the minimal (Bayesian) error probability when discriminating the “a priori” and “observations”, the sampling theorem translates into the following statement: The majority-voting allows us to identify the true solution $s$ with at least $M = \frac{3}{\epsilon'^2} \ln \frac{2}{\delta}$ repetitions of (A.1)–(A.3), if the following condition is satisfied [25].

\[ \epsilon' < P_{S,\text{inf}} - P_{F,\text{sup}}. \]
(22)

Then, by noting that $S$ is the number of repetitions of (A.1)–(A.3), we achieve our third result:

**Resource Estimation (RE) 3.** Given the constants $t$, $\epsilon$, and $\delta$, the number of repetitions $S$ is expressed as
\[ S = O\left(4^n - q^2 |2\eta - t|^{-1} \ln \delta^{-1}\right), \]
(23)
where we assume that $S = 2M$ [36]. The following crucial conditions should be satisfied:
\[ t < \eta \text{ and } \epsilon < 1 - \frac{P_{F,\text{sup}}}{P_{S,\text{inf}}}, \]
(24)
where the former is acquired from the majority-voting condition in Eq. (79), and the latter is derived using $\epsilon' = \epsilon P_S \geq P_{S,\text{inf}}$ and Eq. (22).

Note that $P_{A(\psi)}$ boots up only when $P_A$ runs with a single usage of $|\psi\rangle$, and it is straightforward that $S$ corresponds to the quantum-sample complexity. Accordingly, RE [3] shows that the reduction in complexity depends on the size of the superposition, that is, $|S| = 2^n$. For example, if we use the fullest (exponential-scale) superposition of the sample with $|S| = 2^n$ (or equivalently, $q = n$), $S$ becomes $O(\epsilon^{-2} |2\eta - t|^{-1} \ln \delta^{-1})$, which is consistent with the results of Refs. [5, 6]. By contrast, the opposite extreme case can also be considered, that is, using a non-superoposed sample $|\psi\rangle = |a\rangle |b_a\rangle$ with $|S| = 1$ (or equivalently, $q = 0$), which still allows quantum parallelism to be processed by the BV kernel. However, in this case, $P_S$ becomes exponentially small with $n$ [ Eq. (15)] and is negligible (based on the definition [D]). Thus, a majority-voting condition cannot be established. Moreover, the order of $q$ is at least $O(\log_2 n)$ [37].

**Results.**—The results of RE [1] and [5] can draw the following conclusion: The cost $C$, defined in Eq. (2), can be a polynomial for the problem size $n$. To achieve this, however, the amount of superposed samples in $|\psi\rangle$ should be exponentially large with $n$; thus, $|S|$ must be $O(2^n)$ with $q = O(n)$. This suggests an exponential scale of the circuit width (RE [1]). If the superposition size of $|\psi\rangle$ is small, for example, if $|S| = O(n)$ (or $q = o(\log n)$), the polynomial scaling of $C$ will be impractical, whereas the number of logical qubits can be polynomial in $n$.

Another emerging insight is the possibility of a depth-versus-width trade-off in NBLP. In our study, such a trade-off is specified by RE [1] and [5]. If this trade-off is intrinsic to the NBLP, the quantum solvability may be limited. We provide some technical details of this in the Supplementary Material.

Further improvement can be achieved by developing a more efficient error-correcting code or QRAM scheme, thus lowering the level of noisy physical qubits.
Acknowledgements.—W.S. and J.B. thank Nana Liu for the discussions. This work was supported by the National Research Foundation of Korea (Nos. 2021M3E4A1038213, 2021R1I1A1A01042199, 2020M3E4A1077861, and 2019M3E4A1079666), and the Ministry of Science, ICT and Future Planning (MSIP) by the Institute of Information and Communications Technology Planning and Evaluation grant funded by the Korean government (No. 2020-0-00890, “Development of technology Planning and Evaluation grant, KIAS visiting professorship, and EPSRC Quantum Computing and Simulations Hub grant.

* The first three authors contributed equally to this work
† Electronic address: m.kim@imperial.ac.uk
‡ Electronic address: jbang@etri.re.kr

[1] L. N. Trefethen and D. Bau III, *Numerical linear algebra*, vol. 50 (Siam, 1997).
[2] A. W. Harrow, A. Hassidim, and S. Lloyd, Physical Review Letters 103, 150502 (2009).
[3] A. Blum, A. Kalai, and H. Wasserman, Journal of the ACM (JACM) 56, 34 (2009).
[4] O. Regev, Journal of the ACM (JACM) 50, 506 (2003).
[5] A. W. Cross, G. Smith, and J. A. Smolin, Physical Review A 92, 012327 (2015).
[6] A. B. Grilo, I. Kerenidis, and T. Zijlstra, Physical Review A 99, 032314 (2019).
[7] S. Aaronson, Nature Physics 11, 291 (2015).
[8] V. Giovannetti, S. Lloyd, and L. Maccone, Physical Review A 78, 052310 (2008).
[9] V. Giovannetti, S. Lloyd, and L. Maccone, Physical Review Letters 100, 160501 (2008).
[10] M. Amy, D. Maslov, M. Mosca, and M. Roetteler, IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems 32, 818 (2013).
[11] X. Zhou, D. W. Leung, and I. L. Chuang, Physical Review A 62, 052316 (2000).
[12] A. G. Fowler, A. M. Stephens, and P. Groszowski, Physical Review A 80, 052312 (2009).
[13] M. Howard and E. Campbell, Physical Review Letters 118, 090501 (2017).
[14] R. C. Jaeger and T. N. Blalock, Microelectronic circuit design (McGraw-Hill New York, 1997).
[15] S. Arunachalam, V. Gheorghiu, T. Jochym-O’Connor, M. Mosca, and P. V. Srinivasan, New Journal of Physics 17, 123010 (2015).
[16] M. A. Nielsen and I. L. Chuang, *Quantum computation and quantum information* (Cambridge University Press, 2000).
[17] A. Barenco, A. Ekert, K.-A. Suominen, and P. Törmä, Physical Review A 54, 139 (1996).
[18] W. Hoeffding, in *The collected works of Wassily Hoeffding* (Springer, 1994), pp. 409–426.
[19] M. Mitzenmacher and E. Upfal, *Probability and computing: Randomization and probabilistic techniques in algorithms and data analysis* (Cambridge university press, 2017).
[20] E. Tang, Physical Review Letters 127, 060503 (2021).
[21] O. Di Matteo, V. Gheorghiu, and M. Mosca, IEEE Transactions on Quantum Engineering 1, 1 (2020).
[22] A. Paler, O. Oumarou, and R. Basmadjian, Physical Review A 102, 032608 (2020).
[23] P. Selinger, Physical Review A 87, 042302 (2013).
[24] H. Goto, Physical Review A 90, 052318 (2014).
[25] Y. Nam, Y. Su, and D. Maslov, NPJ Quantum Information 6, 1 (2020).
[26] A result from a more extreme perspective has recently been reported [20].
[27] Here, $[x]$ is the ceil function of $x$ being the smallest number greater than or equal to $x$.
[28] The most exact form of the probability is $P(k = s) = 1, (e_{a})$. However, we drop the dependence on $(e_{a})$ because the errors occur completely at random.
[29] Note that the summation $\sum_{a}^s$ can be replaced by $\sum_{a \in \mathbb{R}}$ because each input $a$ can be matched to its corresponding address $\gamma$. Accordingly, the space $\mathbb{R} \subseteq \{0, 1\}^n$ is equivalent to the space of the inputs in $S$.
[30] The design of conventional (multi-bit) QRAM demands a large number of switching operations (c.f., a transistor in classical DRAM) to run a QRAM. This is because a processor should seek the addresses (or registers) of the corresponding data and decode them by repeating the logic operations. Such a large cost indicates a slow speed, high energy usage, and high decoherence rate during decoding. More specifically, a single run of $\log_2 n$-bit QRAM requires $O(N^{1/d})$ operations with a $d$-dimensional memory array (for more details, see Ref. [9] or Chap. 8 in Ref. [13]).
[31] For example, gate-level parallelisation [21, 22] and/or “one T-depth, seven qubits” Toffoli [23].
[32] Here, we indicate that a semi-classical AQFT can reduce the $T$-count complexity from $O(n \log^2 n)$ to $O(n \log n)$ [23]. Very recently, Nam et al. proposed a fully coherent AQFT that can have a $T$-count of $O(n \log n)$ [24].
[33] More specifically, such a problem can be generalised using $a, s \in \{0, 1, \ldots, 2^d - 1\}^n$. In this case, the error is drawn from a discrete distribution $\mathcal{B}$. Here, we consider a slightly weaker bound. The tight bound is given by $M \geq \frac{2^{\gamma} - 1}{\gamma} \ln \frac{\gamma}{\epsilon}$. Note that $P_{\text{err}} - P_{\text{err}}$ is larger than 0 owing to the majority-voting condition in Eq. [16].
[34] This is because half of the trials of (A.1)–(A.3) will return a failure with $k^* = 0$. Note that factor 2 has no influence on the order of $S$.
[35] If $q = O(\log_2 n)$, the polynomial quantum-sample complexity cannot be achieved, that is, $S = O(4^n - \log n)$.
S1. FURTHER DETAILS ON THE ALGORITHM

A. In the absence of noise: Linear function learning

To understand the operation of the algorithm, let us consider the case of no noise, which is often referred to as “linear function learning.” Given the sample state with \( e_a = 0 \) (or equivalently, \( \eta = -\frac{1}{2} \)),

\[
|\psi\rangle = \frac{1}{\sqrt{2^q}} \sum_a' |a\rangle |a \cdot \mathbf{s} \pmod{2}\rangle,
\]

(S25)

the QFTs are applied, such that

\[
\left( \text{QFT}_{d=2} \otimes \text{QFT}_{d=2} \otimes \cdots \otimes \text{QFT}_{d=2} \otimes \text{QFT}_{d=2} \right) |\psi\rangle.
\]

(S26)

where QFT\(_{d=2}\) is the Hadamard transform: \( |j\rangle \rightarrow \frac{1}{\sqrt{2}} \sum_k (-1)^{jk} |k\rangle \) \((j,k = 0,1)\). The output state is expressed as follows:

\[
\text{QFT}_{d=2}^{\otimes n+1} |\psi\rangle = \frac{1}{\sqrt{2^{q+n}}} \sum_a' \left[ \left( \frac{1}{\sqrt{2}} \sum_{k_0 \in \{0,1\}} (-1)^{a_0 k_0} |k_0\rangle \right) \otimes \left( \frac{1}{\sqrt{2}} \sum_{k_1 \in \{0,1\}} (-1)^{a_1 k_1} |k_1\rangle \right) \right.
\]

\[
\otimes \cdots \otimes \left( \frac{1}{\sqrt{2}} \sum_{k_{n-1} \in \{0,1\}} (-1)^{a_{n-1} k_{n-1}} |k_{n-1}\rangle \right) \left. \otimes \left( \frac{1}{\sqrt{2}} \sum_{k^* \in \{0,1\}} (-1)^{(a \cdot \mathbf{s}) k^*} |k^*\rangle \right) \right]
\]

(S27)

\[
= \frac{1}{\sqrt{2^{q+n+1}}} \sum_a' \sum_{k \in \{0,1\}^n} \sum_{k^* \in \{0,1\}} (-1)^{a \cdot (k+s k^*)} |k\rangle |k^*\rangle.
\]

Subsequently, we measured the state \(|k^*\rangle\). If \(k^* = 1\) is measured using the delta function

\[
\delta_{k_j,-s_j} = \frac{1}{d} \sum_{a_j = 0}^{d-1} \omega^{a_j (k_j + s_j)},
\]

(S28)

we can achieve the final state as the true solution:

\[
|k\rangle = |s_0 s_1 \cdots s_{n-1}\rangle,
\]

(S29)

where \( \omega = e^{i \frac{2\pi}{d}} = (-1) \) with \( d = 2 \), and the probability amplitude \( \frac{1}{\sqrt{2}} \) is eliminated by the measurement of \(|k^*\rangle\).

For a simpler analysis, we assume \( q = n \) (hence, \( \sum_a' = \sum_{a \in \{0,1\}^n} \)). If \( k^* = 0 \) is measured, we cannot retrieve any information of \( \mathbf{s} \); that is, the algorithm returns a failure.

B. In the presence of noise: NBLP

Given the sample state with non-zero noise \( \eta \neq 0 \), i.e.

\[
|\psi\rangle = \frac{1}{\sqrt{2^q}} \sum_a' |a\rangle |a \cdot \mathbf{s} + e_a \pmod{2}\rangle,
\]

(S30)
the \( n + 1 \) QFTs were applied as described above. We then attain the following output state:

\[
\text{QFT}^{d}_{d=2} |\psi\rangle = \frac{1}{\sqrt{2^{n+1}}} \sum_{a} \left[ \left( \frac{1}{\sqrt{2}} \sum_{k_{0} \in \{0,1\}} (-1)^{a_{0}k_{0}} |k_{0}\rangle \right) \otimes \left( \frac{1}{\sqrt{2}} \sum_{k_{1} \in \{0,1\}} (-1)^{a_{1}k_{1}} |k_{1}\rangle \right) \otimes \cdots \right] \otimes \left( \frac{1}{\sqrt{2}} \sum_{k^{*} \in \{0,1\}} (-1)^{(a_{n}s + e_{n})k^{*}} |k^{*}\rangle \right)
\]

\[
= \frac{1}{\sqrt{2^{n+1}}} \sum_{a} \sum' (\omega^{a_{n}s + e_{n}}) (-1)^{a_{n}(k + sk^{*})} |k\rangle |k^{*}\rangle,
\]

which is equal to Eq. (4) in the main manuscript. Note that we cannot use the delta function in Eq. (S28) because \( |k\rangle \) and \( |k^{*}\rangle \) are not perfectly correlated, as in Eq. (S27), with the error term \( e_{n}k^{*} \). Thus, Eq. (S31) allows a candidate \( k = \tilde{s} \), which is generally not equal to the true solution \( s \). Thus, we can calculate the success probability, which is denoted as \( P_{\text{S}} = P(k = s) \), where \( k \) is equal to \( s \) by substituting \( k = sk^{*} \) into Eq. (S31), i.e.,

\[
P(k = s) = \frac{1}{2^{n+q}} \left| \sum_{a} \sum' (\omega^{a_{n}s + e_{n}}) |sk^{*}\rangle |k^{*}\rangle \right|^2
\]

\[
= \frac{1}{2^{n+q}} \sum_{k^{*}} \left| \sum_{a} \omega^{e_{n}k^{*}} \right|^2 \left| \langle sk^{*}|s\rangle \right|^2 \left| \langle k^{*}|1\rangle \right|^2
\]

\[
= \frac{1}{2^{n+q+1}} \left| \sum_{a} (-1)^{e_{n}} \right|^2
\]

\[
(S32)
\]

where we use \( \left| \langle k^{*}|1\rangle \right|^2 = \frac{1}{2} \). This is equal to Eq. (12) in the main manuscript.

S2. OUR DESIGNING RECIPES OF \( P_{\psi} \)

A. Notion of QRAM

RAM is in high demand as a component of high-speed digital computers. A similar operating unit is required for a quantum computer. This was first suggested previously [1] and dubbed as QRAM, which facilitates the entanglement between the address and data registers such that

\[
\frac{1}{\sqrt{|N|}} \sum_{\gamma} |\gamma\rangle_A |\text{null}\rangle_D \rightarrow \frac{1}{\sqrt{|N|}} \sum_{\gamma} |\gamma\rangle_A |D_\gamma\rangle_D,
\]

(S33)
where $\sum \gamma \sqrt{1/3!} |\gamma\rangle$ represents the superposed address qubits, and $|D_\gamma\rangle$ is the state of the $\gamma$-th data register.

QRAM operates in a similar manner as digital RAM, which consists of the main internal components: I/O registers and memory arrays \[2\]. However, QRAM exploits qubits to construct the components. More specifically, the I/O registers can be formed using fully quantum-controlled qubits, whereas the memory arrays can be accessible to classical (or non-superposed) data. Throughout this study, we assume the classical data $D_\gamma$ \[1,3,4\].

**B. Our design and analysis of $P_{|\psi\rangle}$**

Our key idea in designing a model of $P_{|\psi\rangle}$ is to borrow the scheme of a (so-called) bucket-brigade QRAM along with the (classical) data memory cells. Here, we first consider a quantum circuit for *single-bit data* as a unit of the general $n$-bit data structure. The circuit involves several types of qubits for addresses, ancillae, and data, each of which is indexed by $AD$, $A$, and $D$, respectively. The circuit runs the data loading using a three-fold process: coupling-query-decoupling (or fan-out, query, and fan-in using the terminology of RAM). In our circuit, a router is defined as a collection of switches attached to the $j$-th address qubit, and is implemented using CNOT and Toffoli gates \[4\]. In any $j$-th ($q \geq j \geq 2$) stage of the router, a total of $2^{j-1}$ switches are used. Therefore, when $q$ address qubits are introduced, the total number of routers $N_{\text{router}}$ is evaluated by counting the number of stages $N_{\text{router}} = \sum_{j=2}^{q} 2^{j-1}$. Note that $N_{\text{outer}}$ is equal to the number of CNOT and Toffoli gates, which will be used to analyse the cost of $P_{|\psi\rangle}$ later. Fig. S2 illustrates the proposed circuit, where the decoupling part is omitted because the structure of the decoupling is equivalent to that of the coupling part. Here, the number of exploited qubits for the address and memory cells are assumed to be $q$ and $2^q$, respectively. Usually, the size of $q$ is finite and smaller than $n$.

Subsequently, we analyse the cost of implementing $P_{|\psi\rangle}$ in a fault-tolerant manner. We evaluate the $T$-depth complexity (i.e. the number of layers of $T$ gates) by adopting the Clifford+$T$ decomposition of the circuit in Fig. S2. The first theorem is presented as follows.

**Theorem 1.** The $T$-depth for implementing $P_{|\psi\rangle}$ can be a polynomial in $q$ with $O(2^{\log n + q})$ (logical) qubits.

**Proof.** Considering that the resource count of decoupling can be assumed to be identical to the coupling, we analysed the $T$-depth complexity for the coupling and query. First, we consider a single-bit data circuit.

**coupling**—The circuit can be constructed using CNOT and Toffoli gates, which constitute routers. To minimise the $T$-depth of the circuit, a parallelisation of the stages of the switches is considered (Fig. S3). Although various schemes for gate decomposition can be implemented \[3,7\], here we consider the (so-called) “seven $T$-count and four $T$-depth” scheme, where the Toffoli gates can be parallelised by sharing one of the control qubit channels. In this manner, every Toffoli gate in any $j$-th stage can be arranged in a single channel line. Based on this idea, we can parallelise the $T$ (or $T^\dagger$) gates. Consequently, the switches have $2^{j-1} \times 7$ number of $T$ gates; however, the $T$-depth is still only 4. Therefore, for $q$-qubit addresses, the coupling circuit exhibits a $T$-depth of $4(q - 1)$.
FIG. S3: Parallelisation of Toffoli gates.

FIG. S4: Parallelisation of data qubit sharing Toffoli gates in (b) and its Clifford+$T$ decomposition in (c). We denote the extra qubits for parallelisation as $E_1$ and $E_2$.

*Query*—The query process uses $2^n$ Toffoli gates, whose control channel wires are over an address qubit and a memory cell (Fig. S4a). In our case, the Toffoli gates share control channel wires attached to the data qubit. To avoid the restriction of parallelisation owing to the control qubit sharing of the Toffoli gates, we use the extra qubits (denoted by $E_1$ and $E_2$) for parallelisation, as shown in Fig. S4(b). Then, by using the ancilla for the address qubits (excluding the last address qubit), every Toffoli gate can be parallelised. Such a scheme immediately leads to parallelisation of the $T$ gates, as shown in Fig. S4(c). Consequently, the $T$-depth of the query can be optimised as $O(1)$.

Now, we can generalise the architecture of $P|\psi\rangle$ for $n$-bit data, such that the single-bit data circuit runs sequentially.
for each data bit. Thus, the $T$-depth of an $n$-bit data circuit, that is, $P_\psi$, is $O(nq)$. Then, the estimated total number of (logical) qubits is $W = W_{AD} + W_A + W_D + (n + 1)W_{\text{query}} = q + 2^q + (n + 1) + (n + 1)2^{q-1}$, where $W_{AD}$, $W_A$, and $W_D$ denote the number of addresses, ancilla, and data qubits, respectively. Here, $W_{\text{query}}$ is the number of extra qubits in the query. Note that no ancilla qubits are required in the parallelisation of the coupling, whereas $2^{q-1}$ ancilla is required for the query. Accordingly, we confirm that Theorem 1 holds.

S3. TIME-VERSUS-QUBIT TRADE-OFF RELATION IN QUANTUM NBLP ALGORITHM

First, we recall the results of RE 1 and RE 3:

$$
\begin{align*}
W_{P_\psi} &= O(2^{\log n + q}) \quad \text{(number of QRAM logical qubits)}, \\
S &= O \left( 4^{n-q} \epsilon^{-2} |2\eta - t|^{-4} \ln^{-1} \delta \right) \quad \text{(quantum sample complexity)},
\end{align*}
$$

(S34)

where $\epsilon$ and $1 - \delta$ represent the confidence and success probabilities of majority-voting, respectively. Here, $t < \eta$ and $\epsilon < 1 - \frac{P_{\text{P,\text{inst}}}}{P_{\text{P,\text{sup}}}}$. Then, the time-versus-qubit trade-off relation can be specified such that

$$
W_{P_\psi}^2 \times S = O(4^{n+\log n}),
$$

(S35)

where we omit the dependence of $S$ on the other parameters ($\epsilon$, $\eta$, $t$, and $\delta$) for readability. This trade-off may limit the applicability of the algorithm. For example, if $q = n$, the polynomial quantum sample complexity can be obtained (as argued in Refs. [8, 9]) at the expense of $O(2^{n \log n})$ scaling of the logical qubits. By contrast, if we attempt to reduce $W_{P_\psi}$ to a polynomial in $n$, for example, by letting $q = O(\log n)$, we can achieve $S = O(4^{n-\log n})$. However, the exponential reduction in quantum-sample complexity cannot be achieved.

The size of the logical qubits can be a bit more optimised by letting $q = \lceil n - \alpha \log n \rceil$, where $\alpha$ is a constant number exceeding 1. We then obtain $W_{P_\psi} = O(2^{n-(\alpha-1) \log n})$; that is, the required number of logical qubits can be reduced by a factor of $\frac{1}{n^{\alpha-1}}$. The polynomial quantum-sample complexity is expressed as follows:

$$
S = O \left( n^{2\alpha} \epsilon^{-2} |2\eta - t|^{-4} \ln^{-1} \delta \right).
$$

(S36)

* The first three authors contributed equally to this work
† Electronic address: m.kim@imperial.ac.uk
‡ Electronic address: jbang@etri.re.kr

[1] V. Giovannetti, S. Lloyd, and L. Maccone, Physical Review Letters 100, 160501 (2008).
[2] R. C. Jaeger and T. N. Blalock, Microelectronic circuit design (McGraw-Hill New York, 1997).
[3] L. K. Grover, Physical Review Letters 79, 325 (1997).
[4] O. Di Matteo, V. Gheorghiu, and M. Mosca, IEEE Transactions on Quantum Engineering 1, 1 (2020).
[5] P. Niemann, A. Gupta, and R. Drechsler, in 2019 IEEE 49th International Symposium on Multiple-Valued Logic (ISMVL) (IEEE, 2019) pp. 108–113.
[6] M. Amy, D. Maslov, and M. Mosca, IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems 33, 1476 (2014).
[7] P. Selinger, Physical Review A 87, 042302 (2013).
[8] A. W. Cross, G. Smith, and J. A. Smolin, Physical Review A 92, 012327 (2015).
[9] A. B. Grilo, I. Kerenidis, and T. Zijlstra, Physical Review A 99, 032314 (2019).