Resource quantification for programming low-depth quantum circuits

Entong He and Yuxiang Yang

QICI Quantum Information and Computation Initiative, School of Computing and Data Science, The University of Hong Kong, Pokfulam Road, Hong Kong SAR, China

Noisy intermediate-scale quantum (NISQ) devices pave the way for implementing quantum algorithms that offer quantum advantages over their classical counterparts. Due to the intrinsic noise and decoherence in the physical system, NISQ machines are naturally modeled as large-scale, low-depth quantum circuits. In practice, executing such circuits requires sending program states that encode the relevant instructions to a programmable quantum computer, typically through a cloud service. Existing programming approaches designed for generic unitary transformations are computationally inefficient in the low-depth setting, and therefore remain unsatisfactory. As such, to realize NISQ algorithms, it is crucial to find an efficient way to program low-depth circuits as the number of qubits N increases. Here, we investigate the circuit complexity and the size of quantum memory, known as the program cost, required to program low-depth brickwork circuits. We establish a tight worst-case program cost of $\Theta(N\mathrm{polylog}N)$ for universally programming low-depth brickwork circuits in the large-N regime. Moreover, we analyze the trade-off between the cost of describing the layout of local gates and the cost of programming them to implement the target unitaries via the light-cone argument. Our findings suggest that faithful gate-wise programming is essentially optimal in the low-depth regime.

1 Motivation

The world is awaiting the arrival of the Noisy Intermediate-scale Quantum (NISQ) technology era in the near future [1]. With qubit counts on the order of tens to hundreds, quantum computers can run quantum algorithms that offer advantages in time and quantum-memory complexity over their classical counterparts on the same tasks [2, 3, 4]. Despite this optimistic vision, the computational power of quantum computers is intrinsically limited by noise inherent in gate operations and by the limited coherence times of quantum particles interacting with the external environment [5]. To suppress their interference with the computational results, an intuitive approach is to compress the depth of the quantum circuits [6]. These low-depth designs limit the number of sequential operations performed during a computation but can still be provably more powerful than their classical counterparts [7, 8].

Entong He: ethe@cs.hku.hk

Yuxiang Yang: yuxiang@cs.hku.hk

Given the prohibitive resource requirements and costs of local quantum hardware in the NISQ era, clients may delegate computations [9] to cloud servers by transmitting program states that encode target unitary operations. Rather than designing a dedicated circuit for each computational task, the cloud computer uses a fixed circuit architecture capable of processing *any* program state; we refer to this architecture as its *processor*. The processor's ability to implement any target unitary from its corresponding program state is referred to as *universal programmability*. Programming unitaries differs from synthesizing unitaries [10] in that all information about the target unitary is supplied in quantum rather than classical form, as a quantum state that we conventionally refer to as its *program state*. An essential component for implementing NISQ algorithms is a quantum processor capable of universally programming low-depth quantum operations. Although the *No-Programming Theorem* has ruled out the possibility of exact universal programming, as an infinite-size quantum memory and quantum circuit are necessary to store and retrieve the program, its approximate analogue is available [11]. Numerous efforts have been made [12, 13, 14, 15, 16, 17, 18, 19] to circumvent this no-go theorem by allowing nonzero programming error in exchange for reduced quantum-memory complexity; we refer to this complexity as the *program cost*. Optimal trade-offs have been established for unitaries [16], channels [17], and isometries [19], primarily within two frameworks: measure-and-operate (MO) [20, 16, 17], and port-based teleportation (PbT) [14, 18, 19]. Nevertheless, all these works, including [16], assume constant system dimensions and derive the bounds for program cost in the arbitrarily small error regime. Meanwhile, NISQ algorithms might run on many qubits to maintain their potential quantum advantage [7], thereby creating a "wide" circuit architecture. Rather than taking the programming error to be infinitesimal, we allow the programming error to vanish as the system dimension grows. Under these assumptions, the number of qubits controls both the asymptotic error and the circuit architecture, making it natural to ask whether finer-grained lower and upper bounds can be obtained for programming low-depth quantum circuits. In addition, these works focus on an information-theoretic perspective and do not consider the efficiency of implementing the protocols on practical quantum devices, in terms of both gate and program cost.

This article aims to identify the resource requirements for programming QNC circuits, i.e., quantum circuits with polylogarithmic depth [21] and bounded-fan-in gates. This gives rise to a fundamental yet representative quantum circuit architecture – the brickwork circuit [22, 23, 24, 25, 26], which consists of local unitary gates with restricted connectivity. This architecture closely reflects the physical constraints of NISQ devices. Many near-term quantum algorithms are developed with QNC brickwork circuits. Representative examples include the Variational Quantum Algorithms [27], random unitaries [28, 29], and quantum algorithms that efficiently solve the 2D Hidden Linear Function problem [7]. We quantify the overall circuit complexity of preparing and executing the program using the best-known MO universal unitary programming scheme [16]. Building on the constructive techniques employed in [16, 17], we provide a tight lower bound for the worst-case program cost requirement, as a refinement of the previous results [15, 16, 17] in the low-depth regime. A straightforward counting argument of the covering net for all QNC brickwork circuits asymptotically matches this lower bound. Simple as it is, this argument provides a complete characterization of the quantum-storage complexity of programming low-depth brickwork circuits.

Our main results are summarized as follows:

Theorem 1 (Circuit complexity of optimal universal programming, informal)**.** *The optimal universal programming of* $\mathsf{U}(d)$ *where* $d = 2^N$ *with diamond norm error* ϵ *can be implemented with* $\widetilde{\mathcal{O}}\left(\mathrm{poly}(d, 1/\sqrt{\epsilon})\right)$ *quantum gates and* $\mathcal{O}\left(d^2 \log(d^2/\epsilon)\right)$ *ancillae. Furthermore,*

programming low-depth brickwork circuits by programming each local gate optimally can be implemented with $\widetilde{\mathcal{O}}\left(\mathrm{poly}(N, 1/\sqrt{\epsilon})\right)$ *quantum gates and* $\widetilde{\mathcal{O}}\left(N\log(1/\epsilon)\right)$ *ancillae.*

We refer to Section 3.2 for further context. This argument says that, in terms of *universal* programming of unitary operations, the circuit complexity required to optimally synthesize a generic unitary from its corresponding program state is expected to be exponentially high, even approximately. The hardness originates from the fact that a general unitary matrix is characterized by up to $\Theta(d^2)$ free parameters, so is its program state. Notably, the same scaling behavior appears in the context of quantum gate learning and metrology. A related result is stated in [30, Section 4.5.4]; however, the programming setting, however, the execution circuit is oblivious to the target unitary rather than tailored to it.

Theorem 2 (Program cost bounds for programming low-depth circuits, informal version of Theorem 19 and Theorem 25)**.** *Programming a low-depth quantum circuit on* N *qubits to diamond-norm error* $\epsilon \sim 1/\mathrm{polylog}N < \frac{1}{32}$ *requires a quantum processor with program cost* $c_P = \Omega\left(N\mathrm{polylog}N\right)$ *in the worst case. Moreover,* $c_P = \mathcal{O}\left(N\mathrm{polylog}N\right)$ *if we restrict to brickwork circuits.*

The arguments presented in Theorem 1 and Theorem 2 show that the construction in [16] is far from optimal in terms of both circuit complexity and the cost-error trade-off for programming low-depth circuits.

Notably, the scaling of the cost stated in Theorem 2 is tight. Although the error is bounded in our setting, plainly substituting it into the lower and upper bounds for programming general unitaries [12, 13, 14, 15, 16] does not give us the desired scaling. Our result is obtained via an information-theoretic approach and the counting argument based on the structure of brickwork circuits. A compatible scaling for circuit complexity[1] of N-qubit unitaries is reported in [22, Theorem 1]. Compared with the cardinality of the ϵ-net of the topological group $\mathsf{U}(d)$, one can conclude that within $\mathsf{U}(d)$, the low-depth brickwork circuit unitaries are sparsely distributed.

Besides the complexity of individual local unitaries, the program cost is also related to the quantity of unitary gates and the circuit architecture [22, Figure 2]. These factors are inversely correlated: For a circuit with fixed geometry, larger and more complicated local gates often result in a simpler layout involving fewer gates. When the local gates have fixed dimensions, we can still transit to the larger unitary case using the widely adopted light-cone argument [22, 25, 31, 26]. However, when we take into account the overall program cost, in major cases, the optimal programming scenario is to directly program the primitive small unitaries faithfully.

Fact 3 (Summary of Section 4.3, informal)**.** *There exists an approach to trade the cost of programming local unitary gates in a brickwork quantum circuit off for lower circuit architecture complexity. In terms of overall program cost, however, it provides no reduction in major cases.*

2 Preliminaries

We use the following notion in the upcoming context: Let $d = 2^N$ where $N \in \mathbb{N}$, $\mathcal{H} \cong \mathbb{C}^d$ stands for a d-dimensional Hilbert space, $\mathcal{B}(\mathcal{H})$ stands for the set of bounded linear operators on $\mathcal{H}$, and $\mathsf{U}(d)$ stands for the set of $d \times d$ unitary matrices. The set of density

[1] In their definition, the circuit complexity provides a lower bound for the number of small local gates required to synthesize a global unitary.

matrices on the space $\mathcal{H}$ is denoted as $\mathfrak{D}(\mathcal{H})$. A quantum operation connecting two Hilbert spaces $\mathcal{H}_A$ and $\mathcal{H}_B$ can be formalized as a quantum channel, which is a completely positive and trace-preserving (CPTP) map. We denote it by $\mathsf{CPTP}(\mathcal{H}_A, \mathcal{H}_B)$, or $\mathsf{CPTP}(\mathcal{H})$ for short if $\mathcal{H}_A = \mathcal{H}_B = \mathcal{H}$. For a pure state $|\psi\rangle$, we use ψ as a shorthand for its density matrix $|\psi\rangle\langle\psi|$. For any matrix A, we use the double-ket notation $|A\rangle\!\rangle = \sum_{m,n} \langle n|A|m\rangle |n\rangle|m\rangle$, and use $\|A\|$ to indicate its operator norm. For any matrix A, B, $A \preceq B$ indicates $B - A$ is positive-semidefinite. For any space $\mathcal{H}$, we use $|\Phi^+_{\mathcal{H}}\rangle = \frac{1}{\sqrt{\dim \mathcal{H}}}|I_{\mathcal{H}}\rangle\!\rangle$ to denote the maximally entangled state on $\mathcal{H}^{\otimes 2}$. For a unitary operator U, we use the calligraphic font $\mathcal{U}(\cdot)$ to represent the channel $U(\cdot)U^\dagger$.

We will use the conventional notations (big-O, big-Ω and big-Θ) for the asymptotic behavior of functions [21]. Besides, we will write $f = o(g)$ and $f = \omega(g)$ if $f(x)/g(x)$ and $g(x)/f(x)$ is vanishing with large x. For conciseness, we write $f \sim g$ for $f = \Theta(g)$ and $f \lesssim g$ for $f = \mathcal{O}(g)$. The notation $\widetilde{\mathcal{O}}$, as a variant of the big-O, ignores the logarithmic factors. We say a general quantum operation is implemented ϵ-approximately if the circuit produced is ϵ-close to it in diamond norm (see Definition 5).

Definition 4. For any linear operator $E \in \mathcal{B}(\mathcal{H})$, its **Schatten p-norm** is defined by $\|E\|_p = \left(\mathrm{Tr}[(E^\dagger E)^{p/2}]\right)^{1/p}$. For states $\rho, \sigma \in \mathfrak{D}(\mathcal{H})$, their trace distance is defined as

$$d_{\mathrm{Tr}}(\rho, \sigma) = \frac{1}{2}\|\rho - \sigma\|_1.$$

Definition 5. For a quantum channel $\mathcal{E} \in \mathsf{CPTP}(\mathcal{H}_A, \mathcal{H}_B)$, its diamond norm is given by

$$\|\mathcal{E}\|_\diamond = \sup_{\|M\|_1 \le 1} \|(\mathcal{E} \otimes \mathcal{I}_{\mathcal{H}_R})M\|_1 .$$

The diamond norm distance between two quantum channels $\mathcal{E}, \mathcal{F} \in \mathsf{CPTP}(\mathcal{H}_A, \mathcal{H}_B)$ is given by $\|\mathcal{E} - \mathcal{F}\|_\diamond$, where the supremum is attainable with a reference space $\mathcal{H}_R \cong \mathcal{H}_A$. The diamond norm is sub-multiplicative, i.e., $\|\mathcal{E} \circ \mathcal{F}\|_\diamond \le \|\mathcal{E}\|_\diamond \|\mathcal{F}\|_\diamond$; it is non-increasing under quantum operations, i.e., for any channel $\mathcal{E}, \mathcal{F} \in \mathsf{CPTP}(\mathcal{H}_A, \mathcal{H}_B)$, $\mathcal{T} \in \mathsf{CPTP}(\mathcal{H}_B, \mathcal{H}_C)$, $\|\mathcal{T} \circ \mathcal{E} - \mathcal{T} \circ \mathcal{F}\|_\diamond \le \|\mathcal{E} - \mathcal{F}\|_\diamond$.

Lemma 6 (Schur-Weyl duality [32, 33])**.** *Consider the n-tensor replication of a Hilbert space $\mathcal{H}$ with $\dim \mathcal{H} = d$ and $n \in \mathbb{N}$. A partition $\lambda \vdash n$ of any integer $n \ge 0$ is a tuple $\lambda = (\lambda_1, \ldots, \lambda_d)$ such that $\lambda_1 \ge \lambda_2 \ge \cdots \lambda_d \ge 0$ and $\sum_{j=1}^d \lambda_j = n$. We use the notion $\lambda \vdash_d n$ to indicate $\lambda \vdash n$ into at most d rows. Each λ characterizes the shape of a Young diagram. Denote the modules W^d_λ and V_λ as the irreducible subspaces of $\mathsf{U}(d)$ and the permutation group $\mathfrak{S}_n$, respectively. The Schur-Weyl duality states that*

$$\mathcal{H}^{\otimes n} \cong \bigoplus_{\lambda \vdash_d n} W^d_\lambda \otimes V_\lambda.$$

Moreover, the n-wise tensor product of unitary operator $U \in \mathsf{U}(d)$ admits decomposition

$$U^{\otimes n} \cong \bigoplus_{\lambda \vdash_d n} U_\lambda \otimes I_{V_\lambda},$$

where (U_λ, W^d_λ) are irreducible representations of the group $\mathsf{U}(d)$.

Definition 7 (Low-depth brickwork circuit [24, 26])**.** A brickwork quantum circuit on N qubits consists of sequential operations with ℓ unitary gates $\mathscr{G} = \{G_1, \ldots, G_\ell\}$ arranged across at most D layers. The connectivity among qubits, or the geometry, is given by a

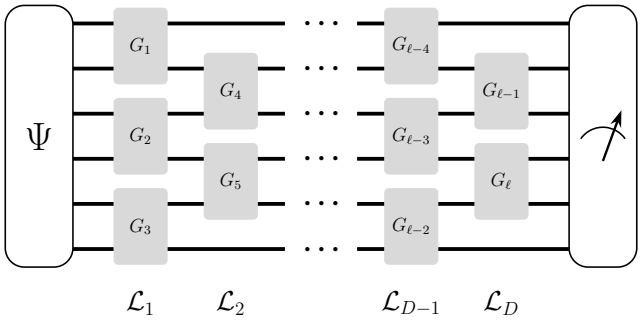


Figure 1: Illustration of a 1D 2-local brickwork quantum circuit consisting of ℓ gates arranged across D layers.

connectivity graph $C = ([N], E)$, where $(j, j') \in E$ if local operations are allowed between qubits j and j'. Each gate is assumed to be k-local with $k = \mathcal{O}(1)$ (bounded constant fan-in), and we denote the set of qubits that G_j acts on as $q_j \subseteq [N]$ subject to $|q_j| = k$ and the subgraph of C induced by qubits in q_j, denoted by $C[q_j]$, is connected. Moreover, we denote $\mathcal{L}_r \subseteq [\ell]$ as the indices of gates applied in the r-th operation. An illustrative example is given in Figure 1.

Following the notion of QNC circuits [21], we say a brickwork circuit has low depth if $D = \mathcal{O}(\text{polylog}N)$.

Definition 8 (Unitary design [34])**.** Let ν be any ensemble of unitary operators on $\mathsf{U}(d)$. Its t-moment operator is defined as

$$\mathcal{M}_\nu^{(t)}(\rho) := \mathop{\mathbb{E}}_{U\sim\nu}\left[U^{\otimes t}\rho U^{\dagger\otimes t}\right].$$

Setting $\nu = \mu$, the Haar measure over $\mathsf{U}(d)$, this is the Haar random t-moment operator: $\mathcal{M}_{\text{Haar}}^{(t)}(\rho) := \int_{\mathsf{U}(d)} U^{\otimes t}\rho U^{\dagger\otimes t}\,\mathrm{d}\mu(U)$. For any $\delta \in (0, 1]$, ν is a diamond norm δ-approximate unitary t-design if $\|\mathcal{M}_\nu^{(t)} - \mathcal{M}_{\text{Haar}}^{(t)}\|_\diamond \leq \delta$, and is a relative δ-approximate unitary t-design if $(1-\delta)\mathcal{M}_{\text{Haar}}^{(t)} \preceq \mathcal{M}_\nu^{(t)} \preceq (1+\delta)\mathcal{M}_{\text{Haar}}^{(t)}$. A t-design is also an s-design for any $s < t$.

Remark 9. On a quantum device, generating Haar random unitaries is inefficient, where the number of gates grows exponentially with the number of qubits [35]. Therefore, in practice, we often employ the aforementioned unitary t-designs to match the first t moments of Haar measures for special computational tasks.

3 Approximate programmability and circuit complexity of MO universal unitary programming

In this section, we comment on the circuit efficiency of the state-of-the-art universal programming scheme, i.e., the learning-based measure-and-operate (MO) programming scheme [20, 16], in terms of its circuit complexity.

3.1 Basic setup of universal programming, and the learning-based MO scheme

Definition 10 (ϵ-universal quantum processor [16, 17])**.** A quantum processor for unitaries in $\mathsf{U}(d)$ is a tuple $(\mathcal{C}, \{\psi_U\}_{U\in\mathsf{U}(d)})$ where $\mathcal{C} \in \mathsf{CPTP}(\mathcal{H}\otimes\mathcal{H}_P)$ and $\psi_U \in \mathfrak{D}(\mathcal{H}_P)$ (the

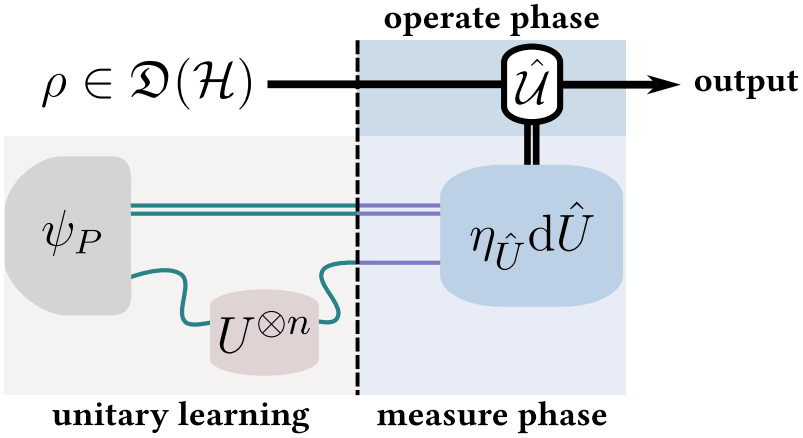


Figure 2: The learning-based MO universal programming scheme.

'quantum chip'). The space $\mathcal{H}_P$ supports the programmability of the processor, and the processor channel $\mathcal{C}$ is independent of the choice of U. The output of the processor when programming $\mathcal{U}$ is given by $\mathcal{E}_U(\cdot) := \mathrm{Tr}_{\mathcal{H}_P}[\mathcal{C}(\cdot \otimes \psi_U)]$. For any $\epsilon \in (0, 1]$, a processor $\mathcal{E}_U$ is ϵ-universal if

$$\forall U \in \mathsf{U}(d), \quad \frac{1}{2} \left\| \mathcal{U} - \mathcal{E}_U \right\|_\diamond \leq \epsilon.$$

The learning-based MO scheme provides a unified framework for programming arbitrary unitary gates in a semi-classical fashion, consisting of **(1)** coherently learning parallel copies of the unitary gate U with a properly entangled probe state $|\psi_P\rangle$ and storing the information in the quantum register in the form of a program state $|\psi_{P,U}\rangle$; and **(2)** retrieving the channel from the register via post-selection with a continuous measurement on $|\psi_{P,U}\rangle$, namely the measure-and-operate operation. A diagrammatic illustration is shown in Figure 2. The probe state and the POVM utilize a representation-theoretic design to achieve sample-optimal ϵ-approximate programming at a Heisenberg limit [16].

It is proved in [20, Lemma 2] that due to symmetry, the optimal probe state for learning n parallel copies of an agnostic unitary is of the form $|\psi_P\rangle = \bigoplus_{\lambda \in \mathsf{S}} \sqrt{q_\lambda} |\Phi^+_{W^d_\lambda}\rangle \otimes |\eta_\lambda\rangle$, where $\mathsf{S} \subseteq \mathsf{S}^d_n := \{\lambda \mid \lambda \vdash_d n\}$ is a subset of all legitimate partitions, $\{q_\lambda\}_{\lambda \in \mathsf{S}}$ is a probability distribution, and $|\eta_\lambda\rangle$ is an arbitrary bipartite state on $V_\lambda^{\otimes 2}$. The learned program state can be expressed by $|\psi_{P,U}\rangle = (U \otimes I)^{\otimes n} |\psi_P\rangle$. To retrieve the unitary, the optimal measurement [20, Theorem 1] is given by $\{|\eta_{\hat{U}}\rangle \langle \eta_{\hat{U}}| \, \mathrm{d}\mu(\hat{U})\}$, where $|\eta_{\hat{U}}\rangle = (\hat{U} \otimes I)^{\otimes n} |\psi_0\rangle$ and the frame vector $|\psi_0\rangle = \bigoplus_{\lambda \in \mathsf{S}} \dim W^d_\lambda |\Phi^+_{W^d_\lambda}\rangle \otimes |\eta_\lambda\rangle$. If we use the notation $\chi_U := \mathrm{Tr}[U]$ to denote the character of $\mathsf{U}(d)$, the MO processor can be expressed by the following quantum channel:

$$\begin{aligned}
\mathcal{E}_{\mathsf{MO},U}(\rho) &= \int_{\mathsf{U}(d)} \mathrm{Tr}\left(\eta_{\hat{U}} \psi_{P,U}\right) \cdot \hat{\mathcal{U}}(\rho) \, \mathrm{d}\mu(\hat{U}) \\
&= \int_{\mathsf{U}(d)} \left| \sum_{\lambda \in \mathsf{S}} \sqrt{q_\lambda} \chi_{\hat{U}_\lambda U_\lambda^{-1}} \right|^2 \cdot \hat{\mathcal{U}}(\rho) \, \mathrm{d}\mu(\hat{U}) \\
&= \mathfrak{p} \cdot \mathcal{U}(\rho) + (1 - \mathfrak{p}) \cdot \frac{I_{\mathcal{H}}}{d}, \\
\mathfrak{p} &= \frac{1}{d^2 - 1} \left(\int_{\mathsf{U}(d)} \left| \sum_{\lambda \in \mathsf{S}} \sqrt{q_\lambda} \sum_{\gamma \in \mathsf{O}_1(\lambda)} \chi_{\hat{U}_\gamma} \right|^2 \mathrm{d}\mu(\hat{U}) - 1 \right),
\end{aligned} \tag{1}$$

where the set of partitions $\mathsf{O}_1(\lambda) \subseteq \mathsf{S}^d_{n+1}$ is induced by the tensor product of shapes $\lambda \otimes (\square)$,

and ($\square$) corresponds to the trivial representation $\hat{U}_{\square} = \hat{U}$. Note that the last equality follows from Schur's lemma [36], since the composite channel $\mathcal{E}_{\mathsf{MO},U} \circ \mathcal{U}^{\dagger}$ is covariant. The expression of the "depolarizing" coefficient $\mathfrak{p}$ is equivalent to the entanglement fidelity [16, Equation B13] up to a scalar, which the authors use as an intermediate identity to derive the diamond norm distance between $\mathcal{E}_{\mathsf{MO},U}$ and $\mathcal{U}$. Alternatively, our formulation in Equation 1 allows direct evaluation.

3.2 Circuit complexity of unitary programming with the optimal MO scheme

Although [20, 16] have demonstrated the information-theoretic optimality of the learning-based MO programming scheme, its operational feasibility, i.e., circuit complexity, remains unexplored. In the following context, we discuss the circuit complexity of implementing the measure-and-operate phase upon obtaining the program state in the optimal MO programming scheme. Firstly, preparing the observable ψ_0 requires applying the Schur transformation to the computational basis of $\mathcal{H}^{\otimes n}$, generating entanglement in the Schur-Weyl basis, and adjusting the amplitudes by a Grover-type algorithm [37]. The Schur transformation can be implemented with $\mathcal{O}\left(\mathrm{poly}(n, \log d, \log(1/\zeta))\right)$ elementary operations with error ζ in operator norm [38, 39], while the subsequent entangling and amplitude-tuning operations require $\Theta(n \log d)$ gates [40]. Suppose that the Schur transformation is ζ-approximate and produces state $|\widetilde{\psi}_0\rangle$. Then $\|\widetilde{\psi}_0 - \psi_0\|_1 \leq \zeta$, and $\left\|\widetilde{\eta}_{\hat{U}} - \eta_{\hat{U}}\right\|_1 \leq \zeta$. The corresponding channel $\widetilde{\mathcal{E}}_{\mathsf{MO},U}$ satisfies

$$\begin{aligned}
\frac{1}{2}\left\|\widetilde{\mathcal{E}}_{\mathsf{MO},U} - \mathcal{E}_{\mathsf{MO},U}\right\|_{\diamond} &\leq \frac{1}{2}\int_{\mathsf{U}(d)} \left|\mathrm{Tr}((\widetilde{\eta}_{\hat{U}} - \eta_{\hat{U}})\psi_{P,U})\right| \,\mathrm{d}\mu(\hat{U}) \\
&\leq \int_{\mathsf{U}(d)} d_{\mathrm{Tr}}\left(\widetilde{\eta}_{\hat{U}}, \eta_{\hat{U}}\right) \,\mathrm{d}\mu(\hat{U}) \\
&\leq \frac{1}{2}\zeta.
\end{aligned}$$

One can readily verify that this inequality holds for any ensemble other than the Haar measure. For simplicity, we restrict $\zeta = \mathcal{O}(\epsilon)$ and ignore this error term in the latter context.

A major computational burden lies in applying the optimal POVM to the program state. Mathematically, this is equivalent to first using Haar random resources to generate the matrix $\hat{U}^{\otimes n}$, use it to synthesize the measurement operator $\eta_{\hat{U}}$, and perform the two-outcome PVM $\{\eta_{\hat{U}}, I - \eta_{\hat{U}}\}$. The quantum processor applies $\hat{\mathcal{U}}$ to the input state if the outcome "$\eta_{\hat{U}}$" occurs, and acts trivially (aborts) otherwise. Therefore, the circuit complexity of implementing the POVM is closely tied to the generation[2] of $\hat{U}$. In practice, we might use unitary designs to mitigate the circuit depth requirement. We evaluate the robustness of the MO scheme's performance in terms of unitary design accuracy in the following context, before which we show a matrix inequality lemma.

Lemma 11. *For two CPTP maps $\mathcal{A}, \mathcal{B} \in \mathsf{CPTP}(\mathcal{H})$ that satisfy $(1-\upsilon)\mathcal{A} \preceq \mathcal{B} \preceq (1+\upsilon)\mathcal{A}$ for some $\upsilon \in [0, 1)$, the tensor product Hilbert space $\mathcal{H} = \mathcal{H}_S \otimes \mathcal{H}_0$, any Hermitian and positive semidefinite operators X, $Y = Y_S \otimes I_0 \in \mathcal{B}(\mathcal{H})$, it holds that*

$$\left\|\mathrm{Tr}_{\mathcal{H}_S}\left[(\mathcal{A} - \mathcal{B})(X)Y\right]\right\|_1 \leq \upsilon \left\|\mathrm{Tr}_{\mathcal{H}_S}\left[\mathcal{A}(X)Y\right]\right\|_1 .$$

[2]For the generation of the n-tensor $\hat{U}^{\otimes n}$, the circuit depth is unaffected, as we can generate identical gates with a fixed configuration on n sites in parallel.

Proof. Note that $(1-\upsilon)\mathcal{A} \preceq \mathcal{B} \preceq (1+\upsilon)\mathcal{A}$ implies

$$-\upsilon\mathcal{A}(X) \preceq (\mathcal{A}-\mathcal{B})(X) \preceq \upsilon\mathcal{A}(X). \tag{2}$$

Denote $Z = (\mathcal{A}-\mathcal{B})(X)$, $W = \mathcal{A}(X)$. Using [30, Equation 9.22] that $\|F\|_1 = \sup_{Q:\|Q\|\leq 1} \mathrm{Tr}\,[FQ]$, we can rewrite the expression of the trace norm on both sides:

$$\begin{aligned}\left\|\mathrm{Tr}_{\mathcal{H}_S}[ZY]\right\|_1 &= \sup_{Q:\|Q\|\leq 1} \mathrm{Tr}\left[\mathrm{Tr}_{\mathcal{H}_S}[ZY]\,Q\right] \\ &= \sup_{Q:\|Q\|\leq 1} \mathrm{Tr}\left[(ZY)(I_S\otimes Q)\right] \\ &= \sup_{Q:\|Q\|\leq 1} \mathrm{Tr}\left[Z\,(Y_S\otimes Q)\right].\end{aligned}$$

Since $Y_S \succeq 0$, we take its square root $\sqrt{Y_S}$. Denote $\widetilde{Z} = \left(\sqrt{Y_S}\otimes I_0\right) Z \left(\sqrt{Y_S}\otimes I_0\right)$, it holds that $\mathrm{Tr}\left[Z\,(Y_S\otimes Q)\right] = \mathrm{Tr}\left[\widetilde{Z}(I_S\otimes Q)\right]$. Analogously, we have $\mathrm{Tr}\left[W(Y_S\otimes Q)\right] = \mathrm{Tr}\left[\widetilde{W}(I_S\otimes Q)\right]$ if we denote $\widetilde{W} = \left(\sqrt{Y_S}\otimes I_0\right) W \left(\sqrt{Y_S}\otimes I_0\right)$. The condition stated in Equation 2 implies $-\upsilon\widetilde{W} \preceq \widetilde{Z} \preceq \upsilon\widetilde{W}$, and therefore $\left|\mathrm{Tr}_{\mathcal{H}_S}(\widetilde{Z})\right| \preceq \mathrm{Tr}_{\mathcal{H}_S}(|\widetilde{Z}|) \preceq \upsilon\mathrm{Tr}_{\mathcal{H}_S}(\widetilde{W})$[3]. If we substitute in $\widetilde{Z}$ in the supremum, we have

$$\sup_{Q:\|Q\|\leq 1} \mathrm{Tr}\left[\widetilde{Z}(I_S\otimes Q)\right] = \sup_{Q:\|Q\|\leq 1} \mathrm{Tr}\left[\mathrm{Tr}_{\mathcal{H}_S}(\widetilde{Z})Q\right] = \left\|\mathrm{Tr}_{\mathcal{H}_S}(\widetilde{Z})\right\|_1,$$

while the same holds for $\widetilde{W}$. Combining the previous statements and that $\|M\|_1 = \||M|\|_1$ for any operator M, we have

$$\begin{aligned}\left\|\mathrm{Tr}_{\mathcal{H}_S}\left[(\mathcal{A}-\mathcal{B})(X)Y\right]\right\|_1 &= \sup_{Q:\|Q\|\leq 1} \mathrm{Tr}\left[\widetilde{Z}(I_S\otimes Q)\right] = \left\|\mathrm{Tr}_{\mathcal{H}_S}(\widetilde{Z})\right\|_1 = \left\|\left|\mathrm{Tr}_{\mathcal{H}_S}(\widetilde{Z})\right|\right\|_1 \\ &\leq \upsilon\left\|\mathrm{Tr}_{\mathcal{H}_S}(\widetilde{W})\right\|_1 = \upsilon \sup_{Q:\|Q\|\leq 1} \mathrm{Tr}\left[\widetilde{W}(I_S\otimes Q)\right] \\ &= \upsilon\left\|\mathrm{Tr}_{\mathcal{H}_S}\left[\mathcal{A}(X)Y\right]\right\|_1.\end{aligned}$$

This completes the proof. □

Theorem 12. *A relative δ-approximate unitary $(n+1)$-design implements the measure phase δ-approximately using the MO scheme that learns n copies of U, assuming the observable ψ_0 is prepared perfectly.*

Proof. Note that the channel $\mathcal{E}_{\mathsf{MO},U}$ can be rewritten as

$$\begin{aligned}\mathcal{E}_{\mathsf{MO},U} &= \int_{\mathsf{U}(d)} \mathrm{Tr}\left(\eta_{\hat{U}}\psi_{P,U}\right)\cdot\hat{\mathcal{U}}\,\mathrm{d}\mu(\hat{U}) \\ &= \int_{\mathsf{U}(d)} \mathrm{Tr}\left((\hat{U}\otimes I)^{\otimes n}\psi_0(\hat{U}\otimes I)^{\otimes n\dagger}\psi_{P,U}\right)\cdot\hat{\mathcal{U}}\,\mathrm{d}\mu(\hat{U}).\end{aligned}$$

When replacing the Haar random ensemble with a δ-approximate ensemble, we first sample $S\sim\nu_\delta$ and apply the n-wise tensor to obtain the POVM operator. Denote the n ancilla registers into which the program state is inserted by $\mathsf{A}_1,\ldots,\mathsf{A}_n$, for any state $\rho\in\mathfrak{D}(\mathcal{H})$,

$$\begin{aligned}&\mathrm{Tr}\left((\hat{U}^{\otimes n}\otimes I)\psi_0(\hat{U}^{\otimes n}\otimes I)^\dagger\psi_{P,U}\right)\cdot\hat{\mathcal{U}}(\rho) \\ =&\,\mathrm{Tr}_{\mathsf{A}_1,\ldots,\mathsf{A}_n}\left[(\hat{U}\otimes I)^{\otimes n}\psi_0(\hat{U}\otimes I)^{\otimes n\dagger}\psi_{P,U}\otimes\hat{\mathcal{U}}(\rho)\right] \\ =&\,\mathrm{Tr}_{\mathsf{A}_1,\ldots,\mathsf{A}_n}\left[\left(\left(\left(\hat{\mathcal{U}}\otimes\mathcal{I}\right)^{\otimes n}\otimes\hat{\mathcal{U}}\right)(\psi_0\otimes\rho)\right)(\psi_{P,U}\otimes I)\right].\end{aligned} \tag{3}$$

[3] By the fact that the partial trace operation is a completely positive map [41].

Ref.	Depth	Condition	Type
[42]	$\mathcal{O}\left(\mathrm{poly}\left(t, \log(1/\varrho)\right) \cdot N^{1/\mathfrak{K}}\right)$	C is a $\mathfrak{K}$-lattice	diamond
[43]	$\mathcal{O}\left(\left(Nt^2 + t\log(1/\varrho)\right)\log N\right)$	–	diamond & relative
[34]	$\mathcal{O}\left(t\,\mathrm{poly}N + t\log(1/\varrho)\right)$	$t \leq 2^{N/4}$	diamond
[34]	$\mathcal{O}\left(t^2\,\mathrm{poly}N + t^2\log(1/\varrho)\right)$	$t \leq 2^{N/4}$	relative
[44]	$\mathcal{O}\left(\left(Nt + \log(1/\varrho)\right)\log^7 t\right)$	$t = \mathcal{O}\left(2^{2N/5}\right)$	diamond & relative
[29]	$\mathcal{O}\left(\left(\xi t + \log\left(N/\varrho\right)\right)\log^7 t\right)$	$t = \mathcal{O}\left(2^{2\xi/5}\right)$, $\exists\,\xi \geq 1$	diamond & relative

Table 1: Circuit depth upper bound for ϱ-approximate unitary t-designs on N qubits.

For clarity, we append the channel with a superscript to indicate the underlying unitary ensemble, i.e., $\mathcal{E}^{\nu}_{\mathsf{MO},U}$ when ν is utilized. It will be convenient to define the channel

$$\mathcal{Q}^{(t)}_{\nu}(\rho) = \mathop{\mathbb{E}}_{U\sim\nu}\left[(\mathcal{U}\otimes\mathcal{I})^{\otimes t}(\rho)\right]$$

for any unitary ensemble ν and $t \in \mathbb{N}$. Observe that $\mathcal{Q}^{(t)}_{\nu} \cong \mathcal{M}^{(t)}_{\nu}\otimes\mathcal{I}^{\otimes t}$, it follows that $(1-\delta)\mathcal{Q}^{(t)}_{\mathrm{Haar}} \preceq \mathcal{Q}^{(t)}_{\nu} \preceq (1+\delta)\mathcal{Q}^{(t)}_{\mathrm{Haar}}$ when ν is a relative δ-approximate unitary t-design. Denote the auxiliary ancillae in the definition of diamond norm by R, and the working register by O, we obtain

$$\left\|\mathcal{E}^{\nu_\delta}_{\mathsf{MO},U} - \mathcal{E}^{\mathrm{Haar}}_{\mathsf{MO},U}\right\|_{\diamond} = \max_{|\rho_{\mathsf{OR}}\rangle}\left\|\left((\mathcal{E}^{\nu_\delta}_{\mathsf{MO},U} - \mathcal{E}^{\mathrm{Haar}}_{\mathsf{MO},U})\otimes\mathcal{I}_{\mathsf{R}}\right)(\rho_{\mathsf{OR}})\right\|_1.$$

Note that with the equality in Equation 3 and that $\mathcal{H}_{\mathsf{R}} \cong \mathcal{H}$,

$$\begin{aligned}
&\left\|\left((\mathcal{E}^{\nu_\delta}_{\mathsf{MO},U} - \mathcal{E}^{\mathrm{Haar}}_{\mathsf{MO},U})\otimes\mathcal{I}_{\mathsf{R}}\right)(\rho_{\mathsf{OR}})\right\|_1 \\
=\ &\left\|\mathop{\mathbb{E}}_{S\sim\nu_\delta}\left[\mathrm{Tr}\left(\eta_S\psi_{P,U}\right)\cdot(\mathcal{S}\otimes\mathcal{I}_{\mathsf{R}})(\rho_{\mathsf{OR}})\right] - \int_{\mathsf{U}(d)}\mathrm{Tr}\left(\eta_{\hat{U}}\psi_{P,U}\right)\cdot(\hat{\mathcal{U}}\otimes\mathcal{I}_{\mathsf{R}})(\rho_{\mathsf{OR}})\,\mathrm{d}\mu(\hat{U})\right\|_1 \\
=\ &\left\|\mathrm{Tr}_{\mathsf{A}_1,\ldots,\mathsf{A}_n}\left[\left(\left(\mathcal{Q}^{(n+1)}_{\nu_\delta} - \mathcal{Q}^{(n+1)}_{\mathrm{Haar}}\right)(\psi_0\otimes\rho_{\mathsf{OR}})\right)(\psi_{P,U}\otimes I_{\mathsf{OR}})\right]\right\|_1 \\
\leq\ &\delta\cdot\left\|\mathrm{Tr}_{\mathsf{A}_1,\ldots,\mathsf{A}_n}\left[\left(\mathcal{Q}^{(n+1)}_{\mathrm{Haar}}(\psi_0\otimes\rho_{\mathsf{OR}})\right)(\psi_{P,U}\otimes I_{\mathsf{OR}})\right]\right\|_1 \\
=\ &\delta\cdot\left\|\left(\mathcal{E}^{\mathrm{Haar}}_{\mathsf{MO},U}\otimes\mathcal{I}_{\mathsf{R}}\right)(\rho_{\mathsf{OR}})\right\|_1 \\
=\ &\delta,
\end{aligned}$$

where we have used Lemma 11 in the last inequality. This completes the proof. □

For an ϵ-universal quantum processor constructed from the MO scheme, by a simple additive argument, $\frac{1}{2}\|\mathcal{E}^{\nu_\delta}_{\mathsf{MO},U} - \mathcal{U}\|_{\diamond} \leq \epsilon + \frac{1}{2}\delta$. To ensure the performance of the retrieval of U from the program state, we require $\delta = \mathcal{O}(\epsilon)$. Note that a sufficiently accurate $(n+1)$-design is necessary, as a large deviation from the optimal covariant POVM [45, 20, 16] affects the density of the measurement outcomes, thus reducing the retrieval precision. Moreover, the query complexity of either learning or programming a d-dimensional unitary scales as $n = \Omega(d^2)$ [16, 46, 24], forcing the design to match the Haar measure to a high order[4]. As presented in Table 1, the circuit depth bound becomes trivial as $t \sim 2^{2N}$. Therefore, the circuit depth requirement for implementing the unitary design in the measure phase is almost

[4]Most constructions of unitary t-designs are gate-efficient only when $t = \mathcal{O}\left(2^{N/2}\right)$.

identical to that of implementing a genuine Haar measure, which requires an elementary gate sequence with depth d^2 [47], resulting in an $\mathcal{O}(nd^2 \log d + \mathrm{poly}(n, \log d, \log(1/\zeta)))$ gate complexity in implementing the Schur-transformed n-wise unitary $\hat{U}^{\otimes n}$. Upon obtaining the measurement outcome, it suffices to construct the unitary transformation $\hat{\mathcal{U}}$ from its classical description. We need extra $\mathcal{O}(d^2 \log^3(d^2/\tau))$ elementary operations to synthesize it up to $\tau = \mathcal{O}(\epsilon)$ in diamond norm error[5] by the Solovay-Kitaev theorem [48]. The POVM using a $(n+1)$-design ensemble $\mathsf{X}_{n+1,d}$ requires $\log_2 |\mathsf{X}_{n+1,d}| = \mathcal{O}(d^2 \log n)$ ancillae [49]. Therefore, the overall gate complexity scales as $\mathrm{poly}(n, d, \log(1/\zeta), \log(1/\tau), \log(1/\delta))$, with $\mathcal{O}(d^2 \log n)$ ancillae.

Finally, physically implementing the MO scheme yields an ϵ_{MO}-universal processor, where ϵ_{MO} is composed of the following terms:

$$\epsilon_{\mathsf{MO}} = \underbrace{\epsilon}_{\text{optimal retrieval error}} + \underbrace{\frac{1}{2}\zeta}_{\text{POVM application error}} + \underbrace{\frac{1}{2}\delta}_{\text{measure phase error}} + \underbrace{\tau}_{\text{operate phase error}},$$

and uses $\widetilde{\mathcal{O}}\left(\mathrm{poly}(d, 1/\sqrt{\epsilon})\right)$ quantum gates and $\mathcal{O}\left(d^2 \log(d^2/\epsilon)\right)$ ancillae beyond the queries to the unknown unitary if we take $\zeta, \tau, \delta \sim \epsilon$ and set $n \sim d^2/\sqrt{\epsilon}$ [16, Theorem 2]. A lower circuit complexity is reported in [46, Theorem 1.1.3]; however, the $\mathrm{poly}(d)$ factor remains unavoidable, which is prohibitive for the purpose of designing an operationally efficient universal programming scheme for $\mathsf{U}(d)$. To the best of our knowledge, no gate-efficient algorithm has been proposed. Although there are unitaries that require deeper circuits to generate [30], recovering a unitary U from the compact program state $\psi_{P,U}$ is generally more challenging than generating it [22].

3.3 Overview of circuit complexity of unitary programming with PbT scheme

To compare the circuit complexity analysis for the MO scheme in Section 3.2, we briefly review the circuit complexity of implementing unitary programming with the PbT scheme. As is remarked in [19], the PbT scheme is a coherent programming strategy, while the MO scheme breaks the coherence via the covariant measurement. Compared to the MO scheme, the PbT scheme requires an even larger amount of multi-copy entanglement between spatial registers [50]. The computational burden of the PbT scheme also stems from measurement: A Pretty-Good Measurement (PGM) is applied according to the average density matrix over all ports. The primitive Ishizaka-Hiroshima protocol [14] naïvely applies the raw PGM and yields an $\mathcal{O}(d^{2n})$ circuit complexity, where n is the number of queries to U. Subsequent works [51, 52, 53] have introduced various methods to reduce circuit complexity by exploiting the symmetry across different ports. The asymptotic scalings of the complexity of these optimized approaches are consistent with those established in Section 3.2, all requiring a gate complexity of $\mathcal{O}(\mathrm{poly}(n, d))$ and an ancillae complexity of $\mathcal{O}(d^2 \log n)$.

3.4 Efficient programming of low-depth brickwork circuits with MO scheme

Although programming an arbitrary unitary by preparing and retrieving a large programming state is resource-intensive, programming a low-depth brickwork circuit can be efficient. Recall from Definition 7 that the circuit can be decomposed into local unitary operations that apply either sequentially or in parallel. If we approximately program each

[5] The diamond norm and the operator norm are equivalent for unitary channels, since $\|U - V\| \leq \|\mathcal{U} - \mathcal{V}\|_\diamond \leq 2\|U - V\|$ [24].

unitary up to a small error, the whole circuit can be programmed approximately with a satisfying performance, given that the number of gates is bounded. The following statement characterizes the error propagation across a brickwork circuit.

Fact 13 ([30])**.** *For a brickwork circuit on ℓ gates, if each gate is programmed up to error ϵ in diamond norm, then the unitary generated by the circuit is programmed up to error $\ell\epsilon$ in diamond norm.*

Since the local gates on the brickwork circuit have a constant dimension $2^k = \mathcal{O}(1)$, using the results from Section 3.2, it requires only $\widetilde{\mathcal{O}}(\mathrm{poly}(1/\sqrt{\epsilon'}))$ quantum gates to program each gate ϵ'-approximately from the collection of their program state by invoking the MO scheme as a subroutine. As per Fact 13, to program a low-depth circuit ϵ-approximate in diamond norm, setting $\epsilon' = \epsilon/\ell$ would suffice. Therefore, programming a brickwork circuit with a fixed architecture ϵ-close in diamond norm requires

$$\ell \cdot \widetilde{\mathcal{O}}\left(\mathrm{poly}(1/\sqrt{\epsilon'})\right) = \widetilde{\mathcal{O}}\left(\ell \cdot \mathrm{poly}(\sqrt{\ell/\epsilon})\right) = \widetilde{\mathcal{O}}\left(\mathrm{poly}\left(N, 1/\sqrt{\epsilon}\right)\right),$$

quantum gates and $\mathcal{O}(\ell \log(1/\sqrt{\epsilon'})) = \widetilde{\mathcal{O}}(N \log(1/\epsilon))$ space, by noting that $\ell \leq \frac{ND}{k} = \mathcal{O}(N\mathrm{polylog}N)$.

Remark 14. Low-depth unitary programming via PbT schemes yields an analogous circuit complexity, as detailed in Section 3.3.

4 Bounds for quantum memory complexity of programming low-depth circuits

Having discussed the hardness of recovering information about general unitaries from the quantum register in the programming scheme, we are also concerned with how large the register should be to achieve approximate programmability. Precisely, we utilize the program states $|\psi_{P,U}\rangle$ [cf. Section 3] to encode the parallel queries to $\mathcal{U}$, henceforth store them in the quantum memory before we ever call the processor to program the unitary. Mathematically, the *program dimension* d_P is defined as the dimension of the subspace where the program states are supported:

$$d_P := \dim\left(\overline{\mathrm{span}}\left\{\psi_{P,U} \mid U \in \mathsf{U}(d)\right\}\right),$$

where $\overline{\mathrm{span}}$ is the closure of the spanned subspace. This identity quantifies the (quantum) memory capacity required to store these quantum states. If we restrict to approximate programmability, there is a finite set of program states, and thus we can remove the closure operation. Equivalently, the base-2 logarithm of the dimension $c_P = \log_2 d_P$, or the *program cost*, quantifies how many qubits are needed in the memory, which is the main figure of merit of [15, 16, 17, 19]. Although prior works have reported either upper and lower bounds for the program cost [12, 13, 14, 15, 16, 17, 18, 19], they are non-trivial only when the quantum circuit resides on a constant number of qubits, i.e., $N = \mathcal{O}(1)$. However, low-depth quantum circuits often extend their registers to preserve their computational power [7]. Therefore, we will instead derive the bounds in the large N regime, while the feasible programming error ϵ yields a lower bound related to the circuit architecture.

4.1 Lower bound for the program cost

We start by presenting several useful lemmas.

Lemma 15 ([16, Appendix A])**.** *For any ϵ-universal processor $(\mathcal{C}, \{\psi_{P,U}\}_{U\in\mathsf{U}(d)})$, where $\mathcal{C}$ is some operations that constructs $\mathcal{U}$ from each $\psi_{P,U}$, for each $U \in \mathsf{U}(d)$, there exists a channel $\mathcal{P}_U(\cdot) = \mathcal{K}_{\mathcal{C}}((\cdot)\otimes\psi_{P,U}) \in \mathsf{CPTP}(\mathcal{H}^{\otimes 2n})$ that acts trivially on the multiplicity subspace, where $\mathcal{K}_{\mathcal{C}}$ is a $\mathcal{C}$-dependent quantum operation, such that*

$$\left\|\mathcal{P}_U - (\mathcal{U}\otimes\mathcal{I})^{\otimes n}\right\|_\diamond \leq 4n\sqrt{2\epsilon}.$$

We will derive the lower bound with an information-theoretic approach, by quantifying the Holevo information of an ensemble generated by a low-depth quantum circuit.

Lemma 16 ([54, 30])**.** *For any quantum state $\rho \in \mathfrak{D}(\mathcal{H})$, its von Neumann entropy is given by $S(\rho) = -\mathrm{Tr}\,(\rho\log\rho)$. For an ensemble of quantum state $\{\rho_x\,\mathrm{d}x\}_{x\in\mathcal{X}}$, its Holevo information is given by*

$$\chi\left(\{\rho_x\,\mathrm{d}x\}_{x\in\mathcal{X}}\right) = S\left(\int_{\mathcal{X}}\rho_x\,\mathrm{d}x\right) - \int_{\mathcal{X}} S(\rho_x)\,\mathrm{d}x.$$

Suppose the ensemble is supported on a $d_{\mathcal{X}}$-dimensional subspace; then its Holevo information is bounded above by

$$\chi\left(\{\rho_x\,\mathrm{d}x\}_{x\in\mathcal{X}}\right) \leq \log d_{\mathcal{X}}.$$

The data processing inequality holds for the Holevo information, i.e.,

$$\chi\left(\{\mathcal{E}(\rho_x)\,\mathrm{d}x\}_{x\in\mathcal{X}}\right) \leq \chi\left(\{\rho_x\,\mathrm{d}x\}_{x\in\mathcal{X}}\right), \quad \forall\,\mathcal{E}\in\mathsf{CPTP}(\mathcal{H}).$$

Lemma 17 (Alicki–Fannes–Winter [55], reformulated)**.** *For any quantum states $\rho,\sigma \in \mathfrak{D}(\mathcal{H})$ that differ on a $\mathfrak{d}$-dimensional subspace of $\mathcal{H}$, it holds that*

$$|S(\rho) - S(\sigma)| \leq \log\mathfrak{d}\cdot d_{\mathrm{Tr}}(\rho,\sigma) + \log 2.$$

Lemma 18. *For any $m,k\in\mathbb{N}$, it holds that*

$$\binom{m+k}{k} \geq \frac{1}{m+k+1}\left(1+\frac{k}{m+1}\right)^{m+1}\left(1+\frac{m}{k+1}\right)^{k+1}.$$

Proof. Taking the logarithm of the left-hand side, we have

$$\log\binom{m+k}{k} = \log\prod_{j=1}^{k}\left(1+\frac{m}{j}\right) = \sum_{j=1}^{k}\log\left(1+\frac{m}{j}\right).$$

Since the function $\log\left(1+\frac{m}{x}\right)$ is convex on $\mathbb{R}_{>0}$, it holds that

$$\begin{aligned}\sum_{j=1}^{k}\log\left(1+\frac{m}{j}\right) &\geq \int_0^k \log\left(1+\frac{m}{x+1}\right)\mathrm{d}x \\ &= m\log\left(\frac{m+x}{e}\right) + x\log\left(1+\frac{m}{x}\right)\Bigg|_1^{k+1} \\ &= (k+1)\log\left(1+\frac{m}{k+1}\right) + (m+1)\log\left(1+\frac{k}{m+1}\right) - \log(m+k+1).\end{aligned}$$

Taking the exponential on both sides of the inequality yields the desired result. □

Now we are ready to present the lower bound for the program cost with bounded programming error.

Theorem 19. *Programming a quantum circuit on N qubits and depth $D = \mathcal{O}(\mathrm{polylog}N)$ up to error $\epsilon = \Omega\left(1/\mathrm{polylog}N\right) < \frac{1}{32}$ in diamond norm distance requires a quantum processor with program cost c_P that satisfies*

$$c_P \geq \varpi \left(1 - \frac{\kappa}{2}\right)^2 \left(\frac{1-\varpi}{4\sqrt{2\epsilon}} - 1\right) \log_2 \frac{4e\sqrt{2\epsilon}d}{(1-\kappa/2)\varpi} - c_0$$

for any $\varpi \in (0, 1-4\sqrt{2\epsilon})$, $\kappa = \Omega(2^{-\mathrm{polylog}(N)}) < 1$, and the constant $c_0 = 5 + \frac{1}{2\log 2} \approx 5.72$.

Proof. Suppose we have a diamond norm κ-approximate unitary n-design ν_κ, where the error parameter $\kappa = \Omega(2^{-\mathrm{polylog}N})$ and $n = \lceil \frac{(1-\kappa/2)\varpi}{4\sqrt{2\epsilon}} \rceil = \mathcal{O}(1/\sqrt{\epsilon}) = \mathcal{O}(\mathrm{polylog}N)$ for some constant ϖ such that $0 < \varpi < 1 - 4\sqrt{2\epsilon}$. Using Schuster, Haferkamp, and Huang's construction [29] [cf. Table 1], the unitary design can be generated on a 1D circuit with $\mathcal{O}(\mathrm{polylog}N)$ depth, which is generalizable to any circuit geometry with the same order of depth [29, Appendix D.1] using local SWAP operations. Therefore, the Holevo information of an ensemble of states generated by n parallel uses of the unitary design on a fixed initial state presents a lower bound for the information required to program our low-depth circuits. Our derivation is based on analyzing the amount of information in the collection of program states $\chi\left(\{\psi_{P,V}\, \mathrm{d}\nu_\kappa(V)\}\right)$. Using the property of the Holevo information, when inserting any quantum state Φ_n into the channel introduced in Lemma 15, it holds that

$$\begin{aligned}
\chi\left(\{\mathcal{P}_V(\Phi_n)\, \mathrm{d}\nu_\kappa(V)\}\right) &= \chi\left(\{\mathcal{K}_\mathcal{C}\left(\Phi_n \otimes \psi_{P,V}\right)\, \mathrm{d}\nu_\kappa(V)\}\right) \\
&\leq \chi\left(\{\Phi_n \otimes \psi_{P,V}\, \mathrm{d}\nu_\kappa(V)\}\right) \\
&= \chi\left(\{\psi_{P,V}\, \mathrm{d}\nu_\kappa(V)\}\right).
\end{aligned} \tag{4}$$

We first note that for any $U \in \mathsf{U}(d)$, both $\mathcal{P}_U$ and $\mathcal{U}^{\otimes n}$ act trivially on the symmetric subspaces. Therefore, on the same input state, their outputs only differ on the subspaces that correspond to the irreducible representations of $\mathsf{U}(d)$, which admit dimension

$$d_n = \sum_{\lambda \vdash_d n} (\dim W_\lambda^d)^2.$$

Using the inequalities in Lemma 15, 16 and 17, for an ϵ-universal quantum processor,

$$\begin{aligned}
&\left|\chi\left(\{\mathcal{P}_V(\Phi_n)\, \mathrm{d}\nu_\kappa(V)\}\right) - \chi\left(\{(\mathcal{V}\otimes\mathcal{I})^{\otimes n}(\Phi_n)\, \mathrm{d}\nu_\kappa(V)\}\right)\right| \\
\leq\ & \left|S\left(\mathop{\mathbb{E}}_{V\sim\nu_\kappa}\left[\mathcal{P}_V(\Phi_n)\right]\right) - S\left(\mathop{\mathbb{E}}_{V\sim\nu_\kappa}\left[(\mathcal{V}\otimes\mathcal{I})^{\otimes n}(\Phi_n)\right]\right)\right| \\
&+ \left|\mathop{\mathbb{E}}_{V\sim\nu_\kappa}\left[S\left(\mathcal{P}_V(\Phi_n)\right) - S\left((\mathcal{V}\otimes\mathcal{I})^{\otimes n}(\Phi_n)\right)\right]\right| \\
\leq\ & \frac{\log d_n}{2}\left\|\mathop{\mathbb{E}}_{V\sim\nu_\kappa}\left[\left(\mathcal{P}_V - (\mathcal{V}\otimes\mathcal{I})^{\otimes n}\right)(\Phi_n)\right]\right\|_1 + \log 2 \\
&+ \mathop{\mathbb{E}}_{V\sim\nu_\kappa}\left[\frac{\log d_n}{2}\left\|\left(\mathcal{P}_V - (\mathcal{V}\otimes\mathcal{I})^{\otimes n}\right)(\Phi_n)\right\|_1 + \log 2\right] \\
\leq\ & \log d_n \cdot \mathop{\mathbb{E}}_{V\sim\nu_\kappa}\left[\left\|\left(\mathcal{P}_V - (\mathcal{V}\otimes\mathcal{I})^{\otimes n}\right)(\Phi_n)\right\|_1\right] + 2\log 2 \\
\leq\ & 4n\sqrt{2\epsilon}\log d_n + 2\log 2.
\end{aligned} \tag{5}$$

Furthermore, we consider the ensemble $\chi\left(\{(\mathcal{U}\otimes\mathcal{I})^{\otimes n}(\Phi_n)\, \mathrm{d}\mu(U)\}\right)$, where μ is the Haar measure on $\mathsf{U}(d)$. Analogous to the formulation in Equation 5, if we take $\Phi_n = |\Phi_n\rangle\langle\Phi_n|$

as a pure state, the von Neumann entropy in the second term vanishes, and thus

$$\begin{aligned}
&\left|\chi\left(\{(\mathcal{V}\otimes\mathcal{I})^{\otimes n}(\Phi_n)\,\mathrm{d}\nu_\kappa(V)\}\right)-\chi\left(\{(\mathcal{U}\otimes\mathcal{I})^{\otimes n}(\Phi_n)\,\mathrm{d}\mu(U)\}\right)\right| \\
&= \left|S\left(\mathop{\mathbb{E}}_{V\sim\nu_\kappa}\left[(\mathcal{V}\otimes\mathcal{I})^{\otimes n}(\Phi_n)\right]\right)-S\left(\int_{\mathsf{U}(d)}(\mathcal{U}\otimes\mathcal{I})^{\otimes n}(\Phi_n)\,\mathrm{d}\mu(U)\right)\right| \\
&\leq \frac{\log d_n}{2}\left\|\mathop{\mathbb{E}}_{V\sim\nu_\kappa}\left[(\mathcal{V}\otimes\mathcal{I})^{\otimes n}(\Phi_n)\right]-\int_{\mathsf{U}(d)}(\mathcal{U}\otimes\mathcal{I})^{\otimes n}(\Phi_n)\,\mathrm{d}\mu(U)\right\|_1+\log 2 \\
&= \frac{\log d_n}{2}\left\|\left(\mathcal{Q}^{(n)}_{\nu_\kappa}-\mathcal{Q}^{(n)}_{\mathrm{Haar}}\right)(\Phi_n)\right\|_1+\log 2 \\
&\leq \frac{\log d_n}{2}\left\|\mathcal{M}^{(n)}_{\nu_\kappa}-\mathcal{M}^{(n)}_{\mathrm{Haar}}\right\|_\diamond+\log 2 \\
&\leq \frac{\kappa}{2}\log d_n+\log 2.
\end{aligned} \tag{6}$$

Combining Equation 4, 5 and 6, we have

$$\begin{aligned}
\chi\left(\{\psi_{P,V}\,\mathrm{d}\nu_\kappa(V)\}\right) &\geq \chi\left(\{\mathcal{P}_V(\Phi_n)\,\mathrm{d}\nu_\kappa(V)\}\right) \\
&\geq \chi\left(\{(\mathcal{V}\otimes\mathcal{I})^{\otimes n}(\Phi_n)\,\mathrm{d}\nu_\kappa(V)\}\right)-4n\sqrt{2\epsilon}\log d_n-2\log 2 \\
&\geq \chi\left(\{(\mathcal{U}\otimes\mathcal{I})^{\otimes n}(\Phi_n)\,\mathrm{d}\mu(U)\}\right)-\left(4n\sqrt{2\epsilon}+\frac{\kappa}{2}\right)\log d_n-3\log 2 \\
&= S\left(\int_{\mathsf{U}(d)}(\mathcal{U}\otimes\mathcal{I})^{\otimes n}(\Phi_n)\,\mathrm{d}\mu(U)\right)-\left(4n\sqrt{2\epsilon}+\frac{\kappa}{2}\right)\log d_n-3\log 2.
\end{aligned}$$

If we take $|\Phi_n\rangle$ as the maximally entangled state

$$|\Phi_n\rangle=\bigoplus_{\lambda\in\mathsf{S}_n^d}\frac{\dim W_\lambda^d}{\sqrt{d_n}}\,|\Phi^+_{W_\lambda^d}\rangle\otimes|\eta_\lambda\rangle\,,$$

where $|\eta_\lambda\rangle$ is again an arbitrary bipartite state on $V_\lambda^{\otimes 2}$. Using [34, Lemma 3.3], the Haar random moment operator turns Φ_n into

$$\mathcal{Q}^{(n)}_{\mathrm{Haar}}(\Phi_n)=\bigoplus_{\lambda\in\mathsf{S}_n^d}\frac{I_{W_\lambda^{d\otimes 2}}}{d_n}\otimes|\eta_\lambda\rangle\langle\eta_\lambda|\,.$$

Therefore, $S\left(\mathcal{Q}^{(n)}_{\mathrm{Haar}}(\Phi_n)\right)=\log d_n$. Putting everything together, and using the upper bound [cf. Lemma 16] $\chi\left(\{\psi_{P,V}\,\mathrm{d}\nu_\kappa(V)\}\right)\leq\log d_P$, we arrive at the following bound for the cost:

$$\log d_P\geq\left(1-4n\sqrt{2\epsilon}-\frac{\kappa}{2}\right)\log d_n-3\log 2. \tag{7}$$

The explicit expression for d_n [56] is given by

$$d_n=\binom{n+d^2-1}{d^2-1}.$$

Using Lemma 18, we have

$$d_n\geq\frac{1}{n+d^2}\left(1+\frac{n}{d^2}\right)^{d^2}\left(1+\frac{d^2-1}{n+1}\right)^{n+1}\geq\frac{d^2}{(n+d^2)^2}\left(1+\frac{n}{d^2}\right)^{d^2}\left(1+\frac{d^2}{n}\right)^{n},$$

having used $\left(1+\frac{d^2-1}{n+1}\right)^{n+1} \geq \left(1+\frac{d^2-1}{n}\right)^{n} = \left(1+\frac{d^2}{n}\right)^{n}\left(1-\frac{1}{n+d^2}\right)^{n} \geq \frac{d^2}{n+d^2}\left(1+\frac{d^2}{n}\right)^{n}$. Moreover, using the Taylor series, we have

$$\left(1+\frac{n}{d^2}\right)^{d^2} = \exp\left\{d^2 \log\left(1+\frac{n}{d^2}\right)\right\} \geq \exp\left\{n-\frac{n^2}{2d^2}\right\},$$
$$\left(1+\frac{d^2}{n}\right)^{n} \geq \left(\frac{d^2}{n}\right)^{n}.$$

Since $n = \mathcal{O}(\mathrm{polylog}N)$ and $d = 2^N$, the quotient n/d is vanishing for sufficiently large N. When $d \geq n$, it holds that

$$d_n \geq \left(\frac{1}{d+\frac{n}{d}}\right)^2 \exp\left(-\frac{n^2}{2d^2}\right)\left(\frac{ed^2}{n}\right)^n \geq \frac{e^{-\frac{1}{2}}}{4d^2}\left(\frac{ed^2}{n}\right)^n \geq \frac{e^{-\frac{1}{2}}}{4}\left(\frac{ed}{n-1}\right)^n.$$

Recall that we have set $n = \lceil \frac{(1-\kappa/2)\varpi}{4\sqrt{2\epsilon}} \rceil$, and thus

$$1-4n\sqrt{2\epsilon}-\frac{\kappa}{2} \geq 1-4\sqrt{2\epsilon}\left(\frac{(1-\kappa/2)\varpi}{4\sqrt{2\epsilon}}+1\right)-\frac{\kappa}{2} = \left(1-\frac{\kappa}{2}\right)\left(1-\varpi-4\sqrt{2\epsilon}\right).$$

Using the above inequality and Equation 7,

$$\begin{aligned}\log d_P &\geq \left(1-4n\sqrt{2\epsilon}-\kappa/2\right) n \log\left(\frac{ed}{n-1}\right) - \left(5\log 2+\frac{1}{2}\right) \\ &\geq \varpi\,(1-\kappa/2)^2\left(\frac{1-\varpi}{4\sqrt{2\epsilon}}-1\right)\log\left(\frac{4e\sqrt{2\epsilon}d}{(1-\kappa/2)\varpi}\right) - \left(5\log 2+\frac{1}{2}\right),\end{aligned}$$

the desired result follows immediately from the relation $c_P = \log_2 d_P = \log d_P / \log 2$. □

Remark 20. The memory size lower bound presented in Theorem 19 has asymptotic $\log_2 d_P = \Omega\left(N\mathrm{polylog}N\right)$ when the parameters ϖ and κ are fixed as constants, $D \sim \mathrm{polylog}N$ and the programming error $\epsilon \sim 1/\mathrm{polylog}N$, which improves the bound $c_P = \Omega\left(N \log\log N\right)$ stated in [26, Theorem 7] for storing the quantum states prepared by low-depth brickwork quantum circuits.

Note that our result does not take the conventional form of the No-Programming Theorem, as is announced in [13, 30, 15, 16]. Instead of fixing the dimension of the unitaries and setting an infinitesimal programming error, we herein assume the error is bounded from below in asymptotics, yet is allowed to vanish as the scale of the circuit grows, or equivalently, $N \to \infty$. This somewhat resembles the notion of "faithful" quantum operations defined in [57, 26].

Observation 21 (High randomness implies large program cost)**.** The main proof idea of Theorem 19 is that we can quantify the Holevo information (and thus the program cost) of a low-depth circuit unitary ensemble via its capability to approximate Haar randomness (up to a circuit size-dependent moment n). We observe that a family of unitaries is hard to program if it scrambles the information well, that is, generates sufficient randomness. From the perspective of randomness generation, we can informally recover the no-programming theorem in a new way. Take a minimal-size exact unitary t-design ensemble $\{U_x\,\mathrm{d}x\}_{x\in\mathcal{X}}$, and the corresponding s-moment Choi state ensemble $\{U_x^{\otimes s}\left|\Phi_s\right\rangle\left\langle\Phi_s\right| U_x^{\dagger\otimes s}\,\mathrm{d}x\}_{x\in\mathcal{X}}$, where

$|\Phi_s\rangle = \frac{1}{\sqrt{d^2}}|I^{\otimes s}\rangle\!\rangle$. When $d^2 \lesssim s \le t$, its Holevo information reads

$$\begin{aligned}
\log_2 d_P &\ge \chi\left(\left\{U_x^{\otimes s}\,|\Phi_s\rangle\,\langle\Phi_s|\,U_x^{\dagger\otimes s}\,\mathrm{d}x\right\}_{x\in\mathcal{X}}\right) \\
&= S\left(\int_{\mathcal{X}} U_x^{\otimes s}\,|\Phi_s\rangle\,\langle\Phi_s|\,U_x^{\dagger\otimes s}\,\mathrm{d}x\right) - \int_{\mathcal{X}} S\left(U_x^{\otimes s}\,|\Phi_s\rangle\,\langle\Phi_s|\,U_x^{\dagger\otimes s}\right)\mathrm{d}x \\
&= \sum_{\lambda\vdash_d s} \frac{\dim W_\lambda^d \dim V_\lambda}{d^s}\cdot\log\left(d^s\cdot\frac{\dim W_\lambda^d}{\dim V_\lambda}\right) \\
&\ge -\log\left(\frac{1}{d^{2s}}\sum_{\lambda\vdash_d s}(\dim V_\lambda)^2\right) = 2s\log d - \log\left(\sum_{\lambda\vdash_d s}(\dim V_\lambda)^2\right) \\
&\gtrsim \frac{d^2-1}{2}\log(2s) + \frac{d-1}{2}\log(2\pi) - \frac{d^2}{2}\log d - \sum_{j=1}^{d}\log(j-1)! = \Omega(d^2\log s),
\end{aligned} \tag{8}$$

where the first inequality is due to the concavity of the logarithm, and the last asymptotic is due to Regev [56, Corollary 4.4] in the large s regime. When $s > t$, the Holevo information yields a simple upper bound: Take any t-design X_t^d, by minimality,

$$\chi\left(\left\{U_x^{\otimes s}\,|\Phi_s\rangle\,\langle\Phi_s|\,U_x^{\dagger\otimes s}\,\mathrm{d}x\right\}_{x\in\mathcal{X}}\right) \le \log\left|\left\{U_x^{\otimes s}\,|\Phi_s\rangle\,\langle\Phi_s|\,U_x^{\dagger\otimes s}\,\mathrm{d}x\right\}_{x\in\mathcal{X}}\right| \le \log|\mathsf{X}_t^d|.$$

By [49, Theorem 21], there exists a unitary t-design on $\mathsf{U}(d)$ with size $|\mathsf{X}_t^d| \le \binom{d^2+t-1}{t}^2$. Therefore, when $s > t$, by minimality of the ensemble,

$$\chi\left(\left\{U_x^{\otimes s}\,|\Phi_s\rangle\,\langle\Phi_s|\,U_x^{\dagger\otimes s}\right\}_{x\in\mathcal{X}}\right) \le \log\binom{d^2+t-1}{t}^2 = \mathcal{O}\left(d^2\log t\right).$$

This formulation reveals that the information capacity of a state ensemble carrying s copies of a unitary t-design grows with respect to s when $s \le t$, and ceases to grow when s exceeds t. Specifically, the amortized Holevo information[6] $\frac{1}{\log s}\chi(\{U_x^{\otimes s}\,|\Phi_s\rangle\,\langle\Phi_s|\,U_x^{\dagger\otimes s}\,\mathrm{d}x\}_{x\in\mathcal{X}})$ is non-vanishing for all $s \le t$. Notably, requiring the ensemble of unitaries to form an exact t-design for every $t \in \mathbb{N}$ allows taking $s \to \infty$ in Equation 8, which yields a request for infinite-size memory. Since $\mathsf{U}(d)$ is the only ensemble that satisfies the design requirement, this recovers the No-Programming Theorem.

4.2 Upper bound for the program cost

A generic upper bound can be derived easily via a metric-space covering-net argument. The idea is intuitive: If the quantum memory can store the discretizations of the space of all unitary transformations with an appropriate resolution, then a carefully designed retrieval algorithm could program any unitary approximately. We formalize the terminology below.

Definition 22. Let $(\mathscr{X}, \mathfrak{d})$ be a metric space, let $\mathscr{K} \subseteq \mathscr{X}$, and fix $\epsilon > 0$. A subset $\mathscr{N}(\mathscr{K}, \mathfrak{d}, \epsilon) \subseteq \mathscr{K}$ is called an ϵ-net of $\mathscr{K}$ if, for every $x \in \mathscr{K}$, there exists $y \in \mathscr{N}(\mathscr{K}, \mathfrak{d}, \epsilon)$ such that $\mathfrak{d}(x, y) \le \epsilon$; its cardinality, $|\mathscr{N}(\mathscr{K}, \mathfrak{d}, \epsilon)|$, is called the ϵ-covering number of $\mathscr{K}$.

Lemma 23 ([24, Lemma 9, reformulated])**.** *For any error $0 < \epsilon \le 1$, the ϵ-covering number of the d-dimensional unitary group $\mathsf{U}(d)$ with respect to the diamond norm (distance) $\|\cdot\|_\diamond$ of the induced unitary channel satisfies the following inequality:*

$$|\mathscr{N}(\mathsf{U}(d), \|\cdot\|_\diamond, \epsilon)| \le \left(\frac{12}{\epsilon}\right)^{2d^2}.$$

[6] The amortization is to exclude the effect of the growth of system size.

To derive an upper bound for the program cost, we first build an ϵ-net for the unitaries that can be generated by a low-depth quantum circuit. Using a slightly modified version of Fact 13, we can arrive at the following statement.

Lemma 24. *Let $\mathscr{U}_{k,\ell,D} \subseteq \mathsf{U}(2^N)$ be the set of N-qubit unitaries that can be generated by a brickwork circuit on ℓ k-qubit unitaries with depth D. For any $0 < \epsilon \leq 1$, the ϵ-covering number of unitary channels that corresponds to unitaries in $\mathscr{U}_{k,\ell,D}$ in diamond norm distance satisfies*

$$|\mathscr{N}(\mathscr{U}_{k,\ell,D}, \|\cdot\|_\diamond, \epsilon)| \leq \left[\left(\frac{eN}{k}\right)^k \left(\frac{12\ell}{\epsilon}\right)^{2^{2k+1}}\right]^\ell.$$

Proof. Recall Definition 7. The set of unitaries in $\mathscr{U}_{k,\ell,D}$ can be written explicitly as

$$\mathscr{U}_{k,\ell,D} = \left\{\prod_{r=1}^{D}\prod_{j\in\mathcal{L}_r} G_j^{q_j} : \quad G_j \in \mathsf{U}(2^k),\ \sum_{r=1}^{D}|\mathcal{L}_r| = \ell,\ |q_j| = k\right\},$$

where we use the superscript q_j to indicate that G_j acts on the qubits in q_j and suppress the identity on other idle qubits. The proof idea resembles that of [24, Theorem 8], as we can rewrite the product operation $\prod_{r=1}^{D}\prod_{j\in\mathcal{L}_r} G_j^{q_j} = \prod_{j=1}^{\ell} G_j^{q_j}$ with a proper indexing of the gates. Using the error propagation relation in Fact 13 together with the union bound. Let $\mathscr{Q}(k, C) = \{q \subseteq [N] : |q| = k, C[q] \text{ is connected}\}$ denote the valid k-qubit supports on which the k-local gates may act. When the circuit is all-to-all (i.e., the connectivity graph $C = K_N$, the N-complete graph), there are $\binom{N}{k}$ pairs of qubits that each local unitary gate can act on. Therefore, the cardinality of the covering net $\mathscr{N}(\mathscr{U}_{k,\ell,D}, \|\cdot\|, \epsilon)$ yields an upper bound

$$\begin{aligned}
|\mathscr{N}(\mathscr{U}_{k,\ell,D}, \|\cdot\|_\diamond, \epsilon)| &\leq \left[\left|\mathscr{Q}(k, C) \times \mathscr{N}(\mathsf{U}(2^k), \|\cdot\|_\diamond, \epsilon/\ell)\right|\right]^\ell \\
&\leq \left[|\mathscr{Q}(k, K_N)|\left|\mathscr{N}(\mathsf{U}(2^k), \|\cdot\|_\diamond, \epsilon/\ell)\right|\right]^\ell \\
&= \left[\binom{N}{k}\left|\mathscr{N}(\mathsf{U}(2^k), \|\cdot\|_\diamond, \epsilon/\ell)\right|\right]^\ell.
\end{aligned}$$

The result follows immediately via the handy inequality $\binom{N}{k} \leq \left(\frac{eN}{k}\right)^k$ and Lemma 23. □

We can readily obtain the following upper bound for the program cost.

Theorem 25. *Programming a brickwork quantum circuit on N qubits with depth D and ℓ k-local gates up to error $\epsilon \in (0, 1]$ in diamond norm distance requires a quantum processor with program cost c_P that satisfies*

$$c_P \leq k\ell \log_2\left(\frac{eN}{k}\right) + 2^{2k+1}\ell \log_2\left(\frac{12\ell}{\epsilon}\right). \tag{9}$$

Proof. By definition, it suffices to construct an ϵ-universal processor $(\mathcal{C}, \{\psi_{P,U}\}_{U\in\mathscr{U}_{k,\ell,D}})$ that can coherently program the brickwork circuit unitaries from the ϵ-net $\mathscr{N}(\mathscr{U}_{k,\ell,D}, \|\cdot\|_\diamond, \epsilon) = \{U_1, \ldots, U_M\}$. The processor can be explicitly constructed via postselection:

$$\begin{aligned}
\forall\, \rho \in \mathfrak{D}(\mathcal{H}),\ \mathcal{C}(\rho \otimes \psi_{P,U}) &= \sum_{j=1}^{M} \langle j|\, \psi_{P,U}\, |j\rangle \cdot \mathcal{U}_j(\rho), \\
\psi_{P,U} &= |t\rangle\langle t| :\ \|\mathcal{U}_t - \mathcal{U}\|_\diamond \leq \epsilon.
\end{aligned} \tag{10}$$

The set of desired programming states $\{\psi_{P,U}\}_{U\in\mathscr{U}_{k,\ell,D}}$ exists due to the definition of the ϵ-net. Finally, in the sense of ϵ-approximate programmability, the program dimension for the above scheme can be bounded from above by

$$\begin{aligned} d_P &= \dim\left(\operatorname{span}\left\{\psi_{P,U} \mid U \in \mathscr{N}(\mathscr{U}_{k,\ell,D}, \|\cdot\|_\diamond, \epsilon)\right\}\right) \\ &\leq \left|\mathscr{N}(\mathscr{U}_{k,\ell,D}, \|\cdot\|_\diamond, \epsilon)\right|. \end{aligned}$$

Taking the logarithm on both sides of the inequality, the proof is completed with the upper bound presented in Lemma 24. □

Remark 26. With the constraint $\ell \leq \frac{ND}{k}$, if we set the asymptotics of the parameters identical to those in Theorem 19, then $c_P \leq ND\log_2\left(\frac{eN}{k}\right) + \frac{2^{2k+1}}{k}ND\log_2\left(\frac{12ND}{k\epsilon}\right) = \mathcal{O}\left(N\text{polylog}N\right)$. Combining the lower bound presented in Theorem 19, we obtain a tight characterization of the asymptotic program cost for programming unitaries in $\mathscr{U}_{k,\ell,D}$ when $D \sim \text{polylog}N$, that is, $c_P = \Theta\left(N\text{polylog}N\right)$. Although learning a low-depth brickwork circuit in $\mathscr{U}_{k,\ell,D}$ can be exponentially hard in the worst case [25, Theorem 3], programming it can be efficient with prior knowledge of the circuit architecture.

4.3 Trade-off between circuit architecture complexity and local gate program cost

Although the upper bound presented in Section 4.2 is a coarse estimation, it suggests the following fact: The program state of a unitary generated by a quantum circuit encodes information not only of the parameters of the local unitary gates, but also how the gates are applied among the registers legitimately [cf. Definition 7].

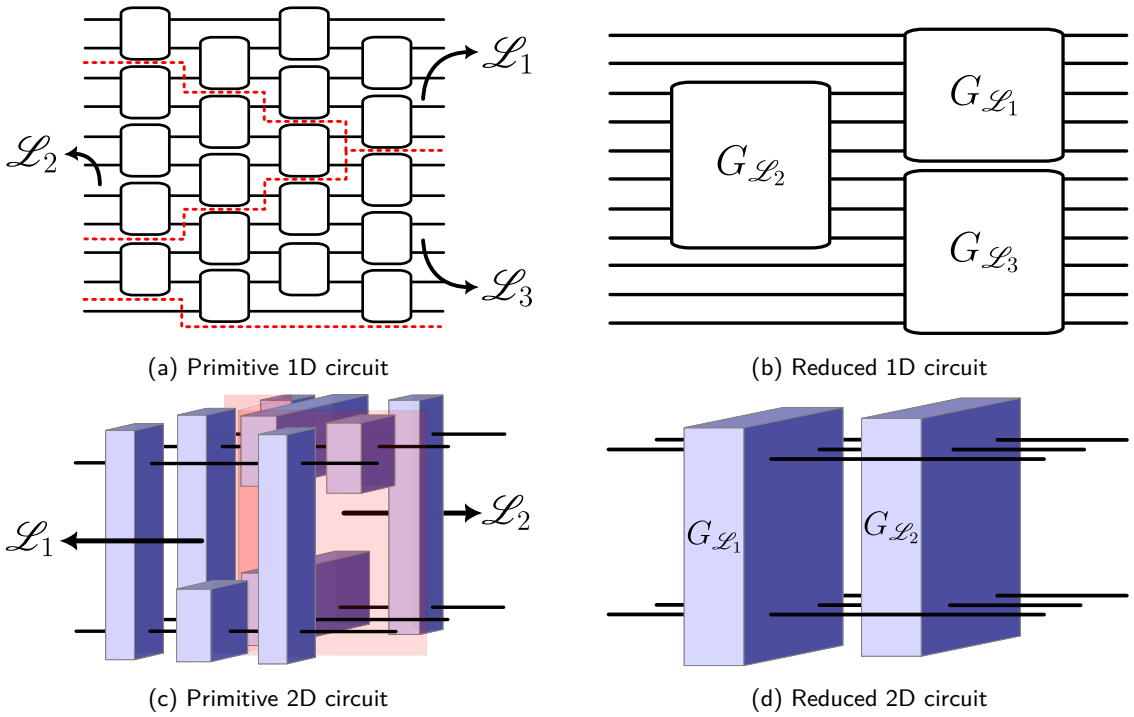
Assume that the location $(r : j \in \mathcal{L}_r;\ q_j)$ about each unitary gate G_j is encoded in a bit-string $\mathsf{L}_j \in \{0,1\}^m$, where m is a constant dependent on the circuit architecture[7], and the sender (the user who prepares the program state) and receiver (the quantum processor) have reached a consensus on the circuit architecture (the connectivity graph, circuit depth, gate count, etc.). At such, a candidate of the program state is given by the following tensor product state:

$$\psi_{P,U} = \bigotimes_{j=1}^{\ell}\left(\left|\mathsf{L}_j\right\rangle\left\langle\mathsf{L}_j\right| \otimes \psi_{P,G_j}\right).$$

While the state does not yield the most compact form, its expression reveals an underlying regularity: When the local gates become larger, the program cost of each gate increases, while there are fewer gates allowed to be placed in the circuit; thus, both the gate number ℓ and the bit-string length $|\mathsf{L}_j|$ decrease. This highlights a trade-off in program cost between describing the circuit architecture and storing individual program states of local gates.

How can this flexibility be utilized, given that the gates have fixed dimensions? We invoke the well-applied light-cone argument for general quantum circuits [22, 25, 31, 26], which indicates that the local gates can be grouped into non-intersecting light-cones without violating their relative order of implementation. Each group of small unitaries is combined into a larger unitary, while the layout of these resulting unitaries can be significantly simpler. To specify, the light-cones are defined by their first layer of gates. Starting from the first layer, each subsequent layer is constructed inductively by including the set of gates that act on the qubits affected by the gates in the previous layer. The remaining

[7]For instance, we can take m such that 2^m is greater than or equal to the number of valid pairs of qubits on any layer of the circuit, where the k-local gates can be applied. Note that the argument applies to quantum brickwork circuits with any geometry.



(a) Primitive 1D circuit (b) Reduced 1D circuit

(c) Primitive 2D circuit (d) Reduced 2D circuit

Figure 3: Examples of light-cone reduction of 1D and 2D 2-local brickwork circuits.

separated blocks between the forward light-cones are also grouped to form the backward light-cones. A schematic illustration of this reduction is provided in Figure 3.

One would naturally ask whether the reduction help reduce the program cost of brickwork circuits. To address this, we analyze the cost bounds in two scenarios: when the local unitaries are either generic or have a specific structure.

4.4 Reduction of circuits with general local unitaries

Assume that each light-cone yields depth W, and we compose the gates in light-cone $\mathscr{L}$ to form the light-cone gate $G_{\mathscr{L}}$. The circuit is thus reduced to $\lceil \frac{D}{W} \rceil$ layers. Suppose that the gates are interlaced so that the width of each light cone (i.e., the number of qubits it occupies) scales as $\Theta(W)$. After reduction, the circuit contains $\frac{D}{W} \cdot \Theta(\frac{N}{W}) = \Theta(\frac{ND}{W^2})$ gates. In comparison, the primitive circuit contains $\Theta(ND)$ gates, each being $\mathcal{O}(1)$-local. Let c_P and c_P^r denote the ϵ-universal program costs of the primitive and reduced all-to-all circuits, respectively. Applying Theorem 25 gives

$$c_P \lesssim ND \log_2 N + ND \log_2 \left(\frac{ND}{\epsilon}\right),$$
$$c_P^r \lesssim \frac{ND}{W} \log_2 \left(\frac{N}{W}\right) + 2^{\Theta(W)} \frac{ND}{W^2} \log_2 \left(\frac{ND}{W^2\epsilon}\right).$$

In the large N regime, the second terms in the bounds above dominate asymptotically. To ensure that the reduced quantum circuit is less costly, i.e., $c_P^r = o(c_P)$, one forces

$$\frac{2^{\Theta(W)}}{W^2} \log_2 \left(\frac{ND}{W^2\epsilon}\right) = o\left(\log_2 \left(\frac{ND}{\epsilon}\right)\right) \quad \Longrightarrow \quad \epsilon = \omega\left(\frac{ND}{W^{\frac{2}{1-W^2/2^{\Theta(W)}}}}\right).$$

To validate the bound with at most constant error, we require that $W^{\frac{2}{1-W^2/2^{\Theta(W)}}} = \omega(ND)$. Since the quotient $W^2/2^{\Theta(W)}$ is vanishing, for sufficiently large W, $W^{\frac{2}{1-W^2/2^{\Theta(W)}}} \leq W^{2+\varsigma}$ for some $\varsigma < 1$. Thus, the previous condition gives $W = \omega\left((ND)^{\frac{1}{2+\varsigma}}\right)$. Under the low-depth circuit assumption [cf. Definition 7], $N = \omega(D^z)$ for any $z \in \mathbb{N}$. Substitute this into the lower bound, $W = \omega(D^{\frac{z+1}{2+\varsigma}})$, contradicting the fundamental constraint $W \leq D$ when $z \geq 2$ even in the shallow-circuit setting $D = \mathcal{O}(1)$ [25].

We hereby conclude that the reduction does not save the asymptotic cost of storing the program state in the general case[8]. This is not a surprising result, as unitaries are continuous objects with an exponentially large number of free parameters. In contrast, the layout of the quantum gates within the circuit is a discrete object that yields simple and compact descriptions. In more general cases, the width of the light-cones can grow exponentially in their depth [25, Definition 10], making it even more costly to program the light-cone unitaries. Furthermore, the analysis above ensures that the bound presented in Theorem 25 is majorly optimal, without any prior assumption on the local unitaries.

4.5 Light-cone argument reduces the cost for structured local unitaries

Although the discussion provided in Section 4.4 shows that for programming brickwork circuits with generic local unitaries, the light-cone argument does not reduce program cost for us, there are certainly examples where it works. By relaxing the requirement of universality and restricting attention to a specific family of unitary operators that admit an efficient parametrization, the light-cone unitary can be described using a set of parameters whose growth rate is lower than that of the parameters defining the local unitaries it contains. Mathematically, the set $\mathsf{P}_k \subseteq \mathbb{R}^{2^{2k}}$ depicts the free real parameters of those k-local unitaries, and similarly $\mathsf{P}_{\mathscr{L}}$ depicts that of a single light-cone unitary $G_{\mathscr{L}}$[9]. Denote the program cost of the local and light-cone unitaries as $c_P(\mathsf{P}_k)$ and $c_P(\mathsf{P}_{\mathscr{L}})$ respectively. Suppose a brickwork circuit on ℓ gates is reduced to light-cones $\{\mathscr{L}_1, \ldots, \mathscr{L}_h\}$. The asymptotic program cost of the circuit is reduced when the following condition is satisfied:

$$\sum_{j=1}^{h} c_P(\mathsf{P}_{\mathscr{L}_j}) = o\left(\ell \cdot c_P(\mathsf{P}_k)\right). \tag{11}$$

For intuitiveness, we provide a concrete example[10] herein.

Example 27. Define the following set of unitaries on k qubits:

$$\mathscr{U}_{k,P} = \left\{ e^{\mathrm{i}\theta \prod_{j=1}^{k} P^{q_j}} \;\middle|\; \theta \in [0, 2\pi),\ q \subset [N],\ |q| = k \right\}, \quad P \in \{X, Y, Z\}.$$

We restrict the brickwork circuit of interest to be chosen from the set $\langle \mathscr{U}_{k,P} \rangle$, i.e., $\mathscr{U}_{k,P}$ serves as its generating set, and assume $C = K_N$. Still, the program cost originates from the uncertainty about the local gate type and which set of qubits it applies to. Thus, $\mathsf{P}_k = [0, 2\pi) \times \mathscr{Q}(k, K_N)$. It can be readily verified that $[\prod_{j=1}^{k} P^{q_j}, \prod_{j=1}^{k} P^{q'_j}] = 0$ for any set of indices q, q'. Moreover, we assume that in the light-cone there are $T_{\mathscr{L}} = o(m_{\mathscr{L}})$ distinct

[8] The light-cone argument is also invalidated in bounding the computational power of polylog$N$-depth quantum circuits [31, 58].

[9] Note that this notion is ad hoc, but we can instead treat $\mathsf{P}_{\mathscr{L}}$ as a subset of $\mathsf{P}_k^{|\mathscr{L}|}$.

[10] We believe that these examples are sparse within the universe of parameterized unitaries.

values of q, expressed as $\{q_1, \ldots, q_T\} \subseteq \mathscr{Q}(k, K_N)$. Therefore, if $\mathscr{L} = \{G_1, \ldots, G_{m_{\mathscr{L}}}\}$ where $G_t = e^{\mathrm{i}\theta_t \prod_{j=1}^{k} P^{q_{t,j}}}$, the light-cone gate can be expressed as

$$G_{\mathscr{L}} = e^{\mathrm{i}\left(\sum_{t=1}^{m_{\mathscr{L}}} \theta_t \prod_{j=1}^{k} P^{q_{t,j}}\right)}.$$

To perform an ϵ-approximate programming of each local unitary gate, we build an ϵ-net $\mathscr{N}([0, 2\pi), |\cdot|, \epsilon)$ for $[0, 2\pi)$. For any $\theta \in [0, 2\pi)$, there exists an angle $\theta_\epsilon \in \mathscr{N}([0, 2\pi), |\cdot|, \epsilon)$ such that $|\theta - \theta_\epsilon| \leq \epsilon$. Similar to Equation 10, the processor is constructed as

$$\begin{aligned}
\mathcal{C}(\rho \otimes \psi_{P,G_t})(\cdot) &= \sum_{\hat{\theta} \in \mathscr{N}([0,2\pi), |\cdot|, \epsilon)} \sum_{q \in \mathscr{Q}(k,C)} \langle \hat{\theta}, q | \, \psi_{P,G_t} \, | \hat{\theta}, q \rangle \cdot e^{\mathrm{i}\hat{\theta} \prod_{j=1}^{k} P^{q_j}} (\cdot) e^{-\mathrm{i}\hat{\theta} \prod_{j=1}^{k} P^{q_j}}, \\
\psi_{P,G_t} &= |\widetilde{\theta}_t, q_t\rangle \langle \widetilde{\theta}_t, q_t| : \ |\widetilde{\theta}_t - \theta_t| \leq \epsilon.
\end{aligned}$$

Using [59, Example 5], it follows that

$$\begin{aligned}
\frac{1}{2} \left\| \mathcal{C}(\rho \otimes \psi_{P,G_t}) - \mathcal{G}_t \right\|_\diamond &\leq \left\| e^{\mathrm{i}\widetilde{\theta}_t \prod_{j=1}^{k} P^{q_{t,j}}} - e^{\mathrm{i}\theta_t \prod_{j=1}^{k} P^{q_{t,j}}} \right\| \\
&\leq |\widetilde{\theta}_t - \theta_t| \left\| e^{\mathrm{i} \prod_{j=1}^{k} P^{q_{t,j}}} \right\| \\
&\leq \epsilon.
\end{aligned} \tag{12}$$

For the light-cone gate $G_{\mathscr{L}}$, we can rewrite its expression by our assumption, by

$$e^{\mathrm{i}\left(\sum_{t=1}^{m_{\mathscr{L}}} \theta_t \prod_{j=1}^{k} P^{q_{t,j}}\right)} = e^{\mathrm{i}\left(\sum_{r=1}^{T_{\mathscr{L}}} \left(\sum_{t:q_t=q_r} \theta_t \bmod 2\pi\right) \prod_{j=1}^{k} P^{q_{r,j}}\right)},$$

, thereby reducing the number of free parameters. Assume that $G_{\mathscr{L}}$ acts on $k_{\mathscr{L}}$ qubits and depends on $T_{\mathscr{L}}$ angles. Accordingly, we can write $\mathsf{P}_{\mathscr{L}} = [0, 2\pi)^{T_{\mathcal{L}}} \times \mathscr{Q}(k_{\mathscr{L}}, N)$. Still, we assume that the circuit is sufficiently dense with $\ell \sim ND$, $m_{\mathscr{L}_j} \sim W^2$, $h \sim \frac{ND}{W^2}$, and $k_{\mathscr{L}_j} \sim W$ for any $j \in [h]$. To ensure that the circuit is ϵ-universally programmed, the local unitaries are ϵ/ℓ-approximate, while the light-cone unitaries are ϵ/h-approximate. Specifically, the free angles for local unitaries and the light-cone unitaries are to be approximated to error ϵ/ℓ and $\epsilon/hT_{\mathscr{L}_j}$, in light of Equation 12. Therefore, the primitive circuit and the reduced circuit can be bounded from above by

$$\begin{aligned}
\sum_{j=1}^{h} c_P(\mathsf{P}_{\mathscr{L}_j}) &\lesssim \sum_{j=1}^{h} T_{\mathscr{L}_j} \log_2 \left(\frac{2\pi h T_{\mathscr{L}_j}}{\epsilon} \right) + T_{\mathscr{L}_j} k_{\mathscr{L}_j} \log_2 \left(\frac{eN}{k_{\mathscr{L}_j}} \right), \\
\ell \cdot c_P(\mathsf{P}_k) &\lesssim \ell \cdot \left(\log_2 \left(\frac{2\pi\ell}{\epsilon} \right) + k \log_2 \left(\frac{eN}{k} \right) \right).
\end{aligned}$$

One can readily verify that the condition presented in Equation 11 is satisfied. Therefore, we arrive at a processor that fully encodes not only the gate parameters but also the "light-cone-reduced" circuit architecture, operating at a lower overall program cost.

4.6 Going beyond light-cone: The landscape of reducing the program cost for parameterized quantum circuits

The failure to reduce program cost via the light-cone argument reveals a fundamental difference between the parameterization of a quantum circuit and that of the distribution it generates. The former grants us full knowledge of the circuit geometry, while the latter is

merely a "shadow" of the unitary circuit[11] [60]. Although the light-cone argument enables us to prove lower bounds on the capability of solving promise problems [7, 31] and generating complex distributions [61], simplifying the circuit layout introduces an increase in complexity inside the light-cone unitaries, as demonstrated in Section 4.4. Consequently, within the programming framework, this simplification does not fundamentally cause the parameter space to degenerate, thus maintaining the program cost.

More broadly, one may ask how to optimize more practically relevant NISQ circuit families, such as the Hardware-Efficient Ansätze (HEA) [62]. While the light-cone argument can be applied to HEA circuits with bounded gate locality (which resembles a QNC circuit), it does not reduce the parameter count [62, Section 3.2]. For a generic family of parameterized quantum circuits, we present the following proposition, as an extension to Section 4.1 and 4.2. It suggests that finding an exact parameter space of it directly implies a lower and upper bound for the program cost for its ϵ-approximate processor.

Proposition 28 (Program cost ≈ parameter count)**.** *Let $\mathscr{U}_{\mathsf{P}} = \{U(\boldsymbol{\theta})\}_{\boldsymbol{\theta}\in\mathsf{P}} \subseteq \mathsf{U}(d)$ be a unitary family parameterized by a compact subset P of a real vector space T, and suppose that P is contained in a ball of radius R centered at the origin of T. If the parameter-to-unitary map $\pi : \mathsf{P} \to \mathscr{U}_{\mathsf{P}}$ is L-Lipschitz, then for all $0 < \delta < 1 - 4\sqrt{2\epsilon}$, the program cost of ϵ-approximate programming of the family $\mathscr{U}_{\mathsf{P}}$ satisfies*

$$\left(1-\delta-4\sqrt{2\epsilon}\right)(\dim T-1)\log_2\left(1+\frac{1}{\dim T-1}\left\lceil\frac{\delta}{4\sqrt{2\epsilon}}\right\rceil\right)-\mathcal{O}(1)$$
$$\leq \log_2 d_P = c_P \leq \dim T\cdot\log_2\left(1+\frac{\mathcal{O}(RL)}{\epsilon}\right).$$

Proof. The proof idea resembles that of Theorem 19 and Theorem 25, with slight modifications. For the lower bound, applying the Schur-Weyl duality on the n-fold parameter space yields $T^{\otimes n} \cong \bigoplus_{\lambda\vdash_{\dim T} n} T_\lambda \otimes V_\lambda$. The dimension d_n in the formulation of the Holevo information is replaced by

$$d_{n,T} = \sum_{\lambda\vdash_{\dim T} n} (\dim T_\lambda)^2 = \dim\left(\vee^n(T)\right) = \binom{n+\dim T-1}{\dim T-1},$$

where $\vee^n T$ is the symmetric subspace of $T^{\otimes n}$. Then, by an analogous processor-reusing argument in Theorem 19, but without the randomness truncation due to circuit depth constraints, we can obtain a lower bound

$$\log_2 d_P \geq \left(1-4n\sqrt{2\epsilon}\right)\log_2 d_{n,T} - \mathcal{O}(1) = \left(1-4n\sqrt{2\epsilon}\right)\log_2\binom{n+\dim T-1}{\dim T-1} - \mathcal{O}(1).$$

Taking $n = \lceil\frac{\delta}{4\sqrt{2\epsilon}}\rceil$, and using the inequality $\binom{a+b}{b} \geq (1+\frac{a}{b})^b$ concludes the proof for the lower bound. As for the upper bound, the Lipschitz condition yields for $\boldsymbol{\theta}, \boldsymbol{\theta}' \in \mathsf{P}$,

$$\left\|\mathcal{U}(\boldsymbol{\theta}) - \mathcal{U}(\boldsymbol{\theta}')\right\|_\diamond \leq 2\left\|U(\boldsymbol{\theta}) - U(\boldsymbol{\theta}')\right\| \leq 2L\left\|\boldsymbol{\theta}-\boldsymbol{\theta}'\right\|_2.$$

Therefore, by Fact 13, an $\epsilon' = \epsilon/(2L)$ covering net of P in the ℓ_2 norm suffices to achieve ϵ-approximate programmability. By a standard volumetric argument [63], if we denote

[11] That is, we are only concerned about the parameters revealed by a limited collection of measurements, as per the case in quantifying the capability of quantum circuits in solving decision problems.

the leading constant $C_x = \frac{\pi^{x/2}}{\Gamma(x/2+1)}$, and use $\mathcal{B}(r)$ to denote the radius-r ball centered at the origin, it holds that

$$\begin{aligned}
&\left|\mathcal{N}(\mathsf{P}, \|\cdot\|_2, \epsilon')\right| \cdot \mathbf{Vol}\left(\mathcal{B}(\epsilon'/2)\right) \leq \mathbf{Vol}\left(\mathcal{B}(R+\epsilon'/2)\right) \\
&\quad \Longrightarrow \left|\mathcal{N}(\mathsf{P}, \|\cdot\|_2, \epsilon')\right| \leq \frac{C_{\dim T}\left(R+\epsilon'/2\right)^{\dim T}}{C_{\dim T}\left(\epsilon'/2\right)^{\dim T}} = \left(1+\frac{2R}{\epsilon'}\right)^{\dim T} = \left(1+\frac{4RL}{\epsilon}\right)^{\dim T}.
\end{aligned}$$

The upper bound is obtained readily by taking the logarithm on both sides of the inequality, concluding the proof. As a sanity check, this is consistent with Observation 21: a unitary design that matches sufficiently high moments must have a high-dimensional parameter space in which to encode its "random seed". □

Proposition 28 has established the resource equivalence between the program cost and the parameter count of unitary families. Notably, when $\dim T = d^2$ and $L, R = \mathcal{O}(1)$, it recovers the quantitative No-Programming Theorem in [15, 16]. The analysis of optimizing program cost for unitaries generated by quantum circuits of specific structures naturally reduces to optimizing the parameter count for them. However, this does not immediately apply to our analysis of the programmability of QNC circuits, as the minimal parameterization of a bounded-size quantum circuit is still open[12].

We remark that this is a difficult task, as demonstrated by several prior works. In [64], the authors define the UMCSP problem: given a complete description of a unitary U, decide whether there exists a small (in gate count) circuit that approximately implements U. Adapting to the programming task, if we are given access to the description of a generic unitary family $\mathscr{U}_\mathcal{X} = \{U_x\}_{x\in\mathcal{X}}$ and a UMCSP oracle, we can decide whether there exists a compact circuit family that approximately generates $\mathscr{U}_\mathcal{X}$, yielding a small parameter space, and thus small program cost. In [65, 66], the authors investigate the hardness of deciding whether the parameters of a given parameterized quantum circuit can be fused to reduce the parameter count. The main result of [64] shows that UMCSP with a small promise gap is in QCMA and, assuming the existence of computationally secure cryptography, is not in BQP. Even given a prior guarantee on the circuit being a parameterized circuit, [65, 66] shows that deciding whether the program cost of a parameterized circuit family can be reduced is NP-hard, if no further structural assumptions are made. In both cases, the search-to-decision reduction requires exponential queries to the decision oracle to synthesize an explicit programming scheme. This explains why constructing a general protocol for reducing program cost is intrinsically hard beyond specific examples such as that in Example 27.

Observing these hardness results, and the relations presented in Proposition 28, we raise the following question:

Problem 29. What is the hardness of deciding whether the optimal program cost of ϵ-approximate programming of a low-depth circuit unitary family $\{U_x\}_{x\in\mathcal{X}}$ is beyond or below two gapped thresholds, under different choices of ϵ? Can we synthesize the cost-optimal programming scheme via search-to-decision reduction by solving this decision problem?

[12] We conjecture that solving this leads to solving several open problems in circuit lower bound problems, for instance, unconditional quantum advantage of QNC^0 over circuits other than NC^0 [7], e.g., NC^1.

5 Conclusion

5.1 Concluding this work

In this work, we present a comprehensive analysis of the computational and storage resources required to program the unitaries generated by low-depth brickwork quantum circuits. Notably, the scaling of these requirements for programming generic unitaries is far from optimal in the low-depth regime. We find that programming N-qubit low-depth brickwork circuits can be efficient with respect to both circuit depth and program cost. When the circuit depth is polylogarithmic, information-theoretic arguments show that the program-cost scaling $\sim N\text{polylog}N$ is tight for macroscopic N, corresponding to a large-scale quantum device. This result paves the way for programmable quantum computers capable of running NISQ algorithms. We further examine whether the conventional light-cone argument can reduce the upper bound for program costs, assuming generic and unstructured local unitaries, and find that gate-wise programming is essentially optimal in the low-depth regime while the light-cone grouping only helps for certain structured circuit families.

The above discovery allows us to rethink the discussion about the performance – or resource-error trade-off – equivalence programming $\approx$ metrology $\approx$ learning [16], from the setting of universal unitaries to that of low-depth circuit unitaries. By [24, Theorem 18], there exists a unitary generated by a polylog$N$-depth brickwork circuit that can not be learned efficiently unless RingLWE is polynomial-time solvable, even with non-vanishing error, while programming it is efficient. As suggested by [67], greater difficulty in synthesizing states and unitaries corresponds to greater difficulty in learning them. Informally speaking, the hardness of programming and learning unitaries coincide in the worst and universal case but largely separate when restricted to NISQ circuits. Therefore, the previous conjecture is likely to break down.

5.2 Future perspectives

The derived program-cost bounds may be extended to noisy quantum circuits using the Stinespring dilation theorem [30]. However, restricting the resulting dilation to low-depth quantum channels and deriving a corresponding program-cost lower bound are highly nontrivial. For the upper bound, applying the covering/packing argument in our construction, or that in [15], would suffice. Meanwhile, we establish the lower bound via the randomness-program cost relation [cf. Observation 21], which is not only restricted to low-depth circuits but also applies to quantum circuits of arbitrary geometry. Our proof techniques can also be utilized to recover results in some prior works that have similar settings to unitary programming. For instance, the gate compression protocol [68] involves an input state holder (Alice) and a gate holder (Bob) who are to apply the gate on the state collaboratively with minimum communication[13], allowing failure on some randomly chosen state with negligible probability. Our techniques can be utilized to provide the upper and lower bounds for the communication cost, using concentration of the Haar measure under the Schur-Weyl basis (see, e.g., [68, 69]).

Further research on NISQ-circuit programming could include **(1)** investigating whether succinctly encoding the structural information of the circuit into the program reduces the cost; **(2)** developing efficient programming schemes for local unitary gates subject to algebraic constraints, such as stabilizer gates [70] and locally symmetric unitaries [71].

[13]The program cost is essentially a resource metric of quantum communication between the commander and the processor.

(3) developing programming schemes for noisy NISQ devices, in which the program state and the processor's POVM may be significantly degraded by Markovian or non-Markovian noise [72]. To achieve precision comparable to that of ideal, noiseless programming, explicit error correction and additional unitary queries might be required during the learning phase, thereby increasing both circuit complexity and overall program cost.

Acknowledgments

Both the authors contributed substantially to the research presented in this paper and to the preparation of the manuscript.

E.H. thanks Mingnan Zhao and Minglong Qin for insightful discussions on unitary designs. This work is supported by the National Natural Science Foundation of China via the Excellent Young Scientists Fund (Hong Kong and Macau) Project 12322516, the National Natural Science Foundation of China (NSFC)/Research Grants Council (RGC) Joint Research Scheme via Project N_HKU7107/24, the Hong Kong Research Grant Council (RGC) through the General Research Fund (GRF) grant 17303923, and the Guangdong Provincial Quantum Science Strategic Initiative via Project GDZX2403008.

References

[1] John Preskill. "Quantum Computing in the NISQ era and beyond". Quantum **2**, 79 (2018).

[2] Lov K. Grover. "A fast quantum mechanical algorithm for database search". In Proceedings of the 28th Annual ACM Symposium on Theory of Computing. Pages 212–219. STOC '96. Association for Computing Machinery (1996).

[3] Peter W. Shor. "Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer". SIAM Journal on Computing **26**, 1484–1509 (1997).

[4] Aram W. Harrow, Avinatan Hassidim, and Seth Lloyd. "Quantum algorithm for linear systems of equations". Physical Review Letters **103**, 150502 (2009).

[5] Kishor Bharti, Alba Cervera-Lierta, Thi Ha Kyaw, Tobias Haug, Sumner Alperin-Lea, Abhinav Anand, Matthias Degroote, Hermanni Heimonen, Jakob S. Kottmann, Tim Menke, Wai-Keong Mok, Sukin Sim, Leong-Chuan Kwek, and Alán Aspuru-Guzik. "Noisy intermediate-scale quantum algorithms". Reviews of Modern Physics **94**, 015004 (2022).

[6] Yunchao Liu. "Shallow quantum circuits: Algorithms, complexity, and fault tolerance". PhD thesis. University of California, Berkeley. (2024).

[7] Sergey Bravyi, David Gosset, and Robert König. "Quantum advantage with shallow circuits". Science **362**, 308–311 (2018).

[8] Adam Bene Watts, Robin Kothari, Luke Schaeffer, and Avishay Tal. "Exponential separation between shallow quantum circuits and unbounded fan-in shallow classical circuits". In Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing. Pages 515–526. STOC '19. Association for Computing Machinery (2019).

[9] Anne Broadbent. "Delegating private quantum computations". Canadian Journal of Physics **93**, 941–946 (2015).

[10] Alex Lombardi, Fermi Ma, and John Wright. "A one-query lower bound for unitary synthesis and breaking quantum cryptography". In Proceedings of the 56th Annual ACM Symposium on Theory of Computing. Pages 979–990. STOC '24. Association for Computing Machinery (2024).

[11] M. A. Nielsen and Isaac L. Chuang. "Programmable quantum gate arrays". Physical Review Letters **79**, 321–324 (1997).

[12] Mark Hillery, Vladimír Bužek, and Mário Ziman. "Probabilistic implementation of universal quantum processors". Physical Review A **65**, 022301 (2002).

[13] D. Gross, J. Eisert, N. Schuch, and D. Perez-Garcia. "Measurement-based quantum computation beyond the one-way model". Physical Review A **76**, 052315 (2007).

[14] Satoshi Ishizaka and Tohya Hiroshima. "Asymptotic teleportation scheme as a universal programmable quantum processor". Physical Review Letters **101**, 240501 (2008).

[15] Aleksander M. Kubicki, Carlos Palazuelos, and David Pérez-García. "Resource quantification for the no-programing theorem". Physical Review Letters **122**, 080505 (2019).

[16] Yuxiang Yang, Renato Renner, and Giulio Chiribella. "Optimal universal programming of unitary gates". Physical Review Letters **125**, 210501 (2020).

[17] Martina Gschwendtner, Andreas Bluhm, and Andreas Winter. "Programmability of covariant quantum channels". Quantum **5**, 488 (2021).

[18] Garazi Muguruza and Florian Speelman. "Port-Based State Preparation and Applications". Quantum **8**, 1573 (2024).

[19] Satoshi Yoshida, Jisho Miyazaki, and Mio Murao. "Quantum advantage in storage and retrieval of isometry channels". Physical Review Letters **136**, 190601 (2026).

[20] Alessandro Bisio, Giulio Chiribella, Giacomo Mauro D'Ariano, Stefano Facchini, and Paolo Perinotti. "Optimal quantum learning of a unitary transformation". Physical Review A **81**, 032324 (2010).

[21] Sanjeev Arora and Boaz Barak. "Computational complexity: A modern approach". Cambridge University Press. (2009).

[22] Jonas Haferkamp, Philippe Faist, Naga B. T. Kothakonda, Jens Eisert, and Nicole Yunger Halpern. "Linear growth of quantum circuit complexity". Nature Physics **18**, 528–532 (2022).

[23] Daniel Belkin, James Allen, Soumik Ghosh, Christopher Kang, Sophia Lin, James Sud, Frederic T. Chong, Bill Fefferman, and Bryan K. Clark. "Approximate t-designs in generic circuit architectures". PRX Quantum **5**, 040344 (2024).

[24] Haimeng Zhao, Laura Lewis, Ishaan Kannan, Yihui Quek, Hsin-Yuan Huang, and Matthias C. Caro. "Learning quantum states and unitaries of bounded gate complexity". PRX Quantum **5**, 040306 (2024).

[25] Hsin-Yuan Huang, Yunchao Liu, Michael Broughton, Isaac Kim, Anurag Anshu, Zeph Landau, and Jarrod R. McClean. "Learning shallow quantum circuits". In Proceedings of the 56th Annual ACM Symposium on Theory of Computing. Pages 1343–1351. STOC '24. Association for Computing Machinery (2024).

[26] Yuxiang Yang. "Compression of quantum shallow-circuit states". Physical Review Letters **134**, 010603 (2025).

[27] M. Cerezo, Andrew Arrasmith, Ryan Babbush, Simon C. Benjamin, Suguru Endo, Keisuke Fujii, Jarrod R. McClean, Kosuke Mitarai, Xiao Yuan, Lukasz Cincio, and Patrick J. Coles. "Variational quantum algorithms". Nature Reviews Physics **3**, 625–644 (2021).

[28] Fernando G. S. L. Brandão, Aram W. Harrow, and Michał Horodecki. "Local random quantum circuits are approximate polynomial-designs". Communications in Mathematical Physics **346**, 397–434 (2016).

[29] Thomas Schuster, Jonas Haferkamp, and Hsin-Yuan Huang. "Random unitaries in extremely low depth". Science **389**, 92–96 (2025).

[30] Michael A. Nielsen and Isaac L. Chuang. "Quantum computation and quantum information: 10th anniversary edition". Cambridge University Press. (2012).

[31] Shivam Nadimpalli, Natalie Parham, Francisca Vasconcelos, and Henry Yuen. "On the Pauli spectrum of QAC^0". In Proceedings of the 56th Annual ACM Symposium on Theory of Computing. Pages 1498–1506. STOC '24. Association for Computing Machinery (2024).

[32] Roe Goodman and Nolan R. Wallach. "Symmetry, representations, and invariants". Springer New York. (2009).

[33] Aram W. Harrow. "Applications of coherent classical communication and the Schur transform to quantum information theory". PhD thesis. Massachusetts Institute of Technology. (2005).

[34] Tony Metger, Alexander Poremba, Makrand Sinha, and Henry Yuen. " Simple Constructions of Linear-Depth t-Designs and Pseudorandom Unitaries ". In 2024 IEEE 65th Annual Symposium on Foundations of Computer Science (FOCS). Pages 485–492. IEEE Computer Society (2024).

[35] Christoph Dankert, Richard Cleve, Joseph Emerson, and Etera Livine. "Exact and approximate unitary 2-designs and their application to fidelity estimation". Physical Review A **80**, 012304 (2009).

[36] Joseph Emerson, Robert Alicki, and Karol Zyczkowski. "Scalable noise estimation with random unitary operators". Journal of Optics B: Quantum and Semiclassical Optics **7**, S347–S352 (2005).

[37] Lov Grover and Terry Rudolph. "Creating superpositions that correspond to efficiently integrable probability distributions" (2002). arXiv:quant-ph/0208112.

[38] Hari Krovi. "An efficient high dimensional quantum Schur transform". Quantum **3**, 122 (2019).

[39] Adam Burchardt, Jiani Fei, Dmitry Grinko, Martin Larocca, Maris Ozols, Sydney Timmerman, and Vladyslav Visnevskyi. "High-dimensional quantum Schur transforms" (2025). arXiv:2509.22640.

[40] Xiao-Ming Zhang, Tongyang Li, and Xiao Yuan. "Quantum state preparation with optimal circuit depth: Implementations and applications". Physical Review Letters **129**, 230504 (2022).

[41] John Watrous. "The theory of quantum information". Cambridge University Press. (2018).

[42] Aram W. Harrow and Saeed Mehraban. "Approximate unitary t-designs by short random quantum circuits using nearest-neighbor and long-range gates". Communications in Mathematical Physics **401**, 1531–1626 (2023).

[43] Jeongwan Haah, Yunchao Liu, and Xinyu Tan. "Efficient approximate unitary designs from random Pauli rotations". In 2024 IEEE 65th Annual Symposium on Foundations of Computer Science (FOCS). Pages 463–475. (2024).

[44] Chi-Fang Chen, Jeongwan Haah, Jonas Haferkamp, Yunchao Liu, Tony Metger, and Xinyu Tan. "Incompressibility and spectral gaps of random circuits". In 2025 IEEE 66th Annual Symposium on Foundations of Computer Science (FOCS). Pages 1304–1312. (2025).

[45] G. Chiribella, G. M. D'Ariano, and M. F. Sacchi. "Optimal estimation of group transformations using entanglement". Physical Review A **72**, 042338 (2005).

[46] Jeongwan Haah, Robin Kothari, Ryan O'Donnell, and Ewin Tang. "Query-optimal estimation of unitary channels in diamond distance". In 2023 IEEE 64th Annual Symposium on Foundations of Computer Science (FOCS). Pages 363–390. (2023).

[47] Vivek V. Shende, Stephen S. Bullock, and Igor L. Markov. "Synthesis of quantum logic circuits". In Proceedings of the 2005 Asia and South Pacific Design Automation Conference. Pages 272–275. ASP-DAC '05. Association for Computing Machinery (2005).

[48] A. Kitaev, A. Shen, and M. Vyalyi. "Classical and quantum computation". American Mathematical Society. (2002).

[49] Aidan Roy and A. J. Scott. "Unitary designs and codes". Designs, Codes and Cryptography **53**, 13–31 (2009).

[50] Marek Mozrzymas, Michał Studziński, and Piotr Kopszak. "Optimal Multi-port-based Teleportation Schemes". Quantum **5**, 477 (2021).

[51] Jiani Fei, Sydney Timmerman, and Patrick Hayden. "Efficient quantum algorithm for port-based teleportation" (2023). arXiv:2310.01637.

[52] Dmitry Grinko, Adam Burchardt, and Maris Ozols. "Efficient quantum circuits for port-based teleportation" (2024). arXiv:2312.03188.

[53] Adam Wills, Min-Hsiu Hsieh, and Sergii Strelchuk. "Efficient algorithms for all port-based teleportation protocols". PRX Quantum **5**, 030354 (2024).

[54] Alexander Semenovich Holevo. "Bounds for the quantity of information transmitted by a quantum communication channel". Problemy Peredachi Informatsii **9**, 3–11 (1973). url: https://www.mathnet.ru/eng/ppi903.

[55] Andreas Winter. "Tight uniform continuity bounds for quantum entropies: Conditional entropy, relative entropy distance and energy constraints". Communications in Mathematical Physics **347**, 291–313 (2016).

[56] Amitai Regev. "Asymptotic values for degrees associated with strips of Young diagrams". Advances in Mathematics **41**, 115–136 (1981).

[57] Yuxiang Yang, Giulio Chiribella, and Masahito Hayashi. "Optimal compression for identically prepared qubit states". Physical Review Letters **117**, 090502 (2016).

[58] Anurag Anshu, Yangjing Dong, Fengning Ou, and Penghui Yao. "On the computational power of QAC^0 with barely superlinear ancillae". In Proceedings of the 57th Annual ACM Symposium on Theory of Computing. Pages 1476–1487. STOC '25. Association for Computing Machinery (2025).

[59] A B Aleksandrov and V V Peller. "Operator Lipschitz functions". Russian Mathematical Surveys **71**, 605–702 (2016).

[60] Jonathan Kunjummen, Minh C. Tran, Daniel Carney, and Jacob M. Taylor. "Shadow process tomography of quantum channels". Physical Review A **107**, 042403 (2023).

[61] Anurag Anshu, Nikolas P. Breuckmann, and Chinmay Nirkhe. "NLTS Hamiltonians from good quantum codes". In Proceedings of the 55th Annual ACM Symposium on Theory of Computing. Pages 1090–1096. STOC '23. Association for Computing Machinery (2023).

[62] Lorenzo Leone, Salvatore F.E. Oliviero, Lukasz Cincio, and M. Cerezo. "On the practical usefulness of the hardware efficient ansatz". Quantum **8**, 1395 (2024).

[63] Roman Vershynin. "Introduction to the non-asymptotic analysis of random matrices". Pages 210–268. Cambridge University Press. (2012).

[64] Nai-Hui Chia, Chi-Ning Chou, Jiayu Zhang, and Ruizhe Zhang. "Quantum Meets the Minimum Circuit Size Problem". In Mark Braverman, editor, 13th Innovations in Theoretical Computer Science Conference (ITCS 2022). Volume 215 of Leibniz International Proceedings in Informatics (LIPIcs), pages 47:1–47:16. Schloss Dagstuhl - Leibniz-Zentrum für Informatik (2022).

[65] John van de Wetering and Matt Amy. "Optimising quantum circuits is generally hard" (2024). arXiv:2310.05958.

[66] John van de Wetering, Richie Yeung, Tuomas Laakkonen, and Aleks Kissinger. "Optimal compilation of parametrised quantum circuits". Quantum **9**, 1828 (2025).

[67] Nai-Hui Chia, Daniel Liang, and Fang Song. "Quantum state and unitary learning implies circuit lower bounds". In Nika Haghtalab and Ankur Moitra, editors, Proceedings of Thirty Eighth Conference on Learning Theory. Volume 291 of Proceedings of Machine Learning Research, pages 1194–1252. PMLR (2025). url: https://proceedings.mlr.press/v291/chia25a.html.

[68] G. Chiribella, Y. Yang, and C. Huang. "Universal superreplication of unitary gates". Physical Review Letters **114**, 120504 (2015).

[69] Giulio Chiribella and Yuxiang Yang. "Quantum superreplication of states and gates". Frontiers of Physics **11**, 110304 (2016).

[70] Héctor J. García, Igor L. Markov, and Andrew W. Cross. "On the geometry of stabilizer states". Quantum Information and Computation **14**, 683–720 (2014).

[71] Iman Marvian. "Restrictions on realizable unitary operations imposed by symmetry and locality". Nature Physics **18**, 283–289 (2022).

[72] Satvik Singh and Nilanjana Datta. "Information storage and transmission under Markovian noise". PRX Quantum **7**, 020312 (2026).