

Approximate Quantum Error Correction with 1D Log-Depth Circuits

Guoding Liu^{1,*}, Zhenyu Du^{1,*}, Zi-Wen Liu^{2,†} and Xiongfeng Ma^{1,‡}

¹Center for Quantum Information, Institute for Interdisciplinary Information Sciences, *Tsinghua University*, Beijing 100084, China

²Yau Mathematical Sciences Center, *Tsinghua University*, Beijing 100084, China

 (Received 7 April 2025; revised 14 September 2025; accepted 6 January 2026; published 13 February 2026)

Efficient and high-performance quantum error correction is essential for achieving fault-tolerant quantum computing. Low-depth random circuits offer a promising approach to identifying effective and practical encoding strategies. In this work, we rigorously prove through information-theoretic analysis that one-dimensional logarithmic-depth random Clifford encoding circuits can achieve high quantum error correction performance. We demonstrate that these random codes typically exhibit good approximate quantum error correction capability by proving that their encoding rate achieves the hashing bound for Pauli noise and the channel capacity for erasure errors. We show that the error correction inaccuracy decays once a threshold of logarithmic depth is exceeded, resulting in negligible recovery errors. This threshold is shown to be lower than that of the simple separate block encoding, and the decay rate is higher. We further establish that these codes are optimal by proving that logarithmic depth is necessary to maintain a constant encoding rate and high error correction performance. To prove our results, we propose decoupling theorems tailored for one-dimensional low-depth circuits. These results also imply strong decoupling and rapid thermalization properties in low-depth random circuits and have potential applications in quantum information science and physics.

DOI: [10.1103/7rzk-2jyh](https://doi.org/10.1103/7rzk-2jyh)

I. INTRODUCTION

Quantum error correction [1–7] is crucial for protecting quantum computers from noise, so as to enable universal fault-tolerant quantum computing [8–13], and ultimately realize scalable quantum speedups. Significant experimental and theoretical progress has been made to reduce resource requirements and enhance the performance of quantum error correction. In experiments, surface codes on a two-dimensional system have shown remarkable success in preserving logical information over extended timescales [6,7]. Nonetheless, the surface code has a fundamental drawback of a low encoding rate. The overhead of physical qubits to make one logical qubit is large for near-term devices [14].

Recent research on quantum low-density parity-check (qLDPC) codes [11,14–18] establishes constant-rate

encoding with robust error-correcting capability. Nonetheless, due to the requirement of substantial long-range connectivity, the realization of high-performance qLDPC codes needs all-to-all circuit architectures or prohibitively large circuit depth in low-dimensional systems, which makes practical implementation challenging. Finding encoding schemes with low overhead and minimal device requirements is crucial for achieving practical quantum error correction in near-term quantum devices. This leads to the central question that underlies this work: What circuit architecture and depth can achieve good quantum error correction performance? Ideally, we desire simple circuit geometry and low-depth implementation. Understanding this would be crucial for both the theory and experimental implementation of quantum error correction.

Random quantum circuits offer a promising approach to evaluating the typical performance of various encoding circuits and identifying practical encoding schemes [13, 19–23]. Here, we conduct the analysis using the approximate quantum error correction (AQEC) [24–27] framework, which enables rigorous quantitative analysis for the average behaviors of random circuit ensembles and facilitates analyzing the logical error and channel capacity of generic noise models [26,28]. Under suitable conditions, AQEC ensures negligible recovery errors between

*These authors contributed equally to this work.

†Contact author: zwliu0@tsinghua.edu.cn

‡Contact author: xma@tsinghua.edu.cn

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

the initial logical state and the state after decoding, which guarantees feasibility for practical applications and allows more flexible design of codes with various advantages [24,29]. Moreover, it provides insights into the role of symmetry in quantum information [22] and deepens our understanding of quantum complexity and order in many-body quantum systems [30].

To address the motivating question, we investigate the AQEC behaviors of random circuit encodings with locality constrained by a one-dimensional (1D) geometry. Previous studies numerically found the 1D $O(\log n)$ -depth local random circuits with a brickwork structure can approximately correct a linear number of random erasure errors while preserving a constant encoding rate [21]. Further numerical results show that such circuits can approach the hashing bound for local Pauli noise [23], matching the performance of random stabilizer codes [28]. A simple block-encoding argument provides intuition for the robustness against local noise: the n -qubit system is divided into $O(n/\log n)$ blocks, each encoding $\Omega(\log n)$ qubits as a random stabilizer code. This structure corrects noise within individual blocks, making it effective against local noise.

Despite certain numerical evidence and intuition, a rigorous understanding of the encoding properties under general noises has been elusive. Intuitively, 1D random circuits with interconnections between blocks can potentially further enhance the error correction capability due to a stronger scrambling power. A comprehensive analysis would be important, but it poses significant mathematical challenges.

In this work, we present the construction of random codes implemented in 1D and logarithmic-depth circuits, along with information-theoretically rigorous proofs of their powerful AQEC performance. We specifically analyze two types of 1D random circuit architectures: the standard 1D local random circuit with a brickwork structure [31,32] and the double-layer blocked circuit composed of two layers of blocked gates [33,34]. We will refer to these as the 1D brickwork circuit and the double-layer blocked circuit, respectively. Compared to the separate block model discussed before, there exist connections between blocks in these two models. There have been extensive prior studies on the approximate unitary design generation properties of such circuits [31–34]. Here we aim to examine their application as encoding schemes, which is morally relevant to design generation but essentially a different problem (as will be further explained later). For example, the 1D brickwork circuit is known to form an approximate 2-design at $O(n)$ depth, and the double-layer blocked circuit achieves this at $O(\log n)$ depth. However, this finding alone does not guarantee a good encoding scheme at the same depth. To protect logical information with a constant encoding rate, the size of the logical space is exponential in n , which requires an exponentially small error in the approximate design. Since the approximate 2-design

error at $O(\log n)$ depth is only polynomially small [33,34], our proofs for the performance of these encoding circuits require new technical work. By demonstrating their effectiveness in decoupling and error correction tasks, our research complements the common study from a design perspective [33,34].

We show that at a logarithmic depth, two encoding schemes can tolerate constant-rate stochastic noises, including local Pauli, erasure, amplitude damping noises, and weak-correlated noises, and maintain a constant encoding rate. For the double-layer blocked circuit, we prove that they can tolerate a higher encoding rate for local noises at $\omega(\log n)$ depth. For instance, they tolerate strength- $\vec{p}$ local Pauli noise with an encoding rate k/n , provided that $1 - k/n - h(\vec{p}) > 0$. This achieves the well-known hashing bound of the channel capacity for local Pauli noise [28,35,36]. For independently and identically distributed (i.i.d.) erasure errors with error rate p , the encoding rate can achieve the quantum channel capacity $1 - 2p$.

Beyond the achievable encoding rate, we show that the recovery error of the two encoding schemes decays exponentially with the depth of encoding circuits. The proof is based on transforming the recovery error after noise and decoding into a decoupling problem, and we establish decoupling theorems for two types of 1D low-depth circuits. Particularly, we show that the error polynomially decays for $O(\log n)$ -depth circuits at a constant encoding rate and vanishes for $\omega(\log n)$ -depth circuits at a higher encoding rate. Compared to a previous result that $O(\log^3 n)$ -depth all-to-all Clifford circuits establish decoupling against a linear-size adversarial error [20,37], our work focuses on decoupling against constant-rate stochastic noise. This relaxation is more practically relevant, and we show that 1D $O(\log n)$ -depth circuits generally suffice for decoupling in this case. Moreover, we compare the performance between the double-layer blocked scheme and the intuitive separate block approach. We find that for certain noise and encoding rate regimes, the double-layer blocked scheme requires a lower depth to make the recovery error decay. Also, the decay rate associated with the double-layer blocked scheme is better than the separate block approach. This advantage also holds for 1D brickwork circuits.

Furthermore, we establish lower bounds on the circuit depth required to achieve high AQEC performance. Specifically, for a D -dimensional circuit, we prove that the recovery error decays polynomially or that a constant encoding rate is achieved, the circuit depth must be at least $\Omega[(\log n)^{1/D}]$. Therefore, our encoding schemes are optimal since they attain the 1D lower bound $\Omega(\log n)$. This lower bound also generalizes existing results for the circuit depth required to achieve a high code distance or for other quantum error correction and fault-tolerant quantum computing metrics [30,38].

Beyond the crucial role in quantum computing and quantum error correction, our results on low-depth random circuits are expected to find applications in quantum many-body physics. The decoupling property and the robustness of random circuits against errors relate to their capacity to scramble information and generate entanglement [20,37,39]. Haar-random unitary and random Clifford operations exhibit strong scrambling power and naturally have good quantum error correction performance. For instance, random stabilizer codes [19] achieve the well-known Singleton bound, $n - k - 2(d - 1) \geq 0$ [40], with a constant encoding rate k/n and a linear code distance $d = \Theta(n)$. However, Haar-random operations require circuit depth exponential in the number of qubits for implementation. Emulating the utility of Haar-random unitaries using low-dimensional, low-depth random circuits is essential for many practical applications like random circuit sampling [41–46], quantum device benchmarking [47–53], and quantum gravity [54–56]. Our results also imply that subsystems of 1D low-depth random circuit states rapidly thermalize, providing a powerful tool for investigating quantum many-body dynamics [57–59].

The paper is organized as follows: In Sec. II, we introduce the low-depth random circuit models and summarize our main results. We also outline the theoretical approach to analyze the AQEC performance. In Sec. III, we introduce the preliminaries for this work. In Secs. IV and V, we elaborate on the main results about the AQEC performance of random codes, as well as the lower bounds for the circuit depth required for AQEC. Finally, in Sec. VI, we conclude with a discussion of our results and implications for future research.

II. CIRCUIT MODELS AND SUMMARY OF RESULTS

In this work, we consider randomized code constructions, where the encoding operation of a code is randomly drawn from a circuit ensemble. As mentioned earlier, we mainly analyze two types of 1D random circuit ensembles and prove their decoupling properties and AQEC performance at a logarithmic depth.

The 1D brickwork circuit represents a relatively standard circuit architecture. We consider the case with the periodic boundary condition, and let q be the local dimension. The open boundary condition is similar. Without loss of generality, here we consider the total number of qudits n to be even. For an n -qudit random unitary operation U composed of s two-qudit gates, $\{U^{(1)}, U^{(2)}, \dots, U^{(s)}\}$, with $U^{(i)}$ randomly sampled from the unitary group of dimension q^2 with support $A^{(i)} \subset [n] = \{1, 2, \dots, n\}$, U has an expression

$$U = U_{A^{(s)}}^{(s)} U_{A^{(s-1)}}^{(s-1)} \cdots U_{A^{(1)}}^{(1)}. \quad (1)$$

For the task of decoupling, each gate $U^{(i)}$ can be sampled from a unitary 2-design group, like the Clifford group, instead of the whole unitary group.

The 1D depth- d brickwork circuit is composed of alternating layers of two-qudit gates in a staggered arrangement, with the formal definition shown below:

$$U = \prod_{l=1}^d U^{[l]}, \quad (2)$$

where

$$U^{[l]} = \begin{cases} \bigotimes_{i=1}^{n/2} U_{\{2i-1, 2i\}}^{((l-1)n/2+i)}, & \text{if } \text{mod}(l, 2) = 1, \\ \bigotimes_{i=1}^{n/2} U_{\{2i, 2i+1\}}^{((l-1)n/2+i)}, & \text{if } \text{mod}(l, 2) = 0. \end{cases} \quad (3)$$

Here, $U^{[l]}$ is the unitary gate on layer l . The qudit index $n+1$ is identified with 1 to match the periodic boundary condition. In our work, we set depth $d = \log(n/\varepsilon) + \log n / \log((q^2 + 1)/2q) + \log(e-1) / \log((q^2 + 1)/2q) + 1$. The local dimension q is a prefixed constant, and the depth is specified by the parameter ε . We denote this 1D brickwork circuit ensemble by $\mathfrak{B}_n^\varepsilon$. When the number of qubits and circuit depth are not of primary concern, we simply refer to this architecture as $\mathfrak{B}$. A diagram of a 1D brickwork circuit is shown in Fig. 1(a). This brickwork structure is particularly well suited for experimental implementation, as it can be realized by applying gates between nearest-neighbor qudits without complex gate compilation. Furthermore, this circuit is an excellent candidate for generating randomness and entanglement. After a depth of d , the information initially contained in any single qudit is scrambled across $2d$ qudits.

In this work, we will use circuit U as the encoding map which encodes 2^k -dimension logical space into 2^n physical space. To use this circuit for encoding, we set the local dimension of each qudit to $q = 2^{n/k}$, where a single qubit of information is encoded per qudit. By selecting n/k as a constant, the local dimension q also remains constant. This configuration yields a constant encoding rate of k/n . The total number of qudits is correspondingly changed to k , which is equivalent to n qubits.

The diagram of the double-layer blocked circuit is shown in Fig. 1(b). This circuit has a two-layer structure, with each layer consisting of blocks of random gates sampled from a unitary 2-design group, where each block has size ξ and depth $d = O(\xi)$ [33]. In the first layer, qubits are partitioned into blocks, and random gates are applied within each block to generate entanglement. The second layer then applies random gates between these blocks, thereby creating interblock connections. Similar to the 1D brickwork circuit, this double-layer blocked circuit scrambles the information of a single qubit across a number of

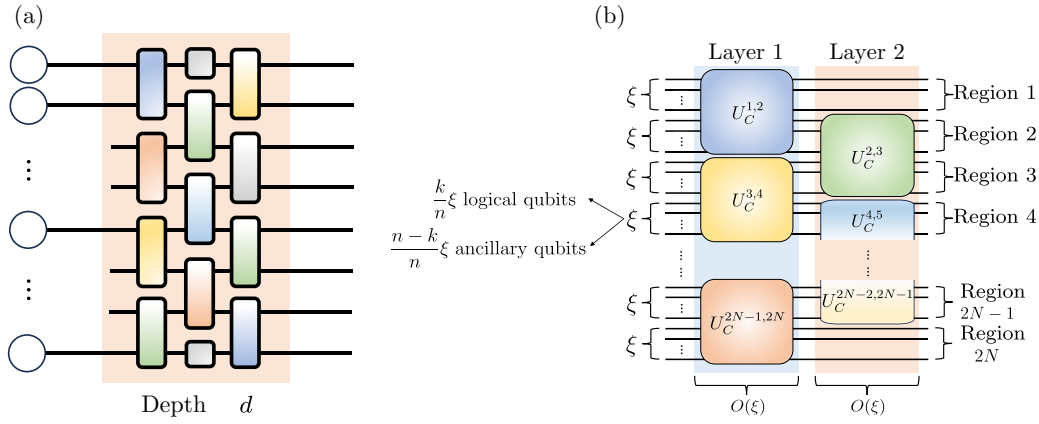


FIG. 1. (a) Diagram of the 1D brickwork circuit. The circuit is constructed by brickwise two-qudit gates. The total qudit number is k , and the circuit depth is d . When using this circuit for encoding, we set the local dimension as $q = 2^{n/k}$ while only one qubit of information is encoded in one qudit. In our work, n/k is chosen as a constant, so q is a constant. The encoding rate maintains k/n in this scheme, and the total physical qubit number is n . Note that our results apply to arbitrary local dimensions, including 1D brickwork circuits composed of two-qubit gates acting on n qubits, where each qubit encodes logical information with probability k/n . The Choi error still decays polynomially, though with a different scaling. Further discussions are provided in Appendix C3. (b) Diagram of the 1D double-layer blocked circuit. Two layers of blocked unitary gates construct the circuit. The term $U_C^{i,i+1}$ is a random gate from a unitary 2-design ensemble on regions i and $i + 1$. The circuit is divided into $2N$ regions, with each region having $\xi = n/2N$ qubits. When using this circuit for encoding, the logical qubits and ancillary qubits are evenly distributed across each region. Each region contains $(k/n)\xi$ logical qubits and $((n - k)/n)\xi$ ancillary qubits.

qubits that is proportional to the circuit's depth. Theoretically, the double-layer blocked structure is simpler than the 1D brickwork circuit, which facilitates further theoretical analysis.

Specifically, in the double-layer blocked circuit, the n qubits are divided into $2N$ regions, with each region containing $\xi = n/2N$ qubits. In the first layer, unitaries $U_C^{2i-1,2i}$ are applied in parallel on regions $2i - 1$ and $2i$ for $1 \leq i \leq N$. In the second layer, unitaries $U_C^{2i,2i+1}$ are applied in parallel on regions $2i$ and $2i + 1$ for $1 \leq i < N$. The entire encoding unitary circuit can be written as

$$U = U_2 \cdot U_1, \quad (4)$$

where

$$U_1 = \bigotimes_{i=1}^N U_C^{2i-1,2i}, \quad U_2 = \bigotimes_{i=1}^{N-1} U_C^{2i,2i+1}. \quad (5)$$

The unitaries $U_C^{i,i+1}$ are sampled from a unitary 2-design ensemble. In this work, we take it to be the Clifford group. Since a ξ -qubit Clifford gate can be realized in depth $O(\xi)$ with basic CNOT, Hadamard, and phase gates in 1D circuits [60,61], the depth of the random circuit U is $O(\xi)$. In this work, we set $\xi = \log(n/\varepsilon)$ and denote the ensemble generated by this 1D random double-layer blocked circuit as $\mathcal{C}_n^\varepsilon$, where ε is a parameter. We would refer to this circuit architecture as $\mathcal{C}$ when parameters are not essential.

We use the double-layer blocked circuit as the encoding map in the following way. The circuit U encodes

the k -qubit logical system L to an n -qubit physical system S . In the encoding procedure, the logical qubits and ancillary qubits are evenly distributed across each region, such that each region contains $(k/n)\xi$ logical qubits and $((n - k)/n)\xi$ ancillary qubits as shown in Fig. 1(b).

We focus on the performance of codes when the encoding circuits are of low depth, that is, when depth $d \ll n$. Due to the light-cone constraint, each qubit can interact with only $O(d)$ other qubits, limiting the spread of information across the system. Consequently, any encoding unitary from the 1D random circuits can only achieve a code distance on the order of circuit depth [30]. Fortunately, by considering AQEC, the number of qubits protected against local noise and limited correlated noise can be much larger than the code distance. Here, the AQEC performance is quantified by the Choi error ϵ_{choi} between the identity operation and $\mathcal{D} \circ \mathcal{N} \circ \mathcal{E}$, where $\mathcal{E}$, $\mathcal{N}$, and $\mathcal{D}$ represent the encoding, noise, and decoding channels, respectively. This error measures the average distance between the initial and recovered states over all inputs, using an optimal decoding channel $\mathcal{D}$. A detailed introduction to AQEC is provided in Sec. III B.

We rigorously analyze a host of important noise models, including local Pauli, erasure, amplitude damping noises, and nearest-neighbor ZZ-coupling noise. The common message is that our two 1D circuit models exhibit good quantum error correction performance at $O(\log n)$ depth for all these practical noise types. We also show that a better performance exists for the double-layer blocked circuit

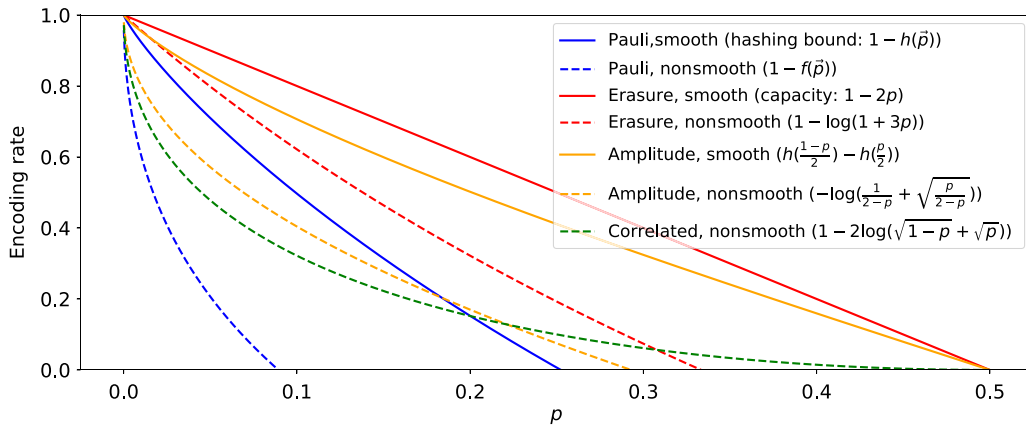


FIG. 2. The encoding rates of different bounds. For Pauli noise, the encoding rates given by the smooth and nonsmooth decoupling theorems are bounded by $1 - h(\vec{p})$ and $1 - f(\vec{p})$, respectively. They are depicted in blue with solid and dashed lines, respectively. For the strength- p depolarizing noise, the four parameters of the Pauli noise are $\vec{p} = (1 - 3p/4, p/4, p/4, p/4)$. The hashing bound $1 - h(\vec{p})$ is higher than the bound $1 - f(\vec{p})$. For erasure errors, the encoding rates given by the smooth and nonsmooth decoupling theorems are bounded by $1 - 2p$ and $1 - \log(1 + 3p)$ and depicted in red with solid and dashed lines, respectively. The bound $1 - 2p$ is the quantum channel capacity of the erasure error, and it is higher than the bound $1 - \log(1 + 3p)$. For amplitude damping noise, the encoding rates given by the smooth and nonsmooth decoupling theorems are bounded by $h((1 - p)/2) - h(p/2)$ and $-\log(1/(2 - p) + \sqrt{p/(2 - p)})$ and depicted in orange with solid and dashed lines, respectively. For strength- p nearest-neighbor correlated noise, only the nonsmooth decoupling theorem is applicable. The achievable encoding rate is $1 - 2 \log(\sqrt{1 - p} + \sqrt{p})$, depicted in green.

at $\omega(\log n)$ depth. In the following, we mainly present the results for this type of circuit.

Remarkably, we prove that the encoding schemes from the double-layer blocked circuit at $\omega(\log n)$ depth can establish a Choi error that goes to zero with an encoding rate k/n under strength- $\vec{p}$ local Pauli noise, provided that $1 - k/n - h(\vec{p}) > 0$. This error decay implies that these encoding schemes can achieve the well-known hashing bound of the channel capacity for local Pauli noise [28,35,36]. Here, $\vec{p} = (p_I, p_X, p_Y, p_Z)$ represents the four parameters of the local Pauli noise where $p_I + p_X + p_Y + p_Z = 1$. The entropy function is defined as $h(\vec{p}) = -p_I \log p_I - p_X \log p_X - p_Y \log p_Y - p_Z \log p_Z$. When depth is $O(\log n)$, the encoding rate can achieve a smaller one, $1 - f(\vec{p})$, against the same strength of Pauli noise where $f(\vec{p}) = 2 \log(\sqrt{p_I} + \sqrt{p_X} + \sqrt{p_Y} + \sqrt{p_Z})$. For strength- p local erasure error, where each qubit is i.i.d. erased with probability p , the encoding rate can also be constant and achieve the quantum channel capacity $1 - 2p$ [62]. For strength- p local amplitude damping noise, the encoding rate can reach $h((1 - p)/2) - h(p/2)$. For nearest-neighbor ZZ-coupling noise, we show that the encoding rate can achieve a constant $1 - 2 \log(\sqrt{1 - p} + \sqrt{p})$. The results for 1D brickwork circuits are the same when circuit depth is $O(\log n)$. The results are summarized in Result 1 and depicted in Fig. 2. We present the formal results in Corollaries 1, 2, 3, 4, 5, 6, 7, and Appendix C3.

Result 1 (Encoding rate of the double-layer blocked circuit $\mathfrak{C}$ and 1D brickwork circuit $\mathfrak{B}$, informal).

- (1) In the large n limit, the expected Choi error for the random codes from $\mathfrak{C}$ and $\mathfrak{B}$ vanishes with encoding rate $k/n < 1 - f(\vec{p})$ for strength- $\vec{p}$ local Pauli noise, $k/n < 1 - \log(1 + 3p)$ for strength- p local erasure errors, $k/n < -\log(1/(2 - p) + \sqrt{p/(2 - p)})$ for strength- p local amplitude damping noise, and $k/n < 1 - 2c \log(\sqrt{1 - p} + \sqrt{p})$ for strength- p correlated noise at circuit depth $O(\log n)$. Here, $c = 1$ for $\mathfrak{C}$ and $c = (1 + k/n)$ for $\mathfrak{B}$.
- (2) In the large n limit, the expected Choi error for the random codes from $\mathfrak{C}$ vanishes with encoding rate $k/n < 1 - h(\vec{p})$ for strength- $\vec{p}$ local Pauli noise, $k/n < 1 - 2p$ for strength- p local erasure errors, and $k/n < h((1 - p)/2) - h(p/2)$ for strength- p local amplitude damping noise at circuit depth $\omega(\log n)$.

We further strengthen the result by showing that these codes exhibit a Choi error that decays exponentially with circuit depth.

Result 2 (Choi error of encoding schemes from the double-layer blocked circuit $\mathfrak{C}$ and 1D brickwork circuit $\mathfrak{B}$, informal). The expected Choi error of the random codes

formed by $\mathfrak{C}_n^\epsilon$ or $\mathfrak{B}_n^\epsilon$ with depth $O(\log n/\epsilon)$ satisfies

$$\begin{aligned}\mathbb{E}_{U \sim \mathfrak{C}_n^\epsilon} \epsilon_{\text{Choi}} &= O\left((\epsilon^{1-\frac{k}{n}} n^{\frac{k}{n}})^{\frac{1}{4}}\right), \\ \mathbb{E}_{U \sim \mathfrak{B}_n^\epsilon} \epsilon_{\text{Choi}} &= O\left(\left(\frac{\epsilon}{k}\right)^{\frac{n}{4k}-\frac{1}{2}}\right),\end{aligned}\quad (6)$$

with an encoding rate $k/n < 1 - f(\vec{p})$ for strength- $\vec{p}$ local Pauli noise, $k/n < 1 - \log(1 + 3p)$ for strength- p local erasure errors, $k/n < -\log\left(1/(2-p) + \sqrt{p/(2-p)}\right)$ for strength- p local amplitude damping noise, and $k/n < 1 - 2c \log(\sqrt{1-p} + \sqrt{p})$ for strength- p correlated noise.

The formal version of this result is given in Corollaries 1, 3, 5, 7, and Appendix C 3. A direct consequence of Result 2 is that 1D encoding schemes with depth $O(\log n)$ establish a Choi error $\epsilon_{\text{Choi}} = 1/\text{poly}(n)$. This small Choi error is particularly valuable for quantum error correction over extended periods. When the encoding rate is above the value mentioned in Result 2, the scaling of the Choi error will be $O((\log n/\log(n/\epsilon))^{1/4})$, as discussed in Appendix B 4.

Furthermore, we establish the lower bound on the circuit depth required to achieve high AQEC performance under the local depolarizing channel, a type of local Pauli noise. This result provides a lower bound for encoding circuit depth against local Pauli noise and can be easily extended to other local errors.

Result 3 (Lower bound for circuit depth, informal). For a circuit U that encodes k logical qubits with depth d and has a Choi error ϵ against a constant-strength local depolarizing channel, then:

- (1) If U is a D-dimensional circuit, then $d = \Omega((\log(1/\epsilon) + \log k)^{1/D})$.
- (2) If U is an all-to-all circuit, then $d = \Omega(\log[\log(1/\epsilon) + \log k])$.

This result matches the upper bound in Results 1 and 2, confirming the optimality of our encoding scheme. We present the formal version of this result in Theorem 4.

To prove the AQEC performance established in Results 1 and 2, we employ the complementary channel approach [27], which evaluates the Choi error by characterizing how much information leaks from the system to the environment after encoding and noise. The issue becomes how decoupled the environment and the other copy of the system are after inputting an entangled state across two copies of systems, with one undergoing encoding and the complementary channel from system to environment. This decoupling task is illustrated in Fig. 3. To address this issue, we prove decoupling theorems tailored for $\mathfrak{C}$ and $\mathfrak{B}$. Compared to previous decoupling results [37,63,64],

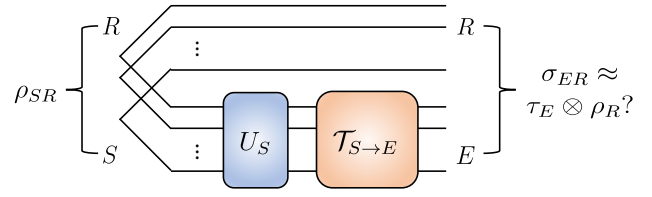


FIG. 3. The diagram of the decoupling task. The process begins with an entangled state, ρ_{SR} , where R is a reference system identical to S . The system S undergoes an encoding unitary operation, U_S , followed by a channel $\mathcal{T}_{S \rightarrow E}$ that maps S to the environment E . The resulting output state is σ_{ER} . The decoupling task examines whether σ_{ER} is sufficiently close to a tensor-product state $\tau_E \otimes \rho_R$ for certain choices of U_S and $\mathcal{T}_{S \rightarrow E}$. The role of U_S is to spread information within S , making it difficult for the environment E to extract information about S and R through $\mathcal{T}_{S \rightarrow E}$. When E has no information about SR , it becomes decoupled from R .

the new decoupling theorem considers initial states that have a tensor-product structure. This relaxation enables the establishment of the decoupling for 1D low-depth circuits instead of high-dimensional or long-depth circuits. Meanwhile, this relaxation is sufficient for evaluating the Choi error. We present the detailed results in Theorems 1, 2, and 3.

Result 4 (Decoupling theorems for the double-layer blocked circuit $\mathfrak{C}$ and 1D brickwork circuit $\mathfrak{B}$, informal). Given three quantum systems S , R , and E , denote the reduced density matrix of ρ_{SR} on system R as ρ_R and the Choi-Jamiolkowski representation of $\mathcal{T}_{S \rightarrow E}$ as τ_{SE} with reduced density matrix on system E as τ_E . Suppose U_S is drawn from $\mathfrak{C}_n^\epsilon$ or $\mathfrak{B}_n^\epsilon$, and ρ_{SR} has a tensor-product structure over different regions. Then in the large n limit,

$$\mathbb{E}_{U_S \sim \mathfrak{C}_n^\epsilon} \|\mathcal{T}_{S \rightarrow E}(U_S \rho_{SR} U_S^\dagger) - \tau_E \otimes \rho_R\|_1 = O(\epsilon^{c_1} n^{c_2}), \quad (7)$$

$$\mathbb{E}_{U_S \sim \mathfrak{B}_n^\epsilon} \|\mathcal{T}_{S \rightarrow E}(U_S \rho_{SR} U_S^\dagger) - \tau_E \otimes \rho_R\|_1 = O(\epsilon^{c_3} n^{c_4}), \quad (8)$$

if $H_2(S|E)_{\tau_{SE}} + H_2(S|R)_{\rho_{SR}} > 0$, and $\epsilon = O(1/\text{poly}(n))$. Meanwhile,

$$\mathbb{E}_{U_S \sim \mathfrak{C}_n^\epsilon} \|\mathcal{T}_{S \rightarrow E}(U_S \rho_{SR} U_S^\dagger) - \tau_E \otimes \rho_R\|_1 = O(\delta + \epsilon^{c'_1} n^{c'_2}) \quad (9)$$

if $H_2^\delta(S|E)_{\tau_{SE}} + H_2(S|R)_{\rho_{SR}} > 0$, and $\epsilon = o(1/\text{poly}(n))$. The coefficients c_1, c_2, c_3, c_4, c'_1 , and c'_2 are constant. Here, H_2 and H_2^δ are the conditional collision entropy and the smoothed version, respectively, which we will introduce in Sec. III A.

In the above theorem, the term $H_2(S|R)_{\rho_{SR}}$ represents the initial correlation between system S and reference R and equals $-k$ for k EPR pairs. The term $H_2(S|E)_{\tau_{SE}}$ evaluates

how the quantum channel from S to E preserves the correlation or the amount of information leakage from S to E . If there is no information leakage, then environment E knows nothing about S and R , $H_2(S|E)_{\tau_{SE}}$ achieves the maximum, and E and R will be decoupled. Normally, a higher noise error rate implies a higher information leakage. Thus, $H_2(S|E)_{\tau_{SE}}$ decays when noise strength increases. It gives the lower encoding rate bounds like $n(1 - f(\bar{p}))$ for Pauli noise. The smoothed entropy $H_2^\delta(S|E)_{\tau_{SE}}$ gives the higher encoding rate bounds. The nonsmooth version also implies the exponential decay of Choi error with circuit depth, as stated in Result 2. Thus, $O(\log n)$ -depth circuits exhibit negligibly small Choi error. On the other hand, the smooth version shows the Choi error approaches 0 for an $\omega(\log n)$ -depth circuit with a larger encoding rate.

The new decoupling theorems are expected to have important implications in quantum many-body physics and quantum information tasks, as outlined in the discussion section. We also generalize these results to bound the error for an arbitrary input state where ρ_{SR} does not have a tensor-product structure (Proposition 2). Additionally, we propose a randomized encoding scheme that exhibits good recovery fidelity for arbitrary input states (Proposition 1).

III. PRELIMINARIES

In this section, we introduce the preliminaries of our work. First, we present the basic notations and quantities used throughout this paper. Then, we introduce fundamental concepts and results in approximate quantum error correction. Next, we discuss the related concept of random unitaries and quantum design.

A. Basic notations and quantities

Let us start with the basic notations used in this work. We denote the Hilbert space as $\mathcal{H}$, the set of the states on $\mathcal{H}$ as $\mathcal{D}(\mathcal{H})$, and the set of the general matrices on $\mathcal{H}$ as $\mathcal{M}(\mathcal{H})$. For a single-qudit system with local dimension q , we denote the Hilbert space as $\mathcal{H}_q$ where $\dim \mathcal{H}_q = q$. The basis of $\mathcal{H}_q$ is $\{|1\rangle, |2\rangle, \dots, |q\rangle\}$. The Hilbert space of the n -qudit system is $\mathcal{H}_q^{\otimes n}$. The Hilbert space $\mathcal{H}_{AB}$ of the combined system AB of subsystems A and B is $\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B$. The dimension of $\mathcal{H}_A$ is denoted as $|A|$. For a quantum state $\rho_{AB} \in \mathcal{D}(\mathcal{H}_{AB})$, its reduced density matrix on system B is denoted as $\rho_B = \text{tr}_A(\rho_{AB})$, where tr_A is the partial trace over subsystem A . The unitary operation U on system A is denoted as U_A with the subscript to represent the support of U . A completely positive and trace-preserving (CPTP) map $\mathcal{T}$ that maps a state from $\mathcal{H}_A$ to $\mathcal{H}_B$ is denoted as $\mathcal{T}_{A \rightarrow B}$. The Choi-Jamiołkowski representation of a CPTP map $\mathcal{T}_{A \rightarrow B}$ is a quantum state τ_{AB} , which is defined as

$$\tau_{AB} = \mathcal{T}_{A' \rightarrow B}(|\hat{\phi}\rangle_{AA'}\langle\hat{\phi}|_{AA'}), \quad (10)$$

where A' is a copy of system A , and $|\hat{\phi}\rangle_{AA'}$ is the maximally entangled state on system AA' with

$$|\hat{\phi}\rangle_{AA'} = \frac{1}{\sqrt{|A|}} \sum_{i=0}^{|A|-1} |i\rangle_A |i\rangle_{A'}. \quad (11)$$

We use several measures to quantify the distance between two states. The fidelity is defined as

$$F(\rho, \sigma) = \|\sqrt{\rho}\sqrt{\sigma}\|_1 = \text{tr}\sqrt{\sqrt{\rho}\sigma\sqrt{\rho}}. \quad (12)$$

For pure state $|\psi\rangle$ and mixed state σ , the fidelity is given by

$$F(|\psi\rangle\langle\psi|, \sigma) = \sqrt{\langle\psi|\sigma|\psi\rangle}. \quad (13)$$

From the fidelity, we can define the purified distance,

$$P(\rho, \sigma) = \sqrt{1 - F(\rho, \sigma)^2}, \quad (14)$$

which satisfies [65]

$$\frac{1}{2}\|\rho - \sigma\|_1 \leq P(\rho, \sigma) \leq \sqrt{\|\rho - \sigma\|_1}. \quad (15)$$

The trace distance between ρ and σ is $\frac{1}{2}\|\rho - \sigma\|_1$. The definitions of fidelity F and trace distance can also be generalized to arbitrary positive matrices beyond states by substituting ρ and σ with general matrices. For two quantum channels $\mathcal{T}_{A \rightarrow B}$ and $\mathcal{T}'_{A \rightarrow B}$, we normally use diamond distance to characterize their distance, defined as

$$\|\mathcal{T}_{A \rightarrow B} - \mathcal{T}'_{A \rightarrow B}\|_\diamond = \max_{\rho_{AR}} \|\mathcal{T}_{A \rightarrow B}(\rho_{AR}) - \mathcal{T}'_{A \rightarrow B}(\rho_{AR})\|_1, \quad (16)$$

where R is a reference system without dimension restriction. Nonetheless, it can be shown that considering a reference system R with the same dimension as A is sufficient for maximization. The definition of the diamond distance can also be generalized for two completely positive maps.

In this work, we will massively use the conditional entropy. The conditional collision entropy of A given B is defined as

$$H_2(A|B)_\rho = -\log \text{tr}[(\mathbb{I}_A \otimes \rho_B)^{-1/4} \rho_{AB} (\mathbb{I}_A \otimes \rho_B)^{-1/4}]^2. \quad (17)$$

Its smoothed version is

$$H_2^\delta(A|B)_\rho = \sup_{\hat{\rho}_{AB} \in \mathcal{B}^\delta(\rho_{AB})} H_2(A|B)_{\hat{\rho}}, \quad (18)$$

where $\mathcal{B}^\delta(\rho) = \{\rho', P(\rho', \rho) \leq \delta\}$ is the set of all δ -close states of ρ . For more details about quantum entropies refer to Appendix A.

Below, we introduce a mathematical lemma used in this work, which turns the evaluation of the 1-norm of a matrix into a second-moment quantity.

Lemma 1 (Lemma 3.7 in Ref. [63]). Given a Hilbert space $\mathcal{H}$, let $M \in \mathcal{M}(\mathcal{H})$ and $\sigma \in \mathcal{D}(\mathcal{H})$. We have that

$$\|M\|_1 \leq \sqrt{\text{tr}(\sigma^{-1/4} M \sigma^{-1/2} M^\dagger \sigma^{-1/4})}. \quad (19)$$

If M is Hermitian, this simplifies to

$$\|M\|_1 \leq \sqrt{\text{tr}((\sigma^{-1/4} M \sigma^{-1/4})^2)}. \quad (20)$$

B. Approximate quantum error correction

In this subsection, we introduce basic concepts in approximate quantum error correction. To protect logical information, the logical system L , consisting of k qubits, is encoded into the n -qubit physical system S via an encoding channel $\mathcal{E}_{L \rightarrow S}$. The logical system may also be entangled with a k -qubit reference system R . The code rate is given by k/n . After encoding, the physical system S undergoes a noise channel $\mathcal{N}$. Finally, a decoding operation $\mathcal{D}_{S \rightarrow L}$ is applied to recover the information stored in the logical system. For an initial state ρ_{LR} , the recovered state is given by

$$[(\mathcal{D} \circ \mathcal{N} \circ \mathcal{E})_L \otimes I_R](\rho_{LR}). \quad (21)$$

For a unitary encoding, $\mathcal{E}_{L \rightarrow S}(\rho) = U_S \rho_L \otimes |0\rangle\langle 0|_{S \setminus L} U_S^\dagger$, as demonstrated by Fig. 4.

One can quantify the performance of quantum error correction by the fidelity between the initial and final states. In

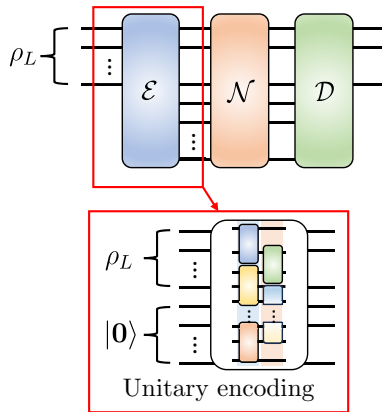


FIG. 4. Diagram of the procedure of quantum error correction. The initial logical information is stored in a quantum state ρ_L . Before undergoing noise $\mathcal{N}$, ρ_L is first encoded by $\mathcal{E}$ with $\mathcal{E}(\rho) = U \rho \otimes |0\rangle\langle 0| U^\dagger$ for unitary encoding operation U . Finally, one applies decoding operation $\mathcal{D}$ to retrieve logical information. More generally, ρ_L may be entangled with a reference system R .

this work, we focus on the entanglement fidelity, or Choi fidelity, which is defined as

$$F_{\text{Choi}} = \max_{\mathcal{D}} F\left(\hat{\phi}_{LR}, [(\mathcal{D} \circ \mathcal{N} \circ \mathcal{E})_L \otimes I_R](\hat{\phi}_{LR})\right), \quad (22)$$

where $\hat{\phi}_{LR}$ is the maximally entangled state between the logical system and the reference system. The Choi error is defined as

$$\begin{aligned} \epsilon_{\text{Choi}} &= \sqrt{1 - F_{\text{Choi}}^2} \\ &= \min_{\mathcal{D}} P\left(\hat{\phi}_{LR}, [(\mathcal{D} \circ \mathcal{N} \circ \mathcal{E})_L \otimes I_R](\hat{\phi}_{LR})\right). \end{aligned} \quad (23)$$

The Choi fidelity quantifies how well the code can preserve entanglement between the logical and reference systems. It is also closely related to the max-average state fidelity F_{ave} , which reflects the average performance of the AQEC codes over all input states,

$$F_{\text{ave}}^2 = \max_{\mathcal{D}} \int d\psi F^2(\psi, (\mathcal{D} \circ \mathcal{N} \circ \mathcal{E})_L(\psi)). \quad (24)$$

Here, $d\psi$ denotes the Haar-random distribution of pure states over $\mathcal{H}_L$. The average fidelity F_{ave} and Choi fidelity satisfy the relation [66,67]

$$F_{\text{ave}}^2 = \frac{d F_{\text{Choi}}^2 + 1}{d + 1}, \quad (25)$$

where $d = |L| = 2^k$ is the dimension of the logical system.

Directly evaluating Eq. (23) is challenging. The Choi error is often analyzed using the complementary channel formalism [22,27]. In essence, the complementary channel formalism quantifies recovery performance by examining the information leakage contained in the environment. Specifically, according to Stinespring dilation, the noise channel can be modeled by interacting the physical system S with the environment E through a unitary U , followed by tracing out the environment,

$$\mathcal{N}_S(\rho_S) = \text{tr}_E(U_{SE} \rho_S \otimes |0\rangle\langle 0|_E U_{SE}^\dagger). \quad (26)$$

If we trace out the system instead of the environment, we obtain the complementary channel of $\mathcal{N}_S$,

$$\hat{\mathcal{N}}_{S \rightarrow E} = \text{tr}_S(U_{SE} \rho_S \otimes |0\rangle\langle 0|_E U_{SE}^\dagger). \quad (27)$$

For a noise channel with Kraus operators $\{K_i\}$,

$$\mathcal{N}_S(\rho) = \sum_i K_i \rho K_i^\dagger, \quad (28)$$

the complementary channel $\hat{\mathcal{N}}_{S \rightarrow E}$ can be written as [27, 68]

$$\hat{\mathcal{N}}_{S \rightarrow E} = \sum_{ij} \text{tr}(K_i \rho K_j^\dagger) |i\rangle\langle j|. \quad (29)$$

Note that the complementary channel of $\mathcal{N} \circ \mathcal{E}$ can be obtained by

$$\widehat{\mathcal{N} \circ \mathcal{E}}_{L \rightarrow E} = \hat{\mathcal{N}}_{S \rightarrow E} \circ \mathcal{E}_{L \rightarrow S}. \quad (30)$$

By applying the complementary channel $\hat{\mathcal{N}}_{S \rightarrow E}$, we obtain the state in the environment, which captures the information leaked from the system to the environment. The Choi error can then be expressed as [22, 27]

$$\epsilon_{\text{Choi}} = \min_{\zeta} P \left(\left(\widehat{\mathcal{N} \circ \mathcal{E}}_{L \rightarrow E} \otimes I_R \right) \left(|\hat{\phi}\rangle\langle\hat{\phi}|_{\text{LR}} \right), \right. \\ \left. \left(\mathcal{T}_{L \rightarrow E}^{\zeta} \otimes I_R \right) \left(|\hat{\phi}\rangle\langle\hat{\phi}|_{\text{LR}} \right) \right), \quad (31)$$

where $\mathcal{T}^{\zeta}$ is a quantum channel that maps any input state into a fixed state ζ , satisfying

$$(\mathcal{T}_{L \rightarrow E}^{\zeta} \otimes I_R)(\rho_{\text{LR}}) = \zeta_E \otimes \text{tr}_L(\rho_{\text{LR}}). \quad (32)$$

Equation (31) indicates that if the state in the environment is close to a fixed state in the tensor product with the reference system, the Choi error will be small, implying that entanglement is well preserved.

Based on the Choi error, one can define the achievable rate and the channel capacity of local noise channels [26]. Note that the Choi error is the function of the logical qubit number k , the encoding map $\mathcal{E}$, and the noise channel $\mathcal{N}$. We can denote it as $\epsilon_{\text{Choi}} = \epsilon_{\text{Choi}}(k, \mathcal{E}, \mathcal{N})$. For a single-qubit noise channel $\mathcal{N}_1$, if there exists a sequence of logical qubit numbers $\{k_n\}$ and encoding maps $\{\mathcal{E}_n\}$ such that

$$\lim_{n \rightarrow \infty} \epsilon_{\text{Choi}}(k_n, \mathcal{E}_n, \mathcal{N}_1^{\otimes n}) = 0, \quad (33)$$

then we call

$$R = \limsup_{n \rightarrow \infty} \frac{k_n}{n} \quad (34)$$

as an achievable rate of $\mathcal{N}_1$. The quantum channel capacity of $\mathcal{N}_1$, denoted as $Q(\mathcal{N}_1)$, is the supremum over all achievable rates.

C. Noise model

In this work, we examine a wide variety of important noise models: Pauli noises [69], erasure errors [70], amplitude damping noises [71], and correlated noises. First, we

introduce Pauli noises. For a single qubit, the Pauli noise is defined as

$$\mathcal{N}_p(\rho) = p_I \rho + p_X X \rho X + p_Y Y \rho Y + p_Z Z \rho Z. \quad (35)$$

We denote $\vec{p} = (p_I, p_X, p_Y, p_Z)$ and refer to this noise as the strength- $\vec{p}$ Pauli noise. When the probabilities satisfy $p_X = p_Y = p_Z = p/3$, the Pauli noise simplifies to the depolarizing noise,

$$\mathcal{N}_d(\rho) = (1 - p)\rho + p \frac{\mathbb{I}}{2}. \quad (36)$$

which we refer to as the strength- p depolarizing noise. For an n -qubit state, we consider the i.i.d. local Pauli noise on each qubit. The overall noise channel is then the tensor product of n independent Pauli noises, expressed as $\mathcal{N} = \mathcal{N}_p^{\otimes n}$.

Another noise model of interest is the erasure error, which completely traces out a subsystem T while simultaneously recording the location of the erased subsystem in a separate register C . The erasure error model is important due to its strong connection with code distance [65, 70]. For an erasure noise act on a fixed subsystem T , the register C can be ignored and the noise is expressed as

$$\mathcal{N}(\rho) = \text{tr}_T(\rho). \quad (37)$$

The complementary channel of the erasure error is

$$\hat{\mathcal{N}}(\rho) = \text{tr}_{\bar{T}}(\rho), \quad (38)$$

where $\bar{T} = S \setminus T$ denotes the complement of subsystem T . This implies that the environment obtains the state in the traced-out subsystem. For the random t -erasure error, the locations of the erased subsystem are randomly selected among all possible t -qubit subsystems, and the noise is expressed as

$$\mathcal{N}(\rho) = \frac{1}{\binom{n}{t}} \sum_{|T|=t} |T\rangle\langle T|_C \otimes \text{tr}_T(\rho). \quad (39)$$

Here, each T represents a distinct t -qubit subsystem being erased, and $|T\rangle_C$ records the specific location of the erasure in the register C . The complementary channel of this random t -erasure noise is [72]

$$\hat{\mathcal{N}}(\rho) = \frac{1}{\binom{n}{t}} \sum_{|T|=t} |T\rangle\langle T|_C \otimes \text{tr}_{\bar{T}}(\rho), \quad (40)$$

where $|T\rangle_C$ continues to label the location of the erasure errors.

Instead of considering erasure errors with a fixed number of erasures, one can also study i.i.d. erasure errors.

Particularly, we consider that each qubit experiences an erasure error i.i.d. with probability p . For a single-qubit state ρ , the noise is described by

$$\mathcal{N}_e(\rho) = (1-p)\rho + p|2\rangle\langle 2|, \quad (41)$$

where state $|2\rangle$ is orthogonal to the space supporting ρ and represents the erasure. For an n -qubit system, the total erasure noise is modeled as $\mathcal{N}_e^{\otimes n}$.

The local amplitude damping noise is also an i.i.d. noise model. We first define the strength- p amplitude damping noise on a single qubit,

$$\mathcal{N}_a(\rho) = K_0\rho K_0^\dagger + K_1\rho K_1^\dagger, \quad (42)$$

with

$$K_0 = \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{1-p} \end{pmatrix}, \quad K_1 = \begin{pmatrix} 0 & \sqrt{p} \\ 0 & 0 \end{pmatrix}. \quad (43)$$

For an n -qubit system, the total amplitude damping noise is $\mathcal{N}_a^{\otimes n}$.

The quantum channel capacity of these noise models is of fundamental importance in quantum information theory. For Pauli noise, a well-known lower bound on the quantum channel capacity is the hashing bound,

$$Q(\mathcal{N}_p) \geq \max\{0, 1 - h(\vec{p})\}, \quad (44)$$

where

$$h(\vec{p}) = -p_I \log p_I - p_X \log p_X - p_Y \log p_Y - p_Z \log p_Z \quad (45)$$

is the Shannon entropy of the probability vector $\vec{p}$. Moreover, it has been shown that the inequality in Eq. (44) is strict in some regimes [28,35,36]. Existing upper and lower bounds for the quantum capacity of Pauli noise do not match, and the exact value of the quantum channel capacity for Pauli noise remains unknown, even for the depolarizing noise. The hashing bound is the most famous lower bound for a generic Pauli noise, which is achievable for random stabilizer codes [28].

Finding encoding schemes simpler than the random stabilizer construction to achieve the hashing bound deserves further exploration.

For erasure errors, the quantum channel capacity is much simpler and can be computed exactly as [62]

$$Q(\mathcal{N}_e) = \max\{0, 1 - 2p\}. \quad (46)$$

This result matches the Singleton bound $n - k - 2t \geq 0$, which yields $k/n \leq 1 - 2t/n$. The erasure probability p in the i.i.d. case corresponds to t/n .

For amplitude damping noise, the quantum channel capacity can also be computed exactly [71]. For $p > \frac{1}{2}$, $Q(\mathcal{N}_a) = 0$. For $p \leq \frac{1}{2}$,

$$Q(\mathcal{N}_a) = \max_{0 \leq x \leq 1} \{h((1-p)x) - h(px)\}, \quad (47)$$

where h is the binary entropy function.

The above noise models are all i.i.d. single-qubit noises, which enjoy a tensor-product structure. In this work, we also study correlated noise models, which are also practically important. Particularly, we introduce strength- p 1D nearest-neighbor ZZ-coupling noise on n qubits,

$$\mathcal{N}_{zz}(\rho) = \circ_{i=1}^{n-1} ((1-p)\mathcal{I} + p\mathcal{Z}_i\mathcal{Z}_{i+1})(\rho), \quad (48)$$

where $\mathcal{I}$ is the identity channel, and $\mathcal{Z}_i$ is the Pauli Z channel on qubit i . Note that for correlated noise, currently, there is no corresponding quantum channel capacity. But there can exist a threshold for p such that the logical error rate goes to 0 for this noise model.

D. Twirling operations and approximate design

In this part, we introduce the concept of twirling and design associated with random unitary operations. The twirling operation is defined over an ensemble of quantum unitary operations, $\mathfrak{S}$. We consider the following t -th order twirling operation over $\mathfrak{S}$,

$$\Phi_{\mathfrak{S}}^t(O) = \mathbb{E}_{U \sim \mathfrak{S}} U^{\otimes t} O U^{\dagger \otimes t}, \quad (49)$$

where U is drawn from the ensemble $\mathfrak{S}$; O is the matrix being twirled, and its support may overpass $U^{\otimes t}$. The expectation is taken over the random unitary operations from $\mathfrak{S}$. When the ensemble $\mathfrak{S}$ corresponds to the Haar-random unitaries $\mathfrak{U}_n$ on n qubits, the first-order twirling operation over $\mathfrak{U}_n$ can be obtained from the Schur-Weyl duality [73],

$$\mathbb{E}_{U \sim \mathfrak{U}_n} U O U^\dagger = \text{tr}_U(O) \frac{\mathbb{I}_q}{q}, \quad (50)$$

where tr_U denotes the trace over the subsystem being twirled, $q = \dim U$ is the dimension of Hilbert space that the unitary acts on, and $\mathbb{I}_q$ is the identity operator with dimension q . The second-order twirling operation over $\mathfrak{U}_n$ is

$$\begin{aligned} \mathbb{E}_{U \sim \mathfrak{U}_n} U^{\otimes 2} O U^{\dagger \otimes 2} &= B_1 \text{tr}_{U^{\otimes 2}}(O B_1) + B_2 \text{tr}_{U^{\otimes 2}}(O B_2) \\ &= \text{tr}_{U^{\otimes 2}} \left(O \frac{\mathbb{I}_{q^2} - q^{-1} F}{q^2 - 1} \right) \mathbb{I}_{q^2} + \text{tr}_{U^{\otimes 2}} \left(O \frac{F - q^{-1} \mathbb{I}_{q^2}}{q^2 - 1} \right) F, \end{aligned} \quad (51)$$

where

$$B_1 = \frac{\mathbb{I}_{q^2}}{q}, \quad B_2 = \frac{F - B_1}{\sqrt{q^2 - 1}}, \quad (52)$$

with $\text{tr}_{U^{\otimes 2}}$ denoting the trace over the subsystem of being twirled, $\mathbb{I}_{q^2}$ is the identity operator on the q^2 -dimension subsystem, and F denotes the SWAP operator on the q^2 -dimension subsystem.

Though the Haar-random unitary ensemble possesses good twirling properties, it requires exponential resources to implement in practice. An approximate version of the Haar random unitary ensemble is usually used for practical applications. An ensemble $\mathfrak{S}$ is called a unitary t -design if $\Phi'_{\mathfrak{S}} = \Phi'_{\mathfrak{U}_n}$ and an ε -approximate unitary t -design [31,74–76] if

$$(1 - \varepsilon)\Phi'_{\mathfrak{U}_n} \leq \Phi'_{\mathfrak{S}} \leq (1 + \varepsilon)\Phi'_{\mathfrak{U}_n}, \quad (53)$$

where $\Phi_1 \leq \Phi_2$ means that $\Phi_2 - \Phi_1$ is a completely positive map [31]. The term ε is referred to as the relative error. An approximate design can also be defined in an additive error version,

$$\|\Phi'_{\mathfrak{S}} - \Phi'_{\mathfrak{U}_n}\|_{\diamond} \leq \varepsilon. \quad (54)$$

It has been shown that an approximate design with relative error ε is an approximate design with additive error 2ε [33]. Hereafter, unless stated otherwise, the error of an approximate design refers to relative error. An exact unitary t -design is an approximate unitary t -design with $\varepsilon = 0$. It was shown that the n -qubit Clifford group is an exact unitary 3-design [77,78], and when the local dimension is odd prime instead of 2, the Clifford group is an exact unitary 2-design [74–76,79,80]. As a remark, $\mathfrak{C}$ and $\mathfrak{B}$ are exact unitary 1-designs since they contain the local Pauli group, which is already a unitary 1-design.

IV. LOG-DEPTH CIRCUIT TO ACHIEVE DECOUPLING AND HIGH AQEC PERFORMANCE

In this section, we investigate the AQEC performance of the unitary operations generated by 1D log-depth circuits using the Choi error metric.

Before elaborating on the detailed results, we note that a recent surprising result shows that the 1D double-layer blocked circuit $\mathfrak{C}_n^\varepsilon$ forms a 9ε -approximate unitary 3-design and 4ε -approximate unitary 2-design [33]. However, the property of approximate design, in general, does not guarantee good AQEC performance unless ε is smaller than an exponential decay function. The decoupling theorem for approximate unitary designs [64] includes an exponential factor, as stated in Lemma A6 in Appendix A. The exponentially small ε implies a circuit depth linear in the qubit number, which is not desired. To address this issue, we develop new decoupling theorems. We use these results to show that $\mathfrak{C}_n^\varepsilon$ exhibits a logical qubit count that scales linearly with n while maintaining a small Choi error. Particularly, $\mathfrak{C}_n^\varepsilon$ can achieve the hashing bound for Pauli noises and the quantum capacity for erasure errors even if ε is not exponentially small.

We begin by reformulating the analysis of the Choi error as a decoupling problem. Next, we present the decoupling theorems for 1D double-layer blocked circuit $\mathfrak{C}$ and 1D brickwork circuit $\mathfrak{B}$ as a key technical contribution. By applying the decoupling theorem to Pauli noise, erasure errors, amplitude damping noise, and correlated noise, we establish the favorable AQEC performance. Furthermore, we prove a better AQEC performance of $\mathfrak{C}$ and $\mathfrak{B}$ than the block-encoding scheme. We also extend these results to more general error metrics beyond the Choi error. In the next section, we derive the lower bound of the circuit depth required to achieve a small Choi error in AQEC, which matches the circuit depth needed for $\mathfrak{C}$ and $\mathfrak{B}$.

A. Approximate quantum error correction and decoupling theorem

We start by analyzing the expected Choi error. With the complementary channel approach, we have that

$$\begin{aligned} \epsilon_{\text{Choi}} &= \min_{\zeta} P \left(\hat{\mathcal{N}}_{S \rightarrow E} \left(U \left| \hat{\phi} \right\rangle \left\langle \hat{\phi} \right|_{\text{LR}} \otimes |0^{n-k}\rangle \langle 0^{n-k}| U^\dagger \right), \zeta_E \otimes \frac{\mathbb{I}_R}{2^k} \right) \\ &\leq P \left(\hat{\mathcal{N}}_{S \rightarrow E} \left(U \left| \hat{\phi} \right\rangle \left\langle \hat{\phi} \right|_{\text{LR}} \otimes |0^{n-k}\rangle \langle 0^{n-k}| U^\dagger \right), \mathbb{E}_{U \sim \mathfrak{U}_n} \hat{\mathcal{N}}_{S \rightarrow E} \left(U \left| \hat{\phi} \right\rangle \left\langle \hat{\phi} \right|_{\text{LR}} \otimes |0^{n-k}\rangle \langle 0^{n-k}| U^\dagger \right) \right) \\ &\quad + \min_{\zeta} P \left(\mathbb{E}_{U \sim \mathfrak{U}_n} \hat{\mathcal{N}}_{S \rightarrow E} \left(U \left| \hat{\phi} \right\rangle \left\langle \hat{\phi} \right|_{\text{LR}} \otimes |0^{n-k}\rangle \langle 0^{n-k}| U^\dagger \right), \zeta_E \otimes \frac{\mathbb{I}_R}{2^k} \right) \\ &= P \left(\hat{\mathcal{N}}_{S \rightarrow E} \left(U \left| \hat{\phi} \right\rangle \left\langle \hat{\phi} \right|_{\text{LR}} \otimes |0^{n-k}\rangle \langle 0^{n-k}| U^\dagger \right), \hat{\mathcal{N}}_{S \rightarrow E} \left(\frac{\mathbb{I}_S}{2^n} \right) \otimes \frac{\mathbb{I}_R}{2^k} \right) + \min_{\zeta} P \left(\hat{\mathcal{N}}_{S \rightarrow E} \left(\frac{\mathbb{I}_S}{2^n} \right) \otimes \frac{\mathbb{I}_R}{2^k}, \zeta_E \otimes \frac{\mathbb{I}_R}{2^k} \right) \\ &= P \left(\hat{\mathcal{N}}_{S \rightarrow E} \left(U \left| \hat{\phi} \right\rangle \left\langle \hat{\phi} \right|_{\text{LR}} \otimes |0^{n-k}\rangle \langle 0^{n-k}| U^\dagger \right), \hat{\mathcal{N}}_{S \rightarrow E} \left(\frac{\mathbb{I}_S}{2^n} \right) \otimes \frac{\mathbb{I}_R}{2^k} \right). \end{aligned} \quad (55)$$

The first inequality comes from the triangle inequality of the purified distance. The second equality uses the property of the twirling over Haar-random unitary gates. The third equality holds by taking the state ζ as $\hat{\mathcal{N}}_{S \rightarrow E}(\mathbb{I}_S/2^n)$.

Our target is to bound $\mathbb{E}_{U \in \mathcal{C}_{\text{Choi}}}$. We utilize the relation between purified distance and trace distance in Eq. (15) to get

$$\begin{aligned} \mathbb{E}_{U \in \mathcal{C}_{\text{Choi}}} &\leq \mathbb{E}_U \sqrt{\left\| \hat{\mathcal{N}}_{S \rightarrow E} \left(U \left| \hat{\phi} \right\rangle \left\langle \hat{\phi} \right|_{\text{LR}} \otimes |0^{n-k}\rangle \langle 0^{n-k}| U^\dagger \right) - \hat{\mathcal{N}}_{S \rightarrow E} \left(\frac{\mathbb{I}_S}{2^n} \right) \otimes \frac{\mathbb{I}_R}{2^k} \right\|_1} \\ &\leq \sqrt{\mathbb{E}_U \left\| \hat{\mathcal{N}}_{S \rightarrow E} \left(U \left| \hat{\phi} \right\rangle \left\langle \hat{\phi} \right|_{\text{LR}} \otimes |0^{n-k}\rangle \langle 0^{n-k}| U^\dagger \right) - \hat{\mathcal{N}}_{S \rightarrow E} \left(\frac{\mathbb{I}_S}{2^n} \right) \otimes \frac{\mathbb{I}_R}{2^k} \right\|_1}. \end{aligned} \quad (56)$$

The remaining part is evaluating the term in the square root, which corresponds to the decoupling issue: how the system E and R decouple after being acted on by a unitary and channel $\hat{\mathcal{N}}_{S \rightarrow E}$.

We now present our decoupling theorems. For the double-layer blocked ensemble $\mathcal{C}$, we establish both nonsmooth and smooth versions, and for the 1D brickwork circuit ensemble $\mathcal{B}$, we prove the nonsmooth version. We demonstrate how the environmental and reference systems are decoupled after the physical system, encoded by unitaries from $\mathcal{C}$ or $\mathcal{B}$, interacts with a noise channel. This step lays the foundation for proving that the encoding rate can be maintained at a constant encoding rate with a small Choi error. Subsequently, we introduce the smooth decoupling theorem for $\mathcal{C}$. This refinement allows us to demonstrate that the encoding rate for the local Pauli noise channel can reach the hashing bound, and the encoding rate for the local erasure error can reach the channel capacity.

We first present the nonsmooth decoupling theorem for the double-layer blocked circuit $\mathcal{C}$ below.

Theorem 1 (Nonsmooth decoupling theorem for 1D double-layer blocked circuits). Given three quantum systems $S = S_1 S_2 \cdots S_{2N}$, $R = R_1 R_2 \cdots R_{2N}$, and $E = E_1 E_2 \cdots E_{2N}$, we denote S_i as the subsystem of region i inside S , and the same for R_i and E_i . Without loss of generality, we assume that S is composed of n qubits, and each S_i has an equal size $\xi = n/2N$ qubits. Let quantum state ρ_{SR} be $\rho_{\text{SR}} = \bigotimes_{i=1}^{2N} \rho_{S_i R_i}$ with a tensor-product structure and $\mathcal{T}_{S \rightarrow E}$ be a completely positive map. Denote the reduced density matrix of ρ_{SR} on system R as ρ_R and the Choi-Jamiołkowski representation of $\mathcal{T}_{S \rightarrow E}$ as τ_{SE} with reduced density matrix on system E as τ_E . Suppose U_S is drawn from $\mathcal{C}_n^\varepsilon$ as shown in Fig. 1(b). We obtain that

$$\mathbb{E}_{U_S \sim \mathcal{C}_n^\varepsilon} \|\mathcal{T}_{S \rightarrow E}(U_S \rho_{\text{SR}} U_S^\dagger) - \tau_E \otimes \rho_R\|_1 \leq \sqrt{2^{-H_2(S|E)\tau_{\text{SE}} - H_2(S|R)\rho_{\text{SR}}} + c \max_{A \subseteq [2N]} 2^{-H_2(A|R)\rho_{\text{AR}} - H_2(A|E)\tau_{\text{AE}}}}, \quad (57)$$

where

$$c = 2 \left(\left(1 + \frac{2^{\xi+1}}{2^{2\xi} + 1} \max_i \left(\max \left(2^{-H_2(S_i|R_i)\rho_{S_i R_i}}, 2^{H_2(S_i|R_i)\rho_{S_i R_i}} \right) \right) \right)^{N-1} - 1 \right). \quad (58)$$

Furthermore, if $\mathcal{T}_{S \rightarrow E}$ has a tensor-product structure, $\mathcal{T}_{S \rightarrow E} = \bigotimes_{i=1}^{2N} \mathcal{T}_{S_i \rightarrow E_i}$, we obtain that

$$\begin{aligned} &\mathbb{E}_{U_S \sim \mathcal{C}_n^\varepsilon} \|\mathcal{T}_{S \rightarrow E}(U_S \rho_{\text{SR}} U_S^\dagger) - \tau_E \otimes \rho_R\|_1 \\ &\leq \sqrt{2^{-\sum_{i=1}^{2N} H_2(S_i|E_i)\tau_{S_i E_i} - \sum_{i=1}^{2N} H_2(S_i|R_i)\rho_{S_i R_i}} + c \prod_{i=1}^{2N} \max(1, 2^{-H_2(S_i|E_i)\tau_{S_i E_i} - H_2(S_i|R_i)\rho_{S_i R_i}})}. \end{aligned} \quad (59)$$

Proof sketch. The proof of Theorem 1 begins with Lemma 1, where the inequality $\|M\|_1 \leq \sqrt{\text{tr}((\sigma^{-1/4} M \sigma^{-1/4})^2)}$ is used to convert the 1-norm

distance into the square of a matrix. By expanding this square, we apply both the first- and second-order twirling operations conducted by $\mathcal{C}$. The analysis of the first-order twirling is straightforward, as $\mathcal{C}$ is a unitary 1-design.

However, the second-order twirling requires evaluating $\mathbb{E}_{U^{\text{tr}}} \left[\left(\tilde{\mathcal{T}}_{S \rightarrow E} \left(U_S \tilde{\rho}_{\text{SR}} U_S^\dagger \right) \right)^2 \right]$ and bounding the difference between the expectation over $\mathfrak{C}$ and the Haar-random unitary group $\mathfrak{U}_n$. The property of $\mathfrak{C}$ being an approximate 2-design does not guarantee that this term is small unless ε is exponentially small. To address this, we refine the evaluation by leveraging the double-layer blocked structure of $\mathfrak{C}$. The additional condition that the state ρ_{SR} exhibits a tensor-product structure simplifies the calculation, and this condition holds for EPR input states. The full proof can be found in Appendix B 1. ■

The nonsmooth decoupling theorem for the 1D brickwork circuits $\mathfrak{B}_n^\varepsilon$ is similar to that for $\mathfrak{C}_n^\varepsilon$, but with a different decoupling error. The results are shown below.

$$\mathbb{E}_{U_S \sim \mathfrak{B}_n^\varepsilon} \|\mathcal{T}_{S \rightarrow E}(U_S \rho_{\text{SR}} U_S^\dagger) - \tau_E \otimes \rho_R\|_1 \leq \sqrt{2^{-H_2(S|E)\tau_{\text{SE}} - H_2(S|R)\rho_{\text{SR}}} + \left(\frac{\varepsilon}{n}\right)^{\log \frac{1}{2\eta\rho_m}} \max_{A \subseteq [n]} 2^{-H_2(A|R)\rho_{\text{AR}} - H_2(A|E)\tau_{\text{AE}}}}, \quad (60)$$

where

$$\rho_m = \max_{i \in [n]} (2^{H_2(S_i|R_i)\rho_{S_i R_i}}, 2^{-H_2(S_i|R_i)\rho_{S_i R_i}}). \quad (61)$$

Proof sketch. The proof of Theorem 2 is similar to that of Theorem 1. The key proof is to evaluate the second-order twirling $\mathbb{E}_{U^{\text{tr}}} \left[\left(\tilde{\mathcal{T}}_{S \rightarrow E} \left(U_S \tilde{\rho}_{\text{SR}} U_S^\dagger \right) \right)^2 \right]$. To address this for $\mathfrak{B}$, we utilize the domain-wall technique [32], which enables us to evaluate the second-moment twirling induced by 1D brickwork circuits. The full proof can be found in Appendix C. ■

With the nonsmooth decoupling theorem, one can show that the Choi error can be polynomially small when ε is polynomially small for Pauli noise, erasure errors, amplitude damping noise, and correlated noise, as demonstrated later. In this case, the depths of the double-layer blocked circuit $\mathfrak{C}_n^\varepsilon$ and 1D brickwork circuit $\mathfrak{B}_n^\varepsilon$ are still $O(\log n)$.

For the double-layer blocked ensemble $\mathfrak{C}$, we can further introduce a smooth decoupling theorem. Particularly,

Theorem 2 (Decoupling with 1D brickwork circuits). Given three quantum systems, $S = S_1 S_2 \cdots S_n$, $R = R_1 R_2 \cdots R_n$, and $E = E_1 E_2 \cdots E_n$, we denote S_i as the subsystem of region i inside S , and the same for R_i and E_i . Here, S_i is a qudit with local dimension q . Let quantum state ρ_{SR} be $\rho_{\text{SR}} = \bigotimes_{i=1}^n \rho_{S_i R_i}$ with a tensor-product structure and $\mathcal{T}_{S \rightarrow E}$ be a completely positive map. Denote the reduced density matrix of ρ_{SR} on system R as ρ_R and the Choi-Jamiołkowski representation of $\mathcal{T}_{S \rightarrow E}$ as τ_{SE} with reduced density matrix on system E as τ_E . Set $\eta = q/(q^2 + 1)$. Suppose U_S is drawn from $\mathfrak{B}_n^\varepsilon$ with depth $d = \log(n/\varepsilon) + \log n / \log(12\eta) + \log(e-1)/\log(1/2\eta) + 1$. We obtain that

the smooth decoupling theorem replaces the conditional collision entropy H_2 with the smoothed conditional collision entropy H_2^δ . The proof follows a standard smoothing technique, with a full derivation provided in Appendix B 2. This smooth version allows us to demonstrate a higher achievable encoding rate for $\mathfrak{C}$ when the noise channel is local. The collision entropy $H_2^\delta(\rho^{\otimes \xi})$ can approximate the tighter von Neumann entropy $H(\rho^{\otimes \xi})$ in the large n limit. In contrast, this approach cannot increase the achievable encoding rate for 1D brickwork circuit $\mathfrak{B}$, since the upper bound contains the term $2^{H_2(S_i R_i)\rho_{S_i R_i}}$ evaluated on a single qudit. In the following, we mainly present the results of $\mathfrak{C}$.

Theorem 3 (Smooth decoupling theorem for 1D double-layer blocked circuits). With the same setting in Theorem 1 and the tensor-product structure of $\mathcal{T}_{S \rightarrow E}$, for any $0 < \delta < 1$, we have that

$$\mathbb{E}_{U \sim \mathfrak{C}_n^\varepsilon} \|\mathcal{T}_{S \rightarrow E}(U_S \rho_{\text{SR}} U_S^\dagger) - \tau_E \otimes \rho_R\|_1 \leq 16N\delta + \sqrt{2^{-\sum_{i=1}^{2N} H_2^\delta(S_i|E_i)\tau_{S_i E_i} - \sum_{i=1}^{2N} H_2^\delta(S_i|R_i)\rho_{S_i R_i}} + c' \prod_{i=1}^{2N} \max(1, 2^{-H_2^\delta(S_i|E_i)\tau_{S_i E_i} - H_2^\delta(S_i|R_i)\rho_{S_i R_i}})}, \quad (62)$$

where

$$c' = 2 \left(\left(1 + \frac{2^{\xi+1}}{2^{2\xi} + 1} \max_i (\max(2^{-H_2^\delta(S_i|R_i)\rho_{S_i R_i}}, 2^{H_2^\delta(S_i|R_i)\rho_{S_i R_i}})) \right)^{N-1} - 1 \right). \quad (63)$$

If we only smooth $H_2(S_i|E_i)$, we will obtain

$$\begin{aligned} \mathbb{E}_{U \sim \mathfrak{C}_n^\varepsilon} \|\mathcal{T}_{S \rightarrow E}(U_S \rho_{SR} U_S^\dagger) - \tau_E \otimes \rho_R\|_1 &\leq 8N\delta \\ &+ \sqrt{2^{-\sum_{i=1}^{2N} H_2^\delta(S_i|E_i)\tau_{S_i E_i} - \sum_{i=1}^{2N} H_2(S_i|R_i)\rho_{S_i R_i}} + c \prod_{i=1}^{2N} \max(1, 2^{-H_2^\delta(S_i|E_i)\tau_{S_i E_i} - H_2(S_i|R_i)\rho_{S_i R_i}})}, \end{aligned} \quad (64)$$

where

$$c = 2 \left(\left(1 + \frac{2^{\xi+1}}{2^{2\xi} + 1} \max_i (\max(2^{-H_2(S_i|R_i)\rho_{S_i R_i}}, 2^{H_2(S_i|R_i)\rho_{S_i R_i}})) \right)^{N-1} - 1 \right). \quad (65)$$

In the following, we introduce the results of applying the decoupling theorems to prove the AQEC performance of $\mathfrak{C}$ and $\mathfrak{B}$. We mainly present the results associated with $\mathfrak{C}$ since its proven achievable encoding rate is better. We also show the advantage of $\mathfrak{C}$ and $\mathfrak{B}$ compared to the block-encoding method in Sec. IV F.

B. Pauli error

We now analyze the performance of random codes under local Pauli noise. This type of noise is particularly important as it is frequently used to model noise encountered in quantum computing [81,82]. Notably, any noise can be turned into Pauli noise by randomized compiling [83]. Specifically, we analyze the Pauli noise model with parameter $\vec{p} = (p_I, p_X, p_Y, p_Z)$, as defined in Eq. (35). By applying the nonsmooth decoupling theorem, we prove the following result.

Corollary 1 (AQEC performance of $\mathfrak{C}$ for Pauli noise). In the large n limit, suppose k and $\vec{p}$ satisfy $k/n \leq 1 - f(\vec{p})$ and $(\varepsilon/n)^{1-k/n} = o(1/n)$, the expected Choi error of the random codes formed by $\mathfrak{C}_n^\varepsilon$ against the strength- $\vec{p}$ local Pauli noise channel satisfies

$$\mathbb{E}_{U \sim \mathfrak{C}_n^\varepsilon} \epsilon_{\text{Choi}} \leq \left(2^{-n(1-f(\vec{p})-\frac{k}{n})} + \frac{4\varepsilon^{1-\frac{k}{n}} n^{\frac{k}{n}}}{\log(n/\varepsilon)} \right)^{\frac{1}{4}}, \quad (66)$$

where $f(\vec{p}) = 2 \log(\sqrt{p_I} + \sqrt{p_X} + \sqrt{p_Y} + \sqrt{p_Z})$. That is, the 1D $O(\log n)$ -depth double-layer blocked circuits can achieve an encoding rate $1 - f(\vec{p})$.

Proof sketch. To prove this corollary, we straightforwardly apply the nonsmooth decoupling theorem or

Theorem 1. We substitute the state $\tau_{S_i E_i}$ and $\rho_{S_i R_i}$ in Theorem 1 with the Choi state of local Pauli noise and the maximally entangled state, respectively. For local Pauli noise, the conditional collision entropy $H_2(S|E)_{\tau_{SE}}$ is given by $n(1 - f(\vec{p}))$. Meanwhile, the entropy $H_2(S|R)_{\rho_{SR}}$ captures the entanglement of the initial state and is equal to $-k$, where k denotes the number of Bell states in $|\hat{\phi}\rangle\langle\hat{\phi}|$. Then, we get the first term $2^{-n(1-f(\vec{p})-k/n)}$ in the square root. The second term comes from the bound of coefficient c with standard inequalities. See Appendix B 3 for the detailed proof. ■

Corollary 1 shows that the Choi error can be made polynomially small by choosing ε to be polynomially small. The exponential decay term $2^{-n(1-f(\vec{p})-k/n)}$ corresponds to the result associated with random stabilizer codes, as discussed in Appendix B 12. The polynomial decay term arises from the approximation error of $\mathfrak{C}$ compared to the Haar-random ensemble. As a consequence, we show that 1D double-layer blocked $O(\log n)$ -depth random circuits maintain a constant encoding and error-correcting rate under local Pauli noise. The encoding rate reaches $1 - 2 \log(\sqrt{p_I} + \sqrt{p_X} + \sqrt{p_Y} + \sqrt{p_Z})$ for local Pauli noise with a polynomially small Choi error. The polynomial decay of the Choi error is particularly beneficial when the logical state needs to be protected over extended periods.

Similar to Corollary 1, for the 1D brickwork circuits $\mathfrak{B}_n^\varepsilon$, we can also show that the encoding scheme can reach the encoding rate $1 - 2 \log(\sqrt{p_I} + \sqrt{p_X} + \sqrt{p_Y} + \sqrt{p_Z})$ for local Pauli noise. The difference is that the polynomial decay term within the bracket changes to $(\varepsilon/k)^{n/k-2}$. The full analysis is shown in Appendix C 3.

When we apply the smooth decoupling theorem, or Theorem 3, we obtain a higher achievable encoding rate,

known as the hashing bound. Specifically, we use Eq. (64) by smoothing only $H_2(S_i|E_i)$.

Corollary 2 (Random codes from $\mathcal{C}$ achieve the hashing bound for Pauli noise). For any $\delta > 0$, in the large n limit, suppose k and $\vec{p}$ satisfy $k/n \leq 1 - h(\vec{p}) - \delta$ and $\lim_{n \rightarrow \infty} \log n / \log(n/\varepsilon) = 0$, the expected Choi error of the random codes formed by $\mathcal{C}_n^\varepsilon$ against the strength- $\vec{p}$ local Pauli noise channel satisfies

$$\mathbb{E}_{U \sim \mathcal{C}_n^\varepsilon} \epsilon_{\text{Choi}} \leq \sqrt{8\delta + \sqrt{2^{-n(1-h(\vec{p})-\frac{k}{n}-\delta)} + \frac{4\varepsilon^{1-\frac{k}{n}} n^{\frac{k}{n}}}{\log(n/\varepsilon)}}}. \quad (67)$$

Here, $h(\vec{p}) = -p_I \log p_I - p_X \log p_X - p_Y \log p_Y - p_Z \log p_Z$. That is, the 1D $\omega(\log n)$ -depth double-layer blocked circuits can achieve the hashing bound $1 - h(\vec{p})$.

Proof sketch. The proof of Corollary 2 closely follows the proof of Corollary 1, with the key distinction being the evaluation of the smoothed entropy. By the asymptotic equipartition property (AEP) of conditional entropy, the smoothed collision entropy converges to the von Neumann entropy. For local Pauli noise, the von Neumann entropy $H(S|E)_{\tau_{\text{SE}}}$ is given by $n(1 - h(\vec{p}))$. The entropy $H_2(S|R)_{\rho_{\text{SR}}}$ captures the initial entanglement and is equal to $-k$. Combining these results, we obtain that the encoding rate achievable under local Pauli noise approaches the hashing bound. The detailed analysis is provided in Appendix B 4. ■

Note that, compared to Corollary 1, Corollary 2 introduces an additional requirement: $\lim_{n \rightarrow \infty} \log n / \log(n/\varepsilon) = 0$. This condition arises from the AEP when the smooth conditional collision entropy approximates the conditional Shannon entropy. The condition ensures that the approximation error in AEP diminishes as n increases. It can be satisfied by choosing $\log(n/\varepsilon)$ to be any function that grows slightly faster than $\log n$, such as $(\log n) \log \log \log n$.

While smoothing the conditional collision entropy introduces an additional error term of 8δ , the results remain significant. Since δ can be chosen arbitrarily small, we can take the limit $\delta \rightarrow 0$ and $n \rightarrow \infty$. In this limit, the Choi error approaches 0. Therefore, for Pauli noise parameterized by $\vec{p} = (p_I, p_X, p_Y, p_Z)$, the Choi error diminishes to zero if the encoding rate k/n is less than $1 - h(\vec{p})$. Hence, we conclude that the $\omega(\log n)$ -depth encoding circuit from $\mathcal{C}$ can achieve the hashing bound. The hashing bound $1 - h(\vec{p})$ can be much higher than the previous bound $1 - f(\vec{p})$ from the nonsmooth decoupling theorem.

C. Erasure error

We now analyze the performance of random codes under erasure errors, an important error model extensively

studied in prior work [21,22] and usually used to assess the performance of quantum codes. We first consider strength- p i.i.d. erasure errors defined in Eq. (41). The derivation is similar to that for Pauli noise, with the differences arising from the noise channel and the evaluation of $H_2(S|E)_{\tau_{\text{SE}}}$ and $H(S|E)_{\tau_{\text{SE}}}$. We provide the results below.

Corollary 3 (AQEC performance of $\mathcal{C}$ for i.i.d. erasure error). In the large n limit, suppose k and p satisfy $k/n \leq 1 - \log(1 + 3p)$ and $(\varepsilon/n)^{1-k/n} = o(1/n)$, the expected Choi error of the random codes formed by $\mathcal{C}_n^\varepsilon$ against the strength- p i.i.d. erasure error satisfies

$$\mathbb{E}_{U \sim \mathcal{C}_n^\varepsilon} \epsilon_{\text{Choi}} \leq \left(2^{-n(1-\log(1+3p)-\frac{k}{n})} + \frac{4\varepsilon^{1-\frac{k}{n}} n^{\frac{k}{n}}}{\log(n/\varepsilon)} \right)^{\frac{1}{4}}. \quad (68)$$

That is, the 1D $O(\log n)$ -depth double-layer blocked circuits can achieve an encoding rate $1 - \log(1 + 3p)$ for strength- p erasure error.

The detailed proof is available in Appendix B 5. Corollary 3 demonstrates that the Choi error can be made polynomially small when ε is polynomially small. The exponential decay term $2^{-n(1-\log(1+3p)-\frac{k}{n})}$ arises from the result associated with random stabilizer codes, while the polynomial decay term comes from the approximation error of $\mathcal{C}_n^\varepsilon$. Consequently, we show that 1D double-layer blocked $O(\log n)$ -depth random circuits achieve an encoding rate of $1 - \log(1 + 3p)$ for i.i.d. erasure error with a vanishing Choi error.

Similar to Corollary 3, for the 1D brickwork circuits $\mathfrak{B}_n^\varepsilon$, the encoding rate $1 - \log(1 + 3p)$ is also achievable for i.i.d. erasure error with a small Choi error $O((\varepsilon/k)^{n/4k-\frac{1}{2}})$. The full analysis is shown in Appendix C 3.

If we apply the smooth decoupling theorem, or Theorem 3, we find that $\mathcal{C}$ can further achieve the quantum channel capacity for the erasure error, which is $1 - 2p$. This result is shown below. Specifically, we use Eq. (64) by smoothing only $H_2(S_i|E_i)$.

Corollary 4 (Random codes from $\mathcal{C}$ achieve the quantum channel capacity for i.i.d. erasure error). Given $\delta > 0$, in the large n limit, suppose k and p satisfy $k/n \leq 1 - 2p - \delta$ and $\lim_{n \rightarrow \infty} \log n / \log(n/\varepsilon) = 0$, the expected Choi error of the random codes formed by $\mathcal{C}_n^\varepsilon$ against the strength- p i.i.d. erasure error satisfies

$$\mathbb{E}_{U \sim \mathcal{C}_n^\varepsilon} \epsilon_{\text{Choi}} \leq \sqrt{8\delta + \sqrt{2^{-n(1-2p-\frac{k}{n}-\delta)} + \frac{4\varepsilon^{1-\frac{k}{n}} n^{\frac{k}{n}}}{\log(n/\varepsilon)}}}. \quad (69)$$

That is, the 1D $\omega(\log n)$ -depth double-layer blocked circuits can achieve the erasure channel capacity $1 - 2p$.

The detailed proof is provided in Appendix B 6. The additional requirement $\lim_{n \rightarrow \infty} \log n / \log(n/\varepsilon) = 0$ arises from AEP, just as it does for Pauli noise. By taking the limits $\delta \rightarrow 0$ and $n \rightarrow \infty$, the Choi error diminishes to zero if the encoding rate k/n is less than $1 - 2p$. Hence, we conclude that the $\omega(\log n)$ -depth encoding circuit from $\mathfrak{C}$ can achieve the quantum channel capacity $1 - 2p$ for i.i.d. erasure error.

In addition to i.i.d. erasure, we also consider the fixed-number erasure error, which was studied in Ref. [22]. This noise model is defined in Eq. (39), where the number of erasures is t . In this error model, a subsystem of size t is randomly selected and traced out. Similar to Corollary 3, we show that these codes maintain a constant rate of information storage and error correction capabilities. Specifically, given the logical qubit count k and the number of erasure errors t , we have that for sufficiently large n , as long as $1 - k/n - \log(1 + 3t/n) \geq 0$, and $(\varepsilon/n)^{1-k/n} = o(1/n)$, the expected Choi error over $\mathfrak{C}_n^\varepsilon$ against the random t erasure errors satisfies

$$\mathbb{E} \epsilon_{\text{Choi}} \leq \left(2^{-(n-2t-k)} + \frac{4\varepsilon^{1-\frac{k}{n}} n^{\frac{k}{n}}}{\log(n/\varepsilon)} \right)^{\frac{1}{4}}. \quad (70)$$

Note that ε is only required to be polynomially small to satisfy $(\varepsilon/n)^{1-k/n} = o(1/n)$. Consequently, the circuit depth is $O(\log(n/\varepsilon)) = O(\log n)$, and the Choi error decays polynomially as the number of qubits n increases. The maximal number of erasure errors can be $n/3$. The full detail is available in Appendix B 10.

For random stabilizer code, the Choi error can be bounded by the first term $2^{-(n-2t-k)/4}$. Thus, random stabilizer codes perform well when $n - 2t - k > 0$, which saturates the quantum Singleton bound, $n - k - 2(d - 1) \geq 0$, for an $[n, k, d]$ code, as distance- d code can guarantee to correct $d - 1$ erasure errors [40]. Here, the codes from $\mathfrak{C}$ require the condition $n - k - n \log(1 + 3t/n) \geq 0$, which is close to the Singleton bound $n - k - 2t \geq 0$ up to a constant factor. This raises a question regarding whether the low-depth codes can achieve the “Singleton bound” for AQEC under random t -erasure errors. Meanwhile, notice that the condition $n - k - n \log(1 + 3t/n) \geq 0$ is in accord with the requirement when applying the nonsmooth decoupling theorem to i.i.d. erasure error. Also, applying the smooth decoupling theorem to i.i.d. erasure error gives an encoding rate $1 - 2p$, which is exactly the Singleton bound. It is possible to enhance Eq. (70) by “smoothing” the result, which we leave for future work.

D. Amplitude damping noise

We now analyze the performance of random codes under amplitude damping noise [71]. The derivation is similar to that for Pauli noise and erasure errors, with

differences in the noise channel and the evaluation of $H_2(S|E)_{\tau_{\text{SE}}}$ and $H(S|E)_{\tau_{\text{SE}}}$. We provide the results below.

Corollary 5 (AQEC performance of $\mathfrak{C}$ for amplitude damping noise). In the large n limit, suppose k and p satisfy $k/n \leq -\log(1/(2-p) + \sqrt{p/(2-p)})$ and $(\varepsilon/n)^{1-k/n} = o(1/n)$, the expected Choi error of the random codes formed by $\mathfrak{C}_n^\varepsilon$ against the strength- p local amplitude damping noise satisfies

$$\mathbb{E}_{U \sim \mathfrak{C}_n^\varepsilon} \epsilon_{\text{Choi}} \leq \left(2^{-n(-\log(\frac{1}{2-p} + \sqrt{\frac{p}{2-p}}) - \frac{k}{n})} + \frac{4\varepsilon^{1-\frac{k}{n}} n^{\frac{k}{n}}}{\log(n/\varepsilon)} \right)^{\frac{1}{4}}, \quad (71)$$

That is, the 1D $O(\log n)$ -depth double-layer blocked circuits can achieve an encoding rate $-\log(1/(2-p) + \sqrt{p/(2-p)})$ for strength- p local amplitude damping noise.

The detailed proof is available in Appendix B 7. Corollary 5 demonstrates that the Choi error can be made polynomially small when ε is polynomially small.

Similar to Corollary 5, for the 1D brickwork circuits $\mathfrak{B}_n^\varepsilon$, the encoding rate $-\log(1/(2-p) + \sqrt{p/(2-p)})$ is also achievable for local amplitude damping noise with a small Choi error $O((\varepsilon/k)^{\frac{n}{4k} - \frac{1}{2}})$. The full analysis is shown in Appendix C 3.

If we apply the smooth decoupling theorem, we find that $\mathfrak{C}$ can achieve a better encoding rate $h((1-p)/2) - h(p/2)$. This result is shown below.

Corollary 6 (Better AQEC performance of $\mathfrak{C}$ for amplitude damping noise). Given $\delta > 0$, in the large n limit, suppose k and p satisfy $k/n \leq h((1-p)/2) - h(p/2) - \delta$ and $\lim_{n \rightarrow \infty} \log n / \log(n/\varepsilon) = 0$, the expected Choi error of the random codes formed by $\mathfrak{C}_n^\varepsilon$ against the strength- p local amplitude damping noise satisfies

$$\mathbb{E}_{U \sim \mathfrak{C}_n^\varepsilon} \epsilon_{\text{Choi}} \leq \sqrt{8\delta + \sqrt{2^{-n(h(\frac{1-p}{2}) - h(\frac{p}{2}) - \frac{k}{n} - \delta))} + \frac{4\varepsilon^{1-\frac{k}{n}} n^{\frac{k}{n}}}{\log(n/\varepsilon)}}. \quad (72)$$

That is, the 1D $\omega(\log n)$ -depth double-layer blocked circuits can achieve the encoding rate $h((1-p)/2) - h(p/2)$.

The detailed proof is provided in Appendix B 8. The bound of the encoding rate is $h((1-p)/2) - h(p/2)$, which is the same as the bound for random stabilizer codes. Nonetheless, this rate may be lower than the quantum capacity of local amplitude damping noise given in

Eq. (47). The discrepancy arises because both $\mathfrak{C}$ and random stabilizer codes are symmetric when interchanging the computational basis, whereas amplitude damping noise is inherently asymmetric.

E. Correlated noise

In the above, we show the AQEC performance of 1D log-depth circuits for local noises. Below, we demonstrate that our analysis extends beyond this type. By applying Theorem 1, we show that codes from $\mathfrak{C}$ can achieve a constant rate of encoding against correlated noise. The derivation is similar to that for local noise, with the differences arising from the noise channel and the evaluation of $H_2(S|E)_{\text{TSE}}$. The result is summarized below.

Corollary 7 (Random codes from $\mathfrak{C}$ possess a threshold for 1D nearest-neighbor ZZ-coupling noise). In the large n limit, suppose k and p satisfy $k/n < 1 - 2 \log(\sqrt{1-p} + \sqrt{p})$ and $(\varepsilon/n)^{1-k/n} = o(1/n)$, the expected Choi error of the random codes formed by $\mathfrak{C}_n^\varepsilon$ against the strength- p nearest-neighbor ZZ-coupling noise satisfies

$$\mathbb{E}_{U \sim \mathfrak{C}_n^\varepsilon} \epsilon_{\text{Choi}} \leq \left(2^{-n(1-2 \log(\sqrt{1-p} + \sqrt{p}) - \frac{k}{n})} + \frac{4\varepsilon^{1-\frac{k}{n}} n^{\frac{k}{n}}}{\log(n/\varepsilon)} \right)^{\frac{1}{4}}. \quad (73)$$

That is, the 1D $O(\log n)$ -depth double-layer blocked circuits can achieve an encoding rate $1 - 2 \log(\sqrt{1-p} + \sqrt{p})$ for strength- p 1D nearest-neighbor ZZ-coupling noise.

The detailed proof is available in Appendix B9. Corollary 7 demonstrates that the Choi error is polynomially small when ε is polynomially small. The exponential decay term $2^{-n(1-2 \log(\sqrt{1-p} + \sqrt{p}) - \frac{k}{n})}$ is associated with random stabilizer codes. Consequently, we show that 1D double-layer blocked $O(\log n)$ -depth random circuits achieve an encoding rate of $1 - 2 \log(\sqrt{1-p} + \sqrt{p})$ for nearest-neighbor ZZ-coupling noise with a polynomially small Choi error. For 1D brickwork circuit $\mathfrak{B}$, we prove a similar result with a modification of the bound to $1 - 2(1 + k/n) \log(\sqrt{1-p} + \sqrt{p})$. The modification originates from a finite local dimension for $\mathfrak{B}$. The full analysis is shown in Appendix C3.

F. Advantages of the 1D double-layer blocked encoding scheme over the block-encoding scheme

In addition to proving that our 1D circuit encoding scheme can achieve a high AQEC performance against local noises and correlated noise, we now specifically discuss its advantage over the block-encoding scheme.

Previous works show that the one-dimensional block-encoding scheme can have a threshold against random

erasure error [21] and can achieve the hashing bound against Pauli noise [23] at a logarithmic depth. Here, we show that although the achievable encoding rate for the block-encoding method is the same as that for the double-layer blocked circuit $\mathfrak{C}$, $\mathfrak{C}$ can have advantages in the required circuit depth and logical error rate.

Particularly, we will show that to correct a given noise with a fixed encoding rate, the required circuit depth of $\mathfrak{C}$ is lower, and the Choi error is lower for $\mathfrak{C}$. In other words, given certain noise channel the circuit depth the double-layer blocked encoding scheme can accomplish error correction but the block-encoding scheme cannot.

Using our notation, the encoding unitary operation of the block-encoding method can be written as

$$U = \bigotimes_{i=1}^{\frac{N}{2}} U_C^{Ai-3, 4i-2, 4i-1, 4i}, \quad (74)$$

where $U_C^{Ai-3, 4i-2, 4i-1, 4i}$ is a random Clifford gate on regions $4i-3, 4i-2, 4i-1, 4i$. The location of regions is shown in Fig. 1(b). The encoding unitary U is a tensor product of parallel Clifford gates. Note that the depth to realize an n -qubit Clifford gate depends linearly on n . To make a fair comparison, we let the block size be 4ξ so that the depths to realize the block encoding and the double-layer blocked encoding schemes are the same. We denote the set of the random block-encoding unitary operations as $\mathfrak{S}_n^B$.

In Appendix B13, we compute the Choi error of the block-encoding scheme and compare it to the double-layer blocked encoding scheme. For local erasure error, we rigorously establish an advantage of the double-layer blocked encoding scheme over the block-encoding scheme. We evaluate the lower bound of the Choi error for the block-encoding method as follows:

$$\mathbb{E}_{U \sim \mathfrak{S}_n^B} \epsilon_{\text{Choi}} \geq \min \left\{ n \left(\frac{\varepsilon}{n} \right)^{8D(\frac{1-k}{2} \|p\| + o(1))}, \Omega(1) \right\}, \quad (75)$$

where D is the binary relative entropy. Particularly, the $\Omega(1)$ term is a positive constant. When $n(\varepsilon/n)^{8D(\frac{1-k}{2} \|p\| + o(1))}$ vanishes, the Choi error also vanishes. If $n(\varepsilon/n)^{8D(\frac{1-k}{2} \|p\| + o(1))}$ does not converge, the Choi error will have a nonvanishing value, and the error correction fails.

Recalling Corollary 3, the Choi error for the double-layer blocked encoding scheme is

$$\mathbb{E}_{U \sim \mathfrak{C}_n^\varepsilon} \epsilon_{\text{Choi}} \leq \left(2^{-n(1-\log(1+3p) - \frac{k}{n})} + \frac{4\varepsilon^{1-\frac{k}{n}} n^{\frac{k}{n}}}{\log(n/\varepsilon)} \right)^{\frac{1}{4}}. \quad (76)$$

Based on Eqs. (75) and (76), there exist regimes of p , encoding rate k/n , and ε such that

$$\lim_{n \rightarrow \infty} \mathbb{E}_{U \sim \mathfrak{C}_n^B \in \text{Choi}} \geq \lim_{n \rightarrow \infty} \min \left\{ n \left(\frac{\varepsilon}{n} \right)^{8D(\frac{1-k}{2n} \|p\| + o(1))}, \Omega(1) \right\} = \Omega(1), \quad (77)$$

$$\lim_{n \rightarrow \infty} \mathbb{E}_{U \sim \mathfrak{C}_n^\varepsilon \in \text{Choi}} \leq \lim_{n \rightarrow \infty} \left(2^{-n(1 - \log(1+3p) - \frac{k}{n})} + \frac{4\varepsilon^{1-\frac{k}{n}} n^{\frac{k}{n}}}{\log(n/\varepsilon)} \right)^{\frac{1}{4}} = 0. \quad (78)$$

A explicit example is when $p = 0.25$, $k/n = 0.2$, $\varepsilon = n^{-0.375}$. In this case,

$$\mathbb{E}_{U \sim \mathfrak{C}_n^B \in \text{Choi}} \geq \min\{n^{0.14}, \Omega(1)\} = \Omega(1), \quad (79)$$

$$\mathbb{E}_{U \sim \mathfrak{C}_n^\varepsilon \in \text{Choi}} \leq n^{-0.025}. \quad (80)$$

Thus, given $p = 0.25$ and $k/n = 0.2$, the block-encoding method requires a smaller ε to make the Choi error vanish, inducing a deeper circuit depth. Hence, we demonstrate that the required circuit depth of $\mathfrak{C}_n^\varepsilon$ is lower, and the Choi error is also lower.

We show the full analysis of the regime where Eqs. (77) and (78) hold in the Appendix. One can also find the regime where $\lim_{n \rightarrow \infty} (\mathbb{E}_{U \sim \mathfrak{C}_n^\varepsilon \in \text{Choi}} / \mathbb{E}_{U \sim \mathfrak{C}_n^B \in \text{Choi}}) = 0$. In this regime, the double-layer blocked scheme outperforms the separate block approach. Thus, the double-layer blocked encoding scheme is proven to have an advantage in certain error regimes against erasure errors. We expect this result to hold for more generic noise models, which is numerically verified for local Pauli noise [23]. We provide more details and discussions in Appendix B 13.

The above comparison can be generalized for 1D brickwork circuits $\mathfrak{B}$ and block-encoding ensemble $\mathfrak{C}_n^B$, where the former is upper-bounded by $O((\varepsilon/k)^{n/4k - \frac{1}{2}})$. Evaluating the regime where this upper bound is smaller than $n(\varepsilon/n)^{8D(\frac{1-k}{2n} \|p\| + o(1))}$ suffices for showing the advantage for $\mathfrak{B}$. Clearly, when the encoding rate k/n is a small constant, $\mathfrak{B}$ will outperform the block approach.

One can observe that the advantages of $\mathfrak{C}$ and $\mathfrak{B}$ are both manifested in the low encoding rate regime. This phenomenon is expected to be related to the stronger scrambling power of $\mathfrak{C}$ and $\mathfrak{B}$ compared to $\mathfrak{C}_n^B$. If the encoding rate is large, like 0.5, a one-layer gate suffices to scramble the logical information to the whole system to protect it. A stronger scrambling power will be useless. In contrast, when the logical qubit number is small, $\mathfrak{C}$ and $\mathfrak{B}$ can scramble their information to the whole system faster and then manifest an advantage.

G. Approximate quantum error correction and decoupling for arbitrary input states

We have previously established decoupling theorems for codes from $\mathfrak{C}$ and $\mathfrak{B}$ with input states that exhibit a product structure across different regions. Using these theorems, we demonstrated the AQEC performance of random codes with the maximally entangled state, corresponding to the case of random inputs. In this part, we extend these results to show that similar decoupling theorems and error-correcting properties hold for arbitrary input states.

We begin by noting that if a unitary encoding U achieves a small Choi error, it also produces a comparably small recovery error for any input state when the encoding operation is combined with a local random unitary operation. Specifically, we select the local random unitary to be a random Pauli matrix and define the encoding and decoding operations as

$$\mathcal{E}_P(\psi) = U[(P\psi P^\dagger)_L \otimes |0\rangle\langle 0|_{S \setminus L}]U^\dagger, \quad (81)$$

$$\mathcal{D}_P(\psi) = P\mathcal{D}(\psi)P^\dagger. \quad (82)$$

Such a random Pauli encoding twirls the original channel $\mathcal{D} \circ \mathcal{N} \circ \mathcal{E}$ into a Pauli channel, ensuring that the Choi error provides a lower bound on the recovery error for arbitrary input states. We establish this result using entanglement fidelity.

Proposition 1 (Recovery fidelity via Pauli twirling). Consider an encoding and decoding procedure $\mathcal{D} \circ \mathcal{N} \circ \mathcal{E}$ with entanglement fidelity F_{Choi} . For any input state ψ_{LR} , we have

$$\mathbb{E}_P F(\psi_{\text{LR}}, [(\mathcal{D}_P \circ \mathcal{N} \circ \mathcal{E}_P)_L \otimes I_R](\psi_{\text{LR}})) \geq F_{\text{Choi}}, \quad (83)$$

where the expectation $\mathbb{E}_P$ is taken over all Pauli matrices P uniformly at random.

Proof sketch. Applying Uhlmann's theorem [27,84], we express the fidelity of the two states in terms of the fidelity between corresponding purifications. By performing a Pauli twirl on these purifications using random Pauli matrices, we reduce the expression to the entanglement fidelity, thereby establishing the lower bound. See Appendix B 11 for the detailed derivation. ■

The encoding and decoding procedures using random Pauli matrices are independent of the input state; thus, they consistently provide high recovery fidelity for arbitrary input states. This quality offers additional insight into the AQEC performance of random codes. We define a unitary ensemble to be right-invariant under Pauli matrices if, for any Pauli gate P and unitary $U \in \mathfrak{U}$, the probability

$p_{\mathfrak{U}}(U)$ of drawing U from $\mathfrak{U}$ is the same as the probability $p_{\mathfrak{U}}(UP)$ of drawing UP from $\mathfrak{U}$,

$$p_{\mathfrak{U}}(U) = p_{\mathfrak{U}}(UP). \quad (84)$$

The Pauli twirling operation is naturally incorporated in a right-invariant ensemble without incurring additional overhead. Consequently, we expect such an ensemble to exhibit robust AQEC performance for any input state. Consider the recovery error ϵ_{ψ} for an arbitrary input state ψ ,

$$\epsilon_{\psi} = \min_{\mathcal{D}} P(\psi_{\text{LR}}, [(\mathcal{D} \circ \mathcal{N} \circ \mathcal{E})_L \otimes I_R](\psi_{\text{LR}})). \quad (85)$$

A direct consequence of Proposition 1 is that the expected recovery error for any arbitrary input state is bounded above by the expected Choi error for any right-invariant unitary ensemble $\mathfrak{U}$.

Corollary 8. If the unitary ensemble $\mathfrak{U}$ is right-invariant under Pauli rotations, then for any state ψ_{LR} , the average error $\mathbb{E}_{\mathfrak{U}}\epsilon_{\psi}$ does not exceed the average Choi error,

$$\mathbb{E}_{\mathfrak{U}}\epsilon_{\psi} \leq \mathbb{E}_{\mathfrak{U}}\epsilon_{\text{Choi}}. \quad (86)$$

Proof. This comes from direct calculation,

$$\begin{aligned} \mathbb{E}_{\mathfrak{U}}\epsilon_{\psi} &= \mathbb{E}_{\mathfrak{U}} \min_{\mathcal{D}} P(\psi_{\text{LR}}, [(\mathcal{D} \circ \mathcal{N} \circ \mathcal{E})_L \otimes I_R](\psi_{\text{LR}})) \\ &= \mathbb{E}_{\mathfrak{U}} \mathbb{E}_P \min_{\mathcal{D}} P(\psi_{\text{LR}}, [(\mathcal{D} \circ \mathcal{N} \circ \mathcal{E})_L \otimes I_R](P\psi_{\text{LR}}P^{\dagger})) \\ &\leq \mathbb{E}_{\mathfrak{U}} \mathbb{E}_P P(\psi_{\text{LR}}, [(\mathcal{D}_P^{(0)} \circ \mathcal{N} \circ \mathcal{E})_L \otimes I_R](P\psi_{\text{LR}}P^{\dagger})) \\ &= \mathbb{E}_{\mathfrak{U}} \mathbb{E}_P P(\psi_{\text{LR}}, [(\mathcal{D}_P^{(0)} \circ \mathcal{N} \circ \mathcal{E}_P)_L \otimes I_R](\psi_{\text{LR}})) \\ &\leq \mathbb{E}_{\mathfrak{U}}\epsilon_{\text{Choi}}. \end{aligned} \quad (87)$$

The second line follows from the fact that $\mathfrak{U}$ is right-invariant under Pauli matrices. In the third line, we choose $\mathcal{D}_P^{(0)}(\psi) = P\mathcal{D}^{(0)}(\psi)P^{\dagger}$ as the recovery channel, where $\mathcal{D}^{(0)}$ is the optimal recovery channel that maximizes the entanglement fidelity of $\mathcal{D} \circ \mathcal{N} \circ \mathcal{E}$. This inequality holds because the specific recovery channel $\mathcal{D}_P^{(0)}$ cannot have a smaller recovery error than the optimal recovery channel. The fifth line comes from Proposition 1. ■

The recovery error for arbitrary input state established in Corollary 8 is closely related to the decoupling between the reference system and environment, as described by the complementary channel formalism [see Eqs. (31) and (32)]. The recovery error can be expressed as

$$\epsilon_{\psi} = \min_{\zeta} P\left(\widehat{(\mathcal{N} \circ \mathcal{E}_{L \rightarrow E} \otimes I_R)}(\psi_{\text{LR}}), \zeta_E \otimes \text{tr}_L(\psi_{\text{LR}})\right), \quad (88)$$

where a low value of ϵ_{ψ} indicates that $\widehat{(\mathcal{N} \circ \mathcal{E}_{L \rightarrow E} \otimes I_R)}(\psi_{\text{LR}})$ is close to a tensor-product state $\zeta_E \otimes \text{tr}_L(\psi_{\text{LR}})$.

By Combining Eq. (88), Eq. (15) with Corollary 8, we conclude that the environment is approximately decoupled from the reference system, with the closeness bounded by the Choi error.

Proposition 2 (Decoupling for arbitrary input states). If the unitary ensemble $\mathfrak{U}$ is right-invariant under Pauli rotations, then for any state ψ_{LR} , the expected closeness of decoupling is bounded above by the expected Choi error,

$$\begin{aligned} \mathbb{E}_{U_S \sim \mathfrak{U}} \min_{\tau_E} \|\hat{\mathcal{N}}_{S \rightarrow E}[U_S(\psi_{\text{LR}} \otimes |0\rangle\langle 0|_{S \setminus L})U_S^{\dagger}] - \tau_E \otimes \rho_R\|_1 \\ \leq 2\mathbb{E}_{U_S \sim \mathfrak{U}}\epsilon_{\text{Choi}}. \end{aligned} \quad (89)$$

Here, ϵ_{Choi} represents the Choi error for the unitary encoding U_S under the noise channel $\mathcal{N}$. The coefficient 2 comes from the relationship between the 1-norm distance and the purified distance.

Note that the ensembles $\mathfrak{C}$ and $\mathfrak{B}$ are right-invariant under Pauli rotations. This decoupling result extends the established decoupling theorems, Theorem 1, Theorem 2, and Theorem 3, by eliminating the requirement for the input state to be in a tensor-product form. Different from the previous two theorems, the environmental state τ_E does not have an explicit form in this generalization.

Note that the state in Proposition 1 is arbitrary. One can take the minimal one over all states. Then, $\min_{\psi} \mathbb{E}_P F(\psi_{\text{LR}}, [(\mathcal{D}_P \circ \mathcal{N} \circ \mathcal{E}_P)_L \otimes I_R](\psi_{\text{LR}}))$ is close to 1 when the Choi fidelity is close to 1. We remark that although the encoding and decoding procedure $\mathcal{E}_P$ and $\mathcal{D}_P$ do not depend on the input state, the recovery fidelity derived in Proposition 1 do not provide a bound for the normal worst-case fidelity or worst-case error. The definition of the worst-case error requires the encoding and decoding operations to be fixed, without any probability mixing,

$$\begin{aligned} F_{\text{worst}} &= \max_{\mathcal{D}} \min_{R, \rho_{\text{LR}}} F(\rho_{\text{LR}}, [(\mathcal{D} \circ \mathcal{N} \circ \mathcal{E})_L \otimes I_R](\rho_{\text{LR}})), \\ \epsilon_{\text{worst}} &= \sqrt{1 - F_{\text{worst}}^2} \\ &= \min_{\mathcal{D}} \max_{R, \rho_{\text{LR}}} P(\rho_{\text{LR}}, [(\mathcal{D} \circ \mathcal{N} \circ \mathcal{E})_L \otimes I_R](\rho_{\text{LR}})). \end{aligned} \quad (90)$$

That is, F_{worst} is defined for a specific encoding $\mathcal{E}$. In contrast, $\min_{\psi} \mathbb{E}_P F(\psi_{\text{LR}}, [(\mathcal{D}_P \circ \mathcal{N} \circ \mathcal{E}_P)_L \otimes I_R](\psi_{\text{LR}}))$ considers an ensemble of encodings and decodings $\{\mathcal{E}_P, \mathcal{D}_P\}_P$ and shows that this ensemble achieves high expected recovery fidelity for any input state. The worst-case fidelity is not guaranteed for a particular choice of $\mathcal{E}_P$.

Nonetheless, when the Choi error is small, it can be shown that there exists a code subspace that exhibits good worst-case error for a fixed encoding procedure. Specifically, we use the code space $C = \text{span}\{U_S|i\rangle_L|0\rangle_{S \setminus L}\}$ spanned by the logical state to represent a coding scheme,

and the worst-case error is only dependent on C . It has been shown that when the Choi fidelity of the original encoding is $1 - \epsilon$, there exists a large subcode $C' \subset C$ such that

$$\dim C' \geq \left\lfloor \frac{1}{2} \dim C \right\rfloor, \quad (91)$$

and the worst-case fidelity for this subcode satisfies [26]

$$F_{\text{worst}}(C') > 1 - 2\epsilon. \quad (92)$$

This implies that as long as the Choi error of the encoding is small, the worst-case error in a large subspace $C' \subset C$ is also small, thus ensuring good error correction performance over a significant portion of the code space.

V. LOWER BOUND OF CIRCUIT DEPTH IN AQEC

After establishing the AQEC performance of a logarithmic-depth circuit for local noise, we now derive lower bounds on the circuit depth required to achieve good AQEC performance. Here, we focus on the lower bound for local Pauli noise, specifically i.i.d. depolarizing noise, which can be easily extended to other local noise models. We show that the scaling of circuit depth in the double-layer blocked encoding $\mathcal{C}$ and 1D brickwork encoding $\mathcal{B}$ is asymptotically optimal. Our analysis also provides a general result encompassing both D -dimensional and all-to-all circuit architectures.

Theorem 4 (Lower bound for circuit depth). For a circuit U that encodes k logical qubits with depth d , if U has a Choi error $\epsilon < 0.1$ against the strength- p local depolarizing channel, then:

- (1) If U is a D -dimensional circuit, then

$$(2d)^D \log\left(\frac{1}{p}\right) + 2D \log(2d) \geq \log\left(\frac{3}{8\epsilon_{\text{Choi}}^2}\right) + \log k. \quad (93)$$

In the large n and d limit, this implies

$$d \geq \frac{1}{2} \left[\frac{\log\left(\frac{3}{8\epsilon^2}\right) + \log k}{(1 + o(1)) \log\left(\frac{1}{p}\right)} \right]^{1/D}. \quad (94)$$

- (2) If U is an all-to-all circuit, then

$$2^d \log\left(\frac{1}{p}\right) + 2d \geq \log\left(\frac{3}{8\epsilon_{\text{Choi}}^2}\right) + \log k. \quad (95)$$

In the large n and d limit, this implies

$$d \geq \log \left[\frac{\log\left(\frac{3}{8\epsilon^2}\right) + \log k}{(1 + o(1)) \log\left(\frac{1}{p}\right)} \right]. \quad (96)$$

Proof sketch. We outline the main idea of the proof and refer to Appendix D for detailed arguments. The proof leverages the analysis of the light cones of each qubit within the circuit. For a circuit of depth d , the light cone of each qubit is bounded by a size M that depends on the circuit depth d and circuit architecture. For instance, in a 1D circuit, $M = 2d$.

We first prove that there exist at least k/M^2 logical qubits with disjoint light cones, and we denote this subset as J . That is, the light cones of any two qubits in J do not intersect with each other. If any of these light cones are totally affected by noise, then any recovery channel cannot have a small recovery error. Since the light cones in J are disjoint, the events of them being totally affected are independent. The probability that none of these light cones are completely traced out and replaced by the maximally mixed state is upper-bounded by

$$p_{\text{None}} \leq (1 - p^M)^{|J|}. \quad (97)$$

We also prove that the Choi error is lower-bounded by the probability that at least one of these light cones is entirely traced out,

$$\epsilon^2 \geq \frac{3}{4} (1 - p_{\text{None}}). \quad (98)$$

Using standard inequalities and substituting M with a function of the circuit depth d that depends on the circuit's structure, we derive the desired lower bounds on the circuit depth. ■

The light-cone argument has been utilized in previous works to establish lower bounds for quantum error-correcting codes, such as determining the number of physical qubits required for fault-tolerant quantum computation [38] and the circuit depth necessary to achieve specific subsystem variance properties of codes [30]. In our work, we extend these lower bounds by considering the AQEC performance under a practical noise model, specifically establishing circuit depth lower bounds under local noise. While our results are derived for local depolarizing noise, they can be easily extended to other error models. Furthermore, our analysis applies to arbitrary circuit architectures characterized by an adjacency graph G . In such cases, the upper bound M on the size of the light cones depends on both the circuit depth and the adjacency graph, $M = M(G, d)$, while the remaining parts of the proof remain unchanged.

Using Theorem 4, we now show that the circuit depth scaling of $\mathcal{C}$ and $\mathcal{B}$ is optimal up to constant factors. Theorem 4 shows that the depth of a 1D circuit should be $\Omega(\log n)$ to achieve a polynomially small Choi error or a constant encoding rate. Meanwhile, 1D circuits $\mathcal{C}$ and $\mathcal{B}$

can achieve such error scaling and encode $k = \Theta(n)$ logical qubits with circuit depth $O(\log n)$. The depth is optimal up to a constant factor.

Moreover, 1D double-layer blocked circuits $\mathfrak{C}$ are nearly optimal for achieving the hashing bound, as the circuit depth $\omega(\log n)$ given in Corollary 2 is close to the lower bound $\Omega(\log n)$ given by Theorem 4. The scaling $\omega(\log n)$ arises from suppressing errors in AEP concentration, suggesting that improving the circuit depth scaling to $O(\log n)$ may be inherently challenging. Exploring potential methods to enhance this bound is an interesting question for future research.

Moreover, we can implement the codes from $\mathfrak{C}$ using an all-to-all circuit architecture to further reduce the circuit depth. In an all-to-all architecture, any ξ -qubit Clifford unitary can be implemented in $O(\log \xi)$ depth with $O(\xi^2)$ ancillary qubits [85,86]. Consequently, the Clifford encoding $\mathfrak{C}$ can be realized with a depth of $O(\log \xi) = O(\log \log(n/\epsilon))$, utilizing a total of $O(n\xi) = O(n \log(n/\epsilon))$ ancillary qubits. This encoding scheme effectively encodes $k = O(n)$ logical qubits with a circuit depth of $O(\log \log(n/\epsilon))$, which is comparable to the circuit depth lower bound $\Omega(\log \log n)$ established in Theorem 4. This comparison indicates that our results provide a tight lower bound applicable to a broad range of circuit architectures. Notably, in such an all-to-all encoding strategy, the encoding rate becomes $O(k/n\xi) = O(\log^{-1}(n/\epsilon))$, which has a gap to a constant encoding rate. Whether all-to-all circuits with $O(\log \log n)$ depth can achieve a constant encoding rate while maintaining good AQEC performance remains an intriguing direction for future research.

VI. CONCLUSION AND DISCUSSION

In this work, we establish the AQEC performance of 1D logarithmic-depth encoding circuits with brickwork and double-layer blocked architecture against a wide range of practical noise models. We show that logarithmic depth is both necessary and sufficient to achieve a constant encoding rate and error-correcting capability while keeping the AQEC error polynomially small. In particular, we prove that for the double-layer blocked circuit, $\omega(\log n)$ depth suffices to achieve the well-known hashing bound under local Pauli noise and the channel capacity under local erasure errors. Although the double-layer blocked circuit is shown to be able to form an approximate design [33] with a polynomially small error, our analysis does not rely on this property. The decoupling theorem for approximate designs [64] generally requires an exponentially small error to achieve a negligible decoupling error. This raises an interesting question of whether more general conditions, such as forming an approximate design with polynomially small error, suffice to ensure good AQEC performance for random codes. Meanwhile, one

TABLE I. The table summarizes the circuit depth scaling of three types of circuits with respect to decoupling against local and weakly correlated noise, as well as their approximate design properties. The second column presents our results on decoupling. The third column reports results on approximate designs, while the second row indicates areas where further improvements are possible.

	Decoupling	Approx. design
1D brickwork	$\Theta(\log n)$	$O(n)$ [31,32]
1D Double-layer blocked	$\Theta(\log n)$	$\Theta(\log n)$ [33,34]
All-to-all	$\Theta(\log \log n)$	$\Theta(\log \log n)$ [33,34]

can further unify the properties of approximate design and decoupling as the scrambling power of a random circuit ensemble and investigate more properties of $\mathfrak{C}$ like out-of-time-ordered correlation [87,88]. For clarity, Table I summarizes the depth scaling of two classes of 1D circuits and all-to-all circuits with respect to decoupling and design properties.

We also introduce a random Pauli rotation before unitary encoding to show that approximate decoupling and small recovery errors are maintained for arbitrary input states. Developing new techniques to bound the worst-case recovery error for random codes without resorting to random Pauli matrices, possibly by further exploiting the internal structure of these encoding circuits, remains an interesting avenue for future work.

Moreover, exploring low-depth circuits in higher-dimensional architectures, such as 2D or all-to-all circuits, is also important. With improved connectivity in these architectures, the required circuit depth can be expected to decrease. For instance, as discussed in Sec. V, all-to-all circuits can use $O(n \log n)$ physical qubits to encode $O(n)$ logical qubits with $O(\log \log n)$ circuit depth while preserving good AQEC performance. An intriguing question is whether the encoding rate can be made constant in the all-to-all circuit while still achieving good AQEC performance. Additionally, it is interesting to investigate whether the hashing bound for 1D circuits can be improved from $\omega(\log n)$ to $O(\log n)$. We also conjecture that, in a D -dimensional circuit architecture, circuit depths on the order of $O((\log n)^{1/D})$ suffice to achieve a constant encoding rate and error-correcting capability in the AQEC setting.

It is worth noting that working with AQEC instead of focusing on the exact code distance offers promising avenues for improvements in various aspects of code performance such as encoding power and logical operation sets. For example, AQEC may allow enhanced parameters for geometrically local codes beyond known fundamental limits such as the Bravyi-Poulin-Terhal bound [89]. Moreover, certain covariant codes under AQEC possess a continuous set of logical operations, a feature that is highly desirable for fault tolerance [22]. Beyond evaluating AQEC performance, achieving fault-tolerant

implementations of codes from low-depth Clifford encoding circuits [13] is also crucial in practice, which requires further analysis of the fault-tolerant implementation of quantum gates and the development of improved decoding algorithms for such codes.

Beyond AQEC, our analysis of 1D random circuits extends the previous domain-wall technique [32] and can be applied to study other properties of random circuits, such as their generic second-moment behavior [90,91]. We have developed decoupling theorems for low-depth circuits, which potentially have important applications in quantum information theory [63,92]. Decoupling theorems have applications in quantum communication tasks like quantum state merging [93] and in quantum cryptography, such as privacy amplification [94]. We also envision a wide range of exciting applications in many-body physics and quantum thermodynamics [95,96]. For instance, a linear number of random erasure errors can be corrected, implying that the subsystem of a 1D log-depth circuit state is highly likely to thermalize. Our results may also offer insights for studies of the black hole information paradox [54] and quantum many-body entanglement [97].

ACKNOWLEDGMENTS

We thank Junjie Chen, Chushi Qin, Zitai Xu, and Xiangran Zhang for their insightful discussions. G.L., Z.D., and X.M. are supported by the National Natural Science Foundation of China (Grant No. 12575023), and the Innovation Program for Quantum Science and Technology (Grants No. 2021ZD0300804 and No. 2021ZD0300702). Z.-W.L. is supported in part by NSFC under Grant No. 12475023, Dushi Program, and startup funding from YMSC.

DATA AVAILABILITY

No data were created or analyzed in this study.

APPENDIX A: QUANTUM ENTROPIES AND DECOUPLING THEOREMS

In this Appendix, we review the related concepts of quantum entropies and the decoupling theorem, which is an important tool for analyzing the AQEC performance of random circuits. First, we introduce several important entropies that we will use. Here, we mainly consider the sandwiched conditional Rényi entropy [98],

$$\tilde{H}_\alpha^\downarrow(A|B)_\rho = -\tilde{D}_\alpha(\rho_{AB} \parallel \mathbb{I}_A \otimes \rho_B), \quad (\text{A1})$$

where $\tilde{D}_\alpha$ denotes the sandwiched Rényi entropy,

$$\tilde{D}_\alpha(\rho \parallel \sigma) = \frac{1}{\alpha - 1} \log \text{tr}[(\sigma^{\frac{1-\alpha}{2\alpha}} \rho \sigma^{\frac{1-\alpha}{2\alpha}})^\alpha]. \quad (\text{A2})$$

Here and in the whole work, the base of the logarithm is 2. The notation of $\downarrow$ in $\tilde{H}_\alpha^\downarrow(A|B)_\rho$ is used to distinguish it

from another definition of sandwiched conditional Rényi entropy,

$$\tilde{H}_\alpha^\uparrow(A|B)_\rho = -\inf_{\sigma_B} \tilde{D}_\alpha(\rho_{AB} \parallel \mathbb{I}_A \otimes \sigma_B). \quad (\text{A3})$$

Clearly, $\tilde{H}_\alpha^\uparrow(A|B)_\rho \geq \tilde{H}_\alpha^\downarrow(A|B)_\rho$, and these two quantities are different in general. When α takes value ∞ , 2, and $1/2$, we get the conditional min-entropy, the conditional collision entropy, and the conditional max-entropy, respectively, as shown below.

Definition A1. Given a quantum state ρ_{AB} , the conditional min-entropy of A given B is defined as

$$\tilde{H}_{\min}^\downarrow(A|B)_\rho = \sup\{\lambda \in \mathbb{R} : 2^{-\lambda} \cdot \mathbb{I}_A \otimes \rho_B - \rho_{AB} \geq 0\}, \quad (\text{A4})$$

$$\tilde{H}_{\min}^\uparrow(A|B)_\rho = \sup_{\sigma_B} \sup\{\lambda \in \mathbb{R} : 2^{-\lambda} \cdot \mathbb{I}_A \otimes \sigma_B - \rho_{AB} \geq 0\}; \quad (\text{A5})$$

the conditional collision entropy of A given B is defined as

$$\tilde{H}_2^\downarrow(A|B)_\rho = -\log \text{tr}[(\mathbb{I}_A \otimes \rho_B)^{-1/4} \rho_{AB} (\mathbb{I}_A \otimes \rho_B)^{-1/4}]^2, \quad (\text{A6})$$

$$\begin{aligned} \tilde{H}_2^\uparrow(A|B)_\rho = \sup_{\sigma_B} & -\log \text{tr} \left[\left((\mathbb{I}_A \otimes \sigma_B)^{-1/4} \right. \right. \\ & \times \left. \left. \rho_{AB} (\mathbb{I}_A \otimes \sigma_B)^{-1/4} \right)^2 \right]; \end{aligned} \quad (\text{A7})$$

and the conditional max-entropy of A given B is defined as

$$\tilde{H}_{\max}^\downarrow(A|B)_\rho = \log F(\rho_{AB}, \mathbb{I}_A \otimes \rho_B)^2, \quad (\text{A8})$$

$$\tilde{H}_{\max}^\uparrow(A|B)_\rho = \sup_{\sigma_B} \log F(\rho_{AB}, \mathbb{I}_A \otimes \sigma_B)^2. \quad (\text{A9})$$

When α goes to 1, $\tilde{H}_\alpha^{\uparrow/\downarrow}(A|B)_\rho$ will converge to the conditional von Neumann entropy, which is defined as

$$H(A|B) = H(AB) - H(B), \quad (\text{A10})$$

where H is the von Neumann entropy with $H(\rho) = -\text{tr} \rho \log \rho$.

We also consider the smoothing version of the conditional entropy. We call two quantum states ρ and σ δ -close if their purified distance $P(\rho, \sigma) \leq \delta$, and we denote $\mathcal{B}^\delta(\rho) = \{\rho', P(\rho', \rho) \leq \delta\}$ as the set of all δ -close states of ρ . Then, we have the following smooth conditional entropies.

Definition A2. The δ -smooth sandwiched conditional Rényi entropy of A given B for state ρ_{AB} is the maximal sandwiched conditional Rényi entropy of states in

$\mathcal{B}^\delta(\rho_{AB})$:

$$\tilde{H}_\alpha^{\downarrow\delta}(A|B)_\rho = \sup_{\hat{\rho}_{AB} \in \mathcal{B}^\delta(\rho_{AB})} \tilde{H}_\alpha^{\downarrow}(A|B)_{\hat{\rho}}; \quad (\text{A11})$$

$$\tilde{H}_\alpha^{\uparrow\delta}(A|B)_\rho = \sup_{\hat{\rho}_{AB} \in \mathcal{B}^\delta(\rho_{AB})} \tilde{H}_\alpha^{\uparrow}(A|B)_{\hat{\rho}}. \quad (\text{A12})$$

Particularly, the δ -smooth conditional min-entropy of A given B is defined as

$$\tilde{H}_{\min}^{\downarrow\delta}(A|B)_\rho = \sup_{\hat{\rho}_{AB} \in \mathcal{B}^\delta(\rho_{AB})} \tilde{H}_{\min}^{\downarrow}(A|B)_{\hat{\rho}}, \quad (\text{A13})$$

$$\tilde{H}_{\min}^{\uparrow\delta}(A|B)_\rho = \sup_{\hat{\rho}_{AB} \in \mathcal{B}^\delta(\rho_{AB})} \tilde{H}_{\min}^{\uparrow}(A|B)_{\hat{\rho}}; \quad (\text{A14})$$

the δ -smooth conditional collision entropy of A given B is defined as

$$\tilde{H}_2^{\downarrow\delta}(A|B)_\rho = \sup_{\hat{\rho}_{AB} \in \mathcal{B}^\delta(\rho_{AB})} \tilde{H}_2^{\downarrow}(A|B)_{\hat{\rho}}, \quad (\text{A15})$$

$$\tilde{H}_2^{\uparrow\delta}(A|B)_\rho = \sup_{\hat{\rho}_{AB} \in \mathcal{B}^\delta(\rho_{AB})} \tilde{H}_2^{\uparrow}(A|B)_{\hat{\rho}}; \quad (\text{A16})$$

and the δ -smooth conditional max-entropy of A given B is defined as

$$\tilde{H}_{\max}^{\downarrow\delta}(A|B)_\rho = \sup_{\hat{\rho}_{AB} \in \mathcal{B}^\delta(\rho_{AB})} \tilde{H}_{\max}^{\downarrow}(A|B)_{\hat{\rho}}, \quad (\text{A17})$$

$$\tilde{H}_{\max}^{\uparrow\delta}(A|B)_\rho = \sup_{\hat{\rho}_{AB} \in \mathcal{B}^\delta(\rho_{AB})} \tilde{H}_{\max}^{\uparrow}(A|B)_{\hat{\rho}}. \quad (\text{A18})$$

Below, we list several relationships among different entropies.

Lemma A1 (Corollary 4 in Ref. [99]). Given quantum state ρ_{AB} , we have that $\tilde{H}_2^{\downarrow}(A|B)_\rho \geq \tilde{H}_{\min}^{\uparrow}(A|B)_\rho$.

Based on Lemma A1, we have $\tilde{H}_2^{\downarrow\delta}(A|B)_\rho \geq \tilde{H}_{\min}^{\uparrow\delta}(A|B)_\rho$. Meanwhile, since $\tilde{D}$ is monotonically increasing in α [100,101], we have:

Lemma A2. Given quantum state ρ_{AB} , we have that $\tilde{H}_2^{\downarrow}(A|B)_\rho \leq \tilde{H}_{\max}^{\downarrow}(A|B)_\rho \leq \tilde{H}_{\max}^{\uparrow}(A|B)_\rho$.

Thus, $\tilde{H}_2^{\downarrow\delta}(A|B)_\rho \leq \tilde{H}_{\max}^{\uparrow\delta}(A|B)_\rho$. Note that the conditional min-entropy and the conditional max-entropy both have an asymptotic equipartition property [102,103], as shown below.

Lemma A3 (AEP [102,103]). For quantum state ρ_{AB} , we have

$$\lim_{\delta \rightarrow 0} \lim_{n \rightarrow \infty} \frac{1}{n} \tilde{H}_{\min}^{\uparrow\delta}(A|B)_{\rho^{\otimes n}} = H(A|B)_\rho, \quad (\text{A19})$$

$$\lim_{\delta \rightarrow 0} \lim_{n \rightarrow \infty} \frac{1}{n} \tilde{H}_{\max}^{\downarrow\delta}(A|B)_{\rho^{\otimes n}} = H(A|B)_\rho. \quad (\text{A20})$$

Since $\tilde{H}_{\min}^{\uparrow\delta}(A|B)_\rho \leq \tilde{H}_2^{\downarrow\delta}(A|B)_\rho \leq \tilde{H}_{\max}^{\uparrow\delta}(A|B)_\rho$, we have

$$\lim_{\delta \rightarrow 0} \lim_{n \rightarrow \infty} \frac{1}{n} \tilde{H}_2^{\downarrow\delta}(A|B)_{\rho^{\otimes n}} = H(A|B)_\rho. \quad (\text{A21})$$

More exactly, for any $0 < \delta < 1$, as long as $n \geq -\frac{8}{\delta} \log(1 - \sqrt{1 - \delta^2})$, we have that

$$\frac{1}{n} \tilde{H}_{\min}^{\uparrow\delta}(A|B)_{\rho^{\otimes n}} \geq H(A|B)_\rho - \frac{4 \log \gamma \sqrt{\log \frac{2}{\delta^2}}}{\sqrt{n}}, \quad (\text{A22})$$

$$\frac{1}{n} \tilde{H}_{\max}^{\uparrow\delta}(A|B)_{\rho^{\otimes n}} \leq H(A|B)_\rho + \frac{4 \log \gamma \sqrt{\log \frac{2}{\delta^2}}}{\sqrt{n}}, \quad (\text{A23})$$

where

$$\gamma \leq \sqrt{2^{-\tilde{H}_{\min}^{\uparrow\delta}(A|B)_\rho}} + \sqrt{2^{\tilde{H}_{\max}^{\uparrow\delta}(A|B)_\rho}} + 1. \quad (\text{A24})$$

Thus,

$$\begin{aligned} H(A|B)_\rho - \frac{4 \log \gamma \sqrt{\log \frac{2}{\delta^2}}}{\sqrt{n}} \\ \leq \frac{1}{n} \tilde{H}_2^{\downarrow\delta}(A|B)_{\rho^{\otimes n}} \leq H(A|B)_\rho + \frac{4 \log \gamma \sqrt{\log \frac{2}{\delta^2}}}{\sqrt{n}}. \end{aligned} \quad (\text{A25})$$

Below, we present a lemma for calculating quantum conditional entropies with the purification of a quantum state.

Lemma A4 (Dual relations for quantum conditional entropies, Theorem 2 in Ref. [99]). Given tripartite pure state ρ_{ABC} with reduced density matrices on system AB and

AC as ρ_{AB} and ρ_{AC} , respectively, we have that

$$\tilde{H}_\alpha^\downarrow(A|B)_{\rho_{AB}} = -\frac{1}{\alpha-1} \log \text{tr}_C((\text{tr}_A \rho_{AC}^\alpha)^\alpha). \quad (\text{A26})$$

In Ref. [63], the authors prove a decoupling theorem, which describes how the two systems are decoupled after one system is acted on by a random unitary and a channel. This theorem is useful in analyzing the performance of a random quantum error correction scheme, which is shown below.

Lemma A5 (Decoupling theorem [63]). Given quantum state ρ_{SR} and quantum channel $\mathcal{T}_{S \rightarrow E}$, denote the reduced density matrix of ρ_{SR} on system R as ρ_R and the Choi-Jamiołkowski representation of $\mathcal{T}_{S \rightarrow E}$ as τ_{SE} with reduced density matrix on system E as τ_E . We obtain that

$$\begin{aligned} \mathbb{E}_{U \sim \mathcal{U}_n} \|\mathcal{T}_{S \rightarrow E}(U_S \rho_{SR} U_S^\dagger) - \tau_E \otimes \rho_R\|_1 \\ \leq 2^{-\frac{1}{2} \tilde{H}_2^\uparrow(S|R)_{\rho_{SR}} - \frac{1}{2} \tilde{H}_2^\uparrow(S|E)_{\tau_{SE}}} \end{aligned} \quad (\text{A27})$$

and

$$\begin{aligned} \mathbb{E}_{U \sim \mathcal{U}_n} \|\mathcal{T}_{S \rightarrow E}(U_S \rho_{SR} U_S^\dagger) - \tau_E \otimes \rho_R\|_1 \\ \leq 2^{-\frac{1}{2} \tilde{H}_{\min}^\uparrow(S|R)_{\rho_{SR}} - \frac{1}{2} \tilde{H}_{\min}^\uparrow(S|E)_{\tau_{SE}}} + 12\delta, \end{aligned} \quad (\text{A28})$$

where $\mathbb{E}_{U \sim \mathcal{U}_n}$ denotes the expectation over the Haar-random unitary operations. This result also holds for a unitary 2-design group instead of $\mathcal{U}_n$ [64].

The term $\tilde{H}_2^\uparrow(S|R)_{\rho_{SR}}$ characterizes the initial correlation between the system and the reference part, and $\tilde{H}_2^\uparrow(S|E)_{\tau_{SE}}$ characterizes how the channel $\mathcal{T}_{S \rightarrow E}$ preserves the correlation. When applying the decoupling theorem to AQEC, one can view the state ρ_{SR} to be the initial logical state, the unitary operation U_S to be the encoding map, and the channel $\mathcal{T}_{S \rightarrow E}$ to be the complementary channel from system to environment. Then the decoupling error on the right-hand side can be viewed as the Choi error.

If the system is twirled by an ε -approximate 2-design $\mathfrak{S}$ instead of $\mathcal{U}_n$, we also have a result [64].

Lemma A6 (Decoupling theorem for an approximate 2-design [64]). Given quantum state ρ_{SR} and quantum channel $\mathcal{T}_{S \rightarrow E}$, denote the reduced density matrix of ρ_{SR} on system R as ρ_R and the Choi-Jamiołkowski representation of $\mathcal{T}_{S \rightarrow E}$ as τ_{SE} with reduced density matrix on system

E as τ_E . We obtain that

$$\begin{aligned} \mathbb{E}_{U \sim \mathfrak{S}} \|\mathcal{T}_{S \rightarrow E}(U_S \rho_{SR} U_S^\dagger) - \tau_E \otimes \rho_R\|_1 \\ \leq \sqrt{1 + 4\varepsilon |S|^4 2^{-\frac{1}{2} \tilde{H}_{\min}^\uparrow(S|R)_{\rho_{SR}} - \frac{1}{2} \tilde{H}_{\min}^\uparrow(S|E)_{\tau_{SE}}}}, \end{aligned} \quad (\text{A29})$$

where $\mathbb{E}_{U \sim \mathfrak{S}}$ denotes the expectation over $\mathfrak{S}$, and ε is the additive error of the approximate design.

Due to the coefficient $4\varepsilon |S|^4$, for a k -qubit system S , ε needs to be exponentially small with k to make the decoupling error vanish. Thus, an approximate 2-design ensemble with polynomially decreasing error is not guaranteed to have a good performance in quantum error correction.

In this work, for the collision entropy, we will usually consider the case of arrow down. For the min-entropy and the max-entropy, we will usually consider the case of arrow up. For simplicity, below without additional clarification, we denote

$$H_{\min}(A|B)_\rho = \tilde{H}_{\min}^\uparrow(A|B)_\rho, \quad (\text{A30})$$

$$H_2(A|B)_\rho = \tilde{H}_2^\downarrow(A|B)_\rho, \quad (\text{A31})$$

$$H_{\max}(A|B)_\rho = \tilde{H}_{\max}^\uparrow(A|B)_\rho. \quad (\text{A32})$$

The smooth conditional entropies are modified correspondingly.

APPENDIX B: DECOUPLING THEOREM AND AQEC FOR 1D DOUBLE-LAYER BLOCKED LOW-DEPTH CIRCUITS

In this Appendix, we first prove the nonsmooth decoupling theorem or Theorem 1 and then use Theorem 1 to prove the smooth decoupling theorem or Theorem 3. Then, we apply the decoupling theorem for 1D double-layer blocked low-depth circuits to local and correlated noises. After that, we prove the AQEC performance against the fixed-number random erasure errors for 1D low-depth circuits. Subsequently, we give the proof of Proposition 1. We also present the AQEC results with random Clifford encoding and block-encoding methods as a comparison.

1. Proof of nonsmooth decoupling theorem or Theorem 1

Proof. Based on Lemma 1, for any quantum states $\sigma_E \in \mathcal{D}(\mathcal{H}_E)$ and $\zeta_R \in \mathcal{D}(\mathcal{H}_R)$,

$$\|\mathcal{T}_{S \rightarrow E}(U_S \rho_{SR} U_S^\dagger) - \tau_E \otimes \rho_R\|_1 \leq \sqrt{\text{tr}[(\sigma_E \otimes \zeta_R)^{-1/4} (\mathcal{T}_{S \rightarrow E}(U_S \rho_{SR} U_S^\dagger) - \tau_E \otimes \rho_R) (\sigma_E \otimes \zeta_R)^{-1/4}]^2}. \quad (\text{B1})$$

We define completely positive map $\tilde{\mathcal{T}}_{S \rightarrow E} = \sigma_E^{-1/4} \mathcal{T}_{S \rightarrow E} \sigma_E^{-1/4}$ with the corresponding Choi-Jomiołkowski representation $\tilde{\tau}_{SE} = \sigma_E^{-1/4} \tau_{SE} \sigma_E^{-1/4}$ and $\tilde{\rho}_{SR} = \zeta_R^{-1/4} \rho_{SR} \zeta_R^{-1/4}$. Recall that $S = S_1 S_2 \cdots S_{2N}$ and $R = R_1 R_2 \cdots R_{2N}$. In the following, we may use $A \subseteq [2N]$ or S_A to represent the subsystem $\bigcup_{i \in A} S_i$. Also, ρ_{SR} has a structure of tensor product, $\rho_{SR} = \bigotimes_{i=1}^{2N} \rho_{S_i R_i}$. Below, we let ζ_R also have a tensor-product structure,

$$\zeta_R = \bigotimes_{i=1}^{2N} \zeta_{R_i}. \quad (\text{B2})$$

Thanks to the tensor-product structure of ρ_{SR} and ζ_R , we have that

$$\tilde{\rho}_{SR} = \bigotimes_{i=1}^{2N} \tilde{\rho}_{S_i R_i}, \quad (\text{B3})$$

where $\tilde{\rho}_{S_i R_i} = \zeta_{R_i}^{-1/4} \rho_{S_i R_i} \zeta_{R_i}^{-1/4}$.

We then get

$$\begin{aligned} & \|\mathcal{T}_{S \rightarrow E}(U_S \rho_{SR} U_S^\dagger) - \tau_E \otimes \rho_R\|_1 \\ & \leq \sqrt{\text{tr} \left[\left(\tilde{\mathcal{T}}_{S \rightarrow E}(U_S \tilde{\rho}_{SR} U_S^\dagger) - \tilde{\tau}_E \otimes \tilde{\rho}_R \right)^2 \right]}. \end{aligned} \quad (\text{B4})$$

Using Jensen's inequality, we have that

$$\begin{aligned} & \mathbb{E}_U \|\mathcal{T}_{S \rightarrow E}(U_S \rho_{SR} U_S^\dagger) - \tau_E \otimes \rho_R\|_1 \\ & \leq \sqrt{\mathbb{E}_U \text{tr} \left[\left(\tilde{\mathcal{T}}_{S \rightarrow E}(U_S \tilde{\rho}_{SR} U_S^\dagger) - \tilde{\tau}_E \otimes \tilde{\rho}_R \right)^2 \right]}. \end{aligned} \quad (\text{B5})$$

Below, we focus on the term inside the square root,

$$\begin{aligned} & \mathbb{E}_U \text{tr} \left[\left(\tilde{\mathcal{T}}_{S \rightarrow E}(U_S \tilde{\rho}_{SR} U_S^\dagger) - \tilde{\tau}_E \otimes \tilde{\rho}_R \right)^2 \right] \\ & = \mathbb{E}_U \text{tr} \left[\left(\tilde{\mathcal{T}}_{S \rightarrow E}(U_S \tilde{\rho}_{SR} U_S^\dagger) \right)^2 \right] + \text{tr}[(\tilde{\tau}_E \otimes \tilde{\rho}_R)^2] \\ & \quad - 2 \mathbb{E}_U \text{tr}[\tilde{\mathcal{T}}_{S \rightarrow E}(U_S \tilde{\rho}_{SR} U_S^\dagger)(\tilde{\tau}_E \otimes \tilde{\rho}_R)]. \end{aligned} \quad (\text{B6})$$

Since U is randomly sampled from $\mathcal{C}_n^e$, which forms a unitary 1-design, we have

$$\mathbb{E}_U \tilde{\mathcal{T}}_{S \rightarrow E}(U_S \tilde{\rho}_{SR} U_S^\dagger) = \tilde{\tau}_E \otimes \tilde{\rho}_R \quad (\text{B7})$$

and

$$\begin{aligned} & \text{tr}[(\tilde{\tau}_E \otimes \tilde{\rho}_R)^2] - 2 \mathbb{E}_U \text{tr}[\tilde{\mathcal{T}}_{S \rightarrow E}(U_S \tilde{\rho}_{SR} U_S^\dagger)(\tilde{\tau}_E \otimes \tilde{\rho}_R)] \\ & = -\text{tr}(\tilde{\tau}_E^2) \text{tr}(\tilde{\rho}_R^2). \end{aligned} \quad (\text{B8})$$

For the remaining part $\mathbb{E}_U \text{tr} \left[\left(\tilde{\mathcal{T}}_{S \rightarrow E}(U_S \tilde{\rho}_{SR} U_S^\dagger) \right)^2 \right]$, we use $\text{tr} A^2 = \text{tr} A^{\otimes 2} F$ to evaluate it with F a SWAP operator

and separate the two layers of the encoding unitary gate. Then,

$$\begin{aligned} & \mathbb{E}_U \text{tr} \left[\left(\tilde{\mathcal{T}}_{S \rightarrow E}(U_S \tilde{\rho}_{SR} U_S^\dagger) \right)^2 \right] \\ & = \text{tr}_S(\mathbb{E}_{U_2} [U_2^{\otimes 2} \mathbb{E}_{U_1} [U_1^{\otimes 2} \text{tr}_R(\tilde{\rho}_{SR}^{\otimes 2} F_R) U_1^{\dagger \otimes 2}] U_2^{\dagger \otimes 2}] \\ & \quad \times (\tilde{\mathcal{T}}_{S \rightarrow E}^\dagger)^{\otimes 2}(F_E)) \\ & = \text{tr}_S(M_2 M_1 (\text{tr}_R(\tilde{\rho}_{SR}^{\otimes 2} F_R)) (\tilde{\mathcal{T}}_{S \rightarrow E}^\dagger)^{\otimes 2}(F_E)), \end{aligned} \quad (\text{B9})$$

where U_1 and U_2 are the first and second layers of the encoding unitary U , respectively, and $\tilde{\mathcal{T}}_{S \rightarrow E}^\dagger$ is the adjoint channel of $\tilde{\mathcal{T}}_{S \rightarrow E}$. We also define $M_i = \mathbb{E}_{U_i} U_i^{\otimes 2} \cdot U_i^{\dagger \otimes 2}$ as the second-order twirling operation of layer i . Thus, the key technical part is to evaluate the action of $M_2 M_1$.

Note that when $\xi = 1$, U_i is generated by two-qubit gates. In this case, $(M_2 M_1)^l$ simulates $2l$ -depth 1D brickwork circuits [32]. In this part, we focus mainly on the double-layer blocked structure and $l = 1$, which enables a simpler evaluation. The case of 1D brickwork circuits can be analyzed using the domain-wall approach [32], which we will discuss in Appendix C.

In the following, for simplicity, we set $O_{SR} = \text{tr}_R(\tilde{\rho}_{SR}^{\otimes 2} F_R)$ and $O_{SE} = (\tilde{\mathcal{T}}_{S \rightarrow E}^\dagger)^{\otimes 2}(F_E)$. We also denote $O_{S_i R_i} = \text{tr}_{R_i}(\tilde{\rho}_{S_i R_i}^{\otimes 2} F_{R_i})$. Thus, $O_{SR} = \bigotimes_i O_{S_i R_i}$. Our task is to evaluate $\text{tr}(M_2 M_1 (O_{SR}) O_{SE})$.

Note that $S = S_1 S_2 \cdots S_{2N}$ and $R = R_1 R_2 \cdots R_{2N}$. Based on Eq. (51) and the tensor-product structures of U_1 , we have that

$$\begin{aligned} M_1(O_{SR}) & = \mathbb{E}_{U_1} [U_1^{\otimes 2} O_{SR} U_1^{\dagger \otimes 2}] \\ & = \frac{1}{[(2^{2\xi})^2 - 1]^N} \sum_{\vec{\gamma} \in \{\mathbb{I} \otimes \mathbb{I}, F \otimes F\}^{\otimes N}} \text{tr}(O_{SR} g_{\vec{\gamma}}) \vec{\gamma} \\ & = \sum_{\vec{\gamma} \in \{\mathbb{I} \otimes \mathbb{I}, F \otimes F\}^{\otimes N}} c_{\vec{\gamma}} \vec{\gamma}. \end{aligned} \quad (\text{B10})$$

Here, $\mathbb{I}$ is the identity operation on two copies of a subsystem i , and F is the SWAP operation on two copies of a subsystem i ; $\vec{\gamma}$ can be viewed as a vector with $2N$ components. The j th component of $g_{\vec{\gamma}}$, denoted as g_{γ_j} satisfies: $g_{\gamma_{2j-1}} \otimes g_{\gamma_{2j}} = \mathbb{I} \otimes \mathbb{I} - 2^{-2\xi} F \otimes F$ if $\gamma_{2j-1} \otimes \gamma_{2j} = \mathbb{I} \otimes \mathbb{I}$, and $g_{\gamma_{2j-1}} \otimes g_{\gamma_{2j}} = F \otimes F - 2^{-2\xi} \mathbb{I} \otimes \mathbb{I}$ if $g_{\gamma_{2j-1}} \otimes g_{\gamma_{2j}} = F \otimes F$ where γ_j is the j th component of $\vec{\gamma}$. Using $O_{SR} = \bigotimes_i O_{S_i R_i}$ and Lemma B1 below, it is straightforward to upper bound the coefficient $c_{\vec{\gamma}}$,

$$\begin{aligned} c_{\vec{\gamma}} & = \frac{1}{[(2^{2\xi})^2 - 1]^N} \text{tr}(O_{SR} g_{\vec{\gamma}}) \leq (2^{-2\xi})^{2N} \text{tr}(O_{SR} \vec{\gamma}) \\ & = 2^{-2n} \text{tr}(O_{SR} \vec{\gamma}). \end{aligned} \quad (\text{B11})$$

Lemma B1 (Lemma 3.6 in Ref. [63]). Given a Hilbert space $\mathcal{H}_{AB}$, let $\tilde{\rho}_{AB}$ be a non-negative operation on $\mathcal{H}_{AB}$. We have that

$$\frac{1}{|A|} \leq \frac{\text{tr}(\tilde{\rho}_{AB}^2)}{\text{tr}(\tilde{\rho}_B^2)} \leq |A|. \quad (\text{B12})$$

The proof of Eq. (B11) is mainly based on

$$\begin{aligned} \text{tr}(O_{S_{2i-1}R_{2i-1}} \otimes O_{S_{2i}R_{2i}}(\mathbb{I} \otimes \mathbb{I} - 2^{-2\xi} F \otimes F)) &= \text{tr}(\tilde{\rho}_{R_{2i-1}}^2 \tilde{\rho}_{R_{2i}}^2) - 2^{-2\xi} \text{tr}(\tilde{\rho}_{S_{2i-1}R_{2i-1}}^2 \tilde{\rho}_{S_{2i}R_{2i}}^2) \\ &\leq (1 - (2^{-2\xi})^2) \text{tr}(\tilde{\rho}_{R_{2i-1}}^2 \tilde{\rho}_{R_{2i}}^2), \\ \text{tr}(O_{S_{2i-1}R_{2i-1}} \otimes O_{S_{2i}R_{2i}}(F \otimes F - 2^{-2\xi} \mathbb{I} \otimes \mathbb{I})) &= \text{tr}(\tilde{\rho}_{S_{2i-1}R_{2i-1}}^2 \tilde{\rho}_{S_{2i}R_{2i}}^2) - 2^{-2\xi} \text{tr}(\tilde{\rho}_{R_{2i-1}}^2 \tilde{\rho}_{R_{2i}}^2) \\ &\leq (1 - (2^{-2\xi})^2) \text{tr}(\tilde{\rho}_{S_{2i-1}R_{2i-1}}^2 \tilde{\rho}_{S_{2i}R_{2i}}^2). \end{aligned} \quad (\text{B13})$$

where the inequalities are based on Lemma B1.

The next step is to evaluate $M_2 M_1(O_{\text{SR}}) = \sum_{\vec{\gamma}} c_{\vec{\gamma}} M_2(\vec{\gamma})$. Based on the tensor-product structure of U_2 , we can analyze the action of each block part of M_2 on $\vec{\gamma}$. Since $U_2 = \bigotimes_{i=1}^{N-1} U_C^{2i,2i+1}$, M_2 can be written as $\bigotimes_{i=1}^{N-1} M^{2i,2i+1}$ where $M^{2i,2i+1}$ is the second-order twirling operation on subsystems $S_{2i}S_{2i+1}$. Based on Eq. (51), the action of $M^{2i,2i+1}$ on $\{\mathbb{I}, F\}^{\otimes 2}$ is

$$M^{2i,2i+1}(\mathbb{I}_{2i} \otimes \mathbb{I}_{2i+1}) = \mathbb{I}_{2i} \otimes \mathbb{I}_{2i+1}, \quad (\text{B14})$$

$$M^{2i,2i+1}(F_{2i} \otimes F_{2i+1}) = F_{2i} \otimes F_{2i+1}, \quad (\text{B15})$$

$$M^{2i,2i+1}(\mathbb{I}_{2i} \otimes F_{2i+1}) = \eta(\mathbb{I}_{2i} \otimes \mathbb{I}_{2i+1} + F_{2i} \otimes F_{2i+1}), \quad (\text{B16})$$

$$M^{2i,2i+1}(F_{2i} \otimes \mathbb{I}_{2i+1}) = \eta(\mathbb{I}_{2i} \otimes \mathbb{I}_{2i+1} + F_{2i} \otimes F_{2i+1}). \quad (\text{B17})$$

Here, $\eta = 2^\xi / (2^{2\xi} + 1)$. Note that only if $\gamma_{2i} \neq \gamma_{2i+1}$, $M^{2i,2i+1}$ can alter $\vec{\gamma}$.

Note that $\vec{\gamma} \in \{\mathbb{I} \otimes \mathbb{I}, F \otimes F\}^{\otimes N}$. We always have $\gamma_{2i-1} = \gamma_{2i}$. Given a vector $\vec{\gamma} \in \{\mathbb{I} \otimes \mathbb{I}, F \otimes F\}^{\otimes N}$, we count all indexes j such that $\gamma_{2j} \neq \gamma_{2j+1}$ and denote this set as $\mathcal{G} = \{j_1, j_2, \dots, j_g\}$. Obviously, $0 \leq g \leq N-1$, and there is $\binom{N-1}{g}$ choices of $\mathcal{G}$ after fixing g . We also denote the index set $\mathcal{A} = \{2j, 2j+1 | j \in \mathcal{G}\}$. We also define $S_{\mathcal{A}} = \bigcup_{i \in \mathcal{A}} S_i$ and $S_{\bar{\mathcal{A}}} = \bigcup_{i \in \bar{\mathcal{A}}} S_i$. Then, M_2 only alters $\vec{\gamma}$ within the support of $S_{\mathcal{A}}$ and leaves the remained part unchanged. We use $\gamma_{\bar{\mathcal{A}}}$ to denote this unchanged part. Then,

$$M_2(\vec{\gamma}) = \eta^g \gamma_{\bar{\mathcal{A}}} \otimes \bigotimes_{j \in \mathcal{G}} (\mathbb{I}_{2j} \otimes \mathbb{I}_{2j+1} + F_{2j} \otimes F_{2j+1}). \quad (\text{B18})$$

Thus, the final target can be evaluated by

$$\begin{aligned} \text{tr}(M_2 M_1(O_{\text{SR}}) O_{\text{SE}}) &= \sum_{\vec{\gamma} \in \{\mathbb{I} \otimes \mathbb{I}, F \otimes F\}^{\otimes N}} c_{\vec{\gamma}} \text{tr}(M_2(\vec{\gamma}) O_{\text{SE}}) \\ &= \sum_{\gamma_1 = \mathbb{I}, F} \sum_{g=0}^{N-1} \sum_{\mathcal{G}} c_{\vec{\gamma}} \text{tr}(M_2(\vec{\gamma}) O_{\text{SE}}) \\ &= \sum_{\gamma_1 = \mathbb{I}, F} \sum_{g=0}^{N-1} \eta^g \sum_{\mathcal{G}} c_{\vec{\gamma}} \text{tr} \left(\gamma_{\bar{\mathcal{A}}} \otimes \bigotimes_{j \in \mathcal{G}} (\mathbb{I}_{2j} \otimes \mathbb{I}_{2j+1} + F_{2j} \otimes F_{2j+1}) O_{\text{SE}} \right). \end{aligned} \quad (\text{B19})$$

Note that given a vector $\vec{\gamma} = F_{\mathcal{A}} \otimes \mathbb{I}_{\bar{\mathcal{A}}}$, we always have

$$\begin{aligned} \text{tr}(\vec{\gamma} O_{\text{SE}}) &= \text{tr}(\tilde{\mathcal{T}}_{S \rightarrow E}^{\otimes 2}(\vec{\gamma}) F_E) \\ &= 2^{2n} \text{tr}(\tilde{\tau}_{\text{SE}}^{\otimes 2} F_{\mathcal{A}} \otimes \mathbb{I}_{\bar{\mathcal{A}}} \otimes F_E) \\ &= 2^{2n} \text{tr}_{\text{AE}}(\text{tr}_{\bar{\mathcal{A}}} \tilde{\tau}_{\text{SE}})^2 \\ &\geq 0. \end{aligned} \quad (\text{B20})$$

Thus, using Eq. (B11), we have

$$\begin{aligned}
\text{tr}(M_2 M_1 (O_{\text{SR}}) O_{\text{SE}}) &\leq \sum_{\gamma_1=\mathbb{I}, F}^{N-1} \sum_{g=0} \eta^g \sum_{\mathcal{G}} 2^{-2n} \text{tr}(O_{\text{SR}} \vec{\gamma}) \text{tr} \left(\gamma_{\vec{A}} \otimes \bigotimes_{j \in \mathcal{G}} (\mathbb{I}_{2j} \otimes \mathbb{I}_{2j+1} + F_{2j} \otimes F_{2j+1}) O_{\text{SE}} \right) \\
&\leq 2^{-2n} (\text{tr}(O_{\text{SR}} \mathbb{I}^{\otimes 2N}) \text{tr}(O_{\text{SE}} \mathbb{I}^{\otimes 2N}) + \text{tr}(O_{\text{SR}} F^{\otimes 2N}) \text{tr}(O_{\text{SE}} F^{\otimes 2N})) \\
&\quad + \sum_{\gamma_1=\mathbb{I}, F}^{N-1} \sum_{g=1} (2\eta)^g \sum_{\mathcal{G}} 2^{-2n} \max_{\gamma'_{\vec{A}}=\gamma_{\vec{A}}, \gamma'_{\vec{A}} \in \{\mathbb{I} \otimes \mathbb{I}, F \otimes F\}^{\otimes g}} \text{tr}(O_{\text{SR}} \gamma_{\vec{A}} \otimes \gamma'_{\vec{A}}) \text{tr}(O_{\text{SE}} \gamma'_{\vec{A}} \otimes \gamma'_{\vec{A}}) \\
&\leq \text{tr}(\tilde{\tau}_E^2) \text{tr}(\tilde{\rho}_R^2) + \text{tr}(\tilde{\tau}_{\text{SE}}^2) \text{tr}(\tilde{\rho}_{\text{SR}}^2) \\
&\quad + \sum_{\gamma_1=\mathbb{I}, F}^{N-1} \sum_{g=1} (2\eta)^g \sum_{\mathcal{G}} 2^{-2n} \max_{\gamma'_{\vec{A}}=\gamma_{\vec{A}}, \gamma'_{\vec{A}} \in \{\mathbb{I} \otimes \mathbb{I}, F \otimes F\}^{\otimes g}} \text{tr}(O_{\text{SR}} \vec{\gamma}') \text{tr}(O_{\text{SE}} \vec{\gamma}') \frac{\text{tr}(O_{S_{\vec{A}} R_{\vec{A}}} \gamma_{\vec{A}})}{\text{tr}(O_{S_{\vec{A}} R_{\vec{A}}} \gamma'_{\vec{A}})} \\
&\leq \text{tr}(\tilde{\tau}_E^2) \text{tr}(\tilde{\rho}_R^2) + \text{tr}(\tilde{\tau}_{\text{SE}}^2) \text{tr}(\tilde{\rho}_{\text{SR}}^2) \\
&\quad + \sum_{\gamma_1=\mathbb{I}, F}^{N-1} \sum_{g=1} (2\eta)^g \sum_{\mathcal{G}} 2^{-2n} \max_{\vec{\gamma} \in \{\mathbb{I}, F\}^{\otimes 2N}} (\text{tr}(O_{\text{SR}} \vec{\gamma}) \text{tr}(O_{\text{SE}} \vec{\gamma})) \left(\max_i \left(\frac{\text{tr}(O_{S_i R_i} \mathbb{I}_i)}{\text{tr}(O_{S_i R_i} F_i)}, \frac{\text{tr}(O_{S_i R_i} F_i)}{\text{tr}(O_{S_i R_i} \mathbb{I}_i)} \right) \right)^g \\
&\leq \text{tr}(\tilde{\tau}_E^2) \text{tr}(\tilde{\rho}_R^2) + \text{tr}(\tilde{\tau}_{\text{SE}}^2) \text{tr}(\tilde{\rho}_{\text{SR}}^2) \\
&\quad + 2 \sum_{g=1}^{N-1} \left(2\eta \max_i \left(\frac{\text{tr}(O_{S_i R_i} \mathbb{I}_i)}{\text{tr}(O_{S_i R_i} F_i)}, \frac{\text{tr}(O_{S_i R_i} F_i)}{\text{tr}(O_{S_i R_i} \mathbb{I}_i)} \right) \right)^g \binom{N-1}{g} \max_{\vec{\gamma}} (2^{-2n} \text{tr}(O_{\text{SR}} \vec{\gamma}) \text{tr}(O_{\text{SE}} \vec{\gamma})) \\
&= \text{tr}(\tilde{\tau}_E^2) \text{tr}(\tilde{\rho}_R^2) + \text{tr}(\tilde{\tau}_{\text{SE}}^2) \text{tr}(\tilde{\rho}_{\text{SR}}^2) \\
&\quad + 2 \sum_{g=1}^{N-1} \left(2\eta \max_i \left(\frac{\text{tr}(\tilde{\rho}_{R_i}^2)}{\text{tr}(\tilde{\rho}_{S_i R_i}^2)}, \frac{\text{tr}(\tilde{\rho}_{S_i R_i}^2)}{\text{tr}(\tilde{\rho}_{R_i}^2)} \right) \right)^g \binom{N-1}{g} \max_{A \subseteq [2N]} (\text{tr}_{\text{AR}}(\text{tr}_{\vec{A}} \tilde{\rho}_{\text{SR}})^2 \text{tr}_{\text{AE}}(\text{tr}_{\vec{A}} \tilde{\tau}_{\text{SE}})^2) \\
&= \text{tr}(\tilde{\tau}_E^2) \text{tr}(\tilde{\rho}_R^2) + \text{tr}(\tilde{\tau}_{\text{SE}}^2) \text{tr}(\tilde{\rho}_{\text{SR}}^2) + c \max_{A \subseteq [2N]} (\text{tr}_{\text{AR}}(\text{tr}_{\vec{A}} \tilde{\rho}_{\text{SR}})^2 \text{tr}_{\text{AE}}(\text{tr}_{\vec{A}} \tilde{\tau}_{\text{SE}})^2), \tag{B21}
\end{aligned}$$

where

$$c = 2 \left(\left(1 + 2\eta \max_i \left(\max \left(\frac{\text{tr}(\tilde{\rho}_{R_i}^2)}{\text{tr}(\tilde{\rho}_{S_i R_i}^2)}, \frac{\text{tr}(\tilde{\rho}_{S_i R_i}^2)}{\text{tr}(\tilde{\rho}_{R_i}^2)} \right) \right) \right)^{N-1} - 1 \right). \tag{B22}$$

Consequently,

$$\begin{aligned}
&\mathbb{E}_U \| \mathcal{T}_{S \rightarrow E}(U_S \rho_{\text{SR}} U_S^\dagger) - \tau_E \otimes \rho_R \|_1 \\
&\leq \sqrt{\mathbb{E}_U \text{tr} \left[\left(\tilde{\mathcal{T}}_{S \rightarrow E}(U_S \tilde{\rho}_{\text{SR}} U_S^\dagger) - \tilde{\tau}_E \otimes \tilde{\rho}_R \right)^2 \right]} \\
&\leq \sqrt{-\text{tr}(\tilde{\tau}_E^2) \text{tr}(\tilde{\rho}_R^2) + \text{tr}(\tilde{\tau}_E^2) \text{tr}(\tilde{\rho}_R^2) + \text{tr}(\tilde{\tau}_{\text{SE}}^2) \text{tr}(\tilde{\rho}_{\text{SR}}^2) + c \max_{A \subseteq [2N]} (\text{tr}_{\text{AR}}(\text{tr}_{\vec{A}} \tilde{\rho}_{\text{SR}})^2 \text{tr}_{\text{AE}}(\text{tr}_{\vec{A}} \tilde{\tau}_{\text{SE}})^2)} \\
&\leq \sqrt{\text{tr}(\tilde{\tau}_{\text{SE}}^2) \text{tr}(\tilde{\rho}_{\text{SR}}^2) + c \max_{A \subseteq [2N]} (\text{tr}_{\text{AR}}(\text{tr}_{\vec{A}} \tilde{\rho}_{\text{SR}})^2 \text{tr}_{\text{AE}}(\text{tr}_{\vec{A}} \tilde{\tau}_{\text{SE}})^2)}. \tag{B23}
\end{aligned}$$

Note that $\tilde{\rho}_{\text{SR}} = \bigotimes_{i=1}^{2N} \tilde{\rho}_{S_i R_i}$ and $\tilde{\rho}_{S_i R_i} = \zeta_{R_i}^{-1/4} \rho_{S_i R_i} \zeta_{R_i}^{-1/4}$. We are free to choose any state as ζ_{R_i} . By the definition of conditional collision entropy in Definition A1, we can choose $\zeta_{R_i} = \rho_{R_i}$ or $\zeta_R = \rho_R$ to let

$$\text{tr}(\tilde{\rho}_{R_i}^2) = 1, \text{tr}(\tilde{\rho}_{S_i R_i}^2) = 2^{-H_2(S_i | R_i) \rho_{S_i R_i}}. \tag{B24}$$

Thus,

$$\text{tr}_{\text{AR}}(\text{tr}_{\tilde{A}} \tilde{\rho}_{\text{SR}})^2 = \text{tr}_{\text{AR}}(\tilde{\rho}_{\text{AR}}^2) = 2^{-H_2(A|R)\rho_{\text{AR}}}. \quad (\text{B25})$$

Also, we can choose $\sigma_E = \tau_E$ such that

$$\text{tr}(\tilde{\tau}_E^2) = 1, \text{tr}(\tilde{\tau}_{\text{SE}}^2) = 2^{-H_2(S|E)\tau_{\text{SE}}}. \quad (\text{B26})$$

In this case,

$$\text{tr}_{\text{AE}}(\text{tr}_{\tilde{A}} \tilde{\tau}_{\text{SE}})^2 = \text{tr}_{\text{AE}}(\tilde{\tau}_{\text{AE}}^2) = 2^{-H_2(A|E)\tau_{\text{AE}}}. \quad (\text{B27})$$

Finally, we get

$$\begin{aligned} & \mathbb{E}_U \| \mathcal{T}_{S \rightarrow E}(U_S \rho_{\text{SR}} U_S^\dagger) - \tau_E \otimes \rho_R \|_1 \\ & \leq \sqrt{2^{-H_2(S|E)\tau_{\text{SE}}} 2^{-H_2(S|R)\rho_{\text{SR}}} + c \max_{A \subseteq [2N]} 2^{-H_2(A|R)\rho_{\text{AR}}} 2^{-H_2(A|E)\tau_{\text{AE}}}}, \end{aligned} \quad (\text{B28})$$

where

$$c = 2 \left((1 + 2\eta \max_i (\max(2^{-H_2(S_i|R_i)\rho_{S_i R_i}}, 2^{H_2(S_i|R_i)\rho_{S_i R_i}}))^{N-1} - 1 \right). \quad (\text{B29})$$

Furthermore, if $\mathcal{T}_{S \rightarrow E}$ has a tensor-product structure, $\mathcal{T}_{S \rightarrow E} = \bigotimes_{i=1}^{2N} \mathcal{T}_{S_i \rightarrow E_i}$, then $\tilde{\tau}_{\text{SE}} = \bigotimes_{i=1}^{2N} \tilde{\tau}_{S_i E_i}$ with $\tilde{\tau}_{S_i E_i} = \tau_{E_i}^{-1/4} \tau_{S_i E_i} \tau_{E_i}^{-1/4}$. In this case,

$$H_2(A|R)\rho_{\text{AR}} = \sum_{i \in A} H_2(S_i|R_i)\rho_{S_i R_i}, \quad (\text{B30})$$

$$H_2(A|E)\tau_{\text{AE}} = \sum_{i \in A} H_2(S_i|E_i)\tau_{S_i E_i}. \quad (\text{B31})$$

Thus, we get the following decoupling result:

$$\begin{aligned} & \mathbb{E}_U \| \mathcal{T}_{S \rightarrow E}(U_S \rho_{\text{SR}} U_S^\dagger) - \tau_E \otimes \rho_R \|_1 \\ & \leq \sqrt{2^{-\sum_{i=1}^{2N} H_2(S_i|E_i)\tau_{S_i E_i} - \sum_{i=1}^{2N} H_2(S_i|R_i)\rho_{S_i R_i}} + c \prod_{i=1}^{2N} \max(1, 2^{-H_2(S_i|E_i)\tau_{S_i E_i} - H_2(S_i|R_i)\rho_{S_i R_i}})}. \end{aligned} \quad (\text{B32})$$

Proof is completed. ■

2. Proof of smooth decoupling theorem or Theorem 3

Proof. Theorem 1 already tells us that

$$\begin{aligned} & \mathbb{E}_U \| \mathcal{T}_{S \rightarrow E}(U_S \rho_{\text{SR}} U_S^\dagger) - \tau_E \otimes \rho_R \|_1 \\ & \leq \sqrt{2^{-\sum_{i=1}^{2N} H_2(S_i|E_i)\tau_{S_i E_i} - \sum_{i=1}^{2N} H_2(S_i|R_i)\rho_{S_i R_i}} + c \prod_{i=1}^{2N} \max(1, 2^{-H_2(S_i|E_i)\tau_{S_i E_i} - H_2(S_i|R_i)\rho_{S_i R_i}})}, \end{aligned} \quad (\text{B33})$$

where

$$c = 2 \left(\left(1 + 2\eta \max_i (\max(2^{-H_2(S_i|R_i)\rho_{S_i R_i}}, 2^{H_2(S_i|R_i)\rho_{S_i R_i}}))^{N-1} - 1 \right) \right). \quad (\text{B34})$$

The remaining part considers the smoothing of the conditional collision entropy. Take $\hat{\rho}_{S_i R_i} \in \mathcal{B}^\delta(\rho_{S_i R_i})$ such that $H_2^\delta(S_i|R_i)\rho_{S_i R_i} = H_2(S_i|R_i)\hat{\rho}_{S_i R_i}$ and $\hat{\tau}_{S_i E_i} \in \mathcal{B}^\delta(\tau_{S_i E_i})$ such that $H_2^\delta(S_i|E_i)\tau_{S_i E_i} = H_2(S_i|E_i)\hat{\tau}_{S_i E_i}$. By Eq. (15), we have

$\|\rho_{S_i R_i} - \hat{\rho}_{S_i R_i}\|_1 \leq 2\delta$ and $\|\tau_{S_i E_i} - \hat{\tau}_{S_i E_i}\|_1 \leq 2\delta$. Thus,

$$\begin{aligned}
& \sqrt{2^{-\sum_{i=1}^{2N} H_2^\delta(S_i|E_i)\tau_{S_i E_i} - \sum_{i=1}^{2N} H_2^\delta(S_i|R_i)\rho_{S_i R_i}} + c' \prod_{i=1}^{2N} \max(1, 2^{-H_2^\delta(S_i|E_i)\tau_{S_i E_i} - H_2^\delta(S_i|R_i)\rho_{S_i R_i}})} \\
&= \sqrt{2^{-\sum_{i=1}^{2N} H_2(S_i|E_i)\hat{\tau}_{S_i E_i} - \sum_{i=1}^{2N} H_2(S_i|R_i)\hat{\rho}_{S_i R_i}} + c' \prod_{i=1}^{2N} \max(1, 2^{-H_2(S_i|E_i)\hat{\tau}_{S_i E_i} - H_2(S_i|R_i)\hat{\rho}_{S_i R_i}})} \\
&\geq \mathbb{E}_U \|\hat{\mathcal{T}}_{S \rightarrow E}(U_S \hat{\rho}_{\text{SR}} U_S^\dagger) - \hat{\tau}_E \otimes \hat{\rho}_R\|_1 \\
&\geq \mathbb{E}_U \|\hat{\mathcal{T}}_{S \rightarrow E}(U_S \hat{\rho}_{\text{SR}} U_S^\dagger) - \tau_E \otimes \rho_R\|_1 - \|\tau_E \otimes \rho_R - \hat{\tau}_E \otimes \hat{\rho}_R\|_1 \\
&\geq \mathbb{E}_U \|\hat{\mathcal{T}}_{S \rightarrow E}(U_S \hat{\rho}_{\text{SR}} U_S^\dagger) - \tau_E \otimes \rho_R\|_1 - \sum_{i=1}^{2N} \|\tau_{E_i} \otimes \rho_{R_i} - \hat{\tau}_{E_i} \otimes \hat{\rho}_{R_i}\|_1 \\
&\geq \mathbb{E}_U \|\hat{\mathcal{T}}_{S \rightarrow E}(U_S \hat{\rho}_{\text{SR}} U_S^\dagger) - \tau_E \otimes \rho_R\|_1 - \sum_{i=1}^{2N} (\|\tau_{E_i} - \hat{\tau}_{E_i}\|_1 + \|\rho_{R_i} - \hat{\rho}_{R_i}\|_1) \\
&\geq \mathbb{E}_U \|\hat{\mathcal{T}}_{S \rightarrow E}(U_S \hat{\rho}_{\text{SR}} U_S^\dagger) - \tau_E \otimes \rho_R\|_1 - \sum_{i=1}^{2N} (\|\tau_{S_i E_i} - \hat{\tau}_{S_i E_i}\|_1 + \|\rho_{S_i R_i} - \hat{\rho}_{S_i R_i}\|_1) \\
&\geq \mathbb{E}_U \|\hat{\mathcal{T}}_{S \rightarrow E}(U_S \hat{\rho}_{\text{SR}} U_S^\dagger) - \tau_E \otimes \rho_R\|_1 - 8N\delta \\
&\geq \mathbb{E}_U \|\mathcal{T}_{S \rightarrow E}(U_S \rho_{\text{SR}} U_S^\dagger) - \tau_E \otimes \rho_R\|_1 - \mathbb{E}_U \|\hat{\mathcal{T}}_{S \rightarrow E}(U_S \hat{\rho}_{\text{SR}} U_S^\dagger) - \hat{\mathcal{T}}_{S \rightarrow E}(U_S \rho_{\text{SR}} U_S^\dagger)\|_1 \\
&\quad - \mathbb{E}_U \|\hat{\mathcal{T}}_{S \rightarrow E}(U_S \rho_{\text{SR}} U_S^\dagger) - \mathcal{T}_{S \rightarrow E}(U_S \rho_{\text{SR}} U_S^\dagger)\|_1 - 8N\delta, \tag{B35}
\end{aligned}$$

where the coefficient c' is

$$\begin{aligned}
c' &= 2 \left(\left(1 + 2\eta \max_i (\max(2^{-H_2(S_i|R_i)\hat{\rho}_{S_i R_i}}, 2^{H_2(S_i|R_i)\hat{\rho}_{S_i R_i}})) \right)^{N-1} - 1 \right) \\
&= 2 \left(\left(1 + 2\eta \max_i (\max(2^{-H_2^\delta(S_i|R_i)\rho_{S_i R_i}}, 2^{H_2^\delta(S_i|R_i)\rho_{S_i R_i}})) \right)^{N-1} - 1 \right). \tag{B36}
\end{aligned}$$

In the derivation, we repeatedly use the triangle inequality of the one norm. In the fifth line, we utilize the fact that for quantum states ρ_i and σ_i ,

$$\begin{aligned}
\left\| \bigotimes_{i=1}^r \rho_i - \bigotimes_{i=1}^r \sigma_i \right\|_1 &= \left\| \sum_{i=1}^r \bigotimes_{j=1}^{i-1} \rho_j \otimes (\rho_i - \sigma_i) \otimes \bigotimes_{j=i+1}^r \sigma_j \right\|_1 \\
&\leq \sum_{i=1}^r \left\| \bigotimes_{j=1}^{i-1} \rho_j \otimes (\rho_i - \sigma_i) \otimes \bigotimes_{j=i+1}^r \sigma_j \right\|_1 \\
&\leq \sum_{i=1}^r \prod_{j=1}^{i-1} \|\rho_j\|_\infty \|\rho_i - \sigma_i\|_1 \prod_{j=i+1}^r \|\sigma_j\|_\infty \\
&\leq \sum_{i=1}^r \|\rho_i - \sigma_i\|_1. \tag{B37}
\end{aligned}$$

Note that $\|AB\|_1 \leq \|A\|_1 \|B\|_\infty$ and $\|\rho\|_\infty \leq 1$ for quantum state ρ .

Meanwhile, let $\delta_{\text{SR}}^+ - \delta_{\text{SR}}^- = \hat{\rho}_{\text{SR}} - \rho_{\text{SR}}$ where δ_{SR}^+ and δ_{SR}^- have orthogonal support and they are both positive. Since the two have orthogonal support, we have

$$\text{tr}(\delta_{\text{SR}}^+) = \text{tr}(\delta_{\text{SR}}^-) = \frac{1}{2} \|\hat{\rho}_{\text{SR}} - \rho_{\text{SR}}\|_1 \leq \frac{1}{2} \sum_{i=1}^{2N} \|\hat{\rho}_{S_i R_i} - \rho_{S_i R_i}\|_1 \leq 2N\delta. \quad (\text{B38})$$

Thus,

$$\begin{aligned} & \mathbb{E}_U \|\hat{\mathcal{T}}_{S \rightarrow E}(U_S \hat{\rho}_{\text{SR}} U_S^\dagger) - \hat{\mathcal{T}}_{S \rightarrow E}(U_S \rho_{\text{SR}} U_S^\dagger)\|_1 \\ &= \mathbb{E}_U \|\hat{\mathcal{T}}_{S \rightarrow E}(U_S \delta_{\text{SR}}^+ U_S^\dagger) - \hat{\mathcal{T}}_{S \rightarrow E}(U_S \delta_{\text{SR}}^- U_S^\dagger)\|_1 \\ &\leq \mathbb{E}_U \|\hat{\mathcal{T}}_{S \rightarrow E}(U_S \delta_{\text{SR}}^+ U_S^\dagger)\|_1 + \mathbb{E}_U \|\hat{\mathcal{T}}_{S \rightarrow E}(U_S \delta_{\text{SR}}^- U_S^\dagger)\|_1 \\ &= \mathbb{E}_U \text{tr}[\hat{\mathcal{T}}_{S \rightarrow E}(U_S \delta_{\text{SR}}^+ U_S^\dagger)] + \mathbb{E}_U \text{tr}[\hat{\mathcal{T}}_{S \rightarrow E}(U_S \delta_{\text{SR}}^- U_S^\dagger)] \\ &= \text{tr}(\delta_{\text{SR}}^+ + \delta_{\text{SR}}^-) \\ &\leq 4N\delta. \end{aligned} \quad (\text{B39})$$

In the fifth line, we utilize that $\hat{\mathcal{T}}_{S \rightarrow E}$ is trace-preserving. Similarly, let $\Delta_{\text{SE}}^+ - \Delta_{\text{SE}}^- = \hat{\tau}_{\text{SE}} - \tau_{\text{SE}}$ where Δ_{SE}^+ and Δ_{SE}^- have orthogonal support and they are both positive. We also have

$$\text{tr}(\Delta_{\text{SE}}^+) = \text{tr}(\Delta_{\text{SE}}^-) = \frac{1}{2} \|\hat{\tau}_{\text{SE}} - \tau_{\text{SE}}\|_1 \leq \frac{1}{2} \sum_{i=1}^{2N} \|\hat{\tau}_{S_i E_i} - \tau_{S_i E_i}\|_1 \leq 2N\delta. \quad (\text{B40})$$

Meanwhile, we denote the Choi-Jamiołkowski representations of Δ_{SE}^+ and Δ_{SE}^- as $\mathcal{D}_{S \rightarrow E}^+$ and $\mathcal{D}_{S \rightarrow E}^-$, respectively. Thus,

$$\begin{aligned} & \mathbb{E}_U \|\hat{\mathcal{T}}_{S \rightarrow E}(U_S \rho_{\text{SR}} U_S^\dagger) - \mathcal{T}_{S \rightarrow E}(U_S \rho_{\text{SR}} U_S^\dagger)\|_1 \\ &= \mathbb{E}_U \|\mathcal{D}_{S \rightarrow E}^+(U_S \rho_{\text{SR}} U_S^\dagger) - \mathcal{D}_{S \rightarrow E}^-(U_S \rho_{\text{SR}} U_S^\dagger)\|_1 \\ &\leq \mathbb{E}_U \|\mathcal{D}_{S \rightarrow E}^+(U_S \rho_{\text{SR}} U_S^\dagger)\|_1 + \mathbb{E}_U \|\mathcal{D}_{S \rightarrow E}^-(U_S \rho_{\text{SR}} U_S^\dagger)\|_1 \\ &= \mathbb{E}_U \text{tr}[\mathcal{D}_{S \rightarrow E}^+(U_S \rho_{\text{SR}} U_S^\dagger)] + \mathbb{E}_U \text{tr}[\mathcal{D}_{S \rightarrow E}^-(U_S \rho_{\text{SR}} U_S^\dagger)] \\ &= \text{tr}[\mathcal{D}_{S \rightarrow E}^+ \left(\frac{\mathbb{I}_S}{|S|} \otimes \rho_R \right)] + \text{tr}[\mathcal{D}_{S \rightarrow E}^- \left(\frac{\mathbb{I}_S}{|S|} \otimes \rho_R \right)] \\ &= \text{tr}(\Delta_{\text{SE}}^+ \otimes \rho_R) + \text{tr}(\Delta_{\text{SE}}^- \otimes \rho_R) \\ &= \text{tr}(\Delta_{\text{SE}}^+ + \Delta_{\text{SE}}^-) \\ &\leq 4N\delta. \end{aligned} \quad (\text{B41})$$

In the fifth line, we utilize the 1-design property of the random unitary such that

$$\mathbb{E}_U U_S \rho_{\text{SR}} U_S^\dagger = \frac{\mathbb{I}_S}{|S|} \otimes \rho_R. \quad (\text{B42})$$

Finally, we get

$$\begin{aligned} & \mathbb{E}_U \|\mathcal{T}_{S \rightarrow E}(U_S \rho_{\text{SR}} U_S^\dagger) - \tau_E \otimes \rho_R\|_1 \leq 16N\delta \\ &+ \sqrt{2^{-\sum_{i=1}^{2N} H_2^\delta(S_i|E_i)\tau_{S_i E_i}} - \sum_{i=1}^{2N} H_2^\delta(S_i|R_i)\rho_{S_i R_i}} + c' \prod_{i=1}^{2N} \max(1, 2^{-H_2^\delta(S_i|E_i)\tau_{S_i E_i} - H_2^\delta(S_i|R_i)\rho_{S_i R_i}}), \end{aligned} \quad (\text{B43})$$

where

$$c' = 2 \left(\left(1 + 2\eta \max_i (\max(2^{-H_2^\delta(S_i|R_i)\rho_{S_i R_i}}, 2^{H_2^\delta(S_i|R_i)\rho_{S_i R_i}})) \right)^{N-1} - 1 \right). \quad (\text{B44})$$

■

Note that we can only smooth part of conditional entropies in the smoothing process. For instance, we only smooth $H_2(S_i|E_i)$ while maintaining $H_2(S_i|R_i)$. Then, we get

$$\begin{aligned} \mathbb{E}_U \|\mathcal{T}_{S \rightarrow E}(U_S \rho_{\text{SR}} U_S^\dagger) - \tau_E \otimes \rho_R\|_1 &\leq 8N\delta \\ &+ \sqrt{2^{-\sum_{i=1}^{2N} H_2^\delta(S_i|E_i)\tau_{S_i E_i}} - \sum_{i=1}^{2N} H_2(S_i|R_i)\rho_{S_i R_i}} + c \prod_{i=1}^{2N} \max(1, 2^{-H_2^\delta(S_i|E_i)\tau_{S_i E_i} - H_2(S_i|R_i)\rho_{S_i R_i}}), \end{aligned} \quad (\text{B45})$$

where

$$c = 2 \left(\left(1 + 2\eta \max_i (\max(2^{-H_2(S_i|R_i)\rho_{S_i R_i}}, 2^{H_2(S_i|R_i)\rho_{S_i R_i}})) \right)^{N-1} - 1 \right). \quad (\text{B46})$$

Similarly, we only smooth $H_2(S_i|R_i)$ while maintaining $H_2(S_i|E_i)$ to get

$$\begin{aligned} \mathbb{E}_U \|\mathcal{T}_{S \rightarrow E}(U_S \rho_{\text{SR}} U_S^\dagger) - \tau_E \otimes \rho_R\|_1 &\leq 8N\delta \\ &+ \sqrt{2^{-\sum_{i=1}^{2N} H_2(S_i|E_i)\tau_{S_i E_i}} - \sum_{i=1}^{2N} H_2^\delta(S_i|R_i)\rho_{S_i R_i}} + c' \prod_{i=1}^{2N} \max(1, 2^{-H_2(S_i|E_i)\tau_{S_i E_i} - H_2^\delta(S_i|R_i)\rho_{S_i R_i}}), \end{aligned} \quad (\text{B47})$$

where

$$c' = 2 \left(\left(1 + 2\eta \max_i (\max(2^{-H_2^\delta(S_i|R_i)\rho_{S_i R_i}}, 2^{H_2^\delta(S_i|R_i)\rho_{S_i R_i}})) \right)^{N-1} - 1 \right). \quad (\text{B48})$$

The proof of the above formulas is the same as that of Theorem 3.

3. Proof of Corollary 1 by applying nonsmooth decoupling theorem to Pauli noise

Proof. We adopt the complementary channel to analyze the Choi error. Specifically,

$$\epsilon_{\text{Choi}} = \min_{\zeta} P \left(\left(\widehat{\mathcal{N} \circ \mathcal{E}_{L \rightarrow E}} \otimes I_R \right) \left(|\hat{\phi}\rangle\langle\hat{\phi}|_{\text{LR}} \right), \left(\mathcal{T}_{L \rightarrow E}^\zeta \otimes I_R \right) \left(|\hat{\phi}\rangle\langle\hat{\phi}|_{\text{LR}} \right) \right). \quad (\text{B49})$$

Since $\mathcal{T}_{L \rightarrow E}^\zeta$ maps all states to a fixed state ζ , $(\mathcal{T}_{L \rightarrow E}^\zeta \otimes I_R) (|\hat{\phi}\rangle\langle\hat{\phi}|_{\text{LR}}) = \zeta_E \otimes \text{tr}_L (|\hat{\phi}\rangle\langle\hat{\phi}|_{\text{LR}}) = \zeta_E \otimes \mathbb{I}_R/2^k$. Meanwhile, $\widehat{\mathcal{N} \circ \mathcal{E}_{L \rightarrow E}} = \hat{\mathcal{N}}_{S \rightarrow E} \circ \mathcal{E}_{L \rightarrow S}$ with $\mathcal{E}_{L \rightarrow S}$ the encoding map satisfying

$$\mathcal{E}_{L \rightarrow S}(\rho) = U_S(\rho_L \otimes |0\rangle\langle 0|^{n-k})U_S^\dagger, \quad (\text{B50})$$

where U_S is the encoding unitary, generated by 1D $\log(n/\varepsilon)$ -depth Clifford circuit ensemble $\mathfrak{C}_n^\varepsilon$ in our case. The term $|0\rangle\langle 0|^{n-k}$ is the ancillary qubits used for quantum error correction. Hence,

$$\begin{aligned}
\epsilon_{\text{Choi}} &= \min_{\zeta} P \left(\hat{\mathcal{N}}_{S \rightarrow E} \left(U \left| \hat{\phi} \right\rangle \left\langle \hat{\phi} \right|_{\text{LR}} \otimes |0^{n-k}\rangle\langle 0^{n-k}| U^\dagger \right), \zeta_E \otimes \frac{\mathbb{I}_R}{2^k} \right) \\
&\leq P \left(\hat{\mathcal{N}}_{S \rightarrow E} \left(U \left| \hat{\phi} \right\rangle \left\langle \hat{\phi} \right|_{\text{LR}} \otimes |0^{n-k}\rangle\langle 0^{n-k}| U^\dagger \right), \mathbb{E}_{U \sim \mathfrak{U}_n} \hat{\mathcal{N}}_{S \rightarrow E} \left(U \left| \hat{\phi} \right\rangle \left\langle \hat{\phi} \right|_{\text{LR}} \otimes |0^{n-k}\rangle\langle 0^{n-k}| U^\dagger \right) \right) \\
&\quad + \min_{\zeta} P \left(\mathbb{E}_{U \sim \mathfrak{U}_n} \hat{\mathcal{N}}_{S \rightarrow E} \left(U \left| \hat{\phi} \right\rangle \left\langle \hat{\phi} \right|_{\text{LR}} \otimes |0^{n-k}\rangle\langle 0^{n-k}| U^\dagger \right), \zeta_E \otimes \frac{\mathbb{I}_R}{2^k} \right) \\
&= P \left(\hat{\mathcal{N}}_{S \rightarrow E} \left(U \left| \hat{\phi} \right\rangle \left\langle \hat{\phi} \right|_{\text{LR}} \otimes |0^{n-k}\rangle\langle 0^{n-k}| U^\dagger \right), \hat{\mathcal{N}}_{S \rightarrow E} \left(\frac{\mathbb{I}_S}{2^n} \right) \otimes \frac{\mathbb{I}_R}{2^k} \right) + \min_{\zeta} P \left(\hat{\mathcal{N}}_{S \rightarrow E} \left(\frac{\mathbb{I}_S}{2^n} \right) \otimes \frac{\mathbb{I}_R}{2^k}, \zeta_E \otimes \frac{\mathbb{I}_R}{2^k} \right) \\
&= P \left(\hat{\mathcal{N}}_{S \rightarrow E} \left(U \left| \hat{\phi} \right\rangle \left\langle \hat{\phi} \right|_{\text{LR}} \otimes |0^{n-k}\rangle\langle 0^{n-k}| U^\dagger \right), \hat{\mathcal{N}}_{S \rightarrow E} \left(\frac{\mathbb{I}_S}{2^n} \right) \otimes \frac{\mathbb{I}_R}{2^k} \right). \tag{B51}
\end{aligned}$$

Here, in the first inequality, we utilize the triangle inequality of the purified distance. In the second equality, we utilize the property of the twirling over Haar-random unitary gates. In the third equality, we take the state ζ as $\hat{\mathcal{N}}_{S \rightarrow E}(\mathbb{I}_S/2^n)$.

To further bound $\mathbb{E}_{U \sim \mathfrak{C}_n^\varepsilon} \epsilon_{\text{Choi}}$, we utilize the relation between purified distance and trace distance in Eq. (15) to get

$$\begin{aligned}
\mathbb{E}_{U \in \text{Choi}} &\leq \mathbb{E}_U \sqrt{\left\| \hat{\mathcal{N}}_{S \rightarrow E} \left(U \left| \hat{\phi} \right\rangle \left\langle \hat{\phi} \right|_{\text{LR}} \otimes |0^{n-k}\rangle\langle 0^{n-k}| U^\dagger \right) - \hat{\mathcal{N}}_{S \rightarrow E} \left(\frac{\mathbb{I}_S}{2^n} \right) \otimes \frac{\mathbb{I}_R}{2^k} \right\|_1} \\
&\leq \sqrt{\mathbb{E}_U \left\| \hat{\mathcal{N}}_{S \rightarrow E} \left(U \left| \hat{\phi} \right\rangle \left\langle \hat{\phi} \right|_{\text{LR}} \otimes |0^{n-k}\rangle\langle 0^{n-k}| U^\dagger \right) - \hat{\mathcal{N}}_{S \rightarrow E} \left(\frac{\mathbb{I}_S}{2^n} \right) \otimes \frac{\mathbb{I}_R}{2^k} \right\|_1}. \tag{B52}
\end{aligned}$$

For simplicity, we omit $\mathfrak{C}_n^\varepsilon$ in the subscript of expectation. For i.i.d. Pauli noise, we have

$$\hat{\mathcal{N}}_{S \rightarrow E} = (\hat{\mathcal{N}}_p)^{\otimes n}, \tag{B53}$$

with $\hat{\mathcal{N}}_p$ as the complementary channel of the single-qubit Pauli noise $\mathcal{N}_p(\rho) = p_I \rho + p_X X \rho X + p_Y Y \rho Y + p_Z Z \rho Z$.

Now we apply the nonsmooth decoupling theorem. Based on Theorem 1, the term in the square root in Eq. (B52) satisfies

$$\begin{aligned}
&\mathbb{E}_U \left\| \hat{\mathcal{N}}_{S \rightarrow E} \left(U \left| \hat{\phi} \right\rangle \left\langle \hat{\phi} \right|_{\text{LR}} \otimes |0^{n-k}\rangle\langle 0^{n-k}| U^\dagger \right) - \hat{\mathcal{N}}_{S \rightarrow E} \left(\frac{\mathbb{I}_S}{2^n} \right) \otimes \frac{\mathbb{I}_R}{2^k} \right\|_1 \\
&\leq \sqrt{2^{-\sum_{i=1}^{2N} H_2(S_i|E_i) \tau_{S_i E_i} - \sum_{i=1}^{2N} H_2(S_i|R_i) \rho_{S_i R_i}} + c \prod_{i=1}^{2N} \max(1, 2^{-H_2(S_i|E_i) \tau_{S_i E_i} - H_2(S_i|R_i) \rho_{S_i R_i}})}, \tag{B54}
\end{aligned}$$

where $\tau_{S_i E_i} = \left(\hat{\mathcal{N}}_p \left(\left| \hat{\phi} \right\rangle \left\langle \hat{\phi} \right| \right) \right)^{\otimes \xi}$, $\rho_{S_i R_i} = \left| \hat{\phi} \right\rangle \left\langle \hat{\phi} \right|^{(k/n)\xi} \otimes |0^{(n-k)/n}\rangle\langle 0^{(n-k)/n}|$, and

$$c = 2 \left(\left(1 + 2\eta \max_i (\max(2^{-H_2(S_i|R_i) \rho_{S_i R_i}}, 2^{H_2(S_i|R_i) \rho_{S_i R_i}})) \right)^{N-1} - 1 \right). \tag{B55}$$

Note that $H_2(S_i|R_i)_{\rho_{S_i R_i}}$ is given by

$$\begin{aligned}
H_2(S_i|R_i)_{\rho_{S_i R_i}} &= H_2(L_i|R_i)_{\left| \hat{\phi} \right\rangle \left\langle \hat{\phi} \right|_{L_i R_i}} \\
&= \frac{k}{n} \xi H_2(L|R)_{\left| \hat{\phi} \right\rangle \left\langle \hat{\phi} \right|} \\
&= -\frac{k}{n} \xi. \tag{B56}
\end{aligned}$$

Also, given state $|\psi\rangle\langle\psi|_{SS'}$ and the purification of $\hat{\mathcal{N}}_{S' \rightarrow E}(|\psi\rangle\langle\psi|_{SS'})$ denoted as $|\phi\rangle_{SS'E}$, from Lemma A4 we can get

$$\begin{aligned} H_2(S|E)_{\hat{\mathcal{N}}_{S' \rightarrow E}(|\psi\rangle\langle\psi|_{SS'})} &= H_2(S|E)_{\text{tr}_{S'}(|\phi\rangle\langle\phi|)} \\ &= -\log \text{tr}_{S'}((\text{tr}_S(\sqrt{\text{tr}_E(|\phi\rangle\langle\phi|)}))^2) \\ &= -\log \text{tr}_{S'}((\text{tr}_S(\sqrt{\mathcal{N}_{S'}(|\psi\rangle\langle\psi|_{SS'})}))^2). \end{aligned} \quad (\text{B57})$$

Thus,

$$\begin{aligned} H_2(S_i|E_i)_{\tau_{S_i E_i}} &= \xi H_2(S|E)_{\hat{\mathcal{N}}_p(|\hat{\phi}\rangle\langle\hat{\phi}|)} \\ &= -\xi \log \text{tr}_{S'} \left(\left(\text{tr}_S(\sqrt{\mathcal{N}_p(|\hat{\phi}\rangle\langle\hat{\phi}|_{SS'})}) \right)^2 \right) \\ &= -\xi \log \text{tr}_{S'} \left(\left(\text{tr}_S \left(\sqrt{p_I |\hat{\phi}\rangle\langle\hat{\phi}| + p_X X |\hat{\phi}\rangle\langle\hat{\phi}| X + p_Y Y |\hat{\phi}\rangle\langle\hat{\phi}| Y + p_Z Z |\hat{\phi}\rangle\langle\hat{\phi}| Z} \right) \right)^2 \right) \\ &= -\xi \log \text{tr}_{S'} \left(\left(\text{tr}_S \left(\sqrt{p_I} |\hat{\phi}\rangle\langle\hat{\phi}| + \sqrt{p_X} X |\hat{\phi}\rangle\langle\hat{\phi}| X + \sqrt{p_Y} Y |\hat{\phi}\rangle\langle\hat{\phi}| Y + \sqrt{p_Z} Z |\hat{\phi}\rangle\langle\hat{\phi}| Z \right) \right)^2 \right) \\ &= -\xi \log \text{tr}_{S'} \left(\left((\sqrt{p_I} + \sqrt{p_X} + \sqrt{p_Y} + \sqrt{p_Z}) \frac{\mathbb{I}}{2} \right)^2 \right) \\ &= \xi (1 - 2 \log(\sqrt{p_I} + \sqrt{p_X} + \sqrt{p_Y} + \sqrt{p_Z})). \end{aligned} \quad (\text{B58})$$

Denote $f(\vec{p}) = 2 \log(\sqrt{p_I} + \sqrt{p_X} + \sqrt{p_Y} + \sqrt{p_Z})$, we have $H_2(S_i|E_i)_{\tau_{S_i E_i}} = 1 - f(\vec{p})$. Below, we consider the region of $\vec{p}$ such that $0 < f(\vec{p}) < 1$. For sufficiently large n , utilizing $(1 + x/n)^n \leq 1 + 2x$ for small x , we have that

$$\begin{aligned} c &\leq 8N \max(2^{-\xi(1+\frac{k}{n})}, 2^{-\xi(1-\frac{k}{n})}) \\ &\leq \frac{4n2^{-\xi(1-\frac{k}{n})}}{\xi} \\ &= \frac{4\varepsilon^{1-\frac{k}{n}} n^{\frac{k}{n}}}{\log(n/\varepsilon)}. \end{aligned} \quad (\text{B59})$$

The condition that x is small can be achieved by $(\varepsilon/n)^{1-k/n} = o(1/n)$.

Thus, for sufficiently large n , we have that

$$\begin{aligned} &\mathbb{E}_U \left\| \hat{\mathcal{N}}_{S \rightarrow E}(U |\hat{\phi}\rangle\langle\hat{\phi}|_{\text{LR}} \otimes |0^{n-k}\rangle\langle 0^{n-k}| U^\dagger) - \hat{\mathcal{N}}_{S \rightarrow E} \left(\frac{\mathbb{I}_S}{2^n} \right) \otimes \frac{\mathbb{I}_R}{2^k} \right\|_1 \\ &\leq \sqrt{2^{-\sum_{i=1}^{2N} H_2(S_i|E_i)_{\tau_{S_i E_i}} - \sum_{i=1}^{2N} H_2(S_i|R_i)_{\rho_{S_i R_i}}} + c \prod_{i=1}^{2N} \max(1, 2^{-H_2(S_i|E_i)_{\tau_{S_i E_i}} - H_2(S_i|R_i)_{\rho_{S_i R_i}}})} \\ &\leq \sqrt{2^{-n(1-f(\vec{p})-\frac{k}{n})} + \frac{4\varepsilon^{1-\frac{k}{n}} n^{\frac{k}{n}}}{\log(n/\varepsilon)} \prod_{i=1}^{2N} \max(1, 2^{-\xi(1-f(\vec{p})-\frac{k}{n})})}. \end{aligned} \quad (\text{B60})$$

From the above equation, as long as ε decays faster than $n^{-\frac{k}{n-k}}$, and encoding rate $k/n \leq 1 - f(\vec{p})$, the Choi error decays polynomially. Hence, we conclude that 1D $O(\log(n))$ -depth circuit achieves encoding rate $1 - f(\vec{p})$. As a reminder, this rate is lower than the hashing bound $1 - h(\vec{p})$. ■

4. Proof of Corollary 2 that 1D low-depth circuit achieves the hashing bound with smooth decoupling theorem

Proof. We follow the derivation in the last subsection but substitute the nonsmooth decoupling theorem with the smooth decoupling theorem. Based on Eq. (B45), the term in the square root in Eq. (B52) satisfies

$$\mathbb{E}_U \left\| \hat{\mathcal{N}}_{S \rightarrow E}(U \left| \hat{\phi} \right\rangle \left\langle \hat{\phi} \right|_{\text{LR}} \otimes |0^{n-k}\rangle \langle 0^{n-k}| U^\dagger) - \hat{\mathcal{N}}_{S \rightarrow E} \left(\frac{\mathbb{I}_S}{2^n} \right) \otimes \frac{\mathbb{I}_R}{2^k} \right\|_1 \leq 8\delta$$

$$+ \sqrt{2^{-\sum_{i=1}^{2N} H_2^{\delta/N}(S_i|E_i)\tau_{S_i E_i} - \sum_{i=1}^{2N} H_2(S_i|R_i)\rho_{S_i R_i}} + c \prod_{i=1}^{2N} \max(1, 2^{-H_2^{\delta/N}(S_i|E_i)\tau_{S_i E_i} - H_2(S_i|R_i)\rho_{S_i R_i}})}. \quad (\text{B61})$$

where $\tau_{S_i E_i} = \hat{\mathcal{N}}_{S' \rightarrow E}(|\hat{\phi}\rangle\langle\hat{\phi}|_{SS'}) = (\hat{\mathcal{N}}_p(|\hat{\phi}\rangle\langle\hat{\phi}|))^{\otimes \xi}$, $\rho_{S_i R_i} = |\hat{\phi}\rangle\langle\hat{\phi}|_{L_i R_i} \otimes |0^{((n-k)/n)\xi}\rangle\langle 0^{((n-k)/n)\xi}| = |\hat{\phi}\rangle\langle\hat{\phi}|^{(k/n)\xi} \otimes |0^{((n-k)/n)\xi}\rangle\langle 0^{((n-k)/n)\xi}|$, and

$$c = 2 \left(\left(1 + 2\eta \max_i (\max(2^{-H_2(S_i|R_i)\rho_{S_i R_i}}, 2^{H_2(S_i|R_i)\rho_{S_i R_i}})) \right)^{N-1} - 1 \right). \quad (\text{B62})$$

By Eq. (B56), we obtain that $H_2(S_i|R_i)\rho_{S_i R_i} = -(k/n)\xi$. Below, we analyze $H_2^{\delta/N}(S_i|E_i)\tau_{S_i E_i}$. Note that for any $0 < \delta < 1$ as long as $\xi \geq -\frac{8}{5} \log(1 - \sqrt{1 - \delta^2/N^2})$, we have

$$H(S|E)_\rho - \frac{4 \log \gamma \sqrt{\log \frac{2N^2}{\delta^2}}}{\sqrt{\xi}} \leq \frac{1}{\xi} H_2^{\delta/N}(S|E)_{\rho^{\otimes \xi}} \leq H(S|E)_\rho + \frac{4 \log \gamma \sqrt{\log \frac{2N^2}{\delta^2}}}{\sqrt{\xi}}, \quad (\text{B63})$$

with

$$\gamma \leq \sqrt{2^{-H_{\min}^\delta(S|E)_\rho}} + \sqrt{2^{H_{\max}^\delta(S|E)_\rho}} + 1. \quad (\text{B64})$$

The condition $\xi \geq -\frac{8}{5} \log(1 - \sqrt{1 - \delta^2/N^2})$ can be fulfilled as long as $n^{\frac{11}{8}} \epsilon^{\frac{5}{8}}$ decays with n . In this case, given any $0 < \delta < 1$, for sufficiently large n , we have

$$2N^2 \left(\frac{\epsilon}{n} \right)^{\frac{5}{8}} \leq \delta^2, \quad (\text{B65})$$

which implies

$$\left(1 - \left(\frac{\epsilon}{n} \right)^{\frac{5}{8}} \right)^2 \geq 1 - \frac{\delta^2}{N^2} \quad (\text{B66})$$

and

$$\xi \geq -\frac{8}{5} \log(1 - \sqrt{1 - \delta^2/N^2}). \quad (\text{B67})$$

Note that based on the definition of the complementary channel [27], given state $|\psi\rangle\langle\psi|_{SS'}$ and the purification of $\hat{\mathcal{N}}_{S' \rightarrow E}(|\psi\rangle\langle\psi|_{SS'})$ denoted as $|\phi\rangle_{SS'E}$, we can get

$$\text{tr}_E |\phi\rangle\langle\phi|_{SS'E} = \hat{\mathcal{N}}_{S'}(|\psi\rangle\langle\psi|_{SS'}). \quad (\text{B68})$$

Thus,

$$\begin{aligned} H(S|E)_{\hat{\mathcal{N}}_{S' \rightarrow E}(|\psi\rangle\langle\psi|_{SS'})} &= H(SE) - H(E) \\ &= H(S') - H(SS') \\ &= -H(S|S')_{\hat{\mathcal{N}}_{S'}(|\psi\rangle\langle\psi|_{SS'})}. \end{aligned} \quad (\text{B69})$$

With direct calculation, we have that

$$\mathcal{N}_p \left(\left| \hat{\phi} \right\rangle \left\langle \hat{\phi} \right|_{SS'} \right) = p_I \left| \hat{\phi} \right\rangle \left\langle \hat{\phi} \right| + p_X X \left| \hat{\phi} \right\rangle \left\langle \hat{\phi} \right| X + p_Y Y \left| \hat{\phi} \right\rangle \left\langle \hat{\phi} \right| Y + p_Z Z \left| \hat{\phi} \right\rangle \left\langle \hat{\phi} \right| Z \quad (\text{B70})$$

and

$$\text{tr}_S \mathcal{N}_p \left(\left| \hat{\phi} \right\rangle \left\langle \hat{\phi} \right|_{SS'} \right) = \frac{\mathbb{I}}{2}. \quad (\text{B71})$$

Note that $\left| \hat{\phi} \right\rangle \left\langle \hat{\phi} \right|$, $X \left| \hat{\phi} \right\rangle \left\langle \hat{\phi} \right| X$, $Y \left| \hat{\phi} \right\rangle \left\langle \hat{\phi} \right| Y$, and $Z \left| \hat{\phi} \right\rangle \left\langle \hat{\phi} \right| Z$ are mutually orthogonal. Hence, $H(S|E)_{\hat{\mathcal{N}}_p(\left| \hat{\phi} \right\rangle \left\langle \hat{\phi} \right|)} = H \left(\text{tr}_S \mathcal{N}_p \left(\left| \hat{\phi} \right\rangle \left\langle \hat{\phi} \right|_{SS'} \right) \right) - H \left(\mathcal{N}_p \left(\left| \hat{\phi} \right\rangle \left\langle \hat{\phi} \right|_{SS'} \right) \right) = 1 - h(\bar{p})$. Meanwhile, for any $0 < \delta < 1$,

$$\lim_{n \rightarrow \infty} 4 \log \gamma \sqrt{\frac{\log \frac{2N^2}{\delta^2}}{\xi}} = \lim_{n \rightarrow \infty} 4 \log \gamma \sqrt{\frac{\log \frac{2N^2}{\delta^2}}{\log \frac{n}{\epsilon}}} = 4 \log \gamma \sqrt{2 \lim_{n \rightarrow \infty} \frac{\log n}{\log \frac{n}{\epsilon}}}. \quad (\text{B72})$$

As long as $\lim_{n \rightarrow \infty} \log n / \log(n/\epsilon) = 0$, for sufficiently large n , we have that

$$4 \log \gamma \sqrt{\frac{\log \frac{2N^2}{\delta^2}}{\xi}} \leq \delta. \quad (\text{B73})$$

This can be achieved by choosing $\log n / \epsilon$ as a function increasing slightly faster than $\log n$ like $(\log n) \log \log \log n$. This additional factor $\log \log \log n$ is normally negligible for a practical number n . Also, recall that by Eq. (B59), $c \leq \frac{4\epsilon^{1-\frac{k}{n}} n^{\frac{k}{n}}}{\log(n/\epsilon)}$. Thus, for sufficiently large n and any $0 < \delta < 1$, we have that

$$\begin{aligned} & \mathbb{E}_U \left\| \hat{\mathcal{N}}_{S \rightarrow E} \left(U \left| \hat{\phi} \right\rangle \left\langle \hat{\phi} \right|_{\text{LR}} \otimes |0^{n-k}\rangle \langle 0^{n-k}| U^\dagger \right) - \hat{\mathcal{N}}_{S \rightarrow E} \left(\frac{\mathbb{I}_S}{2^n} \right) \otimes \frac{\mathbb{I}_R}{2^k} \right\|_1 \\ & \leq 8\delta + \sqrt{2^{-\sum_{i=1}^{2N} \xi(1-h(\bar{p})-\frac{k}{n}-\delta)} + \frac{4\epsilon^{1-\frac{k}{n}} n^{\frac{k}{n}}}{\log(n/\epsilon)} \prod_{i=1}^{2N} \max(1, 2^{-\xi(1-h(\bar{p})-\frac{k}{n}-\delta)})} \\ & = \delta + \sqrt{2^{-n(1-h(\bar{p})-\frac{k}{n}-\delta)} + \frac{4\epsilon^{1-\frac{k}{n}} n^{\frac{k}{n}}}{\log(n/\epsilon)} \prod_{i=1}^{2N} \max(1, 2^{-\xi(1-h(\bar{p})-\frac{k}{n}-\delta)})}. \end{aligned} \quad (\text{B74})$$

When $1 - h(\bar{p}) - k/n - \delta > 0$, we have that

$$\mathbb{E}_U \left\| \hat{\mathcal{N}}_{S \rightarrow E} \left(U \left| \hat{\phi} \right\rangle \left\langle \hat{\phi} \right|_{\text{LR}} \otimes |0^{n-k}\rangle \langle 0^{n-k}| U^\dagger \right) - \hat{\mathcal{N}}_{S \rightarrow E} \left(\frac{\mathbb{I}_S}{2^n} \right) \otimes \frac{\mathbb{I}_R}{2^k} \right\|_1 \leq 8\delta + \sqrt{2^{-n(1-h(\bar{p})-\frac{k}{n}-\delta)} + \frac{4\epsilon^{1-\frac{k}{n}} n^{\frac{k}{n}}}{\log(n/\epsilon)}}. \quad (\text{B75})$$

Substituting the above equation into Eq. (B52) completes the proof. The term $2^{-n(1-h(\bar{p})-\frac{k}{n}-\delta)}$ is the error associated with the random Clifford encoding, as demonstrated in Appendix B 12.

As a remark, the decoupling error within Eq. (B75) is dominated by 8δ . Thus, the scaling of δ controls the decay rate of the decoupling error and the associated Choi error. From Eq. (B73), we obtain that δ can be chosen as the scaling of $O(\sqrt{\log n / \log(n/\epsilon)})$. Thus, the scaling of the Choi error is $O((\log n / \log(n/\epsilon))^{\frac{1}{4}})$. ■

5. Proof of Corollary 3 by applying nonsmooth decoupling theorem to i.i.d. erasure error

Proof. Following the previous derivation, the expectation of Choi error over $\mathfrak{C}_n^\varepsilon$ can be bounded by

$$\mathbb{E}_{U \in \text{Choi}} \leq \sqrt{\mathbb{E}_U \left\| \hat{\mathcal{N}}_{S \rightarrow E} \left(U \left| \hat{\phi} \right\rangle \left\langle \hat{\phi} \right|_{\text{LR}} \otimes |0^{n-k}\rangle \langle 0^{n-k}| U^\dagger \right) - \hat{\mathcal{N}}_{S \rightarrow E} \left(\frac{\mathbb{I}_S}{2^n} \right) \otimes \frac{\mathbb{I}_R}{2^k} \right\|_1}. \quad (\text{B76})$$

For i.i.d. erasure error, we have

$$\hat{\mathcal{N}}_{S \rightarrow E} = (\hat{\mathcal{N}}_e)^{\otimes n}, \quad (\text{B77})$$

with $\hat{\mathcal{N}}_e$ as the complementary channel of the single-qubit erasure error $\mathcal{N}_e(\rho) = (1-p)\rho + p|2\rangle\langle 2|$.

Applying the nonsmooth decoupling theorem or Theorem 1, the term in the square root in Eq. (B52) satisfies

$$\begin{aligned} & \mathbb{E}_U \left\| \hat{\mathcal{N}}_{S \rightarrow E} \left(U \left| \hat{\phi} \right\rangle \left\langle \hat{\phi} \right|_{\text{LR}} \otimes |0^{n-k}\rangle \langle 0^{n-k}| U^\dagger \right) - \hat{\mathcal{N}}_{S \rightarrow E} \left(\frac{\mathbb{I}_S}{2^n} \right) \otimes \frac{\mathbb{I}_R}{2^k} \right\|_1 \\ & \leq \sqrt{2^{-\sum_{i=1}^{2N} H_2(S_i|E_i)\tau_{S_i E_i} - \sum_{i=1}^{2N} H_2(S_i|R_i)\rho_{S_i R_i}} + c \prod_{i=1}^{2N} \max(1, 2^{-H_2(S_i|E_i)\tau_{S_i E_i} - H_2(S_i|R_i)\rho_{S_i R_i}})}, \end{aligned} \quad (\text{B78})$$

where $\tau_{S_i E_i} = \left(\hat{\mathcal{N}}_e \left(\left| \hat{\phi} \right\rangle \left\langle \hat{\phi} \right| \right) \right)^{\otimes \xi}$, $\rho_{S_i R_i} = \left| \hat{\phi} \right\rangle \left\langle \hat{\phi} \right|^{(k/n)\xi} \otimes |0^{(n-k)/n}\rangle \langle 0^{(n-k)/n}|$, and

$$c = 2 \left(\left(1 + 2\eta \max_i (\max(2^{-H_2(S_i|R_i)\rho_{S_i R_i}}, 2^{H_2(S_i|R_i)\rho_{S_i R_i}})) \right)^{N-1} - 1 \right). \quad (\text{B79})$$

Note that $H_2(S_i|R_i)\rho_{S_i R_i} = -(k/n)\xi$ given by Eq. (B56). Also, from Eq. (B57) we can get

$$\begin{aligned} H_2(S_i|E_i)\tau_{S_i E_i} &= \xi H_2(S|E)_{\hat{\mathcal{N}}_e(|\hat{\phi}\rangle\langle\hat{\phi}|)} \\ &= -\xi \log \text{tr}_{S'} \left(\left(\text{tr}_S \left(\sqrt{\mathcal{N}_e(|\hat{\phi}\rangle\langle\hat{\phi}|_{SS'})} \right) \right)^2 \right) \\ &= -\xi \log \text{tr}_{S'} \left(\left(\text{tr}_S \left(\sqrt{(1-p)|\hat{\phi}\rangle\langle\hat{\phi}|_{SS'} + p \frac{\mathbb{I}_S}{2} \otimes |2\rangle\langle 2|_{S'}} \right) \right)^2 \right) \\ &= -\xi \log \text{tr}_{S'} \left(\left(\text{tr}_S \left(\sqrt{1-p} |\hat{\phi}\rangle\langle\hat{\phi}|_{SS'} + \sqrt{\frac{p}{2}} \mathbb{I}_S \otimes |2\rangle\langle 2|_{S'} \right) \right)^2 \right) \\ &= -\xi \log \text{tr}_{S'} \left(\left(\sqrt{1-p} \frac{\mathbb{I}_{S'}}{2} + \sqrt{2p} |2\rangle\langle 2|_{S'} \right)^2 \right) \\ &= -\xi \log \text{tr}_{S'} \left(\frac{1-p}{4} \mathbb{I}_{S'} + 2p |2\rangle\langle 2|_{S'} \right) \\ &= -\xi \log \left(\frac{1-p}{2} + 2p \right) \\ &= \xi(1 - \log(1 + 3p)). \end{aligned} \quad (\text{B80})$$

By Eq. (B59), $c \leq \frac{4e^{1-\frac{k}{n}} n^{\frac{k}{n}}}{\log(n/\varepsilon)}$. Thus, for sufficiently large n , we have that

$$\begin{aligned} & \mathbb{E}_U \left\| \hat{\mathcal{N}}_{S \rightarrow E} \left(U \left| \hat{\phi} \right\rangle \left\langle \hat{\phi} \right|_{\text{LR}} \otimes |0^{n-k}\rangle \langle 0^{n-k}| U^\dagger \right) - \hat{\mathcal{N}}_{S \rightarrow E} \left(\frac{\mathbb{I}_S}{2^n} \right) \otimes \frac{\mathbb{I}_R}{2^k} \right\|_1 \\ & \leq \sqrt{2^{-\sum_{i=1}^{2N} H_2(S_i|E_i) \tau_{S_i E_i} - \sum_{i=1}^{2N} H_2(S_i|R_i) \rho_{S_i R_i}} + c \prod_{i=1}^{2N} \max(1, 2^{-H_2(S_i|E_i) \tau_{S_i E_i} - H_2(S_i|R_i) \rho_{S_i R_i}})} \\ & \leq \sqrt{2^{-n(1-\log(1+3p)-\frac{k}{n})} + \frac{4e^{1-\frac{k}{n}} n^{\frac{k}{n}}}{\log(n/\varepsilon)} \prod_{i=1}^{2N} \max(1, 2^{-\xi(1-\log(1+3p)-\frac{k}{n})})}. \end{aligned} \quad (\text{B81})$$

Substituting the above equation into Eq. (B76) completes the proof. From the above equation, we have that as long as ε decays faster than $n^{-k/(n-k)}$, and encoding rate $k/n \leq 1 - \log(1+3p)$, the Choi error decays polynomially. Hence, we conclude that 1D $O(\log(n))$ -depth circuit achieves encoding rate $1 - \log(1+3p)$ for i.i.d. erasure error. ■

6. Proof of Corollary 4 by applying smooth decoupling theorem to i.i.d. erasure error

Proof. We follow the derivation in the last subsection but substitute the nonsmooth decoupling theorem with the smooth decoupling theorem. Based on Eq. (B45), we have

$$\begin{aligned} & \mathbb{E}_U \left\| \hat{\mathcal{N}}_{S \rightarrow E} \left(U \left| \hat{\phi} \right\rangle \left\langle \hat{\phi} \right|_{\text{LR}} \otimes |0^{n-k}\rangle \langle 0^{n-k}| U^\dagger \right) - \hat{\mathcal{N}}_{S \rightarrow E} \left(\frac{\mathbb{I}_S}{2^n} \right) \otimes \frac{\mathbb{I}_R}{2^k} \right\|_1 \leq 8\delta \\ & + \sqrt{2^{-\sum_{i=1}^{2N} H_2^{\delta/N}(S_i|E_i) \tau_{S_i E_i} - \sum_{i=1}^{2N} H_2(S_i|R_i) \rho_{S_i R_i}} + c \prod_{i=1}^{2N} \max(1, 2^{-H_2^{\delta/N}(S_i|E_i) \tau_{S_i E_i} - H_2(S_i|R_i) \rho_{S_i R_i}})}. \end{aligned} \quad (\text{B82})$$

where $\tau_{S_i E_i} = \hat{\mathcal{N}}_{S' \rightarrow E} \left(\left| \hat{\phi} \right\rangle \left\langle \hat{\phi} \right|_{SS'} \right) = \left(\hat{\mathcal{N}}_e \left(\left| \hat{\phi} \right\rangle \left\langle \hat{\phi} \right| \right) \right)^{\otimes \xi}$, $\rho_{S_i R_i} = \left| \hat{\phi} \right\rangle \left\langle \hat{\phi} \right|_{L_i R_i} \otimes |0^{((n-k)/n)\xi}\rangle \langle 0^{((n-k)/n)\xi}| = \left| \hat{\phi} \right\rangle \left\langle \hat{\phi} \right|^{(k/n)\xi} \otimes |0^{((n-k)/n)\xi}\rangle \langle 0^{((n-k)/n)\xi}|$, and

$$c = 2 \left(\left(1 + 2\eta \max_i (\max(2^{-H_2(S_i|R_i) \rho_{S_i R_i}}, 2^{H_2(S_i|R_i) \rho_{S_i R_i}})) \right)^{N-1} - 1 \right). \quad (\text{B83})$$

Note that $H_2(S_i|R_i) \rho_{S_i R_i} = -(k/n)\xi$ given by Eq. (B56). Same with Appendix B 4, as long as $\lim_{n \rightarrow \infty} \log n / \log(n/\varepsilon) = 0$, for sufficiently large n , we have that

$$\left| \frac{1}{\xi} H_2^{\delta/N}(S|E)_{\rho^{\otimes \xi}} - H(S|E)_\rho \right| \leq \delta. \quad (\text{B84})$$

Meanwhile, from Eq. (B69) we have

$$\begin{aligned} H(S_i|E_i) \tau_{S_i E_i} &= \xi H(S|E)_{\hat{\mathcal{N}}_e(|\hat{\phi}\rangle\langle\hat{\phi}|)} \\ &= -\xi H(S|S')_{\mathcal{N}_e(|\hat{\phi}\rangle\langle\hat{\phi}|)} \\ &= -\xi H(S|S')_{(1-p)|\hat{\phi}\rangle\langle\hat{\phi}|_{SS'} + p \frac{\mathbb{I}_S}{2} \otimes |2\rangle\langle 2|_{S'}} \\ &= \xi \left(H(S')_{\frac{1-p}{2}\mathbb{I}_{S'} + p|2\rangle\langle 2|_{S'}} - H(SS')_{(1-p)|\hat{\phi}\rangle\langle\hat{\phi}|_{SS'} + p \frac{\mathbb{I}_S}{2} \otimes |2\rangle\langle 2|_{S'}} \right) \\ &= \xi \left(-p \log p - (1-p) \log \frac{1-p}{2} + (1-p) \log(1-p) + p \log \frac{p}{2} \right) \\ &= \xi(1-2p). \end{aligned} \quad (\text{B85})$$

Also, by Eq. (B59), $c \leq \frac{4\varepsilon^{1-\frac{k}{n}}n^{\frac{k}{n}}}{\log(n/\varepsilon)}$. Thus, for sufficiently large n and $0 < \delta < 1$, we have that

$$\begin{aligned} & \mathbb{E}_U \left\| \hat{\mathcal{N}}_{S \rightarrow E} \left(U \left| \hat{\phi} \right\rangle \left\langle \hat{\phi} \right|_{\text{LR}} \otimes |0^{n-k}\rangle \langle 0^{n-k}| U^\dagger \right) - \hat{\mathcal{N}}_{S \rightarrow E} \left(\frac{\mathbb{I}_S}{2^n} \right) \otimes \frac{\mathbb{I}_R}{2^k} \right\|_1 \\ & \leq 8\delta + \sqrt{2^{-\sum_{i=1}^{2N} \xi(1-2p-\frac{k}{n}-\delta)} + \frac{4\varepsilon^{1-\frac{k}{n}}n^{\frac{k}{n}}}{\log(n/\varepsilon)} \prod_{i=1}^{2N} \max(1, 2^{-\xi(1-2p-\frac{k}{n}-\delta)})} \\ & = 8\delta + \sqrt{2^{-n(1-2p-\frac{k}{n}-\delta)} + \frac{4\varepsilon^{1-\frac{k}{n}}n^{\frac{k}{n}}}{\log(n/\varepsilon)} \prod_{i=1}^{2N} \max(1, 2^{-\xi(1-2p-\frac{k}{n}-\delta)})}. \end{aligned} \quad (\text{B86})$$

When $1 - 2p - k/n - \delta > 0$, we have that

$$\mathbb{E}_U \left\| \hat{\mathcal{N}}_{S \rightarrow E} \left(U \left| \hat{\phi} \right\rangle \left\langle \hat{\phi} \right|_{\text{LR}} \otimes |0^{n-k}\rangle \langle 0^{n-k}| U^\dagger \right) - \hat{\mathcal{N}}_{S \rightarrow E} \left(\frac{\mathbb{I}_S}{2^n} \right) \otimes \frac{\mathbb{I}_R}{2^k} \right\|_1 \leq 8\delta + \sqrt{2^{-n(1-2p-\frac{k}{n}-\delta)} + \frac{4\varepsilon^{1-\frac{k}{n}}n^{\frac{k}{n}}}{\log(n/\varepsilon)}}. \quad (\text{B87})$$

Substituting the above equation into Eq. (B76) completes the proof. From the above equation, we have that the encoding rate k/n can achieve $1 - 2p$ for i.i.d. erasure error with 1D $\omega(\log(n))$ -depth circuit. This encoding rate is also the quantum channel capacity of the erasure error. ■

7. Proof of Corollary 5 by applying nonsmooth decoupling theorem to local amplitude damping noise

Proof. Following the previous derivation for local Pauli noise and erasure errors, the key point is to evaluate $H_2(S_i|R_i)_{\rho_{S_i R_i}}$ and $H_2(S_i|E_i)_{\tau_{S_i E_i}}$. Similar to previous analysis, $H_2(S_i|R_i)_{\rho_{S_i R_i}} = -(k/n)\xi$.

For local amplitude damping noise, we have

$$\hat{\mathcal{N}}_{S \rightarrow E} = (\hat{\mathcal{N}}_a)^{\otimes n}, \quad (\text{B88})$$

with $\hat{\mathcal{N}}_a$ as the complementary channel of the amplitude damping noise. Thus,

$$\begin{aligned} & H_2(S_i|E_i)_{\tau_{S_i E_i}} \\ & = -\xi \log \text{tr}_{S'} \left(\left(\text{tr}_S \left(\sqrt{\mathcal{N}_a \left(\left| \hat{\phi} \right\rangle \left\langle \hat{\phi} \right|_{SS'} \right)} \right) \right)^2 \right) \\ & = -\xi \log \text{tr}_{S'} \left(\left(\text{tr}_S \left(\sqrt{\frac{1}{2}(|00\rangle + \sqrt{1-p}|11\rangle)(\langle 00| + \sqrt{1-p}\langle 11|)_{SS'} + \frac{p}{2}|10\rangle\langle 10|_{SS'}}} \right) \right)^2 \right) \\ & = -\xi \log \text{tr}_{S'} \left(\left(\text{tr}_S \left(\sqrt{\frac{2-p}{2} \left(\sqrt{\frac{1}{2-p}}|00\rangle + \sqrt{\frac{1-p}{2-p}}|11\rangle \right) \left(\sqrt{\frac{1}{2-p}}\langle 00| + \sqrt{\frac{1-p}{2-p}}\langle 11| \right)_{SS'} + \frac{p}{2}|10\rangle\langle 10|_{SS'}}} \right) \right)^2 \right) \\ & = -\xi \log \text{tr}_{S'} \left(\left(\text{tr}_S \left(\sqrt{\frac{2-p}{2} \left(\sqrt{\frac{1}{2-p}}|00\rangle + \sqrt{\frac{1-p}{2-p}}|11\rangle \right) \left(\sqrt{\frac{1}{2-p}}\langle 00| + \sqrt{\frac{1-p}{2-p}}\langle 11| \right)_{SS'} + \sqrt{\frac{p}{2}}|10\rangle\langle 10|_{SS'}}} \right) \right)^2 \right) \\ & = -\xi \log \text{tr}_{S'} \left(\left(\left(\frac{1}{\sqrt{4-2p}} + \sqrt{\frac{p}{2}} \right) |0\rangle\langle 0|_{S'} + \frac{1-p}{\sqrt{4-2p}} |1\rangle\langle 1|_{S'} \right)^2 \right) \\ & = -\xi \log \text{tr}_{S'} \left(\left(\left(\frac{1}{\sqrt{4-2p}} + \sqrt{\frac{p}{2}} \right)^2 |0\rangle\langle 0|_{S'} + \frac{(1-p)^2}{4-2p} |1\rangle\langle 1|_{S'} \right) \right) \\ & = -\xi \log \left(\left(\frac{1}{\sqrt{4-2p}} + \sqrt{\frac{p}{2}} \right)^2 + \frac{(1-p)^2}{4-2p} \right) \\ & = -\xi \log \left(\frac{1}{2-p} + \sqrt{\frac{p}{2-p}} \right). \end{aligned} \quad (\text{B89})$$

By Eq. (B59), $c \leq \frac{4\varepsilon^{1-\frac{k}{n}} \frac{k}{n}}{\log(n/\varepsilon)}$. Thus, for sufficiently large n , we have that

$$\begin{aligned} & \mathbb{E}_U \left\| \hat{\mathcal{N}}_{S \rightarrow E}(U \left| \hat{\phi} \right\rangle \left\langle \hat{\phi} \right|_{\text{LR}} \otimes |0^{n-k}\rangle \langle 0^{n-k}| U^\dagger) - \hat{\mathcal{N}}_{S \rightarrow E} \left(\frac{\mathbb{I}_S}{2^n} \right) \otimes \frac{\mathbb{I}_R}{2^k} \right\|_1 \\ & \leq \sqrt{2^{-\sum_{i=1}^{2N} H_2(S_i|E_i) \tau_{S_i E_i} - \sum_{i=1}^{2N} H_2(S_i|R_i) \rho_{S_i R_i}} + c \prod_{i=1}^{2N} \max(1, 2^{-H_2(S_i|E_i) \tau_{S_i E_i} - H_2(S_i|R_i) \rho_{S_i R_i}})} \\ & \leq \sqrt{2^{-n(-\log(\frac{1}{2-p} + \sqrt{\frac{p}{2-p}}) - \frac{k}{n})} + \frac{4\varepsilon^{1-\frac{k}{n}} \frac{k}{n}}{\log(n/\varepsilon)} \prod_{i=1}^{2N} \max(1, 2^{-\xi(-\log(\frac{1}{2-p} + \sqrt{\frac{p}{2-p}}) - \frac{k}{n}))}. \end{aligned} \quad (\text{B90})$$

From the above equation, we have that as long as ε decays faster than $n^{-k/(n-k)}$, and encoding rate $k/n \leq -\log\left(1/(2-p) + \sqrt{p/(2-p)}\right)$, the Choi error decays polynomially. Hence, we conclude that 1D $O(\log(n))$ -depth circuit achieves encoding rate $-\log\left(1/(2-p) + \sqrt{p/(2-p)}\right)$ for local amplitude damping noise. ■

8. Proof of Corollary 6 by applying smooth decoupling theorem to local amplitude damping noise

Proof. Following the previous derivation, the key point is to evaluate $H_2(S_i|R_i)_{\rho_{S_i R_i}}$ and $H(S_i|E_i)_{\tau_{S_i E_i}}$. Similar to previous analysis, $H_2(S_i|R_i)_{\rho_{S_i R_i}} = -(k/n)\xi$. Meanwhile, from Eq. (B69) we have

$$\begin{aligned} H(S_i|E_i)_{\tau_{S_i E_i}} &= -\xi H(S|S')_{\mathcal{N}_d(|\hat{\phi}\rangle\langle\hat{\phi}|)} \\ &= -\xi H(S|S')_{\frac{1}{2}(|00\rangle + \sqrt{1-p}|11\rangle)(\langle 00| + \sqrt{1-p}\langle 11|)_{SS'} + \frac{p}{2}|10\rangle\langle 10|_{SS'}} \\ &= \xi \left(H(S')_{\frac{1+p}{2}|0\rangle\langle 0|_{S'} + \frac{1-p}{2}|1\rangle\langle 1|_{S'}} - H(SS')_{\frac{1}{2}(|00\rangle + \sqrt{1-p}|11\rangle)(\langle 00| + \sqrt{1-p}\langle 11|)_{SS'} + \frac{p}{2}|10\rangle\langle 10|_{SS'}} \right) \\ &= \xi \left(h\left(\frac{1-p}{2}\right) - h\left(\frac{p}{2}\right) \right). \end{aligned} \quad (\text{B91})$$

Also, by Eq. (B59), $c \leq \frac{4\varepsilon^{1-\frac{k}{n}} \frac{k}{n}}{\log(n/\varepsilon)}$. Thus, for sufficiently large n and $0 < \delta < 1$, we have that

$$\begin{aligned} & \mathbb{E}_U \left\| \hat{\mathcal{N}}_{S \rightarrow E}(U \left| \hat{\phi} \right\rangle \left\langle \hat{\phi} \right|_{\text{LR}} \otimes |0^{n-k}\rangle \langle 0^{n-k}| U^\dagger) - \hat{\mathcal{N}}_{S \rightarrow E} \left(\frac{\mathbb{I}_S}{2^n} \right) \otimes \frac{\mathbb{I}_R}{2^k} \right\|_1 \\ & \leq 8\delta + \sqrt{2^{-n(h(\frac{1-p}{2}) - h(\frac{p}{2}) - \frac{k}{n} - \delta)} + \frac{4\varepsilon^{1-\frac{k}{n}} \frac{k}{n}}{\log(n/\varepsilon)} \prod_{i=1}^{2N} \max\left(1, 2^{-\xi(h(\frac{1-p}{2}) - h(\frac{p}{2}) - \frac{k}{n} - \delta)}\right)}. \end{aligned} \quad (\text{B92})$$

When $h((1-p)/2) - h(p/2) - k/n - \delta > 0$, we have that

$$\mathbb{E}_U \left\| \hat{\mathcal{N}}_{S \rightarrow E}(U \left| \hat{\phi} \right\rangle \left\langle \hat{\phi} \right|_{\text{LR}} \otimes |0^{n-k}\rangle \langle 0^{n-k}| U^\dagger) - \hat{\mathcal{N}}_{S \rightarrow E} \left(\frac{\mathbb{I}_S}{2^n} \right) \otimes \frac{\mathbb{I}_R}{2^k} \right\|_1 \leq 8\delta + \sqrt{2^{-n(h(\frac{1-p}{2}) - h(\frac{p}{2}) - \frac{k}{n} - \delta)} + \frac{4\varepsilon^{1-\frac{k}{n}} \frac{k}{n}}{\log(n/\varepsilon)}}. \quad (\text{B93})$$

From the above equation, we have that the encoding rate k/n can achieve $h((1-p)/2) - h(p/2)$ for local amplitude damping noise with 1D $O(\log(n))$ -depth circuit, which is the same as the bound with random stabilizer codes. ■

9. Proof of Corollary 7 by applying the nonsmooth decoupling theorem to nearest-neighbor ZZ-coupling noise

Proof. Following the previous derivation, the expectation of Choi error over $\mathfrak{C}_n^\varepsilon$ can be bounded by

$$\mathbb{E}_{U \in \mathfrak{C}_n^\varepsilon} \leq \sqrt{\mathbb{E}_U \left\| \hat{\mathcal{N}}_{S \rightarrow E}(U \left| \hat{\phi} \right\rangle \left\langle \hat{\phi} \right|_{\text{LR}} \otimes |0^{n-k}\rangle \langle 0^{n-k}| U^\dagger) - \hat{\mathcal{N}}_{S \rightarrow E} \left(\frac{\mathbb{I}_S}{2^n} \right) \otimes \frac{\mathbb{I}_R}{2^k} \right\|_1}. \quad (\text{B94})$$

For 1D nearest-neighbor ZZ-coupling noise, we have

$$\mathcal{N}_S = \mathcal{N}_{zz} = \circ_{i=1}^{n-1} ((1-p)\mathcal{I} + p\mathcal{Z}_i\mathcal{Z}_{i+1}). \quad (\text{B95})$$

Applying the nonsmooth decoupling theorem or Theorem 1, the term in the square root in Eq. (B52) satisfies

$$\begin{aligned} & \mathbb{E}_U \left\| \hat{\mathcal{N}}_{S \rightarrow E}(U \left| \hat{\phi} \right\rangle \left\langle \hat{\phi} \right|_{\text{LR}} \otimes |0^{n-k}\rangle \langle 0^{n-k}| U^\dagger) - \hat{\mathcal{N}}_{S \rightarrow E} \left(\frac{\mathbb{I}_S}{2^n} \right) \otimes \frac{\mathbb{I}_R}{2^k} \right\|_1 \\ & \leq \sqrt{2^{-H_2(S|E)\tau_{\text{SE}} - H_2(S|R)\rho_{\text{SR}}} + c \max_{A \subseteq [2N]} 2^{-H_2(A|R)\rho_{\text{AR}} - H_2(A|E)\tau_{\text{AE}}}}, \end{aligned} \quad (\text{B96})$$

where $\tau_{\text{SE}} = \mathcal{N}_{zz}(|\hat{\phi}\rangle\langle\hat{\phi}|^{\otimes n})$, $\rho_{S_i R_i} = |\hat{\phi}\rangle\langle\hat{\phi}|^{(k/n)\xi} \otimes |0^{(n-k)/n}\rangle\langle 0^{(n-k)/n}|$, $\tau_{\text{AE}} = \text{tr}_{\bar{A}} \tau_{\text{SE}}$, $\rho_{\text{SR}} = \bigotimes_{i=1}^{2N} \rho_{S_i R_i}$, and

$$c = 2 \left(\left(1 + 2\eta \max_i (\max(2^{-H_2(S_i|R_i)\rho_{S_i R_i}}, 2^{H_2(S_i|R_i)\rho_{S_i R_i}})) \right)^{N-1} - 1 \right). \quad (\text{B97})$$

Note that $H_2(S_i|R_i)_{\rho_{S_i R_i}} = -(k/n)\xi$ given by Eq. (B56). Thus, $H_2(A|R)_{\rho_{\text{AR}}} = -(k|A|/n)\xi$.

Also, from Eq. (B57) we can get

$$\begin{aligned} H_2(A|E)_{\tau_{\text{AE}}} &= -\log \text{tr}_{S'} \left[\text{tr}_A \sqrt{\text{tr}_{\bar{A}} \mathcal{N}_{zz} \left(|\hat{\phi}\rangle\langle\hat{\phi}|^{\otimes n} \right)} \right] \\ &= -\log \text{tr}_{A'} \left[\text{tr}_A \sqrt{\prod_{i: i \in S_A, i+1 \in S_{\bar{A}}} [(1-p)\mathcal{I} + p\mathcal{Z}_i] \prod_{\{i, i+1\} \subseteq S_A} [(1-p)\mathcal{I} + p\mathcal{Z}_i\mathcal{Z}_{i+1}]} \left(|\hat{\phi}\rangle_{AA'}\langle\hat{\phi}|_{AA'} \right)} \right] \\ &= |A| \xi - 2 \log \left(\sum_{\vec{Z} \in \{\mathcal{I}, \mathcal{Z}\}^{\otimes |A|\xi}} \text{Pr}(\vec{Z}) \right) \\ &\geq |A| \xi - 2 \log \left((\sqrt{1-p} + \sqrt{p})^{|A|(\xi+1)} \right) \\ &= \xi |A| \left[1 - 2(1 + \xi^{-1}) \log \left(\sqrt{1-p} + \sqrt{p} \right) \right] \\ &\approx \xi |A| \left[1 - 2 \log \left(\sqrt{1-p} + \sqrt{p} \right) \right]. \end{aligned} \quad (\text{B98})$$

Here, $\text{Pr}(\vec{Z})$ is the probability of pattern $\vec{Z}$ generated by $\prod_{i: i \in S_A, i+1 \in S_{\bar{A}}} [(1-p)\mathcal{I} + p\mathcal{Z}_i] \prod_{\{i, i+1\} \subseteq S_A} [(1-p)\mathcal{I} + p\mathcal{Z}_i\mathcal{Z}_{i+1}]$. Note that this probability can be obtained by coarse-graining the binomial distribution $B(|A|(\xi+1), p)$. This can be seen by counting the generators of this probability distribution,

$$|A|(\xi+1) \geq |\{i \in S_A, i+1 \in S_{\bar{A}}\}| + |\{i \in S_A, i+1 \in S_A\}|. \quad (\text{B99})$$

The coarse-graining operation can only decrease $\sum \text{Pr}(\vec{Z})$, which gives us the inequality in the fourth line.

By Eq. (B59), $c \leq \frac{4\varepsilon^{1-\frac{k}{n}} \frac{k}{n}}{\log(n/\varepsilon)}$. Thus, for sufficiently large n , we have $\xi^{-1} \approx 0$ and obtain that

$$\begin{aligned} & \mathbb{E}_U \left\| \hat{\mathcal{N}}_{S \rightarrow E} \left(U \left| \hat{\phi} \right\rangle \left\langle \hat{\phi} \right|_{\text{LR}} \otimes |0^{n-k}\rangle \langle 0^{n-k}| U^\dagger \right) - \hat{\mathcal{N}}_{S \rightarrow E} \left(\frac{\mathbb{I}_S}{2^n} \right) \otimes \frac{\mathbb{I}_R}{2^k} \right\|_1 \\ & \leq \sqrt{2^{-H_2(S|E)_{\text{TSE}} - H_2(S|R)_{\rho_{\text{SR}}}} + c \max_{A \subseteq [2N]} 2^{-H_2(A|R)_{\rho_{\text{AR}}} - H_2(A|E)_{\tau_{\text{AE}}}}} \\ & \leq \sqrt{2^{-n(1-2\log(\sqrt{1-p}+\sqrt{p})-\frac{k}{n})} + \frac{4\varepsilon^{1-\frac{k}{n}} \frac{k}{n}}{\log(n/\varepsilon)} \prod_{i=1}^{2N} \max(1, 2^{-\xi(1-2\log(\sqrt{1-p}+\sqrt{p})-\frac{k}{n})})}. \end{aligned} \quad (\text{B100})$$

Substituting the above equation into Eq. (B94) completes the proof. From the above equation, we have that as long as ε decays faster than $n^{-k/(n-k)}$, and encoding rate $k/n < 1 - 2\log(\sqrt{1-p} + \sqrt{p})$, the Choi error decays polynomially. Hence, we conclude that a 1D $O(\log(n))$ -depth circuit achieves an encoding rate $1 - 2\log(\sqrt{1-p} + \sqrt{p})$ for nearest-neighbor ZZ-coupling noise, which is the same with that of random Clifford encoding scheme, manifested by Eq. (B170). ■

10. Approximate quantum error correction against the fixed-number random erasure errors for $\mathfrak{C}_n^\varepsilon$

We first present the results of AQEC performance of $\mathfrak{C}_n^\varepsilon$ for random t -erasure error below. Then we give the full proof. The proof idea is similar to that of the nondecoupling theorem.

Theorem B1 (AQEC performance of $\mathfrak{C}_n^\varepsilon$ for random t -erasure error). In the large n limit, suppose k and t satisfy $1 - k/n - \log(1 + 3t/n) \geq 0$, and ε satisfies that $(\varepsilon/n)^{1-k/n} = o(1/n)$, then the expected Choi error of the random codes from $\mathfrak{C}_n^\varepsilon$ against the random t erasure errors satisfies

$$\mathbb{E} \epsilon_{\text{Choi}} \leq \left(2^{-(n-2t-k)} + \frac{4\varepsilon^{1-\frac{k}{n}} \frac{k}{n}}{\log(n/\varepsilon)} \right)^{\frac{1}{4}}. \quad (\text{B101})$$

Proof. Here, we analyze the random erasure error with a fixed number of erasures. We utilize the complementary channel to analyze the Choi error. Specifically,

$$\epsilon_{\text{Choi}} = \min_{\zeta} P \left((\widehat{\mathcal{N}} \circ \mathcal{E}_{L \rightarrow E} \otimes I_R) \left(\left| \hat{\phi} \right\rangle \left\langle \hat{\phi} \right|_{\text{LR}} \right), (\mathcal{T}_{L \rightarrow E}^\zeta \otimes I_R) \left(\left| \hat{\phi} \right\rangle \left\langle \hat{\phi} \right|_{\text{LR}} \right) \right). \quad (\text{B102})$$

Since $\mathcal{T}_{L \rightarrow E}^\zeta$ maps all states to a fixed state ζ , $(\mathcal{T}_{L \rightarrow E}^\zeta \otimes I_R) \left(\left| \hat{\phi} \right\rangle \left\langle \hat{\phi} \right|_{\text{LR}} \right) = \zeta_E \otimes \text{tr}_L \left(\left| \hat{\phi} \right\rangle \left\langle \hat{\phi} \right|_{\text{LR}} \right) = \zeta_E \otimes \mathbb{I}_R/2^k$. Meanwhile, $\widehat{\mathcal{N}} \circ \mathcal{E}_{L \rightarrow E} = \hat{\mathcal{N}}_{S \rightarrow E} \circ \mathcal{E}_{L \rightarrow S}$ with $\mathcal{E}_{L \rightarrow S}$ the encoding map satisfying

$$\mathcal{E}_{L \rightarrow S}(\rho) = U_S(\rho_L \otimes |0\rangle\langle 0|^{n-k})U_S^\dagger, \quad (\text{B103})$$

where U_S is the encoding unitary, generated by 1D $\log(n/\varepsilon)$ -depth Clifford circuits in our case. The term $|0\rangle\langle 0|^{n-k}$ is the ancillary qubits used for quantum error correction. For a fixed t -erasure errors, $\hat{\mathcal{N}}_{S \rightarrow E}(\rho) = \text{tr}_{n-t}(\rho)$. For a random t -erasure error, which is a mixing of t -erasure errors at different locations, $\hat{\mathcal{N}}_{S \rightarrow E}(\rho) = \frac{1}{\binom{n}{t}} \sum_{|T|=t} |T\rangle\langle T| \otimes \text{tr}_{S \setminus T}(\rho)$ [72] where $|T\rangle$ labels the location of the erasure errors. Hence,

$$\begin{aligned} \epsilon_{\text{Choi}} &= \min_{\zeta} P \left(\hat{\mathcal{N}}_{S \rightarrow E} \left(U \left| \hat{\phi} \right\rangle \left\langle \hat{\phi} \right|_{\text{LR}} \otimes |0^{n-k}\rangle \langle 0^{n-k}| U^\dagger \right), \zeta_E \otimes \frac{\mathbb{I}_R}{2^k} \right) \\ &\leq P \left(\hat{\mathcal{N}}_{S \rightarrow E} \left(U \left| \hat{\phi} \right\rangle \left\langle \hat{\phi} \right|_{\text{LR}} \otimes |0^{n-k}\rangle \langle 0^{n-k}| U^\dagger \right), \mathbb{E}_{U \sim \mathcal{U}_n} \hat{\mathcal{N}}_{S \rightarrow E} \left(U \left| \hat{\phi} \right\rangle \left\langle \hat{\phi} \right|_{\text{LR}} \otimes |0^{n-k}\rangle \langle 0^{n-k}| U^\dagger \right) \right) \\ &\quad + \min_{\zeta} P \left(\mathbb{E}_{U \sim \mathcal{U}_n} \hat{\mathcal{N}}_{S \rightarrow E} \left(U \left| \hat{\phi} \right\rangle \left\langle \hat{\phi} \right|_{\text{LR}} \otimes |0^{n-k}\rangle \langle 0^{n-k}| U^\dagger \right), \zeta_E \otimes \frac{\mathbb{I}_R}{2^k} \right) \end{aligned}$$

$$\begin{aligned}
&= P\left(\hat{\mathcal{N}}_{S \rightarrow E}(U|\hat{\phi}\rangle\langle\hat{\phi}|_{\text{LR}} \otimes |0^{n-k}\rangle\langle 0^{n-k}| U^\dagger), \hat{\mathcal{N}}_{S \rightarrow E}\left(\frac{\mathbb{I}_S}{2^n}\right) \otimes \frac{\mathbb{I}_R}{2^k}\right) + \min_{\zeta} P\left(\hat{\mathcal{N}}_{S \rightarrow E}\left(\frac{\mathbb{I}_S}{2^n}\right) \otimes \frac{\mathbb{I}_R}{2^k}, \zeta_E \otimes \frac{\mathbb{I}_R}{2^k}\right) \\
&= P\left(\hat{\mathcal{N}}_{S \rightarrow E}\left(U|\hat{\phi}\rangle\langle\hat{\phi}|_{\text{LR}} \otimes |0^{n-k}\rangle\langle 0^{n-k}| U^\dagger\right), \hat{\mathcal{N}}_{S \rightarrow E}\left(\frac{\mathbb{I}_S}{2^n}\right) \otimes \frac{\mathbb{I}_R}{2^k}\right). \tag{B104}
\end{aligned}$$

Here, in the first inequality, we utilize the triangle inequality of the purified distance. In the second equality, we utilize the property of the twirling over Haar random unitary gates. In the third equality, we take the state ζ as $\hat{\mathcal{N}}_{S \rightarrow E}(\mathbb{I}_S/2^n)$.

In our case, the encoding unitary is generated by random 1D $\log(n/\epsilon)$ -depth Clifford circuits with the ensemble denoted as $\mathfrak{C}_n^\epsilon$. Then, we aim to bound $\mathbb{E}_{U \sim \mathfrak{C}_n^\epsilon} \in \text{Choi}$. For simplicity, we omit $\mathfrak{C}_n^\epsilon$ in the subscript. Utilizing Eq. (15) we get

$$\begin{aligned}
\mathbb{E}_{U \in \text{Choi}} &\leq \mathbb{E}_U \sqrt{\left\| \hat{\mathcal{N}}_{S \rightarrow E}(U|\hat{\phi}\rangle\langle\hat{\phi}|_{\text{LR}} \otimes |0^{n-k}\rangle\langle 0^{n-k}| U^\dagger) - \hat{\mathcal{N}}_{S \rightarrow E}\left(\frac{\mathbb{I}_S}{2^n}\right) \otimes \frac{\mathbb{I}_R}{2^k} \right\|_1} \\
&\leq \sqrt{\mathbb{E}_U \left\| \hat{\mathcal{N}}_{S \rightarrow E}(U|\hat{\phi}\rangle\langle\hat{\phi}|_{\text{LR}} \otimes |0^{n-k}\rangle\langle 0^{n-k}| U^\dagger) - \hat{\mathcal{N}}_{S \rightarrow E}\left(\frac{\mathbb{I}_S}{2^n}\right) \otimes \frac{\mathbb{I}_R}{2^k} \right\|_1}. \tag{B105}
\end{aligned}$$

When the complementary channel is $\mathbb{E}_i |i\rangle\langle i| \otimes \hat{\mathcal{N}}_{S \rightarrow E}^i$ where $\hat{\mathcal{N}}_{S \rightarrow E}^i$ is the complementary channel of $\mathcal{N}_{S \rightarrow E}^i$, we have that

$$\begin{aligned}
&\left\| \hat{\mathcal{N}}_{S \rightarrow E}(U|\hat{\phi}\rangle\langle\hat{\phi}|_{\text{LR}} \otimes |0^{n-k}\rangle\langle 0^{n-k}| U^\dagger) - \hat{\mathcal{N}}_{S \rightarrow E}\left(\frac{\mathbb{I}_S}{2^n}\right) \otimes \frac{\mathbb{I}_R}{2^k} \right\|_1 \\
&= \mathbb{E}_i \left\| \hat{\mathcal{N}}_{S \rightarrow E}^i(U|\hat{\phi}\rangle\langle\hat{\phi}|_{\text{LR}} \otimes |0^{n-k}\rangle\langle 0^{n-k}| U^\dagger) - \hat{\mathcal{N}}_{S \rightarrow E}^i\left(\frac{\mathbb{I}_S}{2^n}\right) \otimes \frac{\mathbb{I}_R}{2^k} \right\|_1, \tag{B106}
\end{aligned}$$

due to that $\|\mathbb{E}_i |i\rangle\langle i| \otimes A_i\|_1 = \mathbb{E}_i \|A_i\|_1$. Below, we use $\mathbb{E}_{\mathcal{N}}$ to represent a case of mixing noise channels. For the mixing noise channel, we have

$$\mathbb{E}_{U \in \text{Choi}} \leq \sqrt{\mathbb{E}_U \mathbb{E}_{\mathcal{N}} \left\| \hat{\mathcal{N}}_{S \rightarrow E}(U|\hat{\phi}\rangle\langle\hat{\phi}|_{\text{LR}} \otimes |0^{n-k}\rangle\langle 0^{n-k}| U^\dagger) - \hat{\mathcal{N}}_{S \rightarrow E}\left(\frac{\mathbb{I}_S}{2^n}\right) \otimes \frac{\mathbb{I}_R}{2^k} \right\|_1}. \tag{B107}$$

For further evaluation, we use Φ_{Haar} to denote $\hat{\mathcal{N}}_{S \rightarrow E}(\mathbb{I}_S/2^n) \otimes \mathbb{I}_R/2^k$. Using $\|A\|_1 \leq \sqrt{|A|}\|A\|_2$, we further get

$$\begin{aligned}
\mathbb{E}_{U \in \text{Choi}} &\leq \sqrt{\mathbb{E}_U \mathbb{E}_{\mathcal{N}} \left\| \hat{\mathcal{N}}_{S \rightarrow E}(U|\hat{\phi}\rangle\langle\hat{\phi}|_{\text{LR}} \otimes |0^{n-k}\rangle\langle 0^{n-k}| U^\dagger) - \Phi_{\text{Haar}} \right\|_1} \\
&\leq \sqrt{\sqrt{\dim E \dim R} \mathbb{E}_{\mathcal{N}} \mathbb{E}_U \left\| \hat{\mathcal{N}}_{S \rightarrow E}(U|\hat{\phi}\rangle\langle\hat{\phi}|_{\text{LR}} \otimes |0^{n-k}\rangle\langle 0^{n-k}| U^\dagger) - \Phi_{\text{Haar}} \right\|_2} \\
&\leq (\dim E \dim R)^{\frac{1}{4}} \left(\mathbb{E}_{\mathcal{N}} \mathbb{E}_U \text{tr}(\hat{\mathcal{N}}_{S \rightarrow E}(U|\hat{\phi}\rangle\langle\hat{\phi}|_{\text{LR}} \otimes |0^{n-k}\rangle\langle 0^{n-k}| U^\dagger) - \Phi_{\text{Haar}})^2 \right)^{\frac{1}{4}} \\
&= (\dim E \dim R)^{\frac{1}{4}} \left(\mathbb{E}_{\mathcal{N}} \left[\text{tr} \mathbb{E}_U \left[\hat{\mathcal{N}}_{S \rightarrow E}(U|\hat{\phi}\rangle\langle\hat{\phi}|_{\text{LR}} \otimes |0^{n-k}\rangle\langle 0^{n-k}| U^\dagger) \right]^2 + \text{tr} \Phi_{\text{Haar}}^2 \right. \right. \\
&\quad \left. \left. - 2 \text{tr} \hat{\mathcal{N}}_{S \rightarrow E} \left(\mathbb{E}_U \left[U|\hat{\phi}\rangle\langle\hat{\phi}|_{\text{LR}} \otimes |0^{n-k}\rangle\langle 0^{n-k}| U^\dagger \right] \right) \Phi_{\text{Haar}} \right] \right)^{\frac{1}{4}}. \tag{B108}
\end{aligned}$$

In the last line, we expand the 2-norm. The error terms can be divided into two parts:

$$\text{tr} \mathbb{E}_U \left[\hat{\mathcal{N}}_{S \rightarrow E}(U|\hat{\phi}\rangle\langle\hat{\phi}|_{\text{LR}} \otimes |0^{n-k}\rangle\langle 0^{n-k}| U^\dagger) \right]^2 \tag{B109}$$

and

$$\text{tr} \Phi_{\text{Haar}}^2 - 2 \text{tr} \hat{\mathcal{N}}_{S \rightarrow E} \left(\mathbb{E}_U \left[U|\hat{\phi}\rangle\langle\hat{\phi}|_{\text{LR}} \otimes |0^{n-k}\rangle\langle 0^{n-k}| U^\dagger \right] \right) \Phi_{\text{Haar}}. \tag{B110}$$

Equations (B109) and (B110) can be interpreted as the error brought by the second and the first moments of random unitary operations, respectively. Provided that $\mathfrak{C}_n^\varepsilon$ forms a unitary 1-design, we have

$$\mathbb{E}_U[U|\hat{\phi}\rangle\langle\hat{\phi}|_{\text{LR}} \otimes |0^{n-k}\rangle\langle 0^{n-k}| U^\dagger] = \frac{\mathbb{I}_S}{2^n} \otimes \frac{\mathbb{I}_R}{2^k}. \quad (\text{B111})$$

Thus,

$$\text{tr}\Phi_{\text{Haar}}^2 - 2\text{tr}\hat{\mathcal{N}}_{S \rightarrow E} \left(\mathbb{E}_U \left[U|\hat{\phi}\rangle\langle\hat{\phi}|_{\text{LR}} \otimes |0^{n-k}\rangle\langle 0^{n-k}| U^\dagger \right] \right) \Phi_{\text{Haar}} = -\text{tr}\Phi_{\text{Haar}}^2, \quad (\text{B112})$$

where

$$\begin{aligned} \text{tr}\Phi_{\text{Haar}}^2 &= \text{tr} \left(\left[\hat{\mathcal{N}}_{S \rightarrow E} \left(\frac{\mathbb{I}_S}{2^n} \right) \right]^2 \otimes \frac{\mathbb{I}_R}{2^{2k}} \right) \\ &= \frac{1}{2^k} \text{tr} \left(\left[\hat{\mathcal{N}}_{S \rightarrow E} \left(\frac{\mathbb{I}_S}{2^n} \right) \right]^2 \right). \end{aligned} \quad (\text{B113})$$

For the erasure channel, $\hat{\mathcal{N}}_{S \rightarrow E} = \text{tr}_{n-t}$, $\text{tr}\Phi_{\text{Haar}}^2 = 1/2^{k+t} = 1/\dim E \dim R$, and $\mathbb{E}_\mathcal{N} \text{tr}\Phi_{\text{Haar}}^2 = 1/2^{k+t}$.

The remained part to bound $\mathbb{E}_{U \in \text{Choi}}$ is $\mathbb{E}_\mathcal{N} \text{tr} \mathbb{E}_U \left[\hat{\mathcal{N}}_{S \rightarrow E} \left(U|\hat{\phi}\rangle\langle\hat{\phi}|_{\text{LR}} \otimes |0^{n-k}\rangle\langle 0^{n-k}| U^\dagger \right) \right]^2$. We first evaluate $\text{tr} \mathbb{E}_U [\hat{\mathcal{N}}_{S \rightarrow E} (U|\hat{\phi}\rangle\langle\hat{\phi}|_{\text{LR}} \otimes |0^{n-k}\rangle\langle 0^{n-k}| U^\dagger)]^2$ and then take the expectation $\mathbb{E}_\mathcal{N}$. We use $\text{tr} A^2 = \text{tr} A^{\otimes 2} F$ to evaluate it with F being a SWAP operator. Hence,

$$\begin{aligned} &\text{tr} \mathbb{E}_U \left[\hat{\mathcal{N}}_{S \rightarrow E} \left(U|\hat{\phi}\rangle\langle\hat{\phi}|_{\text{LR}} \otimes |0^{n-k}\rangle\langle 0^{n-k}| U^\dagger \right) \right]^2 \\ &= \text{tr} \mathbb{E}_U \left[\left(\hat{\mathcal{N}}_{S \rightarrow E} \left(U|\hat{\phi}\rangle\langle\hat{\phi}|_{\text{LR}} \otimes |0^{n-k}\rangle\langle 0^{n-k}| U^\dagger \right) \right)^{\otimes 2} F_E \otimes F_R \right] \\ &= \text{tr} \mathbb{E}_U \left[U^{\otimes 2} \left(|\hat{\phi}\rangle\langle\hat{\phi}|_{\text{LR}} \otimes |0^{n-k}\rangle\langle 0^{n-k}| \right)^{\otimes 2} U^{\dagger \otimes 2} \right] (\hat{\mathcal{N}}_{S \rightarrow E}^{\dagger})^{\otimes 2} (F_E) \otimes F_R, \end{aligned} \quad (\text{B114})$$

where F_E denotes the SWAP operator acting on two copies of system E , and F_R is defined similarly. The term $\hat{\mathcal{N}}_{S \rightarrow E}^\dagger$ is the adjoint map of $\hat{\mathcal{N}}_{S \rightarrow E}$. To further evaluate the above quantity, we need to investigate the structure of $\mathfrak{C}_n^\varepsilon$. Note that the random encoding unitary U can be generated by two layers, as shown in Fig. 1(b). The qubits are divided into $2N$ parts, labeled from 1 to $2N$, with each part having an equal size of $\xi = n/2N$ qubits. The first layer of the encoding unitary is a tensor product of random Clifford gates acting on the $2i - 1$ and $2i$ parts, and the second layer is a tensor product of random Clifford gates acting on the $2i$ and $2i + 1$ parts. That is,

$$U = U_2 \cdot U_1, \quad (\text{B115})$$

where

$$U_1 = \bigotimes_{i=1}^N U_C^{2i-1, 2i}, \quad U_2 = \bigotimes_{i=1}^{N-1} U_C^{2i, 2i+1}. \quad (\text{B116})$$

The gate $U_C^{i, i+1}$ is uniformly and randomly sampled from the Clifford group in the regions i and $i + 1$. As a result,

$$\begin{aligned} &\text{tr} \mathbb{E}_U \left[\hat{\mathcal{N}}_{S \rightarrow E} \left(U|\hat{\phi}\rangle\langle\hat{\phi}|_{\text{LR}} \otimes |0^{n-k}\rangle\langle 0^{n-k}| U^\dagger \right) \right]^2 \\ &= \text{tr} \mathbb{E}_{U_1} \left[U_1^{\otimes 2} \left(|\hat{\phi}\rangle\langle\hat{\phi}|_{\text{LR}} \otimes |0^{n-k}\rangle\langle 0^{n-k}| \right)^{\otimes 2} U_1^{\dagger \otimes 2} \right] \mathbb{E}_{U_2} \left[U_2^{\dagger \otimes 2} (\hat{\mathcal{N}}_{S \rightarrow E}^\dagger)^{\otimes 2} (F_E) U_2^{\otimes 2} \right] \otimes F_R \\ &= \text{tr} \left(\text{tr}_R \left(\mathbb{E}_{U_1} \left[U_1^{\otimes 2} \left(|\hat{\phi}\rangle\langle\hat{\phi}|_{\text{LR}} \otimes |0^{n-k}\rangle\langle 0^{n-k}| \right)^{\otimes 2} U_1^{\dagger \otimes 2} \right] F_R \right) \mathbb{E}_{U_2} \left[U_2^{\dagger \otimes 2} (\hat{\mathcal{N}}_{S \rightarrow E}^\dagger)^{\otimes 2} (F_E) U_2^{\otimes 2} \right] \right). \end{aligned} \quad (\text{B117})$$

We make the two layers of random unitary gates act on different terms for further evaluation. We first focus on the first term,

$$\text{tr}_R \left(\mathbb{E}_{U_1} \left[U_1^{\otimes 2} \left(|\hat{\phi}\rangle\langle\hat{\phi}|_{\text{LR}} \otimes |0^{n-k}\rangle\langle 0^{n-k}| \right)^{\otimes 2} U_1^{\dagger \otimes 2} \right] F_R \right). \quad (\text{B118})$$

For simplicity, we consider each part to have the same number of encoding qubits and ancillary qubits. That is, in part i , there will be $(k/n)\xi = k/2N$ encoding qubits and $((n-k)/n)\xi = (n-k)/2N$ ancillary qubits. Thus, based on Eqs. (51) and (B116), we get

$$\begin{aligned} & \mathbb{E}_{U_1} \left[U_1^{\otimes 2} \left(|\hat{\phi}\rangle\langle\hat{\phi}|_{\text{LR}} \otimes |0^{n-k}\rangle\langle 0^{n-k}| \right)^{\otimes 2} U_1^{\dagger \otimes 2} \right] \\ &= \bigotimes_{i=1}^N \left(\frac{\mathbb{I}_{d_{S_{2i-1,2i}}^2}}{d_{S_{2i-1,2i}}^2} \otimes \frac{\mathbb{I}_{d_{R_{2i-1,2i}}^2}}{d_{R_{2i-1,2i}}^2} + \frac{F_{S_{2i-1,2i}} - \frac{\mathbb{I}_{d_{S_{2i-1,2i}}^2}}{d_{S_{2i-1,2i}}^2}}{d_{S_{2i-1,2i}}^2 - 1} \otimes \left(F_{R_{2i-1,2i}} - \frac{1}{d_{S_{2i-1,2i}}} \right) \frac{\mathbb{I}_{d_{R_{2i-1,2i}}^2}}{d_{R_{2i-1,2i}}^2} \right). \end{aligned} \quad (\text{B119})$$

Since $F_R = \bigotimes_{i=1}^N F_{R_{2i-1,2i}}$, we have

$$\begin{aligned} & \text{tr}_R \left(\mathbb{E}_{U_1} \left[U_1^{\otimes 2} \left(|\hat{\phi}\rangle\langle\hat{\phi}|_{\text{LR}} \otimes |0^{n-k}\rangle\langle 0^{n-k}| \right)^{\otimes 2} U_1^{\dagger \otimes 2} \right] F_R \right) \\ &= \bigotimes_{i=1}^N \left(\frac{\mathbb{I}_{d_{S_{2i-1,2i}}^2}}{d_{S_{2i-1,2i}}^2 d_{R_{2i-1,2i}}^2} + \frac{F_{S_{2i-1,2i}} - \frac{\mathbb{I}_{d_{S_{2i-1,2i}}^2}}{d_{S_{2i-1,2i}}^2}}{d_{S_{2i-1,2i}}^2 - 1} \left(1 - \frac{1}{d_{S_{2i-1,2i}} d_{R_{2i-1,2i}}} \right) \right) \\ &= \bigotimes_{i=1}^N \left(\frac{\mathbb{I}_{d_{S_{2i-1,2i}}^2}}{d_{S_i}^4 d_{R_i}^2} + \frac{F_{S_{2i-1,2i}} - \frac{\mathbb{I}_{d_{S_{2i-1,2i}}^2}}{d_{S_i}^2}}{d_{S_i}^4 - 1} \left(1 - \frac{1}{d_{S_i}^2 d_{R_i}^2} \right) \right) \\ &= \bigotimes_{i=1}^N (a \mathbb{I}_{d_{S_{2i-1}}^2} \otimes \mathbb{I}_{d_{S_{2i}}^2} + b F_{S_{2i-1}} \otimes F_{S_{2i}}) \\ &= \bigotimes_{i=1}^N (a \mathbb{I}_{2i-1} \mathbb{I}_{2i} + b F_{2i-1} F_{2i}). \end{aligned} \quad (\text{B120})$$

Here, the last line is a simple notation of the third line. That is, we use $\mathbb{I}_i$ and F_i to denote $\mathbb{I}_{d_{S_i}^2}$ and F_{S_i} , respectively, and we omit the tensor product; $d_{S_i} = 2^\xi$ represents the dimension of part i and correspondingly, $d_{R_i} = 2^{\frac{k}{n}\xi}$. The coefficients

$$\begin{aligned} a &= \frac{1}{d_{S_i}^4 d_{R_i}^2} - \frac{1}{d_{S_i}^2 (d_{S_i}^4 - 1)} \left(1 - \frac{1}{d_{S_i}^2 d_{R_i}^2} \right) \\ &= \frac{1}{2^{4\xi + \frac{2k}{n}\xi}} - \frac{1}{2^{2\xi} (2^{4\xi} - 1)} \left(1 - \frac{1}{2^{2\xi + \frac{2k}{n}\xi}} \right) \end{aligned} \quad (\text{B121})$$

and

$$\begin{aligned} b &= \frac{1}{d_{S_i}^4 - 1} \left(1 - \frac{1}{d_{S_i}^2 d_{R_i}^2} \right) \\ &= \frac{1}{2^{4\xi} - 1} \left(1 - \frac{1}{2^{2\xi + \frac{2k}{n}\xi}} \right). \end{aligned} \quad (\text{B122})$$

It is easy to see that $a \leq 2^{-4\xi - \frac{2k}{n}\xi}$ and $b \leq 2^{-4\xi}$.

Next, we consider the second term $\mathbb{E}_{U_2}[U_2^{\dagger\otimes 2}(\hat{\mathcal{N}}_{S \rightarrow E}^{\dagger})^{\otimes 2}(F_E)U_2^{\otimes 2}]$ with $\mathcal{N}$ a fixed t -erasure channel. Note that $(\hat{\mathcal{N}}_{S \rightarrow E}^{\dagger})^{\otimes 2}(F_E) = F_E \otimes \mathbb{I}_{\bar{E}}$, where $\mathbb{I}_{\bar{E}}$ denotes the identity operator on two copies of subsystem $\bar{E}$. Combined with Eqs. (51) and (B116), we have

$$\mathbb{E}_{U_2}[U_2^{\dagger\otimes 2}(\hat{\mathcal{N}}_{S \rightarrow E}^{\dagger})^{\otimes 2}(F_E)U_2^{\otimes 2}] = (F_E \otimes \mathbb{I}_{\bar{E}})_1 \bigotimes_{i=1}^{N-1} (a_{2i,2i+1} \mathbb{I}_{2i} \mathbb{I}_{2i+1} + b_{2i,2i+1} F_{2i} F_{2i+1}) (F_E \otimes \mathbb{I}_{\bar{E}})_{2N}, \quad (\text{B123})$$

where

$$a_{2i,2i+1} = \frac{2^{n_I^{2i} + n_I^{2i+1}}}{2^{2\xi}} \frac{2^{4\xi} - 2^{2n_F^{2i} + 2n_F^{2i+1}}}{2^{4\xi} - 1}, \quad b_{2i,2i+1} = \frac{2^{n_I^{2i} + n_I^{2i+1}} (2^{2n_F^{2i} + 2n_F^{2i+1}} - 1)}{2^{4\xi} - 1}. \quad (\text{B124})$$

Here, $(F_E \otimes \mathbb{I}_{\bar{E}})_i$ denotes the restriction of $F_E \otimes \mathbb{I}_{\bar{E}}$ is part i . Note that $(F_E \otimes \mathbb{I}_{\bar{E}})_i$ is a tensor product of identity operators and SWAP operators. The terms n_I^i and n_F^i denote the number of identity operators and SWAP operators in $(F_E \otimes \mathbb{I}_{\bar{E}})_i$, respectively. Hence, $n_I^i + n_F^i = \xi$ and $\sum_i n_F^i = t$.

Combined the results above, we get

$$\begin{aligned} & \text{tr} \mathbb{E}_U \left[\hat{\mathcal{N}}_{S \rightarrow E} \left(U \left| \hat{\phi} \right\rangle \left\langle \hat{\phi} \right|_{\text{LR}} \otimes |0^{n-k}\rangle \langle 0^{n-k}| U^\dagger \right) \right]^2 \\ &= \text{tr} \left(\text{tr}_R \left(\mathbb{E}_{U_1} \left[U_1^{\otimes 2} \left(\left| \hat{\phi} \right\rangle \left\langle \hat{\phi} \right|_{\text{LR}} \otimes |0^{n-k}\rangle \langle 0^{n-k}| \right)^{\otimes 2} U_1^{\dagger\otimes 2} \right] F_R \right) \mathbb{E}_{U_2} \left[U_2^{\dagger\otimes 2} (\hat{\mathcal{N}}_{S \rightarrow E}^{\dagger})^{\otimes 2} (F_E) U_2^{\otimes 2} \right] \right) \\ &= \text{tr} \left(\left(\bigotimes_{i=1}^N (a \mathbb{I}_{2i-1} \mathbb{I}_{2i} + b F_{2i-1} F_{2i}) \right) \left((F_E \otimes \mathbb{I}_{\bar{E}})_1 \bigotimes_{i=1}^{N-1} (a_{2i,2i+1} \mathbb{I}_{2i} \mathbb{I}_{2i+1} + b_{2i,2i+1} F_{2i} F_{2i+1}) (F_E \otimes \mathbb{I}_{\bar{E}})_{2N} \right) \right) \\ &= \begin{pmatrix} 2^{2n_I^{2N} + n_F^{2N}} & 2^{2n_I^{2N} + 2n_F^{2N}} \end{pmatrix} \left(\prod_{i=1}^{N-1} \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} \begin{pmatrix} 2^{2\xi} & 2^\xi \\ 2^\xi & 2^{2\xi} \end{pmatrix} \begin{pmatrix} a_{2i,2i+1} & 0 \\ 0 & b_{2i,2i+1} \end{pmatrix} \begin{pmatrix} 2^{2\xi} & 2^\xi \\ 2^\xi & 2^{2\xi} \end{pmatrix} \right) \begin{pmatrix} a 2^{2n_I^1 + n_F^1} \\ b 2^{n_I^1 + 2n_F^1} \end{pmatrix} \\ &= 2^n \begin{pmatrix} 2^{n_I^{2N}} & 2^{n_F^{2N}} \end{pmatrix} \left(\prod_{i=1}^{N-1} \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} \begin{pmatrix} u_{2i} u_{2i+1} & \eta(u_{2i} u_{2i+1} + v_{2i} v_{2i+1}) \\ \eta(u_{2i} u_{2i+1} + v_{2i} v_{2i+1}) & v_{2i} v_{2i+1} \end{pmatrix} \right) \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} \begin{pmatrix} 2^{n_I^1} \\ 2^{n_F^1} \end{pmatrix}. \quad (\text{B125}) \end{aligned}$$

Here, the third line is a direct substitution. The fourth line is done by direct calculation. Note that

$$\text{tr}_1((a \mathbb{I}_1 \mathbb{I}_2 + b F_1 F_2)(F_E \otimes \mathbb{I}_{\bar{E}})_1) = (a 2^{n_I^1 + n_F^1} \mathbb{I}_2 + b 2^{n_I^1 + 2n_F^1} F_2) \quad (\text{B126})$$

and

$$\text{tr}_1((x \mathbb{I}_1 + y F_1)(a \mathbb{I}_1 \mathbb{I}_2 + b F_1 F_2)) = (a 2^{2\xi} x + a 2^\xi y) \mathbb{I}_2 + (b 2^\xi x + b 2^{2\xi} y) F_2. \quad (\text{B127})$$

By using the above two equalities, we derive the fourth line. For the fifth line, we use that $n_I^i + n_F^i = \xi$ and $2N\xi = n$. Also, we calculate

$$\begin{aligned} & \begin{pmatrix} 2^\xi & 1 \\ 1 & 2^\xi \end{pmatrix} \begin{pmatrix} a_{2i,2i+1} & 0 \\ 0 & b_{2i,2i+1} \end{pmatrix} \begin{pmatrix} 2^\xi & 1 \\ 1 & 2^\xi \end{pmatrix} = \begin{pmatrix} 2^{2\xi} a_{2i,2i+1} + b_{2i,2i+1} & 2^\xi (a_{2i,2i+1} + b_{2i,2i+1}) \\ 2^\xi (a_{2i,2i+1} + b_{2i,2i+1}) & a_{2i,2i+1} + 2^{2\xi} b_{2i,2i+1} \end{pmatrix} \\ &= \begin{pmatrix} u_{2i} u_{2i+1} & \eta(u_{2i} u_{2i+1} + v_{2i} v_{2i+1}) \\ \eta(u_{2i} u_{2i+1} + v_{2i} v_{2i+1}) & v_{2i} v_{2i+1} \end{pmatrix}, \quad (\text{B128}) \end{aligned}$$

where

$$u_i = 2^{n_I^i}, \quad v_i = 2^{n_F^i}, \quad \eta = \frac{2^\xi}{2^{2\xi} + 1} \leq 2^{-\xi}. \quad (\text{B129})$$

Denoting $D_{2i,2i+1} = \begin{pmatrix} u_{2i}u_{2i+1} & 0 \\ 0 & v_{2i}v_{2i+1} \end{pmatrix}$, $E_{2i,2i+1} = \eta(u_{2i}u_{2i+1} + v_{2i}v_{2i+1}) \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, and $C = \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}$, we have

$$\text{tr } \mathbb{E}_U \left[\hat{\mathcal{N}}_{S \rightarrow E} \left(U \left| \hat{\phi} \right\rangle \left\langle \hat{\phi} \right|_{\text{LR}} \otimes \left| 0^{n-k} \right\rangle \left\langle 0^{n-k} \right| U^\dagger \right) \right]^2 = 2^n (u_{2N} \quad v_{2N}) \left(\prod_{i=1}^{N-1} C(D_{2i,2i+1} + E_{2i,2i+1}) \right) C \begin{pmatrix} u_1 \\ v_1 \end{pmatrix}. \quad (\text{B130})$$

This equation is similar to $(D + E)^{N-1}$, which enables us to expand it according to the number of “ E .” We first consider the expansion term with no E . That is,

$$\begin{aligned} 2^n (u_{2N} \quad v_{2N}) \left(\prod_{i=1}^{N-1} C D_{2i,2i+1} \right) C \begin{pmatrix} u_1 \\ v_1 \end{pmatrix} &= 2^n \left(a^N \prod_{i=1}^{2N} u_i + b^N \prod_{i=1}^{2N} v_i \right) \\ &= 2^n (a^N 2^{n-t} + b^N 2^t) \\ &\leq 2^n (2^{-2n-k} 2^{n-t} + 2^{-2n} 2^t) \\ &= 2^{-t-k} + 2^{-n+t}. \end{aligned} \quad (\text{B131})$$

The first term 2^{-t-k} in this equation equals $\text{tr } \Phi_{\text{Haar}}^2$ and will be eliminated in the final result. The second term 2^{-n+t} exponentially decays with respect to n . Combining with the coefficient $\dim E \dim R = 2^{t+k}$, this is $2^{-(n-2t-k)}$ corresponding to a result of random Clifford encoding, as demonstrated in Appendix B 12. In fact, if the encoding unitary is drawn from the whole unitary group instead of $\mathfrak{C}_n^\varepsilon$, η will quickly decay to 0, and it is unnecessary to consider E in the expansion. The expansion terms with E can be viewed as extra errors.

Below, we consider the case there are g numbers of E in the expansion. We set the location of E as $j_1, j_2, \dots, j_g$ with $1 \leq j_1 < j_2 < \dots < j_g \leq N - 1$. That is,

$$\begin{aligned} &2^n (u_{2N} \quad v_{2N}) \left(\prod_{i=j_g+1}^{N-1} C D_{2i,2i+1} \right) C E_{2j_g,2j_g+1} \cdots C E_{2j_2,2j_2+1} \left(\prod_{i=j_1+1}^{j_2-1} C D_{2i,2i+1} \right) C E_{2j_1,2j_1+1} \left(\prod_{i=1}^{j_1-1} C D_{2i,2i+1} \right) C \begin{pmatrix} u_1 \\ v_1 \end{pmatrix} \\ &= 2^n \eta^g \prod_{r=1}^g (u_{2j_r} u_{2j_r+1} + v_{2j_r} v_{2j_r+1}) \\ &\quad \times (b^{N-j_g} v_{2j_g+2 \rightarrow 2N} \quad a^{N-j_g} u_{2j_g+2 \rightarrow 2N}) \left(\prod_{s=1}^{g-1} \begin{pmatrix} 0 & a^{j_{s+1}-j_s} u_{2j_s+2 \rightarrow 2j_{s+1}-1} \\ b^{j_{s+1}-j_s} v_{2j_s+2 \rightarrow 2j_{s+1}-1} & 0 \end{pmatrix} \right) \begin{pmatrix} a^{j_1} u_{1 \rightarrow 2j_1-1} \\ b^{j_1} v_{1 \rightarrow 2j_1-1} \end{pmatrix} \\ &= 2^n \eta^g \prod_{r=1}^g (u_{2j_r} u_{2j_r+1} + v_{2j_r} v_{2j_r+1}) (a^{-j_1+j_2-j_3+\dots} b^{j_1-j_2+j_3-\dots} u_{2j_1+2 \rightarrow 2j_2-1, 2j_3+2 \rightarrow 2j_4-1, \dots} v_{1 \rightarrow 2j_1-1, 2j_2+2 \rightarrow 2j_3-1, \dots} \\ &\quad + a^{j_1-j_2+j_3-\dots} b^{-j_1+j_2-j_3+\dots} u_{1 \rightarrow 2j_1-1, 2j_2+2 \rightarrow 2j_3-1, \dots} v_{2j_1+2 \rightarrow 2j_2-1, 2j_3+2 \rightarrow 2j_4-1, \dots}) \\ &= 2^n \eta^g \prod_{r=1}^g \left(\frac{2^{2\xi}}{v_{2j_r} v_{2j_r+1}} + v_{2j_r} v_{2j_r+1} \right) \\ &\quad \times \begin{cases} a^{N-j} b^j 2^{(2N-2j-g)\xi} \frac{v_{1 \rightarrow 2j_1-1, 2j_2+2 \rightarrow 2j_3-1, \dots, 2j_{g-1}+2 \rightarrow 2j_g-1}}{v_{2j_1+2 \rightarrow 2j_2-1, \dots, 2j_{g-2}+2 \rightarrow 2j_{g-1}-1, 2j_g+2 \rightarrow 2N}} \\ + a^j b^{N-j} 2^{(2j-g)\xi} \frac{v_{2j_1+2 \rightarrow 2j_2-1, \dots, 2j_{g-2}+2 \rightarrow 2j_{g-1}-1, 2j_g+2 \rightarrow 2N}}{v_{1 \rightarrow 2j_1-1, 2j_2+2 \rightarrow 2j_3-1, \dots, 2j_{g-1}+2 \rightarrow 2j_g-1}} \\ a^{N-j} b^j 2^{(2N-2j-g)\xi} \frac{v_{2j_1+2 \rightarrow 2j_2-1, 2j_3+2 \rightarrow 2j_4-1, \dots, 2j_{g-1}+2 \rightarrow 2j_g-1}}{v_{1 \rightarrow 2j_1-1, \dots, 2j_{g-1}+2 \rightarrow 2j_g-1, 2j_g+2 \rightarrow 2N}} \\ + a^j b^{N-j} 2^{(2j-g)\xi} \frac{v_{1 \rightarrow 2j_1-1, \dots, 2j_{g-1}+2 \rightarrow 2j_g-1, 2j_g+2 \rightarrow 2N}}{v_{2j_1+2 \rightarrow 2j_2-1, 2j_3+2 \rightarrow 2j_4-1, \dots, 2j_{g-1}+2 \rightarrow 2j_g-1}} \end{cases}, \quad j = \sum_{s=1}^g (-1)^{s+1} j_s, \quad \text{if } \text{mod}(g, 2) = 1, \\ &\quad j = \sum_{s=1}^g (-1)^s j_s, \quad \text{if } \text{mod}(g, 2) = 0 \end{aligned}$$

$$\begin{aligned}
&= 2^{n-t} \eta^g \prod_{r=1}^g (2^{2\xi} + v_{2j_r}^2 v_{2j_r+1}^2) \\
&\quad \times \begin{cases} a^{N-j} b^j 2^{(2N-2j-g)\xi} v_{1 \rightarrow 2j_1-1, 2j_2+2 \rightarrow 2j_3-1, \dots, 2j_{g-1}+2 \rightarrow 2j_g-1}^2 \\ \quad + a^j b^{N-j} 2^{(2j-g)\xi} v_{2j_1+2 \rightarrow 2j_2-1, \dots, 2j_{g-2}+2 \rightarrow 2j_{g-1}-1, 2j_g+2 \rightarrow 2N}^2, & j = \sum_{s=1}^g (-1)^{s+1} j_s, \quad \text{if } \text{mod}(g, 2) = 1, \\ a^{N-j} b^j 2^{(2N-2j-g)\xi} v_{2j_1+2 \rightarrow 2j_2-1, 2j_3+2 \rightarrow 2j_4-1, \dots, 2j_{g-1}+2 \rightarrow 2j_g-1}^2 \\ \quad + a^j b^{N-j} 2^{(2j-g)\xi} v_{1 \rightarrow 2j_1-1, \dots, 2j_{g-1}+2 \rightarrow 2j_g-1, 2j_g+2 \rightarrow 2N}^2, & j = \sum_{s=1}^g (-1)^s j_s, \quad \text{if } \text{mod}(g, 2) = 0. \end{cases}
\end{aligned} \tag{B132}$$

Here, $u_{i_1 \rightarrow j_1, \dots, i_s \rightarrow j_s}$ means $\prod_{r=1}^s \prod_{t=i_r}^{j_r} u_t$. The meaning is the same for $v_{i_1 \rightarrow j_1, \dots, i_s \rightarrow j_s}$. We also use $t = \sum_{i=1}^{2N} n_F^i$ or $2^t = \prod_{i=1}^{2N} v_i$ in the above derivation.

Now, we come back to considering the condition of random erasure noise channels and take the expectation $\mathbb{E}_{\mathcal{N}}$. Note that the erasure channel $\mathcal{N}$ only influences the value of v_i . For a quantity $v_{i_1 \rightarrow j_1, \dots, i_s \rightarrow j_s}$, we have

$$\mathbb{E}_{\mathcal{N}} v_{i_1 \rightarrow j_1, \dots, i_s \rightarrow j_s}^2 = \mathbb{E}_{\mathcal{N}} 2^{2 \sum_{r=1}^s \sum_{t=i_r}^{j_r} n_F^t}. \tag{B133}$$

Denote $X = \sum_{r=1}^s \sum_{t=i_r}^{j_r} n_F^t$. This is a random variable representing the number of erasures in the region $\{i_1, i_1 + 1 \dots, j_1, i_2, i_2 + 1 \dots, j_2, \dots, i_s, i_s + 1, \dots, j_s\}$. Since we consider the t erasures to be equally distributed on all of the qubits, the random variable X satisfies

$$\Pr_{\mathcal{N}}(X = x) = \frac{\binom{m}{x} \binom{n-m}{t-x}}{\binom{n}{t}} = \frac{\binom{t}{x} \binom{n-t}{m-x}}{\binom{n}{m}}, \tag{B134}$$

where $m = |\{i_1, i_1 + 1 \dots, j_1, i_2, i_2 + 1 \dots, j_2, \dots, i_s, i_s + 1, \dots, j_s\}| \xi$. We prove that

$$\begin{aligned}
\mathbb{E}_{\mathcal{N}} 2^{2X} &= \sum_{x=\max(0, m+t-n)}^{\max(m, t)} 2^{2x} \frac{\binom{t}{x} \binom{n-t}{m-x}}{\binom{n}{m}} \\
&\leq \left(1 + \frac{3t}{n}\right)^m \\
&\leq 2^{m \log\left(1 + \frac{3t}{n}\right)}.
\end{aligned} \tag{B135}$$

This can be seen that $\sum_{x=\max(0, m+t-n)}^{\max(m, t)} 2^{2x} \binom{t}{x} \binom{n-t}{m-x}$ is the coefficient of z^m in the expansion of the function $(1 + 4z)^t (1 + z)^{n-t}$. With Lemma B2 proved later, the coefficients of z^m of $(1 + 4z)^t (1 + z)^{n-t}$ is not larger than that of $(1 + ((4t + n - t)/n)z)^n$, of which the coefficient of z^m is $(1 + 3t/n)^m \binom{n}{m}$.

Thus, under the expectation of the location of the erasures, we get

$$\begin{aligned}
&\mathbb{E}_{\mathcal{N}} 2^n (u_{2N} \quad v_{2N}) \left(\prod_{i=j_g+1}^{N-1} CD_{2i, 2i+1} \right) CE_{2j_g, 2j_g+1} \cdots CE_{2j_2, 2j_2+1} \left(\prod_{i=j_1+1}^{j_2-1} CD_{2i, 2i+1} \right) CE_{2j_1, 2j_1+1} \left(\prod_{i=1}^{j_1-1} CD_{2i, 2i+1} \right) C \begin{pmatrix} u_1 \\ v_1 \end{pmatrix} \\
&\leq 2^{n-t} \eta^g \sum_{h=0}^g \binom{g}{h} (a^{N-j} b^j 2^{(2N-2j-g+2h)\xi + (2j+g-2h)\xi} \log\left(1 + \frac{3t}{n}\right) + a^j b^{N-j} 2^{(2j-g+2h)\xi + (2N-2j+g-2h)\xi} \log\left(1 + \frac{3t}{n}\right)) \\
&\leq 2^{n-t} 2^{-g\xi} \sum_{h=0}^g \binom{g}{h} (2^{-2n - \frac{N-j}{N}k} 2^{(2N-2j-g+2h)\xi + (2j+g-2h)\xi} \log\left(1 + \frac{3t}{n}\right) + 2^{-2n - \frac{j}{N}k} 2^{(2j-g+2h)\xi + (2N-2j+g-2h)\xi} \log\left(1 + \frac{3t}{n}\right))
\end{aligned}$$

$$\begin{aligned}
&\leq 2^{-t-k} 2^{-\frac{g}{2N}n} \sum_{h=0}^g \binom{g}{h} (2^{\frac{2j\xi}{n}} k 2^{-(2j+g-2h)\xi+(2j+g-2h)\xi \log\left(1+\frac{3t}{n}\right)} + 2^{(1-\frac{2j\xi}{n})k} 2^{-(2N-2j+g-2h)\xi+(2N-2j+g-2h)\xi \log\left(1+\frac{3t}{n}\right)}) \\
&= 2^{-t-k} 2^{-\frac{g}{2N}(n-k)} \sum_{h=0}^g \binom{g}{h} (2^{\frac{(2j-g)\xi}{n}} k 2^{-(2j+g-2h)\xi+(2j+g-2h)\xi \log\left(1+\frac{3t}{n}\right)} \\
&\quad + 2^{(1-\frac{(2j+g)\xi}{n})k} 2^{-(2N-2j+g-2h)\xi+(2N-2j+g-2h)\xi \log\left(1+\frac{3t}{n}\right)}) \\
&= 2^{-t-k} 2^{-\frac{g}{2N}(n-k)} \sum_{h=0}^g \binom{g}{h} (2^{\frac{(2h-2g)\xi}{n}} k 2^{-(2j+g-2h)\xi(1-\frac{k}{n}-\log\left(1+\frac{3t}{n}\right))} + 2^{\frac{(2h-2g)\xi}{n}} k 2^{-(2N-2j+g-2h)\xi(1-\frac{k}{n}-\log\left(1+\frac{3t}{n}\right))}). \quad (\text{B136})
\end{aligned}$$

Note that $j = \sum_{s=1}^g (-1)^{s+1} j_s$ or $j = \sum_{s=1}^g (-1)^s j_s$ imply $j \geq g/2$. As long as

$$1 - \frac{k}{n} - \log\left(1 + \frac{3t}{n}\right) \geq 0, \quad (\text{B137})$$

we have

$$\begin{aligned}
&\mathbb{E}_{\mathcal{N}} 2^n (u_{2N} \quad v_{2N}) \left(\prod_{i=j_g+1}^{N-1} CD_{2i,2i+1} \right) CE_{2j_g,2j_g+1} \cdots CE_{2j_2,2j_2+1} \left(\prod_{i=j_1+1}^{j_2-1} CD_{2i,2i+1} \right) CE_{2j_1,2j_1+1} \left(\prod_{i=1}^{j_1-1} CD_{2i,2i+1} \right) C \begin{pmatrix} u_1 \\ v_1 \end{pmatrix} \\
&\leq 2^{1-t-k} 2^{-\frac{g}{2N}(n-k)} \sum_{h=0}^g \binom{g}{h} 2^{\frac{(2h-2g)\xi}{n}k} \\
&= 2^{1-t-k} (2^{-\xi(1+\frac{k}{n})} + 2^{-\xi(1-\frac{k}{n})})^g. \quad (\text{B138})
\end{aligned}$$

Combining Eqs. (B131) and (B138), we obtain

$$\begin{aligned}
&\mathbb{E}_{\mathcal{N}} \text{tr} \mathbb{E}_U \left[\hat{\mathcal{N}}_{S \rightarrow E} \left(U \left| \hat{\phi} \right\rangle \left\langle \hat{\phi} \right|_{\text{LR}} \otimes |0^{n-k}\rangle \langle 0^{n-k}| U^\dagger \right) \right]^2 \\
&= 2^n (u_{2N} \quad v_{2N}) \left(\prod_{i=1}^{N-1} C(D_{2i,2i+1} + E_{2i,2i+1}) \right) C \begin{pmatrix} u_1 \\ v_1 \end{pmatrix} \\
&\leq 2^{-t-k} + 2^{-n+t} + \sum_{g=1}^{N-1} \binom{N-1}{g} 2^{1-t-k} (2^{-\xi(1+\frac{k}{n})} + 2^{-\xi(1-\frac{k}{n})})^g \\
&= 2^{-t-k} + 2^{-n+t} + 2^{1-t-k} ((1 + 2^{-\xi(1+\frac{k}{n})} + 2^{-\xi(1-\frac{k}{n})})^{N-1} - 1). \quad (\text{B139})
\end{aligned}$$

Note that $\xi = \log(n/\varepsilon)$ and $N = n/2\xi$. Utilizing $(1+x/n)^n \leq e^x \leq 1+2x$ for sufficiently small x , we get

$$\begin{aligned}
(1 + 2^{-\xi(1+\frac{k}{n})} + 2^{-\xi(1-\frac{k}{n})})^{N-1} - 1 &\leq 2N(2^{-\xi(1+\frac{k}{n})} + 2^{-\xi(1-\frac{k}{n})}) \\
&\leq 4N2^{-\xi(1-\frac{k}{n})} \\
&= \frac{2n}{\log(n/\varepsilon)} 2^{-\log(n/\varepsilon)(1-\frac{k}{n})} \\
&= \frac{2\varepsilon^{1-\frac{k}{n}} n^{\frac{k}{n}}}{\log(n/\varepsilon)}. \quad (\text{B140})
\end{aligned}$$

The condition that x is small can be achieved by $(\varepsilon/n)^{1-k/n} = o(1/n)$.

Integrating Eqs. (B108), (B112), and (B139), we obtain

$$\begin{aligned} \mathbb{E}_{U \in \text{Choi}} &\leq \left(2^{t+k} (-2^{-t-k} + 2^{-t-k} + 2^{-n+t} + 2^{1-t-k} ((1 + 2^{-\xi(1+\frac{k}{n})} + 2^{-\xi(1-\frac{k}{n})})^{N-1} - 1)) \right)^{\frac{1}{4}} \\ &= \left(2^{-(n-2t-k)} + 2((1 + 2^{-\xi(1+\frac{k}{n})} + 2^{-\xi(1-\frac{k}{n})})^{N-1} - 1) \right)^{\frac{1}{4}} \\ &\leq \left(2^{-(n-2t-k)} + \frac{4\varepsilon^{1-\frac{k}{n}} n^{\frac{k}{n}}}{\log(n/\varepsilon)} \right)^{\frac{1}{4}}. \end{aligned} \quad (\text{B141})$$

Proof is completed. ■

Lemma B2. Given $a, b > 0$, $m, n \in \mathbb{Z}_+$, the coefficient of z^t in the expansion of $(1 + az)^m(1 + bz)^n$ is not larger than that in the expansion of $(1 + (am + bn)z/(m + n))^{m+n}$ for any $0 \leq t \leq m + n$.

Proof. We prove Lemma B2 by induction. For any $a, b > 0$, $m \in \mathbb{Z}_+$, when $n = 1$, the coefficient of z^t is

$$\binom{m}{t} a^t + \binom{m}{t-1} a^{t-1} b, \quad (\text{B142})$$

for $(1 + az)^m(1 + bz)^n$, and is

$$\binom{m+1}{t} \left(\frac{ma+b}{m+1} \right)^t \quad (\text{B143})$$

for $(1 + (am + bn)z/(m + n))^{m+n}$; t takes values from 0 to $m + 1$. Now we need to prove

$$\binom{m}{t} a^t + \binom{m}{t-1} a^{t-1} b \leq \binom{m+1}{t} \left(\frac{ma+b}{m+1} \right)^t, \quad (\text{B144})$$

which is equivalent to

$$1 + \frac{t}{m+1} \left(\frac{b}{a} - 1 \right) \leq \left(1 + \frac{1}{m+1} \left(\frac{b}{a} - 1 \right) \right)^t. \quad (\text{B145})$$

Note that $(1/(m+1))(b/a - 1) > -1$ and when $x > -1$, $(1+x)^r \geq 1+rx$ for any $r \geq 1$ and $r = 0$. This can be obtained by considering $g(r) = (1+x)^r - (1+rx)$. First, $g(0) = g(1) = 0$. Meanwhile, for $r \geq 1$,

$$\begin{aligned} g'(r) &= (1+x)^r \ln(1+x) - x \\ &\geq (1+x)^r \frac{x}{1+x} - x \\ &= ((1+x)^{r-1} - 1)x \\ &\geq 0. \end{aligned} \quad (\text{B146})$$

In the second line, we use the inequality $\ln(1+x) \geq x/(1+x)$. Thus, for any $r \geq 1$, $g(r) \geq 0$. As a consequence, Eq. (B145) is correct, and we prove the case for $n = 1$.

Now assuming that for any $a, b > 0$ and $m \in \mathbb{Z}_+$, the claim is correct for $n \leq n_0$, we try to prove the claim for $n = n_0 + 1$. First, using the case (m, n_0) , we have

$$(1 + az)^m(1 + bz)^{n_0} \leq \left(1 + \frac{am + bn_0}{m + n_0} z \right)^{m+n_0}. \quad (\text{B147})$$

Here, $p_1(z) \leq p_2(z)$ means that each coefficient of z^t in $p_1(z)$ is less or equal to that in $p_2(z)$. Next, using the case $(m + n_0, 1)$, we have

$$\begin{aligned} &\left(1 + \frac{am + bn_0}{m + n_0} z \right)^{m+n_0} (1 + bz) \\ &\leq \left(1 + \frac{a(m+n_0) + bn_0}{m + 1 + n_0} z \right)^{m+n_0+1}. \end{aligned} \quad (\text{B148})$$

Combining the equations above, we prove the case for $n = n_0 + 1$. The whole proof is done by induction. ■

11. Proof of Proposition 1

In this part, we prove Proposition 1, which shows that applying random Pauli matrices before encoding and after decoding ensures that the recovery fidelity for any arbitrary input state is lower-bounded by the Choi fidelity.

Proof of Proposition 1. Consider the Stinespring dilation of the channel $\mathcal{N} \circ \mathcal{E}$, which can be represented as an isometry $V_{\mathcal{N}}$ from L to SE . The pure state $V_{\mathcal{N}}|\psi\rangle_{\text{LR}}$ is a purification of $\mathcal{N} \circ \mathcal{E}(\psi_{\text{LR}})$. A recovery channel $\mathcal{D}$ can also be represented as a unitary $U_{\mathcal{D}}$ acting on S and an ancillary system A . By Uhlmann's theorem [27,84], the Choi fidelity can be expressed as

$$\begin{aligned}
F_{\text{Choi}} &= \max_{U'_{(S \setminus L), AE}} \left| \langle \phi |_{\text{LR}} \otimes \langle 0 |_{(S \setminus L), AE} U'_{(S \setminus L), AE} U_{\mathcal{D}} V_{\mathcal{N}} (|\phi\rangle_{\text{LR}} \otimes |0\rangle_A) \right| \\
&= \frac{1}{d_L} \max_{U'_{(S \setminus L), AE}} \left| \text{tr}_L [(\langle 0 |_{(S \setminus L), AE} U'_{(S \setminus L), AE} U_{\mathcal{D}} V_{\mathcal{N}} (I_L \otimes |0\rangle_A))] \right|. \tag{B149}
\end{aligned}$$

The recovery fidelity for input state ψ using the recovery channel $\mathcal{D}$ can be written as

$$F_{\psi} = \max_{U'_{(S \setminus L), AE}} \left| \langle \psi |_{\text{LR}} \otimes \langle 0 |_{(S \setminus L), AE} U'_{(S \setminus L), AE} U_{\mathcal{D}} V_{\mathcal{N}} (|\psi\rangle_{\text{LR}} \otimes |0\rangle_A) \right|. \tag{B150}$$

Now, we evaluate the expected recovery fidelity when applying a random Pauli matrix. The expression for the expected recovery fidelity is given by

$$\begin{aligned}
\mathbb{E}_{P_L} F_{\psi} &= \mathbb{E}_{P_L} \max_{U'_{(S \setminus L), AE}} \left| \langle \psi |_{\text{LR}} \otimes \langle 0 |_{(S \setminus L), AE} U'_{(S \setminus L), AE} P_L^{\dagger} U_{\mathcal{D}} V_{\mathcal{N}} P_L (|\psi\rangle_{\text{LR}} \otimes |0\rangle_A) \right| \\
&\geq \max_{U'_{(S \setminus L), AE}} \mathbb{E}_{P_L} \left| \langle \psi |_{\text{LR}} \otimes \langle 0 |_{(S \setminus L), AE} U'_{(S \setminus L), AE} P_L^{\dagger} U_{\mathcal{D}} V_{\mathcal{N}} P_L (|\psi\rangle_{\text{LR}} \otimes |0\rangle_A) \right| \\
&\geq \max_{U'_{(S \setminus L), AE}} \left| \mathbb{E}_{P_L} (\langle \psi |_{\text{LR}} \otimes \langle 0 |_{(S \setminus L), AE} U'_{(S \setminus L), AE} P_L^{\dagger} U_{\mathcal{D}} V_{\mathcal{N}} P_L (|\psi\rangle_{\text{LR}} \otimes |0\rangle_A) \right| \\
&= \frac{1}{d_L} \max_{U'_{(S \setminus L), AE}} \left| \text{tr}_L [(\langle 0 |_{(S \setminus L), AE} U'_{(S \setminus L), AE} U_{\mathcal{D}_0} V_{\mathcal{N}} (I_L \otimes |0\rangle_A))] \right| \\
&= F_{\text{Choi}}. \tag{B151}
\end{aligned}$$

In the first line, we apply P_L before encoding and set $P_L^{\dagger} U_{\mathcal{D}}$ as the recovery channel. The second line swaps the maximum outside, which reduces the average fidelity. In the third line, we move the expectation inside the absolute value, further reducing the fidelity. Finally, the fourth line applies Pauli twirling. Specifically, for any state $|\psi_{\text{LR}}\rangle$, we have

$$\mathbb{E}_{P_L} [P_L \psi_{\text{LR}} P_L^{\dagger}] = \text{tr}_L(\psi_{\text{LR}}) \otimes \frac{I_L}{d_L}. \tag{B152}$$

■

12. The Choi error of random Clifford encoding

In this part, we introduce the bound of the Choi error of random Clifford encoding. The results can be viewed as a baseline to measure the performance of other encoding methods. Note that the random Clifford encoding is the random stabilizer code [19], which is a good code with a constant encoding rate and linear code distance in general.

Recall that the Choi error is bounded by

$$\mathbb{E}_{C \in \text{Choi}} \leq \sqrt{\mathbb{E}_C \left\| \hat{\mathcal{N}}_{S \rightarrow E} \left(U |\hat{\phi}\rangle\langle\hat{\phi}|_{\text{LR}} \otimes |0^{n-k}\rangle\langle 0^{n-k}| U^{\dagger} \right) - \hat{\mathcal{N}}_{S \rightarrow E} \left(\frac{\mathbb{I}_S}{2^n} \right) \otimes \frac{\mathbb{I}_R}{2^k} \right\|_1}, \tag{B153}$$

where $\mathbb{E}_C$ is the expectation over the n -qubit Clifford group. We can apply the nonsmooth decoupling theorem for the 2-design group, or Lemma A5 for further evaluation,

$$\mathbb{E}_C \left\| \hat{\mathcal{N}}_{S \rightarrow E} \left(U |\hat{\phi}\rangle\langle\hat{\phi}|_{\text{LR}} \otimes |0^{n-k}\rangle\langle 0^{n-k}| U^{\dagger} \right) - \hat{\mathcal{N}}_{S \rightarrow E} \left(\frac{\mathbb{I}_S}{2^n} \right) \otimes \frac{\mathbb{I}_R}{2^k} \right\|_1 \leq 2^{-\frac{1}{2} H_2(S|R)} |\hat{\phi}\rangle\langle\hat{\phi}|_{\text{LR}} \otimes |0^{n-k}\rangle\langle 0^{n-k}| - \frac{1}{2} H_2(S|E) \hat{\mathcal{N}}_{S' \rightarrow E}(|\hat{\phi}\rangle\langle\hat{\phi}|_{\text{SS}'}). \tag{B154}$$

Another choice is the smooth decoupling theorem shown in Lemma A5:

$$\begin{aligned} \mathbb{E}_C \left\| \hat{\mathcal{N}}_{S \rightarrow E} \left(U \left| \hat{\phi} \right\rangle \left\langle \hat{\phi} \right|_{\text{LR}} \otimes |0^{n-k}\rangle \langle 0^{n-k}| U^\dagger \right) - \hat{\mathcal{N}}_{S \rightarrow E} \left(\frac{\mathbb{I}_S}{2^n} \right) \otimes \frac{\mathbb{I}_R}{2^k} \right\|_1 \\ \leq 2^{-\frac{1}{2} H_2(S|R)} |\hat{\phi}\rangle \langle \hat{\phi}|_{\text{LR}} \otimes |0^{n-k}\rangle \langle 0^{n-k}| - \frac{1}{2} H_{\min}^\delta(S|E) \hat{\mathcal{N}}_{S \rightarrow E}(|\hat{\phi}\rangle \langle \hat{\phi}|_{SS'}) + 8\delta. \end{aligned} \quad (\text{B155})$$

Note that we only smooth $H_2(S|E)$ with $H_{\min}^\delta(S|E)$ while maintaining $H_2(S|R)$ to evaluate a tighter bound like what did in the proof of Corollary 2. Particularly,

$$\begin{aligned} H_2(S|R) |\hat{\phi}\rangle \langle \hat{\phi}|_{\text{LR}} \otimes |0^{n-k}\rangle \langle 0^{n-k}| &= H_2(L|R) |\hat{\phi}\rangle \langle \hat{\phi}|_{\text{LR}} \\ &= -k. \end{aligned} \quad (\text{B156})$$

Below, we continue the evaluation for specific noise models. For the Pauli noise,

$$\hat{\mathcal{N}}_{S \rightarrow E} = (\hat{\mathcal{N}}_p)^{\otimes n}, \quad (\text{B157})$$

with $\hat{\mathcal{N}}_p$ as the complementary channel of the single-qubit Pauli noise $\mathcal{N}_p(\rho) = p_I \rho + p_X X \rho X + p_Y Y \rho Y + p_Z Z \rho Z$. Particularly,

$$\hat{\mathcal{N}}_{S \rightarrow E} \left(|\hat{\phi}\rangle \langle \hat{\phi}|_{SS'} \right) = \left(\hat{\mathcal{N}}_p \left(|\hat{\phi}\rangle \langle \hat{\phi}| \right) \right)^{\otimes n}. \quad (\text{B158})$$

When applying the nonsmooth decoupling theorem, we first calculate

$$H_2(S|E) \left(\hat{\mathcal{N}}_p(|\hat{\phi}\rangle \langle \hat{\phi}|) \right)^{\otimes n} \left(|\hat{\phi}\rangle \langle \hat{\phi}|_{SS'} \right) = n(1 - f(\vec{p})) \quad (\text{B159})$$

by Eq. (B58). Thus,

$$\mathbb{E}_{C \in \text{Choi}} \leq 2^{-\frac{n}{4} (1 - f(\vec{p}) - \frac{k}{n})}. \quad (\text{B160})$$

Next, we apply the smooth decoupling theorem to Pauli noise. Note that

$$\frac{1}{n} H_{\min}^\delta(S|E)_{\rho^{\otimes n}} \geq H(S|E)_\rho - \frac{4 \log \gamma \sqrt{\log \frac{2}{\delta^2}}}{\sqrt{n}} \quad (\text{B161})$$

and

$$\lim_{n \rightarrow \infty} \frac{4 \log \gamma \sqrt{\log \frac{2}{\delta^2}}}{\sqrt{n}} = 0. \quad (\text{B162})$$

For sufficiently large n , we have $\frac{4 \log \gamma \sqrt{\log(2/\delta^2)}}{\sqrt{n}} < \delta$. Meanwhile, $H(S|E)_{\hat{\mathcal{N}}_p(|\hat{\phi}\rangle \langle \hat{\phi}|)} = 1 - h(\vec{p})$. Hence, for any

$\delta > 0$,

$$\begin{aligned} \mathbb{E}_C \left\| \hat{\mathcal{N}}_{S \rightarrow E} \left(U \left| \hat{\phi} \right\rangle \left\langle \hat{\phi} \right|_{\text{LR}} \otimes |0^{n-k}\rangle \langle 0^{n-k}| U^\dagger \right) - \hat{\mathcal{N}}_{S \rightarrow E} \left(\frac{\mathbb{I}_S}{2^n} \right) \otimes \frac{\mathbb{I}_R}{2^k} \right\|_1 \\ \leq 2^{-\frac{n}{2} (1 - h(\vec{p}) - \frac{k}{n} - \delta)} + 8\delta, \end{aligned} \quad (\text{B163})$$

$$\mathbb{E}_{C \in \text{Choi}} \leq \sqrt{8\delta + 2^{-\frac{n}{2} (1 - h(\vec{p}) - \frac{k}{n} - \delta)}}. \quad (\text{B164})$$

The derivation for the i.i.d. erasure error and local amplitude damping noise is similar. For strength- p i.i.d. erasure error, we can get

$$\mathbb{E}_{C \in \text{Choi}} \leq 2^{-\frac{n}{4} (1 - \log(1 + 3p) - \frac{k}{n})}, \quad (\text{B165})$$

with the nonsmooth decoupling theorem, and

$$\mathbb{E}_{C \in \text{Choi}} \leq \sqrt{8\delta + 2^{-\frac{n}{2} (1 - 2p - \frac{k}{n} - \delta)}}, \quad (\text{B166})$$

with the smooth decoupling theorem.

For strength- p local amplitude damping noise, we can get

$$\mathbb{E}_{C \in \text{Choi}} \leq 2^{-\frac{n}{4} (-\log(\frac{1}{2-p} + \sqrt{\frac{p}{2-p}}) - \frac{k}{n})}, \quad (\text{B167})$$

with the nonsmooth decoupling theorem, and

$$\mathbb{E}_{C \in \text{Choi}} \leq \sqrt{8\delta + 2^{-\frac{n}{2} (h(\frac{1-p}{2}) - h(\frac{p}{2}) - \frac{k}{n} - \delta)}}, \quad (\text{B168})$$

with the smooth decoupling theorem.

For strength- p 1D nearest-neighbor ZZ-coupling noise, the analysis is done by the nonsmooth decoupling theorem. The $H_2(S|R)$ term is still equal to $-k$, and the $H_2(S|E)$ term is given by

$$\begin{aligned} H_2(S|E)_{\tau_{SE}} &= -\log \text{tr}_{S'} \left[\text{tr}_S \sqrt{\mathcal{N}_{zz} \left(|\hat{\phi}\rangle \langle \hat{\phi}|^{\otimes n} \right)} \right] \\ &= -\log \text{tr}_{S'} \left[\text{tr}_S \prod_{\{i,i+1\} \subseteq [n]} [(1-p)\mathcal{I} + p \mathcal{Z}_i \mathcal{Z}_{i+1}] \left(|\hat{\phi}\rangle \langle \hat{\phi}|_{SS'} \right) \right] \\ &= n - 2 \log \left((\sqrt{1-p} + \sqrt{p})^{n-1} \right) \\ &\geq n \left[1 - 2 \log \left(\sqrt{1-p} + \sqrt{p} \right) \right]. \end{aligned} \quad (\text{B169})$$

Thus, for ZZ-coupling noise, we have

$$\mathbb{E}_{C \in \text{Choi}} \leq 2^{-\frac{n}{4}(1-2\log(\sqrt{1-p}+\sqrt{p})-\frac{k}{n})}. \quad (\text{B170})$$

Finally, we consider the t -erasure errors. The random Clifford encoding is robust to a fixed t -erasure error rather than a random erasure error. The number of erasure errors should satisfy the Singleton bound $n - 2t - k \geq 0$ [40]. The corresponding complementary channel is $\hat{\mathcal{N}}_{S \rightarrow E} = \text{tr}_{n-t}$. By applying the nonsmooth decoupling theorem, we have that

$$\hat{\mathcal{N}}_{S' \rightarrow E}(|\hat{\phi}\rangle\langle\hat{\phi}|_{SS'}) = |\hat{\phi}\rangle\langle\hat{\phi}|_{\text{TE}}^{\otimes t} \otimes \frac{\mathbb{I}_{\bar{T}}}{2^{n-t}}, \quad (\text{B171})$$

$$\begin{aligned} H_2(S|E)_{\hat{\mathcal{N}}_{S' \rightarrow E}(|\hat{\phi}\rangle\langle\hat{\phi}|_{SS'})} \\ = -\log \text{tr} \left(\left(\left(\frac{\mathbb{I}_E}{2^t} \right)^{-\frac{1}{4}} |\hat{\phi}\rangle\langle\hat{\phi}|_{\text{TE}}^{\otimes t} \otimes \frac{\mathbb{I}_{\bar{T}}}{2^{n-t}} \left(\frac{\mathbb{I}_E}{2^t} \right)^{-\frac{1}{4}} \right)^2 \right) \\ = n - 2t. \end{aligned} \quad (\text{B172})$$

Here, T represents the region of the erasure errors. Combining the results above, we get the upper bound of the Choi error against a fixed t -erasure error for the random

Clifford encoding,

$$\mathbb{E}_{C \in \text{Choi}} \leq 2^{-\frac{1}{4}(n-2t-k)}. \quad (\text{B173})$$

13. Comparison to the block-encoding scheme

In this part, we compare the double-layer blocked encoding scheme and the block-encoding scheme. We consider two types of noise models. For local Pauli noise, we provide evidence that the double-layer blocked encoding scheme might have an advantage. For local erasure error, we rigorously establish an advantage of the double-layer blocked encoding scheme over the block-encoding scheme in certain error regimes.

Local Pauli noise. We first consider the local Pauli noise. As mentioned in the main text, the encoding unitary operation of the block-encoding method is

$$U = \bigotimes_{i=1}^{\frac{N}{2}} U_C^{4i-3, 4i-2, 4i-1, 4i}, \quad (\text{B174})$$

where $U_C^{4i-3, 4i-2, 4i-1, 4i}$ is a random Clifford gate on regions $4i-3, 4i-2, 4i-1, 4i$. The set of the random block-encoding unitary operations is denoted as $\mathfrak{S}_n^B$.

To compare the performance of the two encoding methods, we only need to evaluate

$$\mathbb{E}_U \left\| \hat{\mathcal{N}}_{S \rightarrow E} \left(U |\hat{\phi}\rangle\langle\hat{\phi}|_{\text{LR}} \otimes |0^{n-k}\rangle\langle 0^{n-k}| U^\dagger \right) - \hat{\mathcal{N}}_{S \rightarrow E} \left(\frac{\mathbb{I}_S}{2^n} \right) \otimes \frac{\mathbb{I}_R}{2^k} \right\|_1 \quad (\text{B175})$$

under two encoding ensembles $\mathfrak{C}_n^\varepsilon$ and $\mathfrak{S}_n^B$. For simplicity, here we only analyze the results associated with the nonsmooth decoupling theorem. The case with the smooth decoupling theorem is similar.

The result of $\mathfrak{C}_n^\varepsilon$ is given by Eq. (B60). The result of $\mathfrak{S}_n^B$ can be established by considering the error of each block and the triangle inequality. Specifically,

$$\begin{aligned} \mathbb{E}_{U \sim \mathfrak{S}_n^B} \left\| \hat{\mathcal{N}}_{S \rightarrow E} \left(U |\hat{\phi}\rangle\langle\hat{\phi}|_{\text{LR}} \otimes |0^{n-k}\rangle\langle 0^{n-k}| U^\dagger \right) - \hat{\mathcal{N}}_{S \rightarrow E} \left(\frac{\mathbb{I}_S}{2^n} \right) \otimes \frac{\mathbb{I}_R}{2^k} \right\|_1 \\ \leq \frac{n}{4\xi} \mathbb{E}_{U \sim \mathfrak{S}_{4\xi}^B} \left\| \hat{\mathcal{N}}_{S \rightarrow E} \left(U |\hat{\phi}\rangle\langle\hat{\phi}|_{\text{LR}} \otimes |0^{\frac{4\xi(n-k)}{n}}\rangle\langle 0^{\frac{4\xi(n-k)}{n}}| U^\dagger \right) - \hat{\mathcal{N}}_{S \rightarrow E} \left(\frac{\mathbb{I}_S}{2^{4\xi}} \right) \otimes \frac{\mathbb{I}_R}{2^{\frac{4\xi k}{n}}} \right\|_1 \\ \leq \frac{n}{4\xi} 2^{-\frac{1}{2}H_2(L|R)|\hat{\phi}\rangle\langle\hat{\phi}|^{\otimes 4\xi k/n} - \frac{1}{2}H_2(S|E)_{(\hat{\mathcal{N}}_P(|\hat{\phi}\rangle\langle\hat{\phi}|))^{\otimes 4\xi}}}. \end{aligned} \quad (\text{B176})$$

Using $H_2(S|E)_{\hat{\mathcal{N}}_P(|\hat{\phi}\rangle\langle\hat{\phi}|)} = 1 - f(\vec{p})$, we have that

$$\begin{aligned} \mathbb{E}_{U \sim \mathfrak{S}_n^B} \left\| \hat{\mathcal{N}}_{S \rightarrow E} \left(U |\hat{\phi}\rangle\langle\hat{\phi}|_{\text{LR}} \otimes |0^{n-k}\rangle\langle 0^{n-k}| U^\dagger \right) - \hat{\mathcal{N}}_{S \rightarrow E} \left(\frac{\mathbb{I}_S}{2^n} \right) \otimes \frac{\mathbb{I}_R}{2^k} \right\|_1 &\leq \frac{n}{4\xi} (2^{-2\xi(1-f(\vec{p})-\frac{k}{n})}) \\ &= \frac{n}{4 \log \frac{n}{\varepsilon}} \left(\frac{\varepsilon}{n} \right)^{2(1-f(\vec{p})-\frac{k}{n})}. \end{aligned} \quad (\text{B177})$$

Recall that the result of the double-layer blocked encoding is given by

$$\begin{aligned} & \mathbb{E}_{U \sim \mathcal{C}_n^\varepsilon} \left\| \hat{\mathcal{N}}_{S \rightarrow E} \left(U \left| \hat{\phi} \right\rangle \left\langle \hat{\phi} \right|_{\text{LR}} \otimes |0^{n-k}\rangle \langle 0^{n-k}| U^\dagger \right) - \hat{\mathcal{N}}_{S \rightarrow E} \left(\frac{\mathbb{I}_S}{2^n} \right) \otimes \frac{\mathbb{I}_R}{2^k} \right\|_1 \\ & \leq \sqrt{2^{-n(1-f(\vec{p})-\frac{k}{n})}} + \frac{4\varepsilon^{1-\frac{k}{n}} n^{\frac{k}{n}}}{\log(n/\varepsilon)}. \end{aligned} \quad (\text{B178})$$

For block encoding, the dependence of error on n is given by $(n/4 \log(n/\varepsilon))(\varepsilon/n)^{2(1-f(\vec{p})-k/n)}$. For double-layer blocked encoding, it is approximately given by two terms $\sqrt{2^{-n(1-f(\vec{p})-k/n)}}$ and $\sqrt{4\varepsilon^{1-\frac{k}{n}} n^{\frac{k}{n}} / \log(n/\varepsilon)}$. Note that ε is normally chosen to make $\varepsilon/n = 1/\text{poly}(n)$. Meanwhile, $1 - f(\vec{p}) - k/n$ is often a constant. Thus, an exponentially decay term must vanish faster,

$$\sqrt{2^{-n(1-f(\vec{p})-\frac{k}{n})}} \ll \frac{n}{4 \log \frac{n}{\varepsilon}} \left(\frac{\varepsilon}{n} \right)^{2(1-f(\vec{p})-\frac{k}{n})}. \quad (\text{B179})$$

Also, when $1 - f(\vec{p}) - k/n < \frac{1}{4}(1 - k/n)$, we have

$$\begin{aligned} & \lim_{n \rightarrow \infty} \frac{\sqrt{\frac{4\varepsilon^{1-\frac{k}{n}} n^{\frac{k}{n}}}{\log(n/\varepsilon)}}}{\frac{n}{4 \log \frac{n}{\varepsilon}} \left(\frac{\varepsilon}{n} \right)^{2(1-f(\vec{p})-\frac{k}{n})}} \\ & = 8 \lim_{n \rightarrow \infty} \sqrt{\frac{\log \frac{n}{\varepsilon}}{n}} \left(\frac{\varepsilon}{n} \right)^{2f(\vec{p})-\frac{3}{2}(1-\frac{k}{n})} = 0. \end{aligned} \quad (\text{B180})$$

Thus,

$$\sqrt{\frac{4\varepsilon^{1-\frac{k}{n}} n^{\frac{k}{n}}}{\log(n/\varepsilon)}} \ll \frac{n}{4 \log \frac{n}{\varepsilon}} \left(\frac{\varepsilon}{n} \right)^{2(1-f(\vec{p})-\frac{k}{n})}. \quad (\text{B181})$$

That is, given the noise strength $\vec{p}$, when the encoding rate satisfies $k/n > 1 - \frac{4}{3}f(\vec{p})$, the error given by the double-layer blocked encoding decays much faster. In other words, if we fix a required Choi error decay scaling, the encoding rate offered by the double-layer blocked scheme can be higher. This provides evidence that the double-layer blocked encoding scheme can have an advantage in some error regimes. Nonetheless, it is a comparison of the upper bound of the error. Below, we consider local erasure errors and evaluate a lower bound of the Choi error given by the block-encoding method. This establishes a rigorous advantage of the double-layer blocked encoding scheme.

Local erasure error. To establish the lower bound of the Choi error given by the block-encoding method, we only need to evaluate the lower bound of the Choi error of each block, which corresponds to a random stabilizer encoding.

For a random stabilizer code $[[n, k]]$, correctability under erasure error requires the number of erasure errors t to satisfy the singleton bound,

$$2t + k \leq n \quad \Rightarrow \quad t \leq \left\lfloor \frac{n-k}{2} \right\rfloor. \quad (\text{B182})$$

Define $\tau = \frac{1}{2}(1 - k/n)$. Then any erasure pattern erasing more than τn qubits is uncorrectable. Moreover, as long as the number of erasure errors is $t > \lfloor (n-k)/2 \rfloor$, we can prove that there exists a lower bound of the Choi error for random stabilizer codes. The result is summarized as follows.

Theorem B2. Let $t \leq n/2$. Suppose Q is a subset of n qubits with $|Q| = t$. Then the expectation of Choi error against an erasure error on Q over $[[n, k]]$ random stabilizer codes has a lower bound,

$$\mathbb{E}_C \epsilon_Q \geq \frac{2t - (n-k) - e^{-1}}{8k}, \quad (\text{B183})$$

where ϵ_Q is the Choi error ϵ_{Choi} for erasure on Q .

We postpone the proof of Theorem B2 to the end of this subsection. With the number of erasure errors increasing, the Choi error also monotonically increases. Thus, we have the following corollary.

Corollary B1. For $t > n/2$,

$$\mathbb{E}_C \epsilon_Q \geq \mathbb{E}_C \epsilon_{Q'} = \frac{1}{8} - \frac{1}{ek}, \quad (\text{B184})$$

where $|Q'| = n/2$.

For strength- p local erasure errors, each qubit is erased independently with probability p . The number of erased qubits satisfies a binomial distribution $B(n, p)$, where we denote this number or random variable as E_n . Based on Theorem B2, the Choi error is always lower-bounded by $(1 - e^{-1})/8k \geq 1/16n$ when $E_n > \tau n$. Then the Choi error of a random stabilizer code is lower-bounded by

$(1/16n) \Pr[E_n > \tau n]$. Using Chernoff bounds, for $\tau > p$ or $1 - k/n - 2p > 0$, we have

$$\Pr[E_n > \tau n] = \exp(-n \cdot D(\tau \| p) + o(n)), \quad (\text{B185})$$

where $D(a\|b)$ is the binary relative entropy,

$$D(a\|b) = a \log_2 \left(\frac{a}{b} \right) + (1-a) \log_2 \left(\frac{1-a}{1-b} \right). \quad (\text{B186})$$

Thus, the Choi error satisfies

$$\begin{aligned} \mathbb{E}_{C \in \text{Choi}} &\geq \frac{1}{16n} \exp(-n \cdot D(\tau \| p) + o(n)) \\ &= \exp(-nD(\tau \| p) + o(n)). \end{aligned} \quad (\text{B187})$$

To further give a lower bound of the Choi error for the block-encoding scheme, we introduce the near-optimal fidelity [104], denoted as $\tilde{F}_{\text{Choi}}$, with the corresponding near-optimal Choi error denoted as $\tilde{\epsilon}_{\text{Choi}}$. Compared to the definition of original Choi fidelity,

$$F_{\text{Choi}} = \max_{\mathcal{D}} F(\hat{\phi}_{\text{LR}}, [(\mathcal{D} \circ \mathcal{N} \circ \mathcal{E})_L \otimes I_R](\hat{\phi}_{\text{LR}})), \quad (\text{B188})$$

the near-optimal fidelity uses the transpose channel recovery for decoding,

$$\tilde{F}_{\text{Choi}} = F(\hat{\phi}_{\text{LR}}, [(\mathcal{D}^{TC} \circ \mathcal{N} \circ \mathcal{E})_L \otimes I_R](\hat{\phi}_{\text{LR}})). \quad (\text{B189})$$

The transpose channel recovery $\mathcal{D}^{TC}$ is defined as [104]

$$\mathcal{D}^{TC} = P_L \circ \mathcal{N}^\dagger \circ \mathcal{N}^{-\frac{1}{2}}(P_L). \quad (\text{B190})$$

Here, $\mathcal{N}$ is the noise channel. The map P_L satisfies $P_L(\rho) = \Pi_L \rho \Pi_L$, where Π_L is the projector to the logical space. The map $\mathcal{N}^{-\frac{1}{2}}(P_L)$ satisfies $[\mathcal{N}^{-\frac{1}{2}}(P_L)](\rho) = \mathcal{N}^{-\frac{1}{2}}(\Pi_L) \rho \mathcal{N}^{-\frac{1}{2}}(\Pi_L)$.

It has been proved that there exist two-sided bounds between the optimal fidelity and the near-optimal fidelity [104,105],

$$\frac{1}{2}(1 - \tilde{F}_{\text{Choi}}) \leq 1 - F_{\text{Choi}} \leq 1 - \tilde{F}_{\text{Choi}}. \quad (\text{B191})$$

Using the relation $F = \sqrt{1 - \epsilon^2}$, we have that for sufficiently small ϵ_{Choi} ,

$$\epsilon_{\text{Choi}} \leq \tilde{\epsilon}_{\text{Choi}} \leq \sqrt{2}\epsilon_{\text{Choi}}. \quad (\text{B192})$$

The left inequality follows from the near-optimality. The right inequality is derived by

$$\begin{aligned} \frac{1}{2}(1 - \sqrt{1 - \tilde{\epsilon}_{\text{Choi}}^2}) &\leq 1 - \sqrt{1 - \epsilon_{\text{Choi}}^2} \\ &\Leftrightarrow 4(1 - \epsilon_{\text{Choi}}^2) \leq 2 - \tilde{\epsilon}_{\text{Choi}}^2 + 2\sqrt{1 - \tilde{\epsilon}_{\text{Choi}}^2} \\ &\leq 2 - \tilde{\epsilon}_{\text{Choi}}^2 + 2\left(1 - \frac{1}{2}\tilde{\epsilon}_{\text{Choi}}^2\right) \\ &\Leftrightarrow \tilde{\epsilon}_{\text{Choi}} \leq \sqrt{2}\epsilon_{\text{Choi}}. \end{aligned} \quad (\text{B193})$$

Note that in the second line we use $\sqrt{1-x} \leq 1 - \frac{1}{2}x$ for sufficiently small x .

From the definition of the transpose recovery map, it can be seen that for a block-encoding scheme $\mathcal{E} = \bigotimes_i \mathcal{E}_i$ and block noise model $\mathcal{N} = \bigotimes_i \mathcal{N}_i$, the transpose recovery map is also a tensor product, $\mathcal{D}^{TC} = \bigotimes_i \mathcal{D}_i^{TC}$. Thus, given a block-encoding scheme, since each block is independent, and the noise also acts independently on each block, the near-optimal fidelity for the whole block-encoding scheme is the product of the near-optimal fidelity for each block,

$$\tilde{F}_{\text{Choi}} = \prod_{i=1}^{\frac{n}{4\xi}} \tilde{F}_{\text{block}_i}, \quad (\text{B194})$$

or equivalently,

$$\tilde{F}_{\text{Choi}}^2 = \prod_{i=1}^{\frac{n}{4\xi}} \tilde{F}_{\text{block}_i}^2. \quad (\text{B195})$$

Then,

$$1 - \tilde{\epsilon}_{\text{Choi}}^2 = \prod_{i=1}^{\frac{n}{4\xi}} (1 - \tilde{\epsilon}_{\text{block}_i}^2), \quad (\text{B196})$$

where

$$\begin{aligned} \mathbb{E}_{U \sim \mathfrak{S}_n^B} \tilde{\epsilon}_{\text{block}_i} &\geq \mathbb{E}_{U \sim \mathfrak{S}_n^B} \epsilon_{\text{block}_i} \\ &\geq \exp(-4\xi \cdot D(\tau \| p) + o(\xi)). \end{aligned} \quad (\text{B197})$$

Thus, we obtain the lower bound of the Choi error for the block-encoding scheme,

$$\begin{aligned} \mathbb{E}_{U \sim \mathfrak{S}_n^B} \epsilon_{\text{Choi}} &\geq \frac{1}{\sqrt{2}} \mathbb{E}_{U \sim \mathfrak{S}_n^B} \tilde{\epsilon}_{\text{Choi}} \\ &\geq \frac{1}{\sqrt{2}} \mathbb{E}_{U \sim \mathfrak{S}_n^B} \tilde{\epsilon}_{\text{Choi}}^2 \\ &= \frac{1}{\sqrt{2}} \mathbb{E}_{U \sim \mathfrak{S}_n^B} \left(1 - \prod_i (1 - \tilde{\epsilon}_{\text{block}_i}^2) \right) \\ &= \frac{1}{\sqrt{2}} \left(1 - \prod_i (1 - \mathbb{E}_{U \sim \mathfrak{S}_n^B} \tilde{\epsilon}_{\text{block}_i}^2) \right) \end{aligned}$$

$$\begin{aligned}
&\geq \frac{1}{\sqrt{2}} \left(1 - \prod_i (1 - [\mathbb{E}_{U \sim \mathfrak{C}_n^B} \tilde{\epsilon}_{\text{block}_i}]^2) \right) \\
&\geq \frac{1}{\sqrt{2}} \left(1 - (1 - \exp(-8\xi \cdot D(\tau\|p) + o(\xi)))^{\frac{n}{4\xi}} \right) \\
&\geq \min \left\{ \frac{n}{8\sqrt{2}\xi} \exp(-8\xi \cdot D(\tau\|p) + o(\xi)), \Omega(1) \right\} \\
&= \min \{ n \exp(-8\xi D(\tau\|p) + o(\xi)), \Omega(1) \}. \quad (\text{B198})
\end{aligned}$$

Here, the fourth and fifth lines use the independence of each block and $\mathbb{E}x^2 \geq (\mathbb{E}x)^2$, respectively. The last inequality comes as follows. Let $m := n/4\xi$ and $x := m \exp(-8\xi \cdot D(\tau\|p) + o(\xi))$. Then

$$\begin{aligned}
&(1 - (1 - \exp(-8\xi \cdot D(\tau\|p) + o(\xi)))^{\frac{n}{4\xi}}) \\
&= \left(1 - \left(1 - \frac{x}{m} \right)^m \right), \quad (\text{B199})
\end{aligned}$$

Using $(1 - x/m)^m \leq 1 - x/2$ for $0 \leq x \leq 1$, we obtain the first term; while for $x > 1$, $(1 - (1 - x/m)^m) \geq (1 - (1 - 1/m)^m) = \Omega(1)$, we obtain the second term.

Substituting ξ with $\log(n/\varepsilon)$, we get that

$$\mathbb{E}_{U \sim \mathfrak{C}_n^B} \epsilon_{\text{Choi}} \geq \min \left\{ n \left(\frac{\varepsilon}{n} \right)^{8D(\frac{1-k}{2}\|p) + o(1)}, \Omega(1) \right\}. \quad (\text{B200})$$

This equation shows that, for the block-encoding scheme, given the noise strength p , a trade-off relation between the encoding rate k/n and the Choi error exists. If one fixes an encoding rate, the Choi error will have a lower bound, and vice versa. Below, we prove that there exists a regime of the strength parameter p where the double-layer blocked encoding scheme achieves both a higher encoding rate and a lower Choi error. Moreover, there exist regimes of noise strength and encoding rate such that the double-layer blocked encoding scheme requires a lower depth to make the Choi error decay.

Recall Corollary 3, the Choi error for the double-layer blocked encoding scheme is

$$\mathbb{E}_{U \sim \mathfrak{C}_n^B} \epsilon_{\text{Choi}} \leq \left(2^{-n(1-\log(1+3p)-\frac{k}{n})} + \frac{4\varepsilon^{1-\frac{k}{n}} n^{\frac{k}{n}}}{\log(n/\varepsilon)} \right)^{\frac{1}{4}}. \quad (\text{B201})$$

The main contributions are $2^{-(n/4)(1-\log(1+3p)-k/n)}$ and $(4\varepsilon^{1-(k/n)} n^{k/n} / \log(n/\varepsilon))^{1/4}$. In the following, we discuss the case of $1 - \log(1 + 3p) - k/n > 0$. For simplicity, we set $r = k/n$ and $\varepsilon = n^{-(\alpha+r)/(1-r)}$ where r and α are both positive constants; p is also a constant. Then, $n/\varepsilon = n^{(1+\alpha)/(1-r)}$. The block size parameter $\xi = \log(n/\varepsilon) =$

$((1 + \alpha)/(1 - r)) \log n$. The upper bound of the Choi error for the double-layer blocked encoding scheme is simplified to

$$\begin{aligned}
\mathbb{E}_{U \sim \mathfrak{C}_n^B} \epsilon_{\text{Choi}} &\leq \left(2^{-n(1-\log(1+3p)-r)} + \frac{4(1-r)}{(1+\alpha)n^\alpha \log n} \right)^{\frac{1}{4}} \\
&= O(n^{-\frac{\alpha}{4}}). \quad (\text{B202})
\end{aligned}$$

The lower bound of the Choi error for the block-encoding scheme is simplified to

$$\mathbb{E}_{U \sim \mathfrak{C}_n^B} \epsilon_{\text{Choi}} \geq \min \left\{ n^{1-\frac{8(1+\alpha)}{1-r}D(\frac{1-r}{2}\|p) + o(1)} = n^{c_{\text{block}}}, \Omega(1) \right\}. \quad (\text{B203})$$

In the case that $0 < p < 1/3$, $0 < r < 1$, $\alpha > 0$ and $1 - \log(1 + 3p) - r > 0$, the Choi error of the double-layer blocked encoding scheme always decays. The decay scale is approximately $O(n^{-\alpha/4})$ where we ignore $\log n = n^{o(1)}$. On the other hand, the exponential coefficient within Eq. (B203) is

$$c_{\text{block}} = 1 - \frac{8(1+\alpha)}{1-r} D\left(\frac{1-r}{2}\|p\right) + o(1). \quad (\text{B204})$$

Given noise strength p , under the constraint of $r < 1 - \log(1 + 3p)$ and $\alpha > 0$, c_{block} can be greater than 0. For instance,

$$\begin{aligned}
p &= 0.25, \quad r = 0.2, \\
\alpha &= 0.1 \Rightarrow c_{\text{block}} \approx 0.14 \Rightarrow \mathbb{E}_{U \sim \mathfrak{C}_n^B} \epsilon_{\text{Choi}} \\
&\geq \min\{n^{0.14}, \Omega(1)\} = \Omega(1). \quad (\text{B205})
\end{aligned}$$

That is, given the noise strength $p = 0.25$ and encoding rate $r = 0.1$, we obtain that the double-layer blocked encoding scheme can have an error decay $n^{-0.025}$, and the block-encoding scheme cannot make the Choi error decay. In this case, $\xi = 1.375 \log n$. As mentioned in the main text, the depths for the two encoding schemes are the same after fixing ξ . This manifests that for $p = 0.25$ and $r = 0.1$, the double-layer blocked encoding scheme requires a lower depth to make the Choi error vanish. On the other hand, this implies that at the same circuit depth, the double-layer blocked encoding scheme can achieve a higher encoding rate.

More concretely, one can find parameter regimes such that $c_{\text{block}} > 0$. In this case, the double-layer blocked encoding scheme can correct errors, but the block-encoding scheme cannot. One can also find parameter regimes such that $c_{\text{block}} > -\alpha/4$. In this case, the double-layer blocked encoding scheme always outperforms the block-encoding scheme by manifesting a faster error decay. We plot the regimes that $c_{\text{block}} > 0$ and $c_{\text{block}} > -\alpha/4$ in Figs. 5(a) and 5(b), respectively.

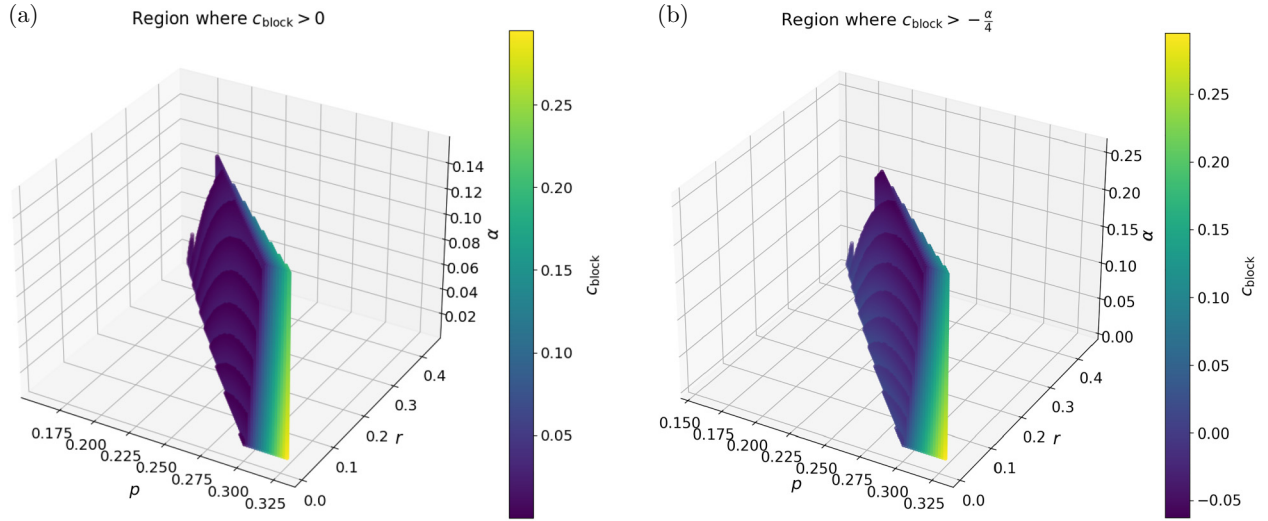


FIG. 5. (a) Parameter regime where $c_{\text{block}} > 0$. In this case, the double-layer blocked encoding scheme can make the Choi error decay, but the block-encoding scheme cannot. (b) Parameter regime where $c_{\text{block}} > -\alpha/4$. In this case, the double-layer blocked encoding scheme outperforms the block-encoding scheme.

Below, we show the proof of Theorem B2.

Proof. We first observe that the choice of subset Q is irrelevant when considering the expectation over the random stabilizer encoding scheme.

Lemma B3. For any fixed subsets $Q_1, Q_2 \subseteq [n]$ with $|Q_1| = |Q_2| = t$,

$$\mathbb{E}_{C \in \mathcal{C}, Q_1} = \mathbb{E}_{C \in \mathcal{C}, Q_2}, \quad (\text{B206})$$

where ϵ_{C, Q_i} is the Choi error for encoding unitary C and erasure errors on Q_i .

Proof. Because the SWAP operation is Clifford, there exists a Clifford S_{12} that swaps Q_1 to Q_2 . Hence

$$\begin{aligned} \mathbb{E}_{C \in \mathcal{C}, Q_2} &= \mathbb{E}_{C \in S_{12} \mathcal{C}, Q_1} \\ &= \mathbb{E}_{C \in \mathcal{C}, Q_1}. \end{aligned} \quad (\text{B207})$$

■

Fix the subsets $Q_1 = [t]$, $Q_2 = \{t+1, \dots, 2t\}$, and $Q_3 = [n] \setminus (Q_1 \cup Q_2)$ so that Q_3 contains the remaining $n - 2t$ physical qubits. We now prove the following.

Lemma B4. For any unitary encoding C ,

$$\frac{1}{2}(\epsilon_{Q_1} + \epsilon_{Q_2}) \geq \frac{2t - (n - k) - e^{-1}}{8k}. \quad (\text{B208})$$

Proof. Let $\epsilon_1 := \epsilon_{Q_1}$ and $\epsilon_2 := \epsilon_{Q_2}$. Also, recall that we denote L , R , and S as the logical, reference, and physical

systems, respectively. The initial input state is the $\hat{\phi}_{\text{LR}}$, and after encoding and noise, the state is denoted as $\hat{\phi}_{\text{LR}}$.

By definition of ϵ_Q , there exists a recovery map $\mathcal{D}_{Q_2 Q_3 \rightarrow L}$ such that

$$\sigma = \mathcal{D}(\hat{\phi}_{\text{LR}}), \quad P(\sigma_{\text{LR}}, \hat{\phi}_{\text{LR}}) = \epsilon_1. \quad (\text{B209})$$

Because the purified distance bounds the trace distance,

$$d_{\text{tr}}(\sigma_{\text{LR}}, \hat{\phi}_{\text{LR}}) \leq \epsilon_1. \quad (\text{B210})$$

Applying the Fannes-Audenaert inequality [65] with subsystem size $|LR| = 2k$ gives

$$S(\sigma_{\text{LR}}) \leq S(\hat{\phi}_{\text{LR}}) + 2\epsilon_1(2k) + \frac{1}{e} = 4k\epsilon_1 + \frac{1}{e} := \delta_1. \quad (\text{B211})$$

Therefore,

$$I_\sigma(Q_1 : R) \leq I_\sigma(Q_1 : LR) \leq 2S(\sigma_{\text{LR}}) \leq 2\delta_1. \quad (\text{B212})$$

Here, $I_\sigma(A : B)$ is the mutual information between subsystems A and B for state σ . The first inequality comes from the data processing inequality for mutual information, and the second uses $I_\sigma(A : B) \leq 2S(\sigma_B)$ for arbitrary σ . Let $\rho := \hat{\phi}_{\text{LR}}$. Because the recovery map $\mathcal{D}$ acts only on $Q_2 Q_3$,

the registers Q_1 and R are untouched, so

$$I_\rho(Q_1 : R) = I_\sigma(Q_1 : R) \leq 2\delta_1. \quad (\text{B213})$$

Expanding mutual information and using that ρ is pure, we obtain

$$\begin{aligned} S(\rho_{Q_1}) + S(\rho_R) &\leq S(\rho_{Q_1 R}) + 2\delta_1 = S(\rho_{Q_2 Q_3}) + 2\delta_1 \\ &\leq S(\rho_{Q_2}) + S(\rho_{Q_3}) + 2\delta_1, \end{aligned} \quad (\text{B214})$$

where the last step uses subadditivity. Repeating the same argument for Q_2 gives

$$\begin{aligned} S(\rho_{Q_2}) + S(\rho_R) &\leq S(\rho_{Q_2 R}) + 2\delta_2 = S(\rho_{Q_1 Q_3}) + 2\delta_2 \\ &\leq S(\rho_{Q_1}) + S(\rho_{Q_3}) + 2\delta_2. \end{aligned} \quad (\text{B215})$$

Adding the two inequalities yields

$$2S(\rho_R) \leq 2S(\rho_{Q_3}) + 2(\delta_1 + \delta_2). \quad (\text{B216})$$

With $S(\rho_R) = k$ and $S(\rho_{Q_3}) \leq |Q_3| = n - 2t$, we obtain

$$\delta_1 + \delta_2 \geq 2t - (n - k). \quad (\text{B217})$$

Substituting (B211), we have

$$\frac{1}{2}(\epsilon_1 + \epsilon_2) \geq \frac{2t - (n - k) - e^{-1}}{8k}. \quad (\text{B218})$$

■

Based on Lemmas B3 and B4, we prove Theorem B2 as follows:

$$\begin{aligned} \mathbb{E}_{C \in Q_1} &= \frac{1}{2}(\mathbb{E}_{C \in Q_1} + \mathbb{E}_{C \in Q_2}) \\ &= \mathbb{E}_C \frac{\epsilon_{Q_1} + \epsilon_{Q_2}}{2} \geq \frac{2t - (n - k) - e^{-1}}{8k}. \end{aligned} \quad (\text{B219})$$

■

APPENDIX C: ANALYSIS FOR 1D BRICKWORK CIRCUITS

In this part, we prove that the decoupling can be achieved by 1D log-depth brickwork circuits. We start with introducing preliminaries, which discuss how to evaluate $\text{tr } O_2 \Phi_{\mathcal{E}}^2(O_1)$ by transforming it into a stochastic process. Then, we present the main results of the decoupling theorem for 1D brickwork circuits and prove Theorem 2. Finally, we apply this decoupling theorem to quantum error correction and show that there exists a constant threshold for encoding schemes with such circuits against local noise. Note that here we prove the result for a generic local dimension q .

1. Random quantum circuits to stochastic processes

Here, we show how to evaluate $\text{tr } O_2 \Phi_{\mathcal{E}}^2(O_1)$ by transforming it into a stochastic process for certain observables O_1 and O_2 , which lays a foundation for further analyzing the decoupling error. The evaluation is equivalent to evaluating a statistical-mechanical partition function. Most of the analysis is similar to that in Ref. [32]. Nonetheless, in Ref. [32], O_1 and O_2 take specific forms of $|1^n\rangle\langle 1^n|^{\otimes 2}$, which makes the boundary condition in the statistical-mechanical mapping easier. For general operators O_1 and O_2 , the boundary condition is more complicated, and a more refined analysis is necessary.

We first evaluate $\Phi_{\mathcal{E}}^2(O_1)$. By definition,

$$\Phi_{\mathcal{E}}^2(O_1) = \mathbb{E}_{U \sim \mathcal{E}} U^{\otimes 2} O_1 U^{\dagger \otimes 2}, \quad (\text{C1})$$

where

$$U = U_{A^{(s)}}^{(s)} U_{A^{(s-1)}}^{(s-1)} \cdots U_{A^{(1)}}^{(1)}. \quad (\text{C2})$$

We would like to evaluate terms like $\mathbb{E}_U U^{\otimes k} O U^{\dagger \otimes k}$. Also, $U^{(i)}$ is a Haar-random gate of dimension q^2 . Inserting random single-qudit operations before all two-qudit gates does not change the expectation. Thus, we further consider the random unitary operation of the following form:

$$U = U_{A^{(s)}}^{(s)} U_{A^{(s-1)}}^{(s-1)} \cdots U_{A^{(1)}}^{(1)} \bigotimes_{i=1}^n U_{\{i\}}^{(-i)}. \quad (\text{C3})$$

Here, $U_{\{i\}}^{(-i)}$ is a random single-qudit operation with dimension $q \times q$ acting on qudit i . It is worth noting that when analyzing 1D brickwork circuits, we also insert a random single-qudit operation $\bigotimes_{i=1}^n U_{\{i\}}^{(-i)}$ before the random two-qudit unitary operations within Eq. (2).

We continue the analysis by evaluating the actions of each part within U . The action of $\bigotimes_{i=1}^n U_{\{i\}}^{(-i)}$ on O_1 can be evaluated by Eq. (51). We obtain that

$$\begin{aligned} &\mathbb{E}_{U^{(-i)}} \left(\bigotimes_{i=1}^n U_{\{i\}}^{(-i) \otimes 2} \right) O_1 \left(\bigotimes_{i=1}^n U_{\{i\}}^{(-i) \dagger \otimes 2} \right) \\ &= \frac{1}{(q^2 - 1)^n} \sum_{\vec{\gamma} \in \{I, S\}^n} \text{tr} \left(O_1 \bigotimes_{j=1}^n g_{\gamma_j} \right) \bigotimes_{j=1}^n \gamma_j. \end{aligned} \quad (\text{C4})$$

Here, γ_j is the j th component of $\vec{\gamma}$ and equals identity operation $\mathbb{I}$ or SWAP operation F . The term $g_{\gamma_j} = \mathbb{I} - q^{-1}F$ if $\gamma_j = \mathbb{I}$, and $g_{\gamma_j} = F - q^{-1}\mathbb{I}$ if $\gamma_j = F$. The vector γ is called a configuration, corresponding to the spin configuration is the statistical-mechanical mapping. For simplicity, we denote $c_{\vec{\gamma}}(O_1) = (1/(q^2 - 1)^n) \text{tr}(O_1 \bigotimes_{j=1}^n g_{\gamma_j})$ and

use $\vec{\gamma}$ to represent $\bigotimes_{j=1}^n \gamma_j$ in the above equation. That is,

$$\begin{aligned} & \mathbb{E}_{U^{(-i)}} \left(\bigotimes_{i=1}^n U_{\{i\}}^{(-i) \otimes 2} \right) O_1 \left(\bigotimes_{i=1}^n U_{\{i\}}^{(-i) \dagger \otimes 2} \right) \\ &= \sum_{\vec{\gamma} \in \{\mathbb{I}, F\}^n} c_{\vec{\gamma}}(O_1) \vec{\gamma}. \end{aligned} \quad (\text{C5})$$

The next step is evaluating the action of a random two-qudit gate $U_{A^{(i)}}^{(i)}$ on $\vec{\gamma}$, which was previously studied in Ref. [32]. Since $A^{(i)}$ only contains two qudits, it suffices

to analyze how $U_{A^{(i)}}^{(i)}$ acts on γ within support $A^{(i)}$. Denote $M^{(i)}(O) = \mathbb{E}_{U^{(i)}} U_{A^{(i)}}^{(i) \otimes 2} O U_{A^{(i)}}^{(i) \dagger \otimes 2}$, we have that

$$\begin{aligned} M^{(i)}(O) &= \text{tr}_{A^{(i)}} \left(O \frac{\mathbb{I}_{A^{(i)}} - q^{-2} F_{A^{(i)}}}{q^4 - 1} \right) \mathbb{I}_{A^{(i)}} \\ &\quad + \text{tr}_{A^{(i)}} \left(O \frac{F_{A^{(i)}} - q^{-2} \mathbb{I}_{A^{(i)}}}{q^4 - 1} \right) F_{A^{(i)}}. \end{aligned} \quad (\text{C6})$$

Particularly, we consider that O equals $\mathbb{I} \otimes \mathbb{I}$, $\mathbb{I} \otimes F$, $F \otimes \mathbb{I}$, and $F \otimes F$ within support $A^{(i)}$. We have that

$$M^{(i)}(O_{A^{(i)}} \otimes (\mathbb{I} \otimes \mathbb{I})_{A^{(i)}}) = O_{A^{(i)}} \otimes (\mathbb{I} \otimes \mathbb{I})_{A^{(i)}}, \quad (\text{C7})$$

$$M^{(i)}(O_{A^{(i)}} \otimes (F \otimes F)_{A^{(i)}}) = O_{A^{(i)}} \otimes (F \otimes F)_{A^{(i)}}, \quad (\text{C8})$$

$$M^{(i)}(O_{A^{(i)}} \otimes (\mathbb{I} \otimes F)_{A^{(i)}}) = O_{A^{(i)}} \otimes \left(\frac{q}{q^2 + 1} (\mathbb{I} \otimes \mathbb{I})_{A^{(i)}} + \frac{q}{q^2 + 1} (F \otimes F)_{A^{(i)}} \right), \quad (\text{C9})$$

$$M^{(i)}(O_{A^{(i)}} \otimes (F \otimes \mathbb{I})_{A^{(i)}}) = O_{A^{(i)}} \otimes \left(\frac{q}{q^2 + 1} (\mathbb{I} \otimes \mathbb{I})_{A^{(i)}} + \frac{q}{q^2 + 1} (F \otimes F)_{A^{(i)}} \right). \quad (\text{C10})$$

Here, $O_{A^{(i)}}$ only has support outside $A^{(i)}$. From the above equation, we obtain that if $O \in \{\mathbb{I}, F\}^n$, then $M^{(i)}(O) \in \{\mathbb{I}, F\}^n$. Particularly, $M^{(i)}$ is a linear transformation over $\{\mathbb{I}, F\}^n$, and we use $M_{\vec{v}\vec{\gamma}}^{(i)}$ to denote the matrix element such that

$$M^{(i)}(\vec{\gamma}) = \sum_{\vec{v} \in \{\mathbb{I}, F\}^n} M_{\vec{v}\vec{\gamma}}^{(i)} \vec{v}. \quad (\text{C11})$$

Suppose $A^{(i)} = \{a_i, b_i\}$, the matrix element $M_{\vec{v}\vec{\gamma}}^{(i)}$ satisfies

$$M_{\vec{v}\vec{\gamma}}^{(i)} = \begin{cases} 1, & \text{if } \gamma_{a_i} = \gamma_{b_i} \text{ and } \vec{\gamma} = \vec{v}, \\ \frac{q}{q^2 + 1}, & \text{if } \gamma_{a_i} \neq \gamma_{b_i}, \gamma_{a_i} = \gamma_{b_i}, \text{ and } \gamma_c = \gamma_c, \forall c \in [n] \setminus \{a_i, b_i\}, \\ 0, & \text{otherwise.} \end{cases} \quad (\text{C12})$$

It can be seen that each action of $M^{(i)}$ maintains the configuration $\vec{\gamma}$ with weight 1, or flips one term in $\vec{\gamma}$ with weight $q/(q^2 + 1)$. Thus, $\Phi_{\mathcal{E}}^2(O_1)$ can be expressed as follows:

$$\begin{aligned} \Phi_{\mathcal{E}}^2(O_1) &= M^{(s)} \dots M^{(2)} M^{(1)} \left(\sum_{\vec{\gamma} \in \{\mathbb{I}, F\}^n} c_{\vec{\gamma}}(O_1) \vec{\gamma} \right) \\ &= \sum_{\vec{\gamma} \in \{\mathbb{I}, F\}^n} c_{\vec{\gamma}}(O_1) M^{(s)} \dots M^{(2)} M^{(1)}(\vec{\gamma}) \\ &= \sum_{\Gamma = (\vec{\gamma}^{(0)}, \dots, \vec{\gamma}^{(s)}) \in \{\mathbb{I}, F\}^{n \times (s+1)}} c_{\vec{\gamma}^{(0)}}(O_1) \vec{\gamma}^{(s)} \prod_{i=1}^s M_{\vec{\gamma}^{(i)} \vec{\gamma}^{(i-1)}}^{(i)}. \end{aligned} \quad (\text{C13})$$

Here, $\Gamma = (\vec{\gamma}^{(0)}, \dots, \vec{\gamma}^{(s)})$ is called a trajectory, which is a sequence of $s + 1$ configurations. We also use Γ_i to denote its i th component. Here, $\Gamma_i = \vec{\gamma}^{(i-1)}$. Furthermore,

$$\begin{aligned} \text{tr}(O_2 \Phi_{\mathcal{E}}^2(O_1)) &= \sum_{\Gamma} \text{weight}(\Gamma) c_{\vec{\gamma}^{(0)}}(O_1) \text{tr}(O_2 \vec{\gamma}^{(s)}) \\ &= \sum_{\Gamma} \text{weight}(\Gamma) c_{\vec{\gamma}^{(0)} \vec{\gamma}^{(s)}}(O_1, O_2), \end{aligned} \quad (\text{C14})$$

where

$$\begin{aligned} \text{weight}(\Gamma) &= \prod_{i=1}^s M_{\vec{\gamma}^{(i)} \vec{\gamma}^{(i-1)}}^{(i)}, \quad (\text{C15}) \\ c_{\vec{\gamma}^{(0)} \vec{\gamma}^{(s)}}(O_1, O_2) &= \frac{1}{(q^2 - 1)^n} \text{tr} \left(O_1 \bigotimes_{j=1}^n g_{\gamma_j^{(0)}} \right) \text{tr} \left(O_2 \bigotimes_{j=1}^n \gamma_j^{(s)} \right). \end{aligned} \quad (\text{C16})$$

When $O_1 = O_2 = |1^n\rangle\langle 1^n|^{\otimes 2}$, $c_{\vec{\gamma}^{(0)} \vec{\gamma}^{(s)}}(O_1, O_2) = 1/(q + 1)^n$, which is the case studied in Ref. [32]. By mapping $\{\mathbb{I}, F\}$ to $\{+1, -1\}$, a configuration $\vec{\gamma}$ is a spin variable, and the trajectory Γ is a two-dimensional Ising spin configuration. Correspondingly, $\text{tr}(O_2 \Phi_{\mathcal{E}}^2(O_1))$ is mapped into a partition function of a classical Ising spin model, which is a weighted summation of different trajectories. For each trajectory Γ , the total weight depends on two terms. The first one is $\text{weight}(\Gamma)$, where the value depends on the whole trajectory. The second term is $c_{\vec{\gamma}^{(0)} \vec{\gamma}^{(s)}}(O_1, O_2)$, and the value only depends on the initial and ending configurations of the trajectory. Note that $\text{weight}(\Gamma)$ is always non-negative since the matrix element $M_{\vec{\gamma}^{(i)} \vec{\gamma}^{(i-1)}}^{(i)}$ is always non-negative.

For further analysis, we write

$$\begin{aligned} \text{tr}(O_2 \Phi_{\mathcal{E}}^2(O_1)) &= \sum_{\vec{\gamma}^{(0)} \vec{\gamma}^{(s)}} c_{\vec{\gamma}^{(0)} \vec{\gamma}^{(s)}}(O_1, O_2) \sum_{\Gamma \setminus \vec{\gamma}^{(0)} \vec{\gamma}^{(s)}} \text{weight}(\Gamma). \end{aligned} \quad (\text{C17})$$

Here $\sum_{\Gamma \setminus \vec{\gamma}^{(0)} \vec{\gamma}^{(s)}}$ takes the summation of trajectories that start with $\vec{\gamma}^{(0)}$ and end with $\vec{\gamma}^{(s)}$. In the following, we denote $Z_{\vec{\gamma}^{(0)} \vec{\gamma}^{(s)}} = \sum_{\Gamma \setminus \vec{\gamma}^{(0)} \vec{\gamma}^{(s)}} \text{weight}(\Gamma)$ and evaluate this term to analyze $\text{tr}(O_2 \Phi_{\mathcal{E}}^2(O_1))$. We start with analyzing $\text{weight}(\Gamma)$ with the method of random walk.

Given a fixed set of supports $\{A^{(1)}, A^{(2)}, \dots, A^{(s)}\}$, the action of $\prod_{i=1}^s M^{(i)}$ on a configuration $\vec{\gamma}^{(0)}$ results in a length- $(s + 1)$ random walk Γ over $\{\mathbb{I}, F\}^n$. The initial location of the random walk is a fixed configuration, $\vec{\gamma}^{(0)}$, and at time step t , the location transitions from $\vec{\gamma}^{(t-1)}$ to $\vec{\gamma}^{(t)}$. The configuration after step 1 may not be fixed. Based on Eq. (C12), the transition must follow the following rule:

the configuration is unchanged from $\vec{\gamma}^{(t-1)}$ to $\vec{\gamma}^{(t)}$ if the operators of $\vec{\gamma}^{(t-1)}$ within support $A^{(t)}$ are $\mathbb{I}\mathbb{I}$ and FF ; there is an identity operation flipped into a SWAP operation or a SWAP operation flipped into an identity operation if the operators of $\vec{\gamma}^{(t-1)}$ within support $A^{(t)}$ are $\mathbb{I}F$ and $F\mathbb{I}$, and the probability of each case of flip is $\frac{1}{2}$. If the configuration is unchanged, the contributed weight of this transition to $\text{weight}(\Gamma)$ is 1, and if there is a flip, the contributed weight is $q/(q^2 + 1)$. As a result, $\text{weight}(\Gamma)$ depends on how many flips within trajectory Γ . That is,

$$\text{weight}(\Gamma) = \left(\frac{q}{q^2 + 1} \right)^{\#F_{\Gamma}}, \quad (\text{C18})$$

where $\#F_{\Gamma}$ is the number of flips within Γ . It can also be written as

$$\begin{aligned} \text{weight}(\Gamma) &= \left(\frac{1}{2} \right)^{\#F_{\Gamma}} \left(\frac{2q}{q^2 + 1} \right)^{\#F_{\Gamma}} \\ &= \text{Pr}(\Gamma) \left(\frac{2q}{q^2 + 1} \right)^{\#F_{\Gamma}}, \end{aligned} \quad (\text{C19})$$

where $\text{Pr}(\Gamma) = (\frac{1}{2})^{\#F_{\Gamma}}$ is the probability of trajectory Γ in this random walk.

To calculate $Z_{\vec{\gamma}^{(0)} \vec{\gamma}^{(s)}}$, we need to count all trajectories satisfying the transition rule and require the start and ending configurations to be $\vec{\gamma}^{(0)}$ and $\vec{\gamma}^{(s)}$, respectively. Then, we obtain that

$$Z_{\vec{\gamma}^{(0)} \vec{\gamma}^{(s)}} = \sum_{\Gamma} \text{weight}(\Gamma) \mathbf{1}_{\Gamma_1 = \vec{\gamma}^{(0)}} \mathbf{1}_{\Gamma_{s+1} = \vec{\gamma}^{(s)}}. \quad (\text{C20})$$

Here, $\mathbf{1}_A$ is the indicator function, which equals 1 if A is satisfied and 0 otherwise. Using Eq. (C19), we have that

$$Z_{\vec{\gamma}^{(0)} \vec{\gamma}^{(s)}} = \mathbb{E}_{\Gamma} \left(\frac{2q}{q^2 + 1} \right)^{\#F_{\Gamma}} \mathbf{1}_{\Gamma_1 = \vec{\gamma}^{(0)}} \mathbf{1}_{\Gamma_{s+1} = \vec{\gamma}^{(s)}}. \quad (\text{C21})$$

Note that the weight of a particular walk or trajectory in the above random walk relates to the number of flips during the walk, which brings challenges in evaluating its value. To tackle this issue, we turn the above unbiased random walk into a biased random walk. In the original unbiased random walk, the flip has an equal probability from $\mathbb{I}$ to F and from F to $\mathbb{I}$. In the new biased random walk, we set the probability from $\mathbb{I}$ to F as $1/(q^2 + 1)$, and the probability from F to $\mathbb{I}$ as $q^2/(q^2 + 1)$. Accordingly, we can modify

the form of the weight of a trajectory into the following:

$$\begin{aligned}
 \text{weight}(\Gamma) &= \left(\frac{q}{q^2 + 1} \right)^{\#F_\Gamma} \\
 &= \left(\frac{q}{q^2 + 1} \right)^{\#F_\Gamma^{\mathbb{I} \rightarrow F}} \left(\frac{q}{q^2 + 1} \right)^{\#F_\Gamma^{F \rightarrow \mathbb{I}}} \\
 &= \left(\frac{1}{q^2 + 1} \right)^{\#F_\Gamma^{\mathbb{I} \rightarrow F}} \left(\frac{q^2}{q^2 + 1} \right)^{\#F_\Gamma^{F \rightarrow \mathbb{I}}} q^{\#F_\Gamma^{\mathbb{I} \rightarrow F} - \#F_\Gamma^{F \rightarrow \mathbb{I}}} \\
 &= \tilde{\text{Pr}}(\Gamma) q^{|\vec{\gamma}^{(s)}| - |\vec{\gamma}^{(0)}|}, \tag{C22}
 \end{aligned}$$

where $\#F_\Gamma^{\mathbb{I} \rightarrow F}$ is the number of flips from $\mathbb{I}$ to F within Γ , $\#F_\Gamma^{F \rightarrow \mathbb{I}}$ is the number of flips from F to $\mathbb{I}$ within Γ , and $\tilde{\text{Pr}}(\Gamma) = (1/(q^2 + 1))^{\#F_\Gamma^{\mathbb{I} \rightarrow F}} (q^2/(q^2 + 1))^{\#F_\Gamma^{F \rightarrow \mathbb{I}}}$ is the probability of trajectory Γ in the biased random walk, and $|\vec{\gamma}^{(i)}|$ is the number of SWAP operators in $\vec{\gamma}^{(i)}$. Then,

$$\begin{aligned}
 Z_{\vec{\gamma}^{(0)} \vec{\gamma}^{(s)}} &= \sum_{\Gamma} \text{weight}(\Gamma) \mathbf{1}_{\Gamma_1 = \vec{\gamma}^{(0)}} \mathbf{1}_{\Gamma_{s+1} = \vec{\gamma}^{(s)}} \\
 &= \sum_{\Gamma} \tilde{\text{Pr}}(\Gamma) q^{|\vec{\gamma}^{(s)}| - |\vec{\gamma}^{(0)}|} \mathbf{1}_{\Gamma_1 = \vec{\gamma}^{(0)}} \mathbf{1}_{\Gamma_{s+1} = \vec{\gamma}^{(s)}}
 \end{aligned}$$

$$\begin{aligned}
 &= q^{|\vec{\gamma}^{(s)}| - |\vec{\gamma}^{(0)}|} \tilde{\mathbb{E}}_{\Gamma} \mathbf{1}_{\Gamma_1 = \vec{\gamma}^{(0)}} \mathbf{1}_{\Gamma_{s+1} = \vec{\gamma}^{(s)}} \\
 &= q^{|\vec{\gamma}^{(s)}| - |\vec{\gamma}^{(0)}|} \tilde{\text{Pr}}(\Gamma : \vec{\gamma}^{(0)} \rightarrow \vec{\gamma}^{(s)}). \tag{C23}
 \end{aligned}$$

Here, $\tilde{\mathbb{E}}_{\Gamma}$ represents the expectation over biased random walk, and $\tilde{\text{Pr}}(\Gamma : \vec{\gamma}^{(0)} \rightarrow \vec{\gamma}^{(s)})$ represents the probability of this biased random walk transitioning from $\vec{\gamma}^{(0)}$ to $\vec{\gamma}^{(s)}$.

For an infinite-size circuit or $s \rightarrow \infty$, the final configuration of the random walk can only take $\mathbb{I}^n$ and F^n since they are the fixed points of the biased random walk. It was found that in this case the probability of $\vec{\gamma}^{(0)}$ transitioning into $\mathbb{I}^n$ or F^n only depends on the weight of $\vec{\gamma}^{(0)}$ or $|\vec{\gamma}^{(0)}|$ [32]. The concrete values are given by

$$\tilde{\text{Pr}}(\Gamma : \vec{\gamma}^{(0)} \rightarrow \mathbb{I}^n) = \frac{1 - q^{-2n+2}|\vec{\gamma}^{(0)}|}{1 - q^{-2n}}, \tag{C24}$$

$$\tilde{\text{Pr}}(\Gamma : \vec{\gamma}^{(0)} \rightarrow F^n) = \frac{q^{-2n+2}|\vec{\gamma}^{(0)}| - q^{-2n}}{1 - q^{-2n}}. \tag{C25}$$

We can use the above results of the infinite-size circuit to derive the $\text{tr } O_2 \Phi_{\text{Haar}}^2(O_1)$ for the Haar-random ensemble. This can be used to check the soundness of the derivation.

From Eq. (51), we have that

$$\begin{aligned}
 \text{tr } O_2 \Phi_{\text{Haar}}^2(O_1) &= \text{tr } O_2 \left(\text{tr } \left(O_1 \frac{\mathbb{I}^n - q^{-n} F^n}{q^{2n} - 1} \right) \mathbb{I}^n + \text{tr } \left(O_1 \frac{F^n - q^{-n} \mathbb{I}^n}{q^{2n} - 1} \right) F^n \right) \\
 &= \frac{1}{q^{2n} - 1} (\text{tr } O_1 \mathbb{I}^n \text{tr } O_2 \mathbb{I}^n + \text{tr } O_1 F^n \text{tr } O_2 F^n - q^{-n} \text{tr } O_1 \mathbb{I}^n \text{tr } O_2 F^n - q^{-n} \text{tr } O_1 F^n \text{tr } O_2 \mathbb{I}^n), \tag{C26}
 \end{aligned}$$

where $\mathbb{I}$ and F are the identity and SWAP operators between two qudits, respectively.

Based on Eqs. (C16), (C17), (C23), (C24), and (C25), we have that

$$\begin{aligned}
 \text{tr } O_2 \Phi_{\text{Haar}}^2(O_1) &= \sum_{\vec{\gamma}^{(0)} \vec{\gamma}^{(s)}} c_{\vec{\gamma}^{(0)} \vec{\gamma}^{(s)}}(O_1, O_2) Z_{\vec{\gamma}^{(0)} \vec{\gamma}^{(s)}} \\
 &= \sum_{\vec{\gamma}^{(0)} \vec{\gamma}^{(s)}} c_{\vec{\gamma}^{(0)} \vec{\gamma}^{(s)}}(O_1, O_2) q^{|\vec{\gamma}^{(s)}| - |\vec{\gamma}^{(0)}|} \tilde{\text{Pr}}(\Gamma : \vec{\gamma}^{(0)} \rightarrow \vec{\gamma}^{(s)}) \\
 &= \sum_{\vec{\gamma}^{(0)}} c_{\vec{\gamma}^{(0)} \mathbb{I}^n}(O_1, O_2) q^{-|\vec{\gamma}^{(0)}|} \tilde{\text{Pr}}(\Gamma : \vec{\gamma}^{(0)} \rightarrow \mathbb{I}^n) + \sum_{\vec{\gamma}^{(0)}} c_{\vec{\gamma}^{(0)} F^n}(O_1, O_2) q^{n - |\vec{\gamma}^{(0)}|} \tilde{\text{Pr}}(\Gamma : \vec{\gamma}^{(0)} \rightarrow F^n) \\
 &= \sum_{\vec{\gamma}^{(0)}} c_{\vec{\gamma}^{(0)} \mathbb{I}^n}(O_1, O_2) q^{-|\vec{\gamma}^{(0)}|} \frac{1 - q^{-2n+2}|\vec{\gamma}^{(0)}|}{1 - q^{-2n}} + \sum_{\vec{\gamma}^{(0)}} c_{\vec{\gamma}^{(0)} F^n}(O_1, O_2) q^{n - |\vec{\gamma}^{(0)}|} \frac{q^{-2n+2}|\vec{\gamma}^{(0)}| - q^{-2n}}{1 - q^{-2n}}
 \end{aligned}$$

$$\begin{aligned}
&= \sum_{\vec{\gamma}^{(0)}} \frac{1}{(q^2 - 1)^n} \left(\text{tr} \left(O_1 \bigotimes_{j=1}^n g_{\gamma_j^{(0)}} \right) \text{tr}(O_2 \mathbb{I}^n) \right) q^{-|\vec{\gamma}^{(0)}|} \frac{1 - q^{-2n+2|\vec{\gamma}^{(0)}|}}{1 - q^{-2n}} \\
&\quad + \sum_{\vec{\gamma}^{(0)}} \frac{1}{(q^2 - 1)^n} \left(\text{tr} \left(O_1 \bigotimes_{j=1}^n g_{\gamma_j^{(0)}} \right) \text{tr}(O_2 F^n) \right) q^{n-|\vec{\gamma}^{(0)}|} \frac{q^{-2n+2|\vec{\gamma}^{(0)}|} - q^{-2n}}{1 - q^{-2n}} \\
&= \frac{1}{q^{2n} - 1} \text{tr}(\text{tr} O_1 \mathbb{I}^n \text{tr} O_2 \mathbb{I}^n + \text{tr} O_1 F^n \text{tr} O_2 F^n - q^{-n} \text{tr} O_1 \mathbb{I}^n \text{tr} O_2 F^n - q^{-n} \text{tr} O_1 F^n \text{tr} O_2 \mathbb{I}^n). \tag{C27}
\end{aligned}$$

In the final equation, we use the fact that

$$\begin{aligned}
\sum_{\vec{\gamma}^{(0)}} q^{-|\vec{\gamma}^{(0)}|} \bigotimes_{j=1}^n g_{\gamma_j^{(0)}} &= (1 - q^{-2})^n \mathbb{I}^n, \\
\sum_{\vec{\gamma}^{(0)}} q^{|\vec{\gamma}^{(0)}|} \bigotimes_{j=1}^n g_{\gamma_j^{(0)}} &= (q - q^{-1})^n F^n. \tag{C28}
\end{aligned}$$

Thus, our derivation from the infinite-size circuit is in accordance with that from the Haar-random ensemble, verifying the soundness of the derivation. In the next section, we will focus on the 1D log-depth brickwork circuits.

2. Decoupling with 1D brickwork circuits and proof of Theorem 2

Below, we give the formal proof of Theorem 2.

As shown in Appendix B 1, since $\mathfrak{B}_n^\varepsilon$ is a unitary-1 design, we have that

$$\begin{aligned}
&\mathbb{E}_U \|\mathcal{T}_{S \rightarrow E}(U_S \rho_{\text{SR}} U_S^\dagger) - \tau_E \otimes \rho_R\|_1 \\
&\leq \sqrt{\mathbb{E}_U \text{tr} \left[\left(\tilde{\mathcal{T}}_{S \rightarrow E}(U_S \tilde{\rho}_{\text{SR}} U_S^\dagger) \right)^2 \right] - \text{tr}(\tilde{\tau}_E^2) \text{tr}(\tilde{\rho}_R^2)} \tag{C29}
\end{aligned}$$

and

$$\begin{aligned}
&\mathbb{E}_U \text{tr} \left[\left(\tilde{\mathcal{T}}_{S \rightarrow E}(U_S \tilde{\rho}_{\text{SR}} U_S^\dagger) \right)^2 \right] \\
&= \text{tr}_S(\Phi_{\mathfrak{B}_n^\varepsilon}^2(\text{tr}_R(\tilde{\rho}_{\text{SR}}^{\otimes 2} F_R))(\tilde{\mathcal{T}}_{S \rightarrow E}^\dagger)^{\otimes 2}(F_E)), \tag{C30}
\end{aligned}$$

where $\tilde{\mathcal{T}}_{S \rightarrow E} = \sigma_E^{-1/4} \mathcal{T}_{S \rightarrow E} \sigma_E^{-1/4}$ with the corresponding Choi-Jomiołkowski representation $\tilde{\tau}_{\text{SE}} = \sigma_E^{-1/4} \tau_{\text{SE}} \sigma_E^{-1/4}$ and $\tilde{\rho}_{\text{SR}} = \zeta_R^{-1/4} \rho_{\text{SR}} \zeta_R^{-1/4}$. In the following, we denote $O_{\text{SR}} = \text{tr}_R(\tilde{\rho}_{\text{SR}}^{\otimes 2} F_R)$ and $O_{\text{SE}} = (\tilde{\mathcal{T}}_{S \rightarrow E}^\dagger)^{\otimes 2}(F_E)$. The following analysis is to give the upper bound of $\text{tr} O_{\text{SE}} \Phi_{\mathcal{L}_n^d}^2(O_{\text{SR}})$.

Note that for 1D brickwork circuits, the random unitary is denoted as

$$U = \prod_{l=1}^d U^{[l]}. \tag{C31}$$

Thus, instead of defining $\vec{\gamma}^{(i)}$ by transforming $\vec{\gamma}^{(i-1)}$ with a two-qudit gate $U^{(i)}$, we define $\vec{\gamma}^i$ by transforming $\vec{\gamma}^{i-1}$ with a layer of two-qudit gates U^i . Particularly, we set $\vec{\gamma}^0$ equaling $\vec{\gamma}^{(0)}$ and define M^i such that

$$M^i(O) = \mathbb{E}_{U^i} U^{i \otimes 2} O U^{i \otimes 2}. \tag{C32}$$

Then,

$$\begin{aligned}
\Phi_{\mathcal{L}_n^d}^2(O_{\text{SR}}) &= M^d \cdots M^2 M^1 \left(\sum_{\vec{\gamma} \in \{\mathbb{I}, F\}^n} c_{\vec{\gamma}}(O_{\text{SR}}) \vec{\gamma} \right) \\
&= \sum_{\vec{\gamma} \in \{\mathbb{I}, F\}^n} c_{\vec{\gamma}}(O_{\text{SR}}) M^d \cdots M^2 M^1(\vec{\gamma}) \\
&= \sum_{\Gamma = (\vec{\gamma}^0, \dots, \vec{\gamma}^d) \in \{\mathbb{I}, F\}^{n \times (d+1)}} c_{\vec{\gamma}^0}(O_{\text{SR}}) \vec{\gamma}^d \prod_{i=1}^d M_{\vec{\gamma}^i \vec{\gamma}^{i-1}}^i. \tag{C33}
\end{aligned}$$

Here, $\Gamma = (\vec{\gamma}^0, \dots, \vec{\gamma}^d)$ is called a trajectory, which is a sequence of $d + 1$ configurations. Note that we view M^i as a transfer matrix as follows:

$$M^i(\vec{\gamma}) = \sum_{\vec{v} \in \{\mathbb{I}, F\}^n} M_{\vec{v} \vec{\gamma}}^i \vec{v}. \tag{C34}$$

Correspondingly,

$$\text{tr} O_{\text{SE}} \Phi_{\mathcal{L}_n^d}^2(O_{\text{SR}}) = \sum_{\Gamma} \text{weight}(\Gamma) c_{\vec{\gamma}^0 \vec{\gamma}^d}(O_{\text{SR}}, O_{\text{SE}}), \tag{C35}$$

where

$$\text{weight}(\Gamma) = \prod_{i=1}^d M_{\vec{\gamma}^i \vec{\gamma}^{i-1}}^i, \tag{C36}$$

and $c_{\vec{\gamma}^0 \vec{\gamma}^d}(O_{\text{SR}}, O_{\text{SE}})$ is given by Eq. (C16).

Following the analysis in Appendix C 1, we have that

$$\text{tr } O_{SE} \Phi_{\mathcal{L}_n^d}^2(O_{SR}) = \sum_{\vec{\gamma}^0 \vec{\gamma}^d} c_{\vec{\gamma}^0 \vec{\gamma}^d}(O_{SR}, O_{SE}) Z_{\vec{\gamma}^0 \vec{\gamma}^d}, \quad (\text{C37})$$

where

$$Z_{\vec{\gamma}^0 \vec{\gamma}^d} = \sum_{\Gamma} \left(\frac{q}{q^2 + 1} \right)^{\#F_{\Gamma}} \mathbf{1}_{\Gamma_1 = \vec{\gamma}^{(0)}} \mathbf{1}_{\Gamma_{d+1} = \vec{\gamma}^d}. \quad (\text{C38})$$

In the following, we will use the domain-wall method to analyze the above quantity.

a. Definition of domain wall

Given a configuration $\vec{\gamma} \in \{\mathbb{I}, F\}^n$, we define the associated set of domain-wall locations as

$$DW(\vec{\gamma}) := \{e \in \{1, 2, \dots, n\} : \gamma_e \neq \gamma_{e+1}\}. \quad (\text{C39})$$

Define $g^i = DW(\vec{\gamma}^i)$. The number of domain walls within g^i is denoted as $|g^i|$. For a trajectory $\Gamma = (\vec{\gamma}^0, \dots, \vec{\gamma}^d)$, we can define its domain-wall trajectory $G(\Gamma) = (g^0, \dots, g^d)$. If $\text{weight}(\Gamma) \neq 0$, then its associated domain-wall trajectory must satisfy the following rule: If there is a domain wall at location e , and a gate acts on qudits $\{e, e+1\}$, then the domain wall must move from e to $e-1$ or $e+1$. If there already exists a domain wall at these locations, the two domain walls will annihilate with each other. On the other hand, a pair of domain walls cannot be created from a vacuum. Meanwhile, due to the boundary condition, the number of domain walls must be even.

Based on Eq. (2), the l th layer quantum gate $U^{[l]}$ only contains two-qudit gates acting on $\{2i - \text{mod}(l, 2), 2i + 1 - \text{mod}(l, 2)\}$. Thus, the domain walls within g^l must have opposite parity with respect to l . For a specific domain wall within g^0 , it may move or stay still at the first step. Starting with the second step, the domain wall will keep moving and survive until the final step or annihilate with another domain wall during the movement. The number of domain walls cannot increase within a trajectory.

For each trajectory Γ , its weight, $\text{weight}(\Gamma) = (q/(q^2 + 1))^{\#F_{\Gamma}}$, is uniquely determined by its associated domain-wall trajectory $G(\Gamma)$. Each movement of a domain wall within $G(\Gamma)$ corresponds to a flip. Hence, we denote $\text{weight}(\Gamma) = \text{weight}(G(\Gamma))$.

A domain-wall trajectory G can always be decomposed into two disjoint parts. That is, we define $G_1 = (g_1^0, \dots, g_1^d)$ and $G_2 = (g_2^0, \dots, g_2^d)$ such that

$$G = G_1 \sqcup G_2 \equiv (g_1^0 \sqcup g_2^0, \dots, g_1^d \sqcup g_2^d), \quad (\text{C40})$$

where $\sqcup$ is the disjoint union. Thus, $g_1^l \cap g_2^l = \emptyset$ for all l . This decomposition satisfies that $\text{weight}(G) =$

$\text{weight}(G_1)\text{weight}(G_2)$ [32]. This relation is essential in evaluating the bound of $\text{weight}(G)$.

For each trajectory Γ , we decompose its domain-wall trajectory $G(\Gamma)$ into two disjoint domain-wall trajectories, $G(\Gamma) = G_U \sqcup G_0$, as shown in Fig. 6. Here, G_U has a conserved number of domain walls, and the domain walls of G_0 will annihilate before the final step. That is, $g_U^d = g^d$, and $g_0^d = \emptyset$. This decomposition is unique [32]. In the following, we separately give the analysis of the upper bound and lower bound for $\text{weight}(\Gamma)$ and $Z_{\vec{\gamma}^0 \vec{\gamma}^d}$. For further elaboration, we denote $\mathcal{G}_U$ and $\mathcal{G}_0$ as the subsets of domain-wall trajectories without annihilation and without a surviving domain wall at the end, respectively. Also, we denote $\mathcal{G}_{U,k} \subset \mathcal{G}_U$ as the set of domain-wall trajectories always with k domain walls. For a 1D periodic circuit structure, k must be even to make $\mathcal{G}_{U,k} \neq \emptyset$.

b. Upper bound of $Z_{\vec{\gamma}^0 \vec{\gamma}^d}$

In this part, we derive the upper bound of $Z_{\vec{\gamma}^0 \vec{\gamma}^d}$, presented in the following lemma.

Lemma C1. Given the set of trajectories Γ with initial and final configurations as $\vec{\gamma}^0$ and $\vec{\gamma}^d$, we divide these trajectories into different sets with fixed G_U ,

$$\mathcal{S}_U^G = \{\Gamma | G(\Gamma) = G_U \sqcup G_0\}. \quad (\text{C41})$$

Given the domain-wall trajectory G_U with initial configuration $g_U^0 = \{e_1^0, e_2^0, \dots, e_{2k_0-1}^0, e_{2k_0}^0\}$. Here, $2k_0 = |DW(\vec{\gamma}^d)|$ is even. Denote $E_{\text{odd}}^0 = \{1, \dots, e_1^0, e_2^0 + 1, \dots, e_{2k_0}^0, \dots, e_{2k_0+1}^0, \dots, n\}$ and $E_{\text{even}}^0 = [n] \setminus E_{\text{odd}}^0 = \overline{E_{\text{odd}}^0}$. We also denote $\vec{\gamma}_A$ as the component of $\vec{\gamma}$ within support A , and denote $|\vec{\gamma}_A| = F$ and $|\vec{\gamma}_A| = \mathbb{I}$ as the number of SWAP and identity operators within $\vec{\gamma}_A$, respectively. Then,

$$Z_{\vec{\gamma}^0 \vec{\gamma}^d} \leq \sum_{\mathcal{S}_U^G} \eta^{(d-1)|DW(\vec{\gamma}^d)|} q^{-\left| \vec{\gamma}_{E_{\text{odd}}^0}^0 = \vec{\gamma}_{(1)}^d \right| - \left| \vec{\gamma}_{E_{\text{even}}^0}^0 = \vec{\gamma}_{(1)}^d \right|}. \quad (\text{C42})$$

Here, we denote $\overline{F} = \mathbb{I}$ and $\overline{\mathbb{I}} = F$. And $\eta = q/(q^2 + 1)$.

Proof. Recall that

$$Z_{\vec{\gamma}^0 \vec{\gamma}^d} = \sum_{\Gamma} \text{weight}(\Gamma) \mathbf{1}_{\Gamma_1 = \vec{\gamma}^0} \mathbf{1}_{\Gamma_{d+1} = \vec{\gamma}^d}. \quad (\text{C43})$$

The summation is over all trajectories from $\vec{\gamma}^0$ to $\vec{\gamma}^d$ with nonzero weight. Note that $\text{weight}(\Gamma) = \text{weight}(G(\Gamma))$. We can analyze the domain-wall trajectories of all Γ for evaluating the weight. Note that Γ has a fixed ending configuration $\vec{\gamma}^d$, implying that $G(\Gamma)$ also has a fixed ending point $DW(\vec{\gamma}^d)$. Following the analysis of the previous subsection, the domain-wall trajectory $G(\Gamma)$ can be uniquely

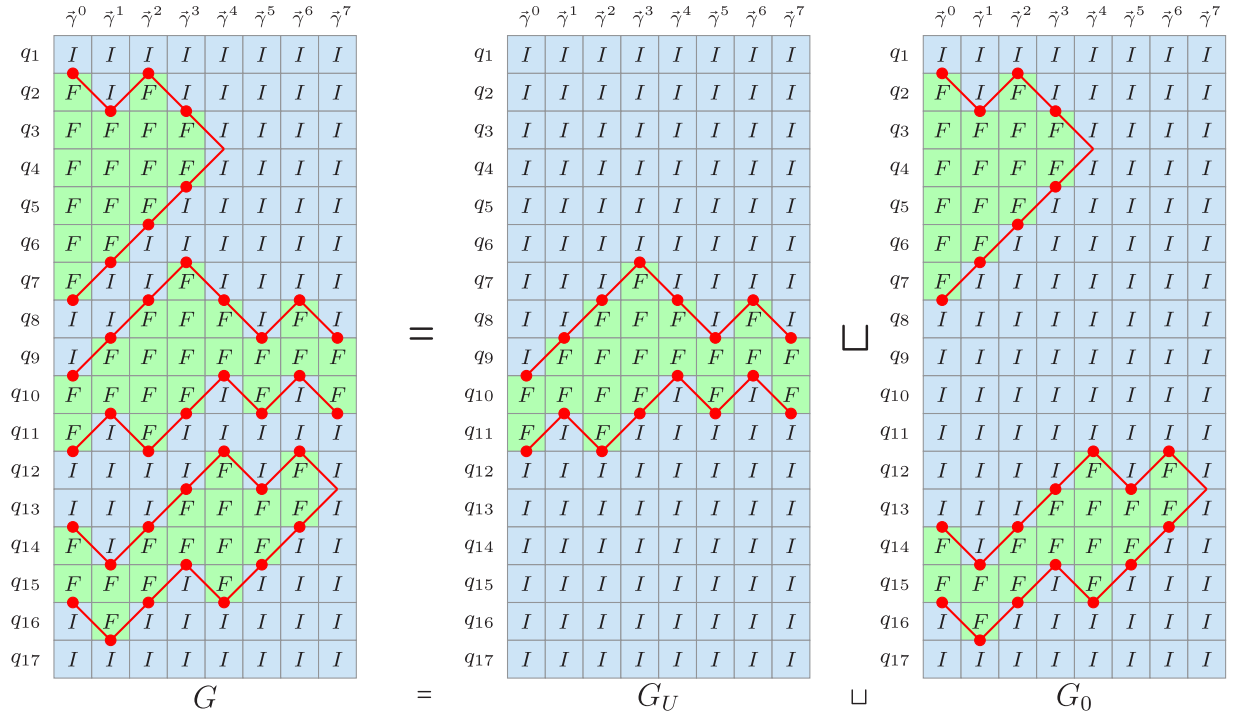


FIG. 6. Diagram for the domain-wall trajectory decomposition. The total qubit number is 17 and the circuit depth is 7. The first figure is a trajectory Γ with corresponding domain-wall trajectory $G(\Gamma)$. The number of domain walls within $G(\Gamma)$ does not conserve. $G(\Gamma)$ can be decomposed into two disjoint domain-wall trajectories, $G(\Gamma) = G_U \sqcup G_0$. G_U always contains two domain walls, and the domain walls of G_0 annihilate at the final depth.

decomposed into the disjoint union of two domain-wall trajectories G_U and G_0 such that $g_U^d = DW(\vec{\gamma}^d)$ and $g_0^d = \emptyset$. After the decomposition, we determine and denote g_U^0 as the initial point of the domain-wall trajectory G_U . Note that $G_U \in \mathcal{G}_{U, |DW(\vec{\gamma}^d)|}$ and $|g_U^0| = |g_U^d| = |DW(\vec{\gamma}^d)|$.

We then divide all trajectories Γ , equivalently, all $G(\Gamma)$, according to G_U . Particularly, we define the set of trajectories with fixed G_U ,

$$\mathcal{S}_U^G = \{\Gamma | G(\Gamma) = G_U \sqcup G_0\}. \quad (\text{C44})$$

Given a fixed $\vec{\gamma}^d$, the number of distinct G_U or different sets $\mathcal{S}_U^G$ is upper-bounded by $2^{|DW(\vec{\gamma}^d)|}$. For each domain wall e inside $DW(\vec{\gamma}^d)$, moving it forward only has two directions. Moving it back to the first layer only has 2^d choices. Since there is only $|DW(\vec{\gamma}^d)|$ domain walls, moving all of them back to the first layer only has $2^{|DW(\vec{\gamma}^d)|}$ options, which gives the upper bound of the number of distinct G_U and $\mathcal{S}_U^G$. Then, we have

$$Z_{\vec{\gamma}^0 \vec{\gamma}^d} = \sum_{\mathcal{S}_U^G} \sum_{\Gamma \in \mathcal{S}_U^G} \text{weight}(\Gamma) \mathbf{1}_{\Gamma_1 = \vec{\gamma}^0} \mathbf{1}_{\Gamma_{d+1} = \vec{\gamma}^d}. \quad (\text{C45})$$

Below, we analyze the bound of $\sum_{\Gamma \in \mathcal{S}_U^G} \text{weight}(\Gamma)$. For each trajectory Γ , we have $G(\Gamma) = G_U \sqcup G_0$. Since

there is no annihilation of domain walls within G_U , the number of domain-wall movements is at least $(d-1) |DW(\vec{\gamma}^d)|$, implying that $\text{weight}(G_U) \leq \eta^{(d-1) |DW(\vec{\gamma}^d)|}$ with $\eta = q/(q^2 + 1)$. Thus,

$$\begin{aligned} \text{weight}(\Gamma) &= \text{weight}(G_U) \text{weight}(G_0) \\ &\leq \eta^{(d-1) |DW(\vec{\gamma}^d)|} \text{weight}(G_0) \end{aligned} \quad (\text{C46})$$

and

$$\sum_{\Gamma \in \mathcal{S}_U^G} \text{weight}(\Gamma) \leq \eta^{(d-1) |DW(\vec{\gamma}^d)|} \sum_{\Gamma \in \mathcal{S}_U^G} \text{weight}(G_0). \quad (\text{C47})$$

The following derivation is to further evaluate the upper bound of $\sum_{\Gamma \in \mathcal{S}_U^G} \text{weight}(G_0)$. Given the domain-wall trajectory G_U with initial configuration g_U^0 and ending configuration $g_U^d = DW(\vec{\gamma}^d)$, we write g_U^i as $\{e_1^i, e_2^i, \dots, e_{2k_0-1}^i, e_{2k_0}^i\}$. Here, $2k_0 = |DW(\vec{\gamma}^d)|$ is even. We denote the support $E_{\text{odd}}^i = \{1, \dots, e_1^i, e_2^i + 1, \dots, e_{2k_0-1}^i, \dots, e_{2k_0}^i + 1, \dots, n\}$ and $E_{\text{even}}^i = [n] \setminus E_{\text{odd}}^i$ as demonstrated in Fig. 7. We also denote $\vec{\gamma}_A$ as the component of $\vec{\gamma}$ within support A , and denote $|\vec{\gamma}_A = F|$ and $|\vec{\gamma}_A = \mathbb{I}|$ as the number of SWAP and identity operators within $\vec{\gamma}_A$, respectively. We prove the following lemma.

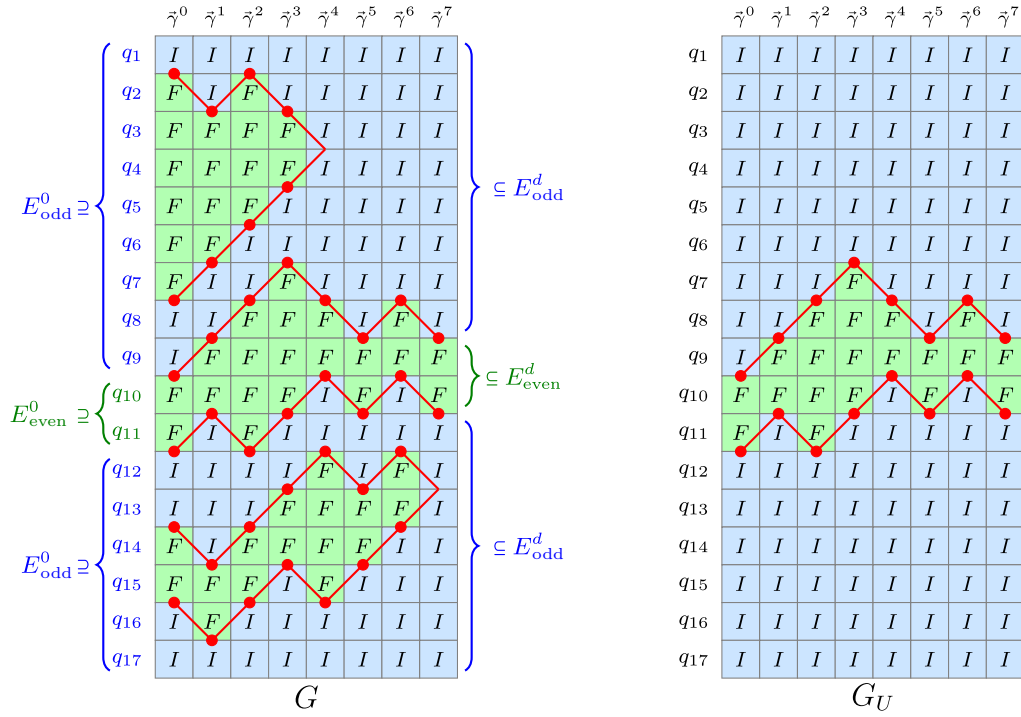


FIG. 7. Diagram to show the supports E^i_{odd} and E^i_{even} within G . We also draw G_U on the right as a reference.

Lemma C2. If $\vec{\gamma}_{\{1\}}^d = \mathbb{I}$,

$$\sum_{\Gamma \in \mathcal{S}_U^G} \text{weight}(G_0) \leq q^{-|\vec{\gamma}_{E^0_{\text{odd}}}^0 = F| - |\vec{\gamma}_{E^0_{\text{even}}}^0 = \mathbb{I}|}; \quad (\text{C48})$$

if $\vec{\gamma}_{\{1\}}^d = F$,

$$\sum_{\Gamma \in \mathcal{S}_U^G} \text{weight}(G_0) \leq q^{-|\vec{\gamma}_{E^0_{\text{odd}}}^0 = \mathbb{I}| - |\vec{\gamma}_{E^0_{\text{even}}}^0 = F|}. \quad (\text{C49})$$

Proof. For $\Gamma \in \mathcal{S}_U^G$, G_U within the decomposition of Γ is fixed. After fixing G_U , we can separate G_0 into $|DW(\vec{\gamma}^d)|$ regions.

In step i , the j th region is defined as $R_j^i = \{e_{j-1}^i + 1, e_{j-1}^i + 2, \dots, e_j^i\}$. And we divide $g_0^i \in G_0$ into the union of its parts in different regions, $g_0^i = g_0^i|_{R_1^i} + g_0^i|_{R_2^i} + \dots + g_0^i|_{R_{|DW(\vec{\gamma}^d)|}^i}$. In the overall picture, the region j is the union of all R_j^i across different steps. Correspondingly, we can decompose G_0 into the disjoint union of trajectories in different regions: $G_0 = \bigsqcup_{j=1}^{|DW(\vec{\gamma}^d)|} G_0^j$ where $G_0^j = (g_0^0|_{R_j^0}, g_0^1|_{R_j^1}, \dots, g_0^d|_{R_j^d})$. An illustration of regions is demonstrated in Fig. 8.

Since $G_0 = \bigsqcup_{j=1}^{|DW(\vec{\gamma}^d)|} G_0^j$, we have that

$$\text{weight}(G_0) = \prod_{j=1}^{|DW(\vec{\gamma}^d)|} \text{weight}(G_0^j) \quad (\text{C50})$$

and

$$\sum_{\Gamma \in \mathcal{S}_U^G} \text{weight}(G_0) = \prod_{j=1}^{|DW(\vec{\gamma}^d)|} \sum_{\Gamma|_{\text{Region } j}} \text{weight}(G_0^j). \quad (\text{C51})$$

In the following, we will derive the upper bound of $\sum_{\Gamma|_{\text{Region } j}} \text{weight}(G_0^j)$. Since the domain walls within G_0 annihilate in the end, $\vec{\gamma}^d$ must be the tensor product of all identity operations or SWAP operations within a region. Depending on which case holds, we call this region a type- $\mathbb{I}$ region or a type- F region. If Region j is a type- $\mathbb{I}$ region, we will show that

$$\sum_{\Gamma|_{\text{Region } j}} \text{weight}(G_0^j) \leq q^{-|\vec{\gamma}_{R_j^0}^0 = F|}. \quad (\text{C52})$$

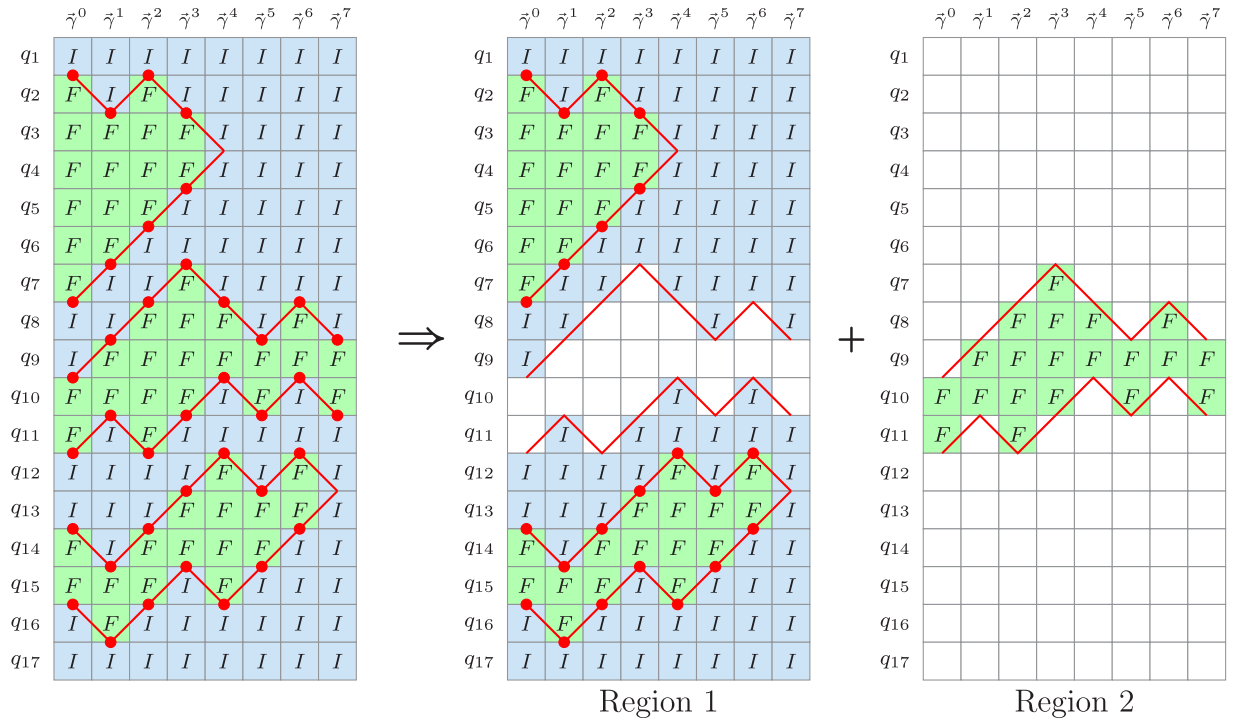


FIG. 8. Diagram of separating a trajectory into two regions.

If Region j is a type- F region, we will show that

$$\sum_{\Gamma|_{\text{Region } j}} \text{weight}(G_0^j) \leq q^{-|\vec{\gamma}_{R_j^0}^0|}. \quad (\text{C53})$$

We first focus on the type-II region and prove Eq. (C52). An example of a type-II region is shown in Fig. 9. It is Region 1 in Fig. 8. Given this region, we first extend it to a wider region. Particularly, we define $e_{j-1}^{\min} = \min_{i \in \{0,1,\dots,d\}} e_j^i$, $e_j^{\max} = \max_{i \in \{0,1,\dots,d\}} e_j^i$, and $R_j = \{e_{j-1}^{\min} + 1, e_{j-1}^{\min} + 2, \dots, e_j^{\max}\}$. Then for each configuration $\vec{\gamma}^i|_{R_j^i}$, we extend it to $\vec{\gamma}'^i|_{R_j}$ by filling the extra components with identity operations. Region j is correspondingly extended to Region j' as shown in Fig. 9. By extending the region, we make sure that the configurations of different steps have equal length. Also, since we extend the region, the number of valid domain-wall configurations within the new region can only increase. Thus, we have

$$\sum_{\Gamma|_{\text{Region } j}} \text{weight}(G_0^j) \leq \sum_{\Gamma'|_{\text{Region } j}} \text{weight}(G_0^j). \quad (\text{C54})$$

Also, the above trajectory is induced by 1D depth- d brickwork circuits. If we consider all trajectories induced by infinite-size circuits, the number of valid domain-wall

configurations also increases. Thus,

$$\sum_{\Gamma'_{1D}, \Gamma'|_{\text{Region } j}} \text{weight}(G_0^j) \leq \sum_{\Gamma'_{\text{inf}}, \Gamma'|_{\text{Region } j}} \text{weight}(G_0^j). \quad (\text{C55})$$

For the extended configurations, using Eqs. (C23) and (C24), we have that

$$\begin{aligned} \sum_{\Gamma'_{\text{inf}}|_{\text{Region } j}} \text{weight}(G_0^j) &= q^{-|\vec{\gamma}^0|_{\text{Region } j}=F}} \frac{1 - q^{-2n+2}|\vec{\gamma}^0|}{1 - q^{-2n}} \\ &\leq q^{-|\vec{\gamma}'^0|_{\text{Region } j}=F}} = q^{-|\vec{\gamma}^0|_{R_j^0}=F}}. \end{aligned} \quad (\text{C56})$$

Hence,

$$\sum_{\Gamma|_{\text{Region } j}} \text{weight}(G_0^j) \leq q^{-|\vec{\gamma}^0|_{R_j^0}=F}}. \quad (\text{C57})$$

Similarly, an example of a type- F region is shown in Fig. 10, which is Region 2 in Fig. 8. Similar to previous arguments, for each configuration $\vec{\gamma}^i|_{R_j^i}$, we extend it to $\vec{\gamma}'^i|_{R_j}$ by filling the extra components with SWAP operations. Region j is correspondingly extended to Region j'

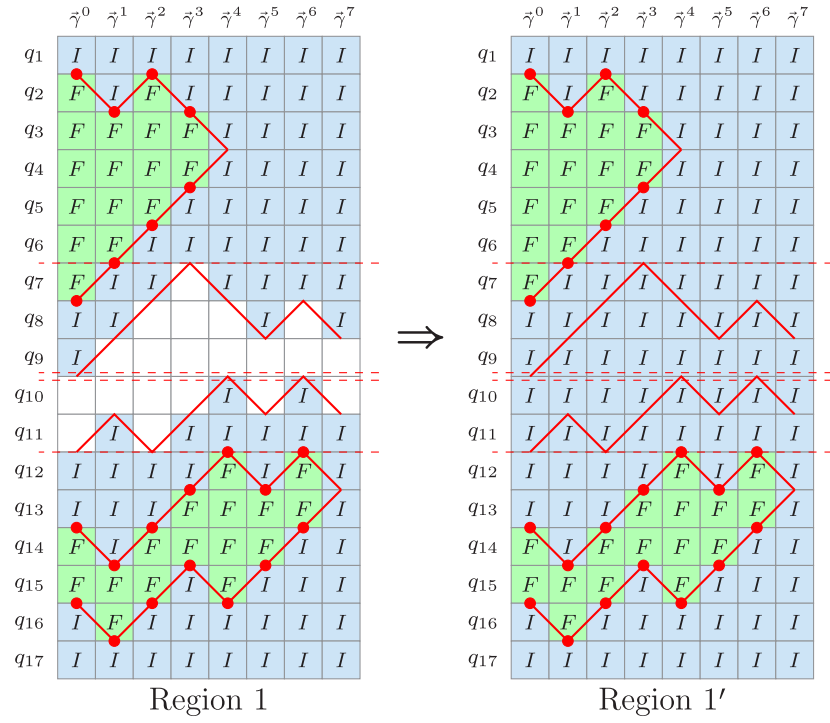


FIG. 9. Extending Region 1 to Region 1' by filling the extra components with identity operations. The configurations in the new region all have equal lengths.

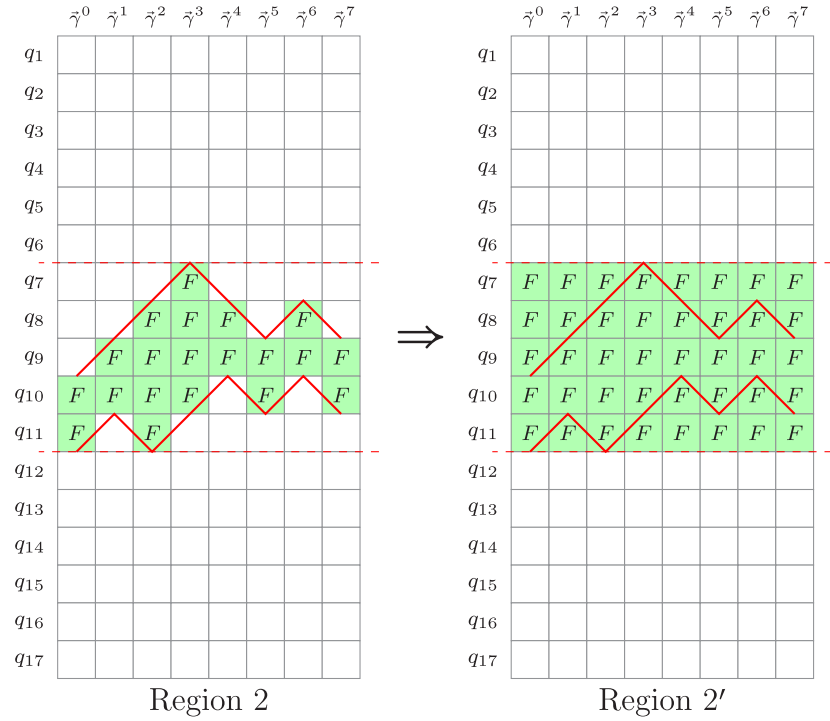


FIG. 10. Extending Region 2 to Region 2' by filling the extra components with SWAP operations. The configurations in the new region all have equal lengths.

as shown in Fig. 9. Then, we have

$$\begin{aligned}
 \sum_{\Gamma | \text{Region } j} \text{weight}(G_0^j) &\leq \sum_{\Gamma' | \text{Region } j} \text{weight}(G_0^j) \\
 &\leq \sum_{\Gamma'_{inf}, \Gamma' | \text{Region } j} \text{weight}(G_0^j) \\
 &= q^{-|\tilde{\gamma}^0|_{\text{Region } j} = \mathbb{I}} \frac{q^{2|\tilde{\gamma}^0|_{\text{Region } j} = \mathbb{I}} - q^{-2n}}{1 - q^{-2n}} \\
 &\leq q^{-|\tilde{\gamma}^0|_{R_j^0} = \mathbb{I}} \\
 &= q^{-|\tilde{\gamma}^0|_{R_j^0} = \mathbb{I}}. \tag{C58}
 \end{aligned}$$

Hence, Eqs. (C52) and (C53) are proved. Combining these two equations with Eq. (C51), we obtain that

$$\sum_{\Gamma \in S_U^G} \text{weight}(G_0) \leq \prod_{j: \text{type-}\mathbb{I}} q^{-|\tilde{\gamma}^0|_{R_j^0} = F} \prod_{j: \text{type-}F} q^{-|\tilde{\gamma}^0|_{R_j^0} = \mathbb{I}}. \tag{C59}$$

When $\tilde{\gamma}_{\{1\}}^d = \mathbb{I}$, Eq. (C59) reduces to

$$\sum_{\Gamma \in S_U^G} \text{weight}(G_0) \leq q^{-|\tilde{\gamma}_{E_{\text{odd}}}^0 = F} - |\tilde{\gamma}_{E_{\text{even}}}^0 = \mathbb{I}}|. \tag{C60}$$

When $\tilde{\gamma}_{\{1\}}^d = F$, Eq. (C59) reduces to

$$\sum_{\Gamma \in S_U^G} \text{weight}(G_0) \leq q^{-|\tilde{\gamma}_{E_{\text{odd}}}^0 = \mathbb{I}} - |\tilde{\gamma}_{E_{\text{even}}}^0 = F}|. \tag{C61}$$

Proof is completed. $\blacksquare$

Applying Lemma C2 and Eq. (C47) to Eq. (C45), we get

$$Z_{\tilde{\gamma}^0 \tilde{\gamma}^d} \leq \sum_{S_U^G} \eta^{(d-1)|DW(\tilde{\gamma}^d)|} q^{-|\tilde{\gamma}_{E_{\text{odd}}}^0 = \overline{\tilde{\gamma}_{\{1\}}^d}| - |\tilde{\gamma}_{E_{\text{even}}}^0 = \tilde{\gamma}_{\{1\}}^d|}. \tag{C62}$$

Note that we denote $\overline{F} = \mathbb{I}$ and $\overline{\mathbb{I}} = F$. $\blacksquare$

c. Proof of Theorem 2

Applying Lemma C1, we get the upper bound of $\text{tr } O_{\text{SE}} \Phi_{\mathcal{L}_n^d}^2(O_{\text{SR}})$. Recall that

$$\text{tr } O_{\text{SE}} \Phi_{\mathcal{L}_n^d}^2(O_{\text{SR}}) = \sum_{\tilde{\gamma}^0 \tilde{\gamma}^d} c_{\tilde{\gamma}^0 \tilde{\gamma}^d}(O_{\text{SR}}, O_{\text{SE}}) Z_{\tilde{\gamma}^0 \tilde{\gamma}^d}, \tag{C63}$$

where

$$\begin{aligned}
 c_{\tilde{\gamma}^0 \tilde{\gamma}^d}(O_{\text{SR}}, O_{\text{SE}}) &= \frac{1}{(q^2 - 1)^n} \text{tr} \left(O_{\text{SR}} \bigotimes_{j=1}^n g_{\gamma_j^{(0)}} \right) \text{tr} \left(O_{\text{SE}} \bigotimes_{j=1}^n \gamma_j^{(s)} \right). \tag{C64}
 \end{aligned}$$

Note that $O_{\text{SR}} = \text{tr}_R(\tilde{\rho}_{\text{SR}}^{\otimes 2} F_R)$ and $O_{\text{SE}} = (\tilde{\mathcal{T}}_{S \rightarrow E}^{\dagger})^{\otimes 2}(F_E)$. Below, we utilize the tensor-product structure of ρ_{SR} , $\rho_{\text{SR}} = \bigotimes_{i=1}^n \rho_{S_i R_i}$. Then, $\tilde{\rho}_{\text{SR}} = \bigotimes_{i=1}^n \tilde{\rho}_{S_i R_i}$ and $O_{\text{SR}} = \bigotimes_{i=1}^n O_{S_i R_i}$. Since $\text{tr}(O_{S_i R_i} g_{\mathbb{I}}) = \text{tr}(\tilde{\rho}_{R_i}^2) - q^{-1} \text{tr}(\tilde{\rho}_{S_i R_i}^2) \geq 0$ and $\text{tr}(O_{S_i R_i} g_F) = \text{tr}(\tilde{\rho}_{S_i R_i}^2) - q^{-1} \text{tr}(\tilde{\rho}_{R_i}^2) \geq 0$, we have $\text{tr}(O_{\text{SR}} \bigotimes_{j=1}^n g_{\gamma_j^{(0)}}) \geq 0$. Meanwhile, $\text{tr}(O_{\text{SE}} F_A \otimes \mathbb{I}_{\tilde{A}}) = q^{2n} \text{tr}_{\text{AE}}(\text{tr}_{\tilde{A}} \tilde{\tau}_{\text{SE}})^2 \geq 0$. Thus, we will always have $c_{\tilde{\gamma}^0 \tilde{\gamma}^d}(O_{\text{SR}}, O_{\text{SE}}) \geq 0$. Then,

$$\begin{aligned}
 \text{tr } O_{\text{SE}} \Phi_{\mathcal{L}_n^d}^2(O_{\text{SR}}) &= \sum_{\tilde{\gamma}^0 \tilde{\gamma}^d} c_{\tilde{\gamma}^0 \tilde{\gamma}^d}(O_{\text{SR}}, O_{\text{SE}}) Z_{\tilde{\gamma}^0 \tilde{\gamma}^d} \\
 &\leq \sum_{\tilde{\gamma}^0 \tilde{\gamma}^d} c_{\tilde{\gamma}^0 \tilde{\gamma}^d}(O_{\text{SR}}, O_{\text{SE}}) \sum_{S_U^G} \eta^{(d-1)|DW(\tilde{\gamma}^d)|} q^{-|\tilde{\gamma}_{E_{\text{odd}}}^0 = \overline{\tilde{\gamma}_{\{1\}}^d}| - |\tilde{\gamma}_{E_{\text{even}}}^0 = \tilde{\gamma}_{\{1\}}^d|} \\
 &= \sum_{\tilde{\gamma}^0 \tilde{\gamma}^d} \frac{1}{(q^2 - 1)^n} \text{tr} \left(O_{\text{SR}} \bigotimes_{j=1}^n g_{\gamma_j^{(0)}} \right) \text{tr} \left(O_{\text{SE}} \bigotimes_{j=1}^n \gamma_j^d \right) \sum_{S_U^G} \eta^{(d-1)|DW(\tilde{\gamma}^d)|} q^{-|\tilde{\gamma}_{E_{\text{odd}}}^0 = \overline{\tilde{\gamma}_{\{1\}}^d}| - |\tilde{\gamma}_{E_{\text{even}}}^0 = \tilde{\gamma}_{\{1\}}^d|}
 \end{aligned}$$

$$\begin{aligned}
&= \sum_{\vec{\gamma}^d, S_U^G} \frac{\eta^{(d-1)|DW(\vec{\gamma}^d)|}}{(q^2 - 1)^n} \sum_{\vec{\gamma}^0} \text{tr} \left(O_{\text{SR}} \bigotimes_{j=1}^n g_{\gamma_j^0} \right) \text{tr} \left(O_{\text{SE}} \bigotimes_{j=1}^n \gamma_j^d \right) q^{-\left| \vec{\gamma}_{\text{odd}}^0 = \vec{\gamma}_{\{1\}}^d \right| - \left| \vec{\gamma}_{\text{even}}^0 = \vec{\gamma}_{\{1\}}^d \right|} \\
&= \sum_{DW(\vec{\gamma}^d), S_U^G} \frac{\eta^{(d-1)|DW(\vec{\gamma}^d)|}}{(q^2 - 1)^n} \sum_{\vec{\gamma}^0} \text{tr} \left(O_{\text{SR}} \bigotimes_{j=1}^n g_{\gamma_j^0} \right) \text{tr} \left(O_{\text{SE}} \bigotimes_{j=1}^n \gamma_j^d \right) \left(q^{-\left| \vec{\gamma}_{\text{odd}}^0 = \mathbb{I} \right| - \left| \vec{\gamma}_{\text{even}}^0 = F \right|} + q^{-\left| \vec{\gamma}_{\text{odd}}^0 = F \right| - \left| \vec{\gamma}_{\text{even}}^0 = \mathbb{I} \right|} \right) \\
&= \sum_{DW(\vec{\gamma}^d), S_U^G} \eta^{(d-1)|DW(\vec{\gamma}^d)|} q^{-2n} \left(\text{tr} O_{\text{SR}} \mathbb{I}_{E_{\text{odd}}^0} F_{E_{\text{even}}^0} \text{tr} O_{\text{SE}} \mathbb{I}_{E_{\text{odd}}^d} F_{E_{\text{even}}^d} + \text{tr} O_{\text{SR}} \mathbb{I}_{E_{\text{even}}^0} F_{E_{\text{odd}}^0} \text{tr} O_{\text{SE}} \mathbb{I}_{E_{\text{even}}^d} F_{E_{\text{odd}}^d} \right) \\
&= q^{-2n} \left(\text{tr} O_{\text{SR}} \mathbb{I}^n \text{tr} O_{\text{SE}} \mathbb{I}^n + \text{tr} O_{\text{SR}} F^n \text{tr} O_{\text{SE}} F^n \right) + q^{-2n} \sum_{|DW(\vec{\gamma}^d)| \in \{2, 4, \dots, 2\lfloor n/4 \rfloor\}} \eta^{(d-1)|DW(\vec{\gamma}^d)|} \\
&\quad \times \sum_{S_U^G} \left(\text{tr} O_{\text{SR}} \mathbb{I}_{E_{\text{odd}}^0} F_{E_{\text{even}}^0} \text{tr} O_{\text{SE}} \mathbb{I}_{E_{\text{odd}}^d} F_{E_{\text{even}}^d} + \text{tr} O_{\text{SR}} \mathbb{I}_{E_{\text{even}}^0} F_{E_{\text{odd}}^0} \text{tr} O_{\text{SE}} \mathbb{I}_{E_{\text{even}}^d} F_{E_{\text{odd}}^d} \right). \tag{C65}
\end{aligned}$$

Here, $\mathbb{I}_A$ and F_A means identity and SWAP operations on the support A . The last inequality comes from Eq. (C28). Note that

$$q^{-2n} \text{tr} O_{\text{SR}} \mathbb{I}^n \text{tr} O_{\text{SE}} \mathbb{I}^n = \text{tr}(\tilde{\tau}_E^2) \text{tr}(\tilde{\rho}_R^2), \tag{C66}$$

$$q^{-2n} \text{tr} O_{\text{SR}} F^n \text{tr} O_{\text{SE}} F^n = \text{tr}(\tilde{\tau}_{\text{SE}}^2) \text{tr}(\tilde{\rho}_{\text{SR}}^2). \tag{C67}$$

Also,

$$\begin{aligned}
q^{-2n} \text{tr} O_{\text{SR}} \mathbb{I}_{E_{\text{odd}}^0} F_{E_{\text{even}}^0} \text{tr} O_{\text{SE}} \mathbb{I}_{E_{\text{odd}}^d} F_{E_{\text{even}}^d} &\leq \max_{E^0} \left(\frac{\text{tr} O_{\text{SR}} \mathbb{I}_{E_{\text{odd}}^0} F_{E_{\text{even}}^0}}{\text{tr} O_{\text{SR}} \mathbb{I}_{E_{\text{odd}}^d} F_{E_{\text{even}}^d}} \right) \max_{E^d} (q^{-2n} \text{tr} O_{\text{SR}} \mathbb{I}_{E_{\text{odd}}^d} F_{E_{\text{even}}^d} \text{tr} O_{\text{SE}} \mathbb{I}_{E_{\text{odd}}^d} F_{E_{\text{even}}^d}) \\
&\leq \max_{i \in [n]} \left(\frac{\text{tr} \tilde{\rho}_{S_i R_i}^2}{\text{tr} \tilde{\rho}_{R_i}^2}, \frac{\text{tr} \tilde{\rho}_{R_i}^2}{\text{tr} \tilde{\rho}_{S_i R_i}^2} \right)^{\left| E_{\text{even}}^d - E_{\text{even}}^0 \right|} \max_{A \subseteq [n]} (\text{tr}_{\text{AR}} (\text{tr}_{\bar{A}} \tilde{\rho}_{\text{SR}})^2 \text{tr}_{\text{AE}} (\text{tr}_{\bar{A}} \tilde{\tau}_{\text{SE}})^2) \\
&\leq \max_{i \in [n]} \left(\frac{\text{tr} \tilde{\rho}_{S_i R_i}^2}{\text{tr} \tilde{\rho}_{R_i}^2}, \frac{\text{tr} \tilde{\rho}_{R_i}^2}{\text{tr} \tilde{\rho}_{S_i R_i}^2} \right)^{d|DW(\vec{\gamma}^d)|} \max_{A \subseteq [n]} (\text{tr}_{\text{AR}} (\text{tr}_{\bar{A}} \tilde{\rho}_{\text{SR}})^2 \text{tr}_{\text{AE}} (\text{tr}_{\bar{A}} \tilde{\tau}_{\text{SE}})^2). \tag{C68}
\end{aligned}$$

Similarly,

$$q^{-2n} \text{tr} O_{\text{SR}} \mathbb{I}_{E_{\text{even}}^0} F_{E_{\text{odd}}^0} \text{tr} O_{\text{SE}} \mathbb{I}_{E_{\text{even}}^d} F_{E_{\text{odd}}^d} \leq \max_{i \in [n]} \left(\frac{\text{tr} \tilde{\rho}_{S_i R_i}^2}{\text{tr} \tilde{\rho}_{R_i}^2}, \frac{\text{tr} \tilde{\rho}_{R_i}^2}{\text{tr} \tilde{\rho}_{S_i R_i}^2} \right)^{d|DW(\vec{\gamma}^d)|} \max_{A \subseteq [n]} (\text{tr}_{\text{AR}} (\text{tr}_{\bar{A}} \tilde{\rho}_{\text{SR}})^2 \text{tr}_{\text{AE}} (\text{tr}_{\bar{A}} \tilde{\tau}_{\text{SE}})^2). \tag{C69}$$

Note that $|\mathcal{S}_U^G| \leq 2^{d|DW(\vec{\gamma}^d)|}$. Then,

$$\begin{aligned}
\text{tr} O_{\text{SE}} \Phi_{\mathcal{L}_n^d}^2(O_{\text{SR}}) &\leq \text{tr}(\tilde{\tau}_E^2) \text{tr}(\tilde{\rho}_R^2) + \text{tr}(\tilde{\tau}_{\text{SE}}^2) \text{tr}(\tilde{\rho}_{\text{SR}}^2) + \sum_{|DW(\vec{\gamma}^d)| \in \{2, 4, \dots, 2\lfloor n/4 \rfloor\}} \eta^{(d-1)|DW(\vec{\gamma}^d)|} \\
&\quad \times (2\rho_m)^{d|DW(\vec{\gamma}^d)|} \max_{A \subseteq [n]} (\text{tr}_{\text{AR}} (\text{tr}_{\bar{A}} \tilde{\rho}_{\text{SR}})^2 \text{tr}_{\text{AE}} (\text{tr}_{\bar{A}} \tilde{\tau}_{\text{SE}})^2) \\
&\leq \text{tr}(\tilde{\tau}_E^2) \text{tr}(\tilde{\rho}_R^2) + \text{tr}(\tilde{\tau}_{\text{SE}}^2) \text{tr}(\tilde{\rho}_{\text{SR}}^2) + \max_{A \subseteq [n]} (\text{tr}_{\text{AR}} (\text{tr}_{\bar{A}} \tilde{\rho}_{\text{SR}})^2 \text{tr}_{\text{AE}} (\text{tr}_{\bar{A}} \tilde{\tau}_{\text{SE}})^2) \sum_{k_0=1}^{\lfloor n/4 \rfloor} \binom{n}{2k_0} ((2\rho_m)^d \eta^{(d-1)2k_0}), \tag{C70}
\end{aligned}$$

where we denote $\rho_m = \max_{i \in [n]} (\text{tr} \tilde{\rho}_{S_i R_i}^2 / \text{tr} \tilde{\rho}_{R_i}^2, \text{tr} \tilde{\rho}_{R_i}^2 / \text{tr} \tilde{\rho}_{S_i R_i}^2)$ for simplicity. In the following, we analyze the coefficient

$$c = \sum_{k_0=1}^{\lfloor n/4 \rfloor} \binom{n}{2k_0} ((2\rho_m)^d \eta^{(d-1)2k_0}). \quad (\text{C71})$$

From Ref. [32], coefficient c has an upper bound,

$$c = \sum_{k_0=1}^{\lfloor n/4 \rfloor} \binom{n}{2k_0} ((2\rho_m)^d \eta^{(d-1)2k_0}) \leq (\eta(2\rho_m)^{\frac{d}{d-1}})^{d-d^*}, \quad (\text{C72})$$

where

$$d^* = \frac{\log n}{\log \frac{1}{2\eta}} + \frac{\log(e-1)}{\log \frac{1}{2\eta}} + 1. \quad (\text{C73})$$

Recall that we choose depth d such that $d - d^* = \log(n/\varepsilon) = O(\log n)$. In the large n limit, $\eta(2\rho_m)^{\frac{d}{d-1}} \leq 2\eta\rho_m + \delta$ with δ an arbitrarily small number. Then,

$$c \leq (2\eta\rho_m + \delta)^{\log \frac{n}{\varepsilon}} = \left(\frac{\varepsilon}{n}\right)^{\log \frac{1}{2\eta\rho_m + \delta}}. \quad (\text{C74})$$

As a result,

$$\text{tr } O_{\text{SE}} \Phi_{\mathcal{L}_n^d}^2(O_{\text{SR}}) \leq \text{tr}(\tilde{\tau}_E^2) \text{tr}(\tilde{\rho}_R^2) + \text{tr}(\tilde{\tau}_{\text{SE}}^2) \text{tr}(\tilde{\rho}_{\text{SR}}^2) + \left(\frac{\varepsilon}{n}\right)^{\log \frac{1}{2\eta\rho_m + \delta}} \max_{A \subseteq [n]} (\text{tr}_{\text{AR}}(\text{tr}_{\tilde{A}} \tilde{\rho}_{\text{SR}})^2 \text{tr}_{\text{AE}}(\text{tr}_{\tilde{A}} \tilde{\tau}_{\text{SE}})^2), \quad (\text{C75})$$

$$\mathbb{E}_U \|\mathcal{T}_{S \rightarrow E}(U_S \rho_{\text{SR}} U_S^\dagger) - \tau_E \otimes \rho_R\|_1 \leq \sqrt{\text{tr}(\tilde{\tau}_{\text{SE}}^2) \text{tr}(\tilde{\rho}_{\text{SR}}^2) + \left(\frac{\varepsilon}{n}\right)^{\log \frac{1}{2\eta\rho_m + \delta}} \max_{A \subseteq [n]} (\text{tr}_{\text{AR}}(\text{tr}_{\tilde{A}} \tilde{\rho}_{\text{SR}})^2 \text{tr}_{\text{AE}}(\text{tr}_{\tilde{A}} \tilde{\tau}_{\text{SE}})^2)}. \quad (\text{C76})$$

Note that $\tilde{\rho}_{\text{SR}} = \bigotimes_{i=1}^{2N} \tilde{\rho}_{S_i R_i}$ and $\tilde{\rho}_{S_i R_i} = \zeta_{R_i}^{-1/4} \rho_{S_i R_i} \zeta_{R_i}^{-1/4}$. We can choose $\zeta_{R_i} = \rho_{R_i}$ or $\zeta_R = \rho_R$ to let

$$\text{tr}(\tilde{\rho}_{R_i}^2) = 1, \text{tr}(\tilde{\rho}_{S_i R_i}^2) = 2^{-H_2(S_i|R_i) \rho_{S_i R_i}}. \quad (\text{C77})$$

Thus,

$$\text{tr}_{\text{AR}}(\text{tr}_{\tilde{A}} \tilde{\rho}_{\text{SR}})^2 = \text{tr}_{\text{AR}}(\tilde{\rho}_{\text{AR}}^2) = 2^{-H_2(A|R) \rho_{\text{AR}}} \quad (\text{C78})$$

and

$$\rho_m = \max_{i \in [n]} (2^{H_2(S_i|R_i) \rho_{S_i R_i}}, 2^{-H_2(S_i|R_i) \rho_{S_i R_i}}). \quad (\text{C79})$$

Also, we can choose $\sigma_E = \tau_E$ such that

$$\text{tr}(\tilde{\tau}_E^2) = 1, \quad \text{tr}(\tilde{\tau}_{\text{SE}}^2) = 2^{-H_2(S|E) \tau_{\text{SE}}}. \quad (\text{C80})$$

In this case,

$$\text{tr}_{\text{AE}}(\text{tr}_{\tilde{A}} \tilde{\tau}_{\text{SE}})^2 = \text{tr}_{\text{AE}}(\tilde{\tau}_{\text{AE}}^2) = 2^{-H_2(A|E) \tau_{\text{AE}}}. \quad (\text{C81})$$

Then, for any $\delta > 0$, we have that

$$\mathbb{E}_U \|\mathcal{T}_{S \rightarrow E}(U_S \rho_{\text{SR}} U_S^\dagger) - \tau_E \otimes \rho_R\|_1 \leq \sqrt{2^{-H_2(S|E) \tau_{\text{SE}} - H_2(S|R) \rho_{\text{SR}}} + \left(\frac{\varepsilon}{n}\right)^{\log \frac{1}{2\eta\rho_m + \delta}} \max_{A \subseteq [n]} 2^{-H_2(A|R) \rho_{\text{AR}} - H_2(A|E) \tau_{\text{AE}}}}. \quad (\text{C82})$$

In the description of Theorem 2, we omit δ for simplicity. Proof is completed.

3. Application of Theorem 2 in quantum error correction

Here, we present the results of applying Theorem 2 in quantum error correction and show that the encoding scheme induced by $\mathfrak{B}$ can have a constant threshold against local noise.

For the quantum error correction task, there are nR physical qudits in total. We first fix the encoding rate $R = k/n$, where R is a constant. Then, we set the local dimension q as $2^{R^{-1}}$ composed of R^{-1} qubits. The total qudit number is correspondingly changed from n to nR . Since R is a constant, the local dimension is still a constant. Then, each S_i is a qudit with local dimension $2^{R^{-1}}$, and R_i is a qubit. Without loss of generality, the state $\rho_{S_i R_i}$ is set as $|\hat{\phi}\rangle\langle\hat{\phi}| \otimes |0\rangle\langle 0|^{R^{-1}-1}$ where $|\hat{\phi}\rangle\langle\hat{\phi}|$ is an EPR pair. Then, $H_2(S_i|R_i) = -1$, $\eta = 2^{R^{-1}}/(2^{2R^{-1}} + 1)$, $\rho_m = 2$, and

$$\frac{1}{2\eta\rho_m} = \frac{2^{2R^{-1}} + 1}{2^{R^{-1}+2}} = \frac{2^{2n/k} + 1}{2^{n/k+2}} > 2^{n/k-2}. \quad (\text{C83})$$

Meanwhile, we consider local noise. The channel $\mathcal{T}_{S \rightarrow E}$ will be the complementary channel of local noise, $\mathcal{T}_{S \rightarrow E} = \hat{\mathcal{N}}^{\otimes n}$. In this case, the decoupling theorem can be simplified to

$$\begin{aligned} & \mathbb{E}_{U_S \sim \mathfrak{B}_n^\varepsilon} \|\mathcal{T}_{S \rightarrow E}(U_S \rho_{SR} U_S^\dagger) - \tau_E \otimes \rho_R\|_1 \\ & \leq \sqrt{2^{-\sum_{i=1}^{nR} H_2(S_i|E_i)_{\tau_{S_i E_i}} - \sum_{i=1}^{nR} H_2(S_i|R_i)_{\rho_{S_i R_i}}} + \left(\frac{\varepsilon}{nR}\right)^{\log \frac{1}{2\eta\rho_m}} \prod_{i=1}^{nR} \max(1, 2^{-H_2(S_i|E_i)_{\tau_{S_i E_i}} - H_2(S_i|R_i)_{\rho_{S_i R_i}})}, \end{aligned} \quad (\text{C84})$$

where $H_2(S_i|E_i)_{\tau_{S_i E_i}} = R^{-1}(1 - f(\vec{p}))$ with $f(\vec{p}) = 2 \log(\sqrt{p_I} + \sqrt{p_X} + \sqrt{p_Y} + \sqrt{p_Z})$ for strength- $\vec{p}$ local Pauli noise, $H_2(S_i|E_i)_{\tau_{S_i E_i}} = R^{-1}(1 - \log(1 + 3p))$ for strength- p local erasure error, and $H_2(S_i|E_i)_{\tau_{S_i E_i}} = -R^{-1} \log(1/(2-p) + \sqrt{p/(2-p)})$ for strength- p local amplitude damping noise.

For local Pauli noise,

$$\begin{aligned} & H_2(S_i|E_i)_{\tau_{S_i E_i}} + H_2(S_i|R_i)_{\rho_{S_i R_i}} \\ & = R^{-1}(1 - f(\vec{p})) - 1 = R^{-1} \left(1 - f(\vec{p}) - \frac{k}{n}\right). \end{aligned} \quad (\text{C85})$$

When $1 - f(\vec{p}) - k/n > 0$, the term $\max(1, 2^{-H_2(S_i|E_i)_{\tau_{S_i E_i}} - H_2(S_i|R_i)_{\rho_{S_i R_i}}}) \leq 1$. It is also true for local erasure error when $1 - \log(1 + 3p) - k/n > 0$ and for local amplitude damping noise when $-\log(1/(2-p) + \sqrt{p/(2-p)}) - k/n > 0$.

Note that the expected Choi error satisfies

$$\mathbb{E}_{U_S \sim \mathfrak{B}_n^\varepsilon} \epsilon_{\text{Choi}} \leq \sqrt{\mathbb{E}_{U_S \sim \mathfrak{B}_n^\varepsilon} \|\mathcal{T}_{S \rightarrow E}(U_S \rho_{SR} U_S^\dagger) - \tau_E \otimes \rho_R\|_1}. \quad (\text{C86})$$

We obtain that the Choi error is upper-bounded by

$$\mathbb{E}_{U_S \sim \mathfrak{B}_n^\varepsilon} \epsilon_{\text{Choi}} \leq \left(2^{-n(1-f(\vec{p})-\frac{k}{n})} + \left(\frac{\varepsilon}{k}\right)^{\frac{n}{k}-2}\right)^{\frac{1}{4}} \quad (\text{C87})$$

for strength- $\vec{p}$ local Pauli noise, upper-bounded by

$$\mathbb{E}_{U_S \sim \mathfrak{B}_n^\varepsilon} \epsilon_{\text{Choi}} \leq \left(2^{-n(1-\log(1+3p)-\frac{k}{n})} + \left(\frac{\varepsilon}{k}\right)^{\frac{n}{k}-2}\right)^{\frac{1}{4}} \quad (\text{C88})$$

for strength- p local erasure error, and upper-bounded by

$$\mathbb{E}_{U_S \sim \mathfrak{B}_n^\varepsilon} \epsilon_{\text{Choi}} \leq \left(2^{-n(-\log(\frac{1}{2-p} + \sqrt{\frac{p}{2-p}}) - \frac{k}{n})} + \left(\frac{\varepsilon}{k}\right)^{\frac{n}{k}-2}\right)^{\frac{1}{4}} \quad (\text{C89})$$

for strength- p local amplitude damping noise.

To let the Choi error vanish, we need to ensure $n/k - 2 > 0$. Thus, we show that there exists a threshold $\min(\frac{1}{2}, 1 - f(\vec{p}))$ for local Pauli noise, a threshold $\min(\frac{1}{2}, 1 - \log(1 + 3p))$ for local erasure error, and a threshold $\min(\frac{1}{2}, -\log(1/(2-p) + \sqrt{p/(2-p)}))$ for local amplitude damping noise. Moreover, by choosing appropriate parameters, $(\varepsilon/k)^{\frac{n}{k}-2}$ will polynomially decay with n .

For strength- p nearest-neighbor ZZ-coupling noise, we apply the original Theorem 2. Similar to the analysis in

Appendix B 9, we obtain that

$$H_2(A|R)_{\rho_{AR}} = -|A|, \quad (C90)$$

$$H_2(A|E)_{\tau_{AE}} \geq R^{-1} |A| \left[1 - 2(1+R) \log(\sqrt{1-p} + \sqrt{p}) \right]. \quad (C91)$$

Thus, when $H_2(A|R) + H_2(A|E) \geq 0$, or equivalently, $1 - R - 2(1+R) \log(\sqrt{1-p} + \sqrt{p}) \geq 0$, we have that

$$\begin{aligned} & \mathbb{E}_{U_S \sim \mathfrak{B}_n^\varepsilon} \epsilon_{\text{Choi}} \\ & \leq \left(2^{-n(1-\frac{k}{n}-2(1+\frac{k}{n}) \log(\sqrt{1-p} + \sqrt{p}))} + \left(\frac{\varepsilon}{k} \right)^{\frac{n}{k}-2} \right)^{\frac{1}{4}}. \end{aligned} \quad (C92)$$

Hence, the achievable encoding rate for $\mathfrak{B}_n^\varepsilon$ is $1 - 2(1 + k/n) \log(\sqrt{1-p} + \sqrt{p})$.

Note that the above analysis also applies to the case of local dimension 2. In this case, the total number of qubits is n , and each state $\rho_{S_i R_i}$ is defined as $(k/n) |\hat{\phi}\rangle\langle\hat{\phi}| + (1 - k/n) |0\rangle\langle 0|$. We then have $H_2(S_i|R_i) = -k/n$, while $H_2(S_i|E_i)$ depends on the noise parameter, typically taking the form $1 - f(\vec{p})$ for local Pauli noise. With $\eta = \frac{2}{5}$ and $\rho_m = 2^{\frac{k}{n}}$, substituting these values into the decoupling theorem yields the same bound, except that the factor $(\varepsilon/k)^{n/k-2}$ is replaced by $(\varepsilon/n)^{\log \frac{5}{4} - k/n}$.

APPENDIX D: LOWER BOUND FOR CIRCUIT DEPTH IN AQEC

Here, we analyze the lower bound on the circuit depth required for a unitary encoding of k logical qubits that achieves a Choi error ϵ_{Choi} . For the $2k$ -qubit maximally entangled state $\hat{\phi}_{LR}$ shared between the reference system R and the logical system L , where both L and R consist of k qubits each, we denote the i th qubit in L by $L_{i,0}$ and the qubit in R that is maximally entangled with $L_{i,0}$ by $R_{i,0}$. We define the physical system S as the union of L and the ancillary system initialized in the state $|0^{n-k}\rangle$.

Consider a unitary circuit U of depth d acting on S , decomposed as

$$U = \prod_{i=1}^d \left(\bigotimes_j V_{i,j} \right), \quad (D1)$$

where each $\{V_{i,j}\}_j$ represents the set of two-qubit gates in the i th layer of U with disjoint supports. For each layer j , where $1 \leq j \leq d$, let $L_{i,j}$ denote the union of the supports of all two-qubit gates that intersect with $L_{i,j-1}$,

$$L_{i,j} = \bigcup \{ \text{supp}(V_{i,j}) \mid \text{supp}(V_{i,j}) \cap L_{i,j-1} \neq \emptyset \}. \quad (D2)$$

In other words, $L_{i,j}$ represents the light cone of the i th qubit in the first j layers of U .

Similarly, we define the backward light cone of the i th physical qubit, $i \in S$, recursively as

$$\begin{aligned} B_{i,d} &= \{i\}, \\ B_{i,j} &= \bigcup \{ \text{supp}(V_{i,j+1}) \mid \text{supp}(V_{i,j+1}) \cap B_{i,j+1} \neq \emptyset \}, \\ 0 &\leq j < d. \end{aligned} \quad (D3)$$

By the definitions of $L_{i,j}$ and $B_{i,j}$, for any logical qubit $i \in L$ and physical qubit $k \in S$, the qubit k is included in $L_{i,d}$ if and only if the logical qubit i is included in $B_{k,0}$.

Finally, we define M as an upper bound on the size of all light cones,

$$M = \max \left(\max_{i \in L} |L_{i,d}|, \max_{i \in S} |B_{i,0}| \right). \quad (D4)$$

With this definition, we can establish a lower bound on the number of physical qubits in L that have disjoint light cones, as stated in the following lemma.

Lemma D1 (Disjoint light cones). For any unitary circuit $U = \prod_{i=1}^d \left(\bigotimes_j V_{i,j} \right)$, there exists a subset $J \subseteq L$ such that:

- (1) For any distinct qubits $i, j \in J$, their light cones are disjoint: $L_{i,d} \cap L_{j,d} = \emptyset$.
- (2) The size of J satisfies $|J| \geq \lceil k/M^2 \rceil$.

Proof. We construct the set J iteratively. Start with an empty set $J_0 = \emptyset$. At each step a , assume we have a set J_a with disjoint light cones. Define $L_{J_a,d} = \bigcup_{i \in J_a} L_{i,d}$, the union of the light cones of qubits in J_a .

For a qubit $j \in L$, the condition that $L_{j,d}$ does not intersect with any $L_{i,d}$ for $i \in J_a$,

$$L_{j,d} \cap L_{J_a,d} = \emptyset, \quad (D5)$$

is equivalent to

$$j \notin \bigcup_{m \in L_{J_a,d}} B_{m,0}. \quad (D6)$$

Since each $L_{i,d}$ has at most M qubits, the size of $L_{J_a,d}$ is at most $|J_a|M$. Consequently, the set $\bigcup_{m \in L_{J_a,d}} B_{m,0}$ has size at most $|J_a|M^2$.

If $k > |J_a|M^2$, there exists at least one logical qubit $i \in L - J_a$ such that $i \notin \bigcup_{m \in L_{J_a,d}} B_{m,0}$. We can then add this qubit to J_a to form $J_{a+1} = J_a \cup \{i\}$, maintaining the property that the light cones of qubits in J_{a+1} remain disjoint.

This process continues until $k \leq |J_a|M^2 = aM^2$. Therefore, the final set J_a satisfies that the light cones of any two distinct qubits in J are disjoint, and $|J_a| = a \geq k/M^2$. ■

Proof of Theorem 4. We now combine Lemma D1 with the properties of depolarizing noise to analyze circuit lower bound. Recall that the Choi fidelity is defined as

$$F_{\text{Choi}} = \max_{\mathcal{D}} F(\hat{\phi}_{\text{LR}}, (\mathcal{D} \circ \mathcal{N} \circ \mathcal{U})_L(\hat{\phi}_{\text{LR}})). \quad (\text{D7})$$

Here, and in the following, we omit the identity channel I_R on the reference system. The encoding channel $\mathcal{U}$ maps a state ρ on system L to another state $\mathcal{U}(\rho) = U(\rho \otimes |0^{n-k}\rangle\langle 0^{n-k}|)U^\dagger$ on system S . The noise channel $\mathcal{N}$ represents i.i.d. local depolarizing noise on system S , specifically $\mathcal{N} = \mathcal{N}_d^S$, where $\mathcal{N}_d^S = \bigotimes_{i \in S} \mathcal{N}_d^{(i)}$ and $\mathcal{N}_d^{(i)}$ is the strength- p depolarizing channel acting on the i th qubit. The recovery channel $\mathcal{D}$ maps a state on system S back to a state on system L .

Consider the subset J obtained in Lemma D1 under the strength- p local depolarizing noise defined in Eq. (36). For a single qubit $i \in J$, the probability that all physical qubits in $L_{i,d}$ are traced out and replaced by the maximally mixed state is $p_L = p^{|L_{i,d}|} \geq p^M$. Since the light cones of logical qubits in J are disjoint, the events that these light cones are traced out are independent. Therefore, the probability that none of these light cones are fully traced out is

$$p_{\text{None}} = \prod_{i \in J} (1 - p^{|L_{i,d}|}) \leq (1 - p^M)^{|J|} \leq \exp(-|J| \cdot p^M). \quad (\text{D8})$$

The second inequality utilizes the bound $(1 - x) \leq e^{-x}$ for $0 \leq x \leq 1$. This probability provides a lower bound for the Choi error, as later obtained in Eq. (D17). To see this, consider the Choi fidelity under the local depolarizing channel,

$$F_{\text{Choi}} = \max_{\mathcal{D}} F\left(\hat{\phi}_{\text{LR}}, \mathcal{D}_{S \rightarrow L} \circ \mathcal{N}_d^S \left[U_d \left(\hat{\phi}_{\text{LR}} \otimes |0^{n-k}\rangle\langle 0^{n-k}| \right) U_d^\dagger \right] \right). \quad (\text{D9})$$

Expanding the state after passing through the noise channel $\mathcal{N}_d^S$, with probability $1 - p_{\text{None}}$, the noise erases all information in at least one of the $L_{i,d}$ systems for $i \in J$, replacing it with a maximally mixed state. Consequently, the state after applying the local depolarizing noise is

$$\mathcal{N}_d^S \left[U_d \left(\hat{\phi}_{\text{LR}} \otimes |0^{n-k}\rangle\langle 0^{n-k}| \right) U_d^\dagger \right] = \sum_l p_l \psi_l + p_{\text{None}} \psi', \quad (\text{D10})$$

where $\sum_l p_l = 1 - p_{\text{None}}$ represents the probability that at least one logical qubit's light cone is traced out. Here, the indices l are used to distinguish states subject to errors on different subsets of physical qubits. For each ψ_l , at least one subsystem $L_{i,d}$ with $i \in J$ is replaced by the maximally mixed state.

The square of the Choi fidelity can then be expressed as

$$\begin{aligned} F_{\text{Choi}}^2 &= \max_{\mathcal{D}} F^2 \left(\hat{\phi}_{\text{LR}}, \mathcal{D}_{S \rightarrow L} \circ \mathcal{N}_d^S \left[U_d \left(\hat{\phi}_{\text{LR}} \otimes |0^{n-k}\rangle\langle 0^{n-k}| \right) U_d^\dagger \right] \right) \\ &= \max_{\mathcal{D}} F^2 \left(\hat{\phi}_{\text{LR}}, \mathcal{D}_{S \rightarrow L} \left[\sum_l p_l \psi_l + p_{\text{None}} \psi' \right] \right) \\ &= \max_{\mathcal{D}} \left\langle \hat{\phi}_{\text{LR}} \left| \mathcal{D}_{S \rightarrow L} \left[\sum_l p_l \psi_l + p_{\text{None}} \psi' \right] \right| \hat{\phi}_{\text{LR}} \right\rangle \\ &= \max_{\mathcal{D}} \left(\sum_l p_l \left\langle \hat{\phi}_{\text{LR}} \left| \mathcal{D}_{S \rightarrow L}(\psi_l) \right| \hat{\phi}_{\text{LR}} \right\rangle + p_{\text{None}} \left\langle \hat{\phi}_{\text{LR}} \left| \mathcal{D}_{S \rightarrow L}(\psi') \right| \hat{\phi}_{\text{LR}} \right\rangle \right) \\ &\leq \max_{\mathcal{D}} \left(\sum_l p_l \left\langle \hat{\phi}_{\text{LR}} \left| \mathcal{D}_{S \rightarrow L}(\psi_l) \right| \hat{\phi}_{\text{LR}} \right\rangle + p_{\text{None}} \right). \end{aligned} \quad (\text{D11})$$

In the third equality, we utilize Eq. (13). The fourth equality follows from the linearity of quantum channels.

Now, consider the quantity $\langle \hat{\phi}_{\text{LR}} | \mathcal{D}_{S \rightarrow L}(\psi_l) | \hat{\phi}_{\text{LR}} \rangle$. Since at least the light cone of one logical qubit in L is fully traced out, we have

$$\psi_l = \text{Tr}_T \left[U \left(\hat{\phi}_{\text{LR}} \otimes |0^{n-k}\rangle\langle 0^{n-k}| \right) U^\dagger \right] \otimes \frac{\mathbb{I}_T}{2^{|T|}}, \quad (\text{D12})$$

for some subset $T \subseteq S$ such that $L_{i,d} \subseteq T$ for some $i \in L$. Crucially, when the light cone $L_{i,d}$ is traced out, the unitaries applied within $L_{i,d}$ are canceled. Therefore, we can replace the initial state $\hat{\phi}_{\text{LR}}$ with $\hat{\phi}_{L-L_{i,0}, R-R_{i,0}} \otimes \mathbb{I}_{R_{i,0}}/2 \otimes \mathbb{I}_{L_{i,0}}/2$. Specifically,

$$\text{tr}_{L_{i,d}} \left[U \left(\hat{\phi}_{\text{LR}} \otimes |0^{n-k}\rangle\langle 0^{n-k}| \right) U^\dagger \right] = \text{tr}_{L_{i,d}} \left[U \left(\hat{\phi}_{L-L_{i,0}, R-R_{i,0}} \otimes \frac{\mathbb{I}_{L_{i,0}}}{2} \otimes |0^{n-k}\rangle\langle 0^{n-k}| \right) U^\dagger \right] \otimes \frac{\mathbb{I}_{R_{i,0}}}{2}. \quad (\text{D13})$$

Thus, the state ψ_l in Eq. (D11) can be written as

$$\psi_l = \hat{\psi}_l \otimes \frac{\mathbb{I}_T}{2^{|T|}} \otimes \frac{\mathbb{I}_{R_0}}{2}. \quad (\text{D14})$$

where $\hat{\psi}_l = \text{tr}_{T \cup R_{i,0}}(\psi_l)$. We then have

$$\left\langle \hat{\phi}_{\text{LR}} \left| \mathcal{D}_{S \rightarrow L} \left(\hat{\psi}_l \otimes \frac{\mathbb{I}_T}{2^{|T|}} \right) \otimes \frac{\mathbb{I}_{R_0}}{2} \right| \hat{\phi}_{\text{LR}} \right\rangle = \max_{\mathcal{D}} \frac{1}{4} \left\langle \hat{\phi}_{L-L_{i,0}, R-R_{i,0}} \left| \mathcal{D}_{S \rightarrow L} \left(\hat{\psi} \otimes \frac{\mathbb{I}_T}{2^{|T|}} \right) \right| \hat{\phi}_{L-L_{i,0}, R-R_{i,0}} \right\rangle \leq \frac{1}{4}. \quad (\text{D15})$$

Therefore, the Choi fidelity in Eq. (D11) satisfies

$$\begin{aligned} F_{\text{Choi}}^2 &\leq \max_{\mathcal{D}} \left(\sum_l p_l \langle \hat{\phi}_{\text{LR}} | \mathcal{D}_{S \rightarrow L}(\psi_l) | \hat{\phi}_{\text{LR}} \rangle + p_{\text{None}} \right) \\ &\leq \sum_l \frac{p_l}{4} + p_{\text{None}} = \frac{1}{4} + \frac{3}{4} p_{\text{None}} \\ &\leq \frac{1}{4} + \frac{3}{4} \exp(-|J| \cdot p^M), \end{aligned} \quad (\text{D16})$$

where the last inequality follows from Eq. (D8). Recall that $1 - \epsilon_{\text{Choi}}^2 = F_{\text{Choi}}^2$. For $\epsilon_{\text{Choi}} < 0.1$, we have $\exp(-|J| \cdot p^M) > 0.98$. In this regime, using the inequality $e^{-x} \leq 1 - x/2$, we obtain

$$\begin{aligned} F_{\text{Choi}}^2 &\leq \frac{1}{4} + \frac{3}{4} \left(1 - \frac{|J|}{2} p^M \right) = 1 - \frac{3|J|}{8} p^M, \\ \epsilon_{\text{Choi}}^2 &\geq \frac{3|J|}{8} p^M. \end{aligned} \quad (\text{D17})$$

Taking the logarithm of both sides, we obtain

$$M \log \left(\frac{1}{p} \right) - \log(|J|) \geq \log \left(\frac{3}{8\epsilon_{\text{Choi}}^2} \right). \quad (\text{D18})$$

According to Lemma D1, $|J| \geq \frac{k}{M^2}$. Therefore, we have

$$M \log \left(\frac{1}{p} \right) + 2 \log M \geq \log \left(\frac{3}{8\epsilon_{\text{Choi}}^2} \right) + \log k. \quad (\text{D19})$$

For a D -dimensional quantum circuit, the size of the light cones can be upper-bounded by $M \leq (2d)^D$, since the length of each edge of the light-cone hypercube is upper-bounded by $2d$. Substituting this bound into the previous

inequality, we obtain

$$(2d)^D \log \left(\frac{1}{p} \right) + 2D \log(2d) \geq \log \left(\frac{3}{8\epsilon_{\text{Choi}}^2} \right) + \log k. \quad (\text{D20})$$

In the case of an all-to-all circuit, the size of the light cones is upper-bounded by $M \leq 2^d$. Therefore, the inequality becomes

$$2^d \log \left(\frac{1}{p} \right) + 2d \geq \log \left(\frac{3}{8\epsilon_{\text{Choi}}^2} \right) + \log k. \quad (\text{D21})$$

■

-
- [1] L. Egan, D. M. Debroy, C. Noel, A. Risinger, D. Zhu, D. Biswas, M. Newman, M. Li, K. R. Brown, M. Cetina, and C. Monroe, Fault-tolerant control of an error-corrected qubit, *Nature* **598**, 281 (2021).
 - [2] M. Gong *et al.*, Experimental exploration of five-qubit quantum error-correcting code with superconducting qubits, *Natl. Sci. Rev.* **9**, nwab011 (2021).
 - [3] L. Postler, S. Heußen, I. Pogorelov, M. Rispler, T. Feldker, M. Meth, C. D. Marciniak, R. Stricker, M. Ringbauer, R. Blatt, P. Schindler, M. Müller, and T. Monz, Demonstration of fault-tolerant universal quantum gate operations, *Nature* **605**, 675 (2022).
 - [4] A. Leverrier and G. Zémor, in *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)* (Institute of Electrical and Electronics Engineers (IEEE), Piscataway, New Jersey, United States, 2022), pp. 872–883.
 - [5] Y. Zhao *et al.*, Realization of an error-correcting surface code with superconducting qubits, *Phys. Rev. Lett.* **129**, 030501 (2022).
 - [6] R. Acharya *et al.*, Suppressing quantum errors by scaling a surface code logical qubit, *Nature* **614**, 676 (2023).
 - [7] R. Acharya *et al.*, Quantum error correction below the surface code threshold, *Nature* **638**, 920 (2024).

- [8] D. Aharonov and M. Ben-Or, in *Proceedings of the Twenty-Ninth Annual ACM Symposium on Theory of Computing*, STOC '97 (Association for Computing Machinery, New York, NY, USA, 1997), pp. 176–188.
- [9] D. Aharonov and M. Ben-Or, Fault-tolerant quantum computation with constant error rate, *SIAM J. Comput.* **38**, 1207 (2008).
- [10] D. Gottesman, An introduction to quantum error correction and fault-tolerant quantum computation, [arXiv:0904.2557](#) [quant-ph].
- [11] D. Gottesman, Fault-tolerant quantum computation with constant overhead, *Quantum Inf. Comput.* **14**, 1338 (2014).
- [12] Q. Xu, P. Zeng, D. Xu, and L. Jiang, Fault-tolerant operation of bosonic qubits with discrete-variable ancillae, *Phys. Rev. X* **14**, 031016 (2024).
- [13] J. Nelson, G. Bentsen, S. T. Flammia, and M. J. Gullans, Fault-tolerant quantum memory using low-depth random circuit codes, *Phys. Rev. Res.* **7**, 013040 (2025).
- [14] A. G. Fowler, M. Mariantoni, J. M. Martinis, and A. N. Cleland, Surface codes: Towards practical large-scale quantum computation, *Phys. Rev. A* **86**, 032324 (2012).
- [15] E. Dennis, A. Kitaev, A. Landahl, and J. Preskill, Topological quantum memory, *J. Math. Phys.* **43**, 4452 (2002).
- [16] A. A. Kovalev and L. P. Pryadko, Fault tolerance of quantum low-density parity check codes with sublinear distance scaling, *Phys. Rev. A* **87**, 020304(R) (2013).
- [17] O. Fawzi, A. Grospellier, and A. Leverrier, Constant overhead quantum fault tolerance with quantum expander codes, *Commun. ACM* **64**, 106 (2020).
- [18] P. Panteleev and G. Kalachev, in *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2022 (Association for Computing Machinery, New York, NY, USA, 2022), pp. 375–388.
- [19] D. Gottesman, *Stabilizer Codes and Quantum Error Correction* (California Institute of Technology, Pasadena, California, United States, 1997).
- [20] W. Brown and O. Fawzi, in *2013 IEEE International Symposium on Information Theory* (Institute of Electrical and Electronics Engineers (IEEE), Piscataway, New Jersey, United States, 2013), pp. 346–350.
- [21] M. J. Gullans, S. Krastanov, D. A. Huse, L. Jiang, and S. T. Flammia, Quantum coding with low-depth random circuits, *Phys. Rev. X* **11**, 031066 (2021).
- [22] L. Kong and Z.-W. Liu, Near-optimal covariant quantum error-correcting codes from random unitaries with symmetries, *PRX Quantum* **3**, 020314 (2022).
- [23] A. S. Darmawan, Y. Nakata, S. Tamiya, and H. Yamasaki, Low-depth random Clifford circuits for quantum coding against Pauli noise using a tensor-network decoder, *Phys. Rev. Res.* **6**, 023055 (2024).
- [24] D. W. Leung, M. A. Nielsen, I. L. Chuang, and Y. Yamamoto, Approximate quantum error correction can lead to better codes, *Phys. Rev. A* **56**, 2567 (1997).
- [25] C. Crépeau, D. Gottesman, and A. Smith, in *Advances in Cryptology—EUROCRYPT 2005*, edited by R. Cramer (Springer, Berlin, 2005), pp. 285–301.
- [26] R. Klesse, Approximate quantum error correction, random codes, and quantum channel capacity, *Phys. Rev. A* **75**, 062315 (2007).
- [27] C. Bény and O. Oreshkov, General conditions for approximate quantum error correction and near-optimal recovery channels, *Phys. Rev. Lett.* **104**, 120501 (2010).
- [28] M. M. Wilde, *Quantum Information Theory* (Cambridge University, Cambridge, United Kingdom, 2013).
- [29] C. Crépeau, D. Gottesman, and A. Smith, in *Preliminary Version in Proceedings of “Advances in Cryptology—EUROCRYPT 2005”* (Springer-Verlag, Berlin, Heidelberg, Germany, 2005).
- [30] J. Yi, W. Ye, D. Gottesman, and Z.-W. Liu, Complexity and order in approximate quantum error-correcting codes, *Nat. Phys.* **20**, 1798 (2024).
- [31] F. G. S. L. Brandão, A. W. Harrow, and M. Horodecki, Local random quantum circuits are approximate polynomial-designs, *Commun. Math. Phys.* **346**, 397 (2016).
- [32] A. M. Dalzell, N. Hunter-Jones, and F. G. S. L. Brandão, Random quantum circuits anticentralize in log depth, *PRX Quantum* **3**, 010333 (2022).
- [33] T. Schuster, J. Haferkamp, and H.-Y. Huang, Random unitaries in extremely low depth, *Science* **389**, 92 (2025).
- [34] N. LaRacuente and F. Leditzky, Approximate unitary k -designs from shallow, low-communication circuits, [arXiv:2407.07876](#).
- [35] D. P. DiVincenzo, P. W. Shor, and J. A. Smolin, Quantum-channel capacity of very noisy channels, *Phys. Rev. A* **57**, 830 (1998).
- [36] J. P. Bonilla Ataides, D. K. Tuckett, S. D. Bartlett, S. T. Flammia, and B. J. Brown, The XZZX surface code, *Nat. Commun.* **12**, 2172 (2021).
- [37] W. Brown and O. Fawzi, Scrambling speed of random quantum circuits, [arXiv:1210.6644](#) [quant-ph].
- [38] N. Baspin, O. Fawzi, and A. Shayeghi, A lower bound on the overhead of quantum error correction in low dimensions, [arXiv:2302.04317](#) [quant-ph].
- [39] S. Choi, Y. Bao, X.-L. Qi, and E. Altman, Quantum error correction in scrambling dynamics and measurement-induced phase transition, *Phys. Rev. Lett.* **125**, 030505 (2020).
- [40] N. J. Cerf and R. Cleve, Information-theoretic interpretation of quantum error-correcting codes, *Phys. Rev. A* **56**, 1721 (1997).
- [41] F. Arute *et al.*, Quantum supremacy using a programmable superconducting processor, *Nature* **574**, 505 (2019).
- [42] A. Bouland, B. Fefferman, C. Nirkhe, and U. Vazirani, On the complexity and verification of quantum random circuit sampling, *Nat. Phys.* **15**, 159 (2019).
- [43] Y. Wu *et al.*, Strong quantum computational advantage using a superconducting quantum processor, *Phys. Rev. Lett.* **127**, 180501 (2021).
- [44] R. Movassagh, The hardness of random quantum circuits, *Nat. Phys.* **19**, 1719 (2023).
- [45] A. Morvan *et al.*, Phase transitions in random circuit sampling, *Nature* **634**, 328 (2024).
- [46] D. Gao *et al.*, Establishing a new benchmark in quantum computational advantage with 105-qubit Zuchongzhi 3.0 processor, *Phys. Rev. Lett.* **134**, 090601 (2025).
- [47] A. Elben, S. T. Flammia, H.-Y. Huang, R. Kueng, J. Preskill, B. Vermersch, and P. Zoller, The randomized measurement toolbox, *Nat. Rev. Phys.* **5**, 9 (2023).

- [48] H.-Y. Huang, R. Kueng, and J. Preskill, Predicting many properties of a quantum system from very few measurements, *Nat. Phys.* **16**, 1050 (2020).
- [49] S. Chen, W. Yu, P. Zeng, and S. T. Flammia, Robust shadow estimation, *PRX Quantum* **2**, 030348 (2021).
- [50] Q. Liu, Z. Li, X. Yuan, H. Zhu, and Y. Zhou, Auxiliary-free replica shadow estimation, [arXiv:2407.20865](#) [quant-ph].
- [51] E. Knill, D. Leibfried, R. Reichle, J. Britton, R. B. Blakestad, J. D. Jost, C. Langer, R. Ozeri, S. Seidelin, and D. J. Wineland, Randomized benchmarking of quantum gates, *Phys. Rev. A* **77**, 012307 (2008).
- [52] E. Magesan, J. M. Gambetta, and J. Emerson, Scalable and robust randomized benchmarking of quantum processes, *Phys. Rev. Lett.* **106**, 180504 (2011).
- [53] G. Liu, Z. Xie, Z. Xu, and X. Ma, Group twirling and noise tailoring for multiqubit controlled phase gates, *Phys. Rev. Res.* **6**, 043221 (2024).
- [54] P. Hayden and J. Preskill, Black holes as mirrors: Quantum information in random subsystems, *J. High Energy Phys.* **2007**, 120 (2007).
- [55] A. Almheiri, X. Dong, and D. Harlow, Bulk locality and quantum error correction in ADS/CFT, *J. High Energy Phys.* **2015**, 163 (2015).
- [56] F. Pastawski, B. Yoshida, D. Harlow, and J. Preskill, Holographic quantum error-correcting codes: Toy models for the bulk/boundary correspondence, *J. High Energy Phys.* **2015**, 149 (2015).
- [57] A. Nahum, J. Ruhman, S. Vijay, and J. Haah, Quantum entanglement growth under random unitary dynamics, *Phys. Rev. X* **7**, 031016 (2017).
- [58] A. Nahum, J. Ruhman, and D. A. Huse, Dynamics of entanglement and transport in one-dimensional systems with quenched randomness, *Phys. Rev. B* **98**, 035118 (2018).
- [59] T. Zhou and A. Nahum, Emergent statistical mechanics of entanglement in random unitary circuits, *Phys. Rev. B* **99**, 174205 (2019).
- [60] D. Maslov and M. Roetteler, Shorter stabilizer circuits via Bruhat decomposition and quantum circuit transformations, *IEEE Trans. Inf. Theory* **64**, 4729 (2018).
- [61] S. Bravyi and D. Maslov, Hadamard-free circuits expose the structure of the Clifford group, *IEEE Trans. Inf. Theory* **67**, 4546 (2021).
- [62] C. H. Bennett, D. P. DiVincenzo, and J. A. Smolin, Capacities of quantum erasure channels, *Phys. Rev. Lett.* **78**, 3217 (1997).
- [63] F. Dupuis, M. Berta, J. Wullschleger, and R. Renner, One-shot decoupling, *Commun. Math. Phys.* **328**, 251 (2014).
- [64] O. Szehr, F. Dupuis, M. Tomamichel, and R. Renner, Decoupling with unitary approximate two-designs, *New J. Phys.* **15**, 053022 (2013).
- [65] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information: 10th Anniversary Edition* (Cambridge University Press, 2010).
- [66] P. Horodecki, M. Horodecki, and R. Horodecki, General teleportation channel, singlet fraction and quasi-distillation, [arXiv:quant-ph/9807091](#) [quant-ph].
- [67] A. Gilchrist, N. K. Langford, and M. A. Nielsen, Distance measures to compare real and ideal quantum processes, *Phys. Rev. A* **71**, 062310 (2005).
- [68] I. Devetak and P. W. Shor, The capacity of a quantum channel for simultaneous transmission of classical and quantum information, *Commun. Math. Phys.* **256**, 287 (2005).
- [69] C. H. Bennett, D. P. DiVincenzo, J. A. Smolin, and W. K. Wootters, Mixed-state entanglement and quantum error correction, *Phys. Rev. A* **54**, 3824 (1996).
- [70] M. Grassl, T. Beth, and T. Pellizzari, Codes for the quantum erasure channel, *Phys. Rev. A* **56**, 33 (1997).
- [71] V. Giovannetti and R. Fazio, Information-capacity description of spin-chain correlations, *Phys. Rev. A* **71**, 032314 (2005).
- [72] P. Faist, S. Nezami, V. V. Albert, G. Salton, F. Pastawski, P. Hayden, and J. Preskill, Continuous symmetries and approximate quantum error correction, *Phys. Rev. X* **10**, 041018 (2020).
- [73] W. Fulton and J. Harris, *Representation Theory: A First Course* (Springer, New York, NY, 2004).
- [74] C. Dankert, Efficient simulation of random quantum states and operators, [arXiv:quant-ph/0512217](#) [quant-ph].
- [75] D. Gross, K. Audenaert, and J. Eisert, Evenly distributed unitaries: On the structure of unitary designs, *J. Math. Phys.* **48**, 052104 (2007).
- [76] C. Dankert, R. Cleve, J. Emerson, and E. Livine, Exact and approximate unitary 2-designs and their application to fidelity estimation, *Phys. Rev. A* **80**, 012304 (2009).
- [77] Z. Webb, The Clifford group forms a unitary 3-design, *Quantum Inf. Comput.* **16**, 1379 (2016).
- [78] H. Zhu, Multiqubit Clifford groups are unitary 3-designs, *Phys. Rev. A* **96**, 062336 (2017).
- [79] D. DiVincenzo, D. Leung, and B. Terhal, Quantum data hiding, *IEEE Trans. Inf. Theory* **48**, 580 (2002).
- [80] W. Dür, M. Hein, J. I. Cirac, and H.-J. Briegel, Standard forms of noisy quantum operations via depolarization, *Phys. Rev. A* **72**, 052326 (2005).
- [81] B. M. Terhal, Quantum error correction for quantum memories, *Rev. Mod. Phys.* **87**, 307 (2015).
- [82] S. T. Flammia and J. J. Wallman, Efficient estimation of Pauli channels, *ACM Trans. Quantum Comput.* **1**, 3 (2020).
- [83] J. J. Wallman and J. Emerson, Noise tailoring for scalable quantum computation via randomized compiling, *Phys. Rev. A* **94**, 052325 (2016).
- [84] A. Uhlmann, The “transition probability” in the state space of a $\star$ -algebra, *Rep. Math. Phys.* **9**, 273 (1976).
- [85] C. Moore and M. Nilsson, Parallel quantum computation and quantum codes, *SIAM J. Comput.* **31**, 799 (2001).
- [86] J. Jiang, X. Sun, S.-H. Teng, B. Wu, K. Wu, and J. Zhang, in *Proceedings of the 2020 ACM-SIAM Symposium on Discrete Algorithms (SODA)* (Society for Industrial and Applied Mathematics (SIAM), Philadelphia, Pennsylvania, United States), pp. 213–229.
- [87] B. Swingle, Unscrambling the physics of out-of-time-order correlators, *Nat. Phys.* **14**, 988 (2018).
- [88] R. J. Garcia, Y. Zhou, and A. Jaffe, Quantum scrambling with classical shadows, *Phys. Rev. Res.* **3**, 033155 (2021).
- [89] S. Bravyi, D. Poulin, and B. Terhal, Tradeoffs for reliable quantum information storage in 2D systems, *Phys. Rev. Lett.* **104**, 050503 (2010).

- [90] M. Heinrich, J. Haferkamp, I. Roth, and J. Helsen, Anti-concentration is (almost) all you need, [arXiv:2510.23719](#) [quant-ph].
- [91] D. Belkin, J. Allen, and B. K. Clark, Apparent universal behavior in second moments of random quantum circuits, [arXiv:2510.23726](#) [quant-ph].
- [92] M. Berta, M. Christandl, and R. Renner, The quantum reverse Shannon theorem based on one-shot information theory, *Commun. Math. Phys.* **306**, 579 (2011).
- [93] M. Horodecki, J. Oppenheim, and A. Winter, Partial quantum information, *Nature* **436**, 673 (2005).
- [94] F. Dupuis, Privacy amplification and decoupling without smoothing, [arXiv:2105.05342](#) [quant-ph].
- [95] N. Linden, S. Popescu, A. J. Short, and A. Winter, Quantum mechanical evolution towards thermal equilibrium, *Phys. Rev. E* **79**, 061103 (2009).
- [96] L. del Rio, J. Åberg, R. Renner, O. Dahlsten, and V. Vedral, The thermodynamic meaning of negative entropy, *Nature* **474**, 61 (2011).
- [97] F. G. S. L. Brandão and M. Horodecki, Exponential decay of correlations implies area law, *Commun. Math. Phys.* **333**, 761 (2015).
- [98] R. Rubboli, M. M. Goodarzi, and M. Tomamichel, Quantum conditional entropies, [arXiv:2410.21976](#) [quant-ph].
- [99] M. Tomamichel, M. Berta, and M. Hayashi, Relating different quantum generalizations of the conditional Rényi entropy, *J. Math. Phys.* **55**, 082206 (2014).
- [100] M. Müller-Lennert, F. Dupuis, O. Szehr, S. Fehr, and M. Tomamichel, On quantum Rényi entropies: A new generalization and some properties, *J. Math. Phys.* **54**, 122203 (2013).
- [101] S. Beigi, Sandwiched Rényi divergence satisfies data processing inequality, *J. Math. Phys.* **54**, 122202 (2013).
- [102] M. Tomamichel, R. Colbeck, and R. Renner, A fully quantum asymptotic equipartition property, *IEEE Trans. Inf. Theory* **55**, 5840 (2009).
- [103] M. Tomamichel, A framework for non-asymptotic quantum information theory, [arXiv:1203.2142](#) [quant-ph].
- [104] G. Zheng, W. He, G. Lee, and L. Jiang, Near-optimal performance of quantum error correction codes, *Phys. Rev. Lett.* **132**, 250602 (2024).
- [105] H. Barnum and E. Knill, Reversing quantum dynamics with near-optimal quantum and classical fidelity, *J. Math. Phys.* **43**, 2097 (2002).